Regulation through the RNA Polymerase Secondary Channel

STRUCTURAL AND FUNCTIONAL VARIABILITY OF THE COILED-COIL TRANScription FACTORS*§

Received for publication, October 4, 2005, and in revised form, October 31, 2005.

Published, JBC Papers in Press, November 18, 2005, DOI 10.1074/jbc.C500405200

Jindrich Symersky 1, Anna Perederina 3, Marina N. Vassylyeva 3, Vladimir Svetlov 1, Irina Artsimovitch 1, and Dmitry G. Vassylyev 2

From the 1Department of Biochemistry and Molecular Genetics, University of Alabama at Birmingham, School of Dentistry, Birmingham, Alabama 35294 and the 2Department of Microbiology, The Ohio State University, Columbus, Ohio 43210

Gre factors enhance the intrinsic endonucleolytic activity of RNA polymerase to rescue arrested transcription complexes and are thought to confer the high fidelity and processivity of RNA synthesis. The Gre factors insert the extended α-helical coiled-coil domains into the RNA polymerase secondary channel to position two invariant acidic residues at the coiled-coil tip near the active site to stabilize the catalytic metal ion. Gfh1, a GreA homolog from Thermus thermophilus, inhibits rather than activates RNA cleavage. Here we report the structure of the T. thermophilus Gfh1 at 2.4 Å resolution revealing a two-domain architecture closely resembling that of GreA. However, the interdomain orientation is strikingly distinct (~162° rotation) between the two proteins. In contrast to GreA, which has two acidic residues on a well fixed stabilized-α-turn, the tip of the Gfh1 coiled-coil is flexible and contains four acidic residues. This difference is likely the key to the Gre functional diversity, while Gfh1 inhibits exo- and endonucleolytic cleavage, RNA synthesis, and pyrophosphorolysis, GreA enhances only the endonucleolytic cleavage. We propose that Gfh1 acidic residues stabilize the RNA polymerase active center in a catalytically inactive configuration through Mg²⁺-mediated interactions. The excess of the acidic residues and inherent flexibility of the coiled-coil tip might allow Gfh1 to adjust its activity to structurally distinct substrates, thereby inhibiting diverse catalytic reactions of RNA polymerase.

Bacterial RNA polymerase (RNAP) catalyzes four chemical reactions: nucleotide addition and its reverse reaction, pyrophosphorolysis (1), as well as two transcript cleavage reactions, exo- and endonucleolytic digestion of the nascent RNA (2, 3). All these reactions are thought to occur in the same active site (AS) and conform to the universal two Mg²⁺-mediated mechanism (4).

The RNAP AS is connected to the enzyme surface by the secondary channel, through which diverse regulatory molecules gain access to the AS and modulate transcript cleavage and elongation. Among these, the bacterial factors GreA, GreB, and DksA all employ a long α-helical coiled-coil (CC) domain protruding through the secondary channel to target functionally crucial Mg²⁺ ions near the RNAP AS with the two invariant acidic residues at the CC tips (8–10, 12). Despite this structural similarity, the functions of these proteins are strikingly distinct; the Gre factors directly remodel the RNAP AS to stimulate the endonucleolytic cleavage of the RNA (8, 10, 12), whereas DksA amplifies the activity of the “magic spot” (ppGpp), the regulator of stringent response in bacteria, but has no direct effect on catalysis (9, 13).

The Thermus thermophilus genome does not encode an ortholog of GreB, instead it contains a gene for a unique GreA paralog, Gfh1 (14). Despite a significant sequence conservation including the two mechanistically important acidic residues (Fig. 1A), this Gre-like factor does not stimulate but instead inhibits the intrinsic and GreA-mediated endonucleolytic RNA cleavage (14) as well as the RNA synthesis (15).

To elucidate the molecular mechanism of Gfh1 and the structural features that determine functionality of the CC transcription factors, we have determined the crystal structure of the T. thermophilus Gfh1 at 2.4 Å resolution. The structural and functional analysis suggests that while CC protein regulators appear to utilize similar structural motifs to gain access to the RNAP AS, their dramatically different effects may result from relatively subtle changes in the amino acid composition and conformation of the tip of the CC-domain.

MATERIALS AND METHODS

Crystalization, Structure Determination, and Refinement—The Gfh1 protein was expressed and purified as described (16). The crystals were obtained by a sitting drop vapor diffusion technique followed by macro-seeding (16). The final well crystallization solution contained 3% polyethylene glycol 8000, 33 mM zinc acetate, and 17 mM sodium cacodylate, pH 6.5. The crystals belong to the P4₃ space group with unit cell dimensions a = b = 59.3, c = 218.9 Å and possess perfect merohedral twinning with the twinning operator (h,k,l). The data for the native protein and three heavy atom derivatives were collected at 100 K on the in house x-ray generator (Rigaku) using a RAXIS-IV+ imaging plate detector and were processed with the HKL2000 program package (17) (supplemental Table 1). The initial phases were obtained by the MLPHARE program and were improved by solvent flattening using the DM program (18). The model was built manually using the O program (19) and refined with the CNS program (20) to a final R-factor of 20.6% (Rmod = 24.5%) at 2.4 Å resolution (supplemental Table 1).

In Vitro Transcription Reactions—Linear DNA template generated by PCR amplification (40 nM), T. thermophilus holo-RNAP (50 nM), and ApU (100 μM; Sigma) were mixed in 45 μl of TGA buffer (20 mM Tris acetate, 20 mM sodium acetate, 2 mM magnesium acetate, 5% glycerol, 14 mM 2-mercaptoethanol, 0.1 mM EDTA, pH 8.0) and incubated for 5 min at 55 °C. ATP, GTP, and UTP (Amersham Biosciences) were added to 2.5 μM, followed by incubation for 5 min at 37 °C and purification through G50 spin columns (Amersham Biosciences) equilibrated in TGA. The obtained A26 transcription elongation complexes (TECs) were incubated with 300 nM 16S rRNA (PerkinElmer Life Sciences) for 3 min at 37 °C to form C27 TECs. C27 TECs were incubated at conditions indicated in the legend to Fig. 2. Samples were analyzed as described previously (21).

RESULTS

The Gfh1 Structure—The Gfh1 structure (Fig. 1B) comprises two domains, globular (G) and CC, which closely resemble the corresponding domains of GreA (22) (root mean square deviation are 1.1 and 0.85 Å for the G- and CC-domains, respectively). A high quality of the experimental electron density allowed unambiguous modeling of all 156 Gfh1 residues in each of the four molecules in the asymmetric unit of the crystal. The only exception is a linker loop (residues 41–45) intervening the α-helices at the CC tip (Fig. 1A and B), which was represented by the relatively weak experimental electron density and exhibited high B-factors in a final model, suggesting that this segment is flexible and therefore might adopt alternative conformations. The four independent Gfh1 monomers are arranged as a pair of nearly identical dimers (supplemental Fig. S1), in which the individual molecules are related by a perfect rotational symmetry, suggesting that the dimers might have some physiological relevance. However, the dimers are stabilized almost exclusively through Zn²⁺-mediated interactions and can represent a crystallization artifact given the high concentration of zinc acetate (33 mM) used for crystallization. On the other hand, judging by 25 well fixed Zn²⁺ ions predominantly bound to the acidic side chains, Gfh1 apparently has affinity for the divalent
**ACCELERATED PUBLICATION: Crystal Structure of the Transcription Factor Gfh1**

**Comparison with GreA and DksA**—Despite their high structural similarity, there are two essential differences between the Gfh1 and GreA structures. First, the interdomain orientations are strikingly distinct (~162° rotation), resulting in completely different orientations of the CC-domains and the interdomain interfaces (Fig. 1C). Interestingly, the Gfh1 interacting surface is more extensive (1238 Å² via 1030 Å² in GreA) and involves mostly hydrophobic interactions, whereas the GreA interface is determined exclusively by hydrogen bonding (supplemental Fig. 2). The Gfh1 structural organization is reminiscent of that of DksA, whereas its G- and CC-domains also form a common hydrophobic core that, similarly to Gfh1, masks a big portion of the CC-domain surfaces (Fig. 1B). Given the potential functional importance of the acidic residues at the CC tip, it is likely that the acidic side chains of Gfh1 are proximal to each other (Fig. 1D). The hydrogen bonds stabilizing the α-turns in GreA, and DksA are shown as white dashed lines.

**Gfh1 Inhibits All Catalytic Activities of RNAP**—The reactions catalyzed by RNAP utilize different substrates and thus likely have different requirements for the "optimal" configuration of the AS. Indeed, alterations in the AS geometry brought by the substitutions of individual Asp residues inhibit different polymerization and cleavage reactions to different extent (23). Similarly, protein factors, like GreA and DksA, may remodel the AS to preferentially impact one but not other activities, thus exerting a specific regulatory effect. In contrast, Gfh1 was shown to inhibit two distinct catalytic reactions (14). To test whether Gfh1 would have a similar effect on all types of catalytic reactions in a single TEK, we prepared C27 TEKs that were uniquely labeled at their 3′-terminal C residue (Fig. 2).

C27 RNA was extended upon addition of GTP and UTP; at low NTP concentrations used here, transcription arrests at U38. C27 RNA was also degraded in the presence of PPi, releasing CTP (which co-migrated with the endonucleolytic activity that leads to the release of short 3′-labeled RNA fragments (cleavage products) and the ability to scavenge trace nucleotides to extend C27 RNA (extension products) became pronounced. Similarly to published studies (14, 15), we found that Gfh1 inhibits all RNAP reactions, the exonucleolytic cleavage is particularly sensitive to inhibition. This selectivity may be relevant to the physiological role of Gfh1, which remains obscure.
ACCELERATED PUBLICATION: Crystal Structure of the Transcription Factor Gfh1

The sequence and structural homology between Gfh1 and GreA and the observation that Gfh1 affects all intrinsic (factor-independent) catalytic activities of RNAP (Fig. 2) suggests that Gfh1 utilizes the Gre/DksA-like structural mechanism of action: protruding the CC-domain through the secondary channel toward the RNAP catalytic center, where the CC-tip acidic residues remodel the AS (perhaps through Mg\(^{2+}\)-mediated interactions with the catalytic residues), thereby modulating the RNAP activities.

A low resolution (15 Å) electron density (12) demonstrates that the RNAP secondary channel accommodates the GreB CC-domain but does not allow one to distinguish between the alternative G-domain orientations: the GreA and Gfh1 G-domains would likely fit equally well to the same electron density if the proteins are superimposed by their CC-domains (supplemental Fig. S3). The published models for RNAP-Gre factor interactions agree on the overall topology of the complex but differ in their details (8, 10, 12), and substitutions in RNAP that directly affect the Gre factors binding have not been reported. The sequence analysis and structural modeling show that with the exception of a few unfavorable amino acid substitutions (Fig. 1A), which likely make the observed conformations preferential for the particular protein in its free form, the residues forming the interdomain interfaces in Gfh1 and GreA are conserved among both protein families. At the same time the interdomain contacts in both proteins are not very extensive and could be disrupted under certain conditions to allow a “flip” between the two observed orientations. In principle, either of the two conformations might be induced for each protein in the course of its recruitment to RNAP. These considerations suggest two alternative modes of Gfh1/GreA binding to RNAP. First, the proteins may maintain their free form conformations. Assuming that in both cases the G-domains contain the major binding determinants (8, 24), while the CC-domains would be positioned essentially similar given the restrictions imposed by the size and shape of the secondary channel on their orientations, this scenario implies distinct RNAP binding modes for Gfh1 and GreA (they may utilize different structural elements to bind to the same site on RNAP or bind to different sites on RNAP); consistently, the G-domains surfaces exposed for the potential interaction with RNAP differ both in shape and sequence. Alternatively, both proteins may adopt one of the observed conformations induced upon binding to RNAP and in this case would likely bind to the same site on RNAP in a similar fashion; this also seems possible in light of their sequence and apparent structural similarity. Interestingly, changes in the relative positions of the two GreB domains upon binding to RNAP and TECs were detected by hydroxyl radical footprinting analysis (24).

Structural comparison of the Gfh1, GreA, and DksA proteins allows us to propose that their dramatically different effects on transcription in vitro (Fig. 2 and Refs. 8, 9, and 14) are likely determined by the sequence and conformation of the short linker between the α-helices at the tip of the CC (Fig. 1D). In GreA and DksA strictly fixed nearly identical α-helices at the tip of the CC would allow for coordination of the two functional acidic side chains, allowing them to coordinate a single Mg\(^{2+}\) ion but at the same time restricting their mobility to focus their activity on a single particular target. The distinct orientations of the α-turns in GreA and DksA (Fig. 1D) likely reflect the differences in the positions of their subdomains thus avoiding functional overlap. We propose that, although the Gre/DksA-like stabilization of the Mg\(^{2+}\) ions in the vicinity of the RNAP AS is also a key element of the Gfh1 function, the flexibility of the Gfh1 CC linker missing the stable α-turn and the excess of the acidic residues in this region might allow Gfh1 to coordinate the Mg\(^{2+}\) ions occupying somewhat distinct positions (likely a characteristic of different reactions catalyzed by RNAP) without significant alterations of the CC orientation. This hypothesis is consistent with the broad functional specificity of Gfh1, which inhibits all RNAP catalytic reactions (Fig. 2). Indeed, the three-residue acidic cluster observed in the Gfh1 structure may readily stabilize one Mg\(^{2+}\) ion providing three coordination bonds (in contrast to only two in GreA/DksA) (Fig. 3A). Under certain conditions, however, one of these three Asp residues may flip its side chain to approach the adjacent fourth Asp, this would allow for coordination of the Mg\(^{2+}\) ion located in a substantially distinct position (Fig. 3B). In fact, the apparent flexibility of the Gfh1 CC linker might allow for additional conformations and subsequent spatial regrouping of the four Asp side chains, resulting in a number of alternative binding modes of the Mg\(^{2+}\) ion or even the nucleotide addition reaction as an example, we suggest two hypothetical pathways, namely competitive and noncompetitive, by which Gfh1 may inhibit RNAP catalysis (Fig. 3C). In the competitive mode, Gfh1 may directly stabilize both catalytic Mg\(^{2+}\) ions that are required for substrate binding and catalysis with its four Asp residues. This would enhance the Gfh1 affinity to RNAP on one hand and sterically exclude substrate binding in the AS on the other. In the noncompetitive pathway, Gfh1 may fix an additional, non-catalytic Mg\(^{2+}\) ion in the site adjacent to, but not overlapping with, the AS. While this would not block substrate binding, competition of this inhibitory Mg\(^{2+}\) ion with the catalytic metal for binding of the substrate phosphates might result in a formation of a stable substrate-bound, but catalytically inactive, transcription complex. A similar mechanism of inhibition was recently proposed for the RNAP inhibitor tagetitoxin, which “donates” an additional Mg\(^{2+}\) ion bound in the vicinity of the AS (11). Given a broad Gfh1 specificity, it may utilize either or both of these pathways to modulate distinct RNAP catalytic reactions. Undoubtedly, the proposed mechanisms remain hypothetical and would require further biochemical and structural verification.

Gfh1 represents a third example of the rapidly growing family of the CC transcription factors operating through the RNAP secondary channel. All of them share a long CC-domain to deliver their active structural elements to the common target, the RNAP AS, but confer strikingly distinct effects on transcription. Remarkably, this functional diversity is achieved through relatively subtle variations in the conformation and sequence of a short linker segment located at the tip of the CC-domain that presumably targets only a few metal ions in the RNAP AS. The CC factors are thus reminiscent of a set of surgical tools, many of which are nearly identical in size and shape and differ only by their tips designed for highly specialized, non-overlapping applications. As transcription is a highly regulated process, we may await discovery of other CC “tools” operating on the AS of RNAP. Moreover, the CC-domain can be
replaced with another extended protein segment (7), further expanding the repertoire of the delivery tools.

Finally, it is worth noting that the studies of the mechanisms of the CC transcription factors are still in their infancy. Although the previous analyses and this work have shed significant light on the structural organization and interactions with RNAP, allowing us and other groups (8–10, 12) to suggest several plausible molecular mechanisms of these factors, a number of important questions concerning their binding sites on RNAP, detailed mechanisms of the AS modulation, and their precise regulatory effects remain unanswered. A combination of biochemical and structural analyses of the CC proteins bound to their respective targets (transcription complexes) is required to address these questions.

Note Added in Proof—While this paper was under review the article describing the homologous structure of the T. aquaticus Gfh1 protein was communicated to us (25). Both the results and implications of this work are in good agreement with our data.

REFERENCES

1. Rozovskaia, T., Chenchik, A., and Bibilashvili, R. (1981) Mol. Biol. (Mosk.) 15, 636–652
2. Borukhov, S., Sagitov, V., and Goldfarb, A. (1993) Cell 72, 459–466
3. Sosunov, V., Sosunova, E., Mustaev, A., Bass, I., Nikiforov, V., and Goldfarb, A. (2003) EMBO J. 22, 2234–2244
4. Steitz, T. A. (1998) Nature 391, 231–232
5. Adelman, K., Yuensonova, I., La Porta, A., Zenkin, N., Lee, J., Lis, J. T., Borukhov, S., Wang, M. D., and Severinov, K. (2004) Mol. Cell 14, 739–751
6. Mukhopadhyay, J., Sineva, E., Knight, J., Levy, R. M., and Ehrlich, R. H. (2004) Mol. Cell 14, 739–751
7. Kettenberger, H., Armache, K. J., and Cramer, P. (2003) Cell 114, 347–357
8. Laptenko, O., Lee, J., Lomakin, I., and Borukhov, S. (2003) EMBO J. 22, 6322–6334
9. Peredernia, A., Svetlov, V., Vassylyeva, M. N., Tabirov, T. H., Yokoyama, S., Artimovitch, I., and Vassylyev, D. G. (2004) Cell 118, 297–309
10. Sosunova, E., Sosunov, V., Kozlov, M., Nikiforov, V., Goldfarb, A., and Mustaev, A. (2003) Proc. Natl. Acad. Sci. U. S. A. 100, 15469–15474
11. Vassylyev, D. G., Svetlov, V., Vassylyeva, M. N., Peredernia, A., Igarashi, N., Matsugaki, N., Watazaki, S., and Artimovitch, I. (2005) Nat. Struct. Mol. Biol. 12, 1086–1093
12. Opalka, N., Chlenov, M., Chacon, P., Rice, W. J., Wriggers, W., and Darst, S. A. (2003) Cell 114, 335–345
13. Paul, B. J., Barker, M. M., Ross, W., Schneider, D. A., Webb, C., Foster, J. W., and Gourse, R. L. (2004) Cell 118, 311–322
14. Hogan, B. P., Hartsch, T., and Erie, D. A. (2002) J. Biol. Chem. 277, 967–975
15. Laptenko, O., and Borukhov, S. (2003) Methods Enzymol. 371, 219–232
16. Peredernia, A., Vassylyeva, N. M., Berezin, I. A., Svetlov, V., Artimovitch, I., and Vassylyev, D. G. (2005) Acta Crystallogr. Sect. F, in press
17. Otwinowski, Z., and Minor, W. (1997) Methods Enzymol. 266, 307–326
18. Collaborative Computational Project Number 4 (1994) Acta Crystallogr. Sect. D Biol. Crystallogr. 50, 760–763
19. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) Acta Crystallogr. Sect. A 47, 110–119
20. Brunger, A. T., Adams, P. D., Clore, G. M., Delano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., and Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Acta Crystallogr. Sect. D Biol. Crystallogr. 54, 905–921
21. Artimovitch, I., Patlan, V., Sekine, S., Vassylyeva, M. N., Hosaka, T., Ochi, K., Yokoyama, S., and Vassylyev, D. G. (2004) Cell 117, 299–310
22. Stebbins, C. E., Borukhov, S., Orlova, M., Polyakov, A., Goldfarb, A., and Darst, S. A. (1995) Nature 373, 636–640
23. Sosunov, V., Zorov, S., Sosunova, E., Nikolaev, A., Zakeyeva, I., Bass, I., Goldfarb, A., Nikiforov, V., Severinov, K., and Mustaev, A. (2003) Nucleic Acids Res. 33, 4202–4211
24. Loizos, N., and Darst, S. A. (1999) J. Biol. Chem. 274, 23378–23386
25. Lamour, V., Hogan, B. P., Eric, D. A., and Darst, S. A. (2005) J. Mol. Biol., in press