

*Review Hypothesis*

# An element of symmetry in yeast TATA-box binding protein transcription factor IID

## Consequence of an ancestral duplication?

Rolf Stucka and Horst Feldmann

*Institut für Physiologische Chemie der Universität München, Schillerstraße 44, D-8000 München 2, FRG*

Received 21 December 1989

TATA-box binding factor TFIID is one of the key factors in transcriptional activation. Surprisingly, the yeast TFDII protein [(1989) *Nature* 341, 299–303; (1989) *Cell* 56, 1173–1181; (1989) *Proc. Natl. Acad. Sci. USA* 86, 7785–7789] reveals only limited homology with other DNA-binding proteins. From computer-assisted searches we infer that yeast TFIID possesses a domain structure in which homologous segments are repeated. The greatest similarity is found between two segments, each 33 amino acids in length, in which the positions of four basic residues are strictly conserved. The high homology is also reflected at the gene level. Implications of this novel type of domain structure for possible interactions in transcriptional activation are discussed.

Transcription factor IID (TFIID); Transcription factor; TATA-box; Transcriptional Regulation; Domain structure; Gene duplication

Recently, several groups have identified [1–4] and sequenced [1–3] the yeast gene encoding the general transcription factor IID (TFIID). Binding TFIID to the TATA-box promoter element initiates assembly of the transcription machinery (see [5] for review). The developmental specificity or the efficiency of transcriptional activity can be determined by regulatory factors that interact with enhancers and upstream promoter elements (see [6,7] for review). The principal question that arises is how TFIID recognizes its target sequence on the DNA and how it interacts with the other proteins that regulate gene activity.

To some extent, the answer should lie in the protein structure of TFIID itself. However, the amino acid sequence deduced from the TFIID gene sequence reveals no strong resemblance to any other protein, including other known DNA-binding proteins (see [1–4,8] for review). The sequence shows slight homologies with the sequences of bacterial transcription factors [1–3], but there are different interpretations of this observation. For several reasons, Hahn et al. [2] think that the resemblance with the sigma subunits is fortuitous, while Horikoshi et al. [1] interpret their findings to mean that general functions of sigma factors are retained in TFIID. Schmidt et al. [3] found a distant similarity between residues 99–177 of yeast TFIID and the

putative calcium-binding regions of calpains [9]; the potential significance of the sequence similarity is thought to be due to similar conformations of the helix-turn-helix motifs in DNA-binding proteins and calcium-binding proteins.

Our interest in the yeast transcription factors arises from their participation in the regulated expression of Ty elements [10], which to date comprise 4 different types of elements, as we recently showed [11]. A search for homologies by the use of the sensitive amino acid sequence comparison program of Rechid et al. [12] revealed a repetition in TFIID of two similar sequences, some 80 amino acid residues in length, between position 67–149 and position 157–240, respectively. From the alignment shown in fig.1, it is obvious that 6 segments forming 3 different 'pairs' of homologous blocks can be distinguished. The program bases the alignment between segments 1 and 4 mainly on the Needleman-Wunsch algorithm [13]. The alignment between segments 3 and 6 (the latter forming the carboxy terminus) has an average score of 4.3 standard deviations above the control mean. Remarkably, 'central' segments 2 and 5 (positions 99–131 and 190–222, respectively) have a score 6.1 standard deviations above the control mean. These values are highly significant.

Segments 1 and 4 contain the two cysteine residues, in similar positions in the sequences aligned, that on experimental grounds have been suggested as forming a disulfide bridge, thus linking together most of the basic region of the polypeptide [2]. Segment 6, which forms

*Correspondence address:* H. Feldmann, Institut für Physiologische Chemie der Universität München, Schillerstraße 44, D-8000 München 2, FRG

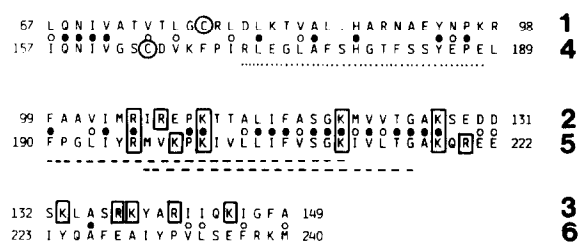


Fig.1. Alignment of two peptide regions from yeast TFIID [1-3] by the sensitive amino acid sequence comparison program [12]. The position in the protein sequence of the first and the last residue in each line is indicated. The segments are numbered consecutively. Identical amino acids are denoted by closed circles, conserved amino acids by open circles, according to the groupings (A,I,L,V,C,M), (H,K,R), (D,E), (F,Y). The two cysteine residues are circled, basic residues in segments 2, 3 and 5 are boxed. The positions in TFIID that were found [1] to be conserved in the sigma factor 2.3 (....) and 2.4 (---) domains, respectively, are underlined as indicated.

the TFIID carboxy terminus, has moderate homology with segment 3. Different from segment 3, which is rich in arginine and lysine residues, segment 6 has no excess positive charge; however, secondary structure predictions (not shown) favor  $\alpha$ -helical conformations for both of these segments. The high homology seen between segments 2 and 5 is due to the fact that the majority of amino acids in equivalent positions within these sequences are identical or of similar chemical nature. Notably, the repeated pattern of 4 basic residues is strictly conserved (fig.1).

The most compelling interpretation of these findings is that TFIID contains a repetition of two peptide domains which are linked by a 'tether' domain, some 24 amino acids in length (position 133-156, including segment 3). The repeated domains would be represented by the moderately homologous segments 1 and 4, and by segments 2 and 5, the two regions of high homology. Strikingly, this domain pattern appears to be predetermined at the gene level. The underlying nucleotide sequences, an alignment of which is shown in fig.2, revealed an overall identity of 55%. The best fit (60%) is scored for the nucleotide sequences corresponding to segments 2 (amino acid positions 94-131) and 5 (amino acid positions 185-222), respectively. These findings strongly suggest that TFIID contains an element of symmetry that might be due to an ancestral duplication of a nucleotide sequence within the gene encoding this protein.

Contrary to what has been suggested by Horikoshi et al. [1], the type of domain structure invoked here would imply that the TFIID protein contains 3 distinct regions with a net positive charge (cf. fig.1). Two of these are the highly conserved domains represented by segments 2 and 5; it is most probable that they will adopt the same conformation in the protein. Secondary structure predictions [14,15] indicate similar hydrophobic  $\beta$ -strand conformation for these segments, between positions 110-124 and 201-215, respectively. The third region rich in basic residues extends from positions 133-156 (including segment 3 in fig.1). It has already been noted by Horikoshi et al. [1] that this latter region has a high potential for  $\alpha$ -helix formation, with a cluster of basic residues being located along one face of the helix. The probability of this region being involved in DNA recognition has been raised by these authors; in their model, however, the helix has been extended into the preceding lysine repeat (positions 120-132).

The presence of the 3 basic domains and the observed symmetry of structural domains in TFIID may be of importance for the central role of this protein in promoter activation and regulation. Several possibilities can be envisaged as to how these structural domains, separately or concomitantly, could be involved in recognition of the TATA-promoter element and in interactions with specific domains of other transcription factors. Indeed, direct effects of GAL4 derivatives [16] or of the mammalian activator protein ATF [17] on mammalian TFIID-promoter interactions have been demonstrated. Given the importance of the acidic domains of upstream activators and the amphipathic nature of many activating factors (e.g. GAL4) (for reviews see [7,18]), one or the other of the basic domains in TFIID might be designed to directly participate in such interactions. Moreover, the observed symmetry of structural domains in TFIID might be of relevance, since some factors, such as GCN4 [19] or GAL4 [7], have been demonstrated to act in transcriptional activation as dimers; it is thought that they bind their recognition sites in a symmetrical fashion. Another possibility is that the two symmetrical basic domains of TFIID interact with TFIIA which has been postulated to exist as a dimer [5]; on the other hand, Hahn et al. [2] have suggested that the C-terminus of TFIID is required for interaction with transcription factor IIA. Alternatively, the two basic domains of

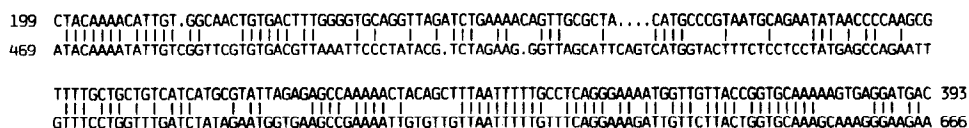


Fig.2. Alignment of two nucleotide sequences from the yeast gene encoding TFIID [1-3]. The positions of the first and the last residues in this alignment are indicated; the first base of the start codon was chosen as position 1. Positions 199-201 represent the codon for amino acid position 67 (fig.1).

TFIID might interact with the C-terminal repeat of the largest RNA polymerase II subunit (see [6] for review) or with the TATA-box promoter element, again in a symmetrical fashion.

The symmetry observed for yeast TFIID may constitute a novel principle for building domain structures in transcription factors; therefore, it is not so surprising that little homology with other motifs in transcriptional regulatory proteins [18] has been observed. In any case, experimental design to elucidate the interactions of TFIID with other components of the transcription machinery should consider this finding.

*Acknowledgements:* This work has been supported by the Deutsche Forschungsgemeinschaft and the Fonds der Chemischen Industrie.

## REFERENCES

- [1] Horikoshi, M., Wang, C.K., Fujii, H., Cromlish, J.A., Weil, P.A. and Roeder, R.G. (1989) *Nature* 341, 299–303.
- [2] Hahn, S., Buratowski, S., Sharp, P.A. and Guarente, L. (1989) *Cell* 56, 1173–1181.
- [3] Schmidt, M.C., Kao, C.C., Pei, R. and Berk, A.J. (1989) *Proc. Natl. Acad. Sci. USA* 86, 7785–7789.
- [4] Eisenmann, D.M., Dollard, C. and Winston, F. (1989) *Cell* 56, 1183–1191.
- [5] Buratowski, S., Hahn, S., Guarente, L. and Sharp, P.A. (1989) *Cell* 56, 549–561.
- [6] Sigler, P.B. (1988) *Nature* 333, 210–212.
- [7] Ptashne, M. (1988) *Nature* 335, 683–689.
- [8] Jean, L.M. (1989) *Science* 245, 1329–1330.
- [9] Kretsinger, R.H. (1987) *Cold Spring Harbor Symp. Quant. Biol.* 52, 499–510.
- [10] Fink, G.R., Boeke, J.D. and Garfinkel, D.J. (1986) *Trends Genet.* 2, 118–124.
- [11] Stucka, R., Lochmüller, H. and Feldmann, H. (1989) *Nucleic Acids Res.* 17, 4993–5001.
- [12] Argos, P. (1987) *J. Mol. Biol.* 193, 385–396.
- [13] Needleman, S.B. and Wunsch, C.D. (1970) *J. Mol. Biol.* 48, 443–453.
- [14] Chou, P. and Fasman, G.D. (1978) *Adv. Enzymol.* 57, 45–148.
- [15] Garnier, J., Osguthorpe, D.J. and Robson, B. (1978) *J. Mol. Biol.* 120, 97–120.
- [16] Horikoshi, M., Carey, M.F., Kakidani, H. and Roeder, R.G. (1988) *Cell* 54, 665–669.
- [17] Horikoshi, M., Hai, T., Lin, Y.-S., Green, M.R. and Roeder, R.G. (1988) *Cell* 54, 1033–1042.
- [18] Struhl, K. (1989) *Trends Biochem. Sci.* 14, 137–140.
- [19] Hope, I.A. and Struhl, K. (1987) *EMBO J.* 6, 2781–2784.