

Transcriptional Activation Modulated by Homopolymeric Glutamine and Proline Stretches

Hans-Peter Gerber, Katja Seipel, Oleg Georgiev,
Manuela Höfner, Martin Hug, Sandro Rusconi, Walter Schaffner*

Many transcription factors contain proline- or glutamine-rich activation domains. Here it is shown that simple homopolymeric stretches of these amino acids can activate transcription when fused to the DNA binding domain of GAL4 factor. In vitro, activity increased with polymer length, whereas in cell transfection assays maximal activity was achieved by 10 to 30 glutamines or about 10 prolines. Similar results were obtained when glutamine stretches were placed within a [GAL4]-VP16 chimeric protein. Because these stretches are encoded by rapidly evolving triplet repeats (microsatellites), they may be the main cause for modulation of transcription factor activity and thus result in subtle or overt genomic effects.

Four distinct types of transcriptional activation domains have been identified by deletion analysis and classified according to their primary sequence: acidic, proline-, glutamine-, and serine-threonine-rich (1). There are at least two functionally different types of transcriptional activation domains: (i) glutamine- and proline-rich domains, which act from a proximal promoter position in conjunction with an enhancer element and (ii) acidic domains (sometimes serine-threonine-rich), which act from proximal and remote (enhancer) positions independently of an additional enhancer (2). Besides glutamine- and proline-rich domains, polyproline and polyglutamine stretches are primarily associated with transcription factors. They are often encoded by repetition of the same codon known as triplet repeat or microsatellite which evolves rapidly. Triplet repeats show pronounced allelic variation and, in some cases, become extremely unstable (3).

The first homopolymeric glutamine repeats to be characterized were found in the Antennapedia transcription factor of *Drosophila* (4). Cross-hybridization experiments showed that the Antennapedia and bithorax proteins share, in addition to the homeobox, M repeats (5). These M repeats were later found to code for glutamine stretches. They are present in many developmentally regulated proteins in *Drosophila* such as engrailed, bithorax, and Notch and were renamed opa repeats (6). Members of the opa sequence family have the general structure (CAX)_n; X is G or A, whereby the triplet CAG or CAA code for glutamine in a specific frame.

On screening the SwissProt protein database for polymeric stretches of at least 20 glutamines and at least 10 prolines (Tables

1 and 2), we found that most of the 40 top-scoring proteins having polymeric glutamine stretches (82%) are transcription factors. Similarly, out of the 44 top-scoring proteins bearing a polyproline stretch, 34 (78%) are thought to be transcription factors (7). From these data we concluded that homopolymeric stretches of glutamines or prolines occur predominantly in transcriptional regulatory proteins.

A nuclear extraction procedure with transiently transfected COS-1 cells provided sufficient amounts of transcription factors for in vitro transcription complementation experiments (8). In this assay, high levels of transcriptional activation by [GAL4] chimeric proteins were obtained with the glutamine-rich domains of OCT-2A, Sp1, TATA binding protein (TBP), and the proline-rich domain from CTF (Fig. 1). In addition, constructs including a 34-oligomer glutamine stretch of human TBP and an artificial 34-oligomer proline

stretch strongly activated transcription (Fig. 1).

We constructed a series of [GAL4] chimeras containing progressively longer stretches of prolines and glutamines (9) to determine their effect on transcription. When analyzed in electrophoretic mobility shift assays, these proteins were all expressed to comparable amounts; no particular instability of larger size homopolymers was observed (Fig. 2, B and D). We found a correlation between the length of the glutamine and proline stretches and the ability to stimulate transcription (Fig. 2, A and C). The observed effect is dependent on DNA binding, because transcription of the reference gene (Figs. 1 and 2) and a reporter template without GAL4 binding sites was not stimulated (7). Extracts with [GAL4] fused to poly-ELQ (Fig. 1), poly-E, poly-D, poly-LLQ, or poly-QQL (7), as well as some domains of proteins that were previously shown to be inactive in vivo (cbl2 and nucleolin) (2), did not increase transcription relative to the background level (Fig. 1) (7).

In cotransfection experiments with HeLa cells, we initially defined the amount of expression vectors giving rise to unsaturated levels of transcriptional stimulation (7). With these conditions, highest transcriptional activity was observed for the 10-oligomer proline (Fig. 3A) and the 10- to 34-oligomer glutamine constructs (Fig. 3B). These findings correlate well with the largest sizes of polymeric proline and glutamine stretches found in transcription factors (Tables 1 and 2) (7). Constructs containing longer stretches of prolines or glutamines showed a reduced ability to mediate transcriptional activation in response to a remote SV40 enhancer in vivo. These findings

Table 1. Transcription factors with polymeric glutamines found in the SwissProt protein database. The search was done for proteins containing at least 20 homopolymeric glutamines with the FASTA program. The list shown represents a selection of the transcription factors found.

Transcription factors	Length of Q-stretch (residues)	Q within stretch (%)	Organism
Embryonic polarity dorsal protein	34	86	<i>Drosophila</i>
Androgen receptor	21	100	Human
Transcription protein GAL11	23	100	Yeast
Nuclear protein SNF5	37	100	Yeast
TATA binding protein (TBP)	38	100	Human
Transcriptional activator DAL81	22	95	Yeast
Grainy head	23	96	<i>Drosophila</i>
Hunchback protein	24	96	<i>Drosophila</i>
Huntingtin*	11-34	100	Human
Glucocorticoid receptor (GR)	22	96	Rat
Suppressor 2 of Zeste protein	26	96	<i>Drosophila</i>
Neurogenic protein mastermind	20	95	<i>Drosophila</i>
Antennapedia homeotic protein	29	76	<i>Drosophila</i>
Neurogenic locus Notch protein†	31	97	<i>Drosophila</i>
Octamer transcription factor N-OCT3	25	96	Human
Homeotic sex combs reduced (SCR)	20	80	<i>Drosophila</i>

Institut für Molekularbiologie II der Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland.

*To whom correspondence should be addressed.

*Protein with unknown function. †Domain including glutamine stretch involved in all transformation.

are not due to decreased protein stability for longer polymers as determined by bandshift experiments (Fig. 2, B and D) (7). To test the effect of poly-Q stretches in the context of a strong transcriptional activation domain that functions irrespective of the presence of an enhancer element, we fused different sizes of poly-Q stretches between the [GAL4] moiety and the activation domain of herpes virus-derived VP16 protein. In transient cotransfection experiments with HeLa cells, we observed again a qualitative correlation between transcriptional stimulation and the size of glutamine stretches up to a length of 40 (Fig. 3C). Control constructs including QA repeats had no strong influence on the transcriptional activation capacity of the VP16 activation domain (Fig. 3C). These experiments revealed strong modulation of transcription factor activity by glutamine stretches of variable size similar to the results obtained when these stretches were tested in combination with the SV40 enhancer.

Instability of triplet repeats has been implicated in several inherited diseases which can be classified into two groups. The first group of diseases contains triplet repeat expansions in the noncoding region of the affected genes. Such is the case in the fragile X syndrome (10), FRAXE mental retardation (11), and myotonic dystrophy (DM) (12). The second group of diseases includes genes with expansions of CAG triplets (coding for glutamine) within putative or established protein-coding regions. Three examples of such cases are spinal and bulbar muscular atrophy or Kennedy's disease (13), Huntington's disease (14), and hereditary spinocerebellar ataxias (SCA1) (15). Similar to Huntington's disease, there is a correlation for SCA1 between the number of glutamines and age of onset, with CAG lengths in excess of 40. For the first group, including fragile X and FRAXE, 200 to 2000 repeats were deleterious, and DM repeats range in number from 60 to 200.

No inherited disease connected with expanded proline-coding triplets has yet been identified. Four triplets code for proline (CCN), and of these, CCG is one of the most common triplet repeats found in human DNA sequences recorded in the Genbank database (16). The sequence poly-CCG and its complementary sequence poly-CGG potentially code for Pro, Ala, Arg, Ser, and Gly. Although the CCG triplet can be subject to rapid expansion or contraction, the proline stretches in transcription factors are usually short (Table 2) (7). We suggest this to be the consequence of strong selection against long poly-P stretches. The *in vivo* experiments in which we observed a more drastic activity drop with longer poly-P stretches compared with poly-Q stretches are in agreement with this notion.

Table 2. Transcription factors with polymeric prolines found in the SwissProt database. The search was done for proteins containing at least nine homopolymeric prolines with the FASTA program. The list shown represents a selection of transcription factors found.

Transcription factors	Length of P-stretch (residues)	P within stretch (%)	Organism
Homeobox protein HOX-2.6	14	100	Mouse
Homeobox protein CHOX-1.4	10	100	Chicken
EBNA-2 nuclear protein	38	97	EB virus
Ecdyson inducible protein	15	100	<i>Drosophila</i>
Wilms tumor protein*	12	92	Mouse
Cellular tumor antigen p53*	11	91	Chicken
Mammalian enhancer binding protein (C/EBP)	9	100	Human
Retinoblastoma protein (p110)*	10	90	Mouse
Homeobox protein HOX-4.3	10	90	Mouse
Octamer binding transcription factor 6	10	80	Rat
Mineralocorticoid receptor	11	82	Human
Androgen receptor	10	90	Human
MYC transforming protein	10	80	Chicken
Huntingtin†	1 × 11 1 × 10	100	Human

*Proteins with tumor suppressor function.

†Protein with unknown function.

Fig. 1. *In vitro* transcription experiments with [GAL4] chimeras having glutamine- or proline-rich domains and homopolymeric glutamine and proline stretches. HeLa nuclear extract (30 to 40 µg in 2 µl) was complemented with COS-1 nuclear extract (10 to 15 µg in 1 to 5 µl) containing [GAL4] chimeric proteins (8, 20). Glutamine-rich activation domains were taken from the ubiquitously expressed transcription factor Sp1 (Sp1:Q, residues 132 to 243, 20% Q), B cell-derived OCT-2A (OCT-2:Q, residues 99 to 161, 26% Q), TATA box binding protein (TBP:Q, residues 17 to 94, 44% Q), and a construct having 34 of 38 total glutamines (TBP:Q, residues 58 to 94, 92% Q). Synthetic constructs (QQS)₁₅ and (ELQ)₁₅ included 15 QQS or ELQ repeats, respectively. The proline-rich activation domain was chosen from CTF-NF1 (CTF:P, residues 399 to 499, 20% P). The 34P construct and all others had the [GAL4] DNA binding domain (DBD) nucleotides 1 to 93 and were cloned according to procedures described (2). P, full length SA probe; M, size marker, pBR322 digested with Hpa II; rt, readthrough transcription signals; βinit, correctly initiated transcription signals; and Ref, reference gene.

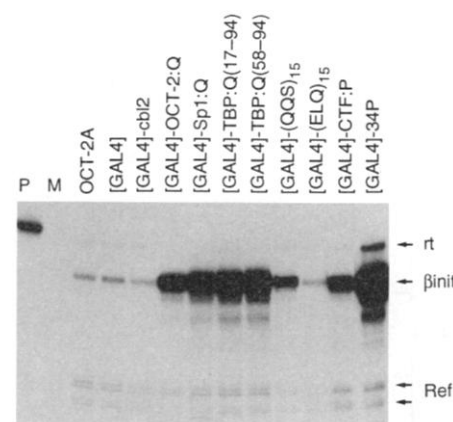
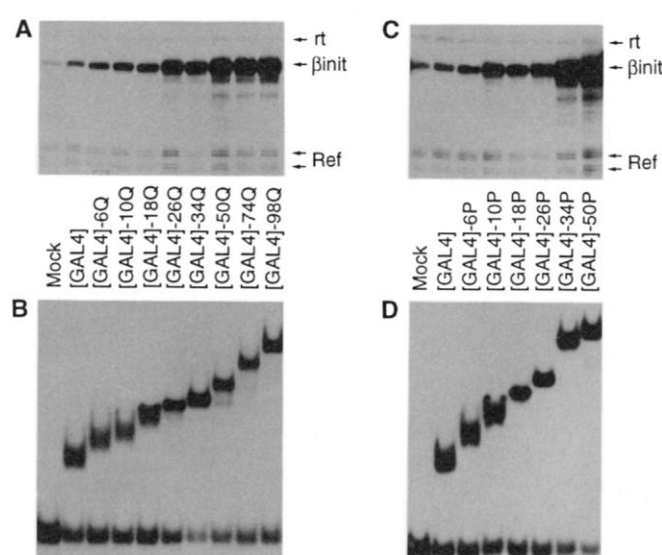


Fig. 2. *In vitro* transcription experiments with [GAL4]-poly-Q or poly-P fusion proteins (9). (A and C) Transcription reactions included 150 ng of 5G-OVEC and 150 ng of OVEC-REF template. *In vitro* transcription conditions were as described for Fig. 1. (B and D) Bandshift experiments of COS-1 nuclear extracts containing [GAL4]-poly-Q (B) or -poly-P proteins (D), respectively. End-labeled oligonucleotide (10 fmol) containing the GAL4 binding site was incubated with 2 to 3 µg of COS-1 mini-extract (1 µl) for 10 min at room temperature before being loaded on a 4% polyacrylamide (1:80) gel (2). Ref, reference gene; rt, read through transcription signals; and βinit, correctly initiated transcription signals.



For yet unknown reasons, transcriptional activation *in vivo* is optimal for a certain number of repeats. The relatively minor effects in the cell culture system might well be related to the clinical effects observed for the three inherited diseases. They show relatively mild and long-lasting manifestations after onset, which typically start at early adulthood. Because neuronal tissues seem to be predominantly affected by CAG expansions (17), we assume that the effect

of glutamine repeats might be most prominent on natural target promoters in neuronal tissues.

The pathological cases notwithstanding, we propose that polymeric amino acid stretches, which are often encoded by triplet repeats of variable size, can have a positive role in evolution. They may be the main cause for modulation of transcription factor activity and thus result in subtle or overt modifications of an organism's genetic program.

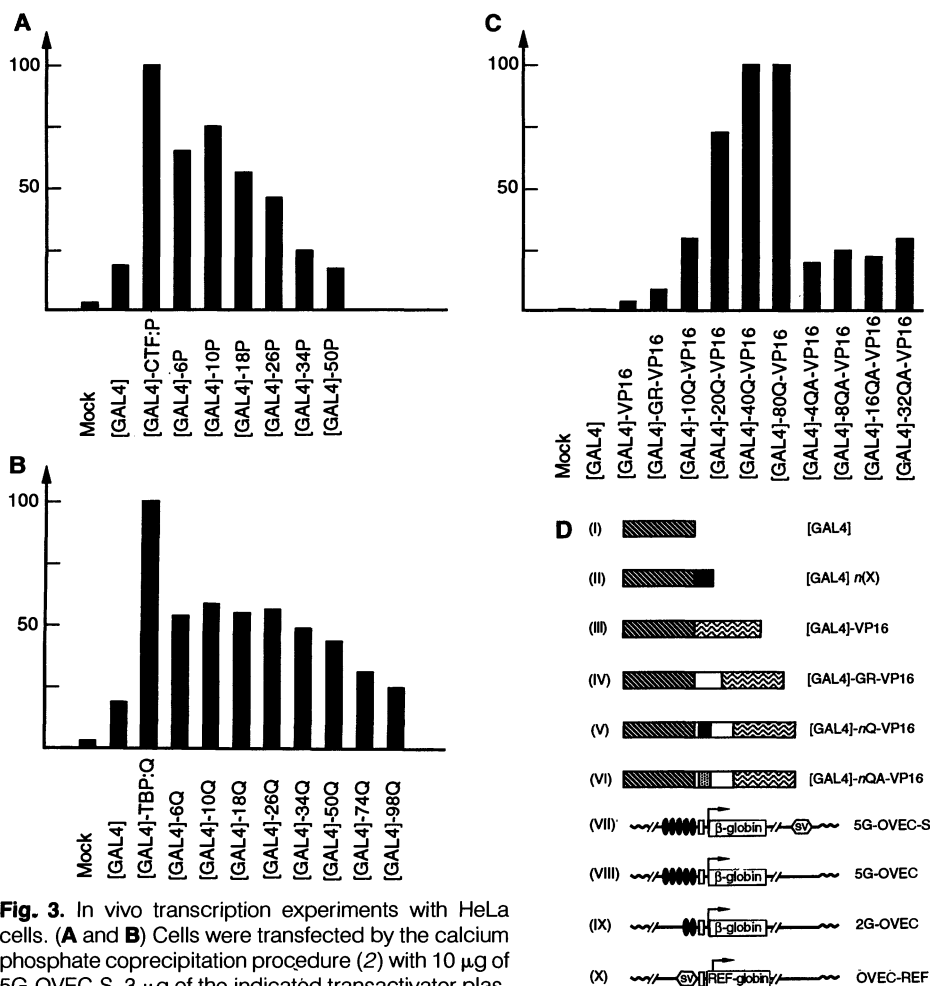


Fig. 3. In vivo transcription experiments with HeLa cells. **(A and B)** Cells were transfected by the calcium phosphate coprecipitation procedure (2) with 10 μ g of 5G-OVEC-S, 3 μ g of the indicated transactivator plasmids, 0.5 μ g of OVEC-REF, and 5 μ g of sonicated salmon sperm carrier DNA. Each experiment was repeated at least three times. Cytoplasmic RNA was extracted and analyzed by ribonuclease protection assay as described (2). For quantification, signals were analyzed by Phosphor Imager with Image Quant software (21). The graphs represent the combined data of two independent transfections. **(C)** Transcriptional activity of [GAL4]-VP16 chimeric proteins. HeLa cells were transfected as described in (A) and (B). Expression vectors (1 μ g) coding for the constructs indicated were cotransfected with 10 μ g of 2G-OVEC and 1 μ g of OVEC-REF template. **(D)** Schematic representation of [GAL4] chimeric proteins and target vectors. The [GAL4] DBD (I) is linked to poly-Q or -P stretches (II) (8) or 77 amino acids specifying the strongest activator of the herpes protein VP16 [residues 413 to 490 (construct III) (2)]. A small segment encoding residues 111 to 157 of the rat GR either devoid (IV) or including polymeric amino acids (V and VI) was sandwiched between the [GAL4] DBD and the VP16 moiety (22). Symbols: diagonally hatched, [GAL4] DBD; wavy hatched, VP16 activator; white, GR moiety; black or heavily shaded, polymeric stretch ($Q_{10,20,40,80}$ or $Q_{4,8,16,32}$). The reporter gene constructs are based on the plasmids OVEC-1 (19) and contained either no enhancer (VIII and IX) or the SV40 enhancer in the 1.5-kb downstream Eco RI site (VII) (2). Construct IX contained two GAL4 binding sites between the Sac I and Sal I sites of OVEC-1 (18). Constructs VII and VIII contained five GAL4 binding sites inserted into the Sal I site of OVEC-1 (18). OVEC-REF (X) contains the SV40 enhancer in the Sal I site (19). Symbols: black, GAL4 binding sites; white, β -globin promoter; hexagon, SV40 enhancer.

REFERENCES AND NOTES

- P. J. Mitchell and R. Tjian, *Science* **245**, 371 (1989).
- K. Seipel, O. Georgiev, W. Schaffner, *EMBO J.* **11**, 4961 (1992).
- C. T. Caskey, A. Pizzuti, Y.-H. Fu, R. G. Fenwick Jr., D. L. Nelson, *Science* **256**, 784 (1992); R. I. Richards and G. R. Sutherland, *Cell* **70**, 709 (1992).
- S. Schneuwly, A. Kuroiwa, P. Baumgartner, W. J. Gehring, *EMBO J.* **5**, 733 (1986).
- W. McGinnis, M. S. Levine, E. Hafen, A. Kuroiwa, W. J. Gehring, *Nature* **308**, 428 (1984).
- K. A. Wharton, B. Yedvobnick, V. G. Finnerty, S. Artavanis-Tsakonas, *Cell* **40**, 55 (1985).
- H. P. Gerber, unpublished results. Abbreviations for the amino acid residues are the following: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
- _____, O. Georgiev, K. Harshman, W. Schaffner, *Nucleic Acids Res.* **20**, 5855 (1992).
- Construction of expression vectors [GAL4]nQ and [GAL4]nP was according to a method that generates multimeric stretches derived from a monomeric element (18). We constructed [GAL4]-6Q by ligating a double-stranded oligomer (5'-CTAGAGAAGACCCTGTGCTGTGCTGCTGTTGGAGACCGGTAC-3' and 5'-CGGTCTC-CACACAGCAACAGCAACAGGTCTTCT-3') coding for six glutamines. Constructs [GAL4]nP were constructed with a double-stranded oligomer (5'-CTAGAGAAGACC GGGAGGTGGGGGAG-GAGGGAGACCGGTAC-3' and 5'-CGGTCTC-CCTCCTCCCCACCTCCGCTCTTCT-3') coding for six consecutive prolines.
- M. Pieretti *et al.*, *Cell* **66**, 817 (1991); M. V. Bell *et al.*, *ibid.* **64**, 861 (1991); R. S. Hansen, T. K. Canfield, M. M. Lamb, S. M. Gartler, C. D. Laird, *ibid.* **73**, 1403 (1993).
- S. J. L. Knight *et al.*, *ibid.* **74**, 127 (1993).
- Y.-H. Fu *et al.*, *Science* **260**, 235 (1993).
- A. R. La Spada, E. M. Wilson, D. B. Lubahn, A. E. Harding, K. H. Fischbeck, *Nature* **352**, 77 (1991).
- T. H. D. C. Group, *Cell* **72**, 971 (1993); A. Norremolle *et al.*, *Hum. Mol. Genet.* **2**, 1475 (1993); C. Zühlke *et al.*, *ibid.* **2**, 1367 (1993).
- H. T. Orr *et al.*, *Nat. Genet.* **4**, 221 (1993).
- J. S. Beckmann and J. L. Webes, *Genomics* **12**, 627 (1992).
- G. A. Ross, M. G. McGinnis, L. M. Russell, S. H. Li, *Trends Neurosci.* **16**, 254 (1993).
- K. Seipel, O. Georgiev, H. P. Gerber, W. Schaffner, *Nucleic Acids Res.* **21**, 5609 (1993).
- G. Westin, T. Gerster, M. M. Müller, G. Schaffner, W. Schaffner, *ibid.* **15**, 6787 (1987).
- We grew COS-1 cells [American Type Culture Collection (ATCC)] in multiflow bottles (Nuncclone, 800 ml, 500 cm²) (8) and transfected them by the lipofection method (Boehringer) with DOTMA [N-[1-(2,3-dioleoyloxy)propyl]-N,N,N-trimethylammonium chloride] reagent (synthesized at the Organisch-Chemisches Institut der Universität Zürich by P. Rüedi). The amount of fusion protein present in each extract preparation was quantified by Phosphor Imager analysis of the signals obtained from bandshift experiments. Equal amounts of DNA binding proteins were added to the transcription reaction, and the volume was adjusted to 5 μ l with mock extract. In vitro transcription conditions (25 μ l) were adjusted with correction buffer to a final solution of 10 mM Hepes (pH 7.9), 10% glycerol, 0.1 mM EDTA, 0.25 mM dithiothreitol, 4 mM MgCl₂, 30 mM KCl, and 4 mM spermidine. Pools containing 150 ng of 5G-OVEC reporter and 150 ng of OVEC-REF template (18, 19), HeLa extract, correction buffer, and ribonucleoside triphosphates (rNTPs), 0.5 mM were preincubated on ice for 10 min before the addition of mini-extracts containing the [GAL4] fusion proteins. Rabbit β -globin reporter RNA was analyzed by S1 mapping (19).
- We determined the relative values of transcriptional activation by dividing the values of the test signals by the reference signals. These values are shown as black bars in Fig. 3.

22. L. Sadowski and M. Ptashne, *Nucleic Acids Res.* 17, 7539 (1989).
23. The GAL4 DBD included sequences (MKLLSSIEQACDICRLKLLKCSKEPKCAKCLKN-NWECRYSPKTKRSLPTRAHLLEVESRLERLE-QLFLIFPREDLMDLMDSLQDIKALL). The VP16 activation domain included sequences (APITD-VSLGDELRLDGEEDVMTADALDDFDLEMLGD-VESPSGPMTHDPVSYGALDVEDDFEFQMFT-DAMGIDDFGG). The GR moiety included sequences (MGLYMGETETKVMGNDLGYPPQA-MGIDDFGG). The sequences of the boundaries

were the following: [GAL4]-VP16: GTPAAASTQF-PGIW; [GAL4]-GR, GTPAAASTLARGS; GR-VP16, GRAPQFPGIW; linkers flanking homopolymers were LEED (amino neighbors) and GDRYP (carboxy neighbors), respectively.

24. We thank D. Amosti for HeLa nuclear extracts and C. Hovens and J. Silke for critical reading of the manuscript. We thank M. Gstaiger for helpful discussions. The preparation of the figures by F. Ochsenbein is gratefully acknowledged.

27 September 1993; accepted 9 December 1993

Promoter-Selective Transcriptional Defect in Cell Cycle Mutant ts13 Rescued by hTAF_{II}250

Edith H. Wang and Robert Tjian

The TAF_{II}250 subunit of the human transcription factor IID (TFIID) rescues the temperature-sensitive hamster cell line ts13 and overcomes a G₁ arrest. Investigation of the transcriptional properties of ts13 nuclear extracts in vitro showed that activation by the site-specific regulators Sp1 and Gal4VP16 is temperature sensitive in ts13 extracts, whereas basal transcription remains unaffected. This transcriptional defect can be rescued by purified human TFIID or by expression of wild-type TAF_{II}250 in ts13 cells. Expression from the cyclin A but not c-fos promoter is temperature sensitive in these mutant cells. Thus, the mutation in TAF_{II}250 appears to have gene-specific effects that may lead to the ts13 cell cycle phenotype.

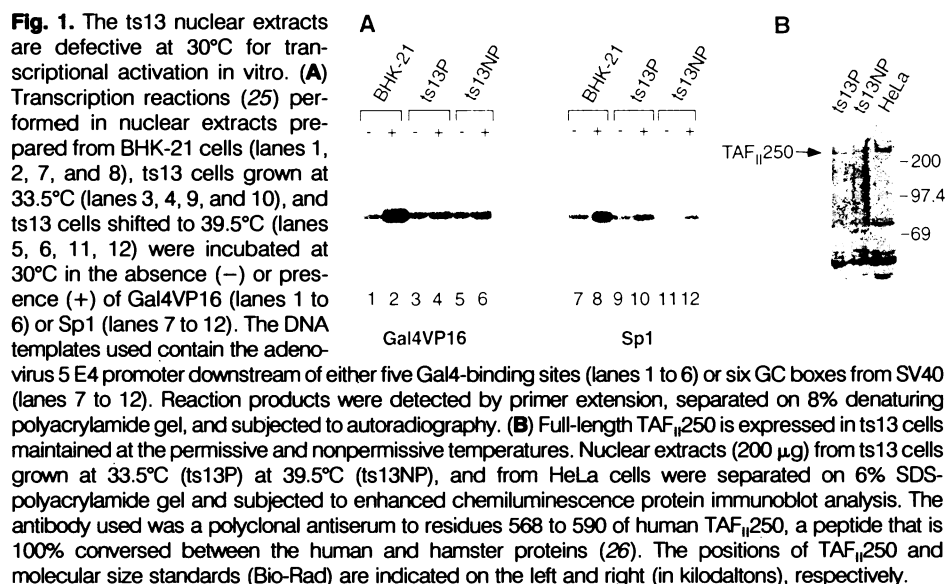
The eukaryotic cell cycle is a highly regulated process that is dependent on the temporal expression of specific genes (1). The cloning of TAF_{II}250 (2, 3), a subunit of the RNA polymerase II transcription factor TFIID (4, 5), has provided a link between cell cycle regulation and the general transcriptional machinery. This largest subunit of human TFIID is encoded by CCG1 (6, 7), a gene that overcomes a G₁ arrest in the temperature-sensitive hamster cell line, ts13 (8). The finding that TAF_{II}250 may play a role in cell cycle regulation prompted us to examine the transcriptional properties of ts13 cells. In particular, we focused on transcriptional activation because TAFs have been postulated to represent a class of factors termed coactivators that are required for regulated but not basal transcription (4, 5, 9, 10). The ts13 cells do not exhibit a global defect in mRNA synthesis at the nonpermissive temperature (11). Thus, the temperature-sensitive cell cycle block in these cells may result from the altered expression of a discrete set of genes.

We have used the mutant hamster cell line ts13 to investigate the potential temperature-sensitive and promoter-specific behavior of the mammalian transcription factor TAF_{II}250 in vitro and in vivo. In addition, we hoped to obtain evidence that TAFs indeed act to mediate transcriptional

activation in the context of intact cells. In order to study regulated transcription in these mutant cells, we used two well-characterized transcriptional activators, Gal4VP16 and Sp1. Because the mechanism of temperature sensitivity in ts13 cells is unknown, we prepared nuclear extracts from cells grown under permissive (33.5°C) as well as nonpermissive (39.5°C) conditions (12) and examined their transcriptional properties in vitro (Fig. 1A). All extracts supported basal transcription. In contrast, nuclear extracts derived from ts13

cells grown at either temperature were defective for Gal4VP16- and Sp1-mediated activation when assayed under conventional assay conditions. As a control, we found that activation was normal in extracts prepared from the parental wild-type BHK-21 hamster cells (Fig. 1A). These results were somewhat unexpected because ts13 cells maintained at 33.5°C presumably contain the functional form of TAF_{II}250. We confirmed by protein immunoblot analysis that TAF_{II}250 is expressed and appears to be intact in ts13 cells grown at both the permissive and nonpermissive temperatures (Fig. 1B). However, it is possible that the transcriptional machinery in ts13 cells is defective for activation when assayed in vitro under standard conditions of 30°C. We therefore performed in vitro transcription at lower temperatures. At 20°C, transcriptional activation was restored in nuclear extracts from ts13 cells (Fig. 2A, lanes 7, 8, 13, 14, and 2B). The levels of activation decreased two- to threefold as the reaction temperature was increased from 20° to 25°C, and became virtually undetectable at 30°C (Fig. 2, A and B), whereas basal transcription remained unchanged or actually increased somewhat at the higher temperature. In contrast, nuclear extracts from BHK-21 cells exhibited comparable levels of activation at all temperatures (Fig. 2A, lanes 1 to 6, and 2B). Therefore, in ts13 extracts, activated but not basal transcription is temperature sensitive in vitro.

To obtain more direct evidence that TAF_{II}250 and therefore TFIID is functionally impaired in ts13 cells, we asked whether human TFIID could restore transcriptional activation in ts13 extracts. Wild-type TFIID from HeLa cells was affinity-purified from the 1.0 M KCl phosphocellulose fraction of LTRα3 cells as described (10) and added to extracts from



Department of Molecular and Cell Biology, Howard Hughes Medical Institute, University of California, Berkeley, CA 94720, USA.