

A sequence motif conserved in diverse nuclear proteins identifies a protein interaction domain utilised for nuclear targeting by human TFIIIS

Yan Ling, Abigail J. Smith and Garry T. Morgan*

Institute of Genetics, University of Nottingham, Queens Medical Centre, Nottingham NG7 2UH, UK

Received March 28, 2006; Accepted March 29, 2006

ABSTRACT

The three structural domains of transcription elongation factor TFIIIS are conserved from yeast to human. Although the N-terminal domain is not needed for transcriptional activity, a similar sequence has been identified previously in other transcription factors. We found this conserved sequence, the LW motif, in another three human proteins that are predominantly nuclear localized. We investigated two examples to determine whether the LW motif is actually a dedicated nuclear targeting signal. However, in one of the newly identified proteins, hIWS1 (human Iws1), a region containing classic nuclear localization signals (NLS) rather than the LW motif was necessary and sufficient for nuclear targeting in HeLa cells. In contrast, human TFIIIS does not possess an NLS and only constructs containing the LW motif were efficiently targeted to nuclei. Moreover, mutations in the motif could cause cytoplasmic accumulation of TFIIIS and enabled a structure/function assay for the domain based on the efficiency of nuclear targeting. Finally, GST pull-down assays showed that the LW motif is part of a protein-binding domain. We suggest that the targeting role the LW motif plays in TFIIIS arises from its more general function as a protein interaction domain, enabling TFIIIS to bind a carrier protein(s) that accomplishes nuclear import.

INTRODUCTION

It has emerged recently that many of the stages of gene expression that occur in the nucleus are mechanistically and/or kinetically coupled [reviewed by (1–3)]. It is also becoming apparent that one important principle underlying

the integration of nuclear events is the organization of the nucleus into distinctive structural and functional compartments, among which the best understood are nucleoli, speckles and Cajal bodies (4,5). Taking the latter as an example, this nuclear organelle is thought, amongst other things, to be a site for interaction, modification and assembly of macromolecular complexes involved both in transcription by RNA polymerase (pol) II and the processing of pre-mRNAs. The activities of Cajal bodies may even extend to the generation of integrated, multifunctional ‘mRNA machines’ or transcriptosomes (6,7).

Another principle underlying the co-ordination of gene expression appears to be the involvement of multifunctional components in different processes or distinctive stages of a process [reviewed by (8–10)]. An example is the transcription factor TFIIIS. This protein was first identified as a transcription elongation factor via biochemical assays that showed it increased the overall transcription rate of pol II by reactivating elongation complexes whose progress had been blocked by a variety of impediments [reviewed by (11,12)]. TFIIIS is thought to accomplish this by activating an intrinsic pol II transcript cleavage activity and then stimulating the arrested elongation complex to re-initiate chain elongation, enabling repeated attempts eventually to clear the transcription block. Recent crystallographic studies of the yeast TFIIIS-pol II complex (13,14) have revealed the likely mechanism of transcript cleavage. TFIIIS binds the surface of pol II and inserts its C-terminal region through the pol II pore such that it contacts the enzyme’s internal active site. Co-ordination of a metal ion by invariant acidic residues in this region of TFIIIS is thought to enable hydrolytic RNA cleavage by the pol II active site (13). In agreement with the crystal structure it is known that the C-terminal two-thirds of TFIIIS is sufficient for *in vitro* transcriptional stimulation and this region is also all that is needed for the activities of nascent RNA cleavage, arrest read-through and pol II binding that have been defined biochemically for intact TFIIIS. Conversely, a function has not yet been assigned to the N-terminal 130 amino acids or so of TFIIIS that

*To whom correspondence should be addressed. Tel: +44 115 823 0390; Fax: +44 115 823 0313; Email: garry.morgan@nottingham.ac.uk
Present address:

Abigail J. Smith, Department of Biochemistry, University of Bristol, School of Medical Sciences, Bristol BS8 1TD, UK

© The Author 2006. Published by Oxford University Press. All rights reserved.

The online version of this article has been published under an open access model. Users are entitled to use, reproduce, disseminate, or display the open access version of this article for non-commercial purposes provided that: the original authorship is properly and fully attributed; the Journal and Oxford University Press are attributed as the original place of publication with the correct citation details given; if an article is subsequently reproduced or disseminated not in its entirety but only in part or as a derivative work this must be clearly indicated. For commercial re-use, please contact journals.permissions@oxfordjournals.org

are not included in the crystal structure. However, one of the few known properties of this region is that it interacts directly with an initiation-competent form of pol II holoenzyme, suggesting TFIIS may be a multifunctional protein that is also involved in controlling an earlier stage of the transcription cycle (15).

The N-terminal region of TFIIS may also reveal another potential mechanism important in co-ordinating nuclear activities, namely the presence of common structural domains among nuclear proteins with distinctive functions. A sequence motif in the N-terminal region that is well conserved between the different isoforms of TFIIS encoded by vertebrate genomes and between vertebrate and yeast TFIIS (16,17), has also been identified in two other transcriptional proteins. These are MED26 (also known as CRSP70 and ARC70), which is a subunit of the human Mediator complex (18,19), and elongin A, a subunit of a transcription elongation factor previously known as SIII (20,21). In the work described here we set out to identify the full range of proteins possessing the conserved TFIIS motif, which we call the LW motif because of the invariant leucine and tryptophan residues it contains. We wanted to do this partly because of the implications a shared motif could have for the *in vivo* properties of TFIIS and also to begin testing the contribution the motif might make to the controlling different aspects of nuclear function. We found that in humans six distinctive proteins contain the motif and all appear to be nuclear. We have tested whether the LW motif functions as a dedicated signal for nuclear localization for two of proteins that contain it. For one such protein nuclear targeting was independent of the LW motif and instead required a region containing classic nuclear localization signals (NLS). However, analysis of mutations in the LW motif of TFIIS demonstrated that here the region was necessary for nuclear targeting. We reasoned that this disparity could be due to the LW motif region being a multifunctional protein-protein interaction domain, i.e. the motif normally functions in nuclear processes but in the case of TFIIS it also operates in the cytoplasm to couple the protein to a carrier that has its own means of nuclear import. Therefore we also examined regions of TFIIS in an *in vitro* protein-protein interaction assay and have found that the LW motif is indeed part of a domain with protein-binding activity.

MATERIALS AND METHODS

Expression constructs

cDNAs encoding all or part of the human TFIIS.o isoform (P23193) and hIWS1 (BAB55198) were cloned into the following vectors. Vector pcDNA3.1/His/6myc was used to create *myc*-tagged constructs. It was derived from pcDNA3.1/HisC (Invitrogen) by cloning into the BamHI site a BglIII/BamHI PCR fragment from MT-6D (22) that encodes six tandem copies of a 13 amino acid sequence containing the *myc* epitope. In addition to full-length cDNAs, *myc*-tagged deletion constructs of TFIIS and hIWS1 were created, all by subcloning PCR-derived fragments into the EcoRI/NotI or EcoRI/XhoI sites, respectively, of pcDNA3.1/His/6myc. This vector was also used for the cloning of site-specific TFIIS mutants. These mutants were made by PCR as described previously (23) using *Pfu* Turbo DNA polymerase (Stratagene) and mutagenic primers that resulted in the substitution

of the residues in the LW motif indicated in Figure 5. Vector pcDNA3.1/His/6myc/PK was made by cloning a BglIII/BamHI PCR fragment encoding residues 20–410 of pyruvate kinase (PK) derived from an existing plasmid (24) into the BamHI site downstream of the *myc*-tag in pcDNA3.1/HisC/*myc*. Plasmids encoding PK-tagged fusions were then created by inserting the required cDNA fragments in-frame downstream of PK in this vector. Vector pGEX-4T-1 was used to create constructs for expression of GST-tagged TFIIS fusions in *Escherichia coli*. These plasmids were constructed by subcloning the respective deletion fragments from the pcDNA3.1/His/6myc series into the EcoRI/NotI sites of pGEX-4T-1. Details of the procedures and sequences of oligonucleotide primers used in these constructions are available upon request. DNA sequences of recombinant plasmids were confirmed using a Big Dye Terminator kit (ABI) and ABI automated sequencer.

A construct that encoded rat PIBP (NP 001014282) cloned in vector pFLAG-CMV5c was kindly provided by Dr Paul Whitley (University of Bath, UK).

Tissue culture and transfection

HeLa cells were maintained in DMEM (Gibco BRL) supplemented with 10% foetal calf serum and penicillin and streptomycin at 100 U/ml and 100 µg/ml, respectively (Gibco BRL) and were passaged at a ratio of 1:10. The day prior to transfection cells were seeded at 2×10^5 cells per well in six-well dishes that also contained a 22 mm² coverslip where required for subsequent cytological analyses. Transfections were carried out either using EffecteneTM transfection reagent (Qiagen) or Lipofectamine plusTM reagent (Gibco BRL) according to the manufacturer's recommendations and the cells incubated for a further 24 h. Treatment of transfected cells with leptomycin B (LMB) was accomplished by incubating cells with 10 ng/ml LMB (Sigma-Aldrich) for 3 h prior to fixation.

Verification of polypeptides produced from transfected constructs was obtained by immunoblotting cell extracts that were first run on 12% SDS-polyacrylamide gels and then electroblotted to nitrocellulose membranes. Blocked membranes were incubated with primary antibody (mAb 9E10; Boehringer) and then with a horseradish peroxidase-conjugated goat secondary antibody (Amersham Biosciences). Blots were developed using an enhanced chemiluminescence (ECL) detection kit (Amersham Bioscience).

Cytological preparations and fluorescence microscopy

For indirect immunofluorescence HeLa cells grown on coverslips were first rinsed in phosphate-buffered saline (PBS) (137 mM NaCl, 2.7 mM KCl, 10.2 mM Na₂HPO₄ and 1.8 mM KH₂PO₄, pH7.4) and then fixed by incubation in 4% paraformaldehyde at room temperature for 10 min. They were then permeabilized in 0.5% Triton X-100 for 10 min at 4°C and incubated in 10% foetal bovine serum (FBS) [First Link (UK) Ltd] in PBS for 30 min to 1 h at room temperature to block non-specific binding of antibodies. Coverslips were then incubated for 1 h at room temperature with primary antibodies diluted in 10% FBS as follows: mAb 9E10 (Boehringer), to 1 µg/ml; mouse anti-FLAG, 1:400; serum 36H against human TFIIS, 1:500 [we generated 36H against

a fusion of hTFIIS.o residues 1–77 and GST that was purified after over-expression in *E.coli* and used to produce antiserum as described previously (25)]. Coverslips were then washed three times for 5 min each with 10% FBS, and then incubated for 1 h at room temperature with the following secondary antibodies (Molecular Probes) diluted to 1–2 µg/ml in PBS: Alexa 488- or 546-conjugated goat anti-mouse IgG; Alexa 594-conjugated goat anti-rabbit IgG. Finally after three 5 min washes in PBS the coverslips were mounted in 50% glycerol/PBS and sealed with clear nail varnish. Differential interference contrast (DIC), and fluorescence observations were made and recorded as described previously (25).

GST pull-down assays

GST-TFIIS fusion proteins were expressed in *E.coli* strain JM105. Glutathione agarose (Sigma) was used for purification of the GST fusions according to the manufacturer's instructions. The purity and concentration of protein samples were estimated by SDS-PAGE. *In vitro* transcription and translation of an hIWS1 derivative (residues 437–819) was carried out with the TNTTM-coupled rabbit reticulocyte lysate system (Promega) on a closed circular template according to the manufacturer's recommendations. The ³⁵S-labelled protein was analysed by electrophoresis through SDS-PAGE before visualization and quantification of bands representing intact proteins by phosphorimaging. Pull-down assays for protein-protein interactions between ³⁵S-labelled hIWS1 and GST-hTFIIS derivatives were carried out as described previously (23,26).

RESULTS AND DISCUSSION

Identification of novel proteins containing the highly conserved LW motif

It has been established by limited proteolysis, NMR spectroscopy and X-ray crystallography that TFIIS is a modular protein of three domains (14,17). The most N-terminal of these (domain I) comprises ~130 amino acids and exhibits two distinct sub-regions. The C-terminal 50 amino acids of TFIIS domain I are not conserved and are unstructured. However the 80 or so N-terminal residues exhibit high cross-species sequence conservation and NMR data for yeast TFIIS shows that this region forms a closely packed four-helix bundle, a structure that has also been predicted for human TFIIS (17). Furthermore, within the N-terminal region a sequence motif of ~50 residues that corresponds to helices 2, 3 and 4 has also been identified in two other proteins, elongin A and MED 26 (17). In order to define this motif more fully and to determine how widespread it might be, we carried out BLAST and Psi-BLAST (27) searches using TFIIS, elongin A and MED 26 query sequences from a range of species. By using invariant leucine and tryptophan residues as key criteria to assess matches to the 'LW motif' (Figure 1A) we found that in addition to MED26, elongin A isoforms, TFIIS isoforms and products of likely pseudogenes, the human proteome contains the following other examples:

- (i) PPP1R10, a nuclear regulatory subunit of protein phosphatase 1 (28) that was previously known as p99 or FB19 (29) or, in rat, as PNUTS (30). Although

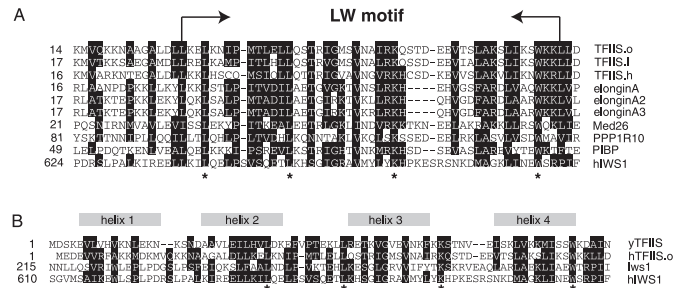


Figure 1. The LW sequence motif. (A) Identification of the LW motif in diverse human polypeptides. The extent of the LW motif is indicated by arrows above a CLUSTALW 1.7 (47) alignment that was manually adjusted to position any gaps to the interhelical linker regions of the structure predicted for this region of human TFIIS (17). Identical or functionally similar residues are outlined in black and the asterisks at the bottom indicate the invariant residues that we used to define the motif. Amino acid co-ordinates are given at the start of each sequence. Conceptual translation products were predicted from sequences with the following nucleotide database accession numbers: TFIIS isoforms, M81601, D50495, AK027024; elongin A isoforms, L47345, AB030834, AB076840; Med26, AF104253; PPP1R10p, Y13247; PIBP, AK056674; hIWS1, AK027561. (B) Sequence conservation of LW motifs between yeast and human in TFIIS and Iws1. Again, identical or functionally similar residues are outlined in black and the asterisks indicate the invariant residues used to define the motif. Above are indicated the positions of the four helices comprising yeast TFIIS domain I as predicted by Booth *et al.* (17).

automated computer annotation has also marked the similarity with TFIIS domain 1 and led to suggestions that PPP1R10 is involved in transcription (28), so far there seems no direct evidence for this involvement and there is some indication of a role in RNA processing (29).

- (ii) PIBP, a small (208 amino acids), hypothetical protein (FLJ32112) that has been predicted from full-length human cDNA and whose rat orthologue was selected as a phosphoinositide binding protein in a phage display screen for binding to PtdIns(3,5)P₂ [(31) and Dr Paul Whitley, University of Bath, UK, personal communication].
- (iii) hIWS1, another hypothetical protein (FLJ14655) predicted from human cDNA, whose C-terminal region has extensive sequence similarity with the protein encoded in yeast by the essential gene *IWS1* (32,33). Within both the human Iws1 protein (hIWS1) and yeast Iws1 the LW motif is located relatively close to the C-terminus rather than being almost N-terminal as in TFIIS and the other LW motif-containing proteins (Figure 1B). Comparison of the LW motifs of Iws1 and TFIIS in yeast and human confirms the invariant and the highly conserved residues that we have used to identify the motif unambiguously even among superficially weak database matches (Figure 1B). Iws1 is known to associate with the elongation factor Spt6 (32) and there are genetic interactions between *IWS1* and genes encoding elongation factors Spt4 and Spt5 (33) as well as an allele of TATA-binding protein (34). It is therefore thought that Iws1 functions in both transcription initiation and elongation (35).

Our analyses therefore identify Iws1 as the second yeast protein in addition to TFIIS that possesses an LW motif and, because this feature has been conserved in the human

orthologues, the motif presumably has an ancient, common function in both proteins. Since the previously identified LW motif proteins are transcription factors, such proteins have been assumed to be nuclear in their distribution, and in some cases a predominantly nuclear localization has been demonstrated (21,36). Of the LW motif proteins newly identified here, PPP1R10 has already been shown to be a nuclear protein by immunolocalization (29,30). In order to determine whether all LW motif containing proteins, and by implication the LW motif itself, are likely to have predominantly nuclear functions, we examined the cellular localization of the other two new examples, PIBP and hIWS1.

Nuclear targeting and subnuclear localization of LW motif containing proteins

HeLa cells were transfected with *myc*-tagged hIWS1 or FLAG-tagged rat PIBP and the localization of the epitope-tagged proteins determined by immunostaining in comparison with that of *myc*-tagged and endogenous TFIIS. Both hIWS1 (Figure 2A) and PIBP (Figure 2A and B) were predominantly nuclear in their distribution and were at least as efficiently targeted to the nucleus as human TFIIS, both with respect to the partitioning of staining between nucleus and cytoplasm within a given cell and with respect to the percentage of transformed cells exhibiting the typical pattern. hIWS1 had a diffuse, somewhat punctate and non-nucleolar distribution pattern (Figure 2A), essentially the same as that of TFIIS. However, in addition to being generally distributed throughout the nucleus, PIBP was also present at high levels in the nucleoli of cells expressing low to moderate amounts of the tagged protein (Figure 2B), and in cells with very high levels of expression PIBP appeared to accumulate at the nucleolar periphery. The distinctiveness of PIBP subnuclear localization is emphasized by double immunostaining for endogenous TFIIS (Figure 2B), in which PIBP localization is highest in nucleoli where TFIIS concentration is at its lowest. Further diversity in subnuclear distribution of LW motif proteins is apparent in the localization described previously for PPP1R10 in HeLa cells (29), which is exemplified by a punctate nucleoplasmic staining pattern that also includes several bright nucleoplasmic bodies and similar discrete accumulations in nucleoli. We did not observe the accumulation of hIWS1, PIBP or TFIIS in such nucleoplasmic bodies, even though we have shown previously that in amphibian oocytes TFIIS is enriched in Cajal bodies (25).

Overall there is no evidence that the LW motif functions as a signal for ensuring the delivery of proteins to a particular subnuclear compartment. However since we can now conclude that LW motif proteins do at least share the property of nuclear localization it seemed possible that the motif could function as a signal for nuclear import. Clearly though if the LW motif were a dedicated nuclear import signal it would be much longer than the classic NLS first characterized in the SV40 T-antigen [reviewed by (37)]. Instead it would be more analogous to examples such as the 38 amino acid M9 domain, which acts as the nuclear import signal for hnRNPA1 (38,39). A further complexity is that computer prediction using the PSORTII algorithm (40) identifies multiple potential NLS patterns in all the LW motif proteins except TFIIS. To determine whether the LW motif has a dedicated nuclear

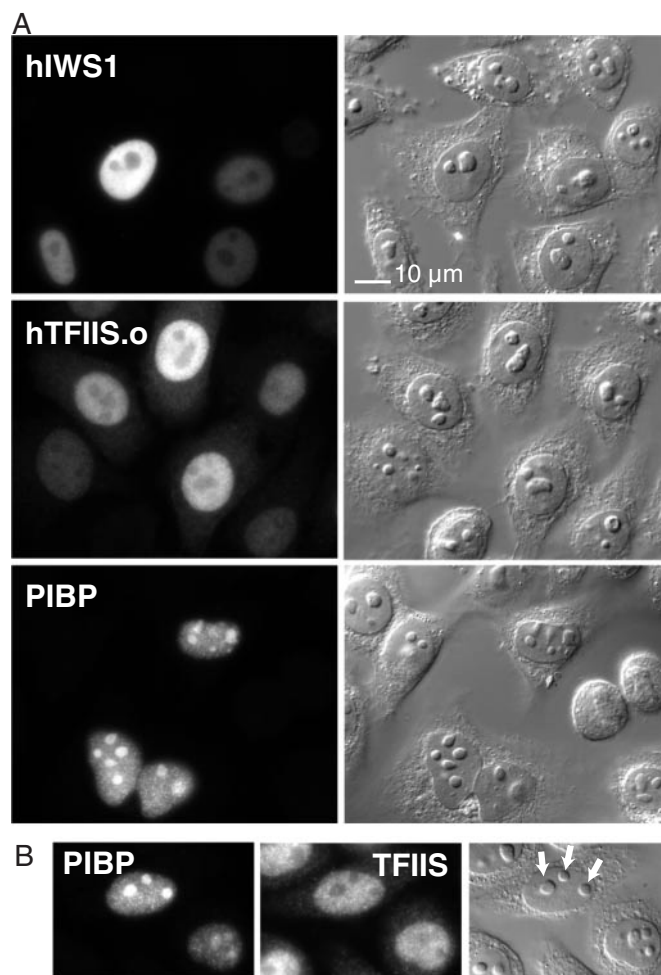


Figure 2. Nuclear targeting and subnuclear localization of hIWS1 and PIBP in HeLa cells. (A) The top panel shows a fluorescent image (and corresponding DIC image) of cells expressing *myc*-tagged hIWS1 that have been immunostained with mAb 9E10 against the tag. For comparison the central panels show the similar, predominantly nuclear localization pattern exhibited by the *myc*-tagged human TFIIS.o isoform. The bottom panels are from cells expressing FLAG-tagged PIBP, which again show a predominantly nuclear distribution of the tagged protein. (B) Cells expressing FLAG-tagged PIBP were co-stained for endogenous TFIIS. Apart from a general nuclear staining, the subnuclear localization patterns of the two proteins appear distinctive, with PIBP being prevalent in nucleoli (arrows in DIC image) whereas TFIIS is relatively under-represented in these organelles.

targeting function we set out to identify experimentally the sequences required for the targeting of hIWS1, which has one of the weaker NLS scores among LW motif proteins, and of the human TFIIS.o isoform, which has no predicted NLS.

Nuclear targeting of hIWS1 is dependent on NLS not LW motifs

hIWS1 (Figure 3A) consists of 819 amino acids, of which an N-terminal region of ~500 residues is highly acidic while a C-terminal region of ~250 residues is well conserved in yeast and includes the LW motif (amino acids 638–688). Two potential NLS patterns occur in the central portion of the polypeptide [KK(X)₁₁KKQK at 358 and KRRR at 542] and we first tested whether a *myc*-tagged hIWS1 fragment containing these signals but not the LW motif was capable

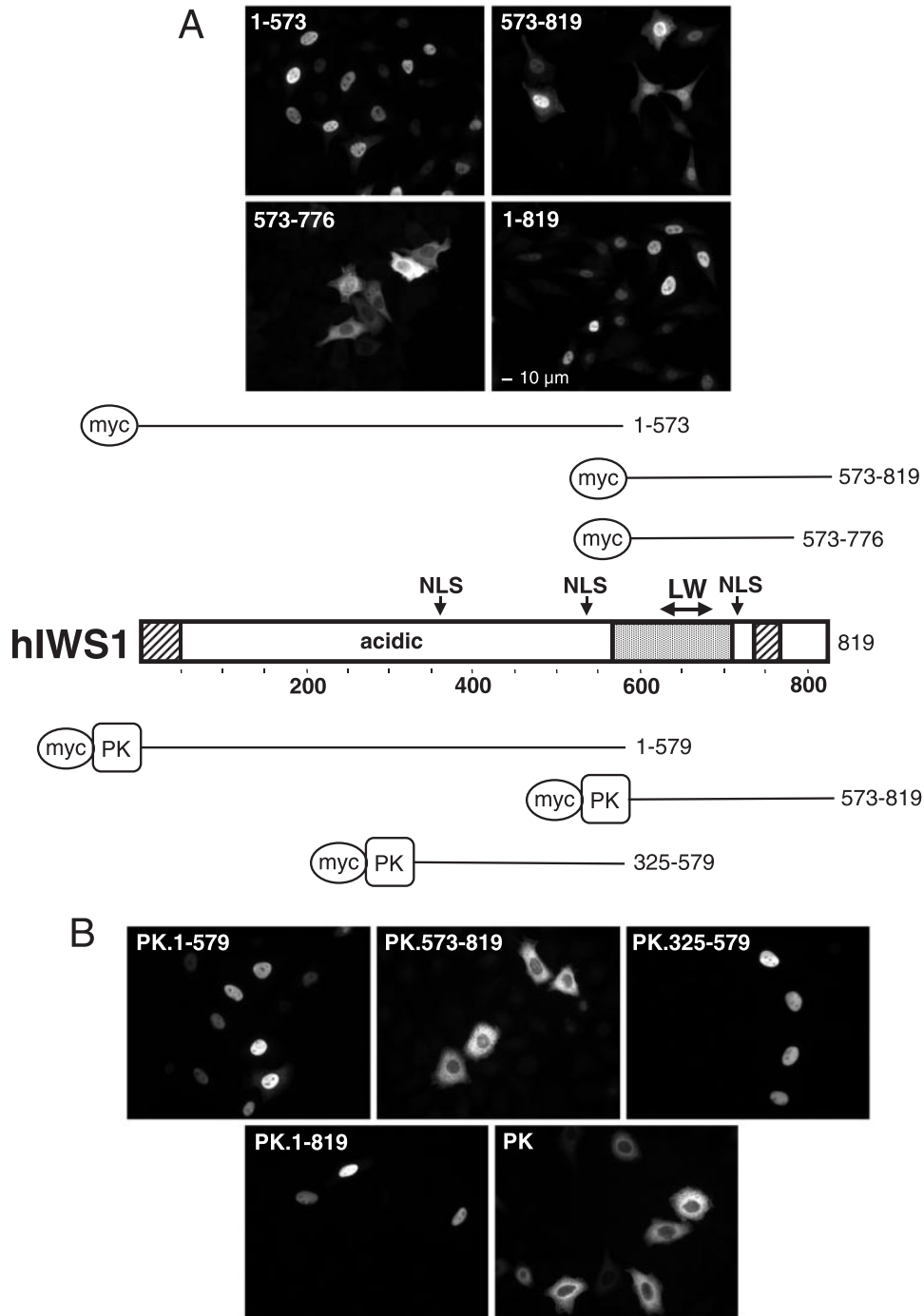


Figure 3. A region containing NLS sequences but not the LW motif suffices for nuclear targeting of hIWS1 (A) Immunolocalization patterns of *myc*-tagged hIWS1 region 1–573 (which does not contain the LW motif) compared with regions 573–819 and 573–776 (which do contain the LW motif). Full-length hIWS1 (1–819) is shown as a control. The diagram below indicates the extent of the tagged polypeptides (numbered according to amino acid residues) and the basic organization of hIWS1. The locations of the three potential NLS signals predicted in the hIWS1 sequence are indicated by downward arrows and the location of the LW motif is shown by horizontal arrows. The region containing the highly conserved sequence present in both yeast and human and that includes the LW motif is shaded, and two other well-conserved regions that flank an extensive region possessing many acidic residues are indicated by hatching. (B) Immunolocalization of *myc*-tagged fusions between regions of hIWS1 and the large cytoplasmic enzyme pyruvate kinase (PK). Immunolocalization of a full-length hIWS1–PK fusion and of PK alone are shown as controls. Above the fluorescent images the hIWS1 fragments used in the fusions are drawn with respect to the overall organization of the polypeptide.

of nuclear targeting. Indeed an N-terminal fragment corresponding to amino acids 1–573 was targeted to the nucleus as efficiently as full-length hIWS1 (Figure 3A), suggesting the LW motif is not required for nuclear targeting. Conversely,

two C-terminal fragments corresponding to amino acids 573–819 and 573–776, respectively, had a more general cellular localization pattern with the smaller fragment even appearing to be predominantly cytoplasmic (Figure 3A). Since

these C-terminal constructs contain the LW motif, as well as a third potential NLS (amino acid 713), it would appear that the LW motif does not accomplish efficient nuclear targeting. In order to test if the central NLS region was necessary and sufficient for targeting, we constructed *myc*-tagged fusions of parts of hIWS1 with the large enzyme pyruvate kinase (PK), which is normally confined to the cytoplasm. As before the C-terminal 573–819 fragment of hIWS1 containing the LW motif did not cause efficient nuclear targeting, and in fact this large PK fusion was confined to the cytoplasm to the same extent as PK alone (Figure 3B). However the N-terminal hIWS1 1–579 fragment was able to target the PK fusion to the nucleus as efficiently as did full-length hIWS1 (Figure 3B). Moreover, a PK fusion with a subfragment of hIWS1 comprising residues 325–579, which includes both the potential NLS patterns of the central region, showed efficient nuclear targeting. Therefore it appears that NLS rather than LW motifs are necessary and sufficient to ensure efficient nuclear targeting of hIWS1.

The LW motif of human TFIIS is necessary for nuclear targeting

Although the previous results showed that the LW motif is not a specialized nuclear targeting domain, we wanted to test whether in some circumstances it could suffice for this purpose, namely in proteins such as human TFIIS.o that appear to lack conventional nuclear targeting signals. As an initial approach we divided hTFIIS.o into three segments, residues 1–77, which include the LW motif, 78–171 and 172–301, based upon the broad patterns of sequence conservation exhibited by vertebrate TFIIS isoforms (Figure 4). A set of five *myc*-tagged constructs each containing a different segment or combination of segments of hTFIIS.o were expressed in HeLa cells, and the cells immunostained to assess targeting efficiency relative to full-length hTFIIS (Figure 4). As before, most cells transfected with full-length hTFIIS.o exhibited predominantly nuclear localization, with the remainder showing about equal staining of nucleus and cytoplasm. The two constructs that contained the LW motif, 1–77 and 1–171 showed a similar degree of nuclear localization to that of full-length TFIIS. However, with the remaining constructs, which lacked segment 1–77, the distribution of the tagged protein was no longer predominantly nuclear and the cells showed a level of cytoplasmic staining that either was similar to or stronger than that of the nucleus. We have confirmed the finding that only *myc* fusion proteins containing segment LW motif exhibit a predominantly nuclear distribution in analogous constructs derived from *Xenopus* TFIIS.o in HeLa cells (data not shown). Therefore it does appear that the LW motif within segment 1–77 could contain the signals responsible for the nuclear targeting of TFIIS.

In order to test further this potential targeting role for the LW motif, we first constructed fusion proteins between PK and all or part of hTFIIS. However nuclear targeting of even full-length TFIIS was disrupted in the fusion proteins, which remained confined to the cytoplasm (data not shown). This effect was seen when PK was fused to either the N-terminus or to the C-terminus of TFIIS and could be due to a dominant influence of PK on structure/function of the LW motif. Alternatively it might point to a diffusion-based mode for

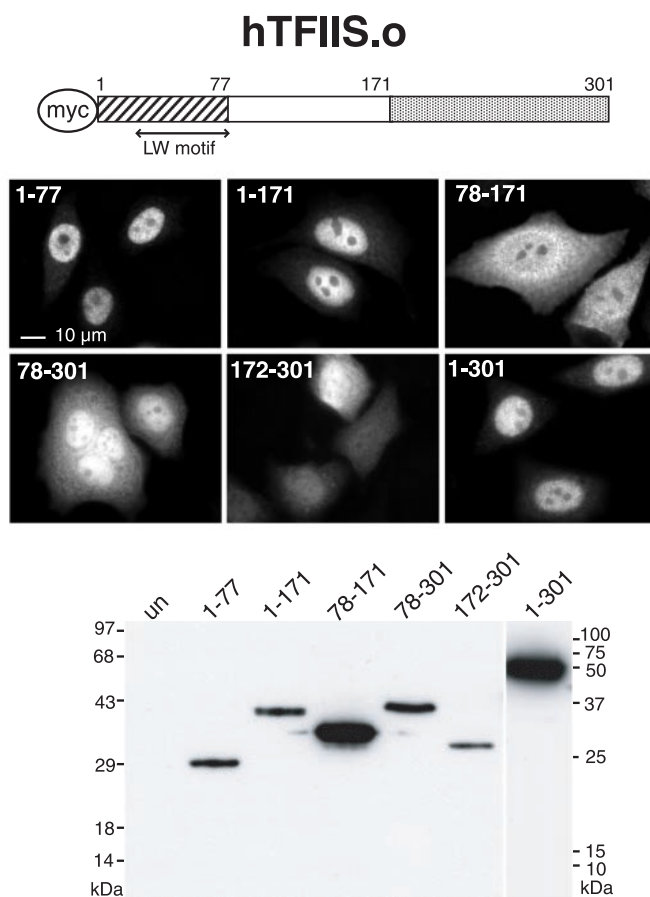
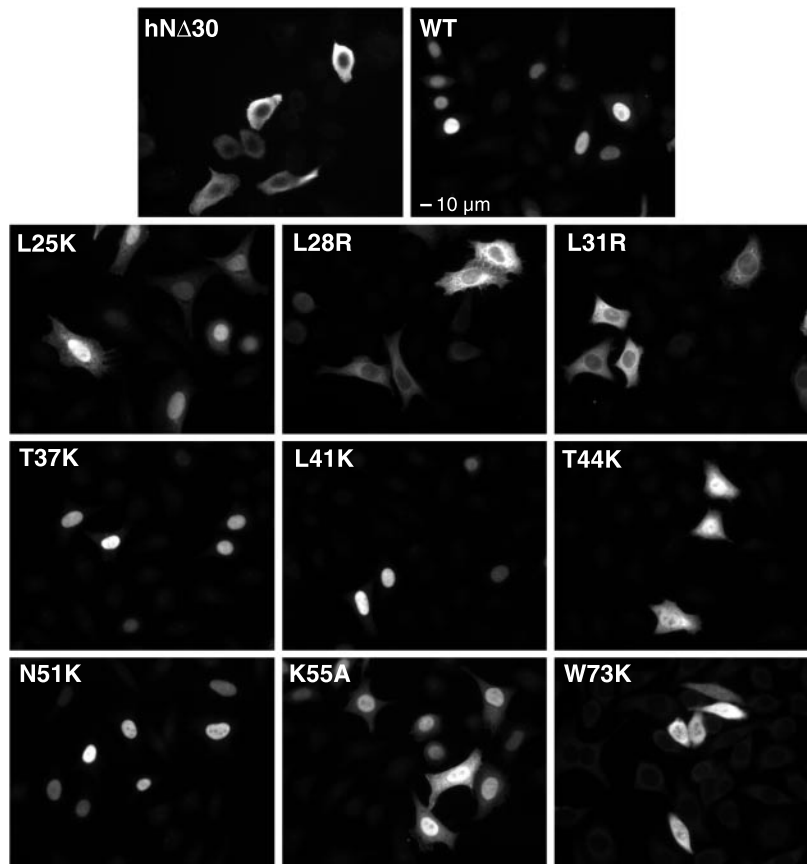
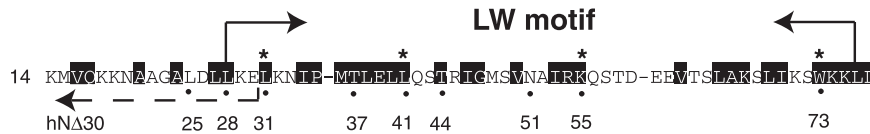


Figure 4. Human TFIIS polypeptides that contain the LW motif show efficient nuclear targeting. The diagram indicates the general organisation of TFIIS and the regions of the hTFIIS.o isoform used in our *myc*-tagged constructs. The C-terminal region 172–301 is highly conserved, and in terms of the structure/function properties of TFIIS [reviewed in (11)] includes the intact Zn ribbon domain, and a portion of domain II. Region 1–77 contains the highly conserved LW motif and represents that part of domain I with a defined structure. Residues 78–301 can confer all known transcriptional functions of TFIIS. The central panel shows immunofluorescent images of cells expressing the corresponding *myc*-tagged constructs stained with mAb 9E10. The lower panel shows immunoblots of deletion products and full-length TFIIS from transfected HeLa cell extracts probed with mAb 9E10. un = control extract from untransfected cells.

TFIIS nuclear entry that cannot operate simply because of the large size of these fusion proteins. In another attempt to examine the potential role of the LW motif in targeting we created finer-scale mutations in the region. In the first of these, hNΔ30, the first 30 amino acids of hTFIIS.o were deleted (Figure 5). This deletion removes most of the first helix of the three into which the LW motif region is folded, and may also disrupt the rest of the structural domain. We found that hNΔ30 was not targeted to the nucleus; instead in 80% of transfected cells it exhibited strong cytoplasmic localization and in the remainder was generally distributed throughout the cell (Figure 5). This distribution pattern suggests that an almost total disabling of the nuclear targeting function of the N-terminal region by the hNΔ30 deletion.

We then used a set of point mutations to determine the effect of substituting amino acids in the hTFIIS.o LW motif that are



<u>Substitution</u>	<u>Type of residue</u>	<u>TFIIS nuclear targeting</u>
L25K	conserved surface	++
L28R	hydrophobic core	-
L31R	hydrophobic core	-
T37K	conserved surface	++(+)
L41K	conserved surface	++(+)
T44K	hydrophobic core	-
N51K	conserved	++(+)
K55A	conserved surface	++
W73K	hydrophobic core	-
None (WT)		+++

Figure 5. Effect of LW motif mutations on the nuclear targeting efficiency of TFIIS. The sequence of hTFIIS.o is shown with conserved and invariant residues indicated as in Figure 1A and with the locations of nine substitution mutations (black dots) and a deletion mutation (hNΔ30; broken line) also shown. Below are representative samples of immunostained cells expressing the mutant forms of *myc*-tagged TFIIS. The relative strength of nuclear targeting by each mutant is given in the table at the bottom together with the structural role predicted for the substituted residue by Booth *et al.* (17). Targeting efficiency was assessed on the basis of the most common localization pattern observed for each mutant as well as the proportion of cells still exhibiting the predominantly nuclear distribution observed for wild-type TFIIS, and is expressed on a scale from no nuclear targeting (–) to wild-type levels (+++).

conserved to different extents and/or that are proposed to play distinctive structural roles in the domain (Figure 5). These hTFIIS mutants were expressed as full-length (Figure 6A) *myc* fusions in HeLa cells and exhibited a range of phenotypes.

In particular substitution of four LW motif residues (L28, L31, T44 and W73) had severe effects on cellular distribution of the fusion proteins and caused the complete absence of nuclear localized hTFIIS (Figure 5). In particular, the L28R and L31R

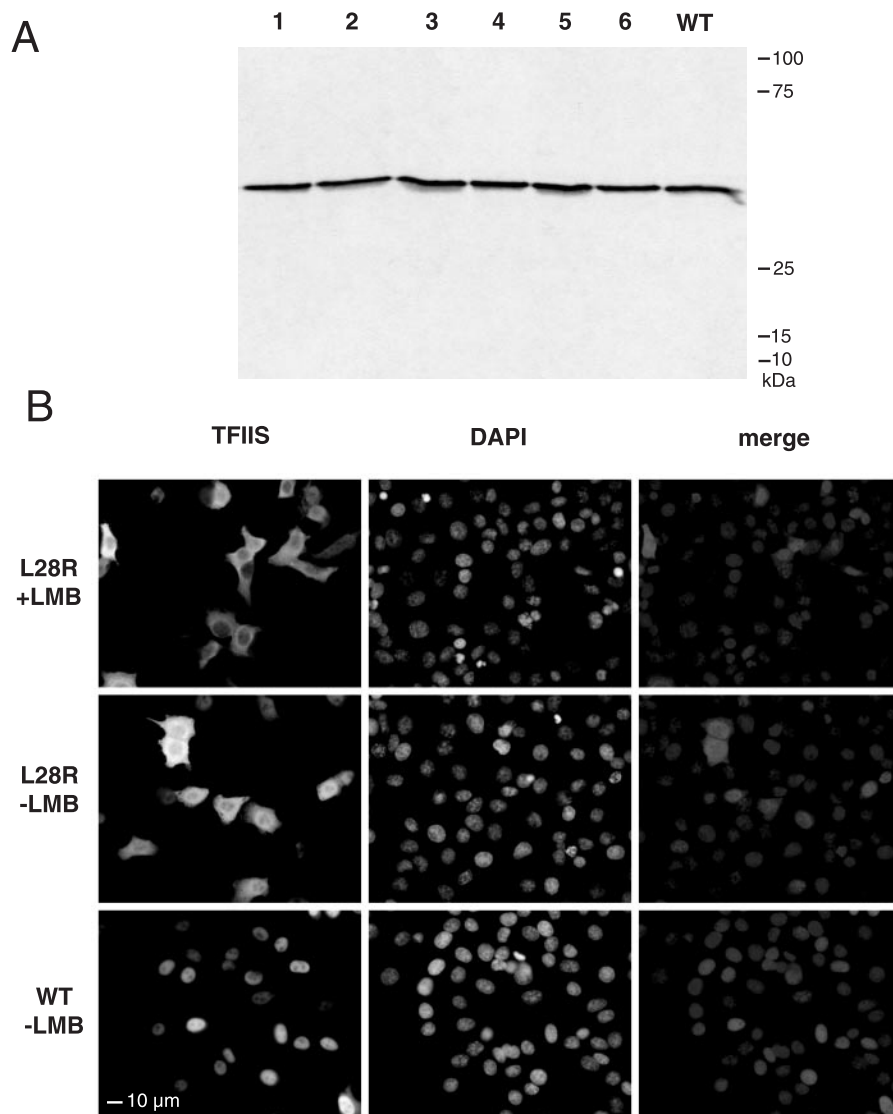


Figure 6. Expression of TFIIIS mutant L28R in HeLa cells. (A) Immunoblot of *myc*-tagged TFIIIS wild type (wt) and point mutants (1–6) from extracts of transfected HeLa cells probed with mAb 9E10. Mutants shown are lane 1, T37K; lane 2, L31R; lane 3, L28R; lane 4, L25K; lane 5, A9P; lane 6, A9D. (B) The effect of leptomycin B (LMB) on the cytoplasmic localization of hTFIIIS mutant L28R. HeLa cells expressing L28R were incubated in the presence (+) or absence (–) of LMB prior to immunostaining TFIIIS with serum 36H. The distributions of overexpressed TFIIIS mutant or wild-type polypeptides are shown in comparison with nuclear DAPI staining. (In comparison with 9E10, 36H does not contribute to detectable cytoplasmic immunostaining, rather endogenous TFIIIS produces only faint background nuclear staining at the imaging exposures used here).

mutations resulted in cytoplasmic accumulation of hTFIIIS in the majority of cells, demonstrating that unlike hIWS1, nuclear targeting of this protein requires the LW motif. However, as discussed below, this targeting role in the context of TFIIIS may be best explained as a secondary consequence of a more general function for the LW motif. These results also suggest that the nuclear targeting mechanism that is disrupted is one of active nuclear import rather than one of passive diffusion followed by active nuclear retention, which can be exhibited by some small proteins. We favour this interpretation because the clear cytoplasmic accumulation of TFIIIS (rather than its free cellular diffusion) as a result of the L28R and L31R point mutations and the h Δ 30 deletion seems difficult to reconcile with a diffusion-based transport mechanism. However, it is conceivable that these cytoplasmically

localized mutants are actively excluded from the nucleus as the result of their becoming substrates, unlike wild-type TFIIIS, for nuclear export. To address this possibility we examined the effect of incubating cells transformed with mis-localizing TFIIIS mutants in leptomycin B (LMB), which inhibits CRM1-dependent nuclear export (41). We found that LMB treatment did not lead to the nuclear accumulation of mutants L28R (Figure 6B) or L31R (data not shown), indeed there appeared an even more frequent occurrence of cytoplasmic localization than in untreated cells examined in parallel. Finally, a role for the LW motif primarily in nuclear retention is not supported by the behaviour of the IWSL 573–776 construct, which contains the LW motif and comprises only ~200 amino acids but which is cytoplasmically localized rather than accumulating in the nucleus (Figure 3A).

Nuclear targeting of TFIIS provides an *in vivo* structure/function assay for the LW motif

The effects of substituting particular amino acids in the LW motif ranged from the complete abolition of nuclear targeting and the generation of either a cytoplasmic or general cellular distribution of TFIIS, to slight effects in which nuclear localization was preserved in many or even the majority of cells (summarized in Figure 5). The level of the effect on nuclear targeting appears to show some correlation with the degree of evolutionary conservation of the substituted residues. For example, mutations L31R and W73K, which change the invariant residues defining the LW motif, markedly reduced nuclear targeting. However in some cases substituting highly conserved or even invariant residues such as L41K caused much less disruption of targeting. There is a better correlation between the effect of a mutation on targeting and the location of the substituted residue in the structure proposed for domain I of human TFIIS. Booth *et al.* (17) predicted the residues forming the hydrophobic core of the domain and identified conserved surface residues that presumably could contribute to interactions with other proteins. As summarized in Figure 5, the more severe examples of disrupted nuclear targeting were observed when hydrophobic core residues were mutated whereas substitution of surface residues was associated with milder, if any, effects on targeting. This correlation can be interpreted as simply reflecting the greater potential of any single substitution when in the hydrophobic core to disrupt the overall structure/function properties of the domain. Nevertheless it is noteworthy that the effect on nuclear targeting provides a first *in vivo* functional assay for domain I of TFIIS and that as such it supports the model for human TFIIS domain I that was modelled on the solution structure determined for yeast TFIIS by Booth *et al.* (17).

Protein–protein interaction: a function for the LW motif in diverse nuclear proteins?

Overall, our experiments with hIWS1 suggest that, despite its conservation in six distinctive nuclear proteins, the LW motif is unlikely to function in vertebrate cells as part of a dedicated nuclear targeting mechanism. However, human TFIIS apparently represents an exception, because in the absence of an NLS-based mechanism of nuclear localization it appears to have developed a reliance on the LW motif for nuclear targeting. This should be contrasted with the situation in yeast where it is well established that the N-terminal region of TFIIS can be removed without damaging its nuclear functions (11) and that the presence of the LW motif is not required for nuclear localization of TFIIS (Rachel Fish and Caroline Kane, personal communication). Interestingly, in yeast the import receptor Nmd5p/Kap119p has been shown to be the main effector of TFIIS nuclear import (42). This protein is one of many members of the importin- β superfamily that provide an alternative to NLS-based nuclear import, although since presumably it acts via a part of yeast TFIIS other than the N-terminal region, we would predict that a role in TFIIS import would not be conserved in any mammalian orthologues. However, it is clearly a possibility that the apparent acquisition by the TFIIS LW motif of a nuclear targeting function in human cells is due to an ability to interact directly with another of the range of dedicated import receptors. However, there is

another possible explanation. Many nuclear proteins, including components of the MCM complex of replication proteins (43) and the basal transcription/DNA repair factor TFIID (44), themselves lack a dedicated import signal and instead utilize a mechanism of nuclear import that relies on a piggyback principle; i.e. in the cytoplasm these proteins bind to a functionally related partner protein(s) that does possess a nuclear import signal such as an NLS and that acts as a carrier for nuclear accumulation of both proteins.

Such an explanation for the nuclear targeting of human TFIIS obviously requires that the region containing the LW motif at least comprises a protein–protein interaction domain. This possibility is supported by the finding that the N-terminal region of TFIIS displays an efficient interaction with a human pol II holoenzyme during protein affinity chromatography (15). In order to examine more directly its potential for protein–protein interactions, we have exploited the fact that in a yeast two-hybrid screen with human TFIIS as the bait one of the proteins we identified was hIWS1 (Y. Ling and G. Morgan, manuscript in preparation). Therefore we were able to use a GST pull-down assay with hIWS1 to test different parts of TFIIS for the possession of a protein interaction domain. As shown in Figure 7, GST–TFIIS fusions that contain the LW motif either alone (TFIIS 1–77) or as part of a larger polypeptide (TFIIS 1–301 and 1–171) were able to bind efficiently to hIWS1, whereas the C-terminal half of TFIIS (172–301) was not. We conclude from this data that the human TFIIS LW motif indeed functions as an *in vitro* protein interaction domain, although we do not know if these proteins ever interact in the cytoplasm. Recently several other proteins binding the N-terminal regions of TFIIS have been identified; the novel human proteins FESTA and EloA-BP1, (36,45) and, in yeast, components of the SAGA complex and of the mediator complexes (46). It will be important to determine which, if any, of

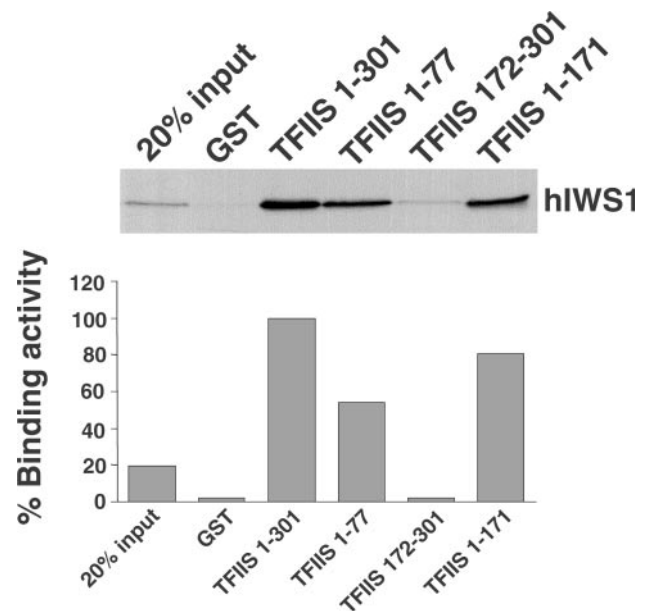


Figure 7. The region of TFIIS containing the LW motif mediates a protein–protein interaction. Pull-down assay of ^{35}S -labelled hIWS1 binding to immobilized GST-fusion proteins containing regions of TFIIS (indicated above the lanes). Graphical representation of this data is shown below in comparison with 20% of the labelled input protein and a pull-down using GST alone.

these TFIIS-binding proteins might serve as the *in vivo* carrier(s) for human TFIIS via the cytoplasmic binding of its LW motif.

In a broader context, the evidence that the LW motif in TFIIS is a protein–protein interaction domain invites the question of whether among the diverse nuclear proteins containing it the motif has a common role. For instance, if all LW motif proteins were involved in different aspects of the same process, namely transcriptional control, they might all interact with common nuclear target protein(s) or other molecule via their LW motif regions. However, the diversity among these proteins with respect to subnuclear localization and their predicted functions or properties could also mean that each individual LW motif is an interaction module utilized for binding different partners involved in distinctive nuclear processes. One useful approach will be to test the extended complement of LW motif containing proteins we have described for their ability to interact with the various TFIIS N-terminal domain binding proteins.

ACKNOWLEDGEMENTS

We are grateful to Dr Paul Whitley for communicating unpublished data and supplying the FLAG-tagged PIBP construct, Dr M. Bellini for the PK construct and Dr Caroline Kane for unpublished data. This work was supported by a BBSRC studentship (A.J.S.) and by a grant (053665) from the Wellcome Trust. Funding to pay the Open Access publication charges for this article was provided by the University of Nottingham.

Conflict of interest statement. None declared.

REFERENCES

- Proudfoot, N. (2000) Connecting transcription to messenger RNA processing. *Trends Biochem. Sci.*, **25**, 290–293.
- Neugebauer, K.M. (2002) On the importance of being co-transcriptional. *J. Cell Sci.*, **115**, 3865–3871.
- Maniatis, T. and Reed, R. (2002) An extensive network of coupling among gene expression machines. *Nature*, **416**, 499–506.
- Lewis, J.D. and Tollervey, D. (2000) Like attracts like: getting RNA processing together in the nucleus. *Science*, **288**, 1385–1389.
- Lamond, A.I. and Spector, D.L. (2003) Nuclear speckles: a model for nuclear organelles. *Nature Rev. Mol. Cell Biol.*, **4**, 605–612.
- Gall, J.G., Bellini, M., Wu, Z. and Murphy, C. (1999) Assembly of the nuclear transcription and processing machinery: Cajal bodies (coiled bodies) and transcriptosomes. *Mol. Biol. Cell*, **10**, 4385–4402.
- Gall, J.G. (2000) Cajal bodies: the first 100 years. *Ann. Rev. Cell Dev. Biol.*, **16**, 273–300.
- Krecic, A.M. and Swanson, M.S. (1999) hnRNP complexes: composition, structure, and function. *Curr. Opin. Cell Biol.*, **11**, 363–371.
- Wilkinson, M.F. and Shyu, A.B. (2001) Multifunctional regulatory proteins that control gene expression in both the nucleus and the cytoplasm. *Bioessays*, **23**, 775–787.
- Manley, J.L. (2002) Nuclear coupling: RNA processing reaches back to transcription. *Nature Struct. Biol.*, **9**, 790–791.
- Wind, M. and Reines, D. (2000) Transcription elongation factor SII. *Bioessays*, **22**, 327–336.
- Fish, R.N. and Kane, C.M. (2002) Promoting elongation with transcript cleavage stimulatory factors. *Biochim. Biophys. Acta*, **1577**, 287–307.
- Kettenberger, H., Armache, K.J. and Cramer, P. (2003) Architecture of the RNA polymerase II–TFIIS complex and implications for mRNA cleavage. *Cell*, **114**, 347–357.
- Kettenberger, H., Armache, K.J. and Cramer, P. (2004) Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. *Mol. Cell*, **16**, 955–965.
- Pan, G., Aso, T. and Greenblatt, J. (1997) Interaction of elongation factors TFIIS and elongin A with a human RNA polymerase II holoenzyme capable of promoter-specific initiation and responsive to transcriptional activators. *J. Biol. Chem.*, **272**, 24563–24571.
- Labhart, P. and Morgan, G.T. (1998) Identification of novel genes encoding transcription elongation factor TFIIS (TCEA) in vertebrates: conservation of three distinct TFIIS isoforms in frog, mouse, and human. *Genomics*, **52**, 278–288.
- Booth, V., Koth, C.M., Edwards, A.M. and Arrowsmith, C.H. (2000) Structure of a conserved domain common to the transcription factors TFIIS, elongin A, and CRSP70. *J. Biol. Chem.*, **275**, 31266–31268.
- Ryu, S., Zhou, S., Ladurner, A.G. and Tjian, R. (1999) The transcriptional cofactor complex CRSP is required for activity of the enhancer-binding protein Sp1. *Nature*, **397**, 446–450.
- Bourbon, H.M., Aguilera, A., Ansari, A.Z., Asturias, F.J., Berk, A.J., Bjorklund, S., Blackwell, T.K., Borggreffe, T., Carey, M., Carlson, M. *et al.* (2004) A unified nomenclature for protein subunits of mediator complexes linking transcriptional regulators to RNA polymerase II. *Mol. Cell*, **14**, 553–557.
- Aso, T., Lane, W.S., Conaway, J.W. and Conaway, R.C. (1995) Elongin (SII): a multisubunit regulator of elongation by RNA polymerase II. *Science*, **269**, 1439–1443.
- Yamazaki, K., Guo, L., Sugahara, K., Zhang, C., Enzan, H., Nakabeppu, Y., Kitajima, S. and Aso, T. (2002) Identification and biochemical characterization of a novel transcription elongation factor, Elongin A3. *J. Biol. Chem.*, **277**, 26444–26451.
- Tuma, R.S., Stolk, J.A. and Roth, M.B. (1993) Identification and characterization of a sphere organelle protein. *J. Cell Biol.*, **122**, 767–773.
- Ling, Y., Lakey, J.H., Roberts, C.E. and Sharrocks, A.D. (1997) Molecular characterization of the B-box protein–protein interaction motif of the ETS-domain transcription factor Elk-1. *EMBO J.*, **16**, 2431–2440.
- Bellini, M. and Gall, J.G. (1999) Coilin shuttles between the nucleus and cytoplasm in *Xenopus* oocytes. *Mol. Biol. Cell*, **10**, 3425–3434.
- Smith, A.J., Ling, Y. and Morgan, G.T. (2003) Subnuclear localization and Cajal body targeting of transcription elongation factor TFIIS in amphibian oocytes. *Mol. Biol. Cell*, **14**, 1255–1267.
- Ling, Y., West, A.G., Roberts, E.C., Lakey, J.H. and Sharrocks, A.D. (1998) Interaction of transcription factors with serum response factor. Identification of the Elk-1 binding surface. *J. Biol. Chem.*, **273**, 10506–10514.
- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
- Ceulemans, H., Stalmans, W. and Bollen, M. (2002) Regulator-driven functional diversification of protein phosphatase-1 in eukaryotic evolution. *Bioessays*, **24**, 371–381.
- Kreivi, J.P., Trinkle-Mulcahy, L., Lyon, C.E., Morrice, N.A., Cohen, P. and Lamond, A.I. (1997) Purification and characterisation of p99, a nuclear modulator of protein phosphatase 1 activity. *FEBS Lett.*, **420**, 57–62.
- Allen, P.B., Kwon, Y.G., Nairn, A.C. and Greengard, P. (1998) Isolation and characterization of PNUTS, a putative protein phosphatase 1 nuclear targeting subunit. *J. Biol. Chem.*, **273**, 4089–4095.
- Whitley, P., Reaves, B.J., Hashimoto, M., Riley, A.M., Potter, B.V. and Holman, G.D. (2003) Identification of mammalian Vps24p as an effector of phosphatidylinositol 3,5-bisphosphate-dependent endosome compartmentalization. *J. Biol. Chem.*, **278**, 38786–38795.
- Krogan, N.J., Kim, M., Ahn, S.H., Zhong, G., Kobor, M.S., Cagney, G., Emili, A., Shilatifard, A., Buratowski, S. and Greenblatt, J.F. (2002) RNA polymerase II elongation factors of *Saccharomyces cerevisiae*: a targeted proteomics approach. *Mol. Cell Biol.*, **22**, 6979–6992.
- Lindstrom, D.L., Squazzo, S.L., Muster, N., Burckin, T.A., Wachter, K.C., Emigh, C.A., McCleery, J.A., Yates, J.R.III and Hartzog, G.A. (2003) Dual roles for Spt5 in pre-mRNA processing and transcription elongation revealed by identification of Spt5-associated proteins. *Mol. Cell Biol.*, **23**, 1368–1378.
- Fischbeck, J.A., Kraemer, S.M. and Stargell, L.A. (2002) SPN1, a conserved gene identified by suppression of a postrecruitment-defective yeast TATA-binding protein mutant. *Genetics*, **162**, 1605–1616.
- Hartzog, G.A. (2003) Transcription elongation by RNA polymerase II. *Curr. Opin. Genet. Dev.*, **13**, 119–126.

36. Saso,K., Ito,T., Natori,S. and Sekimizu,K. (2003) Identification of a novel tissue-specific transcriptional activator FESTA as a protein that interacts with the transcription elongation factor S-II. *J. Biochem.*, **133**, 493–500.
37. Mattaj,I.W. and Englmeier,L. (1998) Nucleocytoplasmic transport: the soluble phase. *Ann. Rev. Biochem.*, **67**, 265–306.
38. Siomi,H. and Dreyfuss,G. (1995) A nuclear localization domain in the hnRNP A1 protein. *J. Cell Biol.*, **129**, 551–560.
39. Weighardt,F., Biamonti,G. and Riva,S. (1995) Nucleo-cytoplasmic distribution of human hnRNP proteins: a search for the targeting domains in hnRNP A1. *J. Cell Sci.*, **108**, 545–555.
40. Nakai,K. and Kanehisa,M. (1992) A knowledge base for predicting protein localization sites in eukaryotic cells. *Genomics*, **14**, 897–911.
41. Fornerod,M., Ohno,M., Yoshida,M. and Mattaj,I.W. (1997) CRM1 is an export receptor for leucine-rich nuclear export signals. *Cell*, **90**, 1051–1060.
42. Albertini,M., Pemberton,L.F., Rosenblum,J.S. and Blobel,G. (1998) A novel nuclear import pathway for the transcription factor TFIIS. *J. Cell Biol.*, **143**, 1447–1455.
43. Pasion,S.G. and Forsburg,S.L. (1999) Nuclear localization of *Schizosaccharomyces pombe* Mcm2/Cdc19p requires MCM complex assembly. *Mol. Biol. Cell*, **10**, 4043–4057.
44. Santagati,F., Botta,E., Stefanini,M. and Pedrini,A.M. (2001) Different dynamics in nuclear entry of subunits of the repair/transcription factor TFIH. *Nucleic Acids Res.*, **29**, 1574–1581.
45. Tamura,K., Miyata,K., Sugahara,K., Onishi,S., Shuin,T. and Aso,T. (2003) Identification of EloA-BP1, a novel Elongin A binding protein with an exonuclease homology domain. *Biochem. Biophys. Res. Commun.*, **309**, 189–195.
46. Wery,M., Shematorova,E., Van Driessche,B., Vandenhautte,J., Thuriaux,P. and Van Mullem,V. (2004) Members of the SAGA and mediator complexes are partners of the transcription elongation factor TFIIS. *EMBO J.*, **23**, 4232–4242.
47. Thompson,J.D., Higgins,D.G. and Gibson,T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **22**, 4673–4680.