

Circular Permutation in Proteins

Spencer Bliven^{1*}, Andreas Prlić^{2*}

1 Bioinformatics Program, University of California, San Diego, La Jolla, California, United States of America, **2** San Diego Supercomputer Center, University of California San Diego, La Jolla, California, United States of America

This is a “Topic Page” article for *PLoS Computational Biology*.

Circular permutation describes a type of relationship between proteins, whereby the proteins have a changed order of amino acids in their protein sequence, such that the sequence of the first portion of one protein (adjacent to the N-terminus) is related to that of the second portion of the other protein (near its C-terminus), and vice versa (see Figure 1). This is directly analogous to the mathematical notion of a cyclic permutation over the set of residues in a protein.

Circular permutation can be the result of evolutionary events, post-translational modifications, or artificially engineered mutations. The result is a protein structure with different connectivity, but overall similar three-dimensional (3D) shape. The homology between portions of the proteins can be established by observing similar sequences between N- and C-terminal portions of the two proteins, structural similarity, or other methods.

History

In 1979, Bruce Cunningham and his colleagues discovered the first instance of a circularly permuted protein in nature [1]. 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 (see Figure 2). Later work confirmed the circular permutation between the pair [2] and showed that concanavalin A is permuted post-translationally [3] through cleavage and an unusual protein ligation [4].

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 [5]. In 1989, Karolin Luger and her colleagues introduced a genetic method for making circular permutations by carefully fragmenting and ligating DNA [6]. This method allowed for permutations to be introduced at arbitrary sites, and is still used today to design circularly permuted proteins in the lab.

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 lipid antigen presentation in humans. Christopher Ponting and Robert Russell identified a circularly permuted version of a saposin inserted into plant aspartic proteinase, which they nicknamed swaposin [7]. 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. The Circular Permutation Database [8] contains 2,238 circularly permuted protein pairs with known structures, and many more are known without structures [9]. The CyBase database collects proteins that are cyclic, some of which are

permuted variants of cyclic wild-type proteins [10]. 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 [11].

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 [12]. Other, less common, mechanisms have been proposed, such as “cut and paste” [13] or “exon shuffling.”

Permutation by Duplication

The earliest model proposed for the evolution of circular permutations is the permutation by duplication mechanism [1]. 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 (see Figure 3).

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 [14].

Saposin and swaposin. An example for permutation by duplication is the relationship between saposin and swaposin. Saposins are highly conserved glycoproteins that consist of an approximately 80 amino acid residue long protein forming a four alpha helical structure. They have a nearly identical placement of cysteine residues and glycosylation sites. The cDNA sequence that

Citation: Bliven S, Prlić A (2012) Circular Permutation in Proteins. *PLoS Comput Biol* 8(3): e1002445. doi:10.1371/journal.pcbi.1002445

Editor: Shoshana Wodak, University of Toronto, Canada

Published: March 29, 2012

Copyright: © 2012 Bliven, Prlić. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The RCSB PDB is managed by two members of the RCSB: Rutgers and UCSD, and is funded by National Science Foundation (NSF), National Institute of General Medical Sciences, Department of Energy (DOE), National Library of Medicine, National Cancer Institute, National Institute of Neurological Disorders and Stroke and National Institute of Diabetes and Digestive and Kidney Diseases. The RCSB PDB is a member of the wwPDB. This work was supported by the RCSB PDB grant NSF DBI 0829586. The funders had no role in the preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: sbliven@ucsd.edu (SB); andreas.prlic@gmail.com (AP)

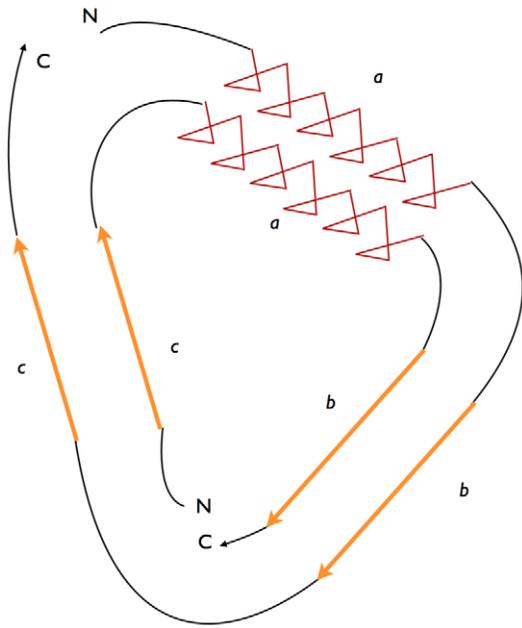


Figure 1. Schematic representation of a circular permutation in two proteins. The first protein (outer circle) has the sequence a-b-c. After the permutation the second protein (inner circle) has the sequence c-a-b. The letters N and C indicate the location of the amino- and carboxy-termini of the protein sequences and how their positions change relative to each other.
doi:10.1371/journal.pcbi.1002445.g001

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 [16]. This repeat suggests a mechanism for the evolution of the relationship with the plant-specific insert (PSI) (see Figure 4). The PSI is a domain exclusively found in plants, consisting of approximately 100 residues and found in plant aspartic

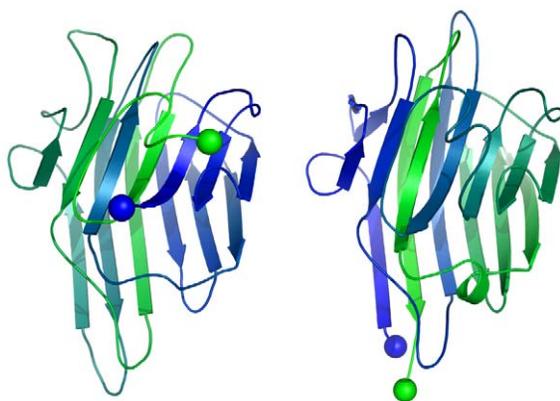


Figure 2. Two proteins that are related by a circular permutation. Concanavalin A (left), from the Protein Data Bank (PDB), 3cna and peanut lectin (right), from PDB 2pel, which is homologous to favin. The termini of the proteins are highlighted by blue and green spheres, and the sequence of residues is indicated by the gradient from blue (N-terminus) to green (C-terminus). The 3D fold of the two proteins is highly similar; however, the N- and C- termini are located on different positions of the protein [1].
doi:10.1371/journal.pcbi.1002445.g002

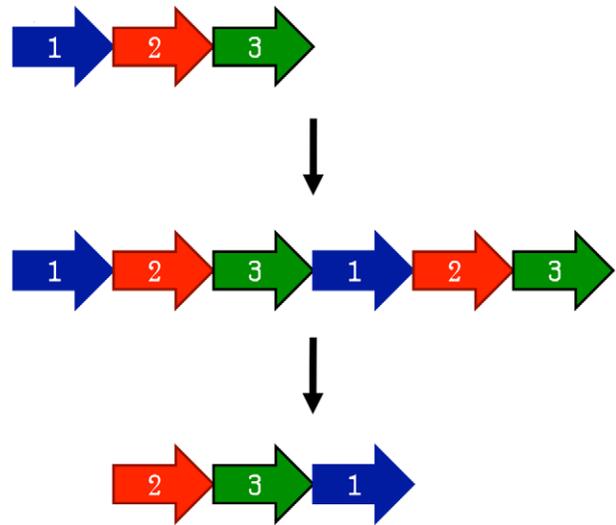


Figure 3. The permutation by duplication mechanism for producing a circular permutation. First, a gene is duplicated in place. Next, start and stop codons are introduced, resulting in a circularly permuted gene.
doi:10.1371/journal.pcbi.1002445.g003

proteases [17]. 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” [7]. For a review on functional and structural features of saposin-like proteins, see Bruhn (2005) [18].

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 (see Figure 5).

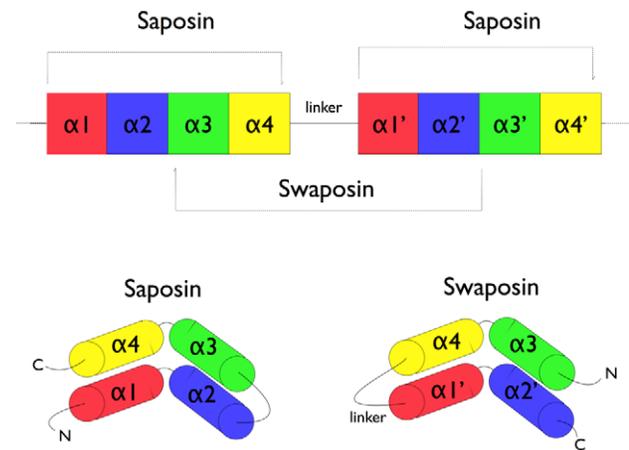


Figure 4. Suggested relationship between saposin and swaposin. They could have evolved from a similar gene [15]. Both consist of four alpha helices with the order of helices being permuted relative to each other.
doi:10.1371/journal.pcbi.1002445.g004

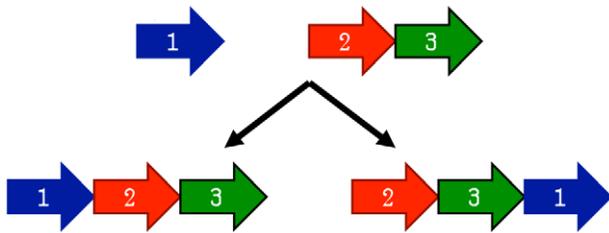


Figure 5. The fission and fusion mechanism of circular permutation. Two separate genes arise (potentially from the fission of a single gene). If the genes fuse together in different orders in two orthologues, a circular permutation occurs. doi:10.1371/journal.pcbi.1002445.g005

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 [19].

Transhydrogenases. An example for the fission and fusion mechanism can be found in nicotinamide nucleotide transhydrogenases [20]. 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 (see Figure 6). Phylogenetic analysis suggests that the three groups of domain arrangements were acquired and fused independently [12].

Other Processes that Can Lead to Circular Permutations

Post-translational modification. The two evolutionary models mentioned above describe ways in which genes may be

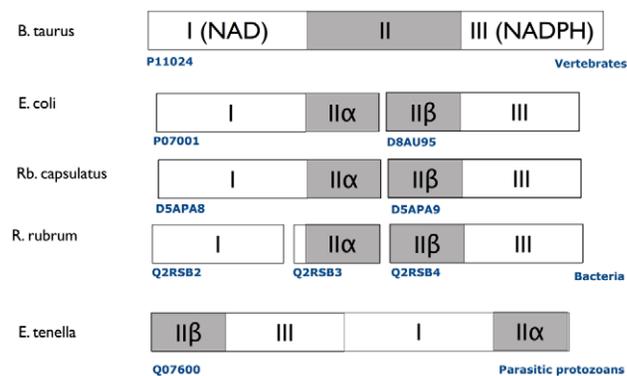


Figure 6. Transhydrogenases in various organisms can be found in three different domain arrangements. In cattle, the three domains are arranged sequentially. In the bacteria *E. coli*, *Rb. capsulatus*, and *R. rubrum*, the transhydrogenase consists of two or three subunits. Finally, transhydrogenase from the protist *E. tenella* consists of a single subunit that is circularly permuted relative to cattle transhydrogenase [20]. doi:10.1371/journal.pcbi.1002445.g006

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 auto-catalysis, as in the case of concanavalin A [4] (see Figure 2). Alternately, permutation may require restriction enzymes and ligases [5].

The Role of Circular Permutations in Protein Engineering

Many proteins have their termini located close together in 3D space [21,22]. 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 [6]. Although some permutation sites prevent the protein from folding correctly, many permuted proteins have been created with nearly identical structure and function to the original protein.

The motivation for creating a circular permuted 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 [23].
- **Improve catalytic activity.** Circularly permuting a protein can sometimes increase the rate at which it catalyzes a chemical reaction, leading to more efficient proteins [24].
- **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 [25].
- **Improve thermostability.** Making proteins active over a wider range of temperatures and conditions can improve their utility [26].

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 [27–29].
- **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 [30].
- **Modify quaternary structure.** Circularly permuted proteins have been shown to take on different quaternary structure than wild-type proteins [31].
- **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 [32]. 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 [33,24], or to detect the

Table 1. Algorithms for comparing pairs of circularly permuted proteins.

| Name | Type | Description | Author | Year | Availability | Reference |
|---------------|---------------------------------|---|------------------------|------|------------------|-----------|
| FBPLOT | Sequence | Draws dot plots of suboptimal sequence alignments. | Zuker | 1991 | | [34] |
| Bachar et al. | Structure, topology independent | Uses geometric hashing for the topology independent comparison of proteins. | Bachar et al. | 1993 | | [35] |
| Uliel et al. | Sequence | First suggestion of how a sequence comparison algorithm for the detection of circular permutations can work. | Uliel et al. | 1999 | | [36] |
| SHEBA | Structure | Duplicates a sequence in the middle; uses SHEBA algorithm for structure alignment; determines new cut position after structure alignment. | Jung, Lee | 2001 | | [37] |
| Multiprot | Structure, topology independent | Calculates a sequence order independent multiple protein structure alignment. | Shatsky | 2004 | Server, download | [38] |
| RASPODOM | Sequence | Modified Needleman and Wunsch sequence comparison algorithm | Weiner et al. | 2005 | Server | [39] |
| CPSARST | Structure | Describes protein structures as one-dimensional text strings by using a Ramachandran sequential transformation (RST) algorithm. Detects circular permutations through a duplication of the sequence representation and "double filter-and-refine" strategy. | Lo, Lyu | 2008 | Server | [40] |
| GANGSTA+ | Structure | Works in two stages: Stage one identifies coarse alignments based on secondary structure elements. Stage two refines the alignment on residue level and extends into loop regions. | Schmidt-Goenner et al. | 2009 | Server, download | [41] |
| SANA | Structure | Detect initial aligned fragment pairs (AFPs). Build network of possible AFPs. Use random-mate algorithm to connect components to a graph. | Wang et al. | 2010 | Download | [42] |
| CE-CP | Structure | Built on top of the combinatorial extension algorithm. Duplicates atoms before alignment, truncates results after alignment. | Bliven et al. | 2010 | Server, download | [43] |

doi:10.1371/journal.pcbi.1002445.t001

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 [32].

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. Table 1 is a collection of such methods.

Further Reading

- David Goodsell (2010) Concanavalin A and Circular Permutation. Research Collaboratory for Structural Biology (RCSB) Protein Data Bank (PDB) Molecule of the Month April 2010.
- Yu and Lutz (2011), for a review of the use of circular permutation in protein design [22].
- Weiner and Bornberg-Bauer (2006), for a review of evolutionary mechanisms for circular permutations [12].
- Cyclic permutation entry in Wikipedia, http://en.wikipedia.org/w/index.php?title=Cyclic_permutation

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.

Supporting Information

Text S1 Version history of the text file (XML); figures are also available in their original formats: Figure 1, Figure 2, Figure 3, Figure 4, Figure 5, Figure 6. (XML)

Text S2 Peer reviews and response to reviews. Human-readable versions of the reviews and authors' responses are available as comments on this article. (XML)

Acknowledgments

We want to thank Daniel Mietchen for carefully reading the manuscript and suggesting improvements. We would like to thank the other editors, Shoshana J. Wodak, Andrew M. Collings, and Philip E. Bourne for their guidance through the publishing process for this non-standard format. Thanks also to the referees for conducting open review. The version history of the text file and the peer reviews (and response to reviews) are available as supporting information in Text S1 and S2.

References

- Cunningham BA, Hemperly JJ, Hopp TP, Edelman GM (1979) Favin versus concanavalin A: Circularly permuted amino acid sequences. *Proc Natl Acad Sci U S A* 76: 3218–3222.
- Einspahr H, Parks EH, Suguna K, Subramanian E, Suddath FL (1986) The crystal structure of pea lectin at 3.0-Å resolution. *J Biol Chem* 261: 16518–16527.
- Carrington DM, Auffret A, Hanke DE (1985) Polypeptide ligation occurs during post-translational modification of concanavalin A. *Nature* 313: 64–67.
- Bowles DJ, Pappin DJ (1988) Traffic and assembly of concanavalin A. *Trends Biochem Sci* 13: 60–64.
- Goldenberg DP, Creighton TE (1983) Circular and circularly permuted forms of bovine pancreatic trypsin inhibitor. *J Mol Biol* 165: 407–413.
- Luger K, Hommel U, Herold M, Hofsteenge J, Kirschner K (1989) Correct folding of circularly permuted variants of a beta alpha barrel enzyme in vivo. *Science* 243: 206–210.
- Ponting CP, Russell RB (1995) Swaposins: circular permutations within genes encoding saposin homologues. *Trends Biochem Sci* 20: 179–180.
- CPDB – the Circular Permutation Database. Available: <http://sarst.life.nthu.edu.tw/cpdb/>. Accessed 16 February 2012.
- Lo WC, Lee CC, Lee CY, Lyu PC (2009) CPDB: a database of circular permutation in proteins. *Nucleic Acids Res* 37: D328–D332.
- Kaas Q, Craik DJ (2010) Analysis and classification of circular proteins in CyBase. *Biopolymers* 94: 584–591.
- Andreeva A, Prlić A, Hubbard TJP, Murzin AG (2007) SISYPHUS—structural alignments for proteins with non-trivial relationships. *Nucleic Acids Res* 35: D253–D259.
- Weiner J, Bornberg-Bauer E (2006) Evolution of circular permutations in multidomain proteins. *Mol Biol Evol* 23: 734–743.
- Bujnicki JM (2002) Sequence permutations in the molecular evolution of DNA methyltransferases. *BMC Evol Biol* 2: 3.
- Jeltsch A (1999) Circular permutations in the molecular evolution of DNA methyltransferases. *J Mol Evol* 49: 161–164.
- Ponting CP, Russell RB (1995) Swaposins: circular permutations within genes encoding saposin homologues. *Trends Biochem Sci* 20: 179–180.
- Hazkani-Covo E, Altman N, Horowitz M, Graur D (2002) The evolutionary history of prosaposin: two successive tandem-duplication events gave rise to the four saposin domains in vertebrates. *J Mol Evol* 54: 30–34.
- Guruprasad K, Törmäkangas K, Kervinen J, Blundell TL (1994) Comparative modelling of barley-grain aspartic proteinase: a structural rationale for observed hydrolytic specificity. *FEBS Lett* 352: 131–136.
- Bruhn H (2005) A short guided tour through functional and structural features of saposin-like proteins. *Biochem J* 389: 249–257.
- Lee J, Blaber M (2011) Experimental support for the evolution of symmetric protein architecture from a simple peptide motif. *Proc Natl Acad Sci U S A* 108: 126–130.
- Hatefi Y, Yamaguchi M (1996) Nicotinamide nucleotide transhydrogenase: a model for utilization of substrate binding energy for proton translocation. *FASEB J* 10: 444–452.
- Thornton JM, Sibanda BL (1983) Amino and carboxy-terminal regions in globular proteins. *J Mol Biol* 167: 443–460.
- Yu Y, Lutz S (2011) Circular permutation: a different way to engineer enzyme structure and function. *Trends Biotechnol* 29: 18–25.
- Whitehead TA, Bergeron LM, Clark DS (2009) Tying up the loose ends: circular permutation decreases the proteolytic susceptibility of recombinant proteins. *Protein Eng Des Sel* 22: 607–613.
- Cheltsov AV, Barber MJ, Ferreira GC (2001) Circular permutation of 5-aminolevulinate synthase. Mapping the polypeptide chain to its function. *J Biol Chem* 276: 19141–19149.
- Qian Z, Lutz S (2005) Improving the catalytic activity of *Candida antarctica* lipase B by circular permutation. *J Am Chem Soc* 127: 13466–13467.
- Topell S, Hennecke J, Glockshuber R (1999) Circularly permuted variants of the green fluorescent protein. *FEBS Lett* 457: 283–289.
- Viguera AR, Serrano L, Wilmanns M (1996) Different folding transition states may result in the same native structure. *Nat Struct Biol* 3: 874–880.
- Capraro DT, Roy M, Onuchic JN, Jennings PA (2008) Backtracking on the folding landscape of the beta-trefoil protein interleukin-1beta? *Proc Natl Acad Sci U S A* 105: 14844–14848.
- Zhang P, Schachman HK (1996) In vivo formation of allosteric aspartate transcarbamoylase containing circularly permuted catalytic polypeptide chains: implications for protein folding and assembly. *Protein Sci* 5: 1290–1300.
- Huang YM, Nayak S, Bystroff C (2011) Quantitative in vivo solubility and reconstitution of truncated circular permutants of green fluorescent protein. *Protein Sci* 20: 1775–1780.
- Beermink PT, Yang YR, Graf R, King DS, Shah SS, Schachman HK (2001) Random circular permutation leading to chain disruption within and near alpha helices in the catalytic chains of aspartate transcarbamoylase: effects on assembly, stability, and function. *Protein Sci* 10: 528–537.
- Baird GS, Zacharias DA, Tsien RY (1999) Circular permutation and receptor insertion within green fluorescent proteins. *Proc Natl Acad Sci U S A* 96: 11241–11246.
- Turner NJ (2009) Directed evolution drives the next generation of biocatalysts. *Nat Chem Biol* 5: 567–573.
- Zuker M (1991) Suboptimal sequence alignment in molecular biology. Alignment with error analysis. *J Mol Biol* 221: 403–420.
- Bachar O, Fischer D, Nussinov R, Wolfson H (1993) A computer vision based technique for 3-D sequence-independent structural comparison of proteins. *Protein Eng* 6: 279–288.
- Uliel S, Fliess A, Amir A, Unger R (1999) A simple algorithm for detecting circular permutations in proteins. *Bioinformatics* 15: 930–936.
- Jung J, Lee B (2001) Circularly permuted proteins in the protein structure database. *Protein Sci* 10: 1881–1886.
- Shatsky M, Nussinov R, Wolfson HJ (2004) A method for simultaneous alignment of multiple protein structures. *Proteins* 56: 143–156.
- Weiner J, Thomas G, Bornberg-Bauer E (2005) Rapid motif-based prediction of circular permutations in multi-domain proteins. *Bioinformatics* 21: 932–937.
- Lo WC, Lyu PC (2008) CPSARST: an efficient circular permutation search tool applied to the detection of novel protein structural relationships. *Genome Biol* 9: R11.
- Schmidt-Goenner T, Guerler A, Kolbeck B, Knapp EW (2010) Circular permuted proteins in the universe of protein folds. *Proteins* 78: 1618–1630.
- Wang L, Wu LY, Wang Y, Zhang XS, Chen L (2010) SANA: an algorithm for sequential and non-sequential protein structure alignment. *Amino Acids* 39: 417–425.
- Prlić A, Bliven S, Rose PW, Bluhm WF, Bizon C, Godzik A, Bourne PE (2010) Pre-calculated protein structure alignments at the RCSB PDB website. *Bioinformatics* 26: 2983–2985.