Phage Like It HOT: Solution Structure of the Bacteriophage P1-Encoded HOT Protein, a Homolog of the \(\theta\) Subunit of \(E.\ coli\) DNA Polymerase III

Eugene F. DeRose,1 Thomas W. Kirby,1 Geoffrey A. Mueller,1 Anna K. Chikova,2 Roel M. Schaaper,2 and Robert E. London1,*

1Laboratory of Structural Biology
2Laboratory of Molecular Genetics
NIEHS
Box 12233
Research Triangle Park, North Carolina 27709

Summary

DNA polymerase III, the main replicative polymerase of \(E.\ coli\), contains a small subunit, \(\theta\), that binds to the \(\epsilon\) proofreading subunit and appears to enhance the enzyme’s proofreading function—especially under extreme conditions. It was recently discovered that \(E.\ coli\) bacteriophage P1 encodes a \(\theta\) homolog, named HOT. The \(^1\text{H}-^{15}\text{N}\) HSQC spectrum of HOT exhibits more extreme conditions. It was recently discovered that \(E.\ coli\) DNA polymerase activity, encoded by the \(dnaQ\) gene, interface could provide significant insights into the biological properties of the interesting P1-encoded gene product. A solution structure of \(\theta\) has been reported based on NMR studies of the U-[\(^1\text{C}, ^{15}\text{N}\)]-labeled protein (Keniry et al., 2000). This structure exhibits a unique fold comprised of three \(\alpha\) helices in the N-terminal two-thirds of the protein, while the C-terminal one-third of the protein is disordered. The reported \(^1\text{H}-^{15}\text{N} \) HSQC spectra of \(\theta\) are characterized by uneven resonance intensities and significant evidence of exchange broadening (Li et al., 1999; Keniry et al., 2000). Keniry et al. (2000) also reported that at pH 6.5, \(\theta\) undergoes significant, presumably irreversible, aggregation within a 24 hr period at temperatures above 30°C, although a stable spectrum could be obtained for months at 4°C. The reported structure (PDB code: 1DU2) represents a unique fold in the protein database and includes three helical segments (Q10–D19, E38–E43, and H47–E54), a long loop (A20–E29), and a short extended segment (P34–A37). There are, however, a number of problematic and unusual characteristics of the reported structure. A Ramachandran plot generated by the program PROCHECK NMR

 Introduction

The catalytic core of \(E.\ coli\) DNA polymerase III holoenzyme is a complex of three subunits: \(\alpha\) (containing the DNA polymerase activity, encoded by the \(dnaE\) gene), \(\epsilon\) (containing the proofreading 3′→5′ exonuclease activity, encoded by the \(dnaQ\) gene), and \(\theta\) (encoded by the \(holE\) gene). These subunits are bound together in the linear order: \(\alpha \rightarrow \epsilon \rightarrow \theta\). Relatively little is known about the function of \(\theta\) (76 amino acids, 8.8 kDa) (Kelman and O’Donnell, 1995). The subunit is not essential, as a \(\Delta holE\) mutant is normally viable (Slater et al. 1994). Biochemical studies indicate that \(\theta\) does not affect the polymerase activity of \(\alpha\) or the \(\alpha\)-\(\epsilon\) complex. However, \(\theta\) was shown to have a modest stimulatory effect on the exonuclease activity of \(\epsilon\) on a mispaired primer terminus (Studwell-Vaughan and O’Donnell, 1993). A stimulatory effect of \(\theta\) on \(\epsilon\) was further indicated by genetic studies with the temperature-sensitive \(dnaQ49\) mutant (Taft-Benz and Schaaper, 2004). This mutant carries a V96G substitution in \(\epsilon\), causing a temperature-dependent proofreading defect. At low temperature (25°C), the mutant effect is modest, \(\sim 10\)-fold, but reaches very high levels, up to 10,000-fold, at 37°C. Interestingly, combination of \(dnaQ49\) with a \(\Delta holE\) mutation (i.e., in a strain lacking \(\epsilon\)) produces a 10,000-fold mutator effect even at the low temperature (Taft-Benz and Schaaper, 2004). These data have been interpreted to indicate that the intrinsically unstable DnaQ49 protein is critically dependent, at least at low temperatures, on the presence of \(\theta\) for its structural integrity. This conclusion is supported by NMR studies demonstrating that the binding of \(\theta\) to the catalytic domain of the \(\epsilon\) proofreading subunit, \(\epsilon 186\), results in significant changes in the \(^1\text{H}-^{15}\text{N} \) HSQC spectra of both proteins, as well as in significant stabilization of \(\epsilon 186\) (Li et al., 1999; DeRose et al., 2003).

Lobocka et al. (2004) recently reported that \(E.\ coli\) phase P1 encodes a \(\theta\) homolog, called HOT, encoded by the \(hot\) (homolog of theta) gene. Note that phase P1 (genome size 94 Kb) relies on the \(E.\ coli\) chromosomal replication machinery (DNA polymerase III holoenzyme) for its replication, and, with the exception of the \(ban\) DNA helicase and \(ssb\) single-stranded DNA binding protein, does not encode any other replication proteins (Lobocka et al., 2004). As discussed elsewhere (Chikova et al., 2004), we have constructed new strains of \(E.\ coli\) in which the \(holE\) gene is replaced by its phase P1 homolog, \(hot\). The construction was done in such a way that \(hot\) was expressed from the \(holE\) promoter. In a \(dnaQ49\) mutant carrying \(hot\) instead of \(holE\), mutability at low temperature is again low, indicating that \(hot\) can substitute for \(\theta\) in stabilizing the DnaQ49 protein (Chikova et al., 2004). Interestingly, \(hot\) appears to be more effective in stabilizing DnaQ49 protein than \(\theta\), because at all temperatures tested, the mutator activity of the \((\Delta holE)\) \(hot\) \(\) \(dnaQ49\) strain is significantly lower than for the \(holE\) \(\) \(dnaQ49\) strain (Chikova et al., 2004). This observation suggests that \(hot\) might bind \(\epsilon\) more tightly, or perhaps that \(hot\) is intrinsically more stable than \(\theta\).

Based on the above discussion, knowledge of the solution structure of HOT and the nature of the HOT-\(\epsilon\) interface could provide significant insights into the biological properties of the interesting P1-encoded gene product. A solution structure of \(\theta\) has been reported based on NMR studies of the U-[\(^1\text{C}, ^{15}\text{N}\)]-labeled protein (Keniry et al., 2000). This structure exhibits a unique fold comprised of three \(\alpha\) helices in the N-terminal two-thirds of the protein, while the C-terminal one-third of the protein is disordered. The reported \(^1\text{H}-^{15}\text{N} \) HSQC spectra of \(\theta\) are characterized by uneven resonance intensities and significant evidence of exchange broadening (Li et al., 1999; Keniry et al., 2000). Keniry et al. (2000) also reported that at pH 6.5, \(\theta\) undergoes significant, presumably irreversible, aggregation within a 24 hr period at temperatures above 30°C, although a stable spectrum could be obtained for months at 4°C. The reported structure (PDB code: 1DU2) represents a unique fold in the protein database and includes three helical segments (Q10–D19, E38–E43, and H47–E54), a long loop (A20–E29), and a short extended segment (P34–A37). There are, however, a number of problematic and unusual characteristics of the reported structure. A Ramachandran plot generated by the program PROCHECK NMR
Figure 1. CLUSTAL W Alignment of εP1 and HOT, Showing 53% Sequence Identity

Figure 2. 1H-15N HSQC Spectrum of U-[13C,15N]HOT

Assignments and Secondary Structure

The 1H-15N HSQC spectrum of HOT (Figure 2) exhibits better resonance dispersion and uniformity than the spectrum of εP1 (Li et al., 1999; Keniry et al., 2000). For the purpose of comparison, an HSQC spectrum of εP1 obtained under similar conditions is included as Supplemental Data (available with this article online). The HOT sample was found to be stable for weeks at a time under the conditions described below. We note, in particular, the importance of adding protease inhibitors to these samples in order to block the activity of contaminating proteases that may be present. All backbone amide chemical shifts, except that of Y2, were assigned. All Cα and Cβ chemical shifts, except those of M1, were assigned. All C' chemical shifts, except those of M1, L21, I34, Q45, L67, and D72, were assigned. The C' chemical shifts of I34, Q45, L67, and D72 were not assigned from the HNCO spectrum because these residues precede proline residues. It is not entirely clear why the cross-peak corresponding to the C' shift of L21 was not observed in the HNCO spectrum; however, cross-peaks involving the amide resonances of A22, A23, and S24 were also very weak in the HNCA, HNCA(CO), and CBCA(CO)NH spectra of HOT. Presumably, this region of the protein experiences conformational exchange leading to resonance broadening. It should also be noted that Keniry et al. (2000) could not

Results

Clustal W Alignment of HOT with εP1

Figure 1 shows the CLUSTAL W (Thompson et al., 1994) alignment of HOT with εP1. The overall sequence identity (residues indicated in red) is 53%, but it rises to ~70% if only the structured region of εP1 is considered. Helical regions of εP1 (Keniry et al., 2000) and HOT (present study) are underlined. As shown in the figure and discussed below, the lengths of helices α1 and α3 are significantly longer in HOT compared with εP1. Interestingly, Keniry et al. (2000) noted that D19 and E54 might not be the true termini of helices α1 and α3 due to conformational exchange broadening at the C termini of the two helices, which reduced the number of NOEs that could be observed and used to define the structure.
assign the amide resonances of A22 and A23, corresponding to A23 and S24, respectively, in HOT (see Figure 1). Due to the degeneracy of the Cα and Cβ chemical shifts of S24, the backbone chemical shifts of this residue were assigned from the HNCA, CBCA (CO)NH, and HNCO spectra. Most of the side chain chemical shifts, including aromatic and amide shifts, were assigned for all the residues, with the N-terminal and C-terminal residues being the few exceptions. These assignments were made as described in the Experimental Procedures.

A CSI analysis of the Hα, Cα, Cβ, and Cγ chemical shifts reveals helical regions extending from residues K16–K29, A38–A42, and T51–Q66. A TALOS analysis of the Hα, Cα, Cβ, and Cγ chemical shifts and the primary sequence yields φ and ψ torsion angles that are indicative of α helices for regions Q11–K29, A38–A42, and R50–Q66. The TALOS predictions are thus seen to be in very good agreement with the CSI analysis, with the exception of the length of the first α helix, i.e., K16–K29 predicted by CSI versus Q11–K29 predicted by TALOS. The longer N-terminal helix predicted by TALOS is supported by sequential dα(φ, i, i + 1), dα(ψ, i, i + 2), as well as dα(ψ, i, i + 3) NOEs between E12 and D15. In addition, the Cα chemical shifts also support the longer helix.

HOT Solution Structure

The solution structure of HOT was determined by using ARIA/CNS as described in the Experimental Procedures. The best structure of HOT was obtained by starting from a fully extended chain, including 430 manually assigned NOE[s], and using residual dipolar couplings (RDCs) measured in a medium containing bacteriophage f1 (Hansen et al., 1998). A similar structure was obtained by starting from a model that was homologous with the reported structure of ς (1DU2; [Keniry et al., 2000]) (see the Experimental Procedures). Prior to refinement with RDCs, a plot of the calculated versus the experimentally determined RDC values yielded a correlation coefficient of 0.55 by using residues 10–67 (n = 28), which exclude the flexible ends of the protein. The correlation coefficient improved to 0.99 with an R factor of 0.10 and a Q factor of 0.84 when the RDC restraints were included in the ARIA/CNS calculation (Clare and Garrett, 1999). Despite the significant improvement in the correlation coefficient, the structures obtained before and after the inclusion of the RDC data are very similar. The average structures obtained from the two calculations superimpose with a backbone rmsd of 1.05 Å. The main differences between the two structures are in loops L1 and L2 between the α helices and in the position of α2.

The solution structure of HOT (Figure 3) forms a three-helix bundle, with the N-terminal helix (Q11–R31, α1) sandwiched between a short second helix (A38–A42, α2) and the C-terminal helix (R50–Q66, α3). A short loop (L32–I37, L1) connects α helices α1 and α2, and a longer loop connects α helices α2 and α3. Residues M1–S10 and L67–K83 are disordered. A ribbon rendering of the backbone of the average structure, residues Q11–Q66, of the seven lowest-energy structures, showing the three α helices, is shown in Figure 3A, and a backbone superposition of the seven lowest-energy structures is shown in Figure 3B. The structures superimpose with backbone and heavy atom rmsds of 0.52 Å and 1.06 Å, respectively, for the structured region of the protein (Q11–Q66). The N-terminal helix α1 is oriented antiparallel to the second, short helix α2 (helix crossing angle of 149°) and parallel to the C-terminal helix α3 (helix crossing angle of 32°). The α2 helix crosses α3 at an angle of 158°.


Comparison of the HOT and ς Structures

The solution structure of HOT represents a completely different fold from that reported for ς. Figures 5A and 5B show the backbone ribbon rendering of the lowest-energy structure of HOT (residues Q11–Q66) and the best representative ς conformer (1DU2, model 1, residues Q10–N65) from the solution structure of Keniry et al. (2000). The structures superimpose with a backbone rmsd of 8.61 Å. It is clear from this figure and from the large backbone rmsd of the two structures that the solution structures of HOT and ς represent completely different folds. This conclusion is further supported by a DALI (Holm and Sander, 1993) search of the PDB (Berman et al., 2000) using the structured region (Q11–Q66) of the lowest-energy HOT structure. The DALI search failed to identify ς as a homolog of HOT and revealed that HOT has limited structural similarity to other proteins in the PDB. The match to HOT with the best Z score (2.7) is a three-helix stretch in the E chain of cytochrome c oxidase (2OCC-E, Yoshikawa et al., 1998). The backbone rmsd for this match is 3.6 Å. Other matches had even lower Z scores and are probably not significant. A DALI search with the best ς structure (i.e., 1DU2, model 1 [Keniry et al., 2000]) returns no matches with a Z score above 2.0.

The folding topologies of the HOT and ς structures (Figures 5A and 5B) show that ς is not as well packed as HOT. Both structures contain three α helices; however, α2 of ς does not pack against the other two helices, whereas α1 of HOT is sandwiched between α2 and α3 (Figures 4A and 4B). In the 1DU2 structure of ς, many of the hydrophobic residues on the two amphipathic helices—M13, V16, and L20 on α1 and L48 and F52 on α3—are not involved in the interhelical interface. In con-
trast with the solution structure of \( \theta \), a homology model of \( \theta \) based on the HOT structure (see the Experimental Procedures) contains a well-defined hydrophobic core.

Analysis of RDC Values Measured on U-[\( ^1\)C,\( ^15\)N]\( \theta \)
Despite the substantial sequence homology and similar determined secondary structure of \( \theta \) and HOT, the structure of HOT obtained in the present study exhibits a folding topology that differs significantly from that reported for \( \theta \) (Keniry et al., 2000). In principle, such structural variation could result from differences between the amino acid sequences of the two proteins or from differences in the solution conditions of the two studies. Alternatively, the substantial degree of exchange broadening that has been observed in the NMR spectra of \( \theta \) (a \( ^1\)H-\( ^15\)N HSQC spectrum of labeled \( \theta \) is included as Supplemental Data) could have led to errors in the structure determination. In order to gain some insight into the significance of the structural differences, we measured the RDC constants of U-[\( ^1\)C,\( ^15\)N]\( \theta \) in a sample containing bacteriophage \( \Phi 1 \). Since the \( ^1\)H-\( ^15\)N HSQC spectra that we obtained were very similar to those reported previously by Keniry et al. (2000), we transferred the amide assignments directly from Figure 2 of this reference, dropping any ambiguous assignments. We were able to obtain 30 couplings in the structured region from residues 10 to 69. Of the 30 couplings, 11 are in helix 1 (residues 10–30), 4 are in helix 2 (residues 37–41), and 7 are in helix 3 (residues 49–65), indicating that the data adequately span the secondary structure elements. These assignments also are generally in agreement with results of our own unpublished assignments of \( \theta \).

Figure 6 shows correlation plots of the measured RDC data, with the results calculated from two different \( \theta \) structures. The fitting was done according to procedures described by Yang et al. (1999). Briefly, the difference between the measured and calculated dipolar couplings is minimized by adjusting the orientation of the molecule (Euler angles \( \alpha \), \( \beta \), and \( \gamma \)) and the magnitude (\( A_\perp \)) and rhombicity (\( R \)) of the alignment tensor. Figure 6A is the best fit using the previously published 1DU2 structure (model 1 from the ensemble). The correlation between the calculated and measured RDC data is poor, with a correlation coefficient near zero (0.036), indicating essentially no correlation between the experimental and calculated values. In contrast, fitting the data to a homology model of \( \theta \) based on the HOT structure showed a considerably stronger correlation coefficient of 0.795 (Figure 6B). The high correlation is particularly impressive given that this model was “automatically” generated from the Swiss-Model web site with no further refinement on our part. These results strongly suggest that the solution structure of \( \theta \) is similar to that of HOT.

CD Analysis of HOT and \( \theta \)
The CD spectrum of HOT is very similar to the previously reported CD spectrum of \( \theta \) (Li et al., 1999; Keniry et al., 2000), with broad minima centered at 208 and 220 nm, typical for \( \alpha \)-helical proteins. Analysis of the temperature dependence of the ellipticity measured at 220 nm indicates a transition to lower (less negative) values as the temperature is increased, as would be anticipated to accompany protein unfolding. In the temperature range of 25°C–80°C, the data obtained for either protein were readily fit to a sigmoidal curve of the form:

\[
\theta = \theta_L + \frac{\Delta \theta}{1 + e^{\frac{T - T_m}{\frac{1}{\Delta T}}}} 
\]

(1)

Figure 4. Side Chain Heavy Atom Positions in the \( \alpha \) Helices of HOT
(A) Ribbon rendering (residues 11–66) of the lowest-energy structure of HOT showing the side chain heavy atoms of all residues in \( \alpha \) helices 1–3.
(B) Ribbon rendering rotated 180° about a vertical axis in the page, with respect to the view in (A).
Solution Structure of HOT, a $\theta$ Homolog

**Figure 7.** Circular Dichroism Spectra of HOT and $\theta$ as a Function of Temperature

The ellipticity at 220 nm was monitored for both HOT (circle) and $\theta$ (square) over the range of 26°C–80°C. Protein concentration was 2.1 $\mu$M in 10 mM NaP buffer (pH 6.5). The smooth curves correspond to the best fits of Equation 1 to the data. From the data fits, the melting temperatures for HOT and $\theta$ were determined as 62.1°C and 56.0°C, respectively. Where $R$ is the gas constant (Breslauer, 1995) and $\alpha$ is the normalized fraction of folded protein that is assumed to be proportional to the measured ellipticity. For a sigmoidal transition curve described by Equation 1, we then obtain:

$$\Delta H_m = RST_m (\text{K}).$$

Using the $S$ and $T_m$ parameter values obtained from the data fits shown in Figure 7, we obtain $\Delta H_m$ values of 42.2 and 38.6 kcal/mol for HOT and $\theta$, respectively. Thus, the temperature-dependent CD behavior indicates that HOT has greater thermal stability than $\theta$, characterized by both a higher $T_m$ and a greater $\Delta H_m$, although the difference is not very great.

**Electrostatic Surface and Possible $\epsilon$ Binding Interface**

Chemical shift mapping of the $\theta$ binding interface of $\epsilon$ demonstrates that the $\epsilon$-$\theta$ interface contains primarily hydrophobic residues (DeRose et al., 2003). These results suggest that hydrophobic regions on the surface of

$$\Delta H_m = -4R\alpha T \left( \frac{\Delta \theta}{T} \right),$$

where $\alpha$ is the ellipticity limit at low temperature, $\Delta \theta$ is the change that accompanies unfolding, $T_m$ is the melting temperature, and $S$ is the slope of the melting curve (Figure 7). Analysis of the data for HOT yielded a melting temperature of 62.1°C. Under the same concentration and buffer conditions, $\theta$ exhibits a slightly smaller ellipticity and a qualitatively similar temperature dependence, with $T_m = 56^\circ$C (Figure 7). For both proteins, the low-temperature CD spectrum was recovered upon lowering the temperature back to 25°C, indicating complete reversibility of the unfolding process. These data also allow an estimate of the enthalpy change for protein folding, $\Delta H_m$, which can be estimated from the relationship:

**Figure 6.** Comparison of Measured and Calculated Residual Dipolar Couplings for $\theta$

(A and B) Comparison of measured RDC constants with values (A) calculated by using model 1 from the ensemble published as 1DU2 (Keniry et al. 2000) or (B) calculated from a HOT-based homology model of $\theta$. The data correspond to 30 amide resonances that were unambiguously assigned based on a comparison of our HSQC spectrum with that reported by Keniry et al. (2000). (A) corresponds to a correlation coefficient of 0.036, while (B) yields a correlation coefficient of 0.795.
V36, I37, L63, and L65 (Figure 8A). This region of the both a higher Tm and a greater hydrogen bond.

HOT interact with $\theta$ and, in fact, appears to be more effective than $\theta$ at stabilizing the temperature-sensitive dnaQ49 mutant (Chikova et al., 2004). As mentioned above, this may result because HOT binds $\epsilon$ more tightly and/or because HOT is intrinsically more stable than $\theta$. In order to address this question, and to further elucidate the function of both $\theta$ and HOT, we have determined the solution structure of HOT.

HOT interact with $\epsilon$. An examination of the electrostatic surface of HOT (Figures 8A and 8B) shows two hydrophobic regions on the surface of the protein that may be involved in the interaction with $\epsilon$. A large contiguous stretch of hydrophobic residues on the front surface of the protein is comprised of residues Y28, L32, I34, P35, V36, I37, L63, and L65 (Figure 8A). This region of the protein includes loop L1 (L32–I37) between $\alpha$1 and $\alpha$2, which implies that L1 may interact with $\epsilon$. The second contiguous region of hydrophobic residues on the back surface of the protein includes residues V26, A38, A42, F53, M54, and L57 (Figure 8B). This region of the protein may also interact with $\epsilon$. Keniry et al. (2000) observed that in $\theta$ residues, L20–N32 of loop L1 became less dynamic upon binding $\epsilon$. These results also suggest that L1 of $\theta$ may be important for binding $\epsilon$. 

Discussion

$\theta$ subunit is the smallest component of the catalytic core of the E. coli replicative DNA polymerase III holoenzyme. Although the solution structure of $\theta$ (Keniry et al., 2000) and the X-ray structure of the catalytic domain of its partner protein, the $\epsilon$ proofreading exonuclease (Hamdan et al. 2002), have been determined, little is known about the function of $\theta$ within the DNA polymerase III core. Genetic experiments have suggested that $\theta$ may function to stabilize the intrinsically unstable $\epsilon$ protein (Taft-Benz and Schaaper, 2004). Recent chemical shift mapping of the $\theta$ binding interface of $\epsilon$ (DeRose et al., 2003) suggests that $\theta$ interacts with and probably stabilizes the positions of $\epsilon$ helix $\alpha$7 and the residues immediately preceding this helix. This region of the protein contains the catalytically important residues H162 and D167, thus providing a possible explanation for the observed $\theta$-mediated enhancement of the $\epsilon$ proofreading activity (Studwell-Vaughan and O’Donnell, 1993; Taft-Benz and Schaaper, 2004), particularly in the case of the unstable $\epsilon$ produced by the dnaQ49 mutant (Taft-Benz and Schaaper, 2004). The bacteriophage P1 hot gene encodes for a homolog of $\theta$ (Lobocka et al., 2004). It has been determined that HOT has a similar function to $\theta$ and, in fact, appears to be more effective than $\theta$ at stabilizing the temperature-sensitive dnaQ49 mutant (Chikova et al., 2004). As mentioned above, this may result because HOT binds $\epsilon$ more tightly and/or because HOT is intrinsically more stable than $\theta$. In order to address this question, and to further elucidate the function of both $\theta$ and HOT, we have determined the solution structure of HOT.

In the present study, we unexpectedly found that the $^1$H-$^1$N HSQC spectrum of HOT is subject to considerably less linewidth and intensity variation than the HSQC spectrum of $\theta$. This result supported further structural analysis, leading to the determination of a well-folded protein containing three $\alpha$ helices. Despite the general similarity in the secondary structure predicted from chemical shifts, the fold determined for HOT differs substantially from that previously reported for $\theta$ (1DU2 [Keniry et al., 2000]), as indicated, for example, by the fact that a DALI search failed to identify $\theta$ as a structural homolog of HOT. In comparison with the reported $\theta$ structure, $\alpha$1, $\alpha$2, and $\alpha$3 form a three-helix bundle, $\alpha$1 and $\alpha$3 are considerably longer, and HOT has a much more substantial hydrophobic core.

Consistent with both the biological results and with expectations based on a structural comparison, the temperature-dependent CD behavior indicates that HOT has greater thermal stability than $\theta$, characterized by both a higher $T_m$ and a greater $\Delta H_m$. For $\theta$, Keniry et al. (2000) did not assign 11 resonances (14%), and they noted that another 13 resonances (17%) exhibited significant broadening. The large fraction of problem resonances also is consistent with greater conformational instability of $\theta$ relative to HOT.

We considered it quite surprising that these two proteins, exhibiting a total sequence identity of 52%, and approaching 70% in the well-folded regions, exhibit significantly different folding topologies. Since the problems with the solution behavior and NMR spectra of $\theta$ are well documented (Li et al., 1999; Keniry et al., 2000), we considered it possible that the previously determined solution structure of $\theta$ was in error. In order to more fully evaluate this possibility, we expressed U-[13C,15N]$\delta$ and measured the RDC constants in a weakly orienting medium that included bacteriophage f1. Indeed, the use of RDC data for rapid protein fold recognition has been proposed by several groups (Valafar and Prestegard,
2003, Annila et al., 1999). The results were then compared with calculations using either the reported structure of \( \theta \) or a homology model of \( \theta \) based on the structure of HOT determined in the present study. The results make it clear that the solution structure of \( \theta \) is similar to that of HOT. The possibility that the two different structures for \( \theta \) could be the result of differences in solution conditions is also unlikely, due to the strong similarity in the HSQC spectrum that we obtained (Supplemental Data) and the previously reported spectrum (Keniry et al., 2000). Indeed, in order to minimize this potential source of error, we utilized the amide resonance assignments from the previous study, which are generally in agreement with our own (unpublished) results.

The conclusion that the solution structure of \( \theta \) is homologous with the HOT structure obtained here provides some insight into several of the spectral characteristics of \( \theta \). In particular, the previous study noted an assignment problem with residues L56–L61 (Keniry et al., 2000). Based on the sequence alignment of Figure 1, this is one of the more poorly conserved residue stretches in the well-structured region of HOT. In the homologous structure of \( \theta \), residues I57 and L61 extend into solution, leading to an energetically less favorable conformation. Such a structure would be expected to result in some conformational instability, leading to conformational exchange broadening and making the resonance assignments difficult. In contrast, the corresponding HOT residues, R58 and Q62, are more easily accommodated at these positions.

Interestingly, despite the major differences in structure between 1DU2-\( \theta \) and the HOT-based \( \theta \) homology model, many of the spectral characteristics noted by Keniry et al. (2000) apply to both structures. For example, in both structures, K15 lies against the indole ring of W51, resulting in an upfield shift of the lysine resonances. Consistent with either model, we observed an unusual inequivalence of the two C\( \alpha \) protons of K15. Similarly, the V18 side chain is predicted to experience a downfield shift due to its location near the edges of W51 and F52 in the homology model. According to the homology model, the Y31 hydroxyl group can hydrogen bond with the A62 carbonyl oxygen and the N65 backbone amide, consistent with the slow hydroxyl proton exchange observed by Keniry et al. (2000).

In contrast, Y31 is located far from E43 and Q44 in the homology model, so we are unable to explain the NOE interactions noted by Keniry et al. (2000). In this context, we note the close proximity of Y31 to E29 in the HOT-based \( \theta \) homology model.

The structure obtained here for HOT provides useful insights into the biological function of this protein; a comparison of the structure of the uncomplexed HOT with the \( \epsilon_{186}\)-HOT complex is currently in progress. The results of studies with weakly oriented \( \theta \) strongly support the conclusion that the solution structure of \( \theta \) is homologous to HOT. Since a direct determination of the \( \theta \) structure is made difficult by its dynamic properties, a determination of the structure of the \( \epsilon_{186}\)-\( \theta \) complex should represent a more attainable goal. Keniry et al. (2000) observed that residues in the region from L20 to N32 of \( \theta \) became less dynamic upon binding \( \epsilon \), and they postulated that these residues interact with \( \epsilon \). Preliminary chemical shift mapping studies of U-\( ^{15}\)C\( ,^{15}\)N\( ^{-}\)HOT complexed with unlabeled \( \epsilon_{186} \) indicate many significantly shifted residues, with the largest shifts observed for the N terminus, helix \( \alpha_1 \), and a region that includes the C-terminal portion of helix \( \alpha_3 \). Some of the shifted \( \alpha_1 \) residues of HOT correspond to the region identified previously by Keniry et al. (2000). The observed residue differences may indicate the different requirements for optimal \( E. \) coli and phage replication. Further studies on the fidelity and processivity of \( \theta \) and HOT-containing DNA polymerase III should provide insight into these differences.

Experimental Procedures

Cloning and Expression

Phage P1 DNA was obtained from an \( E. \) coli strain lysogenic for P1 phage \( \phi \)1.100 Tn9 (Lobocka et al., 2004) by using a Qiagen Lambda kit (Qiagen, Inc.). The hot gene was recovered by PCR with two primers, one complementary to the beginning of the gene (starting at the ATG start codon), and the second complementary to the end of the gene (ending at the TAG stop codon). In addition, the primers created 5' or 3' extensions containing NdeI or BamHI recognition sequences, respectively, enabling insertion of the hot coding sequence into the NdeI-BamHI site of vector pETc3 (Novagen, Inc.). The resulting plasmid, pET3cHot, carried a 252-nt hot gene sequence, encoding 83 amino acids, corresponding to nucleotides 85,380–85,631 of the consolidated \( P1 \) genomic sequence (Lobocka et al., 2004). Plasmid pET30aHOT was then created by transferring the NdeI-BamHI fragment of pET3cHot into pET30a(+) (Novagen, Inc.). The pET30aHOT construct was transformed into \( E. \) coli strain BLR (DE3), and cells containing the plasmid were grown to mid-log phase (\( A_{600} \approx 0.6 \)) at 37°C in M9 minimal medium containing 30 \( \mu \)g/ml kanamycin, \( ^{15}\)N-labeled ammonium chloride, and either \( ^{13}\)C- or \( ^{15}\)N-labeled or -unlabeled glucose. HOT expression was induced by the addition of isopropyl-\( \beta \)-D-thiogalactopyranoside (IPTG) to 1 mM, and growth continued for 3 hr. The cells were harvested by centrifugation and were stored at \(-20^\circ \)C.

Protein Purification

The frozen cell pellet was thawed and resuspended in 50 mM Tris (pH 7.5), 1 mM EDTA, and 75 mM NaCl, and cells were lysed by sonication in a Branson Sonifier 200 with a microtip probe at an output level of 6 for 10 × 30 s with 30 s cooling. The lysate was centrifuged at 30,000 × g for 15 min; then, solid ammonium sulfate was added to the supernatant to 30% saturation. The sample was kept on ice for 30 min and then centrifuged at 30,000 × g for 10 min. The supernatant was brought to 55% saturation with the addition of solid ammonium sulfate and was kept, again, on ice for 30 min and centrifuged as described before. The 55% ammonium sulfate precipitate, containing HOT, was resuspended in a minimal amount of 50 mM MES (pH 5.5), and the protein solution was applied to a 2.6 × 83 cm column of Sephacryl S-100, which was eluted with 50 mM MES (pH 5.5). Fractions containing HOT were identified by SDS polyacrylamide gel electrophoresis and were pooled and loaded onto an FPLC column (2.6 × 15 cm) containing SP-Sepharose (Amersham). The column was washed with 50 mM MES (pH 5.5), and HOT was eluted with a linear gradient of 0–1,000 mM NaCl in 50 mM Tris (pH 8.0). Fractions containing HOT were identified by SDS polyacrylamide gel electrophoresis, pooled, and concentrated by using a Centricon YM-3 filter unit (Millipore). The NMR samples typically contained 1.3 mM HOT protein in 5 mM Tris-\( \beta \)-HCl (pH 7), 100 mM NaCl, 5 mM NaN\(_3\), 10% D\(_2\)O, and 0.5 \( \mu \)M Calbiochem Protease Inhibitor Cocktail Set III per 600 ml sample.

U-\( ^{15}\)C\( ,^{15}\)N\( ^{-}\)Hot was expressed as described by Li et al. (1999) and was purified by following the same procedure as that described above for HOT.

Circular Dichroism Measurements

Circular dichroism studies were performed with a Jasco 810 Spectropolarimeter equipped with a Peltier thermal controller. Both the

Solutions Structure of HOT, a Homolog
and HOT samples were at 2.1 \mu M concentration and were prepared in 10 mM NaP buffer (pH 6.5). For high-quality data from 190 to 260 nM, a 1 nm path length cell was used. For temperature ramping experiments, a 1 cm path length cell was used, and the ellipticity at 220 nm was monitored from 26°C to 80°C. The rate of temperature increase was 1°C per minute, and the ellipticity was sampled every 2°C. The samples were then allowed to equilibrate to room temperature with no thermal control before another spectrum was acquired.

Homology Modeling
The homology models were constructed by using Swiss-PDBViewer and Swiss-Model (Gueux and Peitsch, 1997), based on the CLUSTAL W (Thompson et al., 1994) alignment shown in Figure 1. We used the published structure of φ (1DUZ, model 1 [Kenry et al., 2000]) to generate the φ-based HOT homology model and used the lowest-energy structure of HOT obtained here (PDB code 1SE7) to generate the HOT-based φ homology model.

NMR Studies
The NMR experiments were performed on a Varian 600 MHz UNI- TYPNOVA spectrometer, using a 5 mm Varian H{13C,15N} triple resonance probe with actively shielded z axis gradients and variable temperature capability. All NMR experiments were carried out at 25°C. The NMR data were processed with NMRPipe (Delaglio et al., 1995), and the spectra were analyzed with NMRView (Johnson and Blevins, 1994).

The sequential backbone and Cα resonances were assigned from a combined analysis of HNCA (Ikura et al., 1990; Matsuo et al., 1996), HNCA/CB (Wittekind and Mueller 1993; Muhundiram and Kay, 1994), CBCA(CO)NH (Grzesiek and Bax, 1992; Muhundiram and Kay, 1994), and HNCO (Ikura et al., 1990; Kay et al., 1994) experiments. The HNCA, HNCA/CB, CBCA(CO)NH, and HNCO experiments were acquired by using Varian’s ghn_ca, ghn_cacb, gcbca_co_nh, and ghn_co ProteinPack sequences, respectively. Most side chain proton and carbon chemical shifts were assigned from a combined analysis of H(CO)NH-TOSCY and (H)(CO)NH-TOSCY (Grzesiek et al., 1993; Logan et al., 1992, 1993; Montelione et al., 1992) experiments, which were acquired by using the pulse sequences described by Gardner et al. (1996), obtained from Lewis Kay. An isotropic 13C mixing time of 18.3 ms was used for both TOCSY experiments. The side chain chemical shifts of residues preceding proline residues were assigned from a combined analysis of 3D HCCH-TOSCY (Bax et al., 1990; Kay et al., 1993) and 3D 15N-edited NOESY (Marion et al., 1989; Zhang et al., 1994) experiments. The HCCH-TOSCY experiment was acquired using an isotropic 13C mixing time of 15.6 ms, and the 3D 15N-edited NOESY experiment was acquired using a mixing time of 100 ms. The HCCH-TOSCY and 3D 15N-edited NOESY experiments were acquired by using Varian’s hcch_tocsy and gnoesyNhsqc ProteinPack sequences, respectively. Phenylalanine and tyrosine H5, H6, H7, and C5 and tryptophan H11 and C11 chemical shifts were assigned from a combined analysis of (HB)CB(CGCD)HD, (HB)CB(CGCDCE)HE (Yamazaki et al., 1993), and 1H-13C HSQC experiments. The remaining tryptophan aromatic resonances were assigned from a 1H-15C HSQC spectrum. The side chain amide chemical shifts of asparagine and glutamine residues were assigned from the 3D 15N-edited NOESY spectrum and from the appearance of side chain amide cross-peaks in the HNACB spectrum. All chemical shift assignments have been deposited with the BMRB (University of Wisconsin-Madison, 2004), entry number 6127.

NOE cross-peaks were assigned from 3D 15N- and 13C-edited NOESY-HSQC spectra. The 3D 15N-separated NOE experiment is described above. The 3D 13C-edited NOESY-HSQC experiment was acquired with a mixing time of 100 ms, by using a new version of the CN-NOESY-HSQC experiment (Pascal et al., 1994) obtained from Lewis Kay. In this experiment, the 13C carrier frequency was set to 67.0 ppm in order to allow observation of NOEs to the aromatic protons as well as the aliphatic protons. This greatly facilitates the assignment of NOEs involving the aromatic protons.

### Table 1. Statistics for the HOT Structure Ensembles

<table>
<thead>
<tr>
<th>Ensemble</th>
<th>No. of NOE Cross-Peaks</th>
<th>Average No. of NOEs per Pair</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unambiguous</td>
<td>1140</td>
<td>435</td>
</tr>
<tr>
<td>Ambiguous</td>
<td>1575</td>
<td>46</td>
</tr>
<tr>
<td>Totala</td>
<td>1155</td>
<td>51</td>
</tr>
</tbody>
</table>

### Measurement of Residual Dipolar Coupling Constants

In order to measure residual dipolar couplings (RDCs), a sample of HOT was prepared with the addition of bacteriophage f1 cosolvent (Asla Biotech). Approximately 5 mg of the f1 was diluted 1:10 with the NMR buffer and pelleted (5°C, 95,000 rpm in a TL100.2 rotor). The supernatant was removed, and 0.5 ml of protein in NMR buffer was added. The splitting of the HDO resonance was 10.6 Hz, indicating that the sample was about 10 mg/ml pf1 (Hansen et al., 1998).

For measurements on φ, bacteriophage f1 was diluted 1:10 into the standard buffer (10 mM NaP, pH 6.5), 5 mM sodium azide, 10 mM EDTA, 10% D2O, 100 μM AEFBS, and 0.5 μl of cold benzyl alcohol and then incubated with 3000 units of apo-phi2 and 1000 units of apo-φ1 (Asla Biotech). Approximately 100 ml of NaCl. The additional salt was found to be necessary to prevent aggregation of the phage particles and the HOT structure. The phage was pelleted as described above, the supernatant was discarded, and the HOT-based φ homology model was used to generate the HOT-based φ homology model.
pellet, and the phage was resuspended. This procedure was repeated twice to increase the f1 concentration. The approximate final concentration of f1 was 16 mg/ml, as evidenced by the 16 Hz splitting of the H signal of HDO. The HSQC spectra of in the buffer with extra salt and pI1 showed minimal perturbations compared to the isotropic HSQC spectrum, which was very similar to that reported previously (Li et al., 1999; Keniry et al., 2000).

 Backbone amide RDCs were measured by using Varian’s gHsqc, IPAP, IPAP H1NHQC experiment (Ottiger et al., 1998). The J couplings were compared in the isotropic and anisotropic (phage) solutions, and the RDC is the difference in J couplings (LJ) between the two solutions. The range of RDCs measured in HOT was 38 to 17 Hz for the 55 residues with good dispersion. In 13 cases, it was difficult to determine the peak centers for the upfield peak in the anisotropic phase, possibly due to crosscorrelation effects likely caused by strong interactions with the phage. For these residues the J coupling in the anisotropic phase were measured according to two times the difference in Hz between the downfield peak and the peak in the HSQC spectra. For the other 42 peaks, this alternative method gave identical coupling results; the correlation between measuring the J couplings both ways was 0.99. The digital resolution of the processed data sets was 1.8 Hz/point, and the error in measurement used to fit to the structures and in structure calculations was 0.9 Hz. For the RDCs measured for II, the range of couplings was 22 to 9 Hz.

The H1NH1HSQC spectrum of II was assigned by visual comparison with that published in Figure 2 of Keniry et al. (2000). Residues with any overlap or any question as to the assignment were discarded from further analysis in order to be as conservative as possible in transferring the assignments. In the end, 30 residues were confidently assigned in the structured region (residues 10–69) of the protein.

Structure Calculations

Structure calculations were carried out by using ARIA (Nilges, 1995, 1997; Nilges and O’Donoghue, 1998; Linge and Nilges, 1999), version 1.2, and CNS (Brunger et al., 1998), version 1.1, as described previously (DeRose et al., 2003), with the following modifications. In order to obtain convergence to a unique structure, 430 unique manually assigned NOE distance restraints were included in the calculation. Additional stereospecific assignments of the manually assigned NOEs were made by ARIA (Table 1). Distance restraints from the manually assigned NOEs were calibrated by setting the appropriate qshift and qexclude flags to false. Two calculations were performed. The first calculation (calculation 1) used no RDC restraints, while the second calculation (calculation 2) used 36 backbone amide RDC restraints. RDC restraints from the disordered N-terminal (residues 1–10) and C-terminal (residues 67–83) regions of the protein were not used in the second calculation. The molecular alignment tensor was initially estimated by fitting the RDCs to the NMR structures calculated without RDCs (Yang et al., 1999). The initial fitted values did not change after refinement with the RDCs. The final result was Da = 20.9 Hz and R = 0.45. The ensemble of structures computed with the RDC restraints had the lowest rmsd (Table 1) and was taken to give the most accurate representation of the solution structure (see below). In addition to the NOE and RDC distance restraints, 46 hydrogen bond restraints inferred from CSI (Wishart et al., 1995) and TALOS (Cornilescu et al., 1999) predictions of the locations of the helices and a total of 51 φ and ϕ dihedral angle restraints were also used in all of the calculations. The dihedral angle restraints were set as described previously (DeRose et al., 2003).

Since the solution structure of HOT was found to differ significantly from the solution structure of II, and since HOT has 53% sequence identity with II (Figure 1), an additional ARIA calculation (calculation 3) was carried out starting from a homology model structure based on the best II model (i.e., 1DU2, model 1 [Keniry et al., 2000]). This calculation was carried out with the same restraints as the above calculation, including the RDC restraints, but the qshift and qexclude flags were set to true for all the NOE peaks; thus, all NOE distance restraints were assigned by ARIA by using chemical shift assignments. In this calculation, the NOE violation tolerances were set to 0.5, 0.5, 1.0, 0.5, 0.1, 2.5, 0.1, 0.1, and 0.1 for iterations 0–8, and the partial NOE assignment cutoff probabilities were set to 0.95, 0.95, 0.95, 0.93, 0.92, 0.91, 0.90, and 0.80. In our experience, these parameters typically produce structures with folds similar to the initial starting structures.

Ensembles of 20 structures of HOT were generated for each of the three ARIA/CNS calculations described above. These calculations correspond to (1) the use of an initial random extended structure and 430 manually assigned NOEs, with the remainder assigned by ARIA; (2) the same as the first calculation, with the additional inclusion of RDC data; and (3) the same as the second calculation, but starting from an initial HOT model structure generated as a homolog of the reported structure of II (1DU2 [Keniry et al., 2000]) and allowing full reassignment of NOEs by ARIA. Since calculation 2, using backbone amide RDC restraints and starting from a random extended structure, generated structures with the lowest rmsds in the best agreement with the measured RDCs, we have taken this ensemble of structures to be the best representation of the solution structure of HOT. The structural statistics for the seven lowest-energy structures for calculation 2 are shown in Table 1.

Supplemental Data

Supplemental Data including a H1NH-HSQC spectrum of U14C, U15N) obtained at 14.1 T indicating the resonances used for the RDC analysis are available at http://www.shu.org/cgi/content/full/12/12/2221/DG1/.

Acknowledgments

The authors are grateful to Dr. Robert Petrovich of the NIEHS protein expression laboratory for help with the CD studies and to Dr. Yuan Chen, Beckman Research Institute, City of Hope; Dr. Gary Pielak, Dept. of Chemistry, University of North Carolina; and Dr. Dawei Li, Dept. of Structural Biology, St. Jude Children’s Research Hospital for several helpful discussions.


**Accession Numbers**

The HOT structures have been deposited in the Protein Data Bank under the accession number 1SE7.