Hum Genet (1992) 90:342–345

human genetics

Original investigations

Of palindromes and peptides

Susumu Ohno

Beckman Research Institute of the City of Hope, 1450 E. Duarte Road, Duarte, CA 91010-0269, USA

Received: 22 July 1992

Abstract. On the average, 30% of the residues in a protein are members of peptidic palindromes, tripeptidic and longer. This percentage may go up to 50% in histones and certain other DNA binding proteins. The longest peptidic palindrome encountered thus far was 14 residues in length. However, there is every reason to expect even longer peptidic palindromes in other proteins not yet analyzed.

Introduction

In celebration of the 60th birthday of my old and very dear friend, Professor Ulrich Wolf, I begin by dedicating two palindromes connected with the word "Wolf". One rarely finds a surname that makes a comprehensible word when read from right-to-left. "Wolf" is one such exception in that when read in the opposite direction, it becomes "flow". Thus, "Wolf" is a surname very amenable to being a component of palindromes. "Wolf did flow" is indeed an apt description of one of his more inspired performances on his beloved cello. Nevertheless, it is no easy task to form a long palindrome containing the two words "Wolf" and "flow" in symmetrical positions. The best I have come up with, thus far, is the following: "Wolf's deeds lived was dog DNA, and God saw devil's deeds flow." Indeed, after much disagreement, the ancestry of the domestic dog appears to have been settled on wolves rather than jackals.

Furthermore, it is a fact that in European folklore, the dog is a messenger of Satan, whereas the horse is a harbinger of God.

In spite of the difficulty noted above, mankind the world over has developed a consuming interst in palindromes ever since the invention of writing. The beauty of perfect symmetry embodied in each palindrome is the apparent reason for so persistent a consuming passion.

The beauty is more apparent in numerical palindromes, e.g., 1 2 3 5 8 5 3 2 1. One may assign the value of a full note to numeral "1" and that of an 8th note to numeral "8", or one might do the converse. Either way, one would find that each numerical palindrome becomes an inherently pleasing rhythm, because of its perfect symmetry. It is small wonder that composers of ages past have also experimented with palindromes. The more successful pieces are those by Mozart and Haydn.

In view of their perfect symmetry, there is little doubt that palindromic peptide fragments within a protein should contribute a remarkable versatility to it. Yet, since the notification of their presence (Ohno 1989), not much attention has been paid to peptidic palindromes. Thus, I shall begin by presenting some of the longer peptidic palindromes found in diverse proteins.

Long peptidic palindromes found in diverse proteins

The longest peptidic palindrome that I have encountered, thus far, was 14 residues long. It is shown at the top of Fig. 1 as palindrome no. 1. It occupied positions 190–203 of mouse histone H1 variety 1 (Cheng et al. 1989). Because of its palindromic nature, 6 lysines were spaced apart by 2, 1, 2, 1, 2 intervening residues, thus, achieving perfect symmetry in their distribution. Histone H1 is a component of the nucleosome around which DNA wraps itself. The perfectly symmetrical distribution of basic residues in histone H1 offers two alternative ways in which DNA can bind to this histone. Perhaps for this reason, one tends to find long peptidic palindromes in all histones. For example, mouse histone H1 var-1 contained, in addition to the tetradecapeptidic palindrome noted above, one octapeptidic, two heptapeptidic, two hexapeptidic and as many as five pentapeptidic palindromes. When the tetrapeptidic and tripeptidic palindromes were added together, 110 of the 212 residues (52%) of this histone were involved in palindromes. Of those found in mouse histone H1 var-1, two heptapeptidic palindromes are also shown in Fig. 1 as nos. 8 and 9. These two are included in Fig. 1, for they were found, not only in H1 histones of other species, but also in other histones, i.e., H2 and H3. The same applies to nonapeptidic palindrome no. 3 derived from histone H1 of the rabbit and cattle (Dayhoff 1972) as well as to heptapeptidic palindrome no. 10 derived from human histone H1.3 (Albig et al. 1991).

Other long peptidic palindromes of Fig. 1 were chosen because of their divergent amino acid compositions.

This article is dedicated to Professor Ulrich Wolf in honor of his 60th birthday

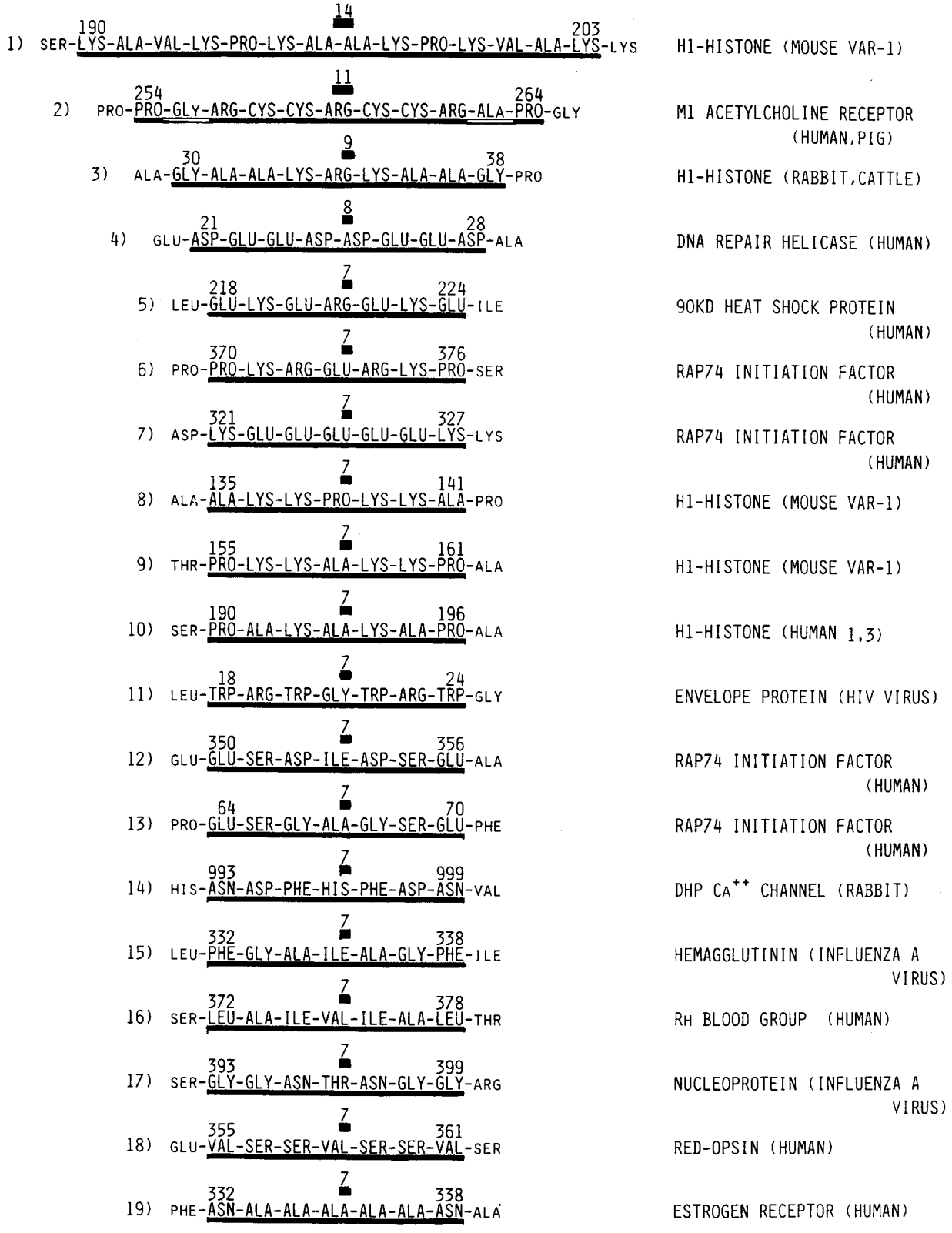

Fig. 1. The list of 19 long divergent peptidic palindromes. They are heptapeptidic and longer in length, differ in amino acid composition and are derived from divergent proteins. Palindromic parts are *underlined* and their sources are indicated at the right. References to these proteins are given in the text. The position of palindromes in a protein is indicated by *numbers* placed above the first and last residues

Monodecapeptidic palindrome no. 2, derived from the human M1 acetylcholine receptor (Allard et al. 1987), for example, contained four cysteine residues. As repeatedly noted, in the average amino acid composition, cysteine is the second rarest residue after tryptophan. Accordingly, the inclusion of four cysteines in one palin-

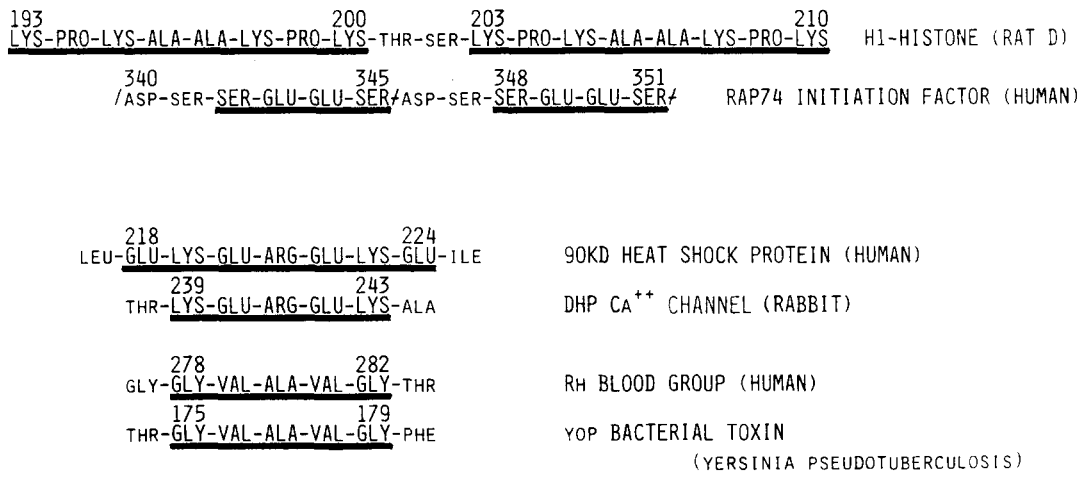

Fig. 2. Two examples of tandem repeats of peptidic palindromes are shown at the *top*, whereas two examples of the ubiquity of pentapeptidic palindromes are shown at the *bottom*

drome was very significant. So were the inclusion of three tryptophan residues in heptapeptidic palindrome no. 11 derived from the envelope protein of HIV-III virus (Ratner et al. 1985). Although this 863-residue-long viral protein contained 22 tryptophans (2.6× greater than the average), 6 additional tryptophans were included in three shorter palindromes. The octapeptidic palindrome no. 4 contained in human DNA repair helicase (Weeda et al. 1990) comprised only acidic residues. However, both acidic and basic residues were included in palindrome no. 5 derived from human 90-kDa heat shock protein (Rebbe et al. 1989), as well as in palindromes no. 6 and 7, both contained within the RAP74 protein, which is a general initiation factor for transcription by RNA polymerase II (Finkelstein et al. 1992). This RAP74 protein also furnished heptapeptidic palindromes no. 12 and 13, each containing a pair of glutamic acids. A pair of phenylalanines were contained in two heptapeptidic palindromes: no. 14 was derived from the rabbit Ca^{++} channel that was recognized as the DHP receptor (Tanabe et al. 1987), while no. 15 was derived from influenza A virus hemagglutinin (Verhoyen et al. 1980). Nucleoprotein of influenza A virus furnished heptapeptidic palindrome no. 17, centered by a polar residue, threonine (Winter and Fields 1981). Palindrome no. 16, derived from human Rh blood group-related protein (Cherif-Zahar et al. 1990), consisted only of hydrophobic residues. Palindrome no. 18, contained in human rhodopsin (Nathans et al. 1986), included four serines that are polar. Heptapeptic palindrome no. 19 containing five successive alanines derived from human estrogen receptor (Greene et al. 1986) is deliberately shown to dispute a widely held misconception that homo-oligopeptides are rare. In fact, homo-oligopeptides, pentapeptidic to heptapeptidic in lenght, were regularly found in divergent proteins. Needless to say, all the homo-oligopeptides are palindromes. Five successive glutamic acids contained in palindrome no. 7 should also be recalled.

Figure 1 shows that long peptidic palindromes are to be found in divergent proteins and that their amino acid compositions are very diverse, often containing rare residues such as tryptophan, cysteine and histidine. Although not shown, another rare residue, methionine, is also found in long palindromes.

I have already noted that histones tend to be exceptionally rich in peptidic palindromes – residues involved in tripeptidic and longer palindromes approaching or exceeding 50% of the total. Excluding histones, the average content of peptidic palindromes was estimated at 28.9% based upon the sampling of 54 diverse proteins. Positioned at the opposite end from histones were various sugar metabolizing enzymes in which the average palindromic content was a mere 18.3%. These sugar metabolizing enzymes evolved soon after the beginning of life on this earth, therefore, long before the division of eukaryotes from prokaryotes, and each has been performing the same task ever since. Perhaps, their venerable old age has something to do with the paucity of peptidic palindromes. Indeed, these sugar metabolizing enzymes were rich in degenerate palindromes, e.g., Val-*Asp*-Phe-*Asn*-Val.

Tandem repeats of peptidic palindromes and their ubiquity

Peptidic palindromes are often found in tandem. Two such examples are shown at the top of Fig. 2. Within rat histone H1.D (Cole et al. 1990), two copies of the octapeptidic palindrome appeared in succession separated by two residues. Needless to say, this octapeptidic palindrome is a core portion of the tetradecapeptidic palindrome shown at the top of Fig. 1. Such tandem repetition can be viewed either as a drive toward a longer palindrome or a degeneracy that has split a longer palindrome into two shorter, identical ones. In palindromes, the beginning is also the end, thus, the end is also the beginning. At any rate, within positions 193–210 of rat histone H1.D, 8 lysines are already spaced symmetrically apart by 1, 2, 1, 2, 1, 2, 1 intervening residues (Fig. 2, top). Changing Thr-Ser at position 201–202 to either Ser-Ser

or Thr-Thr, would result in an octadecapeptidic palindrome. That would be a mere formality, however, for serine and threonine, both having an hydroxyl group, are already brothers in arms.

Shorter hexapeptidic tandem repeats found within human RAP74 general initiation factor are also shown in the second row of Fig. 2. Were the 346th and 447th positions changed from Asp-Ser to Asp-Asp, a new decapeptidic palindrome in which 4 serines are spaced apart by 2 acidic intervening residues would be born.

In all forms of communication devices, such as languages, an identical phrase, sentence and even paragraph can be used in essays dealing with divergent subjects. For example, the blood stirring phrase "toujours l'audace, toujours l'audace" (always with audacity) can be the dictum of military strategists, but it can also serve as a very appropriate dictum for horsemen or tennis players and even for stock brokers.

The same is true of proteins: contrary to popular belief, identical oligopeptides are regularly found in totally unrelated proteins (Ohno 1991), and they are more often than not peptidic palindromes. Since a dictionary of ubiquitous oligopeptides will be compiled in the near future, I shall only give two examples of ubiquitous pentapeptidic palindromes.

As shown for the rabbit in the middle of Fig. 2, the pentapeptidic palindrome Lys-Glu-Arg-Glu-Lys was the core of the no. 5 heptapeptidic palindrome already shown in Fig. 1, as derived from the human 90 KD heat shock protein, and the same was found in the rabbit Ca^{+} channel, identified as the DHP receptor. Human Rh blood group protein and yopA protein of *Yersinia pseudotuberculosis* (Rosqvist et al. 1988) shared the pentapeptidic palindrome: Gly-Val-Ala-Val-Gly. *Y. pseudotuberculosis* is a less virulent relative of *Yersinia pestis*; the pathogen of the dreaded black plague.

Functional significance of peptidic palindromes

Peptides can assume only three secondary structures, forming β-pleated sheets or α-helices, or, failing in the above two alternatives, assuming random coil configurations. Each β-pleated sheet tends to stabilize itself by forming hydrogen bonds with a neighboring β-pleated sheet that is placed side-by-side. Such a pair of β-pleated sheets can either be in parallel or in anti-parallel orientation. If both are peptidic palindromes, being parallel or anti-parallel does not matter, for hydrogen bonds should be formed between the identical pairs of residues in both orientations.

The plasma membrane is 5 mm thick, thus, in order to transverse the plasma membrane, each α-helix must be made of 22 residues. If this transmembrane α-helix is a peptidic palindrome, it matters not whether its amino-terminal or its carboxyl end remains inside the plasma membrane. Indeed, each transmembrane α-helix appears to be a degenerate peptidic palindrome (Ohno 1989).

References

Albig W, Kardalinou E, Drabent B, Mimmer A, Doenecke D (1991) Isolation and characterization of two H1 histone genes within clusters of core histone genes. Genomics 10:940–948

Allard JW, Sigal IS, Dixon RAF (1987) Sequence of the gene encoding the human M1 muscarinin acetylcholine receptor. Nucleic Acids Res 15:10604

Cheng G, Nandy A, Clerk S, Skoultchi AI (1989) Different 3′-end processing produces two independently regulated mRNAs from a single H1 histone gene. Proc Natl Acad Sci USA 86:7002–7006

Chérif-Zahar B, Bloy C, Le Van Kim C, Blanchard D, Bailly P, Hermand P, Salmon C, Cartron J-P, Colin Y (1990) Molecular cloning and protein structure of a human blood group Rh polypeptide. Proc Natl Acad Sci USA 87:6243–6247

Cole KD, Kandala JC, Kremer E, Kistler WS (1990) Isolation of a genomic clone encoding the rat histone variant H1D gene. Gene 89:265–269

Dayhoff MO (1972) Atlas of protein sequence and structure, vol 5. The National Biomedical Foundation, Silver Spring, Md

Finkelstein A, Kostrub CF, Li J, Chavez DP, Wang BQ, Fang SM, Greenblatt J, Buron ZF (1992) A cDNA encoding RAP74, a general initiation factor for transcription by RNA polymerase II. Nature 355:464–467

Greene GL, Gilna P, Waterfield M, Baker A, Hort Y, Shine J (1986) Sequence and expression of human estrogen receptor complementary DNA. Science 231:1150–1154

Nathans J, Thomas D, Hogness DS (1986) Molecular genetics of human color vision: the genes encoding glue, green, and red pigments. Science 232:193–202

Ohno S (1989) Intrinsic evolution of proteins: The role of peptidic palindromes. Rev Biol (Biology Forum) 82:341–343

Ohno S (1991) To be or not to be a responder in T cell responses: ubiquitous oligopeptides in all proteins. Immunogenetics 34:215–221

Ratner L, Haseltine W, Patarca R, Livak KJ, Starcich B, Josephs SF, Doran ER, Rafalski JA, Whitehorn EA, Baumeister K, Ivanoff L, Petteway SR Jr, Pearson ML, Lautenberger JA, Papas TS, Ghrayeb J, Chang NT, Gallo RC, Wong-Staal F (1985) Complete nucleotide sequence of the Aids virus, HTLV-III. Nature 313:277–284

Rebbe NF, Hickman WS, Ley TJ, Stafford DW, Hickman SN (1989) Nucleotide sequence and regulation of a human 90-kDa heat shock protein gene. J Biol Chem 264:15006–15011

Roqvist R, Kurnik M, Wolf-Watz H (1988) Increased virulence of *Yersinia pseudotuberculosis* by two independent mutations. Nature 334:522–525

Tanabe T, Takeshima H, Mikami A, Flockerzi V, Takahashi H, Kanagawa K, Kojima M, Matsuo H, Hirose T, Numa S (1987) Primary structure of the receptor for calcium channel blockers from skeletal muscle. Nature 328:313–318

Verhoeyen M, Fang R, Jou WM, Devous R, Huylebroeck D, Saman E, Fiers W (1980) Antigenic drift between the haemagglutinin of the Hong Kong influenza strains A/Aichi/2/68 and A/Victoria/3/75. Nature 286:771–776

Weeda G, Van Ham RCA, Vermeulen W, Bootsma D, Der Eb AJ, Hoeijmakers JHJ (1990) Identification of the molecular defect involving the human repair disorders xeroderma pigmentosum and Cockayne's syndrome in the ERCC-3-encoding a presumed DNA repair helicase. Mol Cell Biol 10:2570–2581

Winter G, Fields S (1981) The structure of the gene encoding the nucleoprotein of human influenza virus A/PR/8/34. Virology 114:423–428