of the same segment must be reflected in the structure of the protein, perhaps giving TFIID a partly symmetrical configuration. Such a symmetry could be relevant to the role of the protein in DNA recognition and binding; after all, the protein has to recognize a TATA sequence which is itself symmetrical (the same or a very similar nucleotide sequence on each of two opposing strands). It is also possible that the homologous domains bind similarly to two identical or different transcription factors.

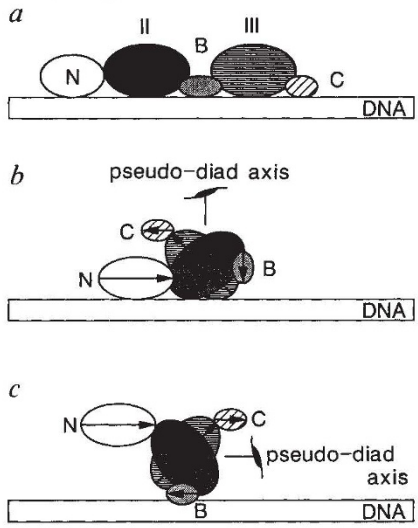
Whatever their function, the amino-acid residues conserved in the homologous repeats must be expected to play an essential role in the proper functioning of TFIID.

JAN H. J. HOEIJMAKERS
Department of Cell Biology and Genetics, Erasmus University, PO Box 1738, 3000 DR Rotterdam, The Netherlands

SIR—The DNA sequence of yeast TFIID has recently been reported[1,2]. By further sequence analysis I have revealed an interesting sequence repeat characteristic of a domain structure. The segment between amino-acid residues 67 and 131 shows striking sequence similarity to that between residues 157 and 222, which the authors of refs 1 and 2 failed to notice. Identical residues occur at 25 positions out of 66 (30 per cent), and if conservative amino-acid replacements such as Ala→Gly, Val→Leu, Asp→Glu are considered, the sequence similarity of these two domains is as high as 60 per cent. The architecture of TFIID is apparent from the presence of these duplicated domains within the protein. These domains have been discovered independently by Hoeijmakers (see above).

Of the residues of the N-terminal region (see figure) 65 per cent are hydrophilic, including Asp- or Glu 14 and Lys- or Arg 10. The B domain, which bridges domains II and III (see above for nomenclature), is rich in Lys and Arg residues. Horikoshi *et al.*[1] noted that part of this region might form an α-helix with Lys and Arg along one face. Weak sequence similarities between residues 171–212 of TFIID (within the proposed III domain) and the conserved 2.3 and 2.4 regions of bacterial σ-factors were noted[1,2]. Horikoshi *et al.*[1] proposed that this region of TFIID might interact with the TATA element, as do bacterial σ factors[4]. The similarity to the 2.3 region is not convincing in the II domain, but both domains II and III show some similarity to the 2.4 region.

This domain structure immediately suggests ways in which TFIID might interact with DNA and other transcription factors. There are three symmetrically distinct ways to arrange domains II and III with respect to DNA. In model *a* (see figure), domains II and III are tandemly placed with respect to the DNA molecule so that the basic region (B domain) is in contact with the TATA box. In models *b* and *c*, the protein chain is folded back at



Schematic models of the TFIID–DNA complex consistent with the five domain structure of TFIID: the N-terminal hydrophilic domain (N); the duplicated II and III domains; the B domain which bridges the II and III domains; and the C-terminal domain (C).

the B region so that II and III are related by a pseudo-diad axis. This axis of symmetry can have two distinct orientations relative to the DNA molecule. If domains II and III both interact with DNA, they should be structurally equivalent with respect to DNA. Symmetry would then require the pseudo-diad axis to point into the minor or major groove (the TATA box has diad symmetry) as is the case for prokaryotic DNA-binding proteins such as λ, 434, Trp and Met repressors[5]. In this case the B domain cannot have close contact with DNA. In model *c*, the pseudo-diad axis is not involved in the interaction with DNA, the B domain binding directly to the TATA box. In this model domains II and III may stabilize the protein or bind to other factors.

The binding of TFIID to the TATA box promotes the binding of transcription factors such as TFIIA and TFIIB and the subsequent binding of RNA polymerase II (refs 6, 7). Yeast TFIID can replace mammalian TFIID to form active transcription-initiation complexes and it is interesting to see how the domain structure and amino-acid sequences have been conserved during evolution. Deletion or swapping of the II and III domains is an obvious experiment to determine their function in TFIID.

KIYOSHI NAGAI
MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

1. Horikoshi, M. *et al. Nature* **341**, 299–303 (1989).
2. Hahn, S. *et al. Cell* **58**, 1173–1181 (1989).
3. Schmidt, M.C., Kao, C.C., Pei, R. & Berk, A. *Proc. natn. Acad. Sci. U.S.A.* **86**, 7785–7789 (1989).
4. Lillie, J.W. & Green, M.R. *Nature* **341**, 279–280 (1989).
5. Otwinowski, Z. *et al. Nature* **335**, 321–329 (1988).
6. Hahn, S. *et al. EMBO J.* **8**, 3379–3382 (1989).
7. Buratowski, S. *et al. Cell* **56**, 549–561 (1989).

Archaebacterial or eocyte tree?

SIR—In response to my study[1], Gouy and Li[2] re-examined the tree of life and concluded that the archaebacterial tree, rather than the eocyte tree, is supported. I would like to point out some flaws in their paper.

First, the authors assert that their preferred methods of phylogenetic analysis (augmented distance and maximum parsimony) are superior to evolutionary parsimony. Felsenstein, however, showed that maximum parsimony can artefactually select trees with juxtaposed long branches (as in the archaebacterial tree) whenever taxa evolve at different rates[3]. Similarly, when rates vary at different positions by orders of magnitude (as in ribosomal RNA sequences), augmented distance can also select an archaebacterial-like tree[4]. By contrast, evolutionary parsimony is the algorithm least sensitive to unequal rate effects[5–7]. The mathematical assumptions of evolutionary parsimony are "mild without precedent"[7].

Although Gouy and Li report that ". . . the eocyte tree is strongly favoured . . ." by evolutionary parsimony analyses of a complete set of small subunit ribosomal RNAs, they use simulations with conspicuously unrealistic parameters to discount this result. For example, in my original analysis of ribosomal RNA data, 1,400,000 nucleotide quartets (17,000 of these were independent) were analysed. Gouy and Li used just 900 quartets for their simulations. Of course, if only 900 quartets are simulated, evolutionary parsimony will support no tree. If a realistic number of quartets are used, evolutionary parsimony functions robustly and the strong support that it gives for the eocyte tree cannot be dismissed.

Likewise, simulations supporting Gouy and Li's favoured methods of reconstruction are erroneous. For example, they calculate branch lengths using augmented distances to estimate the number of substitutions. Shoemaker and Fitch[8] showed that because of invariable sites, these lengths underestimate the actual distances by factors of 3 to 10. Because unequal-rate effects increase rapidly with sequence divergence, and branch lengths have been underestimated, Gouy and Li's simulations are unable to detect if algorithms fail.

Furthermore, Gouy and Li do not discuss why they assigned a non-zero length to the central branch of their eocyte model tree. Their reported χ^2 value for the eocyte