Extremophiles (1998) 2:379–389

ORIGINAL PAPER

Ralf Thoma · Martin Schwander · Wolfgang Liebl
Kasper Kirschner · Reinhard Sterner

A histidine gene cluster of the hyperthermophile *Thermotoga maritima*: sequence analysis and evolutionary significance

Received: November 3, 1997 / Accepted: March 10, 1998

Abstract The sequences of histidine operon genes in hyperthermophiles are informative for understanding high protein thermostability and the evolution of metabolic pathways. Therefore, a cluster of eight *his* genes from the hyperthermophilic and phylogenetically early bacterium *Thermotoga maritima* was cloned and sequenced. The cluster has the gene order *his*DCBdHAFI–E, lacking only *his*G and *his*Bp, and does not contain intercistronic regions. This compact organization of *his* genes resembles the *his* operon of enterobacteria. Sequence analysis downstream of the stop codon of *his*I–E identifies a region with a significantly higher cytosine over guanosine content, which is indicative of a rho-dependent termination of transcription of the *his* operon. Multiple sequence alignments of N^1-((5′-phosphoribosyl)-formimino)-5-aminoimidazole-4-carboxyamide ribonucleotide isomerase (HisA) and of the cycloligase moiety of imidazoleglycerol phosphate synthase (HisF) support the previous assignment of the $(\beta\alpha)_8$-barrel fold to these proteins. The alignments also reveal a second phosphate-binding motif located in the first halves of both enzymes and thereby support the hypothesis that HisA and HisF have evolved by a sequence of two gene duplication events. Comparison of the amino acid compositions of HisA and HisF from mesophiles and thermophiles shows that the thermostable variants of both enzymes contain a significantly increased number of charged amino acid residues and may therefore be stabilized by additional salt bridges.

Key words Histidine operon · $(\beta\alpha)_8$-Barrel enzymes · Phosphate-binding site · Protein thermostability · Salt bridges

Communicated by: K. Horikoshi

R. Thoma · M. Schwander · K. Kirschner · R. Sterner[1] (✉)
Abteilung für Biophysikalische Chemie, Biozentrum der Universität Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland

W. Liebl[1]
Lehrstuhl für Mikrobiologie, Technische Universität München, Arcisstrasse 21, D-80333 Munich, Germany

Present address:
[1] Institut für Mikrobiologie und Genetik der Universität Göttingen, Grisebachstr. 8, D-37077 Göttingen, Germany
Tel. +49-551-399657; Fax +49-551-393805
e-mail: rsterne@Uni-MolGen.gwdg.de

Introduction

Histidine biosynthesis is an ancient metabolic pathway that was presumably assembled before the separation of Bacteria, Archaea, and Eucarya (Alifano et al. 1996; Bult et al. 1996; Charlebois et al. 1997). The pathway requires 11 enzymatic reactions, which are encoded by corresponding functional domains. They are encoded by only eight genes in the enterobacteria (*his*G, *his*I–E, *his*A, *his*H, *his*F, *his*B, *his*C, and *his*D, in order of the anabolic pathway), because the products of *his*D, *his*B, and *his*I–E are bifunctional (Winkler 1996). In all other investigated microorganisms, the two functions of *his*B are encoded by separate *his*Bp and *his*Bd genes (Fani et al. 1995; Bult et al. 1996; Charlebois et al. 1997). The situation found for *his*I–E is more complicated. Bacteria have mostly bifunctional *his*I–E genes, whereas *his*I and *his*E are separate genes in the Archaea (Charlebois et al. 1997; Bult et al. 1996). In the Eukarya, the *his*I–E gene is fused to the N-terminus of the *his*D gene (Hinnebusch and Fink 1983; Fani et al. 1995). Although *his* genes are organized in a single operon in some Bacteria (Carlomagno et al. 1988; Delorme et al. 1992) and in the archaeon *Sulfolobus solfataricus* (Charlebois et al. 1997), they are scattered throughout the chromosome in other organisms (Beckler and Reeve 1986; Limauro et al. 1990; Bult et al. 1996). The physiological and evolutionary implications of the diverse organization of *his* genes are as yet unknown.

Gene duplication events seem to have played an important role during the assembly of the *his* operon. The extensive sequence similarities between the amino acid sequence of HisH with glutamine amidotransferases of tryptophan

biosynthesis (Zalkin and Smith 1998) are indicative of an evolutionary relationship between these key metabolic pathways. Additionally, it was postulated that *his*A and *his*F have evolved by a sequence of two gene duplications, starting from a precursor "half-sequence" of *his*A (Fani et al. 1994). Moreover, both HisA and HisF have been predicted to belong to the class of $(\beta\alpha)_8$-barrel proteins (Wilmanns and Eisenberg 1993, 1995).

To understand these relationships better, we chose the hyperthermophile *Thermotoga maritima* as a new source of *his* genes. It belongs to the domain of Bacteria and represents one of the earliest and most slowly evolving branches. The proximity of *T. maritima* to the root of the phylogenetic tree implies that it has retained more of the characteristic properties of the last common ancestor before diversification into the three extant domains (Woese et al. 1990) than most other known members of the Bacteria or Archaea. Here we report the sequence of the 3′-terminal part of the *his*D gene and the complete sequences of the *his*C, *his*Bd, *his*H, *his*A, *his*F, and *his*I–E genes of *T. maritima*. The organization of these genes and analysis of the deduced amino acid sequences of the gene products give insights into the evolution of the *his* operon, and the relationship between HisA and HisF, as well as the structural basis of the high thermostability of these two enzymes from *T. maritima* and other thermophiles.

Materials and methods

Manipulation, sequencing, and sequence analysis of DNA

DNA was prepared, digested with restriction endonucleases, and ligated according to Sambrook et al. (1989). *Escherichia coli* strains were transformed with plasmid DNA by electroporation (Dower et al. 1988), using a Gene Pulser (Biorad, Munich, Germany). Exonuclease III digestion of cloned DNA for sequencing was performed with a kit from Promega (Madison, WI, USA). Nonradioactive DNA sequencing followed the dideoxynucleotide chain-termination method (Sanger et al. 1977), using the DyeDeoxy Terminator Cycle sequencing kit from Applied Biosystems (Foster City, CA, USA). The gel was run overnight on an ABI 373 stretch sequencer from Applied Biosystems, with on-line data collection. The sequences were analyzed with the 373A DNA sequencer data analysis program. PCRs were performed in the Trio-block from Biometra (Göttingen, Germany), and oligonucleotides were purchased from Microsyn (Windisch, Switzerland). The GCG software package was used for similarity searches in nucleotide and protein databanks (program BLAST), for searches for open reading frames and restriction sites (program MAP), as well as for alignments of amino acid sequences (programs GAP and PILEUP). In analogy to a previous study (Alifano et al. 1991), to search for regions of high C over G content, a computer program was written in Turbo Pascal that computes within a sliding frame of 78 nucleotides both the C and G content and stores the corresponding fraction of these nucleotides at the first position of the frame.

Media and strains

SOC, LB, and M9 minimal media were prepared as described by Sambrook et al. (1989). M9 minimal medium was supplemented with glucose (30 mM), thiamine (7 mM), biotin (5 mM), and a mixture of heavy metals: $MoNa_2O_4$ (10 mg/l), $CoCl_2$ (2 mg/l), $CuSO_4{\cdot}5H_2O$ (2 mg/l), $MnCl_2{\cdot}4H_2O$ (10 mg/l), $MgSO_4{\cdot}7H_2O$ (200 mg/l), $ZnCl_2$ (10 mg/l), $FeSO_4{\cdot}7H_2O$ (5 mg/l), $CaCl_2{\cdot}2H_2O$ (50 mg/l), and H_3BO_3 (10 mg/l).

The *E. coli his*A deletion strain Hfr G6 *his*A323λ^- (Matney et al. 1964) and the *E. coli his*F deletion strain UTH860 ara-14, glnV44 (AS), galK2, λ^-, *his*F860, rpsL145(stR), malT1(λ^R), xglA5, mtl-1 (Goldschmidt et al. 1970) are derivatives of *E. coli* K12. Both strains were kindly provided by the *E. coli* genetic stock center (Stanford, CA, USA).

Cloning and sequencing strategy

Auxotrophic *E. coli* strain Hfr G6 hisA323λ^-, which lacks the *his*A gene (Matney et al. 1964), was transformed with a gene bank of genomic *T. maritima* DNA (Ostendorp et al. 1993; Sterner et al. 1995), which encompasses about 6000 clones. The transformed cells were plated on supplemented Vogel-Bonner minimal medium and incubated at 37°C. Five prototrophic colonies were detected after 60 h and their plasmids (pUN121/I-V) were isolated. Digestion with *Hin*dIII and *Eco*RI indicated that the plasmids pUN121/I, II, IV, and V were identical (data not shown). Subsequently, pUN121/I and pUN121/III were also transformed into *E. coli* strain UTH 860, which lacks the *his*F gene (Goldschmidt et al. 1970). Both plasmids conferred prototrophy to the recipient within 48 h. pUN121/I and pUN121/III, which contain inserts of 4.3 kb and 10 kb, were digested with several endonucleases, and restriction maps were constructed (Fig. 1A). Using these maps, two overlapping restriction fragments of pUN121/I, a 2.7-kb *Hin*dIII-*Hin*dIII and a 2.2-kb *Cla*I-*Pst*I fragment, as well as a 5.65-kb *Pst*I-*Pst*I restriction fragment of pUN121/III, were subcloned into the vector pBlueskript II SK+ (Stratagene, La Jolla, CA, USA). The resulting plasmids were termed SK+/I H-H, SK+/I C-P, and SK+/III P-P (Fig. 1A).

To produce appropriate templates for DNA sequencing, deletion derivatives of these plasmids were constructed by digestion with exonuclease III from both sides. For SK+/I H-H, digestion was started for both orientations of the insert from the *Sal*I restriction site within the polylinker of SK+, and protection of the vector was assured by digestion with *Kpn*I. Similarly, digestion of SK+/I C-P was started from the *Xba*I site within the polylinker of SK+, and protection of the vector was assured by digestion with *Bst*XI. For sequencing the opposite strand, two subclones of SK+/I C-P were prepared using restriction sites deduced from the determined sequence. For SK+/III P-P, digestion was

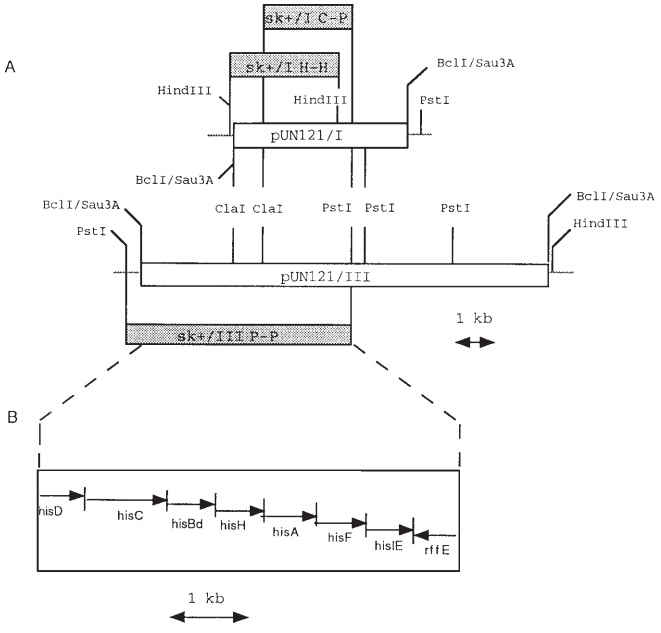

Fig. 1A,B. Sequencing strategy and order of *his* genes of *Thermotoga maritima*. **A** Restriction map of pUN121/I and the sequenced subclones SK$^+$/I C-P and SK$^+$/I H-H, as well as restriction map of pUN121/III and the sequenced subclone SK$^+$/III P-P. *Stippled lines* mark DNA segments of the vector pUN121 that contain restriction sites used for subcloning (see text for details). **B** Order of genes in the sequenced *his* gene cluster from *T. maritima*. HisD codes for L-histidinol dehydrogenase, *his*C for L-histidinol phosphate aminotransferase, *his*Bd for imidazoleglycerol phosphate dehydrogenase, *his*H for the glutamine amidotransferase subunit of the imidazoleglycerol phosphate synthase, *his*A for the N^1-((5′-phosphoribosyl)-formimino)-5-aminoimidazole-4-carboxyamide ribonucleotide isomerase, *his*F for the cycloligase moiety of the imidazoleglycerol phosphate synthase, and *his*I–E for the bifunctional enzyme phosphoribosyl-ATP-pyrophosphohydrolyase fused to phosphoribosyl-AMP cyclohydrolase. *rff*E is not part of the *his* operon and encodes a protein with high sequence similarity to the UDP-*N*-acetylglucosamine-2-epimerase

started from the *Xba*I site of the polylinker of SK+, and protection of the vector was assured by digestion with *Not*I. The *Not*I-site was filled in with α-phosphorothioate deoxynucleotides before exonuclease treatment. After religation, appropriate deletion plasmids were sequenced, using either the T3 or the T7 priming sites of SK+. All isolated SK+/III P-P clones had the insert in the same orientation. Therefore, for sequencing of the opposite strand, either subclones of SK+/III P-P were prepared or internal primers were used that were deduced from the previously determined sequence.

The gene bank of *T. maritima* used for cloning of the cluster of *his* genes contains *Sau*3A I fragments of 3–6 kbp (Ostendorp et al. 1993). Thus, the 10-kbp insert of pUN121/III is unusually large. To unravel the reason for this unusual size, a 2.3-kb *Pst*I-*Pst*I fragment that is located on pUN121/III, downstream of the sequence of SK+/III P-P (Fig. 1A), was cloned into pBlueskript II SK+, and about 500 bp were sequenced from both ends (unpublished data). A database search with the analyzed sequence showed 99% identity to parts of the Tn1000 transposon from *E. coli* (Broom et al. 1995).

Results

Cloning and sequencing of *his* genes from *T. maritima*

The 3′-terminal part of the tryptophan operon from *T. maritima* was cloned recently by functional complementation of *E. coli* auxotrophic strains (Sterner et al. 1995). Thus, regulatory elements that are necessary for expression of *T. maritima* genes are recognized by *E. coli*, and suggested an analogous approach to isolate genes involved in histidine biosynthesis. To this end, the auxotrophic strain of *E. coli* Hfr G6 hisA323λ$^-$, which lacks the *his*A gene (Matney et al. 1964), was transformed with a gene bank of genomic *T. maritima* DNA (Ostendorp et al. 1993; Sterner et al. 1995). Five plasmids were isolated by this approach, two of which were nonidentical (pUN121/I and pUN121/III) and also conferred prototrophy to *E. coli* strain UTH 860, which lacks the gene *his*F (Goldschmidt et al. 1970). Two overlapping restriction fragments of pUN121/I and a restriction fragment of pUN121/III were subcloned into the vector pBlueskript II SK+. The resulting plasmids SK+/I H-H, SK+/I P-C, and SK+/III P-P (Fig. 1A) were digested stepwise with exonuclease III and sequenced in both directions.

Altogether, a continuous stretch of 5269 bp was sequenced, which encodes eight open reading frames (ORFs). From 5′ to 3′, seven ORFs are transcribed in the same direction, whereas the last ORF is transcribed in the reverse direction (Fig. 1B).

Figure 2 presents the complete nucleotide sequence determined in this work and the deduced amino acid sequences of the eight identified ORFs. These were compared with orthologous sequences in several databanks, and significant sequence identities were detected for all eight ORFs. The seven consecutive ORFs correspond to 9 of 11 functional domains of histidine biosynthesis (Alifano et al. 1996). Only the *his*G and *his*Bp genes are missing, and the order of genes is otherwise identical to that found in *E. coli*. Accordingly, these genes were designated *his*D, *his*C, *his*Bd, *his*H, *his*A, *his*F, and *his*I–E. The encoded proteins possess approximately the same number of amino acid residues (HisD, 184, N-terminus missing, HisC 335, HisBd 195, HisH 201, HisA 241, HisF 253, and HisI–E 196) as the counterparts in *E. coli* (Winkler 1996). The eighth ORF, for which the N-terminal region is missing, shows significant sequence identity to the C-terminal regions of the YVYH gene product of *Bacillus subtilis* (58%) and of the *rff*E gene product of *E. coli* (50%) (Fig. 1B). Both genes encode UDP-*N*-acetylglucosamine-2-epimerase, which participates in lipopolysaccharide O-antigen biosynthesis (Kawamura et al. 1982). The succession of *his* genes of *T. maritima* displays most of the features familiar from operons of enterobacteria. All the complete ORFs are preceded by a putative ribosome-binding site, complementary to the 3′-end of the 16S rRNA of *T. maritima* (3′-UCCUUUCU-5′) (Achenbach-Richter et al. 1987).

The incomplete sequence of HisD (L-histidinol–NAD$^+$ oxidoreductase) lacks about 250 residues at the N-terminus.

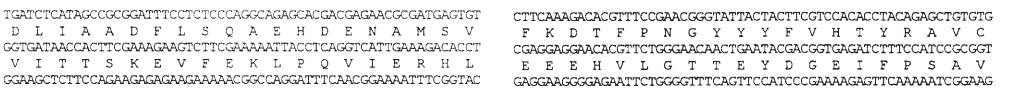

Fig. 2. Nucleotide sequences of *his* genes from *T. maritima* with deduced amino acid sequences of encoded proteins *below*. Stop codons are indicated by *asterisks*. Putative ribosome binding sites are *underlined.* The G > C-rich bubble after the stop codon of *his*I–E is *boxed*

The initial methionine of HisC (L-histidinol phosphate aminotransferase) appears to be encoded by a GTG. This putative start codon overlaps with the stop codon of the preceding *his*D gene. Moreover, the stop codon of *his*C overlaps with the start codon of the subsequent *his*Bd gene. HisBd shows 36% sequence identity to the imidazoleglycerol phosphate dehydratase domain, the C-terminal part of *E. coli* HisBp-Bd. However, as observed in a number of other investigated Bacteria and Archaea (Fani et al. 1995; Charlebois et al. 1997), no ORF with homology to the L-histidinol phosphatase domain HisBp was detected in the sequenced segment of the *T. maritima* genome (Fig. 1B). The initial methionine of HisH, which codes for the glutamine amidotransferase subunit of imidazoleglycerol phosphate synthase, appears to be encoded by TTG, and it overlaps with the stop codon of *his*Bd. The coding region of *his*H overlaps by 22 nucleotides that of *his*A, which encodes N^1-((5′-phosphoribosyl)-formimino)-5-aminoimidazole-4-carboxyamide ribonucleotide isomerase. The coding region of *his*A overlaps by 11 nucleotides that of *his*F, which encodes the cycloligase subunit of imidazoleglycerol phosphate synthase. The stop codon of *his*F overlaps the start codon of *his*I–E, which encodes the bifunctional enzyme phosphoribosyl-ATP-pyrophosphohydrolyase fused to phosphoribosyl-AMP cyclohydrolase.

Rho-independent termination of transcription requires an inverted repeat, followed by a stretch of oligo(dT) (D'Aubenton-Carafa et al. 1990). No such sequence motif was found downstream of the stop codon of *his*I–E, unlike the *his* operons of *E. coli* and *Salmonella typhimurium* (Carlomagno et al. 1988), *Azospirillum brasiliense* (Fani et al. 1989), and *Lactobacillus lactis* (Delorme et al. 1992). Alifano et al. (1991) proposed that rho-dependent termination of transcription requires a region of high cytosine over guanosine content, a so-called C > G-rich "bubble." To search for such a "bubble," the G- and C-content of the complete sequence given in Fig. 2 was compared from 5′ to 3′, using a "window size" of 78 nucleotides. A distinct C > G-rich "bubble" was indeed found downstream of the stop codon of the *his*I–E gene (Fig. 3), which might serve as the termination site of transcription. The three additional small "bubbles" around nucleotides 210, 720, and 2000 are less pronounced.

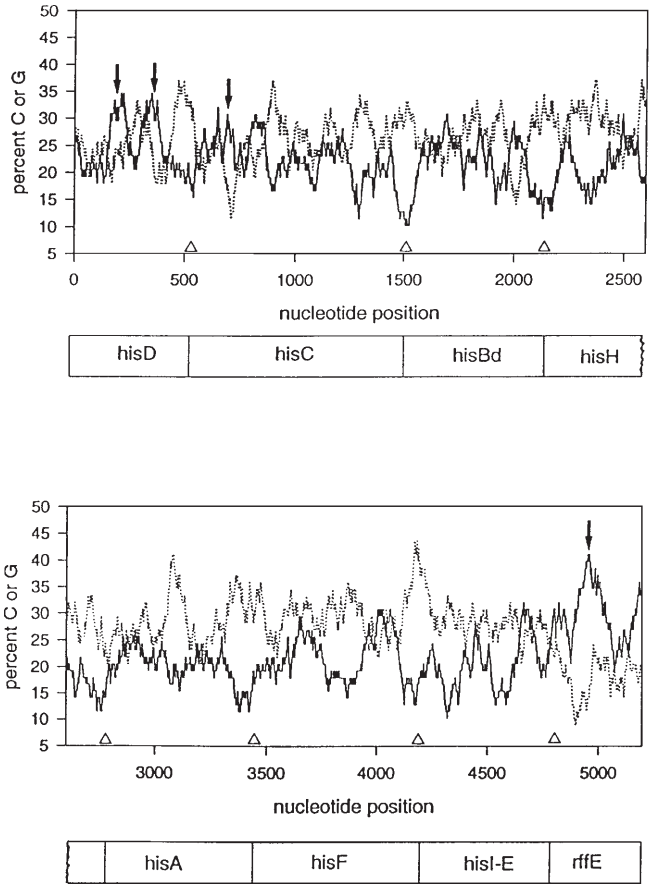

Fig. 3. C > G-rich bubbles (*arrows*) in the sequenced *his* gene cluster. The percentages of C (*solid line*) and G (*dotted line*) are plotted as a function of the nucleotide sequence of the *his* gene cluster. At a given nucleotide position, the percentages of C and G are calculated for the 78 subsequent nucleotides. *Triangles* indicate translation stop codons of the individual *his* genes. The C > G-rich bubble downstream of *his*I–E suggests a rho-dependent termination of transcription (Alifano et al. 1991)

Evidence for two phosphate-binding sites in HisA and HisF

The *his*A and *his*F genes of *T. maritima* are adjacent (Fig. 1B), as observed in most bacterial *his* operons. Moreover, Fani et al. (1994) discovered that the amino acid sequences of HisA and HisF each consist of two duplicated tandem regions with significant sequence identity. Because multiple alignments of several entire HisA and HisF sequences also revealed high similarity, it is likely that they are related by a further gene duplication event. It is important to trace the two putative gene duplication events as far back to the point of phylogenetic separation as possible. Because *T. maritima* is closer to that point than any of the microorganisms studied to date (Woese et al. 1990), we aligned the sequences of the two halves of HisA and HisF with each other (data not shown), following the procedures of Fani et al. (1994, 1995). The observed identities (HisA1/HisA2, 25%; HisF1/HisF2, 25%) and similarities (HisA1/HisA2, 42%; HisF1/HisF2, 41%) are the same as those found in HisA of *Methanococcus voltae* (Fani et al. 1994) and HisF of *Saccharomyces cerevisiae* (Fani et al. 1995).

HisA and HisF have been predicted to belong to the $(\beta\alpha)_8$-barrel class of proteins (Wilmanns and Eisenberg 1993, 1995; Bork et al. 1995). To further test this hypothesis on the basis of an actualized set of data, we performed a multiple joint alignment of all available sequences of HisA and HisF. Moreover, secondary structure predictions were performed separately for HisA and HisF with the program PHD (Rost and Sander 1994). Figure 4 shows that the predicted secondary structural elements of hisA and hisF superimpose well. Furthermore, the predictions are characterized by an alteration of β-strands and α-helices, which is typical of $(\beta\alpha)_8$-barrel enzymes. The active site residues of

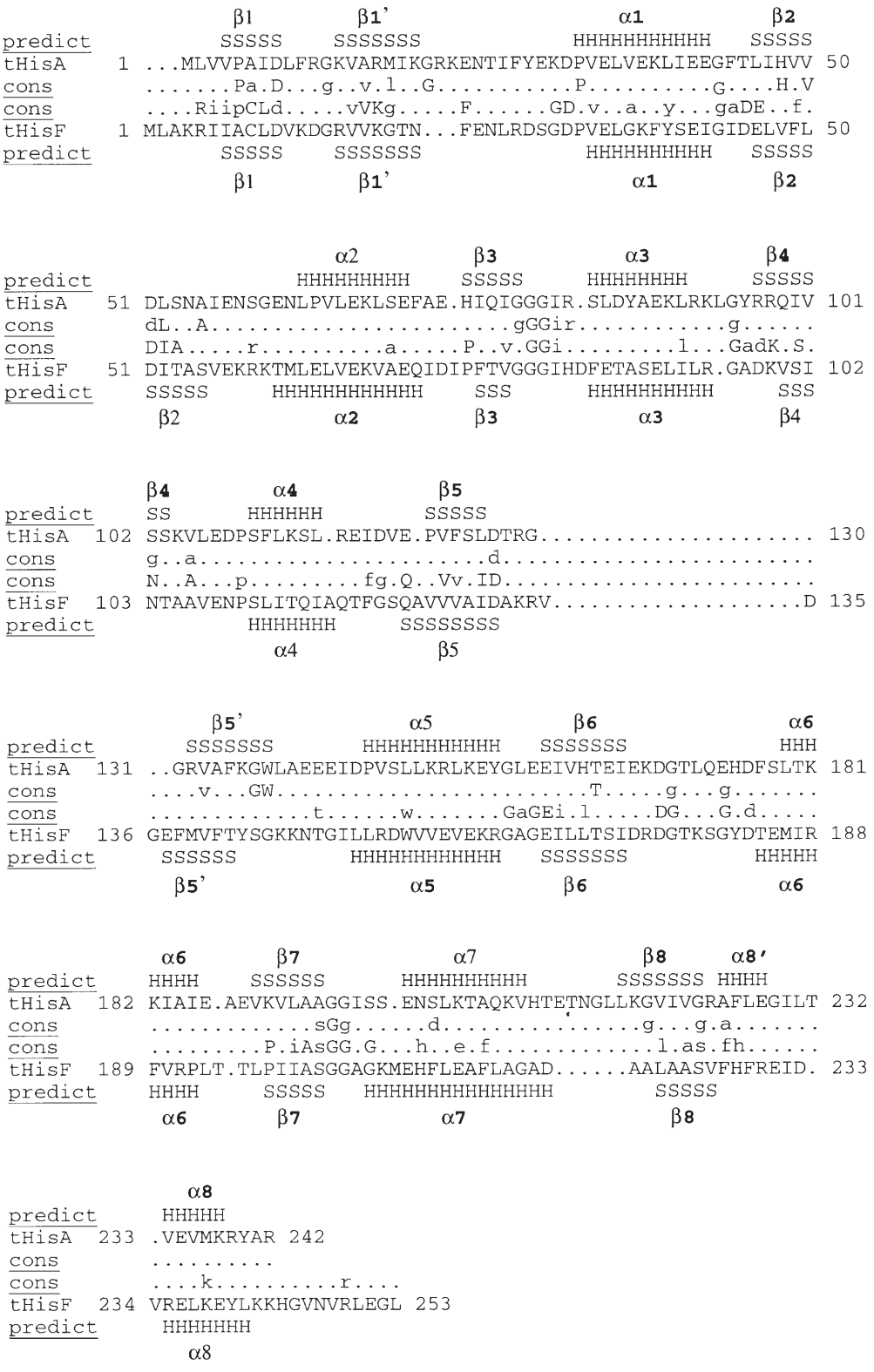

Fig. 4. Predicted secondary structural elements of HisA and HisF are superimposable and indicative of the $(\beta\alpha)_8$-barrel topology. The shown alignment of HisA and HisF from *T. maritima* (tHisA and tHisF) is based on a multiple joint alignment of 17 sequences of His A and 16 sequences of HisF; the sequence numbering ignores gaps. *cons*, invariant residues (*uppercase letters*) and residues that are identical in at least 88% (*lowercase letters*) of all HisA or HisF sequences; *predict*, secondary structural elements (*S*, β-strand; *H*, α-helix) predicted separately for HisA or HisF with the PHD program (Rost and Sander 1994). β1 and β1′, as well as β5 and β5′, indicate alternative sequence stretches that could form the first and the fifth β-strand of the predicted β-barrel. The large gap between β5 and β5′ arises from corresponding sequence insertions in some of the HisF sequences

all known $(\beta\alpha)_8$-barrel enzymes are located at the C-terminal ends of the β-strands and in the loops that connect the β-strands with subsequent α-helices (Reardon and Farber 1995). Accordingly, clusters of conserved residues are found in the sequence stretches of HisA and HisF that comprise strands β1, β2, β3, and β7, as well as the subsequent loop regions.

Many enzymes with the $(\beta\alpha)_8$-barrel topology contain a phosphate-binding site that is located in their C-terminal region, involving the region from β-strand 7 to α-helix 8

Fig. 5A,B. HisA and HisF contain two phosphate-binding motifs. Multiple sequence alignments of the second and fourth quarters of all available HisA (**A**) and HisF (**B**) sequences. *Upper blocks*, second quarters; *lower blocks*, fourth quarters; *motif*, amino acid residues that are found in at least 60% of the phosphate-binding motifs in enzymes with known or assumed $(\beta\alpha)_8$-barrel topology (Bork et al. 1995); *predict*, localization of secondary structure elements (*S*, β-strand; *H*, α-helix) as predicted by the PHD program (Rost and Sander 1994); *first and last lines*, numbering of the secondary structural elements according to Fig. 4. Organisms: *Azospirillium brasilense* (*A. bras*), *Escherichia coli* (*E. coli*), *Haemophilus influenzae* (*H. infl*), *Klebsiella pneumoniae* (*K. pneu*), *Lactococcus lactis* (*L. lact*), *Methanococcus jannaschii* (*M. jann*), *Methanococcus thermolithotrophicus* (*M. ther*), *Methanococcus vannielii* (*M. vann*), *Methanococcus voltae* (*M. volt*), *Pseudomonas aeruginosa* (*P. aerg*), *Rhodobacter sphaeroides* (*R. spha*), *Salmonella typhimurium* (*S. typh*), *Streptomyces coelicolor* (*S. coel*), *Sulfolobus solfataricus* (*S. solf*), *Synechocystis* sp. (*S.* sp), *Thermotoga maritima* (*T. mari*)

A

```
                 β3          α3                  β4          α4
predict    ..SSSSS....HHHHHHHHH.........SSSSSSS......HHHHHH.....
A.bras     .TVPVQLGGGIRDLKTIALWLEK.....G.VSRVILGTVALREPELVREARREFP
E.coli     .NVPVQVGGGVRTEEDVAALLE.....AG.VARVVVGSTAVKSQDMVKGWFERFG
S.typh     .NVPVQVGGGVRTEEDVAALLK.....AG.VARVVIGSTAVKSPDVVKGWFERFG
H.infl     .QCKVQVGGGIRTEQDVADLLA.....VG.ANRVVIGSTAVTHRSMVKNWFIKYG
M.volt     .KVPVQMGGGIRTIQDVEELVD.....LG.INKVIIGTVAVQNPDFVEQLAKKVG
M.ther     .NVPVQIGGGI.SVRCTESNLIE....IG.AKRIILGTVAVENPDIVEEISKKVG
M.vann     .GVPVQIGGGIRSVSDALYFIE.....KG.AEKVILGTVAIQNPKIVREISSIVG
M.jann     .NVPVEVGGGIR.NLEIAKELIS....LG.VDRVIVGTKAILEPKFIDDLNKEIG
R.spha     .AVPAQLGGGIRDMATIALWLEK.....G.LARVILGTVAVENPGLVREAARAFP
L.lact     .RLKIEIGGGIR.DFDTVRMYLEQ....IG.VERVILGTAAVEKPDFLKELLIKYG
S.coel     .DIKVELSGGIR.DDDTLAAALA....TG.CTRVNLGTAALETPEWVAKVIAEHG
S.solf     GFDWIQVGGGIR.DVEKAKRLIS....LD.VNALIFSTIVFTNFNLFYNIVREIG
T.mari     .AEHIQIGGGIR.SLDYAEKLRK....LG.YRRQIVSSKVLEDPSFLKSLREIDV
motif      ...P....GG.....D............G.A.....G.....G............
A.bras     .TTPVIASGAVSSIDDLIALKKEED..TG.IQGVICGRALYDGRIDPKTALDLLS
E.coli     .QVAFQSSGGIGDIDDDVAALRG.....TG.VRGVIVGRALLEGKFTVKEAIACWQ
S.typh     .QIAFQSSGGIGDIDDDVAALRG.....TG.VRGMIVGRALLEGKFTVKEAIQCWQ
H.infl     .PIQFQSSGGIGSLADIEALKG.....TG.VSGVIVGRALLEGKFTLSEAIKCWQ
M.vann     .NIPIIASGGVTTVEDLLKFKE.....IG.VYAVVVGSALYKDMINLKDAILASK
M.ther     .EIPVIASGGVTTIEDLIKLKA.....VG.VEGVVVGSAIYKNLIDLKEAIKACR
M.volt     .NIPIIASGGVTTIEDLIEFKK.....IG.VAGVVVGSALYKNNFKLQDAINTVN
M.jann     .DIPIIYSGGITTLEDIKALKE.....LG.IYGVVIGSALYKGLIDLKKALEIVK
R.spha     .SIPVIASGGVSSLADLLALR.....DTGSIAGAISGRALYAKAFTLEEALQALR
S.coel     .DRPVVASGGVSSLDDLRAIAGLVP..AG.VEGAIVGKALYAKAFTLEEALEATS
L.lact     .SLNVIISGGVKDNSDIQRAT......RSDFYGIIVGKAYYEGKINLEKEFRNAN
S.solf     RGGVKEYAGGVSSDSDITFLK.....NVG.FDYIIVGMAFYLNKIRGTNVV....
T.mari     AEVKVLAAGGISSENSLKTAQKVHTETNGLLKGVIVGRAFLEGILTVEVMK.RYA
predict    ..SSSSSS.....HHHHHHHHHH.......SSSSSSS.HHHH....HHHHH....
                 β7          α7                  β8      α8'     α8
```

B

```
                 β3          α3                  β4          α4
predict    ......SSS......HHHHHHHHHH.....SSS........HHHHHHH.....SS
A.bras     TAEQVFMPLTVGGGVRTVDDIRKLLLA.GADKVSINTAAIHRPEFVQEAAEKFGAQC
E.coli     VAEVIDIPFCVAGGIKSLEDAAKILSF.GADKISINSPALADPTLITRLADRFGVQC
H.infl     IAQVIDIPFCVAGGIKTIEDAEKLFAF.GADKISINSPALADPTLISRLADRFGVQA
K.pneu     VAEVIDIPFCVAGGIKSLDDAAQILSF.GADKISINSPALADPTLITRLADRFGVQC
L.lact     AARELSIPLTVGGGIRSIDDFRRILAR.GADKVSVNSAAIENPELIRQAANEFGVQC
M.jann     TAEKVFIPLTVGGGIKSIEDFRRILRA.GADKVSINTAAVKNPNLIKEASEIFGSQC
P.aerg     AGECF.MPICYGGGIKTLEHAEKIFSL.GVEKVSINTAALMDLSLIRRIADKFGSQS
R.spha     TAEQCFMPLTVGGGVRTHQDVRALLLA.GADKVSFNSAAVADPGVVP.SRRPLRSQC
S.solf     TASVLSIPLTVGGGIRTPDDVSMALRS.GADKVSINTAAVESSQIVKKSAEEFGSQA
S.sp       TAEEVFIPLTVGGGISTLEHIKNLLRA.GADKVSVNSSAVRDPDFISRASDRFGRQC
S.typh     VAEVIDIPFCVAGGIRSIDDAAKILSF.GADKISINSPALADPTLITRLADRFGVQC
T.mari     VAEQIDIPFTVGGGIHDFETASELILR.GADKVSINTAAVENPSLITQIAQTFGSQA
motif      .......P....GG.....D.........GA......G......G............
A.bras     VADGLRIPVIASGGVGTLDHLVEGIREGHATAVLAASI..FHFGTYTIGQAKAALAE
E.coli     VREVCHVPLIASGGAGTMEHFLEAFRDADVDGALAASV..FHKQIINIGELKAYLAT
H.infl     VRGVCRVPLIASGGAGKMVHFRDAFIEAKVDGALAASV..FHKQIIEIGELKSYLVQ
K.pneu     VRAVCRVPLIASGGAGTMEHFLEAFRDASVDGALAASV..FHKQIINIGELKTYLAA
L.lact     VCTAVNIPVVASGGCGKISDIVEVFQNTRSDAALFASL..FHYGEEQLMKLKTNL..
M.jann     ISKSVKLPVIASGGCGKPEHVYEAFVYGKADAALMAGI..LHYREYTIEEIKKYCAD
P.aerg     IASNVHVPVIACGGAGSIADLIDLFERTCVSAVAAGSLFVFHGKHRAVLISYPDVNK
R.spha     ISDAVPIPVIASGGVGTLDHLVEGVTEGGASAVLAASI..FHFGEFTIGEAKAHMAA
S.solf     IVDSVNIPVIASGGAGKMEHFYEVFSLAKADAALAAGI..FHDGIIKIKDLKSYLSQ
S.sp       IAERVEIPVIASGGAGNCQHVYEAFTEGKAEAALLASL..LHYGQLTIGELKTFLEA
S.typh     VRDVCRVPLIASGGAGTMEHFLEAFRDADVDGALAASV..FHKQIINIGELKAYLAG
T.mari     VRPLTTLPIIASGGAGKMEHFLEAF.LAGADAALAASV..FHFREIDVRELKEYLKK
predict    HHH....SSSSS...HHHHHHHHHHHHHHHHH..SSSSS..........HHHHHHH..
                 α6      β7          α7                  β8          α8
```

(Brändén 1991; Wilmanns et al. 1991). Using a sequence motif that is conserved in several $(\beta\alpha)_8$-barrel proteins, Bork et al. (1995) identified putative phosphate-binding sites near the C-termini of both HisA and HisF. Because the substrates of both enzymes each contain two phosphate groups and because of the high sequence similarity of the N- and C-terminal halves (Fani et al. 1994), we asked whether a second phosphate-binding site exists in HisA and HisF. The multiple sequence alignments of the second (βα(3 + 4)) and fourth (βα(7 + 8)) quarters of HisA and HisF are shown in Fig. 5A,B. Although the number of strongly conserved residues is small, the central residues

Table 1. HisA and HisF contain a significantly higher number of charged amino acid residues in thermophiles than in mesophiles

Organism	Percent (%) (D + E + R + H + K)[a]	
	HisA	HisF
Escherichia coli	23.3	27.5
Salmonella typhimurium	22.8	26.7
Azospirillum brasiliense	24.0	28.7
All mesophiles[b]	24.9 ± 3.2 (n = 8)	26.8 ± 1.6 (n = 9)
Methanococcus thermolithotrophicus	28.6	No data available
Methanococcus jannaschii	32.5	32
Thermotoga maritima	33.2	29.2
All thermophiles[b]	31.4 ± 2.5 (n = 3)	30.6 (n = 2)

[a] Fraction of charged amino acid residues in the polypeptide chain
[b] Fraction of charged residues averaged over the currently available protein sequences (n) in the SWISSPROT databank

(P.....GG.....D.....G.A.....G.....G) of the phosphate-binding motif are clearly detectable not only in the fourth quarters (between strand β7 and helix α8), as noted by Bork et al. (1995), but also in the second quarters (between strand β3 and helix α4) of both HisA and HisF. The two highly conserved glycine residues that are located between strand β3 and helix α3, as well as between strand β7 and helix α7 in both HisA and HisF (Fig. 5), are probably crucial for phosphate binding. Two glycine residues at equivalent positions between strand β7 and helix α7 are found in all 20 available amino acid sequences of the $(\beta\alpha)_8$-barrel enzyme phosphoribosyl anthranilate isomerase (data not shown). The X-ray structures of these enzymes from *E. coli* and *T. maritima* show that the peptide amide groups of the invariant glycines are forming hydrogen bonds to oxygen atoms of the bound phosphate ion (Hennig et al. 1997).

Sequence comparisons between several $(\beta\alpha)_8$-barrel enzymes of tryptophan biosynthesis in mesophiles and thermophiles showed that the thermostable variants possess a significantly higher number of charged amino acid residues (Sterner et al. 1995). In an analogous approach, extensive comparisons of all available amino acid sequences in the SWISSPROT databank of HisA and HisF were performed (data not shown). The most striking result is that the variants from thermophilic organisms have a significantly increased number of charged amino acid residues compared with the corresponding variants from mesophilic organisms (Table 1).

Discussion

Organization of the *T. maritima his* gene cluster

A 5.27-kb fragment of *T. maritima* genomic DNA was isolated by functional complementation of auxotrophic *E. coli* strains that are deficient in either *his*A or *his*F. The sequenced fragment contains eight ORFs, seven of which are significantly similar to those of known enzymes of histidine biosynthesis. This *his* gene cluster has the order *his*DCBdHAFI–E and comprises all genes of the *his* operon, with the exception of *his*G and *his*Bp (Fig. 1B). Thus, as in the case of the tryptophan operon (Sterner et al. 1995), the organization of the *his* gene cluster of *T. maritima* is similar to that found in a number of "modern" gram-negative Bacteria such as *E. coli*, where the *his* operon is organized as *his*GDCBpBdHAFI–E (Carlomagno et al. 1988). According to 16S rRNA sequence comparisons (Woese et al. 1990), *T. maritima* is more closely related to the last common ancestor before the separation of the three extant domains of life than any other organism for which the organization of the *his* operon has been revealed. The *his* gene cluster of *T. maritima* is very compact: there are no intercistronic regions (Fig. 2), whereas in the *his* operons of some gram-negative Bacteria such as *A. brasiliense* and *Lactococcus lactis* the *his* genes are separated by additional ORFs (Alifano et al. 1996). This compact operon organization in *T. maritima* suggests that insertion of other ORFs, or distribution of the *his* genes over the whole genome, as found in Eukarya (Fani et al. 1995) or the archaeon *Methanococcus jannaschii* (Bult et al. 1996), were later events in evolution.

A heterogeneous situation is found for the structure of the *his*B gene, which has been well studied both in prokaryotes (Grisolia et al. 1983) and eukaryotes (Struhl 1985). In *E. coli* and *S. typhimurium*, *his*B codes for a protein of 355 amino acids. This protein possesses two enzymatic activities, L-histidinol phosphatase and imidazole glycerolphosphate dehydratase activity. In *S. cerevisiae* and *Neurosporra crassa*, the two activities are encoded by two separate genes (Broach 1981; Fink 1964). A similar situation is found in the bacterium *Streptomyces coelicolor*, where *his*Bd encodes the dehydratase and *his*Bp encodes the phosphatase activity (Limauro et al. 1990). In *T. maritima*, as in other Bacteria such as *A. brasiliense* (Fani et al. 1989) or *L. lactis* (Delorme et al. 1992), and in the

archaeon *S. solfataricus* (Charlebois et al. 1997), only the monofunctional *his*Bd gene has been identified. It appears that *his*Bp is localized elsewhere in the chromosome of these organisms. However, also in *M. jannaschii*, for which the entire genomic DNA has been sequenced (Bult et al. 1996), no region with significant similarity to *his*Bp could be identified. Presumably, *T. maritima* and several other organisms contain an alternative phosphatase to replace the *his*Bp gene product. In support of this notion, the function of a specific L-histidinol phosphatase from *S. cerevisiae* seems to be carried out by a less specific alkaline phosphomonoesterase under certain growth conditions (Gorman and Hu 1969).

The mechanisms of termination of transcription of the his genes seem to be different between *T. maritima* and other Bacteria. Whereas rho-independent termination of transcription is found in the *his* operons of *E. coli*, *S. typhimurium*, *L. lactis*, and *A. brasiliense* (Fani et al. 1995), no inverted repeat followed by an oligo(dT) sequence was detected downstream of the *his*I–E gene of *T. maritima*. Instead, in agreement with the "bubble model" for rho-dependent termination (Alifano et al. 1991), immediately downstream of the stop codon for *his*I–E, a C > G-rich "bubble" is present (see Fig. 3). Several additional, albeit less pronounced, C > G-rich "bubbles" were detected within the *his* operon of *T. maritima*, one of them immediately downstream of the stop codon for *his*D (Fig. 3). Such internal rho-dependent termination signals were shown to constitute additional termination sites for transcription in the *his* operon of *S. typhimurium* (Alifano et al. 1991), and a similar function cannot be excluded for *T. maritima*.

Evolution of *his*A and *his*F

The *his*A and *his*F genes are adjacent in most investigated *his* operons (Fani et al. 1995; also, this study) and catalyze consecutive steps in histidine biosynthesis (Winkler 1996). Based on comparisons of gene and amino acid sequences, it was suggested that *his*A and *his*F have evolved via a series of gene duplication events (Fani et al. 1994). The first assumed tandem duplication and fusion of a small ancestral gene (pre-*his*A) of about 360 base pairs was postulated to generate *his*A, followed by a further tandem duplication and diversification of *his*A that led to generate *his*F. Alignments of the N- and C-terminal halves of *T. maritima* HisA and HisF and the presence of a phosphate-binding motif in both halves (Fig. 5), as well as alignments of the entire sequences of *T. maritima* HisA and HisF (Fig. 4), support the model of two sequential duplication events. Because of the proximity of *T. maritima* to the root of the phylogenetic tree, these duplications probably occurred early in evolution.

Both HisA and HisF have been predicted to belong to the class of $(\beta/\alpha)_8$-barrel proteins (Wilmanns and Eisenberg 1993, 1995). It is not clear which subfamilies of $(\beta/\alpha)_8$-barrel proteins have evolved from a common ancestor by divergent evolution (Farber and Petsko 1990) and which subfamilies arose by convergent evolution to a common fold (Brändén 1991; Hennig et al. 1992). A comparison of HisA and HisF with other $(\beta/\alpha)_8$-barrel enzymes, for example, those involved in tryptophan biosynthesis, may help to answer this question. When the two halves of phosphoribosyl anthranilate isomerase (TrpF), indole glycerol phosphate synthase (TrpC), and the α-subunit of tryptophan synthase (TrpA) were aligned, no significant sequence similarites were detected (data not shown). Also, two of these $(\beta/\alpha)_8$-barrel enzymes fold according to a "6 + 2" mechanism in which the folding of the first six (β/α) modules is followed by the folding and association of the last two (β/α) modules (Miles et al. 1982; Eder and Kirschner 1992). Taken together, the postulated duplication events in HisA and HisF and the absence of a tandem duplication in TrpF, TrpC, and TrpA, as well as the postulated "6 + 2" folding mechanism, are difficult to reconcile with the idea of a common ancestor of $(\beta/\alpha)_8$-barrel proteins, even for the putative subfamily HisA, HisF, TrpF, TrpC, and TrpA.

Determinants of the thermostability of *T. maritima* HisA and HisF

The structural basis of extreme protein thermostability is of scientific and biotechnological interest and has been investigated intensely during recent years (Jaenicke 1996). Early modeling studies (Perutz and Raidt 1975) and comparisons of highly resolved X-ray structures of thermolabile proteins with their thermostable orthologs indicate that, among other structural features, an increased number of polar interactions such as hydrogen bonds or salt bridges are important determinants of thermostability (Kelly et al. 1993; Yip et al. 1995; Korndörfer et al. 1995; Tanner et al. 1996; Macedo-Ribeiro et al. 1996; Lim et al. 1997; Vogt et al. 1997). Thermostable $(\beta\alpha)_8$-barrel enzymes of tryptophan biosynthesis contain an increased number of salt bridges compared to the corresponding enzymes from mesophiles (Hennig et al. 1995, 1997; Knöchel et al. 1996), a feature that had been predicted on the basis of a significantly higher number of charged amino acid residues in the thermostable variants (Sterner et al. 1995). *T. maritima* HisA and HisF are extremely thermostable (unpublished data).

To test whether thermostable variants of HisA and HisF might be stabilized by an additional number of salt bridges, a comparison of all available amino acid sequences in the SWISSPROT databank of HisA and HisF was performed (data not shown), and a significantly increased number of charged amino acid residues compared with the corresponding variants from mesophilic organisms were found (see Table 1). This result suggests that also thermostable enzymes of histidine biosynthesis are stabilized by an additional number of salt bridges. Highly resolved X-ray structures of both thermostable and thermolabile variants of HisA and HisF are needed to test this hypothesis.

Note added in proof. Since the acceptance of this paper, the X-ray structure of *T. maritima* His F has been solved. The structure is indeed a $(\beta\alpha)_8$-barrel with an internal twofold symmetry [Lang D, Obmolova G, Wilmanns M, personal communication].

Acknowledgments We thank Andrea Löschmann for DNA sequencing. This work was supported by grant 31.45853.95 of the Swiss National Science Foundation.

References

Achenbach-Richter L, Gupta R, Stetter KO, Woese CR (1987) Were the original eubacteria thermophiles? Syst Appl Microbiol 9:34–39

Alifano P, Rivellini F, Limauro D, Bruni CB, Caromagno MS (1991) A consensus motif common to all rho-dependent prokaryotic transcription terminators. Cell 64:553–563

Alifano P, Fani R, Liò P, Lazcano A, Bazzicalupo M, Calomagno MS, Bruni CB (1996) Histidine biosynthetic pathway and genes: structure, regulation, and evolution. Microbiol Rev 60:44–69

Beckler GS, Reeve JN (1986) Conservation of primary structure in the *his*I gene of the archaebacterium *Methanococcus vannielii*, the eubacterium *Escherichia coli* and the eucaryote *Saccharomyces cerevisiae*. Mol Gen Genet 204:133–140

Bork P, Gellerich J, Groth H, Hooft R, Martin F (1995) Divergent evolution of a β/α-barrel subclass: detection of numerous phosphate-binding sites by motif search. Protein Sci 4:268–274

Brändén C-I (1991) The TIM barrel – the most frequently occurring folding motif in proteins. Curr Opin Struct Biol 1:978–983

Broach JR (1981) Genes of *Saccharomyces cerevisiae*. In: Strathern JN, Jones EW, Broach JR (eds) The molecular biology of the yeast *Saccharomyces*. Life cycle and inheritance. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp 653–727

Broom JE, Hill DF, Hughes G, Jones WA, McNaughton JC, Stockwell PA, Petersen GB (1995) Sequence of a transposon identified as Tn1000. DNA Seq 5:185–189

Bult CJ, White O, Olsen GJ, Zhou L, Fleischmann RD, Sutton GG, Blake JA, Fitzgerald LM, Clayton RA, Gocayne JD, Kerlavage AR, Dougherty BA, Tomb JF, Adams MD, Reich CI, Overbeek R, Kirkness EF, Weinstock KG, Merrick JM, Glodek A, Scott JL, Geoghagen NSM, Weidman JF, Fuhrmann JL, Nguyen D, Utterback TR, Kelly JM, Peterson JD, Sadow PW, Hanna MC, Cotton MD, Roberts KM, Hurst MA, Kaine BP, Borodovsky M, Klenk HP, Fraser CM, Smith HO, Woese CR, Venter JC (1996) Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. Science 273:1058–1073

Carlomagno MS, Chiariotti L, Alifano P, Nappo AG, Bruni CB (1988) Structure and function of the *Salmonella typhimurium* and *Escherichia coli* K-12 histidine operons. J Mol Biol 203:585–606

Charlebois RL, Sensen CW, Doolittle WF, Brown JR (1997) Evolutionary analysis of the *his*CGAB$_d$FDEHI gene cluster from the archeon *Sulfolobus solfataricus* P2. J Bacteriol 179:4429–4432

D'Aubenton-Carafa Y, Brody E, Thermes C (1990) Prediction of rho-independent *Escherichia coli* transcription terminators. J Mol Biol 216:835–858

Delorme C, Ehrlich SD, Renault P (1992) Histidine biosynthesis genes in *Lactococcus lactis* susbsp. *lactis*. J Bacteriol 174:6571–6579

Dower WJ, Miller JF, Ragsdale CW (1988) High efficiency transformation of *E. coli* by high voltage electroporation. Nucleic Acids Res 16:6127–6145

Eder J, Kirschner K (1992) Stable substructures of eightfold βα-barrel proteins: fragment complementation of phosphoribosylanthranilate isomerase. Biochemistry 31:3617–3625

Fani R, Bazzicalupo M, Damiani G, Bianchi A, Schipani C, Sgaramella V, Polsinelli M (1989) Cloning of histidine genes of *Azospirillum brasilense*: organization of the ABFH gene cluster and nucleotide sequence of the *his*B gene. Mol Gen Genet 216:224–229

Fani R, Liò P, Chiarelli I, Bazzicalupo M (1994) The evolution of the histidine biosynthetic genes in prokaryotes: a common ancestor for the *his*A and *his*F genes. J Mol Evol 38:489–495

Fani R, Liò P, Lazcano A (1995) Molecular evolution of the histidine biosynthetic pathway. J Mol Evol 41:760–774

Farber GK, Petsko GA (1990) The evolution of α/β barrel enzymes. Trends Biochem Sci 15:228–234

Fink GR (1964) Gene-enzyme relations in histidine biosynthesis in yeast. Science 146:525–527

Goldschmidt EP, Cater MS, Matney TS, Butler MA, Greene A (1970) Genetic analysis of the histidine operon in *Escherichia coli* K12. Genetics 66:219–229

Gorman JA, Hu AS (1969) The separation and partial characterization of L-histidinol phosphatase and an alkaline phosphatase of *Saccharomyces cerevisiae*. J Biol Chem 244:1645–1650

Grisolia V, Ricco A, Bruni CB (1983) Structure and function of the internal promotor (hisBp) of the *Escherichia coli* K-12 histidine operon. J Bacteriol 155:1288–1296

Hennig M, Schlesier B, Dauter Z, Pfeffer S, Betzel C, Höhne WE, Wilson KS (1992) A TIM barrel protein without enzymatic activity? FEBS Lett 306:80–84

Hennig M, Darimont B, Sterner R, Kirschner K, Jansonius JN (1995) 2.0 Å structure of indole-3-glycerol phosphate synthase from the hyperthermophile *Sulfolobus solfataricus*: possible determinants of protein stability. Structure (Lond) 3:1295–1306

Hennig M, Sterner R, Kirschner K, Jansonius JN (1997) Crystal structure at 2.0 Å resolution of phosphoribosyl anthranilate isomerase from the hyperthermophile *Thermotoga maritima*: possible determinants of protein stability. Biochemistry 36:6009–6016

Hinnebusch AG, Fink GR (1983) Repeated DNA sequences upstream from HISI also occur at several other co-regulated genes in *Saccharomyces cerevisiae*. J Biol Chem 258:5238–5247

Jaenicke R (1996) Stability and folding of ultrastable proteins: eye lens crystallins and enzymes from thermophiles. FASEB J 10:84–92

Kawamura T, Ishimoto N, Ito E (1982) UDP-*N*-acetyl-D-glucosamine 2′-epimerase from *Escherichia coli*. Methods Enzymol 83:515–519

Kelly CA, Nishiyama M, Ohnishi Y, Beppu T, Birktoft JJ (1993) Determinants of protein thermostability observed in the 1.9-Å crystal structure of malate dehydrogenase from the thermophilic bacterium *Thermus flavus*. Biochemistry 32:3913–3922

Knöchel TR, Hennig M, Merz A, Darimont B, Kirschner K, Jansonius JN (1996) The crystal structure of indole-3-glycerol phosphate synthase from the hyperthermophilic archaeon *Sulfolobus solfataricus* in three different crystal forms: effects of ionic strength. J Mol Biol 262:502–515

Korndörfer I, Steipe B, Huber R, Tomschy A, Jaenicke R (1995) The crystal structure of holo-glyceraldehyde-3-phosphate dehydrogenase from the hyperthermophilic bacterium *Thermotoga maritima* at 2.5 Å resolution. J Mol Biol 246:511–521

Lim JH, Yu YG, Choi IG, Ryu JR, Ahn BY, Kim SH, Han YS (1997) Cloning and expression of superoxide dismutase from *Aquifex pyrophilus*, a hyperthermophilic bacterium. FEBS Lett 406:142–146

Limauro D, Avitabile A, Cappellano C, Puglia AM, Bruni CB (1990) Cloning and characterization of the histidine biosynthetic gene cluster of *Streptomyces coelicolor* A3(2). Gene (Amst) 90:31–41

Macedo-Ribeiro S, Darimont B, Sterner R, Huber R (1996) Small structural changes account for the high thermostability of 1[4Fe-4S] ferredoxin from the hyperthermophilic bacterium *Thermotoga maritima*. Structure (Lond) 4:1291–1301

Matney TS, Goldschmidt EP, Erwin NS, Scroggs A (1964) A preliminary map of genomic sites for F-attachment in *Escherichia coli* K12. Biochem Biophys Res Commun 17:278–281

Miles EW, Yutani K, Ogasahara K (1982) Guanidine hydrochloride induced unfolding of the alpha subunit of tryptophan synthase and of the two alpha proteolytic fragments: evidence for stepwise unfolding of the two alpha domains. Biochemistry 21:2586–2592

Ostendorp R, Liebl W, Schurig H, Jaenicke R (1993) The L-lactate dehydrogenase gene of the hyperthermophilic bacterium *Thermotoga maritima* cloned by complementation in *Escherichia coli*. Eur J Biochem 216:709–715

Perutz MF, Raidt H (1975) Strereochemical basis of heat stability in bacterial ferredoxins and in haemoglobin A2. Nature (Lond) 275:256–259

Reardon D, Farber GK (1995) The structure and evolution of α/β barrel proteins. FASEB J 9:497–503

Rost B, Sander C (1994) Combining evolutionary information and neural networks to predict protein secondary structure. Proteins 19:55–72

Sambrook J, Fritsch EE, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor, NY

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467

Sterner R, Dahm A, Darimont B, Ivens A, Liebl W, Kirschner K (1995) $(\beta\alpha)_8$-barrel proteins of tryptophan biosynthesis in hyperthermophile *Thermotoga maritima*. EMBO J 14:4395–4402

Struhl K (1985) Nucleotide sequence and transcritional mapping of the yeast pet56-his3-ded1 gene region. Nucleic Acids Res 13:8587–8601

Tanner JJ, Hecht RM, Krause KL (1996) Determinants of enzyme thermostability observed in the molecular structure of *Thermus aquaticus* D-glyceraldehyde-3-phosphate dehydrogenase at 2.5 Å resolution. Biochemistry 35:2597–2609

Vogt G, Woell S, Argos P (1997) Protein thermal stability, hydrogen bonds, and ion pairs. J Mol Biol 269:631–643

Wilmanns M, Hyde CC, Davies DR, Kirschner K, Jansonius JN (1991) Structural conservation in parallel β/α-barrel enzymes that catalyze three sequential reactions in the pathway of tryptophan biosynthesis. Biochemistry 30:9161–9169

Wilmanns M, Eisenberg D (1993) Three-dimensional profiles from residue-pair preferences: identification of sequences with β/α-barrel fold. Proc Natl Acad Sci USA 90:1379–1383

Wilmanns M, Eisenberg D (1995) Inverse protein folding by residue pair preference profile method: estimating the correctness of alignments of structurally compatible sequences. Protein Eng 8:627–639

Winkler ME (1996) Biosynthesis of histidine. In: Neidhardt FC, Curtiss R III, Ingraham JL, Lin ECC, Low KB, Magasanik B, Renikoff WS, Riley M, Schaechter M, Umbarger HE (eds) *Escherichia coli* and *Salmonella:* cellular and molecular biology, vol 1. American Society for Microbiology, Washington, DC, pp 485–505

Woese CR, Kandler O, Wheelis ML (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria and Eucarya. Proc Natl Acad Sci USA 87:4576–4579

Yip KS, Stillman TJ, Britton KL, Artymiuk PJ, Baker PJ, Sedelnikova SE, Engel PC, Pasquo A, Chiaraluce R, Consalvi V (1995) The structure of *Pyrococcus furiosus* glutamate dehydrogenase reveals a key role for ion-pair networks in maintaining enzyme stability at extreme temperatures. Structure (Lond) 3:1147–1158

Zalkin H, Smith JL (1998) Enzymes utilizing glutamine as an amide donor. Adv Enzymol Relat Areas Mol Biol 72:87–144