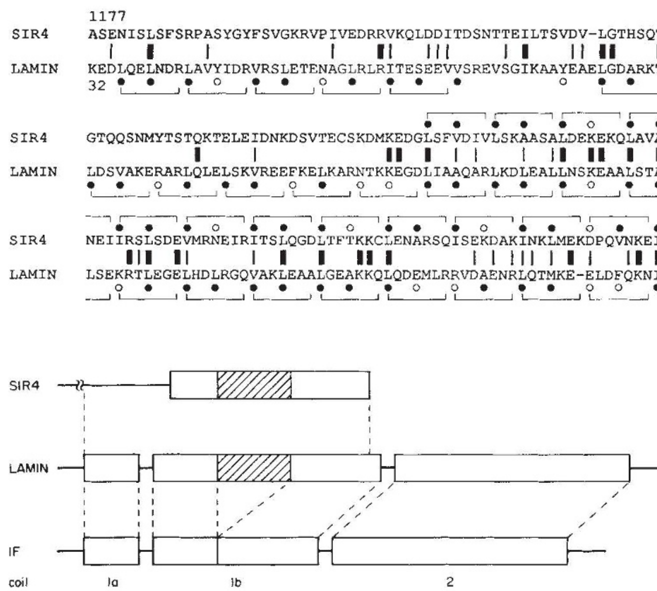
Transcriptional silencing and lamins

SIR—The SIR4 protein in the yeast *Saccharomyces cerevisiae* is required for transcriptional silencing of the *HM* mating-type loci and for high mitotic stability of silencer-containing plasmids. As a result of comparing the sequences of proteins involved in DNA replication and transcriptional control, we have noticed that the C terminus of SIR4 (ref. 1) contains 12 in-phase heptad repeats (brackets in upper figure) within a long region of predicted α-helix (rectangles in lower figure). If the amino acids in the heptad repeat are designated (a-b-c-e-f-g)$_n$, residues a and d (circles in the figure) are generally hydrophobic aliphatic amino acids — alanine (A), isoleucine (I), leucine (L), methionine (M), valine (V) — and form the basis for coiled-coil interactions. Of the 24 residues in the a and d positions of the SIR4 heptad repeats, 20 are hydrophobic aliphatic amino acids (closed circles in figure) typical of the heptad repeats found in proteins like myosin, tropomyosin and the intermediate filament (IF) proteins. A computer search[2] revealed that the C terminus of SIR4 is most closely related to the central rod of the human nuclear lamins A and C (23% identity and 36% conservation between residues 123 and 211 of lamins A and C) and that this similarity is not limited to the a and d positions of the heptad repeat. Although lamins A and C are related to other IF proteins[3], the greatest similarity between the C terminus of SIR4 and coil 1b of the lamins A and C lies within a 43-amino-acid region found only in the lamins and not in other members of the IF family (hatched region, residues 120–162 of lamins A and C) suggesting that SIR4 is specifically related to the nuclear lamins.



Comparison of SIR4, lamin and IF. Identical amino acids are marked by thick bars, conserved amino acids by thin bars.

We note that this region of heptad repeat is an essential functional domain of SIR4 (ref. 1). Furthermore, it is far longer than the 'leucine zipper' found in several transcription factors[4] and the extent and degree of SIR4–lamin similarity is far greater than that noted by Murre *et al.* for the immunoglobin enhancer (kE2) binding protein[5]. This region of SIR4 may serve as a dimerization domain, but it is intriguing to consider that it may also directly interact with the nuclear lamina. In support of this, the REP1 protein encoded by the 2-μM circle plasmid of *Saccharomyces cerevisiae* exhibits a comparable extent and degree of similarity to the lamins and has been shown biochemically to be associated with the karyoskeleton[6]. To explain this, Wu *et al.* postulated that the region of REP1–lamin similarity, found at the C terminus of the REP1 protein, interacts directly with the lamins. It is interesting that the role of REP1 in the mitotic stability of 2-μM plasmids is strikingly similar to the role of SIR4 in the mitotic stability of transcriptional silencer-containing plasmids[7]. In addition to its effect on mitotic stability, however, SIR4 is required for transcriptional silencing by cooperating with silencer-binding proteins[8]. In light of the similarity between SIR4 and the nuclear lamins, it is not unreasonable to suggest that silencing may involve association of the HM loci with the nuclear lamina through SIR4.

JOHN F. X. DIFFLEY
BRUCE STILLMAN

Cold Spring Harbor Laboratory,
Cold Spring Harbor,
New York 11724, USA

1. Marshall, M., Mahoney, D., Rose, A., Hicks, J.B. & Broach, J.R. *Molec. cell. Biol.* **7**, 4441–4452 (1987).
2. Pearson, W.R. & Lipman, D.J. *Proc. natn. Acad. Sci. U.S.A.* **85**, 2444–2448 (1988).
3. McKeon, F.D., Kirschner, M.W. & Caput, D. *Nature* **319**, 463–468 (1986).
4. Landschultz, W.H., Johnson, P.F. & McKnight, S.L. *Science* **240**, 1759–1764 (1988).
5. Murre, C., McCaw, P.S. & Baltimore, D. *Cell* **56**, 777–783 (1989).
6. Wu, L.-C.,C., Fisher, P.A. & Broach, J.R. *J. biol. Chem.* **262**, 883–891 (1986).
7. Kimmerly, W.J. & Rine, J. *Molec. cell. Biol.* **7**, 4225–4237 (1987).
8. Kimmerly, W.J., Buchman, A., Kornberg, R. & Rine, J. *EMBO J.* **7**, 2241–2253 (1988).

Database pollution

SIR—We have noted that a number of entries in the Genbank, EMBL and NBRF nucleic-acid databases contain vector sequences. We compared three vector sequences (plasmid pBR322, phage M13 and λ-phage) with the sequences in the three databases (GenBank release 59.0, EMBL release 18.0, and NBRF-Nucleic release 34.0) and found that overall there are 62 entries that contain unnecessary vector sequence. Most of these entries (52) contain vector sequences which have no serious consequences (that is, vector sequence was left on the end of a submitted sequence). But 10 of the 62 entries contain vector sequences where the authors have made serious mistakes in interpreting their sequence data. These include, a recent entry on GenBank (accession number M19035) that contains a portion of M13, including part of the polycloning site, in the middle of the presumed protein-coding region. Another two (numbers X05793 and X13410) contain vector sequence within supposed coding regions discussed in the papers and six entries (numbers M10384, M14896, M18076, M21517, V00745 and X12660) that are described as unique sequence are almost entirely vector sequence. These problems would certainly have been avoided if the authors had compared their sequence data with available vector sequences.

In addition to these unwarranted inclusions, there are 9 entries on three databases that contain vector sequences which may be warranted but do not have sufficient documentation.

Inclusion of vector sequence in unique sequence entries adds an unnecessary complication to searching the databases. In some cases it is necessary to include vector sequence in a database entry (such as in fusion constructs), but inclusions of this nature should be clearly annotated. It is not the responsibility of database managers or reviewers to weed out unnecessary sequence information. The responsibility rests firmly with the authors of the sequence.

We have informed GenBank of the sequences in question.

DAVID R. STANFORD
JOHN T. PENNISTON
ERIC D. WIEBEN

Department of Biochemistry and Molecular Biology,
Mayo Clinic/Foundation,
Rochester,
Minnesota 55905, USA

© 1989 Nature Publishing Group