A Human RNA Polymerase II Transcription Termination Factor Is a SWI2/SNF2 Family Member

(Received for publication, July 7, 1998, and in revised form, August 3, 1998)

Mingyi Liu‡, Zhi Xie§, and David H. Price¶

From the ‡Department of Biochemistry, University of Iowa, Iowa City, Iowa 52242 and the ¶Howard Hughes Medical Institute, Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, Tennessee 37232

We obtained protein sequence information from Drosophila factor 2, an ATP-dependent RNA polymerase II transcription termination factor, and discovered that it was identical to a SWI2/SNF2 family member called lodestar. Portions of putative human and Caenorhabditis elegans homologues were found in the sequence data bases and a complete cDNA for the human factor was generated using polymerase chain reaction techniques. Recombinant human factor 2 was produced in a baculovirus expression system, purified, and characterized. Similar to the authentic Drosophila factor, the human factor displayed a strong double-stranded DNA-dependent ATPase activity that was inhibited by single-stranded DNA and exhibited RNA polymerase II termination activity. Both factors were able to work on elongation complexes from either species. We discuss the mechanism of termination by factor 2 and the implications for the role of factor 2 in cellular activities.

Communication

TRANSCRIPTION ELONGATION

Transcription elongation has been found to be an important target for control of eukaryotic gene expression (1, 2). After initiating from a promoter, the RNA polymerase II elongation complex is subjected to a block to elongation by the action of components of N-TEF (negative transcription elongation factors).1 One component of N-TEF, Drosophila factor 2, was originally identified by its ability to suppress the generation of long RNA polymerase II transcripts (3). Later it was found that factor 2 caused termination of transcription in an ATP-dependent manner (4–6). DSIF was identified as another negative factor 2 caused termination of transcription in an ATP-dependency (4–6). DSIF was identified as another negative factor 2 caused termination of transcription in an ATP-dependency (4–6). DSIF was identified as another negative factor 2 caused termination of transcription in an ATP-dependency (4–6).

1 Similar to the authentic Drosophila factor, the human factor displayed a strong double-stranded DNA-dependent ATPase activity that was inhibited by single-stranded DNA and exhibited RNA polymerase II termination activity. Both factors were able to work on elongation complexes from either species. We discuss the mechanism of termination by factor 2 and the implications for the role of factor 2 in cellular activities.

EXPERIMENTAL PROCEDURES

Cloning, Expression, and Production of Antibodies to Drosophila Factor 2—Approximately 85 pmol (13 μg) of purified Drosophila factor 2 (DmF2) (4) were excised after SDS-PAGE and subjected to peptide sequencing. The three peptide sequences obtained (NLSQPTIQAVLK, YRWALTGTPIQNK, and KLDLADGVLTGAK) were identical to regions in a known protein, lodestar. A cDNA encoding lodestar was amplified by PCR from cDNA mix (Clonetech) and cloned into a pET21a vector (Novagen). The plasmid was expressed in DE3 cells grown to an OD600 of 0.6 and induced with 0.5 mM isopropyl-β-D-thiogalactopyranoside for 1.5 h. Cells were then harvested and resuspended in 100 mM HEPES (pH 7.6), 15% glycerol, 100 mM KCl, 0.1 mM EDTA, 1 mM dithiothreitol, and 0.1% SDS in isopropanol. Cells were lysed by passing through a French press three times at 4 °C. The cell lysate containing rDmF2 was subjected to centrifugation at 15,000 × g for 20 min, and the supernatant was loaded onto a 30-ml P-11 column equilibrated with 0.1 M NaCl eluted with a 0.1–0.6 M NaCl gradient in TUS buffer. Fractions containing rDmF2 (0.18–0.3 M NaCl) were pooled, dialyzed, and loaded onto a 1-ml Mono S column equilibrated with 0.1 M NaCl elution buffer. The 1 M P-11 step was dialyzed against 15,000 × g for 20 min, and the supernatant was loaded onto a 30-ml P-11 column equilibrated with 0.1 M NaCl eluted with a 0.1–1 M NaCl gradient in TUS buffer. Fractions containing rDmF2 (0.18–0.3 M NaCl) were pooled, dialyzed, and loaded onto a 1-ml Mono S column equilibrated with 0.1 M NaCl in TUS buffer. The column was then eluted with 0.1–0.6 M NaCl gradient in TUS buffer. Fractions containing factor 2 (0.12–0.18 M NaCl) were pooled, dialyzed against phosphate-buffered saline, and used to immunize rabbits (Pocuno Rabbit Farm & Laboratory, Inc.).

Cloning of Human Factor 2—We searched the EST data base for possible homologues using BLAST programs (21) and identified several human ESTs. I.M.A.G.E. Consortium Clone 357834 (22) contained the 3′ end of human factor 2. Using the sequence of this clone and another EST (70720) primers were designed to obtain the 5′ end of the coding region from Marathon-Ready human brain cDNA (Clontech). Single stranded cDNA was synthesized from HeLa total RNA using a primer downstream of the stop codon. A double-stranded cDNA encoding human factor 2 (HuF2) was amplified by PCR and cloned into the modified expression vector pBAC4X-1 (Novagen) to create plasmid pBHHF that encodes an NH2-terminally (His)6-tagged human factor 2.
Sequence Analysis—Pairwise or multiple alignments of full-length factor 2 proteins or their conserved COOH-terminal domains (amino acids 415–1061 for DmF2, 365–1026 for Caenorhabditis elegans factor 2, and 374–990 for HuF2) were made by Gap, Bestfit, andPILEup programs of the GCG program package. The primary sequence analysis was done using SAPP 8 program (28). Secondary structure prediction was done with PSORT program (24). The nuclear localization sequence was predicted by PSORT program combined with sequence alignment. PEST sequence search was done using PESTfind program (25).

Expression and Purification of Human Factor 2—Recombinant baculoviruses were generated using BaculoGold DNA (PharMingen) and pbHFP plasmid. Sf9 cells were harvested 3 days postinfection and were lysed for 40 min in lysis buffer H (10 mM Tris-3HCl at pH 8.0, 150 mM NaCl, 2 mM MgCl2, 20 mM β-mercaptoethanol, 1% Triton X-100, and 0.1% of a saturated solution of PMSF in isopropanol). After centrifugation at 2,000 g for 15 min, the lysate was brought to 1 M NaCl and 5 mM imidazole before a 45 min centrifugation at 250,000 × g. The cleared lysate was loaded onto a Ni2+–nitrilotriacetic acid-agarose column (Qiagen). The column was washed with a solution containing 1 M NaCl, 10 mM Tris-3HCl (pH 8.0), 5 mM imidazole, 10% glycerol, and PMSF and then washed with a similar buffer containing 0.1 M NaCl. HuF2 was eluted by raising the imidazole to 200 mM. After diluting the fractions to 65 mM NaCl using HGEDP (25 mM HEPES at pH 7.6, 15% glycerol, 0.1 mM EDTA, 1 mM dithiothreitol, 0.1% of a saturated solution of PMSF in isopropanol) the sample was cleared by 30-min centrifugation at 2,000 × g and loaded onto a 1-mL Mono S column. A linear gradient from 0.065–0.5 M HGKEDP was used to elute the recombinant protein. The peak fractions (0.11–0.24 M) were pooled, diluted to 0.12 M HGKEDP, and cleared as described above before being loaded onto a 1-mL Mono Q column. The recombinant protein eluted from 0.20 to 0.27 M by a linear gradient from 0.12 to 0.4 M HGKEDP.

APTase Assay and Transcription Termination Assay—APTase assays and transcription termination assays using Drosophila K cell nuclear extract and actin 5C template were carried out essentially as described previously (4, 5). The immobilized CMV template was generated by first cloning the KpnI/HindIII fragment containing the CMV promoter into plG2-basic (Promega) to generate plG2-CMV. After digesting this plasmid with BamHI the resulting DNA fragments were end-filled using Klenow, dCTP, dGTP, dTTP, and biotin-14-dATP. The DNA was digested with EcoRI to produce a 631-nucleotide runoff transcript, phenol-extracted, ethanol-precipitated twice, and dissolved in TE buffer (10 mM Tris-3HCl, pH 8.0, 1 mM EDTA) (pH 8.0). The DNA solution was brought to 0.5 mM NaCl and incubated with Streptavidin-coated beads (Dynal) for 30 min (~40 μg/mg beads). For the termination assay the immobilized CMV template (final concentration ~50 μg of DNA/ml) was incubated with HeLa nuclear extract in the presence of 20 mM HEPES (pH 7.6) and 7 mM MgCl2 for 10 min at 30 °C. The preinitiation complexes were pulse-labeled for 25 s. The early elongation complexes were washed three times with 1 x HMK (20 mM HEPES, 5 mM MgCl2, and 1 x KCl) + 1% Sarkosyl (~4 min for each wash). The complexes were then washed four times with 65 mM HMK (20 mM HEPES, 7 mM MgCl2, 65 mM KCl, and 0.2 mg/ml bovine serum albumin). The beads were resuspended in 65 mM HMKB solution and assayed for termination as described (4). The RNA transcripts were assayed on 15% denaturing polyacrylamide gel and quantitated using Packard InstantImager.

RESULTS

Cloning of Drosophila and Human Factor 2—To further our understanding of the molecular basis of the termination activity of factor 2, we wanted to identify the gene encoding the factor. Sequence information was obtained from purified Drosophila factor 2 (DmF2), and a data base search indicated that it was a known protein, lodestar. Lodestar was first identified as a maternal-effect protein involved in chromosome segregation during mitosis (26). Probing data bases with the Drosophila protein sequence yielded several human ESTs and a cDNA encoding an uncharacterized C. elegans protein. A full-length cDNA encoding the putative human homologue (HuF2) was generated, and its sequence was compared with Drosophila factor 2 and the putative C. elegans homologue (Fig. 1). The three proteins shared seven motifs found in helicase superfamily 1 and 2 and seven additional motifs shared only by SWI2/SNF2 family members (Fig. 1). The full-length DmF2 and HuF2 have 43% identity and 62% similarity, while their more highly conserved carboxyl-terminal regions are 48% identical and 68% similar, significantly higher than the average 31% identity and 53% similarity shared among SWI2/SNF2 family members (27). The putative C. elegans homologue shares similar identity to both HuF2 and DmF2. The three proteins also share three conserved motifs and a predicted nuclear localization signal (between helicase motifs I and II) that are not conserved among the SWI2/SNF2 family members (Fig. 1). Unlike DmF2 (26), HuF2 has no predicted PEST sequence. Except for the COOH-terminal domain and the three factor 2 motifs there is little additional sequence conservation in these three proteins. The NH2 termini of the factor 2 proteins are predicted to be relatively unstructured compared with the highly structured “helicase” domains. The predicted secondary structures of the factor 2-specific motifs are also conserved. Motif B shows high sequence similarity (56% identity and 68% similarity) to a 24-amino acid fragment of a poliovirus capsid protein and is predicted to have secondary structure throughout the region. However, the functional significance of this potential structural homology is not known.

Human Factor 2 Has dsDNA-dependent ATPase Activity—To confirm that the cloned cDNAs encoded factor 2, we generated recombinant forms of both the Drosophila and the human proteins. DmF2 was expressed in E. coli, but the resulting proteins were mostly truncated forms and were insoluble. The longest forms of the recombinant DmF2 had an identical mobility on SDS-PAGE to factor 2 purified from Kc cells (data not shown). Antibodies raised against the recombinant Drosophila protein reacted strongly and specifically with authentic Drosophila factor 2 either in purified form (Fig. 2A) or in crude Kc cell nuclear extract (data not shown). This provides strong evidence that the lodestar protein is DmF2. HuF2 was expressed in a baculovirus expression system as a soluble, NH2-terminally (His)10-tagged protein. The purified recombinant HuF2 was compared with factor 2 purified from Kc cells by silver-stained SDS-PAGE (Fig. 2B). The concentration of DmF2 was known, and the concentration of HuF2 was estimated by densitometric comparison of the two proteins on the stained gel.

The protein sequence for Drosophila lodestar was deposited in the SWISS-PROT data base under SWISS-PROT number P34739 (26). The nucleotide sequence for human factor 2 has been deposited in the GenBank™ data base (accession number AF073771).

FIG. 1. Factor 2 belongs to SWI2/SNF2 DNA-dependent ATPase family. The sequences of Drosophila, human, and a putative C. elegans factor 2 were aligned. Specific protein motifs are shown as indicated in the legend. The percent identity between the factor 2 homologues is indicated.
As further proof of the identity of the cloned proteins we tested them for DNA-dependent ATPase activity. Purified DmF2 has strong dsDNA-dependent ATPase activity that is suppressed by the addition of ssDNA (5, 6). In the presence of dsDNA both DmF2 and recombinant HuF2 displayed an ATPase activity proportional to the amount of protein added (Fig. 3, A and B). Both DmF2 and recombinant HuF2 were significantly stimulated by dsDNA (Fig. 3C). The absolute activity of HuF2 was about 25% of that for DmF2. HuF2 displayed slightly more activity in the absence of dsDNA than DmF2. It is not clear if this small difference is due to a true difference in the two proteins or due to a small amount of contaminating DNA in the recombinant protein preparation. When ssDNA was added to the reactions the ATPase activity of both proteins was suppressed to near-background levels (Fig. 3D). Again, HuF2 retained a significant level of ATPase compared with DmF2, suggesting that it does have a higher intrinsic level of dsDNA-independent ATPase activity. Overall, we conclude from the ATPase assays that HuF2 has very similar properties to authentic DmF2.

**DISCUSSION**

We have extended our studies of factor 2 by cloning the *Drosophila* and human factor and generating functional recombinant human factor 2. We found factor 2 belongs to the SWI2/SNF2 family of proteins that are involved in modulating protein-DNA interactions during transcription, recombination, and repair. Unique biochemical properties and protein se-
quence motifs distinguish factor 2 from other SWI2/SNF2 family members and all other termination factors.

Of all the known termination factors factor 2 is most similar to NPH-I, a viral protein required for VTF/CE-mediated termination of vaccinia polymerase (19, 20). Both factor 2 and NPH-I are ATPases, but NPH-I is stimulated by ssDNA and factor 2 by dsDNA. In the absence of a termination sequence detected by VTF/CE, NPH-I stimulates transcription elongation of vaccinia RNA polymerase (19), but factor 2 has no positive effect on elongation. Comparison of the sequences of the two proteins indicates that they are similar in that they both contain the seven helicase motifs while not exhibiting helicase activity. However, NPH-I contains no SWI2/SNF2 or factor 2-specific motifs and cannot function independently of VTF/CE. Although there are significant differences between the two proteins, further biochemical characterization is required to determine whether they share any steps in their mechanisms of termination.

Comparison of the sequence of factor 2 to the other approximately 150 SWI2/SNF2 family members allows several hypotheses to be made concerning the function of factor 2 and SWI2/SNF2 proteins in general. The ATPase activity of factor 2 is likely encoded by its carboxyl terminus that contains the seven helicase motifs and SWI2/SNF2 motifs shown to be responsible for the ATPase activity and required for the function of other family members (28, 29). The three factor 2-specific motifs surrounding the ATPase domain may be involved in specifying the unique nucleic acid binding properties of the factor or may be involved in specifically coupling ATP hydrolysis to polymerase termination. This idea is supported by the finding that a region outside of the ATPase domain of Mot1p confers its functional specificity (28). Analysis of factor 2 proteins with mutations in these regions using DNA binding and termination assays may be able to resolve these possibilities. It has been suggested that a common feature of SWI2/SNF2 proteins is that they use ATP hydrolysis to disrupt protein-DNA interaction (30), e.g. SNF2 disrupts nucleosome-DNA interactions and MOT1 removes TBP from DNA. The identification of factor 2 as a SWI2/SNF2 family member provides further evidence supporting the ATP driven disruption hypothesis, since the protein couples ATP hydrolysis with removal of RNA polymerase II from the DNA template.

Based on the biochemical characterization of Drosophila factor 2 we suggested that it might be the eukaryotic equivalent of the prokaryotic transcription repair coupling factor (TRCF), Mfd (4–6). Mfd causes termination of RNA polymerase stalled over damaged DNA and then recruits the ABC repair machinery to the damage site (31, 32). The cloning of factor 2 extends the similarity between the two proteins, since both contain helicase motifs. ERCC6 was suggested to be the human TRCF (33); however, several lines of evidence argue against this notion. Mutants of the yeast and mouse homologues of ERCC6 do not affect UV sensitivity (27, 34). ERCC6 cannot dissociate stalled RNA polymerase from the DNA template (35). More decisively, ERCC6 was dispensable for transcription-coupled repair in vitro (36). Although the role of factor 2 as the TRCF is consistent with previous biochemical characterization and the sequence data presented here, it will be necessary to examine the role of factor 2 in defined transcription-coupled repair assays.

The identification of Drosophila factor 2 as the maternal-effect protein, lodestar, may help to explain several previous observations. Lodestar mutants die during early embryogenesis and display chromosome segregation defects. As suggested previously, a reduction in DNA repair could result in this phenotype (26). If so, this supports a role of factor 2 in transcription-coupled repair. Alternatively, chromosome segregation may be hindered by the presence of active transcription complexes during mitosis. Mitotic disruption of elongation complexes on long genes such as Ubx has been observed (37). Consistent with a role of factor 2 in this process, its localization in the nucleus (26) coincides with the mitotic termination of Ubx transcription (37). The possible involvement of factor 2 in elongation control, transcription-coupled repair, and mitotic termination suggests that the activity of factor 2 must be carefully regulated.

Acknowledgment—We thank Ken Williams at the Keck Foundation Biotechnology Resource Laboratory at Yale University for obtaining protein sequence information for Drosophila factor 2.

REFERENCES
1. Bentley, D. L. (1995) Curr. Opin. Genet. Dev. 5, 210–216
2. Reines, D., Conaway, J. W., and Conaway, R. C. (1996) Trends Biochem. Sci. 21, 351–355
3. Price, D. H., Sluder, A. E., and Greenleaf, A. L. (1987) J. Biol. Chem. 262, 3242–3255
4. Xie, Z., and Price, D. H. (1996) J. Biol. Chem. 271, 11043–11046
5. Xie, Z., and Price, D. H. (1997) J. Biol. Chem. 272, 31902–31907
6. Xie, Z., and Price, D. H. (1998) J. Biol. Chem. 273, 3771–3777
7. Hartzog, G. A., Wada, T., Handa, H., and Winston, F. (1996) Genes Dev. 12, 343–356
8. Wada, T., Takagi, T., Yamaguchi, Y., Ferdous, A., Imai, T., Hirose, S., Sugimoto, Yano, K., Hartzog, G. A., Winston, F., Buratowski, S., and Handa, H. (1999) Genes Dev. 13, 533–546
9. Marshall, N. F., and Price, D. H. (1995) J. Biol. Chem. 270, 12335–12338
10. Marshall, N. F., Peng, J., Xie, Z., and Price, D. H. (1996) J. Biol. Chem. 271, 27176–27183
11. Uptain, S. M., Kane, C. M., and Chamberlin, M. J. (1997) Ann. Rev. Biochem. 66, 117–172
12. Richardson, J. P. (1995) Crit. Rev. Biochem. Mol. Biol. 28, 1–30
13. Platt, T. (1984) Mol. Microbiol. 11, 963–990
14. Gottlieb, E., and Steitz, J. A. (1989) EMBO J. 8, 851–861
15. Bachmann, M., Pfeifer, K., Schroder, H. C., and Muller, W. E. (1990) Mol. Microbiol. 4, 85–93
16. Mason, S. W., Sander, E. K., and Grummt, I. (1997) EMBO J. 16, 163–172
17. Mason, S. W., Wallisich, M., and Grummt, I. (1997) J. Mol. Biol. 268, 229–234
18. Reeder, R. H., and Lang, W. H. (1997) Trends Biochem. Sci. 22, 473–477
19. Deng, L., and Shuman, S. (1998) Genes Dev. 12, 533–546
20. Deng, L., Hagar, J., and Shuman, S. (1999) J. Biol. Chem. 274, 19566–19562
21. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) J. Mol. Biol. 215, 403–410
22. Lennon, G. N., Auffray, C., Polymeropoulos, M., and Soares, M. B. (1995) Genomics 33, 151–152
23. Brendel, V., Bucher, P., Nourbaksh, I. R., Blaisdell, B. E., and Karlin, S. (1992) Proc. Natl. Acad. Sci. U.S.A. 89, 2002–2006
24. Rost, B., Sander, C., and Schneider, R. (1994) Comput. Appl. Biosci. 10, 53–60
25. Rogers, S., Wells, R., and Rechsteiner, M. (1986) Science 234, 364–368
26. Girdham, C. H., and Glover, D. M. (1991) J. Biol. Chem. 266, 1786–1799
27. van Gool, A. J., Verhaage, R., Swagemakers, S. M., van de Putte, P., Brouwer, J., Troelstra, C., Bootma, D., and Hoeijmakers, J. H. (1994) EMBO J. 13, 5361–5369
28. Asbelle, D. T., Wang, D., Post, K. W., and Hahn, S. (1998) Trends Genet. 14, 482–485
29. Richmond, E., and Peterson, C. L. (1996) Nucleic Acids Res. 24, 3685–3692
30. Asbelle, D. T., Hansen, K. E., Mueller, C. G., Lane, W. S., Thorner, J., and Hahn, S. (1994) J. Biol. Chem. 269, 1920–1934
31. Selby, C. P., and Sancar, A. (1995) J. Biol. Chem. 270, 4890–4895
32. Selby, C. P., and Sancar, A. (1995) J. Biol. Chem. 270, 4882–4889
33. Troelstra, C., van Gool, A., de Wit, J., Vermeulen, W., Bootma, D., and Hoeijmakers, J. H. (1992) Cell 71, 939–953
34. Collins, A. R., Mitchell, D. L., Zunino, A., Dewit, J., and Busch, D. (1997) Environ. Mutagen. 29, 152–169
35. Selby, C. P., and Sancar, A. (1997) J. Biol. Chem. 272, 1885–1890
36. Selby, C. P., Drapkin, R., Reinberg, D., and Sancar, A. (1997) Nucleic Acids Res. 25, 787–793
37. Shermoen, A. W., and O'Farrell, P. H. (1991) Cell 67, 303–310
38. Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M., Benfield, J., Burton, J., Connell, M., Copsey, T., and Cooper, J. (1994) Nature 368, 32–38