Cloning of a *Drosophila* cDNA with sequence similarity to human transcription factor RAP74

Daniel D. Kephart, Margaret Peele Price, Zachary F. Burton¹, Ann Finkelstein¹, Jack Greenblatt² and David H. Price*

Department of Biochemistry, University of Iowa, Iowa City, IA 52242, ¹Department of Biochemistry and Agricultural Experiment Station, Michigan State University, E. Lansing, MI 48824, USA and ²Banting and Best Department of Medical Research and Department of Molecular and Medical Genetics, University of Toronto, Toronto, Ontario M5G 1L6, Canada

Received January 12, 1993; Accepted January 16, 1993

EMBL accession no. X66982

Factor 5 is a *Drosophila* RNA polymerase II initiation factor which also affects the elongation phase of transcription (1). *Drosophila* factor 5 and its mammalian counterpart, RAP30/74 (also known as TFIIF or ($\beta\gamma$)), have similar biochemical properties including: 1) their requirement for initiation, 2) the ability to stimulate elongation, 3) a high affinity for purified RNA polymerase II, 4) similar chromatographic properties, and 5) a similar subunit composition (1–6). Recently, the sequence of a cDNA coding for RAP74 (7) and the large subunit of TFIIF (8) have been reported. We have used the RAP74 cDNA (7) to isolate a *Drosophila* cDNA with sequence similarity.

When a *Drosophila* K_c cell cDNA library was probed at low stringency with the cDNA encoding RAP74, a 1.1 kb cDNA with sequence similarity to the 5' end of RAP74 was isolated. The K_c cell cDNA was used to screen a *Drosophila* embryonic cDNA library (obtained from J. Kadonaga, University of California, San Diego) and a larger (1929 bp) cDNA with similarity to the entire RAP74 DNA sequence was isolated. This clone encodes a 577 amino acid protein that is 43% identical and 65% similar to human RAP74 (Figure 1). Antibodies raised against peptides predicted from the cDNA sequence react with authentic factor 5 (data not shown). We propose that this cDNA encodes the large subunit of *Drosophila* factor 5 (factor 5a).

An alignment of the two proteins suggests an organization in which globular N- and C-terminal regions are separated by a highly charged central region (51% D + E + K + R), consistent with a model proposed previously for human RAP74 (7). The N-terminal region of factor 5a is the most highly conserved area of the protein. However, a sequence analogous to the first 28 amino acids of factor 5a is not found in the human protein. The central region of the *Drosophila* protein is characterized by stretches of acidic amino acids situated close to regions of basic amino acids. The central, highly charged region of factor 5a contains one region (amino acids 281 to 300) that is 80% identical to RAP74 suggesting that this region is important for biological function.

ACKNOWLEDGEMENT

This work was supported by a grant GM35500 from the National Institutes of Health.

REFERENCES

- Price, D.H., Sluder, A.E. and Greenleaf, A.L. (1987) *J. Biol. Chem.* **262**, 3244–3255.
- Price, D.H., Sluder, A.E. and Greenleaf, A.L. (1989) *Mol. Cell Biol.* **9**, 1465–1475.
- Flores, O., Ha, I. and Reinberg, D. (1990) *J. Biol. Chem.* **265**, 5629–5634.
- Bengal, E., Flores, O., Krauskopf, A., Reinberg, D. and Aloni, Y. (1991) *Mol. Cell Biol.* **11**, 1195–1206.
- Burton, Z.F., Killeen, M., Sopta, M., Ortolan, L.G. and Greenblatt, J. (1988) *Mol. Cell Biol.* **8**, 1602–1613.
- Conaway, J.W. and Conaway, R.C. (1989) *J. Biol. Chem.* **264**, 2357–2362.
- Finkelstein, A., Kostrub, C.F., Li, J., Chavez, D.P., Wang, B.Q., Fang, S.M., Greenblatt, J. and Burton, Z.F. (1992) *Nature* **355**, 464–467.
- Aso, T., Vasavada, H.A., Kawaguchi, T., Germino, F.J., Ganguly, S., Kitajima, S., Weissman, S.M. and Yasukochi, Y. (1992) *Nature* **355**, 461–464.

Fly	1	MSSASKSTPSAASGSSTSAASAAAAASVASGSASSANVOEFKIRVPK.MPKKHVMRFNAT	60
Hum	1	MAALGPSSQNVTEYVVRVPKNTTKYINIMAFNA	34
Fly	61	LNVDFAQNRNVKLERE.NNMKEFRGMEEDOPKFGAGSEYNDQREARRKKFGIIRKRYR	120
Hum	35	DKVNFATWQARLERDLSNKKIYQ..EEEMPESGAGSEFNKLRREARRKKYGVILKEFRP	93
Fly	121	EAQFWILKVGKTKGKKFGIREGGVGENAAFYVTHAPDGAIEAYPLTEWYNFQPIQRYKS	181
Hum	94	EDQFWLLRVNGSKGRKFGIKKGGVTENTSYIYTCQDGRFAPFVHMYNFTFLARHRT	154
Fly	182	LSAEAEQEFGRKKVMNYFSLMLRKRLRGDEEEQDPEEAKLIAATKKSKELKITDME	242
Hum	155	LTAEAEEMERNNKVLNHFSLMQORLKDQDQDEEEKE..KRGRRKASELRIHDLED	212
Fly	243	WID.SEDSDSEDEEDKKKKEQSDSDGKAKGKGGADKKKKRVDVDEAFESDDGDEE	302
Hum	213	DLEMSDASDASGEGR.....IPKAKKAPLARGGRKKKKKGSDEAFESDDGDFE	267
Fly	303	GREMDY...DTSSSEDPDPEAKVDKMGVAEDALRKLTSDEEEDKKKSDSKEDA	360
Hum	268	QOEVDYMSDGSSSQEEPESKAK.....APOQEEGPKGVDEQSDSSEE	310
Fly	361	DGEKKKKDKGDEVSKDKKKKPTKDDKKGKSGSGSDSTDFSSDSDSDDLSNGPPKKK	421
Hum	311	SEEEKPPEEDKEE...EEKKAPTQEKKKRDKSSESD...SSESDIDSASSA...L	361
Fly	422	VVVKDKDKEKEKEKESAAASKVIASSNAN.KSRSATPTLSTDAKRRKMNSLPDLTASDT	481
Hum	362	FMKKKTPPKAKRKPSGGSSRGNSRPGTPSAEGCGSTSLRAAASKLEQKKVSEMPAAKR	422
Fly	482	SNSPTSTPAKRPKNEISTSLPTSFSGG.KVEDYGIIEAVRRYLKRKPLTATELLTKFNK	541
Hum	423	LRLDTGPQSLSGK...STPQPPSGKTPPNSGDVGQVTEDAVRRYLTKRPMITKDLKKFQTK	480
Fly	542	KTPVSSDRLVETHTKILKKINPVKRTIQGKMYLNIK	577
Hum	481	KTGLSSEQTVNVAQLKRLNPERKMKINDKMFSLKE	517

Figure 1. Comparison of the amino acid sequence of *Drosophila* factor 5a (fly) and human RAP74 (hum). Key; identity (I), high similarity (:), similarity (.)

* To whom correspondence should be addressed