J. Mol. Biol. (1993) **230**, 684–688

SEQUENCE NOTES

Prokaryotic Members of a New Family of Putative Helicases with Similarity to Transcription Activator SNF2

Anne-Brit Kolstø[1], Peer Bork[2]†, Kirsti Kvaløy[1], Toril Lindback[1] Anne Grønstadt[1], Tom Kristensen[1] and Chris Sander[2]

[1]*University of Oslo, Biotechnology Centre, N-0137 Oslo 3, Norway*

[2]*European Molecular Biology Laboratory, Meyerhofstr. 1 6900 Heidelberg, Germany*

(Received 23 November 1992; accepted 24 November 1992)

Cloning and sequence analysis of a new open reading frame from *Bacillus cereus* reveals the relationship to a recently identified family of putative eukaryotic transcription activators similar to the yeast SNF2 gene product. As a result of comparative analysis of sequence features conserved in all members of this family, a gene from a chilo iridescent virus, as well as a putative helicase from *Escherichia coli* (*hepA*), can also be grouped into this family. The unexpected presence of prokaryotic and viral sequences in the previously purely eukaryotic SNF2 family suggests a defined subgroup of DNA helicases present in all species, with specific function in transcription activation.

Keywords: homology; helicases; SNF2 family; *Bacillus cereus*; transcription regulation

Recently, many eukaryotic regulatory proteins with similarity to the yeast transcription activator SNF2 (Laurent *et al.*, 1991) have been discovered: (1) an activator for homeotic genes in *Drosophila brahma* (Brm: Tamkun *et al.*, 1992); (2) a gene activator essential for cell growth and viability in yeast, MOT1 (modifier of transcription: Davies *et al.*, 1992); (3) RAD54, involved in both DNA repair and mitotic recombination in yeast (Emery *et al.*, 1991; Davies *et al.*, 1992); (4) STH1 (probably identical with NPS1), involved in G_2 phase control, highly similar to SNF2 but, in contrast to SNF2, essential for viability of yeast (Laurent *et al.*, 1992; Tsuchiya *et al.*, 1992); (5) the *Drosophila* cell-cycle-dependent gene product of *lodestar* (Girdham & Glover, 1991; Laurent *et al.*, 1992); (6) the yeast excision repair gene RAD16 (Mannhaupt *et al.*, 1992); (7) a human (hSNF2) gene highly similar to SNF2, but not capable of complementing SNF2 or STH1-lacking mutants in yeast (Okabe *et al.*, 1992); (8) KYBP, a DNA-binding mouse protein (EMBL accession number X66028; V. Delmas & R. P. Perry, unpublished results); (9) YAL001, a yeast protein located on the left arm of chromosome 1 (Clark *et al.*, 1992); and (10) RAD5, a protein involved in DNA repair (Johnson *et al.*, 1992). Several of these research groups have already identified, based on several consensus motifs (Gorbalenya *et al.*, 1989), a remote relationship of this new family to helicases. Some of these proteins contain inserted DNA-binding domains (Mannhaupt *et al.*, 1992; Johnson *et al.*, 1992) or share a C-terminal domain with otherwise unrelated transcription activators (for review of proteins containing this so-called bromodomain, see PROSITE database and references therein; Bairoch, 1992).

Here, we report cloning and analysis of a partial sequence from *Bacillus cereus*, a new prokaryotic member of this family of SNF2-related proteins. Furthermore, sequence analysis of all members indicates that a viral sequence and the *hepA* gene product from *Escherichia coli* also belong to this subfamily of helicases.

A *B. cereus* cDNA library was prepared in λgt 11 as described by Huynh *et al.* (1988). Clones from this library were subcloned into the *Eco*RI site of pUC19 (Sambrook *et al.*, 1989) and used as anonymous probes in the physical mapping of *B. cereus* ATCC 10987 (Kolstø *et al.*, 1990). The probe BC203, localized on the 840 kb‡ *Not*I fragment of the chromosome, was sequenced using a fluorescence-based sequencer (Voss *et al.*, 1989). Nested deletions (Henikoff, 1984) of BC203 were prepared using the Bal31 deletion kit (Pharmacia, Sweden). The second strand was sequenced after subcloning of BC203 using oligonucleotide primers. Nucleotide sequence analysis (GCG software package: Devereux *et al.*,

† Author to whom all correspondence should be addressed.

‡ Abbreviations used: ORF, open reading frame; Hpb, hypothetical protein from *Bacillus cereus*; bp, base-pair(s); CN, chilo iridescent virus.

0022–2836/93/060684–05 $08.00/0

© 1993 Academic Press Limited

```
  1 LKLAKTYFINHIREFLSKVEKREAFHCSNEFTYTPDVHSFKQETDAIIQQ
 51 FIKIYHNEKMYEDALEVHAKQDESMIFIPPASWNDMLSALSRAEYVQLKQ
101 NEQLFQGLHISKGLLPLHFEFTKGNNGGFTLHIDGLNRVRVMEMYNNALY
151 DGKLYHLPMEDCMRLIELQKMMSRSNSNQFYIPENKMEHFVAKVVPGLMK
201 LGTVRIDEVISDRVETPSLKAKLYLDRVKNRLLAGLEFHYGNVVINPLEE
251 DGQPSVFNRDEKKEKEILDIMSESAFAKTEGGYFMHNEEAEYNFLYHIVP
301 TLKGLVDIYATTAIKLRIHKGDTAPLIRVRRKERIDWLSFRFDIKGIPEA
351 EIKGVLAALEEKRKYYRLANGSLLSLESKEFNEINQFVKESGIRKEFLHG
401 EEVNVPLIRSVKWMNGLHEGNVLSLDESVQDLVESIQNPKKLK-FTVPPT
    ....VPLEAGIADPKGLPE-ELVASRERERDFIQQMMDPSKAKPFKLPIA
MOT
450 LHAVMREYQVYGFEWMKTLAYYRFGGILADDMGLGKTLQSIAYI--DSVL
    IKATLRKYQQDGVNWLAFLNKYHLHGILCDDMGLGKTLQTICIIASDQYL
MOT
498 P--------EIREKKLPILVVSPSSLVYNWFSELKKFAPHIRAVIADGNQ
    RKEDYEKTRSVESRALPSLIICPPSLTGHWENEFDQYAPFLKVVVYAGGP
MOT
540 TERRKILKDVAEFDVVITSYPLLRRDVRSYARP-FHTLFLDEAQAFKNPT
    TVRLTLRPQLSDADIIVTSYDVARNDLAVLNKTEYNYCVLDEGHIIKNSQ
MOT
589 TQTARAVKTIQAEYRFGLTGTPVENSLEELWSIFHVVFPELLPGRK....
    SKLAKAVKEITANHRLILTGTPIQNNVLELWSLFDFLMPGFLGTEK....
MOT
```

Figure 1. Amino acid sequence of the predicted ORF in *Bacillus cereus* (Hpb) and alignment with the closest relative, yeast gene activator MOT1 (35% amino acid identity over 247 residues). No homology has been found yet for the N-terminal part of Hpb.

1984) revealed an open reading frame (ORF), Hpb (hypothetical protein from *B. cereus*: Fig. 1) covering the entire fragment of 1900 bp.

FASTA and TFASTA searches (Pearson & Lipmann, 1988) in SWISSPROT, PIR and EMBL sequence databases revealed a significant homology of the C-terminal part of Hpb (Fig. 1) with most of the proteins shown in Figure 2. Many of them have amino acid identities with Hpb of between 25 and 34% over long segments with only a few insertions or deletions. This is far above the threshold for structural homology (Sander & Schneider, 1991). To verify these results, all detected members were also subjected to FASTA and TFASTA searches. Interestingly, in most of the runs, several sequence segments in different reading frames of chilo iridescent virus (CIV: EMBL accession number M81388; G. Darai & K. C. Sonntag) scored extremely high as well (up to 33% identity over 100 residues). If only the conserved regions of this family are considered (Fig. 2), the similarity of CIV to the SNF2 family increases further (27 to 36% identity). Indeed, the gene product assembled from these fragments appears to encode a putative helicase which belongs to this family (G. Darai *et al.*, personal communication).

The proteins identified were subjected to a number of sequence analysis methods (as described by Bork *et al.*, 1992). Property patterns (Bork & Grunwald, 1990) of conserved boxes, as well as profile searches (Gribskov *et al.*, 1987) of larger fragments, were used to describe all known members and to separate them from other, more distantly related, helicases. Both the property patterns and the profiles significantly detected another prokaryotic protein, a DNA damage-induced putative helicase from *E. coli* (*hepA*), which is located downstream from the *polB* (Lewis *et al.*, 1991). However, the C terminus of the published *hepA* sequence does not match conserved motifs of any helicase subfamily, nor does the C terminus of *lodestar* (Girdham & Glover, 1991). Based on homology searches in DNA databases, we predict frameshifts for both proteins. The alternative translations result in longer proteins which perfectly match all conserved motifs (for details, see Bork & Koonin, 1993).

The most conserved regions of the family (Fig. 2) correspond to the motifs defined for many helicases (Gorbalenya *et al.*, 1989). The part most conserved in the SNF2 family includes motifs V and VI (Fig. 2). Interestingly, this region has the largest differences from the corresponding motifs of other helicase families. These differences may indicate specific DNA-binding functions.

A schematic dendrogram (Fig. 3), based on the multiple alignment of conserved regions (Fig. 2), reveals a clustering of the sequences which do not

```
                [---- I ----]          [----- Ia -----]          Ib
 Rad5   526 GGILSDEMGLGKTVA -47- LIVVPMSLLTQWSNEFTK -33- TVVLTTYGIV -20-
 Rad16  405 GGVLADEMGMGKTIQ -14- LVVAPTVALMQWKNEIEQ -28- DVVLTTYAVL -21-
 Ysnf2  786 NGILADEMGLGKTIQ -19- LVIVPLSTLSNWSSEFAK -31- DVVLTTFEYI - 4-
 Sth1   490 NGILADEMGLGKTIQ -19- LVIVPLSTITNWTLEFEK -31- DVLLTTYEYI - 4-
 YAL01  592 SCILADDMGLGKTCQ -17- LVVVPSSTLENWLREFQK -31- DVIVTTYNLA - 7-
 Brm    792 NGILADEMGLGKTIQ -19- LIIVPLSTLPNWVLEFEK -31- NVLLTTYEYV - 4-
 Hsnf2  174 NGILADEMGLGKTLQ -19- MVLVPKSTLHNWMNEFKR -32- DVCVTSYEMV - 4-
 Rad54  329 GCIMADEMGLGKTLQ -23- IIVCPSSLVNNWANELIK -45- PVLIISYETL - 4-
 Mot1  1291 HGILCDDMGLGKTLQ -30- LIICPPSLTGHWENEFDQ -29- DIIVTSYDVA - 4-
 Lode   460 GGILADDMGLGKTLT -53- LVVCPASLLRQWESEVES -29- DIVVTTYQIV - 7-
 Hepa   171 RVLLADEVGLGKTIE -18- LIIVPETLQHQWLVEMLR -30- QLVICSLDFA - 7-
 Hpb    474 GGILADDMGLGKTLQ -20- LVVSPSSLVYNWFSELKK -29- DVVITSYPLL - 3-
 Civ      ? GGIISLCMGLGKTLT - 0- ALAYSFQNKASFPTLVIT - ?- DIVITTYDVC -46-
 cons       tshhsDpMGLGKTht       hhhhP t h tW  Eh t       thhh oathh
 DEAD       phhhhstoGsGKT         hhhhPo thh Qh   h        thhhso sRh
```

Fig. 2.

```
                [---- II ----]                                  [-- III --]
 Rad5  SGLFSVNFYRIIIDEGHNIRNRTTVTSKAVMAL.QGKCK....WVLTGTPIINRLDDLYSLVKFLELDPWRQ
Rad16  SVLHNIDFYRVILDEAHNIKDRQSNTARAVNNL.KTQKR....WCLSGTPLQNRIGEMYSLIRFLNINPFTK
Ysnf2  ALLSKVKWVHMIIDEGHRMKNAQSKLSLTLNTHYHADYR....LILTGTPLQNNLPELWALLNFVLPKIFNS
 Sth1  SLLSKHDWAHMIIDEGHRMKNAQSKLSFTISHYYRTRNR....LILTGTPLQNNLPELWALLNFVLPKIFNS
YAL01  SFLKNRNFNVVVYDEGHMLKNSTSERFAKLMKI.RANFR....LLLTGTPLQNNLKELMSLLEFIMPNLFIS
  Brm  AVLAKIQWKYMIIDEGHRMKNHHCKLTQVLNTHYIAPYR....LLLTGTPLQNKLPELWALLNFLLPSIFKS
Hsnf2  SVFKKFHWRYLVIDEAHRIKNEKSKLSEIVREF.KSTNR....LLLTGTPLQNNLHELWALLNFLLPDVFNS
Rad54  DQLKNCNVGLMLADEGHRLKNGDS.LTFTALDSISCPRR....VILSGTPIQNDLSEYFALLSFSNPGLLGS
 Mot1  AVLNKTEYNYCVLDEGHIIKNSQSKLAKAVKEI.TANHR....LILTGTPIQNNVLELWSLFDFLMPGFLGT
 Lode  SAVFGVKWRRIILDEAHVVRNHKSQSSLAVCDL.RGKYR....WALTGTPIQNKELDVYALLKFLRCSPFDD
 Hepa  EHLCEAEWDLLVVDEAHHLVWSEDAPSREYQAIEQLAEHVPGVLLLTATPEQLGMESHFARLRLLDPNRFHD
  Hpb  VRSYARPFHTLFLDEAQAFKNPTTQTARAVKTI.QAEYR....FGLTGTPVENSLEELWSIFHVVFPELLPG
  Civ  AVIYGTPWERVICDESQKFANPKTMTYKCIMAV.YGKYK....WCLTGTPIRNYETDIWAQLRFCGYKGVER
 cons    h t ta hhhhDEst hh-tt     hh th   t +    hhLoGTPhtNt  -hashhthh  thh t
 DEAD           thhhDEADthhtsF   h  h              hhhSATh  t

                [------ IV ------]          [---------- V --------]
 Rad5  -268- QVVIFSQFSTYLDILEKELT -38- ILLLSLKAGGVGLNLTCASHAYMMDPWWS
Rad16  -262- KSIVFSQFTSMLDLVEWRLK -32- VFLVSLKAGGVALNLCEASQVFILDPWWN
Ysnf2  -156- RVLIFFQMTQIMDIMEDFLR -33- CFILSTRAGGLGLNLQTADTVIIFDTDWN
 Sth1  -157- RVLMFFQMTQVMDIMEDFLR -33- CFLLSTRAGGLGLNLQTADTVIIFDTDWN
YAL01  -213- KVLIFSLFTQVLDILEMVLS -32- IFILSTKAGGFGINLVCANNVIIFDQSFN
  Brm  -161- RVLLFCQMTQCMTIIEDYLG -33- VFLLSTRAGGLGLNLQTADTVVIFDSDWN
Hsnf2  -138- RVLIFSQMTRLLDILEDYCM -45- IFMLSTRAGGLGINLASADVVILYDSDWN
Rad54  -164- KIVLISNYTQTLDLIEKMCR -33- IFLLSSKAGGCGINLIGANRLILMDPDWN
 Mot1  -185- RALIFCQLKDMLDMVENDLF -35- CLLLTTKVGGLGLNLTGADTVIFVEHDWN
 Lode  -253- KAIVVSQWTSVLDILRDHLS -33- VLLLSLTAGGVGLNLIGANHLLLLDLHWN
 Kybp    ?- RVLIFSQMVRMLDILAEYLK -33- CFLLSTRAGGLGINLASADTVVIFDSDWN
 Hepa  -174- KLPLRCNWSRYCANVKVFAL -27- QVLLCSEIGSEGRNFQFASHMVMFDLPFN
  Civ  -  ?- KIIVFSMFTSCLDLLSEAIK -34- GLFLTYKVGSEGLNLTEATHCICIEPWWT
 cons       +hhhh -atthhthht  h       hhhho thGs GhNL tAtthhhh-  at
 DEAD        hhhh tt   h-hh   h         hhhsthhsRGh-htththhhtat

                [-- VI --]
 Rad5  PSMEDQAIDRLHRIGQTNSVKVMRFIIQDSIEEKMLRIQEKKRTIGE.AMD
Rad16  PSVEWQSGDRVHRIGQYRPVKITRFCIEDSIEARIIELQEKKANMIHATIN
Ysnf2  PHQDLQAQDRAHRIGQKNEVRILRLITTNSVEEVILERAYKKLDIDGKVIQ
 Sth1  PHQDLQAQDRAHRIGQKNEVRILRLITTDSVEEVILERAMQKLDIDGKVIQ
YAL001 PHDDRQAADRAHRVGQTKEVNITTLITKDSIEEKIHQLAKNKLALDSYISE
  Brm  PHQDLQAQDRAHRIGQRNEVRVLRLMTVNSVEERILAAARYKLNMDEKVIQ
Hsnf2  PQVDLQAMDRAHRIGQKKPVRVFRLITDNTVEERIVERAEIKLRLDSIVIQ
Rad54  PAADQQALARVWRDGQKKDCFIYRFISTGTIEEKIFQRQSMKMSLSSCVVD
 Mot1  PMNDLQAMDRAHRIGQKKVVNVYRIITKGTLEEKIMGLQKFKMNIASTVVN
 Lode  PQLEAQAQDRIYRVGQKKNVIIYKFMCVDTVEQRIKGLQDKKLDLADGVLT
 Kybp  PQNDLQAQARAHRIGQKKQVNIYRLVTKGSVEEDILERAKKKMVLDHLVIQ
 Hepa  PDLLEQRIGRLDRIGQAHDIQIHVPYLEKTAQSVLVRWYHEGLDAFEHTCP
  Civ  NAVHNQAKARLWRTGQTKQVYVHNVIIEGSIEEKIVEICKGKDDMAASYLE
 cons  Pt   Qs tRhaR GQ tth hhthhhttohEt hhth   K th t hht
 DEAD      ttahHRhGRtsR tt G s
```

Figure 2. Multiple alignment of all conserved boxes within the SNF2 family of helicases. Large length variation of sequence inserts between the boxes (numbers) are possibly due to insertions of other domains such as zinc fingers or double fingers. The boxes with similarities to other helicase subfamilies are indicated by roman numerals and a consensus is given for both the SNF2 and the DEAD-box family (conventions used: UPPER-CASE LETTERS, strictly conserved amino acid residues; h, hydrophobic residues; a, aromatic residues; o, serine/threonine; −/+, charged position; t, turn-like and probably located at the surface). Rad5 and Rad16 have large insertions between boxes III and VI due to the insertion of DNA-binding domains (Mannhaupt *et al.*, 1992; Johnson *et al.*, 1992). The amino acid sequences in the two C-terminal boxes of *lodestar* (PIR: A40580; Girdham & Glover, 1991) and *hepA* (SWISSPROT: Hepa_Ecoli) come from translated ORFs frameshifted relative to the N-terminal part of the proteins. These frameshifts suggested by homology searches will have to be checked by resequencing. The amino acid sequence segments of the iridescent virus (CIV) result from a translation of three unidentified, putative ORFs in the nucleotide sequence (EMBL accession number M81388). For *B. cereus* Hpb and mouse KYBP, only partial sequences are available.

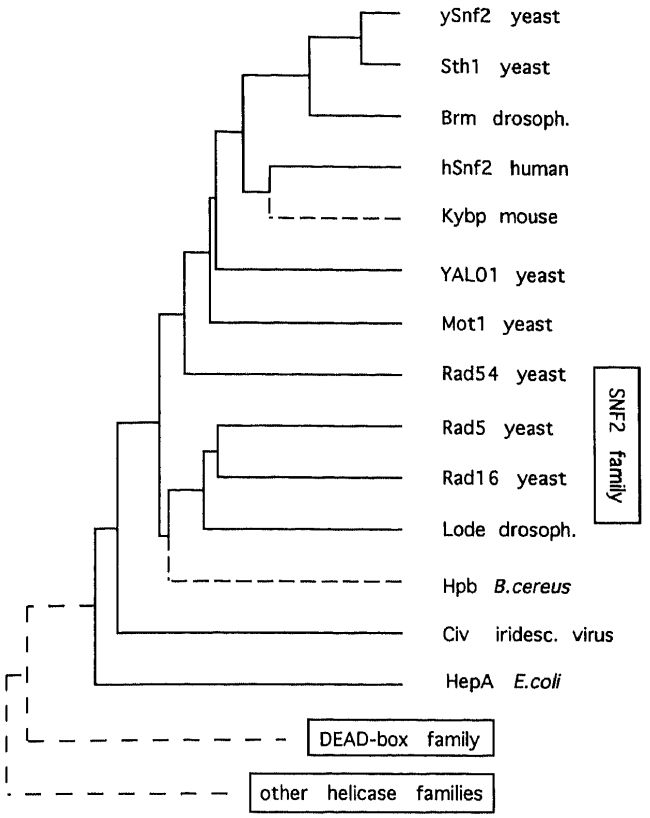

Figure 3. Dendrogram of the SNF2 family based on the conserved boxes shown in Fig. 2. The program PILEUP of the GCG package (Devereux *et al.*, 1984) was used. Dotted lines indicate partial sequences as well as the relation to some other helicase families. Although all members contain a DEAH-box-like motif of helicases, profile searches (Gribskov *et al.*, 1987) with the entire alignment reveal a closer relationship of the SNF2 family to the DEAD-box family (for review and nomenclature, see Schmid & Linder, 1992).

follow the taxonomic grouping of species. This is suggestive of a multigene family in all organisms as it is already known for yeast (SNF2, Rad5, Rad16, Rad54, Sth1, Yal1, Mot1) and *Drosophila* (Brm, Lode). Furthermore, the grouping of the *E. coli* and the *B. cereus* proteins, which seem to be non-orthologous (Fig.3), suggests the presence of more than one SNF2-like helicase in prokaryotes. For a quantitative phylogenetic analysis, at least some of the orthologous genes have to be identified in each species.

In spite of being a multigene family, the SNF2-related proteins can be separated from other helicase families by defined conserved regions (Fig. 2). The presence of prokaryotic and viral sequence in this family, as reported here, suggests a specific function for the SNF family. Indeed, all of the SNF2-related proteins appear to be nuclear proteins, are putative DNA helicases and might even be involved in transcription activation, as shown for SNF2, Mot1 or Brm.

We are grateful to G. Darai for communication of results prior to publication, Amos Bairoch for helpful information and E. Koonin for discussion.

References

Bairoch, A. (1992). Prosite: a dictionary of protein patterns and sites. *Nucl. Acids Res.* **11**, 2013–2018.

Bork, P. & Grunwald, C. (1990). Recognition of different nucleotide-binding sites using a property pattern approach. *Eur. J. Biochem.* **191**, 347–358.

Bork, P. & Koonin, E. (1993). An expanding family of helicases within the "DEAD/H" superfamily. *Nucl. Acids Res.* **21**(3), in the press.

Bork, P., Ouzounis, C., Sander, C., Scharf, M., Schneider, R. & Sonnhammer, E. (1992). Comprehensive sequence analysis of the 182 predicted ORFs of the yeast chromosome III. *Prot. Sci.* **1**, 1677–1690.

Carlson, C. R., Grønstadt, A. & Kolstø, A.-B. (1992). Physical maps of the genome of three *Bacillus cereus* strains. *J. Bacteriol.* **174**, 3750–3756.

Clark, M. W., Zhong, W. W., Keng, T., Stroms, R. K., Barton, A., Kaback, D. & Bussey, H. (1992). Identification of a *Saccharomyces cerevisiae* homolog of the SNF2 transcriptional regulator in the DNA sequence of an 8·6 kb region in the LTE1–CYS1 interval on the left arm of chromosome I. *Yeast*, **8**, 133–145.

Davies, J. L., Kunisawa, R. & Thorner, J. (1992). A presumptive helicase (MOT1 gene product) affects gene expression and is required for viability in the yeast *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **12**, 1879–1892.

Devereux, J., Haeberli, P. & Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucl. Acids Res.* **12**, 387–395.

Emery, H. S., Schild, D., Kellogg, D. E. & Mortimer, R. K. (1991). Sequence of RAD54, a *Saccharomyces cerevisiae* gene involved in recombination and repair. *Gene*, **104**, 103–106.

Girdham, C. H. & Glover, D. M. (1991). Chromosome tangling and breakage at anaphase result from mutations in *lodestar*, a *Drosophila* gene encoding a putative nucleoside triphosphate-binding protein. *Genes Devel.* **5**, 1786–1799.

Gorbalenya, A. E., Koonin, E. V., Domchenko, A. P. & Blinov, V. M. (1989). Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes. *Nucl. Acids Res.* **17**, 4713–4729.

Gribskov, M., McLachlan, A. D. & Eisenberg, D. (1987). Profile analysis: detection of distantly related proteins. *Proc. Nat. Acad. Sci., U.S.A.* **84**, 4355–4358.

Henikoff, S. (1984). Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene*, **28**, 351.

Huynh, T. V., Young, R. A. & Davis, R. W. (1988). Constructing and screening cDNA libraries in λgt10 and λgt11. In *DNA Cloning* (Glover, D. M., ed.), vol. 1, pp. 49–78, IRL Press, Oxford.

Johnson, R. E., Henderson, S. T., Petes, T. D., Prakash, S., Bankmann, M. & Prakash, L. (1992). *Saccharomyces cerevisiae* RAD5-encoded DNA repair protein contains DNA helicase and zinc-binding sequence motifs and affects the stability of simple repetitive sequences in the genome. *Mol. Cell. Biol.* **12**, 3807–3818.

Kolstø, A.-B., Grønstadt, A. & Oppegaard, H. (1990). Physical map of the *Bacillus cereus* chromosome. *J. Bacteriol.* **172**, 3021–3025.

Laurent, B. C., Treitel, M. A. & Carlson, M. (1991). Functional interpedence of yeast SNF2, SNF5, and SNF6 proteins in transcription activation. *Proc. Nat. Acad. Sci., U.S.A.* **88**, 2687–2691.

Laurent, B. C., Yang, X. & Carlson, M. (1992). An essential *Saccharomyces cerevisiae* gene homologous to SNF2 encodes a helicase-related protein in a new family. *Mol. Cell. Biol.* **12**, 1893–1902.

Lewis, L. K., Jenkins, M. E. & Mount, D. W. (1992). Isolation of DNA damage-inducible promoters in *E. coli*: regulation of *polB* (dinA), *dinG*, and *dinH* by LexA repressor. *J. Bacteriol.* **174**, 3377–3385.

Mannhaupt, G., Stucka, R., Ehnle, S., Vetter, I. & Feldmann, H. (1992). Molecular analysis of yeast chromosome II between CMD1 and LYS2: the excision repair gene RAD16 located in this region belongs to a novel group of double-finger proteins. *Yeast*, **8**, 397–408.

Okabe, I., Baily, L. C., Attree, O., Srinivasan, S., Perkel, J. M., Laurent, B. C., Carlson, M., Nelson, D. L. & Nussbaum, R. L. (1992). Cloning of human and bovine homologs of SNF2/SWI2: a global activator of transcription in yeast *S. cerevisiae*. *Nucl. Acids Res.* **20**, 4649–4655.

Pearson, W. R. & Lipmann, D. J. (1988). Improved tools for biological sequence comparison. *Proc. Nat. Acad. Sci., U.S.A.* **85**, 3338–3342.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning. A Laboratory Manual*, 2nd edit, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sander, C. & Schneider, R. (1991). Database of homology-derived protein structures and the structural meaning of sequence alignment. *Proteins*, **9**, 56–68.

Schmid, S. R. & Linder, P. (1992). D-E-A-D protein family of putative RNA helices. *Mol. Microbiol.* **6**, 283–292.

Tamkun, J. W., Deuring, R., Scott, M. P., Kissinger, M., Pattatucci, A. M., Kaufman, T. C. & Kennison, J. A. (1992). *brahma*: a regulator of *Drosophila* homeotic genes structurally related to the yeast transcription activator SNF2/SWI2. *Cell*, **68**, 561–572.

Tsuchia, E., Uno, M., Kiguchi, A., Masuoka, K., Kanemori, Y., Okabe, S. & Mikayawa, T. (1992). The *S. cerevisiae* NPS1 gene, a novel CDC gene which encodes a 160 kDa nuclear protein involved in G_2 phase control. *EMBO J.* **11**, 4017–4026.

Voss, H., Schwager, C., Kristensen, T., Duthie, S., Olsson, A., Erfle, H., Stegemann, J., Zimmermann, J. & Ansorge, W. (1989). One-step reaction protocol for automated DNA sequencing with T7 DNA polymerase results in uniform labeling. *Methods Mol. Cell. Biol.* **1**, 155–159.

Edited by J. Karn

Note added in proof. After acception of this manuscript, the sequence of the human DNA repair gene ERCC6, encoding yet another member of the family described here, has been published (Toelstra, C., Van Gool, A., de Wit, J., Vermeulen, W., Bootsma, D. & Hoejmakers, J. H. J. (1992). ERCC6, a member of a subfamily of putative helicases, is involved in Cockayne's syndrome and preferential repair of active genes. *Cell* **71**, 939–953). This putative helicase is involved in Cockayne's syndrome and preferential repair of active genes. The probable frameshift in LDR has also been noticed.