J. Mol. Biol. (1994) **236**, 1277–1282

Drosophila kelch Motif is Derived from a Common Enzyme Fold

Peer Bork†

European Molecular Biological Laboratory, 69117 Heidelberg, Germany and Max-Delbrück-Centre of Molecular Medicine, 13125 Berlin-Buch

and Russell F. Doolittle

Center for Molecular Genetics, University of California San Diego, La Jolla, CA, 92093-0634, U.S.A.

A systematic screening of sequence databases with a motif hitherto found only in animal and poxvirus proteins has revealed a trail leading back to prokaryotes. Fortuitously, an X-ray structure is available for one of the identified sequences and shows the fundamental fold to be a set of circularly arranged β sheets. This structure may be very widely distributed throughout the biological world in sialidases and some other enzymes. In bacteria, a mobile noncatalytic domain is often associated with these same enzymes.

Keywords: protein modules; homology; propeller fold; galactose oxidase; *kelch* motif

We have been using commonly repeated sequence motifs to screen protein sequence databases in an effort to find where different domains first made their appearance during evolution (Bork, 1992; Doolittle & Bork, 1993). Our general strategy is to make an alignment of amino acid sequences known to be related through common ancestry and then construct a corresponding property pattern (Rohde & Bork, 1993). The pattern is then looked for in various sequence databases. The approach has been especially successful when applied to commonly occurring sequences like the epidermal growth factor (EGF) and fibronectin type III (Fn3) domains (Bork & Doolittle, 1992; Campbell & Bork, 1993). The use of the property pattern instead of a simple sequence, or even a consensus sequence, often allows the detection of related sequences with low degrees of similarity.

The rate of change of most amino acid sequences during evolution has been such that even pattern searching has its limits in such an exploration. Eventually a point is reached where most of the definitive information is lost. Although it is common knowledge that three-dimensional structural attributes are more strongly constrained during protein evolution than their underlying amino acid sequences, the reality is that the number of known three-dimensional structures is dwarfed by the number of known sequences. Clearly, a practical strategy will combine both sorts of information whenever possible and use them in a complementary way. We now report the results of such a study and the uncovering of the three-dimensional folding pattern of a known sequence motif for which no structural information previously existed.

The study was initially begun as a survey of an unusual repetitive pattern, the 50-residue motif *kelch*, named after the *Drosophila melanogaster* mutant in which it was first identified and where it is found six times in an egg chamber regulatory protein (Xue & Cooley, 1993). A conventional sequence search by the discoverers of *kelch* (Xue & Cooley, 1993) had revealed the same motif in the mouse protein MIPP (Chang-Yeh *et al.*, 1991), as well as in varying sized clusters in poxviruses (Goebbel *et al.*, 1990). Based on an initial alignment prepared from six *Drosophila* eggchamber protein (*kelch*) segments, four from the mouse MIPP and 18 assorted segments from various poxviruses, our own search of this motif by pattern analysis has uncovered several other putative *kelch*-like sequences, including a somewhat reduced set in the yeast protein RAL2 (Fukui *et al.*, 1989) and an ectomelia virus (Senkevich *et al.*, 1993). Most startling, however, was a clear relationship to several, so far unidentified, sequence repeats in galactose oxidase from the fungus *Dactylium dendroides* (Figure 1). Fortuitously, an X-ray

† Author to whom all correspondence should be addressed at: EMBL, Meyerhofstr. 1, 69117 Heidelberg, Germany.

0022-2836/94/101277-06 $08.00/0

© 1994 Academic Press Limited

```
VC02_VACCC-1    FYNANVLHNCLYIIGGMINNR.01.VYSVSRVDLE.TKKWKTVTNMSSLKS
VC02_VACCC-2    EVSTCVNDGKLYVIGGLEFSI....STGVAEYLKHGTSKWIRLPNLITPRY
VC02_VACCC-3    IVYAVEYDGDIYVITGITHET....RNYLYKYIVK.EDKWIELYMYFNHVG
VF03_VACCC-1    MCTSTHVGEVVYLIGGWMNNE.01.HNNAIAVNYI.SNNWIPIPPMNNSPR
VF03_VACCC-2    YATGIPANNKLYVVGGLPNPT......SVERWFHG.DAAWVNMPSLLKPRC
VF03_VACCC-3    NPAVASINNVIYVMGGHSETD......TTTEYLLPNHDQWQFGPSTYYPHY
VA55_VACCC-1    SFAVAVLDNIIYMMGGYDQSP.02.SSKVIAYNTC.TNSWIYDIPELKYPR
VA55_VACCC-2    NCGGLADDEYIYCIGGIRDQD.03.TSSIDKWKPS.KPYWQKYAKMREPKC
VA55_VACCC-3    NMSTIVHDGKIYISGGYNNSS.05.SNLVLSYNPI.YDEWTKLSSLNIPRI
Kelc/Drome-1    RKPVG.MPKILLVIGGQAPKA....IRSVEWYDLR.EEKWYQAAEMPNRRC
Kelc/Drome-2    RSGLSVLGDKVYAVGGFNGSL.01.VRTVDVYDPA.TDQWANCSNMEARRS
Kelc/Drome-3    TLGVAALNGCIYAVGGFDGTT.01.LSSAEMYDPK.TDIWRFIASMSTRRS
Kelc/Drome-4    SVGVGVVHGLLYAVGGYDGFT.03.LSSVERYNPD.TDTWVNVAEMSSRRS
Kelc/Drome-5    GAGVGVLNNILYRVGGHDGPM.01.RRSVEAYDCE.TNSWRSVADMSYCRR
Kelc/Drome-6    NAGVVAHDGLLYVVGGDDGTS.01.LASVEVYCPD.SDSWRILPALMTIGR
Est/Hs3033      TVKCAAVGSIVYVLAGFQGVG.01.LGHILEYNTE.TDKWVANSKVRAFPV
Est/Hs0651      GVGIATVMGKIFAVGGHNGNA.01.LNTVEAFEPV.LNRWELVGSVSHCRA
MIPP_Mouse-1    GLGVCVCYGAIYALGGWVGAE.01.GNTIERFDPD.ENKWEVVGSMAVSRY
MIPP_Mouse-2    YFGCCEMQGLIYAVGGISNEG.02.LRSFEVYDPL.SKRWSPLPPMGTRRA
MIPP_Mouse-3    YLGVAALNDCIYAIGGWNETQ.02.LHTVEKYSFE.EEKWVEVASMKVPRA
MIPP_Mouse-4    GMCAVTVNGLLYVSGGRSSSx.10.SDSVEVYNPH.SDTWTEIGNMIVSRC

Oxid/Staur-2    AGHSFLEDGRLLITGGHVDSH.02.VPDAIIFNPK.SGAWDNVPDMNDKRW
Gaoa_Dacde-2    PGISMDGNGQIVVTGGNDAKK.......TSLYDSS.SDSWIPGPDMQVARG
Glyo/Phach-2    ASGALLSNGTMVSMGGTPGGT.10.NQAIRIFEPC7CTLFEDPATVHLLEE
Oxid/Staur-3    PNNTTLANGDVLVLSGETDGE.03.NELPQRYVAA.TELLAEPDHRAEENS
Gaoa_Dacde-3    QSSATMSDGRVFTIGGSWSGG.02.EKNGEVYSPS.SKTWTSLPNAKVNPM
Glyo/Phach-3    PSSVRIFDGSLMIIGGSHVLT......PFYNVDPANSFEFFPSKEQTPRPS
Oxid/Staur-6    HNTTFLPDGKVLVTGGSRLEG.07.VLFPEVWDPE.TNVWKKLASNNAYRG
Gaoa_Dacde-6    HTSVVLPDGSTFITGGQRRGI.07.VFTPEIYVPE.QDTFYKQNPNSIVRV
Glyo/Phach-6    PELVHVPNGQILITNGAGTGF.10.GNSNADHPVL.TPSLYTPDAPLGKRI

consensus       t  hh  tG hYhhGG  tt        h  Y    tt W   tt    R
beta strands     1111   22222               33333     4444
```

Figure 1. Alignment of repetitive *kelch* pattern from *D. melanogaster*, mouse, poxviruses, a fungal galactose oxidase and a bacterial ORF. In general, the most conserved positions are structurally equivalent positions in the X-ray structure of galactose oxidase. Protein names correspond to SWISSPROT codes if available. Amino acids that are conserved in more than 60% of the repeats are shown in bold. On the consensus line: *h*, hydrophobic; *t*, turn. The 4 β-strands as found in galactose oxidase (Ito *et al.*, 1991) are indicated by 1 to 4. In the mouse MIPP sequence (only a fragment is known so far; Chang-Yeh *et al.*, 1991; Xue & Cooley, 1993), x denotes a putative frameshift that extends the open reading frame by 30 residues and continues the high sequence resemblance to *kelch* repeats. The 2 human expressed sequence tags (Ests) are from a brain library (Adams *et al.*, 1993) and hint at further paralogues in animals. When using galactose oxidase as a query sequence for (T)FASTA, (T)BLAST database searches, MIPP ranks at the top (e.g. FASTA opt. score of 128; Pearson & Lipman, 1988; BLAST p value 2.1e-2; Altschul *et al.*, 1990). These scores are only indicative of a relationship. The similarity becomes obvious when searching sequence databases with a consensus pattern (Rohde & Bork, 1993) derived from an alignment of *kelch*, MIPP and motifs from related virus proteins. The repeats in galactose oxidase can be then clearly distinguished from the background of unrelated proteins.

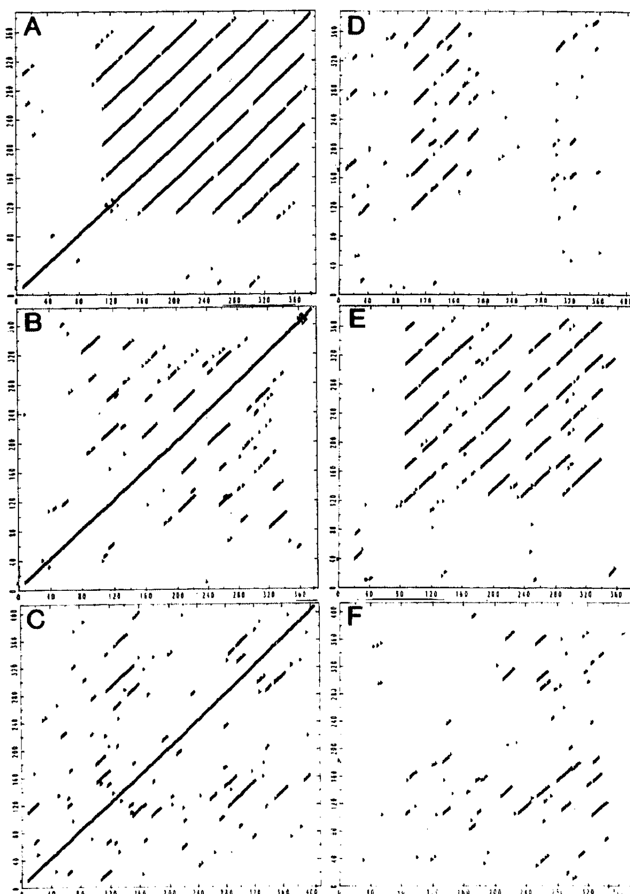

Figure 2. Diagonal plots of *kelch* protein, poxvirus repeats and fungal galactose oxidase sequences. These plots (Doolittle, 1986) use a moving window of 25 residues and have a cutoff score of 220 based on the original Dayhoff Mutation Matrix (Dayhoff, 1972). A, self-diagonal of *kelch*; B, self-diagonal of R55 poxvirus protein; C, self-diagonal of fungal galactose oxidase; D, *kelch versus* galactose oxidase; E, *kelch versus* poxvirus R55; F, poxvirus *versus* galactose oxidase.

structure has been reported for this enzyme (Ito *et al.*, 1991). It is a mostly β structure composed of three domains, the central one of which consists of seven four-stranded sheets, each about 50 residues long. Indeed, the repeats identified by the *kelch* patterns match exactly the seven β sheets, as is effectively visualized by diagonal plots (Figure 2). Thus, the structural features of galactose oxidase are also mirrored at the sequence level. Surprisingly, the domain with the *kelch* patterns is the catalytic unit in the galactose oxidase.

A search of the galactose oxidase sequence (McPherson *et al.*, 1992) by the (T)FASTA and (T)BLAST procedures (Pearson & Lipman, 1988; Altschul *et al.*, 1990) identified two other similar sequences, this time in the EMBL database. The first corresponded to a bacterial DNA sequence in the region adjacent to the nodC gene of *Stigmatella aurantiaca*. It is probably not a mere coincidence that *nodC* encodes a chitin synthase (Atkinson & Long, 1992). In any event, there is little doubt that the bacterial sequence encodes an enzyme equivalent to that found in fungi. Not only is it 28% identical in sequence with the *D. dendroides* enzyme, a similarity typical of eukaryotic–prokaryotic sequence comparisons, but it also retains virtually all the key ligand-binding features of the fungal enzyme (Figure 3A). The only possible exception is the replacement of Trp290 by a Thr, a position thought to be a part of the active site (Ito *et al.*, 1991). The other sequence was the extracellular glyoxal oxidase of the fungus *Phanerochaete chrysosporium* (Kersten & Cullen, 1993). Although no sequence similarity to galactose oxidase (or any other protein) was reported, the similar features of these enzymes prompted us to make an alignment, including the prokaryotic sequence described above (Figure 3A). The alignment reveals that all conserved residues involved in catalysis and ligand binding are conserved; although again with the exception of Trp290 (galactose oxidase) which is replaced by a His in glyoxal oxidase. Thus, in spite

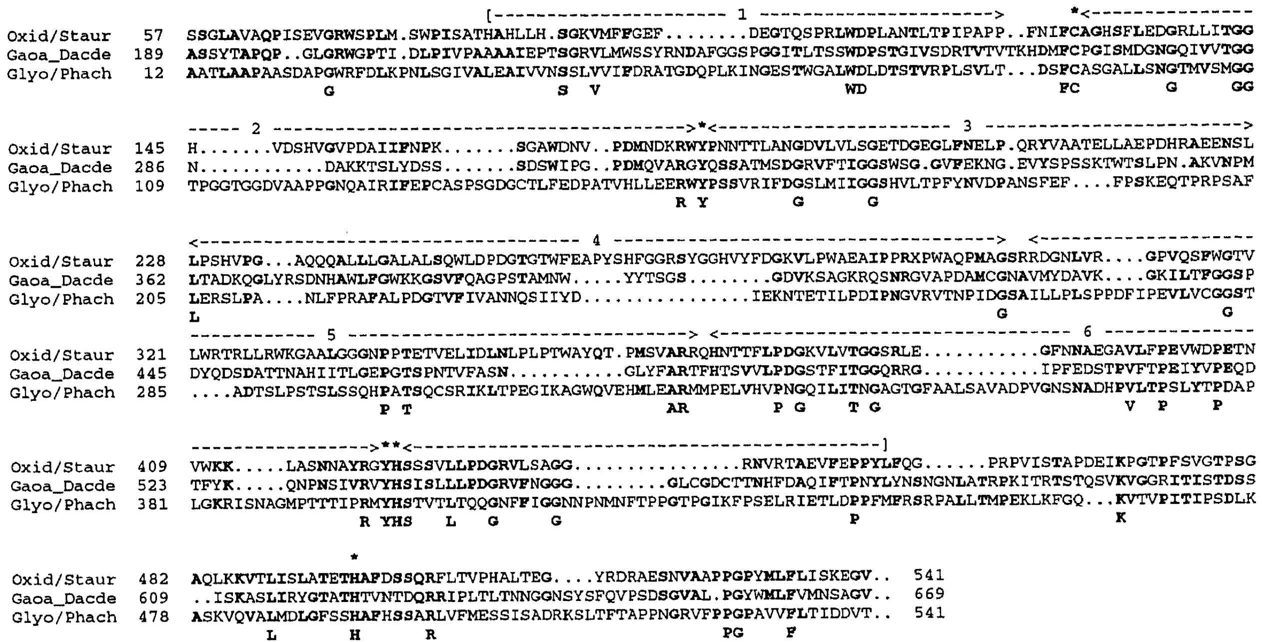

Figure 3. A, Multiple alignment of galactose oxidase sequence (Gaoa_Dacde) from *Dactylium dendroides* with that from an open reading frame (Oxid/Staur) from *Stigmatella aurantiaca* (A. Pospiech *et al.*, unpublished results; EMBL accession number Z11601) and with glyoxal oxidase (Glyo/Phach) from *Phanerochaete chrysosporium* (Kersten & Cullen, 1993). Residues involved in catalysis/ligand binding (Ito *et al.*, 1991) are marked by an asterisk. The 7 regions in GaoxD that correspond to the *kelch* pattern are indicated by arrows. Note that the seventh segment is shorter than the others; the X-ray structure (Ito *et al.*, 1991) makes it clear that this segment closes the barrel and is completed with residues preceding segment 1. The bacterial sequence represents a translation of 2 different frameshifted open reading frames 1701 bp downstream of *nodC* on the opposite strand (x denotes the predicted frameshift). Since the ORF begins next to *nodC*, the presence of an amino-terminal domain similar to the fungal galactose oxidase can be excluded.

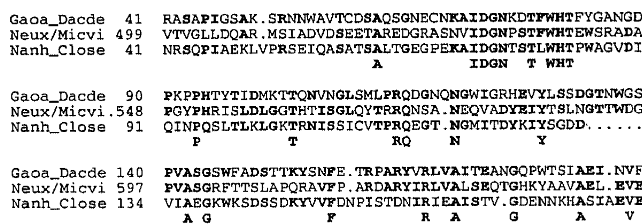

Figure 3. B, Multiple alignment of the amino-terminal domain of galactose oxidase from *D. dendroides* with the noncatalytic domains of *M. viridifaciens* (Sakurada *et al.*, 1992) neuraminidase (Neux/Micvi) and *C. septicum* (Rothe *et al.*, 1991) sialidase (Nanh_Close).

of sequence similarities below 30%, we are able to predict for glyoxal oxidase and the putative ORF in *Stigmatella aurantiaca*: (1) molecular details of the catalytic mechanism and metal binding; (2) functionally important residues; and (3) the topology based on the alignment with galactose oxidase.

The crystal structure of the catalytic domain of galactose oxidase has a fold similar to that observed in several other known enzymes: the heavy chain of methylamine dehydrogenase from the bacteria *Thiobacillus versutus* (Vellieux *et al.*, 1989) and *Paracoccus denitrificans* (Chen *et al.*, 1992), methanol dehydrogenase from *Methylophilus methylotrophus* (Xia *et al.*, 1992) and the neuraminidase (sialidase) from influenza virus (Varghese *et al.*, 1983). All of them are composed of circularly arranged four-stranded β-sheets; the influenza neuraminidase contains six, galactose oxidase and methylamine dehydrogenase are both composed of seven and methanol dehydrogenase even contains eight such β sheets. The structure has been referred to as a "propeller arrangement" (Varghese *et al.*, 1983; Murzin, 1992) or a "superbarrel" (Brånden & Tooze, 1991). There is no significant sequence similarity between any of these enzymes for which this structure has been observed. Since the *kelch* motif shares certain conserved features with the repeats in galactose oxidase, one can also expect common structural features, such as an open channel through the center of the propeller. This channel is closed by the third domain of galactose oxidase; the N-terminal domain of the *kelch* protein might have an equivalent function.

Like the fungal galactose oxidase, the *Stigmatella* sequence has an apparent signal peptide, implying an extracellular role. On the other hand, comparison with the fungal sequence revealed that the bacterial version completely lacks the amino-terminal domain, a structure initially called a β sandwich (Ito *et al.*, 1991), but which is also structurally similar to a "jelly-roll" barrel (Richardson, 1981). In any case, we searched the sequence of this segment separately by the (T)FASTA and (T)BLAST procedures on the chance that the domain was not only expendable but also evolutionarily mobile. Indeed, two other occurrences were found, both in prokaryotic extracellular enzymes

(Figure 3B). One of these was a neuraminidase from *Clostridium septicum* (Rothe *et al.*, 1991) and the other a sialidase from *Micromonospora viridifaciens* (Sakurada *et al.*, 1992). In the case of the sialidase from *C. septicum*, the related domain, which has 33% amino acid identity with the galactose oxidase domain over the course of its 150 residues, occurs at the amino terminus, but in the case of the enzyme from *M. viridifaciens*, where the resemblance amounts to 34% amino acid identity, it occurs at the carboxyl terminus (Figure 4A). Although the catalytic domains of these two bacterial sialidases are clearly similar to each other (32% identical), in neither case is there any obvious sequence resemblance with the catalytic domain of galactose oxidase. The implication is that the subsidiary domains have been shuffled into their present locations after the divergence of the catalytic domains, in the case of the sialidases, and wholly independently, in the case of the galactose oxidases. They might serve as carbohydrate binding domains, thus increasing the affinity for the substrates. Evolutionarily mobile domains have been noted in other bacterial extracellular glycohydrolases (Gilkes *et al.*, 1991; Meinke *et al.*, 1991; Bork & Doolittle, 1992).

There is reason to believe that all sialidases are related to each other (Taylor *et al.*, 1992), and indeed many of these enzymes have amino acid sequences that are so similar that they can be readily aligned. An important exception is the neuraminidase from influenza virus, which has been assigned to its own group in the sequence classification scheme (Henrissat, 1991). For this sialidase, a structure has been known for many years (Varghese *et al.*, 1983). Nonetheless, it has already been predicted (Warner *et al.*, 1992) that all sialidases will have the propeller-like structure observed in the influenza enzyme. The following circumstantial evidence provides support for this hypothesis.

The catalytic domains of the two bacterial sialidases that contain the accessory domain have previously been reported (Rothe *et al.*, 1991; Sakurada *et al.*, 1992) to be related to a wide variety of carbohydrate-involved enzymes and proteins, including neuraminidases (sialidases), invertases, a bacterial hemagglutinin, and a multigene family purported to have sialidase activity in the protozoan *Trypanosoma cruzii* (Uemura *et al.*, 1992). All these proteins have characteristic 12-residue signature sequences, denoted "Asp boxes" or BNR repeats (Rothe *et al.*, 1991) and belong to family 33 in the sequence-based classification of glycohydrolases compiled by Henrissat (1991). What is most intriguing, however, is that many of these boxes are spaced at approximately 50-residue intervals, just as are the repeats from the *kelch* protein and the propeller structures exemplified by galactose oxidase, methylamine dehydrogenase, methanol dehydrogenase and the influenza sialidase. Although it may only be a coincidence that the mobile subsidiary domain observed in galactose oxidase occurs also in some of these sialidases, a protein with the

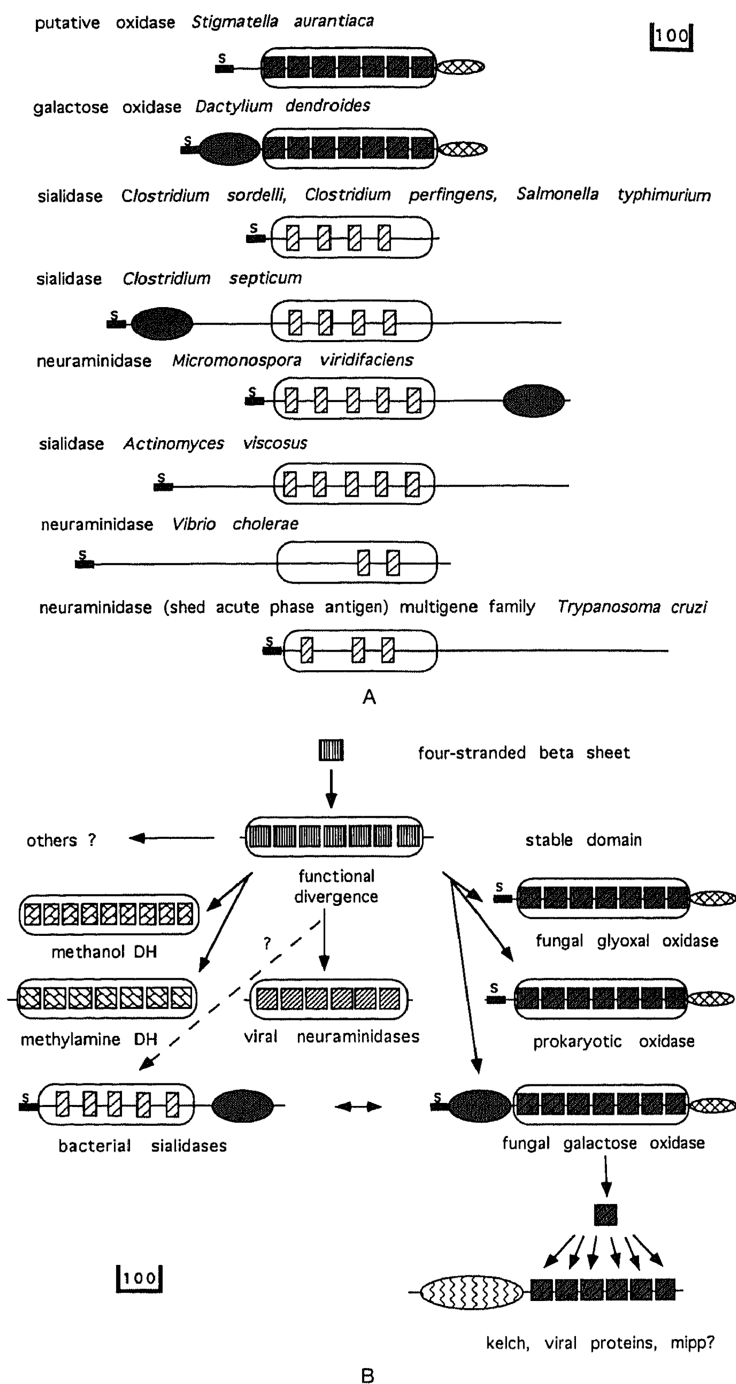

Figure 4. A, Modular architecture of galactose oxidase and various sialidases belonging to Henrissat's (1991) family 33 glycohydrolases. Signal sequences are denoted by s; the shaded ovals depict the evolutionarily mobile domain. B, Possible scenario for the evolution of sialidases, methanol dehydrogenase, methylamine dehydrogenase, galactose oxidase (GAO) and assorted *kelch*-containing proteins.

same three-dimensional fold as influenza virus sialidase, it seems unlikely that those proteins would also contain similarly spaced repeats merely by chance. After submission of this manuscript, the crystal structure of a bacterial sialidase from *Salmonella typhimurium* (Crennell *et al.*, 1993) was deposited in the prerelease of the Brookhaven Protein Databank (Bernstein *et al.*, 1977). It indeed contains a propeller-like structure similar to those

discussed above that contains six four-stranded β-sheets. However, structural comparison (data not shown) using the DALI program (Holm & Sander, 1993) revealed that the conserved residues in sialidases (Asp boxes) are not only different in sequence from the *kelch* motifs but are also located in different regions in the three-dimensional structure. There are at least two explanations as to why the sialidase repeats appear more similar to each other than they are to those in the other proteins under discussion here. First, it may be that gene conversion has been involved, the result of internal duplications of one or more of the units. The second possibility is that one or more of the four-stranded units was acquired by another gene and duplicated in the new setting. In either case, the result would be similar three-dimensional structures with only marginal sequence similarity. Given the complex arrangement of the propeller structure and the need for further duplication of an ancestral four-stranded β-unit, the possibility that all propeller folds share a remote ancestor cannot be excluded, even though their amino acid sequences have diverged to the point where their heritage is not readily apparent.

As in the case of some other proteins that have similar shapes in the absence of significant sequence resemblance (TIM barrels or immunoglobulin-like beta-barrels, for instance), the possibility of structural convergence must be considered. Convergence of any sort is, however, an arguable phenomenon under any circumstances (Doolittle, 1994). In the absence of transitional forms with detectable similar sequences, a final decision about divergence or convergence will have to be posponed.

Quantitative comparison of *kelch* repeats (animal, fungal and those from poxviruses), shows that all the animal sequences are more similar to each other than they are to the rest, implying a more recent divergence. Moreover, the *kelch* repeats within the set of six found in *D. melanogaster* are much more similar to each other than are any of the others (Figures 1 and 2) and are obvious recent descendants of a single four-stranded unit. Comparison of the mouse MIPP and *D. melanogaster kelch* (Xue & Cooley, 1993) allows a rough estimate of how fast these units are changing. At 38% identity, they could be either fast-mutating orthologues or more modestly changing paralogues. It is also possible that the mouse MIPP is a very special case, caught as it is downstream of a long terminal repeat (LTR) from an endogenous retrovirus (Chang-Yeh *et al.*, 1991). However, the repeats are clearly descended from a *kelch*-like ancestor.

As for the repeats in poxviruses, it is likely that they were acquired in recent times from some mammalian host and that the sequence resemblance is now decaying rapidly (Goebbel *et al.*, 1990). The situation is reminiscent of ankyrin repeats, which are also very common in poxviruses (Michaely & Bennett, 1992; Bork, 1993) and are also the likely result of a recent acquisition from an animal host. In any case, the modest degree of similarity among the animal and virus repeats is consistent with a relatively fast-changing sequence where resemblances are blurred beyond recognition when deep divergences such as fungi–animal or prokaryote–eukaryote are considered.

In summary, a search for a repetitive motif found in animal proteins has uncovered a resemblance to a fungal enzyme with a known three-dimensional structure. This in turn led to the discovery of an unknown prokaryotic equivalent. A major difference between the latter two is the presence of a domain in one that is absent in the other. Further searching revealed the missing domain in different locations in other prokaryotic enzymes, allowing immediate comparison of those segments with a known three-dimensional structure. Beyond that, the similarly repetitive structures in the main body of all these proteins suggest they may be all-β propeller or superbarrel structures. A scenario describing how these various structures might have arisen is outlined in Figure 4B.

We are grateful to Dr P. F. Knowles for sending us the coordinates for galactose oxidase from *D. dendroides* to E. V. Koonin for sending us a preprint of their work on ectomelia virus and to Liisa Holm for support in structural superposition. P.B. was supported by grants from Boehringer Ingelheim Fonds and Deutsche Forschung Gemeinschaft, and R.D. by the NIH.

References

Adams, M. D., Kervalage, A. R., Fields, C. & Venter, J. C. (1993). 3400 New expressed sequence tags identify diversity of transcripts in human brain. *Nature Genet.* **4**, 256–267.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410.

Atkinson, E. M. & Long, S. R. (1992). Homology of *Rhizobium meliloti* NodC to polysaccharide polymerizing enzymes. *Mol. Plant–Microbe Interact.* **5**, 439–442.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). The Protein Data Bank: a computer based archival file for macromolecular structures. *J. Mol. Biol.* **112**, 535–542.

Bork, P. (1992). Mobile modules and motifs. *Curr. Opin. Struct. Biol.* **2**, 413–421.

Bork, P. (1993). Hundreds of ankyrin-like repeats in functionally divers proteins: Mobile modules that cross phyla horizontally? *Proteins*, **17**, 363–374.

Bork, P. & Doolittle, R. F. (1992). Proposed acquisition of an animal protein by bacteria. *Proc. Nat. Acad. Sci., U.S.A.* **89**, 8990–8994.

Bränden, C.-I. & Tooze, J. (1991). *Introduction to Protein Structure*, Garland Publishing, Inc., New York & London.

Campbell, I. D. & Bork, P. (1993). EGF-like modules. *Curr. Opin. Struct. Biol.* **3**, 385–393.

Chang-Yeh, A., Mold, D. E. & Huang, R. C. C. (1991). Identification of a novel murine IAP-promoted placenta-expressed gene. *Nucl. Acids Res.* **19**, 3667–3672.

Chen, L., Methews, F. S., Davidson, V. L., Huizinger, E. G., Vellieux, F. M. D. & Hol, W. G. (1992). Three-dimensional structure of the quinoprotein methylamine dehydrogenase from *Paracoccuc denitrificans* determined by molecular replacement at 2·8 Å resolution. *Proteins*, **14**, 288–299.

Crennell, S. J., Garman, E. F., Laver, W. G., Vimr, E. R. & Taylor, G. L. (1993). Crystal structure of a bacterial sialidase (from *Salmonella typhimurium* LT2) shows the same fold as an influenza virus neuraminidase. *Proc. Nat. Acad. Sci., U.S.A.* **90**, 9852–9856.

Dayhoff, M. O. (1972). *An Atlas of Protein Sequence and Structure*, vol. 5, National Biomedical Research Corp., Washington, DC.

Doolittle, R. F. (1986). *Of Urfs and Orfs: a Primer in How to Analyze DNA Sequences*, University Science Press, Mill Valley, CA, U.S.A.

Doolittle, R. F. (1994). Convergent evolution: the need to be explicit. *Trends Biochem. Sci.* **19**, 15–18.

Doolittle, R. F. & Bork, P. (1993). *Sci. Amer.* **269**, 50–56.

Fukui, Y., Miyake, S., Satohand, H. & Yamamoto, M. (1989). Characterization of the *Schizosaccharomyces pombe ral2* gene implicated in the activation of the ras1 gene product. *Mol. Cell Biol.* **9**, 5617–5622.

Gilkes, N. R., Henrissat, B., Kilburn, D. G., Miller, R. C., Jr & Warren, R. A. J. (1991). Domains in microbial 1,4-glycanases: sequence conservation, function and enzyme families. *Microbiol. Rev.* **55**, 303–315.

Goebbel, S. J., Johnson, G. G., Perkins, M. E., Davis, S. W., Winslow, J. P. & Paoletti, E. (1990). The complete DNA sequence of vaccinia virus. *Virology*, **179**, 247–266.

Henrissat, B. (1991). A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.* **280**, 309–316.

Holm, L. & Sander C. (1993). Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123–138.

Ito, N., Phillips, S. E. V., Stevens, C., Ogel, Z. B., McPherson, M. J., Keen, J., Yadav, K. D. S. & Knowles P. F. (1991). Novel thioether bond revealed by a 1·7 Å crystal structure of galactose oxidase. *Nature (London)*, **350**, 87–90.

Kersten, P. J. & Cullen, D. (1993). Cloning and characterization of a cDNA encoding glyoxal oxidase, a H_2O_2-producing enzyme from the lignin-degrading basimycete *Phanerochaete chrysosporium*. *Proc. Nat. Acad. Sci., U.S.A.* **90**, 7411–7413.

McPherson, M. J., Ogel, Z. B., Stevens, C., Yadav, K. D. S., Keen, J. & Knowles, P. F. (1992). Galactose oxidase of *Dactylium dendroides*. *J. Biol. Chem.* **267**, 8146–8152.

Meinke, A., Gilkes, N. R., Kilburn, D. G., Miller, R. C., Jr & Warren, R. A. J. (1991). Bacterial cellulose-binding domain-like sequences in eucaryotic polypeptides. *Prot. Seq. Data Anal.* **4**, 349–353.

Michaely, P. & Bennett, V. (1992). The ANK repeat: an ubiquitous motif involved in macromolecular recognition. *Trends Cell Biol.* **2**, 127–129.

Murzin, A. G. (1992). Structural principles for the propeller assembly of β-sheets: the preference for seven-fold symmetry. *Proteins*, **14**, 191–201.

Pearson, W. R. & Lipman, D. J. (1988). Improved tools for biological sequence comparisons. *Proc. Nat. Acad. Sci., U.S.A.* **85**, 2444–2448.

Richardson, J. S. (1981). Anatomy and taxonomy of protein structure. *Advan. Protein Chem.* **34**, 167–339.

Rohde, K. & Bork, P. (1993). A fast, sensitive pattern-matching approach for protein sequences. *Comp. Appl. Biosci*, **9**, 183–189.

Rothe, B., Rothe, B., Roggentin, P. & Schauer, R. (1991). The sialidase gene from *Clostridium septicum*: cloning, sequencing, expression in *E. coli* and identification of conserved sequences in sialidases and other proteins. *Mol. Gen. Genet.* **226**, 190–197.

Sakurada, K., Ohta, T. & Hasegawa, M. (1992). Cloning, expression, and characterization of the *Micromonospora viridifaciens* neuraminidase gene in *Streptomyces lividans*. *J. Bacteriol.* **174**, 6896–6903.

Senkevich, T. G., Muravnik, G. L., Pozdnyakov, S. G., Chizhikov, V. E., Ryazankina, O. I., Shchelkunov, S. N., Koonin E. V. & Chernos, V. I. (1993). *Virus Res.* in the press.

Taylor, G., Vimr, E., Garman, E. & Laver, G. (1992). Purification, crystallization and preliminary crystallographic study of neuraminidase form *Vibrio cholerae* and *Salmonella typhimurium* LT2. *J. Mol. Biol.* **226**, 1287–1290.

Uemura, H., Schenkman, S., Nussenzweig, V. & Eichinger, D. (1992). Only some members of a gene family in *Trypanosoma cruzi* encode proteins that express both trans-sialidase and neuraminidase activities. *EMBO J.* **11**, 3837–3844.

Varghese, J. N., Laver, W. G., & Colman, P. M. (1983). Structure of the influenza virus glycoprotein antigen neuraminidase at 2·9 Å resolution. *Nature (London)*, **303**, 35–40.

Vellieux, F. M. D., Huitema, F., Groendijk, H., Kalk, H., Frank J., Jongejan, J. A., Duine, J. A., Petratos, K., Drenth, J. & Hol, W. G. J. (1989). Structure of quinoprotein methylamine dehydrogenase at 2·25 Å resolution. *EMBO J.* **8**, 2171–2178.

Warner, T. G., Harris, R., McDowell, R. & Vimr, E. R. (1992). Photolabeling of LT2 sialidase: identification of a peptide with a predicted structural similarity to the active site of influenza-virus sialidase. *Biochem. J.* **285**, 957–964.

Xia, Z., Dai, W., Xiong, J., Hao, Z., Davidson, V. L., White, S. & Mathews, F. S. (1992). The three-dimensional structures of methanol dehydrogenase from two methylotropic bacteria at 2·6 Å resolution. *J. Biol. Chem.* **267**, 22289–22297.

Xue, F. & Cooley, L. (1993). *kelch* encodes a component of intercellular bridges in drosophila egg chambers. *Cell*, **72**, 681–693.

Edited by F. Cohen

(Received 29 September 1993; accepted 29 November 1993)