Evolutionarily Mobile Modules in Proteins

Many proteins consist of a fairly small set of modular elements. How these units spread and multiplied during evolution is not altogether clear, but a pattern may be emerging

by Russell F. Doolittle and Peer Bork

Molecular biologists and biochemists have learned in recent decades that many proteins consist of domains, or discrete blocks of amino acids. Many of these domains have well-defined functions that contribute to the overall activity of a protein. Furthermore, some of these modular units frequently move about within and between proteins during evolution. The evolutionary mobility of these modules is not restricted to hops within the genetic material of a single species: in some cases, the modules are apparently able to travel laterally across species lines—even moving from animal cells to bacteria, for example.

Because a compartmental delineation has also been found in the regions of genes that encode proteins, many biologists have been convinced that these structural features are reflections of the same underlying phenomenon. They believe that each genetic coding region corresponds to a specific structural feature in a protein. We and our colleagues take a somewhat different point of view. The weight of evidence, in our opinion, falls to the argument that the subdivision of genes into separate coding parts is a far more recent development by evolutionary standards.

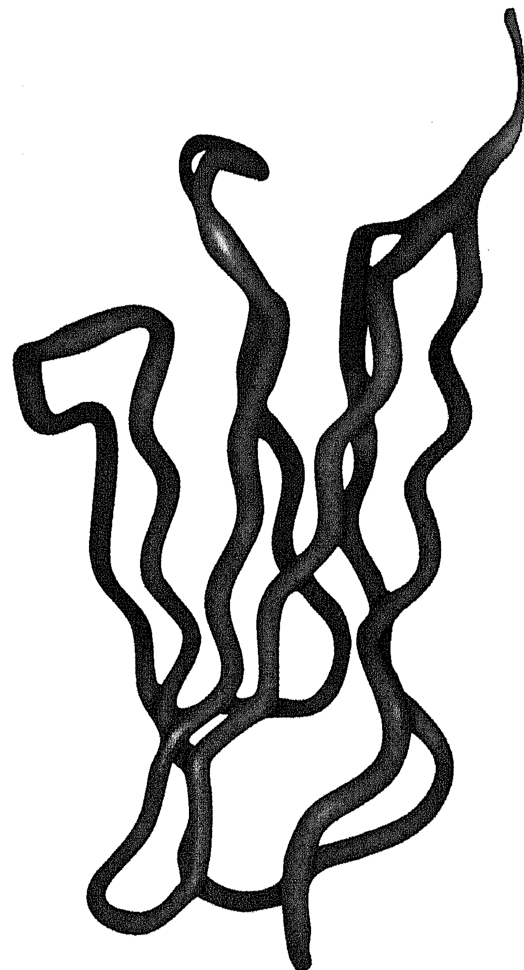
Proteins are long chains of small molecules called amino acids. Twenty different kinds of amino acids, each with its own shape and chemical character, make up all the proteins found in nature. All the properties of a protein depend on which of the 20 amino acids are used in its construction and, particularly, on the order in which they are strung together. Most notably, the amino acid sequence determines how the protein will fold up into an active, three-dimensional body.

The length of a sequence plays an important part in that determination. Chains of amino acids—usually called polypeptides—may be up to many thousands of amino acid units long. (The record to date is for titin, a muscle protein that contains upward of 30,000 amino acid residues.) Short chains, however, are not large enough to have sufficient intramolecular attachments to lock themselves into a single conformation; they tend to flop from one form to another. Ordinarily, when a polypeptide chain contains 30 or 40 residues, it begins to have enough internally cohesive forces to give it a predominant shape, although it may still need additional stabilization from bound metal ions or disulfide bonds between pairs of the amino acid cysteine.

In a constant environment, any protein containing more than some minimum number of amino acid units will always fold itself in the same way. That environment may be the dilute salt solution that constitutes many biological fluids or the greasy confines of a biological membrane; it can also include nearby proteins or even other parts of the same long polypeptide chain.

A sequence that folds spontaneously into a characteristic shape under a defined set of circumstances is called a domain, but that formal definition is seldom applied with rigor. More often than not, the term is applied to any part of a protein that can be defined as structurally distinct from the rest. Some small proteins are completely embodied in a single domain; many others are composed of two or more domains; and some are made up of many domains, the shapes of which may be very similar or very different.

RUSSELL F. DOOLITTLE and PEER BORK analyze the structures of proteins to gain insights into evolution. Doolittle is professor of biology at the Center for Molecular Genetics of the University of California, San Diego. He began his studies of protein evolution while he was a graduate student in biochemistry at Harvard University in the late 1950s. Doolittle is a member of the National Academy of Sciences and other scientific associations. This article is his third for *Scientific American*. Bork is a project leader at the Max Delbrück Center for Molecular Medicine in Berlin and a visiting scientist at the European Molecular Biology Laboratory in Heidelberg. He also holds a teaching appointment at Humboldt University in Berlin.



CD2 DOMAIN 2

The most straightforward way to identify a domain is to determine its structure through x-ray diffraction of protein crystals or nuclear magnetic resonance (NMR) studies. Once the amino acid sequence of a domain has been identified, researchers can find other related domains without recourse to structural studies: they can simply look for amino acid sequences that are similar to those of familiar structures. That shortcut is extremely helpful because many more amino acid sequences are known than are structures from x-ray diffraction or NMR. Even in the absence of experimental determinations, it is often possible to infer the existence of a domain on the basis of sequence alone. By looking at the structural and sequence similarities of proteins in this way, one can learn much about their evolution.

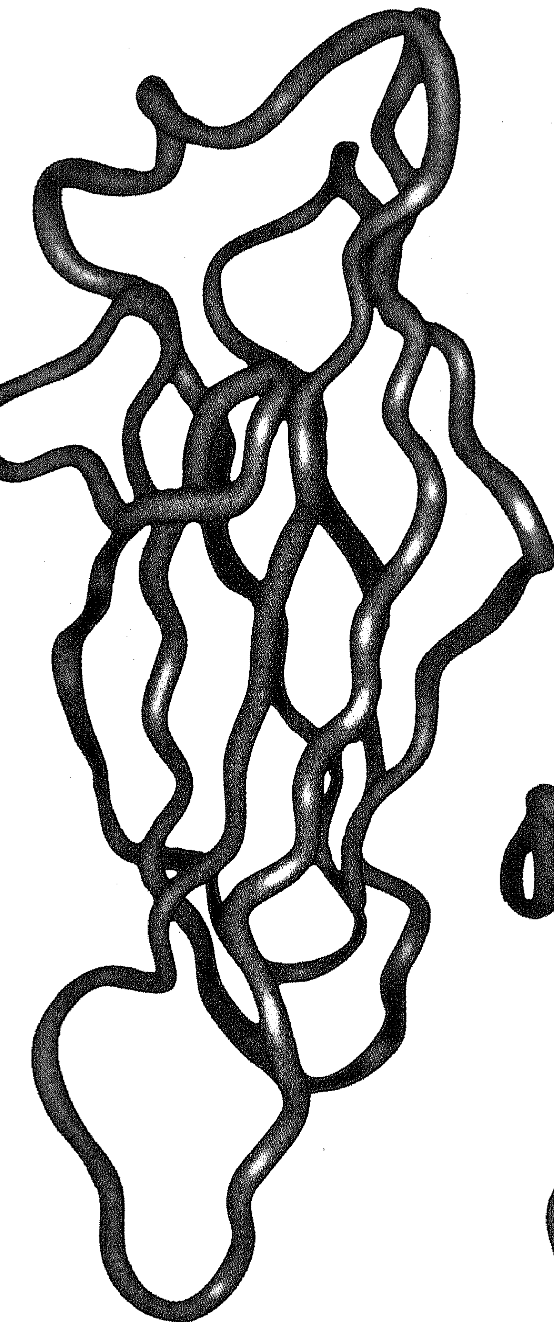
Until the early 1970s, the conventional wisdom about how proteins evolved centered mostly on "duplication and modification." The gene for a given protein is occasionally duplicated by various recombination processes during which genetic information is exchanged between strands of DNA. Sometimes the duplication results in a second gene that can then undergo further modification or mutation to produce a new protein with a novel function. Alternatively, the duplicated DNA can be in tandem, in which case the original protein is elongated and may assume novel properties as a result. As comparisons of amino acid sequences make obvious, these kinds of internal duplications have clearly given rise to many extended proteins, ranging from small ones, such as the bacterial ferredoxins that have only 56 amino acid residues, to large ones, such as bacterial beta-galactosidase, which has more than 1,000.

A hidden aspect of protein evolution came to light about 20 years ago, however, when Michael G. Rossmann of Purdue University determined the three-dimensional structure of the enzyme lactate dehydrogenase by x-ray diffraction. One part of that molecule, he noticed, closely resembled features of some other proteins he had seen. Specifically, a part of the enzyme that bound to a cofactor had obvious counterparts in other dehydrogenases.

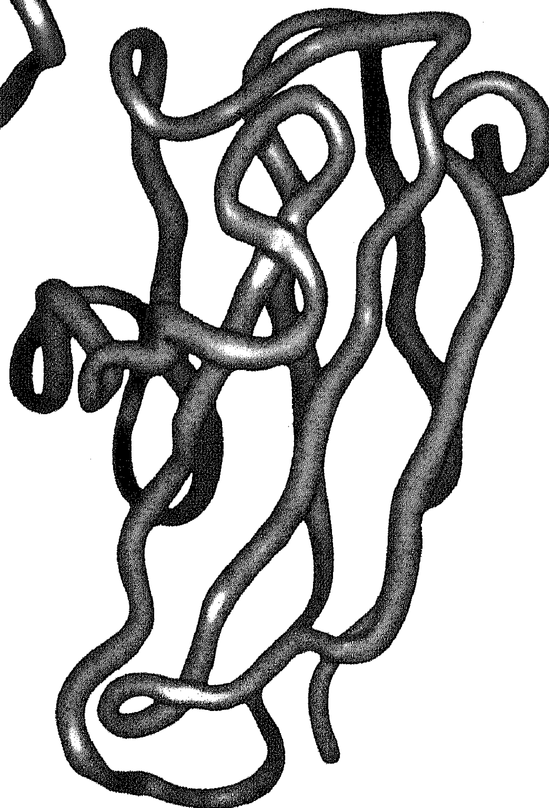
What made those structural similarities noteworthy was that they did not always occur in the same relative parts of the molecules. It seemed as though during evolution the unit had moved around within the linear amino acid sequence without losing its function of binding to a cofactor. Rossmann suggested that proteins were constructed of modules—what we would now consider domains—that had appeared early in life's history and had been assembled in different combinations.

His observation presented a possibility for protein evolution that greatly aug-

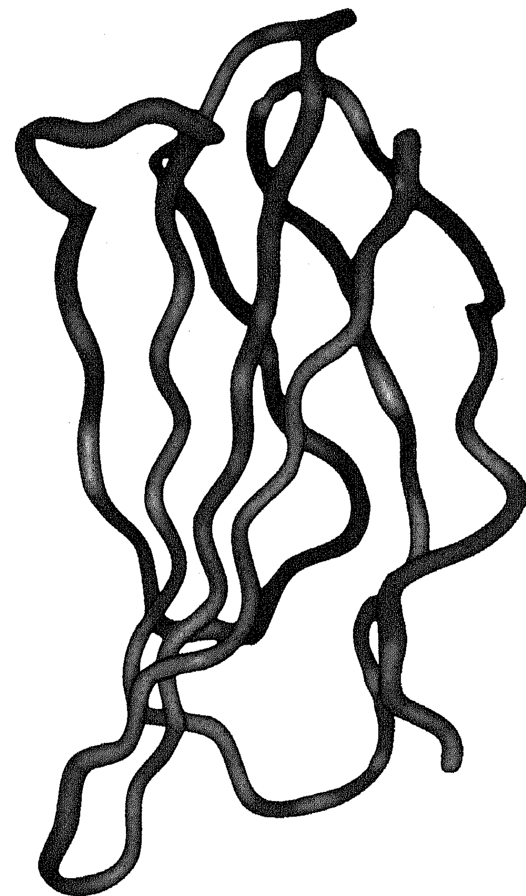
GENETICALLY MOBILE MODULES have been found in many proteins. Two types of these modules, or domains, are shown here. The Fn3 and the GHR domains are examples of fibronectin type III modules. The PapD and CD2 domains are immunoglobulin domains. These modules are linear sequences of amino acids that can fold themselves into consistent, recognizable structures with specific biochemical properties. During evolution, these domains can move as discrete units from one protein to another, which helps new types of proteins to appear.



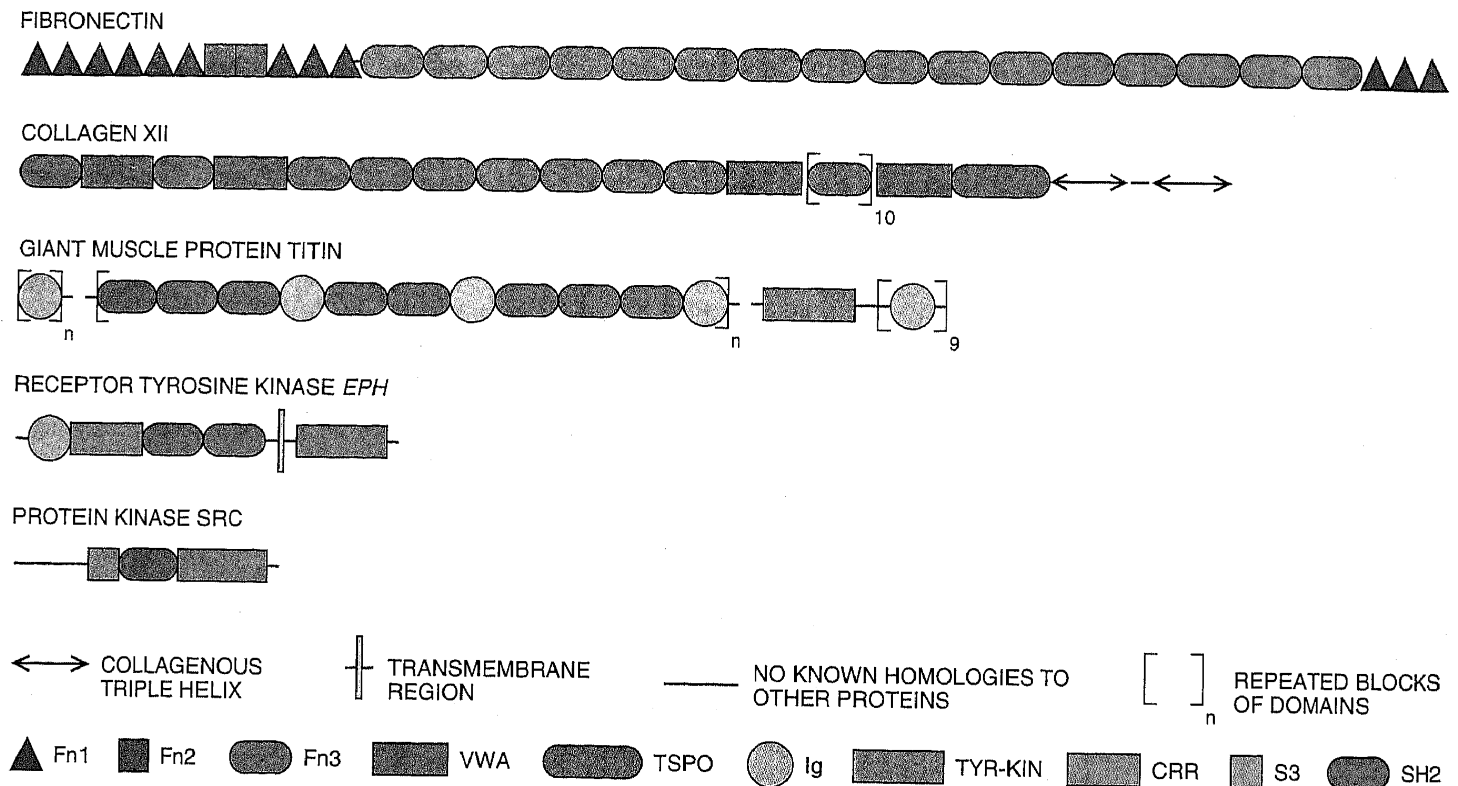
PapD DOMAIN 1



GHR DOMAIN 1



Fn3 (10TH DOMAIN)



LIKE BEADS IN A NECKLACE, domains appear as distinct subunits of modular proteins. Some proteins, such as fibronectin, collagen XII and the muscle protein titin, contain many repeats of the same few domains.

mented what could be accomplished purely through duplication and modification. If new proteins could be created by recombining the components of other ones, then protein diversity could grow explosively.

Rossmann's suggestion proved prescient. The amino acid sequences for numerous large proteins have been determined, and many have the kind of grossly repetitive structure that one might expect of a string of mobile modules. For instance, the protein fibronectin is made up of two long chains, each of which contains more than 2,000 amino acid residues. Casual inspection revealed that the chains of this large extracellular protein consist of several runs of three different types of repeated sequences. The repeats, which are referred to as Fn1, Fn2 and Fn3, have lengths on the order of 45, 60 and 100 amino acids, respectively. (The repeats are "imperfect," which means that all the repeats of a given type are not exactly identical.) Presumably, each type of repeat can fold up independently as a true domain, and the overall protein must be like a long necklace made from three types of beads.

Surprisingly, sequences similar to Fn1, Fn2 and Fn3 were subsequently observed in a huge assortment of other animal proteins. Much the same is also true for a number of other identified domains. The small protein called epidermal growth factor consists of a single domain (53 amino acids long in the human version) that is tightly folded and pinned together by three disulfide bonds. Similar domains have been identified by their sequences in more than 100 proteins, in which they occur from one to more than 30 copies.

The functions of many of these modules are not altogether clear, but many of them do bind to or recognize particular substances. There is a family of lectins that bind to various carbohydrates. Similarly, the immunoglobulin domain, a feature of antibodies and other molecules in the immune system, is well known for its binding capabilities. Some domains may serve as recognition tags that identify a protein as "belonging" to a certain tissue. Many appear to be mere linkers or spacers, innocuous connecting units. Finally, some may have no function at all. It appears, therefore, that many domains can and do move within and between proteins during evolution. As long as no damage or loss of function results from a shuffle, the evolutionary cost of maintaining a domain in a new setting is insignificant. That result is a corollary of the theory of neutral evolution, but it might be phrased as a rule borrowed from professional basketball: "No harm, no foul."

When Rossmann first suggested that modular units might move within and between proteins, no one pondered very seriously the genetic mechanisms that might account for those rearrangements. Not long afterward, however, molecular biologists happened on an unexpected feature of genes that seemed to offer an explanation. As James Watson and Francis Crick had learned in 1953, genetic information is inscribed in the double helices of deoxyribonucleic acid (DNA) molecules. Just as proteins are strings of amino acids, DNA molecules are strings of nucleotide bases. The DNA nucleotide sequences are copied, or transcribed, as complementary strands of messenger ribonucleic acid (RNA). Inside cells, miniature factories called ribosomes assemble proteins from the instructions in the RNA: each codon, or three-base sequence in the RNA, corresponds to an amino acid.

The surprising observation made in the mid-1970s was that the DNA coding for a polypeptide can be interrupted by noncoding sequences—arrays of bases that do not correspond to the sequence of amino acids found in the ultimate protein product. The noncoding sequences are excised by a splicing mechanism before the messenger RNA strand is translated into a polypeptide.

The discovery of those interruptions in genes prompted Walter Gilbert of Harvard University to suggest that the noncoding sequences, which he named introns (for intervening sequences), facilitated the exchange of the coding parts of the genes, which he termed exons (for expressed sequences). The thought was that the additional distance between coding segments would lead to proportionately more opportunities

for recombinations, which depend on random breaks in the DNA. It was suggested that similarities between the sequences of introns might promote misalignments and unequal crossovers of DNA during recombination, which would make the rearrangement of genes easier. Although there was no basis at that time for supposing that introns would have similar sequences, subsequent research has shown that introns are a haven for large numbers of mobile genetic elements. The similar sequences in those mobile elements can contribute to genetic delinquencies during meiosis, the cell division process that gives rise to eggs and sperm.

Of course, many creatures do not engage in meiosis; have those organisms missed out on a splendid way to assemble new proteins? Introns that interrupt protein coding are found only in the DNA of eukaryotes, organisms with discrete nuclei. The genes for bacterial proteins do not contain introns: every set of three bases corresponds to an amino acid in the protein. (A few types of introns that do not interrupt proteins have been found in bacteria, but they are not properly part of this discussion.)

The absence of introns from bacterial protein genes led Ford Doolittle of Dalhousie University and James Darnell of the Rockefeller University to suggest independently that bacteria may have possessed introns in the past but had lost them. Presumably, their genomes had been streamlined during evolution to make their replication more efficient. In short, the introns had been around since the beginning of life; it was the short coding sequences that were engendered separately.

Doolittle and Darnell's proposal led to a still unsettled debate about whether introns appeared "early" and are fundamental to the origin of all proteins or whether they came "late." The latter proposition was advanced by Thomas Cavalier-Smith, now at the University of British Columbia, who theorized that introns might be invasive bits of nucleic acids, referred to as transposable elements, that originated in the symbiotic organisms that eventually became the mitochondria and other organelles of eukaryotic cells. His idea has been extended by Donal Hickey of Ottawa University, among others.

As it turns out, the stretches of DNA that encode the evolutionarily mobile modules in proteins are frequently, but not always, flanked by introns. In other words, the structural units in many proteins are encoded by exons. That observation fostered a widespread belief that all exons are evolutionarily mobile and correspond to potential modular building blocks in proteins.

In our view, this notion is mistaken on two counts. One is that, as Lázló Patthy of the Institute of Enzymology in Budapest first pointed out, all exons can be shuffled, but only a fraction of such shuffled units will be genetically compatible—that is, many of them cannot be sensibly translated in some new positions. When an intron falls in a coding sequence, it will occupy one of three types of positions: squarely between two codons (type 0), between the first and second positions of a codon (type 1) or between the second and third positions (type 2). If that intron and the coding sequence adjacent to it are shuffled into a new location, the intron must adopt the same type of position—otherwise the shifted codons will be translated improperly and a nonsensical amino acid sequence will result. If introns were distributed randomly, we would expect to find that after a shuffle, only one third of the new exon combinations would be in phase. Curiously, the overwhelming majority of the genes for the most frequently shuffled modules are flanked by introns that are of type 1.

The other fundamental reason why only some exons are evolutionarily mobile is that only true domains—those that can fold completely and independently—will be able to survive in new protein settings. Smaller, less self-sufficient sequences would be unable to fold and would lose their identity. Moreover, if a shuffled unit were to land between two exons that were not themselves true domains, then the product of the gene receiving the addition might not be able to fold itself properly either.

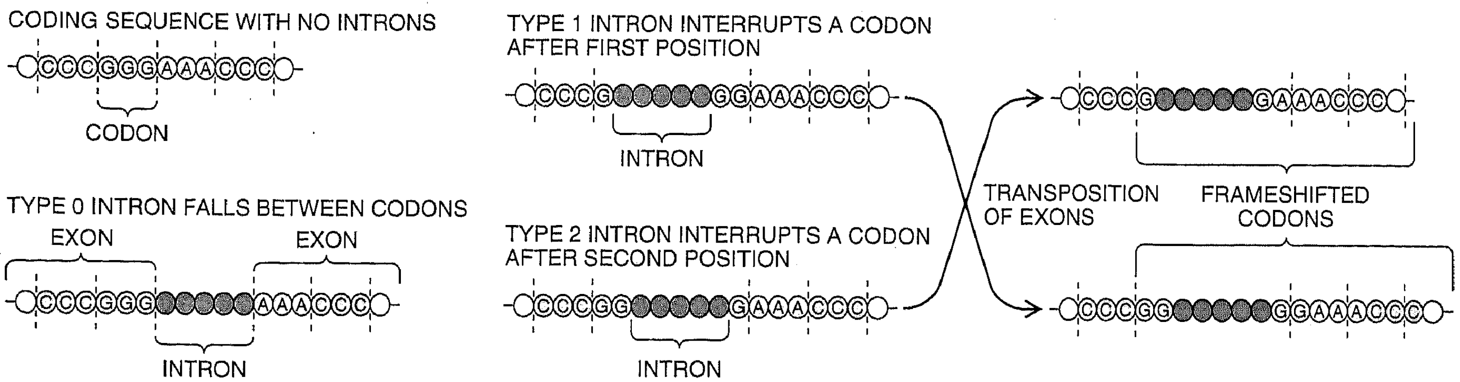
These two factors, one genetic and the other structural, contribute greatly to why mobile domains are so often found in one another's company. Not infrequently, proteins that contain one mobile domain contain others; some proteins are mosaics that contain as many as five different commonly shuffled modules. These types of proteins are both genetically and structurally tolerant of the shuffling process.

The observation that many modules are encoded by exons has been interpreted as support for the idea that the primeval organisms assembled all their proteins from an inventory of exon-encoded primitive structural components. Several points argue against such an interpretation, however. For one, simple arithmetic proves that the hypothetical early exons would have been too small to produce protein components that

Sizes of Some Mobile Domains

Evolutionarily mobile domains vary in size. Some hold their shape in part because of disulfide bonds between pairs of their cysteine residues. Other domains are stable without such bonds.

DOMAINS CONTAINING DISULFIDE BONDS	APPROXIMATE NUMBER OF AMINO ACID UNITS	NUMBER OF CYSTEINES
Somatomedin B	40	8
Complement C9	40	6
EGF	45	6
Fn1	45	4
Fn2	60	4
"Apple" (Sushi) (ßGP26)	61	4
Ovomucoid	68	6
VWF-C	72	8
Kringle	80	6
Kunitz	80	6
Link	~100	4
Scavenger receptor	110	6
Fibrinogen-related domain	250	4 or 6
NO DISULFIDE BONDS		
Collagen	18	
Leucine-rich	25	
"Gla"	35	
Collagen-binding	50	
Lectinlike	100	
Fn3	~100	
SH2	100	
SH3	100	



INTERRUPTED STRUCTURE of some genes that encode proteins divides them into exons and introns (expressed and intervening sequences). Triplets of bases in exons are translated as amino acids in proteins. Introns can interrupt coding sequences in three different positions. If exons and introns moved randomly, transpositions between different types of introns would cause frameshift mutations. That problem counts against the idea that all exons encode movable modules.

could fold on their own. The average size of the known exons in vertebrate genomes today is 135 nucleotide bases, which corresponds to a polypeptide of a mere 45 amino acids. A sequence that short usually needs auxiliary stabilization to fold into a stable conformation.

Keep in mind, too, that the proponents of the "intron early" theory contend that introns are constantly being lost over time. They have been forced to that conclusion by the sporadic occurrence of introns in different species. This inconstant pattern could be the result of either gain or loss of introns, but if one is wedded to the notion that introns were there from the beginning, then the only explanation is loss. Accordingly, the earliest exons would have been even smaller, encoding polypeptides that could not reasonably be expected to fold into domains on their own.

Another argument against the use of modern mosaic proteins as examples of early building blocks concerns the distribution of domains among proteins. Far and away, the majority of mobile modules known to date are found exclusively in animal proteins. At this point, we do not really know when or where most of them arose initially. Perhaps the trail of their evolution has been partly obscured by extensive sequence changes in the related domains of plants, fungi and protozoa. As we shall discuss, the fact that three-dimensional structures are more persistent than sequences in an evolutionary sense may allow this puzzle to be solved.

Besides all the evidence that most exons are not evolutionarily mobile, some mobile domains are also clearly not the products of single exons. One large domain, first observed in the fibrinogen molecule of vertebrate animals, is composed of 250 amino acids. In some proteins, the gene for that domain contains multiple introns. Yet none of these individual exons has ever been found without all the others. None of the exons, it seems, has ever shuffled itself out of the domain. Thus, the mere presence of introns in a gene is apparently not enough to cause exons to be mobile. That the vast majority of identified exons has never been found in more than one setting argues against simple and indiscriminate mobility.

There are other examples of movable units that contain introns within their coding sequences. One of the first movable modules to be identified is referred to as a kringle (because it resembles a type of Danish pastry by that name). It is made of about 80 amino acid units and contains three characteristic disulfide bonds. It is quite similar to the Fn2 domain, differing only in the number of residues between its cysteine residues, and some workers do not distinguish between the two. In some of its settings, the gene for the kringle is split by an intron, but so far no one has found half a kringle in any protein.

A further point in favor of the "intron late" theory is that the introns that interrupt coding regions are much more common in plants and animals than among the earliest diverging eukaryotes. No introns at all have been found in primitive eukaryotes, such as *Giardia lamblia*. Furthermore, modular proteins have been identified in plants for which no recognizable counterpart is known in animals, and vice versa. Finally, there is indirect evidence that the modular assembly of some bacterial proteins occurred so recently that they must have evolved without the aid of introns. All this evidence suggests that protein-interrupting introns appeared after the evolution of eukaryotes.

So some exons encode domains, but most exons do not; those that do can often be genetically duplicated and shuffled. The issue of cause and effect between these phenomena raises a thorny problem. Perhaps the evolution of these introns did facilitate exon shuffling. On the other hand, it is not impossible that these introns often delineate the coding regions for domains because that placement is advantageous for intron propagation. If an intron interrupted a sequence coding for a domain, it might survive in that situation (if it did not violate the phase rule noted earlier). It would not spread further, however, because the exons on its boundaries could not stand alone and thus could not be moved independently. Conversely, if the intron lands between regions coding for independently folding units, it can spread to other locations along with the exons. So exon shuffling may be only incidental to the survival of introns.

One way to learn more about the evolution of movable modules is to look at the structure and dispersal of one in particular. Our favorite is Fn3, the fibronectin type III domain. Like kringles, Fn3 units are sometimes split by single introns, but they have never been observed with fewer than their full complement of 90 to 100 amino acids. The two of us independently followed the discovery of Fn3 in various proteins for several years. In early reports the sequence was turning up only in animal proteins, so we were both surprised when, in 1990, workers at Niigata University in Japan reported the existence of an Fn3 domain in a bacterial protein. Our common interest came to light at a meeting in Italy in 1991, whereupon we decided to join forces in a comprehensive inventory of Fn3 occurrences.

To that end, we screened a data base of protein sequences by various means, including a pattern-searching algorithm that one of us (Bork) had devised with Christian Grunwald when both were at the Central Institute for Molecular Biology in East Berlin. We uncovered well over 300 unique occurrences of the Fn3 se-

quence motif, which established that it was indeed a true domain, should there have been any doubt. The 300 occurrences actually represented 67 different proteins, not counting the same proteins from different species. Sixty of these were from animals and seven from bacteria. None of the sequences identified were from plants, fungi or unicellular eukaryotes.

The obvious questions to be asked were: Did bacteria and animals both inherit the domain from some mutual ancestor, or did one of the groups somehow acquire it from the other? If the domain was present in the common ancestor of prokaryotes and eukaryotes, then why has it not been found in fungi and plants? With the aid of a computer, we aligned all the Fn3 sequences that we could find and constructed a crude phylogenetic tree based on their similarities. Because trying to accomplish that with all 300 sequences was cumbersome, we began by using representative sets of sequences from all the bacterial proteins and from only the most different of the animal ones.

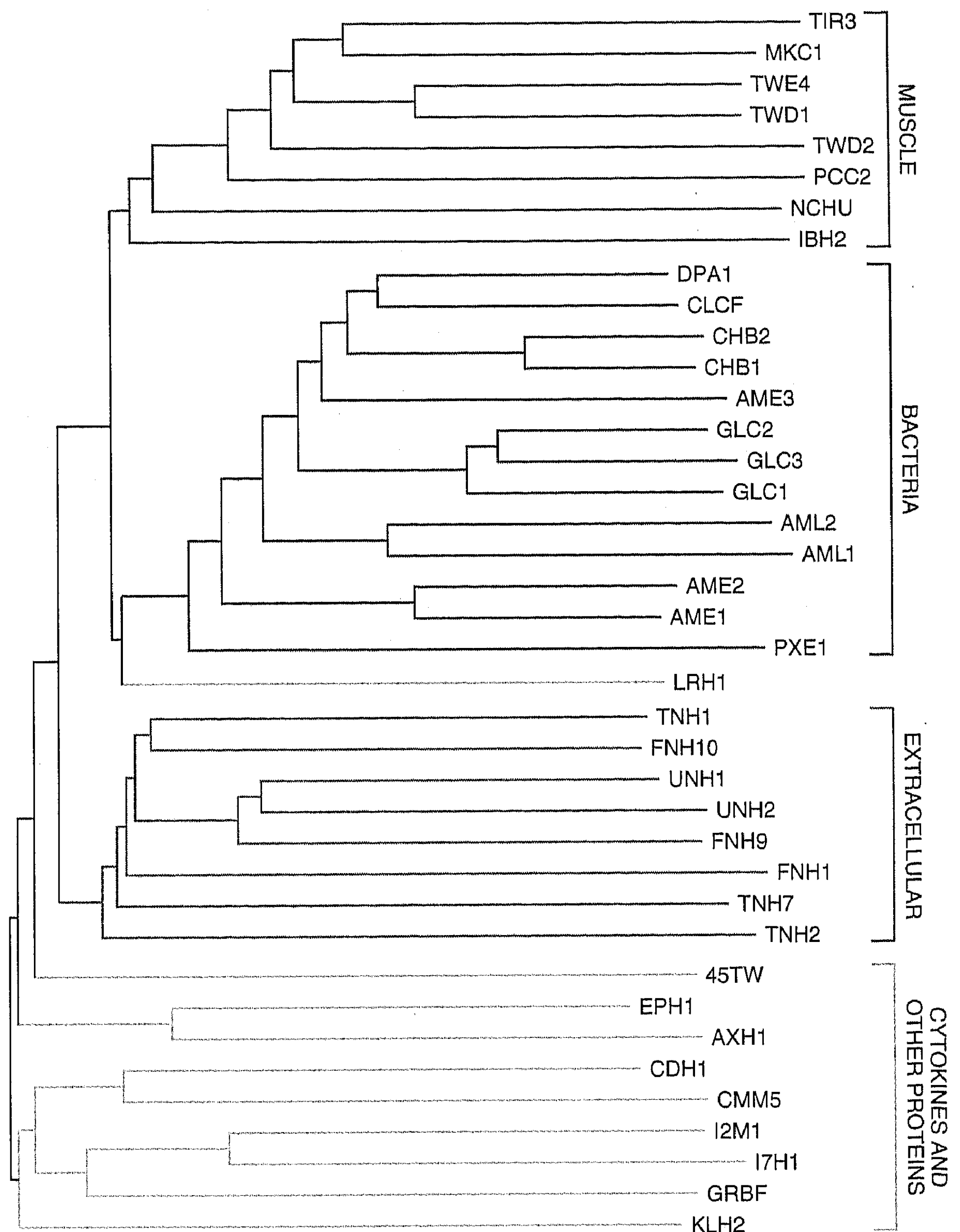
It was soon clear that something was amiss. The bacterial sequences were simply too similar to the animal ones to have descended from a common ancestor two billion years ago. Instead the evidence—including the computer-generated phylogenetic trees—favored the view that somehow the bacteria had acquired the Fn3 domain from an animal source.

There were several reasons to think so. First, it was often the case that an enzyme from one bacterial species carried the domain but that the same enzyme from other species did not, which implied that the domain was structurally and functionally expendable. It must therefore have been a late addition to specific bacteria. Moreover, the Fn3 sequences appeared sporadically and in clusters but always in a characteristic set of extracellular enzymes; if bacteria were losing copies of Fn3 over time, one might expect to find surviving copies in a more varied set of proteins.

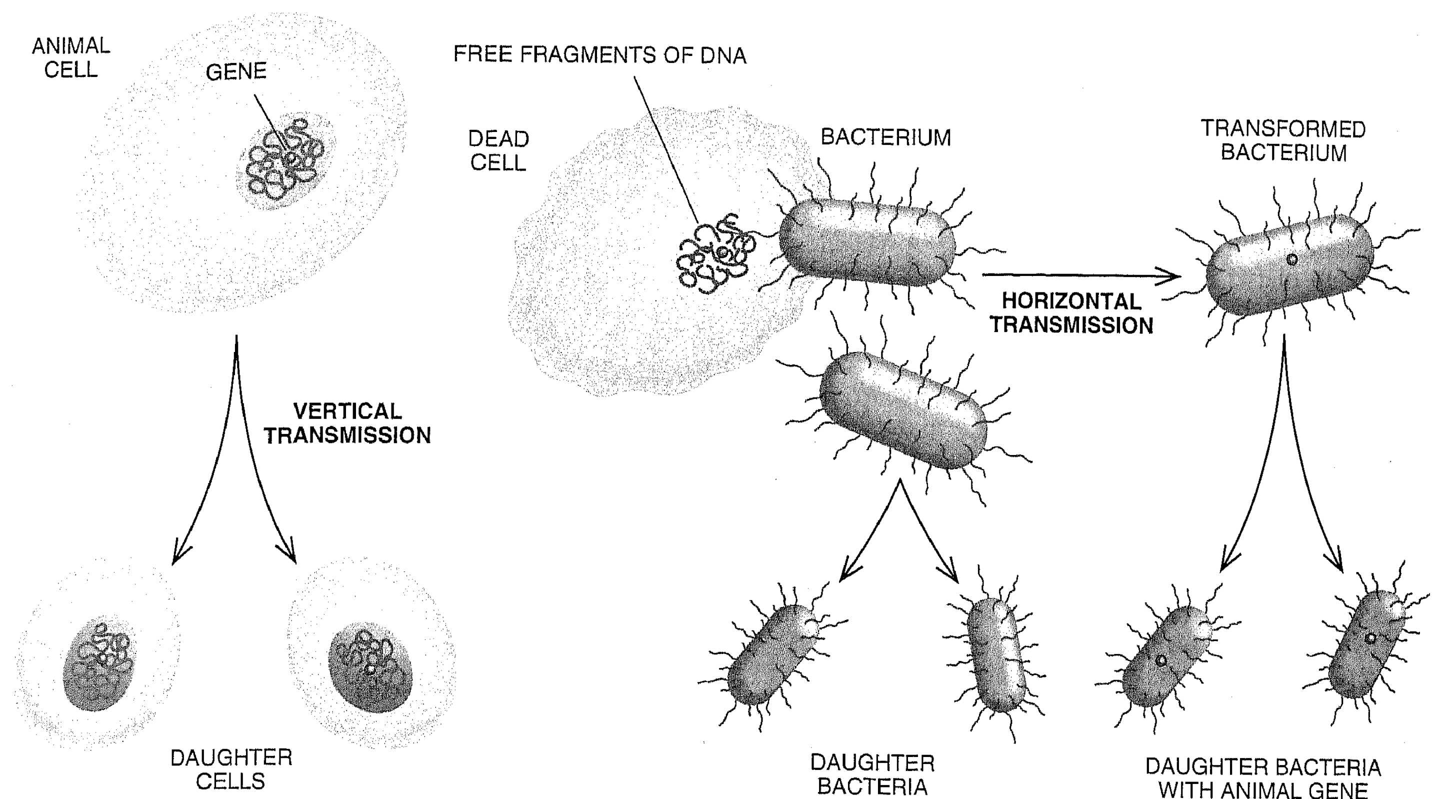
Lastly, although the bacteria that had Fn3 domains were of diverse types, they did have certain features in common. All are soil bacteria that obtain their food from commonly found polymers, such as cellulose and chitin, released during decomposition. Vast numbers of other types of bacteria have been examined, but none of them have the Fn3 domain in any of their proteins. More than half the genome sequence of the intestinal bacterium *Escherichia coli* is known, and not one hint of an Fn3 sequence has surfaced. The same can be said for the large number of studied fungal and plant sequences. If the Fn3 domain had been present in a common ancestor of prokaryotes and eukaryotes, we would have expected its radiation to have continued along the major lines of descent and to be represented in all these groups.

The idea that the gene for a domain can migrate between distantly related organisms may seem outlandish at first. Common experience shows that genes are transmitted vertically, from one generation of organisms to the next. It is nonetheless sometimes possible for genes to be transmitted horizontally as well, not only between species but across distant lineages. Some viruses can pick up small genes from one host and carry them to another host; in rare instances, the transposed gene may assimilate itself into the new host's DNA. Bacteria can be transformed by picking up DNA from their surroundings, which may include decaying animal cells. Many bacteria also carry genes on small rings of DNA, or plasmids, that they can exchange with other bacteria. In theory, all these mechanisms represent opportunities for the horizontal transmission of genes.

Allowing that some bacterium did obtain the gene for an Fn3 domain from an animal source, when did it happen? All that the phylogenetic tree can indicate is that it happened within the past billion years, after the divergence of animals from plants and fungi. To pin down the date, we need to know the average rates of sequence change along both the bacterial and animal lines of



EVOLUTIONARY RELATIONS of 39 Fn3 domains from various sources appear in this phylogenetic tree. Using a computer, the authors created the tree by comparing the sequences of the domains from animals and bacteria. Surprisingly, the bacterial domains are closely related to some from animals, which suggests that bacteria obtained the genes for those domains from animals.



HORIZONTAL TRANSMISSION OF GENES could explain how bacteria obtained domains from animals. Genes pass vertically between generations of cells. Bacteria can also be transformed by the absorption of DNA from their environment. If a bacterium picked up the DNA for a domain from a dead cell, it might have transmitted the gene to its progeny.

divergence. For the animal proteins, we can estimate the rate by comparing sequences from various creatures whose divergence times are shown in the fossil record. Unfortunately, we do not have comparable information about the bacterial sequences. (Whereas some microfossils corresponding to bacteria have been reported, there is certainly not yet an interpretable evolutionary hierarchy for them as there is for animals.) We can see, however, in both the animal and bacterial proteins a tendency for tandem duplication of Fn3 sequences—that is, within any one protein containing more than one Fn3 domain, the sequences of those domains are often adjacent and usually very similar. That observation implies that the duplication of the DNA for the Fn3 domain must be relatively recent.

The timing of the horizontal transmission and of the genetic duplications is critical to our understanding of how these genetic units spread. So far as is known, contemporary bacteria do not have introns in their protein-coding gene sequences. If they ever did have introns in their coding sequences, how long has it been since those introns disappeared? Unless it was fairly recently, the genes for Fn3 must have been spreading without the aid of introns.

One likely possibility is that the Fn3 domain is being spread among soil bacteria by a promiscuous bacteriophage (bacterial virus) or plasmid. We hope eventually to catch the act of transmission in progress by finding a phage that carries a gene for an Fn3 domain. Now that a number of bacterial Fn3 sequences have been found, it may be possible for us to synthesize short tags of DNA that will bind to the Fn3 genetic units. Used in conjunction with the DNA amplification technique called the polymerase chain reaction, those tags could help identify the genes in bacteriophages or other vectors.

Of course, where the Fn3 domain originated is still a mystery. Did it first appear in animals? Or are we just unable to identify its ancestral forms by sequence comparison? Three-dimensional determinations show that the Fn3 structure is suspiciously similar to immunoglobulin domains. Three-dimensional analyses by x-ray crystallography and NMR have made it possible to trace the immunoglobulin domain back to proteins in prokaryotes, including PapD, a "chaperone" that helps other proteins fold, and also to a bacterial enzyme that digests cellulose. It is interesting that the immunoglobulin domain, as originally defined, contains a disulfide bond that pins its two sides together, but the more primitive forms, some of which persist in vertebrate animals today, lack that bond. It is these primitive forms that are most similar to the Fn3 domain.

We can anticipate that other instances of pirated modules will be uncovered. By our census, the Fn3 domain occurs in about one out of every 50 animal proteins (50 out of 2,500 known animal sequences, independent of species redundancies). We estimate that about 25 modules are very commonly strewn throughout animal proteins, much as Fn3 is. More than 100 others are found in more than one setting but less frequently than the first group. Tracing the pedigree and diaspora of these modular units is a major challenge that should reveal much about every aspect of the evolution of all living things.

FURTHER READING

THE GENEALOGY OF SOME RECENTLY EVOLVED VERTEBRATE PROTEINS. Russell F. Doolittle in *Trends in Biochemical Sciences,* Vol. 10, No. 6, pages 233–237; June 1985.

PROTEIN MODULES. Martin Baron, David G. Norman and Iain D. Campbell in *Trends in Biochemical Sciences,* Vol. 16, No. 1, pages 13–17; January 1991.

MODULAR EXCHANGE PRINCIPLE IN PROTEINS. Lázló Patthy in *Current Opinion in Structural Biology,* Vol. 1, No. 3, pages 351–361; June 1991.

MOBILE MODULES AND MOTIFS. Peer Bork in *Current Opinion in Structural Biology,* Vol. 2, No. 3, pages 413–421; June 1992.

PROPOSED ACQUISITION OF AN ANIMAL PROTEIN DOMAIN BY BACTERIA. Peer Bork and Russell F. Doolittle in *Proceedings of the National Academy of Sciences,* Vol. 89, No. 19, pages 8990–8994; October 1, 1992.