= Circular permutation in proteins =

A circular permutation is a relationship between proteins whereby the proteins have a changed order of amino acids in their peptide sequence . The result is a protein structure with different connectivity , but overall similar three @-@ dimensional (3D) shape . In 1979 , the first pair of circularly permuted proteins ? concanavalin A and lectin ? were discovered ; over 2000 such proteins are now known .

Circular permutation can occur as the result of evolutionary events , posttranslational modifications , or artificially engineered mutations . The two main models proposed to explain the evolution of circularly permuted proteins are permutation by duplication and fission and fusion . Permutation by duplication occurs when a gene undergoes duplication to form a tandem repeat , before redundant sections of the protein are removed ; this relationship is found between saposin and swaposin . Fission and fusion occurs when partial proteins fuse to form a single polypeptide , such as in nicotinamide nucleotide transhydrogenases .

Circular permutations are routinely engineered in the laboratory to improve their catalytic activity or thermostability , or to investigate properties of the original protein .

Traditional algorithms for sequence alignment and structure alignment are not able to detect circular permutations between proteins . New non @-@ linear approaches have been developed that overcome this and are able to detect topology @-@ independent similarities .

= = History = =

In 1979 , Bruce Cunningham and his colleagues discovered the first instance of a circularly permuted protein in nature . After determining the peptide sequence of the lectin protein favin , they noticed its similarity to a known protein ? concanavalin A ? except that the ends were circularly permuted . Later work confirmed the circular permutation between the pair and showed that concanavalin A is permuted post @-@ translationally through cleavage and an unusual protein ligation .

After the discovery of a natural circularly permuted protein , researchers looked for a way to emulate this process . In 1983 , David Goldenberg and Thomas Creighton were able to create a circularly permuted version of a protein by chemically ligating the termini to create a cyclic protein , then introducing new termini elsewhere using trypsin . In 1989 , Karolin Luger and her colleagues introduced a genetic method for making circular permutations by carefully fragmenting and ligating DNA . This method allowed for permutations to be introduced at arbitrary sites .

Despite the early discovery of post @-@ translational circular permutations and the suggestion of a possible genetic mechanism for evolving circular permutants , it was not until 1995 that the first circularly permuted pair of genes were discovered . Saposins are a class of proteins involved in sphingolipid catabolism and antigen presentation of lipids in humans . Chris Ponting and Robert Russell identified a circularly permuted version of a saposin inserted into plant aspartic proteinase , which they nicknamed swaposin . Saposin and swaposin were the first known case of two natural genes related by a circular permutation .

Hundreds of examples of protein pairs related by a circular permutation were subsequently discovered in nature or produced in the laboratory . As of February 2012 , the Circular Permutation Database contains 2 @,@ 238 circularly permuted protein pairs with known structures , and many more are known without structures . The CyBase database collects proteins that are cyclic , some of which are permuted variants of cyclic wild @-@ type proteins . SISYPHUS is a database that contains a collection of hand @-@ curated manual alignments of proteins with non @-@ trivial relationships , several of which have circular permutations .

= = Evolution = =

There are two main models that are currently being used to explain the evolution of circularly permuted proteins : permutation by duplication and fission and fusion . The two models have

compelling examples supporting them , but the relative contribution of each model in evolution is still under debate . Other , less common , mechanisms have been proposed , such as " cut and paste " or " exon shuffling " .

= = = Permutation by duplication = = =

The earliest model proposed for the evolution of circular permutations is the permutation by duplication mechanism . In this model , a precursor gene first undergoes a duplication and fusion to form a large tandem repeat . Next , start and stop codons are introduced at corresponding locations in the duplicated gene , removing redundant sections of the protein .

One surprising prediction of the permutation by duplication mechanism is that intermediate permutations can occur . For instance , the duplicated version of the protein should still be functional , since otherwise evolution would quickly select against such proteins . Likewise , partially duplicated intermediates where only one terminus was truncated should be functional . Such intermediates have been extensively documented in protein families such as DNA methyltransferases .

= = = = Saposin and swaposin = = = =

An example for permutation by duplication is the relationship between saposin and swaposin . Saposins are highly conserved glycoproteins , approximately 80 amino acid residues long and forming a four alpha helical structure . They have a nearly identical placement of cysteine residues and glycosylation sites . The cDNA sequence that codes for saposin is called prosaposin . It is a precursor for four cleavage products , the saposins A , B , C , and D. The four saposin domains most likely arose from two tandem duplications of an ancestral gene . This repeat suggests a mechanism for the evolution of the relationship with the plant @-@ specific insert (PSI) . The PSI is a domain exclusively found in plants , consisting of approximately 100 residues and found in plant aspartic proteases . It belongs to the saposin @-@ like protein family (SAPLIP) and has the N- and C- termini " swapped " , such that the order of helices is 3 @-@ 4 @-@ 1 @-@ 2 compared with saposin , thus leading to the name " swaposin " .

= = = Fission and fusion = = =

Another model for the evolution of circular permutations is the fission and fusion model . The process starts with two partial proteins . These may represent two independent polypeptides (such as two parts of a heterodimer) , or may have originally been halves of a single protein that underwent a fission event to become two polypeptides .

The two proteins can later fuse together to form a single polypeptide . Regardless of which protein comes first , this fusion protein may show similar function . Thus , if a fusion between two proteins occurs twice in evolution (either between paralogues within the same species or between orthologues in different species) but in a different order , the resulting fusion proteins will be related by a circular permutation .

Evidence for a particular protein having evolved by a fission and fusion mechanism can be provided by observing the halves of the permutation as independent polypeptides in related species , or by demonstrating experimentally that the two halves can function as separate polypeptides .

= = = = Transhydrogenases = = = =

An example for the fission and fusion mechanism can be found in nicotinamide nucleotide transhydrogenases . These are membrane @-@ bound enzymes that catalyze the transfer of a hydride ion between NAD (H) and NADP (H) in a reaction that is coupled to transmembrane proton translocation . They consist of three major functional units (I , II , and III) that can be found in different arrangement in bacteria , protozoa , and higher eukaryotes . Phylogenetic analysis suggests that the three groups of domain arrangements were acquired and fused independently .

= = = Other processes that can lead to circular permutations = = =

= = = = Post @-@ translational modification = = = =

The two evolutionary models mentioned above describe ways in which genes may be circularly permuted , resulting in a circularly permuted mRNA after transcription . Proteins can also be circularly permuted via post @-@ translational modification , without permuting the underlying gene . Circular permutations can happen spontaneously through autocatalysis , as in the case of concanavalin A. Alternately , permutation may require restriction enzymes and ligases .

= = The role of circular permutations in protein engineering = =

Many proteins have their termini located close together in 3D space . Because of this , it is often possible to design circular permutations of proteins . Today , circular permutations are generated routinely in the lab using standard genetics techniques . Although some permutation sites prevent the protein from folding correctly , many permutants have been created with nearly identical structure and function to the original protein .

The motivation for creating a circular permutant of a protein can vary . Scientists may want to improve some property of the protein , such as :

Reduce proteolytic susceptibility . The rate at which proteins are broken down can have a large impact on their activity in cells . Since termini are often accessible to proteases , designing a circularly permuted protein with less @-@ accessible termini can increase the lifespan of that protein in the cell .

Improve catalytic activity . Circularly permuting a protein can sometimes increase the rate at which it catalyzes a chemical reaction , leading to more efficient proteins .

Alter substrate or ligand binding . Circularly permuting a protein can result in the loss of substrate binding , but can occasionally lead to novel ligand binding activity or altered substrate specificity .

Improve thermostability . Making proteins active over a wider range of temperatures and conditions can improve their utility .

Alternately , scientists may be interested in properties of the original protein , such as :

Fold order . Determining the order in which different parts of a protein fold is challenging due to the extremely fast time scales involved . Circularly permuted versions of proteins will often fold in a different order , providing information about the folding of the original protein .

Essential structural elements . Artificial circularly permuted proteins can allow parts of a protein to be selectively deleted . This gives insight into which structural elements are essential or not .

Modify quaternary structure . Circularly permuted proteins have been shown to take on different quaternary structure than wild @-@ type proteins .

Find insertion sites for other proteins . Inserting one protein as a domain into another protein can be useful . For instance , inserting calmodulin into green fluorescent protein (GFP) allowed researchers to measure the activity of calmodulin via the fluorescence of the split @-@ GFP . Regions of GFP that tolerate the introduction of circular permutation are more likely to accept the addition of another protein while retaining the function of both proteins .

Design of novel biocatalysts and biosensors . Introducing circular permutations can be used to design proteins to catalyze specific chemical reactions , or to detect the presence of certain molecules using proteins . For instance , the GFP @-@ calmodulin fusion described above can be used to detect the level of calcium ions in a sample .

= = Algorithmic detection of circular permutations = =

Many sequence alignment and protein structure alignment algorithms have been developed assuming linear data representations and as such are not able to detect circular permutations

between proteins . Two examples of frequently used methods that have problems correctly aligning proteins related by circular permutation are dynamic programming and many hidden Markov models . As an alternative to these , a number of algorithms are built on top of non @-@ linear approaches and are able to detect topology @-@ independent similarities , or employ modifications allowing them to circumvent the limitations of dynamic programming . The table below is a collection of such methods .

The algorithms are classified according to the type of input they require . Sequence @-@ based algorithms require only the sequence of two proteins in order to create an alignment . Sequence methods are generally fast and suitable for searching whole genomes for circularly permuted pairs of proteins . Structure @-@ based methods require 3D structures of both proteins being considered . They are often slower than sequence @-@ based methods , but are able to detect circular permutations between distantly related proteins with low sequence similarity . Some structural methods are topology independent , meaning that they are also able to detect more complex rearrangements than circular permutation .