

PEPTIDES BINDING TO PARALLEL-STRANDED G-QUADRUPLEXES

FIELD OF THE INVENTION

[0001] The present invention lies in the field of biochemistry and relates to a peptide comprising or consisting of the amino acid sequence of SEQ ID NO:1 or a derivative thereof. The present invention also relates to conjugates of said peptide, the use of the peptide or the conjugate as a medicament and to methods for the detection of parallel-stranded G-quadruplexes.

BACKGROUND OF THE INVENTION

[0002] G-quadruplexes are non-canonical nucleic acid structures, formed by the stacking of G-tetrads and stabilized by cations (Phan, AT (2010) FEBS Journal, 277, 1107). Guanine-rich sequences, found ubiquitously in the genome of various species including the human genome, are able to fold into G-quadruplexes. Interestingly, these sequences are non-randomly distributed; they are localized in specific parts of the genome including the telomeres and promoter regions (Todd, A. K et al. (2005) Nucleic Acids Res, 33, 2901; Huppert, J. L. and Balasubramanian, S. (2005) Nucleic Acids Res, 33, 2908; Huppert, J. L. (2008) Biochimie, 90, 1140).

[0003] For a long time, these structures were believed to have no relevant contribution to biological processes. However, during the past decade strong evidences were found and pointed to significant roles of G-quadruplex structures in biological processes. In fact, it has been now established that G-quadruplexes are involved in essential cellular functions such as transcription, replication and recombination (Piazza, A et al. (2010) NucleicAcidsRes, 38, 4337; Biffi, G et al. (2013) Nat Chem, 5, 182; Rodriguez, R et al. (2012) Nat ChemBiol, 8, 301). For example, the formation and stabilization of G-quadruplexes has been shown to promote genomic instability of mini-satellites in yeast cells (Piazza, A et al. (2010) Nucleic Acids Res, 38, 4337).

[0002] It has been shown that formation and stabilization of G-quadruplexes in telomeres regions, in the promoters of oncogenes, or in the 5'-UTR regions of pre-mRNA oncogenes has anti-cancer effects (Sun, D. et al. (1997) *J Med Chem* 40, 2113; Mergny J.L., Hélène C. (1998) *Nat Med.* 4, 1366; Balasubramanian, S and Neidle, S (2009) *Curr Opin Chem Biol*, 13, 345). These recent findings stimulated the development of synthetic compounds meant to induce/stabilize G-quadruplex structures. For example, Telomestatin (Shin-ya, K et al. (2001) *J Am Chem Soc*, 123, 1262) and Phen-DC3 (De Cian et al. (2007) *J Am Chem Soc*, 129, 1856) are among the most relevant compounds, which bind to G-quadruplexes with a high affinity (with a dissociation constant K_d in the nanomolar range) and a good selectivity against duplex and single-stranded nucleic acids.

[0003] G-quadruplex structures are highly diverse in regard of relative strand orientations and loop types, resulting in different topologies including a (i) parallel-type in which four strands point in the same direction; an (ii) hybrid "3+1" type in which three strands point in one direction and the fourth strand points in the opposite one; (iii) an antiparallel-type in which two strands point in one direction and two strands point in the opposite direction. Those topologies lead to different structural molecular shapes, with various loops and grooves of different size and accessibility (Phan, AT (2010) *FEBS Journal*, 277, 1107). The structural polymorphism of G-quadruplexes depends on their nucleotide sequences and the environmental conditions. In the human genome, G-rich sequences are scattered in different regions of the chromosomes and can form different possible G-quadruplex topologies. So far, most of G-quadruplex binders do not present selectivity against different G-quadruplex topologies and thus exhibit a wide-range genome effects (Biffi, G. et al. (2013) *Nat Chem*, 5, 182; Rodriguez, R et al. (2012) *Nat Chem Biol*, 8, 301).

[0004] Recently several proteins have been reported to interact with G-quadruplexes (Fry, M (2007) *Front Biosci*, 13, 9824; Murat, P et al. (2013) *Curr Opin Genet Dev.* 31, 22). One of such proteins is Rhau (also named DHX36 or G4R1). Rhau is a human helicase of the DEAH-box family, present in all type of cells (Tran, H et al. (2004) *Mol Cell*, 13, 101) and associated with different functions, including the formation of stress granules, interchromatin granule clusters (Chalupnikova, K et al. (2008) *J Biol Chem*, 283, 35186) and the degradation of

urokinase plasminogen activator mRNA. Interestingly, studies by Nagamine and colleagues (Creacy, SD et al. (2008) J BiolChem, 283, 34626; Booy, EP et al. (2012) Nucleic Acids Res, 40, 4110) demonstrated that Rhau protein specifically unwind and bind G-quadruplexes nucleic acids. The G-quadruplex binding domain was identified to be in the N-terminal region of the protein, ranging from residue 53 to 105 (termed Rhau55).

[0005] Due to the structural diversity of G-quadruplexes, it is an object of the present invention to identify binders that selectively discriminate between different G-quadruplex topologies.

SUMMARY OF THE INVENTION

[0006] It is an object of the present invention to meet the above need by providing a peptide comprising or consisting of (i) the amino acid sequence of SEQ ID NO:1 or (ii) an amino acid sequence having at least 80% sequence identity to the amino acid sequence of SEQ ID NO:1 over its entire length. Surprisingly, the inventors have found that such peptide selectively binds to parallel-stranded G-quadruplexes. Hence, said peptide can be used as a medicament. Further, the specific binding to parallel-stranded G-quadruplexes provides the basis for methods to detect such G-quadruplexes. In addition, the present invention relates to conjugates comprising the above peptide.

[0007] In a first aspect, the present invention is thus directed to a peptide comprising or consisting of (i) the amino acid sequence of SEQ ID NO:1 or (ii) an amino acid sequence having at least 80% sequence identity to the amino acid sequence of SEQ ID NO:1 over its entire length.

[0008] In various embodiments of the invention, said peptide is 16 to 50 amino acids in length. In preferred embodiments, said peptide is 18 to 50 amino acids in length.

[0009] In a further aspect, the present invention relates to a conjugate comprising the peptide described above wherein said peptide further comprises at least one functional moiety.

[00010] In various embodiments of the invention, said at least one functional moiety is conjugated to the N-terminus of said peptide. In various other embodiments, said at least one functional moiety is conjugated to the C-terminus of said peptide. In still further embodiments, said at least one functional moiety does not comprise the amino acid sequence set forth in SEQ ID NO:2 or a C-terminal fragment thereof or said at least one functional moiety does not comprise the amino acid sequence set forth in SEQ ID NO:3 or an N-terminal fragment thereof.

[00011] In various embodiments of the invention, the functional moiety is a tag or a pharmaceutically active compound.

[00012] In a further aspect, the present invention relates to a peptide or a conjugate as described above for use as a medicament.

[00013] In a third aspect, the present invention relates to a method for the detection of parallel-stranded G-quadruplexes in a sample wherein said method comprises: contacting the peptide or the conjugate as described above with a sample suspected to contain a parallel-stranded G-quadruplex and detecting the presence or absence of the parallel-stranded G-quadruplex in said sample.

[00014] In a still further aspect, the present invention relates to the use of a peptide or a conjugate as described for the detection of parallel-stranded G-quadruplexes.

BRIEF DESCRIPTION OF THE DRAWINGS

[00015] The invention will be better understood with reference to the detailed description when considered in conjunction with the non-limiting examples and the accompanying drawings.

[00016] **Figure 1** shows the nucleotides and their corresponding sequences (SEQ ID Nos. 4-18) used in the present application.

[00017] **Figure 2** shows the peptides and their corresponding sequences (SEQ ID Nos. 19-31) used in the present application.

[00018] **Figure 3** (A) shows the folding topology of DNA and RNA G-quadruplexes used in this study. A) T95_2T, B) 93del, C) Pu24t, D) CEB25, E) CEB1, F) Htelo1/HT1, G) Htelo2, H) Htelo3 I) Htelo4, J) TERC18, K) 12-nt TER and L) 10-nt TER. Guanines in anti and syn conformation are colored in gray and white respectively. Cytosines are colored in black.

[00019] **Figure 4** shows the NMR imino protons of DNA and RNA G-quadruplexes sequences used in this study, name of the sequences are labeled on top of each spectrum.

[00020] **Figure 5** shows the native gel electrophoresis of 93del, CEB25, Pu24T, CEB1, TERC18, 12nt TER and 10nt TER. The ratio between Rhau55 and Nucleic acids is indicated on top of each lane. DNA concentrations were fixed at 100 μ M, in 70 mM KCL and 20 mM phosphate buffer (pH= 6.8). DNA on the gel was revealed using UV shadowing.

[00021] **Figure 6** shows the native gel electrophoresis of Htelo1, Htelo2, Htelo3, and Htelo4 in dilute and crowding condition (induced by 40% PEG (v/v)). DNA concentrations were fixed at 100 nM and Rhau55 was 1 μ M, in 70 mM KCL and 20 mM phosphate buffer (pH= 6.8). DNA on the gel was revealed using FAM or TAMRA fluorescence dye attached to the extremity of the DNA.

[00022] **Figure 7** shows the NMR imino proton spectra of a mixture of HT1 (filled dot) and T95_2T (open dot), in the absence and the presence of different amount of Rhau55.

[00023] **Figure 8** shows the gel mobility shift assay for T95_2T, Htelo1 in dilute and crowding condition (induced by 40% PEG (v/v)) binding by Rhau55. (A) T95_2T and Htelo1 at a concentration of 100 nM were incubated for 15 minutes at room temperature with an increasing concentration of Rhau55. (B) Quantification of gel electrophoresis assays bands of T95_2T (black dot) and Htelo1 in dilute (open square) and crowding condition (black cross).

[00024] **Figure 9** shows the 2D ^1H - ^{15}N HSQC spectra of free Rhau55 (A) and bound to T95_2T (B). (C) Chemical-shift variations $\Delta\delta$ (ppm) = $[(\Delta^{15}\text{N})^2 + (\Delta^1\text{H})^2]^{1/2}$ between the free and bound Rhau55.

[00025] **Figure 10** shows the native gel electrophoresis of T95_2T. Lane 1 T95_2T, lane 2 to lane 12: addition of Rhau20m1, Rhau20m2, Rhau5, Rhau9, Rhau12, Rhau14, Rhau15',

Rhau16, Rhau20, Rhau18, Rhau29 and Rhau23. DNA concentrations were fixed at 100 μ M, in 70 mM KCL and 20 mM phosphate buffer (pH= 6.8), and the ratio between peptide and DNA was kept at 1:5. DNA on the gel was revealed using UV shadowing.

[00026] **Figure 11** shows the NMR imino proton spectra of T95_2T titrated with increasing concentrations of Rhau55 (A), Rhau18 (B) and Rhau16 (C).

[00027] **Figure 12** shows the native gel electrophoresis of T95_2T, Htelo1, Htelo1 in crowding condition (induced by 40% PEG (v/v)), ST, DX and 93del. The ratio between Rhau18 and DNA is indicated on top of each lane. DNA concentrations were fixed at 100 μ M, in 70 mM KCL and 20 mM phosphate buffer (pH= 6.8). DNA on the gel was revealed using UV shadowing.

[00028] **Figure 13** shows the NMR imino proton spectra of T95_2T (A) with increasing concentrations of Rhau18sm. Peaks arising from the complex are marked with asterix. Complex was obtained with an equimolar ratio of DNA and peptide. (B) G-tetrad representation and folding topology adopted by the T95_2T sequence. (C) NOESY spectrum (mixing time 300ms) at 370C, exchange cross peak are labeled with residue number. (D) Strip plot showing the intermolecular cross peak from ^{13}C - ^1H HSQC-NOESY spectrum.

[00029] **Figure 14** shows the representation of an NMR-based model of the Rhau18-T95_2T complex. Rhau18 is colored in black and T95_2T in white. Rhau18 is presented in surface mode and DNA in sticks mode.

DETAILED DESCRIPTION OF THE INVENTION

[00030] The present inventors surprisingly found that a peptide comprising or consisting of (i) the amino acid sequence of SEQ ID NO:1 or (ii) an amino acid sequence having at least 80% sequence identity to the amino acid sequence of SEQ ID NO:1 over its entire length selectively binds to parallel-stranded G-quadruplexes.

[00031] Thus, in a first aspect, the present invention is thus directed to a peptide comprising or consisting of (i) the amino acid sequence of SEQ ID NO:1 or (ii) an amino acid sequence having at least 80% sequence identity to the amino acid sequence of SEQ ID NO:1 over its entire length.

[00032] In various embodiments of the invention, said peptide comprises or consists of an amino acid sequence having at least 85%, more preferably at least 90%, at least 95%, or at least 97% sequence identity to the amino acid sequence of SEQ ID NO:1 over its entire length.

[00033] In various embodiments of the invention, said peptide is 16 to 50 amino acids in length. In various other embodiments of the invention, said peptide is 16 to 1000, or more preferably 16 to 500, 16 to 200, 16 to 150, 16 to 100, 16 to 75, 16 to 40, 16 to 35, 16 to 30 or 16 to 25 amino acids in length, or said peptide is 18 to 1000, or more preferably 18 to 500, 18 to 200, 18 to 150, 18 to 100, 18 to 75, 18 to 50, 18 to 40, 18 to 35, 18 to 30 or 18 to 25 amino acids in length. In more preferred embodiments of the invention, the peptide of at least 18 amino acids in length comprises or consists of the peptide sequence set forth in SEQ ID NO:23.

[00034] In a further aspect, the present invention relates to a conjugate comprising the peptide described above wherein said peptide further comprises at least one functional moiety.

[00035] In various embodiments of the invention, said at least one functional moiety is conjugated to the N-terminus of said peptide. In various other embodiments, said at least one functional moiety is conjugated to the C-terminus of said peptide. In still various other embodiments, said at least one functional moiety is conjugated to the C-terminus and the N-terminus of said peptide. In further embodiments, said at least one functional moiety does not comprise the amino acid sequence set forth in SEQ ID NO:2 or a C-terminal fragment thereof or said at least one functional moiety does not comprise the amino acid sequence set forth in SEQ ID NO:3 or an N-terminal fragment thereof. In other various embodiments, if the functional moiety is a peptide and said peptide consists of at most 100, preferably 70, 50, 30, 25, 20, 15 or 10 amino acids.

[00036] In various embodiments of the invention, the functional moiety is a tag or a pharmaceutically active compound.

[00037] In a further aspect, the present invention relates to a peptide or a conjugate as described above for use as a medicament.

[00038] In another aspect, the present invention relates to a method for the detection of parallel-stranded G-quadruplexes in a sample wherein said method comprises: contacting the peptide or the conjugate as described above with a sample suspected to contain a parallel-stranded G-quadruplex and detecting the presence or absence of the parallel-stranded G-quadruplex in said sample.

[00039] In a still further aspect, the present invention relates to the use of a peptide or a conjugate as described for the detection of parallel-stranded G-quadruplexes.

[00040] The terms "polypeptide", "protein", and "peptide", which are used interchangeably herein, refers to a polymer of the 20 protein amino acids, or amino acid analogs, regardless of the size or function of the molecule. Although "protein" is often used in reference to relatively large polypeptides, and "peptide" is often used in reference to small polypeptides, usage of these terms in the art overlaps and varies. Thus, the above terms relate to one or more associated molecules, wherein the molecules consist of amino acids coupled by peptide (amide) bonds. The amino acids are preferably the 20 naturally occurring amino acids glycine, alanine, valine, leucine, isoleucine, phenylalanine, cysteine, methionine, proline, serine, threonine, glutamine, asparagine, aspartic acid, glutamic acid, histidine, lysine, arginine, tyrosine and tryptophan.

[00041] The peptides and conjugates of the invention can be synthesized synthetically or can be expressed in an organism or can be produced by *in vitro* transcription/translation. The peptides or conjugates may be expressed in, but such expression is not limited to *Escherichia coli*, *Saccharomyces cerevisiae*, *Candida albicans*, *Pichia Pastoris*, insect cells such as Sf9 (*Spodoptera frugiperda*) cells, Nicotiana (tobacco plant) and CHO (Chinese hamster ovary) cells. Alternatively, the peptide or conjugate of the invention are expressed by an *in vitro* transcription/translation or "IVTT" system. "IVTT reaction" or "*in vitro* transcription translation reaction", as interchangeably used herein, relates to cell-free systems that allow for specific transcription and translation by comprising macromolecular components (RNA polymerase, 70S or 80S ribosomes, tRNAs, aminoacyl-tRNA synthetases, initiation, elongation and termination

factors, etc.) required for transcription and translation. To ensure efficient translation, the system may also be supplemented with amino acids, energy sources (ATP, GTP), energy regenerating systems, and other co-factors (Mg^{2+} , K^{+} , etc.). Such systems or extracts are also known as “coupled” and “linked” systems as they start with DNA templates, which are subsequently transcribed into RNA and then translated. Preferred IVTT reactions comprise the rabbit reticulocyte lysate, the wheat germ extract and the *E. coli* cell-free system, in a more preferred embodiment the IVTT reaction is the rabbit reticulocyte lysate.

[00042] Preferably the synthesis of the peptide or conjugate of the invention is a synthetic synthesis. Methods of synthetic peptide synthesis include, but are not limited to liquid-phase peptide synthesis and solid-phase peptide synthesis (SPPS). Methods to produce peptides synthetically and according protocols are well-known in the art (Nilsson, BL et al. (2005) *Annu Rev BiophysBiomolStruct*, 34, 91). The synthesized peptides may be further modified by the attachment of additional chemical moieties.

[00043] The term “sequence”, as used herein, relates to the primary nucleotide sequence of nucleic acid molecules or the primary amino acid sequence of a protein.

[00044] As used herein, “sequence identity” or “identity” in the context of two peptide sequences makes reference to the residues in the two sequences that are the same position when aligned for maximum correspondence over a specified comparison window. When percentage of sequence identity is used in reference to proteins, it is recognized that residue positions which are not identical often differ by conservative amino acid substitutions, where amino acid residues are substituted for other amino acid residues with similar chemical properties (e.g., charge or hydrophobicity) and therefore do not change the functional properties of the molecule. When sequences differ in conservative substitutions, the percent sequence identity may be adjusted upwards to correct for the conservative nature of the substitution. Sequences that differ by such conservative substitutions are said to have “sequence similarity” or “similarity”. Means for making this adjustment are well known in the art. Typically this involves scoring a conservative substitution as a partial rather than a full mismatch, thereby increasing the percentage sequence identity. Thus, for example, where an identical amino acid is given a score of 1 and a non-conservative substitution is given a score of zero, a conservative substitution is given a score

between zero and 1. The scoring of conservative substitutions is calculated, e.g., as implemented in the program PC/GENE (Intelligenetics, Mountain View, Calif.), BLAST (Jonhson, M et al. (2008) *Nucleic Acids Research*, 1, 36), Clustalw (EMBL-EBI), etc.

[00045] As used herein, “percentage of sequence identity” means the value determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison, and multiplying the result by 100 to yield the percentage of sequence identity. The sequences may be aligned over their entire length or, optionally, only a specific region of one or both sequences may be used to generate the alignment.

[00046] The term “entire length”, as used herein in the context of sequence identity, relates to the primary amino acid sequence of a given peptide ranging from the first amino acid at the N-terminus to the last amino acid at the C-terminus of said given peptide.

[00047] By the term “functional moiety”, as used herein, any chemical molecule or group is meant that can be attached to the N-terminal or C-terminal end, or to any other residue of a peptide including protecting groups, fluorescent or otherwise detectable groups, tags, and/or pharmaceutically active compounds.

[00048] “Pharmaceutically active compound”, as used herein, relates to a substance in a pharmaceutical drug or a pesticide that is biologically active. Such compounds encompass a broad chemical variety and relate to molecules of different classes according to Lipinski’s rule of five such as small molecules, proteins, nucleotides, lipids, sugars, and derivatives thereof.

[00049] The term “medicament”, as used herein, relates to any chemical substance formulated or compounded as single active ingredient or in combination with other pharmacologically active substances. The formulated or compounded composition may comprise, besides the pharmaceutically active compound, a pharmaceutically acceptable carrier.

[00050] “Tag”, as used herein, relates to a group of atoms or a molecule that is attached covalently to a nucleic acid sequence or another biological molecule for the purpose of detection by an appropriate detection system. The term “tagged peptide” refers to a peptide to which a tag has been covalently attached. The term “tag” and “label” may be used interchangeably.

[00051] The term “N-terminus” relates to the start of a protein or polypeptide, terminated by an amino acid with a free amine group (-NH₂).

[00052] The term “an N-terminal fragment” relates to a peptide or protein sequence which is in comparison to a reference peptide or protein sequence C-terminally truncated, such that a contiguous amino acid polymer starting from the N-terminus of the peptide or protein remains. In some embodiments, such fragments may have a length of at least 10, 20, 50, or 100 amino acids.

[00053] The term “C-terminus” relates to the end of an amino acid chain (protein or polypeptide), terminated by a free carboxyl group (-COOH).

[00054] The term “a C-terminal fragment” relates to a peptide or protein sequence which is in comparison to a reference peptide or protein sequence N-terminally truncated, such that a contiguous amino acid polymer starting from the C-terminus of the peptide or protein remains. In some embodiments, such fragments may have a length of at least 10, 20, 50, or 100 amino acids.

[00055] “At least one”, as used herein, relates to one or more, in particular 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more.

[00056] The term “G-quadruplex” refers to a four-stranded helical nucleic acid structure comprising multiple stacked G-tetrads, each of which consists of four guanine bases that associate in a cyclical manner through Hoogsteen hydrogen bonds and are further stabilized through coordination to a cation in the center. The body of stacked G-tetrads, comprising a total of 2–8 layers, is collectively referred to as the G-tetrad core. Each of the four guanine columns constituting the G-tetrad core can arise from a single (continuous column), two, or four (discontinuous column) separate guanine stretch/es. The term “multimolecular” as used herein refers to a G-tetrad core that is formed by at least two separate oligonucleotide strands, each of which comprises at least one G-rich segment or partial G-rich segment. The term “parallel-stranded G-quadruplex”, as used herein, relates to a G-quadruplex structure wherein all four

strands point in the same direction. Such parallel-stranded G-quadruplex is formed, for example by the nucleotide sequence set forth in SEQ ID NO:5.

[00057] The term "conjugate," as used herein, refers to a compound comprising two or more molecules (e.g., peptides, carbohydrates, small molecules, or nucleic acid molecules) that are chemically linked. The two or molecules desirably are chemically linked using any suitable chemical bond (e.g., covalent bond). Suitable chemical bonds are well known in the art and include disulfide bonds, acid labile bonds, photolabile bonds, peptidase labile bonds (e.g. peptide bonds), thioether, and esterase labile bonds.

[00058] The term "detection", as used herein, relates to quantitatively or qualitatively identifying an analyte (e.g., DNA, RNA or a protein) within a sample. In preferred embodiments of the invention, the analyte is a parallel-stranded G-quadruplex. The absence or presence of the parallel-stranded G-quadruplex is determined by binding of the peptide or conjugate of the present invention to this G-quadruplex. The formed complex of the peptide or conjugate of the present invention and the parallel-stranded G-quadruplex can be detected by several different techniques known in the art. Such techniques may include, but are not limited to immunoassay, mass spectrometry, chromatography, Western Blot, or gel electrophoresis.

[00059] In some embodiments, the immunoassay may be, but is not limited to an Enzyme-linked Immunosorbent Assay (ELISA), Western blot, agglutination test, biotin/avidin type assays, radioimmunoassays, immunoelectrophoresis and immunoprecipitation. The reactions generally include revealing labels such as fluorescent, chemiluminescent, radioactive, enzymatic labels or dye molecules, or other methods for detecting the formation of a complex between the antigen and the antibody or antibodies reacted therewith. These and further immunoassays are well known in the art (David Wild (Ed.): The Immunoassay Handbook. 3rd ed. Elsevier Science Publishing Company, Amsterdam 2005).

[00060] More particularly, an ELISA method can be used, wherein the wells of a microtiter plate are coated with an antibody against the protein to be tested. A biological sample containing or suspected of containing the G-quadruplex and the peptide/conjugate of the invention are then added to the coated wells. After a period of incubation sufficient to allow the formation of antibody-antigen complexes, the plate(s) can be washed to remove unbound

moieties and a detectably labelled secondary binding molecule added. The secondary binding molecule is allowed to react with any captured G-quadruplex-peptide/conjugate-complex, the plate washed and the presence of the secondary binding molecule detected using methods well known in the art.

[00061] In certain embodiments of the above detailed methods, if the determination is via mass spectrometry, the mass spectrometry may be selected from the group comprising MS measurements using EI, CI, ESI, APLI, APPI and APCI.

[00062] The analyte determination employing chromatography may be selected from the group comprising liquid chromatography, HPLC, FPLC, Smart chromatography, gel chromatography, size exclusion chromatography, reverse phase chromatography and ion-exchange chromatography (Introduction to Modern Liquid Chromatography, Lloyd R. Snyder, Wiley, 2009).

[00063] In various embodiments, if the analyte is detected via gel electrophoresis, the gel electrophoresis may be selected from the group, but not limited to agarose gel electrophoresis, sodium dodecyl sulfate poly acryl amide gel electrophoresis (SDS-PAGE), 2D-gel electrophoresis, native gel electrophoresis and quantitative preparative native continuous polyacrylamide gel electrophoresis (QPNC-PAGE).

[00064] Of course, in certain embodiments of the methods of the present invention at least two determination methods may be coupled to each other in a subsequent manner. In a variant, a gel electrophoresis may be followed by a mass spectroscopic analysis. Alternatively, a gel electrophoresis may be followed by a Western Blot, a chromatography may be followed by a mass spectroscopic analysis, and a chromatography may be followed by an immune assay, e.g. an ELISA.

[00065] Where in the methods detailed above an analyte is determined with the use of DNA labels, the DNA label may be determined by PCR, gel electrophoresis and/or Southern Blot.

[00066] The term "contacting", as used herein, refers generally to providing access of one component, reagent, analyte or sample to another. For example, contacting can involve mixing a solution comprising the peptide or conjugate of the invention with a sample comprising a G-

quadruplex. The solution comprising one component, reagent, analyte or sample may also comprise another component or reagent, which facilitates mixing, interaction, uptake, or other physical or chemical phenomenon advantageous to the contact between components, reagents, analytes and/or samples.

[00067] In various embodiments of the invention, the sample is a biological sample, for example a body fluid, cell or tissue sample. Body fluids comprise, but are not limited to blood, blood plasma, blood serum, breast milk, cerebrospinal fluid, cerumen (earwax), endolymph and perilymph, gastric juice, mucus (including nasal drainage and phlegm), peritoneal fluid, pleural fluid, saliva, sebum (skin oil), semen, sweat, tears, vaginal secretion, nipple aspirate fluid, vomit and urine. The cell or tissue sample may comprise material originated from any part of the body such as connective tissue, muscle tissue, nervous tissue, and epithelial tissue. The term “obtaining a sample”, as used herein, relates to different methods known in the art that comprise, but not limited to, biopsy, sentinel node biopsy or removal of blood, bone marrow, sputum or bronchial fluids.

EXAMPLES

Example 1: Rhau55 binds preferentially to parallel G-quadruplexes

[00068] Rhau55 was expressed in *Escherichia coli*. The binding of Rhau55 to different G-quadruplexes nucleic acids topology was first tested using gel electrophoresis.

[00069] Rhau55 binds to various parallel G-quadruplexes DNA and RNA (Figures 1 and 3-5) but not to various forms of non-parallel G-quadruplexes (Figures 1 and 6). This result was confirmed by NMR spectroscopy (Figure 7): adding Rhau55 in a mixture containing two well-defined intramolecular G-quadruplexes (a propeller parallel form (termed T95_2T) and a non-parallel 1 (termed HT1) only affect the peaks of the former form.

[00070] To confirm that the observed selective effect was based on the structure rather than sequence, the binding of Rhau55 was tested using the human telomeric sequences in K^+

containing solution, with and without crowded condition (40% v/v PEG200). In diluted solutions the human telomeric sequences Htelo1, Htelo2, Htelo3 and Htelo4 form a distinct and non-parallel G-quadruplex (Luu KN et al. (2006) J Am ChemSoc, 128(30), 9963-9970; Phan AT et al. (2006) Nucleic Acids Res, 34(19), 5715-5719; Lim KW et al. (2009) J Am ChemSoc, 131(12), 4301-430; Lim KW et al. (2009) Nucleic Acids Res, 37(18), 6239-6248), however in crowded conditions those sequences form a propeller-type parallel stranded G-quadruplex. The gel shift assay (Figure 6) showed that Rhau55 binds significantly only to a parallel-type G-quadruplex.

[00071] Subsequently, a quantitative analysis of the interaction between Rhau55 and three G-quadruplexes was performed using gel electrophoresis (Figure 4). The analysis indicated a dissociation constant of 1029 nM and 989 nM for the parallel-type form of Htelo1 (under crowded condition) and the single loop propeller-type T95_2T, respectively, consistent with previous published data for the full-length protein (Tran H et al. (2004) Mol Cell, 13, 101). In contrast an apparent binding of $>10 \mu\text{M}$ was found with the hybrid “3+1” form 1, and an estimated K_d over $10 \mu\text{M}$ was found for Htelo2, Htelo3 and Htelo4 non-parallel forming sequences, highlighting a net preference for the human telomeric sequences forming a parallel-type G-quadruplex. From the data little effect on three (Htelo1, parallel form) and single (T95_2T) loop size was observed on Rhau55 binding affinity.

Example 2: NMR analysis reveals a minimal binding domain

[00072] The interaction between T95_2T and Rhau55 was investigated. Backbone assignment of double labeled ^{15}N , ^{13}C Rhau55 free and in complex were recorded using a series of classical triple resonances experiments. 93% of backbone ^{15}N , ^1HN , $^1\text{H}\alpha$, $^{13}\text{C}\alpha$, and $^{13}\text{C}'$ assignments have been performed for both free and bound Rhau55. Analysis of chemical shift by Talos+ (Shen Y et al. (2009) J. Biomol. NMR, 44, 213-223) and CD spectrum (data not shown) showed a predominant helical structure for the free Rhau55. The bound form of Rhau55 was obtained with an equimolar ratio of Rhau55 and T95_2T (Figure 12A).

[00073] ^1H , ^{15}N -HSQC spectra of the free and bound Rhau55 are presented in Figure 9A and 9B, respectively. Chemical shift variation between the free and bound Rhau55 is shown in Figure 9C. Significant chemical shift perturbations are present only in the first 20 amino acids, and no significant changes are observed for the rest of the sequence, which strongly suggests that the DNA binding occurs in that region.

[00074] Inferring from the NMR chemical shift variation data, several combinations of synthetic peptides containing the N-terminal region of Rhau55 ranging from 5 to 29 amino acids (according to the numbering of the sequence of Rhau55) were tested with T95_2T using gel electrophoresis (Figure 10). These data show that the 16-amino-acid peptide Rhau16 constitutes the minimal domain required for the G-quadruplex binding. It should be noted that this domain contains a highly conserved 13-amino-acid sequence (PGHLKGREIGMWY) that was previously identified as a major determinant for the G-quadruplex binding (Lattmann, S. Nucleic Acids Res 2010, 38, 6219). However, this 13-amino-acid sequence is not sufficient for G-quadruplex binding, as the Rhau14 peptide containing this sequence does not bind G-quadruplexes (Figure 10, lane 6).

[00075] NMR imino proton spectra of T95_2T in complex with Rhau16, Rhau18sm and Rhau55 using equimolar ratio is presented in Figures 11. The NMR spectra of the complexes T95_2T-Rhau55 and T95_2T-Rhau18sm were very similar and occurred in the slow exchange regime, suggesting a similar binding mode, while the NMR spectrum of T95_2T-Rhau16 is different from that of Rhau55 and occurred in a different NMR time scale regime. Thus, the further analysis was drawn to the binding of Rhau18sm with G-quadruplex.

[00076] A similar binding assay was performed with Rhau55. Gel electrophoresis assay of Rhau18sm was performed with a parallel single propeller loop T95_2T, the hybrid "3+1" Htelo1, the parallel stranded G-quadruplex form of Htelo1 obtained under crowded condition, an interlocked dimeric parallel stranded G-quadruplex (termed 93del), the Dickerson B-DNA duplex (DX) and an stem-loop structure (ST).

[00077] The results showed (Figure 12) that Rhau18m binds to parallel-type stranded G-quadruplex (monomeric and dimeric) but not with the hybrid "3+1", the duplex DNA or the

stem-loop structure form. Rhau18sm behave similarly to that of Rhau55. Rhau18sm binds preferentially to parallel stranded G-quadruplex, with an affinity three order of magnitude. Then the interaction of Rhau18sm with T95_2T was studied using NMR.

Example 3: Structural basis for the discrimination of parallel G-quadruplexes by Rhau

[00078] The complex formed by Rhau18sm and T95_2T, using fully ^{15}N and ^{13}C labeled protein and unlabeled DNA showed the presence of 12 imino protons in the bound form which indicated the persistence of three-layered G-quadruplex structure upon protein binding (Figure 13A). Imino protons of T95_2T in the complex were assigned through chemical exchange between the free and bound state (Figure 13C). A parallel topology for the bound T95_2T was found using the NOE imino-aromatics cross-peaks of guanines (data not shown and Figure 13B). Intramolecular NOE cross-peaks for the bound form of Rhau18sm were assigned manually with 3D ^{15}N and ^{13}C NOESY-HSQC and by comparing spectra from Rhau18sm in the free form. Using 2D ^1H NOESY, 3D ^{15}N , ^{13}C HSQC-NOESY and 2D, 3D ^{15}N , ^{13}C -edited NOESY 42 intermolecular NOEs cross-peaks were identified (see example in Figure 12D) between Rhau18sm and T95_2T. Using constraints from previously published data on T95_2T (except restrained involving residues T1 and T2) intramolecular constraints of Rhau18sm and intermolecular, a model of Rhau18sm bound to T95_2T was generated (Figure 14). The top G-tetrad is almost entirely covered by the peptide. The appearance of such binding requires a high accessibility of said G-tetrad. In all non-parallel G-quadruplexes the top G-tetrad are not accessible and covered from the solvent by different types of loops, which may be the major structural determinant for the observed selectivity.

[00079] The invention has been described broadly and generically herein. Each of the narrower species and subgeneric groupings falling within the generic disclosure also form part of the invention. This includes the generic description of the invention with a proviso or negative limitation removing any subject-matter from the genus, regardless of whether or not the excised material is specifically recited herein. Other embodiments are within the following claims. In addition, where features or aspects of the invention are described in terms of Markush groups,

those skilled in the art will recognize that the invention is also thereby described in terms of any individual member or subgroup of members of the Markush group.

[00080] One skilled in the art would readily appreciate that the present invention is well adapted to carry out the objects and obtain the ends and advantages mentioned, as well as those inherent therein. Further, it will be readily apparent to one skilled in the art that varying substitutions and modifications may be made to the invention disclosed herein without departing from the scope and spirit of the invention. The compositions, methods, procedures, treatments, molecules and specific compounds described herein are presently representative of preferred embodiments are exemplary and are not intended as limitations on the scope of the invention. Changes therein and other uses will occur to those skilled in the art which are encompassed within the spirit of the invention are defined by the scope of the claims. The listing or discussion of a previously published document in this specification should not necessarily be taken as an acknowledgement that the document is part of the state of the art or is common general knowledge.

[00081] The invention illustratively described herein may suitably be practiced in the absence of any element or elements, limitation or limitations, not specifically disclosed herein. Thus, for example, the terms "comprising", "including," containing", etc. shall be read expansively and without limitation. The word "comprise" or variations such as "comprises" or "comprising" will accordingly be understood to imply the inclusion of a stated integer or groups of integers but not the exclusion of any other integer or group of integers. Additionally, the terms and expressions employed herein have been used as terms of description and not of limitation, and there is no intention in the use of such terms and expressions of excluding any equivalents of the features shown and described or portions thereof, but it is recognized that various modifications are possible within the scope of the invention claimed. Thus, it should be understood that although the present invention has been specifically disclosed by exemplary embodiments and optional features, modification and variation of the inventions embodied therein herein disclosed may be resorted to by those skilled in the art, and that such modifications and variations are considered to be within the scope of this invention.

[00082] The content of all documents and patent documents cited herein is incorporated by reference in their entirety.

CLAIMS

1. Peptide comprising or consisting of (i) the amino acid sequence of SEQ ID NO:1 or (ii) an amino acid sequence having at least 80% sequence identity to the amino acid sequence of SEQ ID NO:1 over its entire length.
2. The peptide according to claim 1, wherein said peptide is 16 to 50 amino acids in length.
3. The peptide according to claim 1 or 2, wherein said peptide is 18 to 50 amino acids in length.
4. Conjugate comprising a peptide according to any one of claims 1 to 3, wherein said peptide further comprises at least one functional moiety.
5. The conjugate according to claim 4, wherein said at least one functional moiety is conjugated to the N-terminus of said peptide.
6. The conjugate according to claim 4, wherein said at least one functional moiety is conjugated to the C-terminus of said peptide.
7. The conjugate according to claim 5, with the proviso that said at least one functional moiety does not comprise the amino acid sequence set forth in SEQ ID NO:2 or a C-terminal fragment thereof.
8. The conjugate according to claim 6, with the proviso that said at least one functional moiety does not comprise the amino acid sequence set forth in SEQ ID NO:3 or an N-terminal fragment thereof.
9. The conjugate according to any one of claims 4 to 8, wherein the functional moiety is a tag.
10. The conjugate according to any one of claims 4 to 8, wherein the functional moiety is a pharmaceutically active compound.

11. A peptide according to any one of claims 1 to 3, or a conjugate according to any one of claims 4 to 10 for use as a medicament.
12. A method for the detection of parallel-stranded G-quadruplexes in a sample wherein said method comprises:
 - contacting the peptide according to any one of claims 1 to 3 or the conjugate according to any one of claims 4 to 10 with a sample suspected to contain a parallel-stranded G-quadruplex; and
 - detecting the presence or absence of the parallel-stranded G-quadruplex in said sample.
13. Use of a peptide according to any one of claims 1 to 3 or a conjugate according to any one of claims 4 to 10 for the detection of parallel-stranded G-quadruplexes.