

University of Dayton

eCommons

---

Chemistry Faculty Publications

Department of Chemistry

---

11-6-2008

## Patent: Cell-Permeable Miniature Proteins

Alanna S. Schepartz

Douglas S. Daniels

Betsy Smith

Follow this and additional works at: [https://ecommons.udayton.edu/chm\\_fac\\_pub](https://ecommons.udayton.edu/chm_fac_pub)



Part of the [Other Chemistry Commons](#), and the [Physical Chemistry Commons](#)

---

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
6 November 2008 (06.11.2008)

PCT

(10) International Publication Number  
**WO 2008/133929 A2**

(51) **International Patent Classification:**  
**C07K 14/465 (2006.01) C07K 14/47 (2006.01)**

(74) **Agent:** CHEN, Tani; Wolf, Greenfield & Sacks, P.c., Federal Reserve Plaza, 600 Atlantic Avenue, Boston, MA 02210-2206 (US).

(21) **International Application Number:**  
**PCT/US2008/005264**

(81) **Designated States** (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW

(22) **International Filing Date:** 24 April 2008 (24.04.2008)

(25) **Filing Language:** English

(26) **Publication Language:** English

(30) **Priority Data:**  
60/926,379 24 April 2007 (24.04.2007) US  
61/009,905 3 January 2008 (03.01.2008) US  
61/011,311 16 January 2008 (16.01.2008) US  
61/068,259 5 March 2008 (05.03.2008) US

(84) **Designated States** (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(71) **Applicant** (for all designated States except US): **YALE UNIVERSITY** [US/US]; Two Whitney Avenue, New Haven, CT 06511 (US).

(72) **Inventors; and**

(75) **Inventors/Applicants** (for US only): **SCHEPARTZ, Alanna, S.** [US/US]; 4 Borglum Raod, Wilton, CT 06897 (US). **DANIELS, Douglas, S.** [US/US]; 745 Orange Street, Apt. 203, New Haven, CT 06511 (US). **SMITH, Betsy** [US/US]; 17 Clark Street, #4, New Haven, CT 06511 (US).

**Published:**

- without international search report and to be republished upon receipt of that report
- with sequence listing part of description published separately in electronic form and available upon request from the International Bureau

(54) **Title:** CELL PERMEABLE MINIATURE PROTEINS

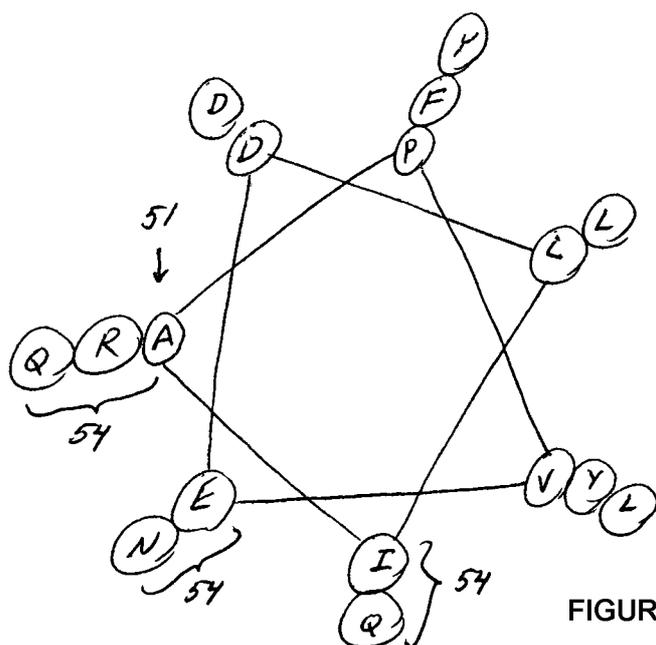


FIGURE ..2

(57) **Abstract:** The present invention generally relates to miniature proteins, including miniature proteins that are permeable to cells. Certain aspects of the invention are generally related to miniature proteins, such as avian pancreatic polypeptide (aPP), modified such that the miniature proteins are permeable to cells. For instance, a portion of the aPP, such as the alpha helix region and/or the type II polyproline helix region, may be modified to render the region substantially cationic. As an example, one or more residues may be substituted with cationic amino acid residues such as arginine. The miniature proteins may also have additional functions, such as the ability to bind to other proteins such as Bcl2 or hDM2. Another aspect of the invention is generally directed to sequences, such as PRR or PPR, that can be added to other proteins in order to increase their cell permeability. Still other aspects of the invention are generally directed to methods of making such proteins, methods of using such proteins, kits involving such proteins, and the like.

WO 2008/133929 A2

## CELL PERMEABLE MINIATURE PROTEINS

### GOVERNMENT FUNDING

5           Research leading to various aspects of the present invention were sponsored, at least in part, by the National Institutes of Health and the National Foundation for Cancer Research. The U.S. Government has certain rights in the invention.           

### RELATED APPLICATIONS

10           This application claims the benefit of U.S. Provisional Patent Application Serial No. 61/068,259, filed March 5, 2008, entitled "Cell Permeable Miniprotein Inhibitors of BCL2 Interactions," by A. Schepartz; U.S. Provisional Patent Application Serial No. 61/011,311, filed January 16, 2008, entitled "Cell-Permeable Miniature Proteins via a Helical Arginine Display," by A. Schepartz, *et al.*; U.S. Provisional Patent Application  
15           Serial No. 61/009,905, filed January 3, 2008, entitled "Intrinsically Cell-Permeable Miniature Proteins Based on a Minimal Cationic PPII Motif," by A. Schepartz, *et al.* and U.S. Provisional Patent Application Serial No. 60/926,379, filed April 24, 2007, entitled "Cell Permeable Miniature Proteins," by A. Schepartz Shrader, *et al.*, each incorporated herein by reference.

20

### FIELD OF INVENTION

          The present invention generally relates to miniature proteins, including miniature proteins that are permeable to cells.

25

### BACKGROUND

          Many proteins recognize nucleic acids, other proteins or macromolecular assemblies using a partially exposed alpha helix. Within the context of a native protein fold, such alpha helices are usually stabilized by extensive tertiary interactions with residues that may be distant in primary sequence from both the alpha helix and from each  
30           other. With some exceptions, removal of these tertiary interactions destabilizes the alpha helix and results in molecules that neither fold nor function in macromolecular recognition. The ability to recapitulate or perhaps even improve on the recognition properties of an alpha helix within the context of a small molecule should find utility in

the design of synthetic mimetics or inhibitors of protein function, or new tools for proteomics research.

Two fundamentally different approaches have been taken to bestow alpha helical structure on otherwise unstructured peptide sequences. One approach makes use of modified amino acids or surrogates that favor helix initiation or helix propagation. Some success has been realized by joining the *i* and *i*+7 positions of a peptide with a long-range disulfide bond to generate molecules whose helical structure was retained at higher temperatures. A second approach is to pare the extensive tertiary structure surrounding a given recognition sequence to generate the smallest possible molecule possessing function. This strategy has generated minimized versions of the Z domain of protein A (fifty-nine amino acids) and atrial natriuretic peptide (twenty-eight amino acids). The two minimized proteins, at thirty-three and fifteen amino acids, respectively, displayed relatively high biological activity. Despite this success, it is difficult to envision a simple and general application of this truncation strategy in the large number of cases where the alpha helical epitope is stabilized by residues scattered throughout the primary sequence.

In light of this limitation, a more flexible approach to protein minimization called protein grafting has been employed. Schematically, protein grafting involves removing residues required for molecular recognition from their native alpha helical context and grafting them on the scaffold provided by small yet stable proteins. Numerous researchers have engineered protein scaffolds to present binding residues on a relatively small peptide carrier. These scaffolds are small polypeptides onto which residues critical for binding to a selected target can be grafted. The grafted residues are arranged in particular positions such that the spatial arrangement of these residues mimics that which is found in the native protein. These scaffolding systems are commonly referred to as miniature proteins or miniproteins. A common feature is that the binding residues are known before the miniprotein is constructed.

Examples of these miniproteins include the thirty-seven amino acid protein charybdotoxin and the thirty-six amino acid protein, avian pancreatic peptide. Avian pancreatic polypeptide (aPP) is a polypeptide in which residues fourteen through thirty-two form an alpha helix stabilized by hydrophobic contacts with an N-terminal type II polyproline (PPII) helix formed by residues one through eight. Because of its small size and stability, aPP is a useful scaffold for protein grafting of alpha helical recognition epitopes.

### SUMMARY OF THE INVENTION

The present invention generally relates to miniature proteins, including miniature proteins that are permeable to cells. The subject matter of the present invention involves, in some cases, interrelated products, alternative solutions to a particular problem, and/or a plurality of different uses of one or more systems and/or articles.

Various aspects of the present invention are directed to compositions, including, in some cases, compositions directed to miniature proteins, for example PYY, aPP, YY2, YY3, YY1, or the like, which may be modified in various ways, as discussed herein. For instance, one set of embodiments of the present invention is generally directed to a cell-permeable miniature protein.

In another set of embodiments, the composition includes a miniature protein including an alpha helix region, where the miniature protein is modified by substitution of three to six amino acid residues, inclusively, with arginine residues. The substitutions may be present at sequential and/or non-sequential positions within the alpha helix region. In another set of embodiments, the composition includes a miniature protein including a type II polyproline helix region, where the miniature protein is modified by substitution of three to six amino acid residues, inclusively, with arginine residues. The substitutions may be present at sequential and/or non-sequential positions within the type II polyproline helix region.

The composition, in yet another set of embodiments, includes a PYY including a type II polyproline helix region. The PYY, in some embodiments, is modified by substitution of three to six amino acid residues, inclusively, with arginine residues. In certain cases, the PYY is modified by substitution of at least a portion of the type II polyproline helix region with the sequence  $(PPR)_n$ , where n is at least 2.

In one set of embodiments of the present invention, the composition comprises the sequence APPLPPNRGEDASPEELSRYYRSLRHLYLNLVTRQRY (SEQ ID NO: 50). In another set of embodiments of the present invention, the composition comprises the sequence APPLPPNRGEDASPRELSRYRSLRHLYLNLVTRQRY (SEQ ID NO: 51). In yet another set of embodiments of the present invention, the composition comprises the sequence APPLPPNRGEDASPRELRRYYRSLRHLYLNL VTRQRY (SEQ ID NO: 52).

According to another set of embodiments, the composition comprises a miniature protein including a type II polyproline helix region, where the miniature protein is modified by substitution of at least a portion of the type II polyproline helix region with the sequence  $(PPR)_n$ , where  $n$  is at least 4. The composition, in yet another set of  
5 embodiments, the composition comprises a miniature protein including a type II polyproline helix region, where the miniature protein is modified by substitution of at least a portion of the type II polyproline helix region with the sequence  $(PRR)_n$ , where  $n$  is at least 3.

In one set of embodiments, the composition includes a miniature protein  
10 modified by substitution of at least three amino acid residues with arginine residues, where the substitutions are chosen from the group consisting of positions 15, 18, 19, 22, 25, and 26 of the avian pancreatic polypeptide.

In another set of embodiments, the composition is directed to an avian pancreatic polypeptide (aPP) including an alpha helix region, where the avian pancreatic  
15 polypeptide is modified by substitution of three to six amino acid residues, inclusively, with arginine residues. The substitutions may be present at sequential and/or non-sequential positions within the alpha helix region.

The composition, in accordance with yet another set of embodiments, is directed to an avian pancreatic polypeptide (aPP) including a type II polyproline helix region,  
20 where the avian pancreatic polypeptide is modified by substitution of three to six amino acid residues, inclusively, with arginine residues. The substitutions may be present at sequential and/or non-sequential positions within the type II polyproline helix region.

The composition, according to one set of embodiments, includes an avian pancreatic polypeptide (aPP) including a type II polyproline helix region, where the  
25 avian pancreatic polypeptide is modified by substitution of at least a portion of the type II polyproline helix region with the sequence  $(PPR)_n$ , where  $n$  is at least 4. In another set of embodiments, the composition includes an avian pancreatic polypeptide (aPP) including a type II polyproline helix region, where the avian pancreatic polypeptide is modified by substitution of at least a portion of the type II polyproline helix region with  
30 the sequence  $(PRR)_n$ , where  $n$  is at least 3.

In accordance with yet another set of embodiments, the composition includes an avian pancreatic polypeptide modified by substitution of at least three amino acid

residues with arginine residues, where the substitutions are chosen from the group consisting of positions 15, 18, 19, 22, 25, and 26 of the avian pancreatic polypeptide.

In one set of embodiments of the present invention, the composition comprises the sequence GPSQPTYPGDDAPVRDLRRFYRDLRRYLNVVTRHRY (SEQ ID NO: 21). In another set of embodiments of the present invention, the composition comprises the sequence GPSQPTYPGDDAPVPJLIRFYRDLRRYLNVVTRHRY (SEQ ID NO: 22). In still another set of embodiments of the present invention, the composition comprises the sequence GPSQPTYPGDDAPVRDLIRFYRDLQRYLNVVTRHRY (SEQ ID NO: 23).

The composition, in still another set of embodiments, is generally directed to an avian pancreatic polypeptide (aPP) including an alpha helix region and/or a type II polyproline helix region that is substantially cationic. In yet another set of embodiments, the composition includes a miniature protein including an alpha helix region and/or a type II polyproline helix region that is substantially cationic. In one set of embodiments, the composition is directed to a cell-permeable avian pancreatic polypeptide.

In another set of embodiments, the composition is directed to a modified avian pancreatic polypeptide that, when exposed to HeLa cells at a concentration of 1 micromolar, equilibrates inside the HeLa cells at a concentration that is at least 75 times greater than the equilibration concentration of an unmodified avian pancreatic polypeptide exposed to the HeLa cells under the same conditions.

The composition, in still another set of embodiments, is directed to a miniature protein that, when exposed to a population of Jurkat cells at a concentration of 10 micromolar, induces apoptosis in at least about 50% of the Jurkat cells. In accordance with yet another set of embodiments, the composition includes a modified miniature protein that, when exposed to HeLa cells at a concentration of 1 micromolar, equilibrates inside the HeLa cells at a concentration that is at least 75 times greater than the equilibration concentration of an unmodified miniature protein exposed to the HeLa cells under the same conditions.

The composition, according to still another set of embodiments of the present invention, is directed to a protein selected from the group consisting of the following: GPRRPRRPRRGDDAPVEDLIRFYNDLQQYLNVVTRHRYC (SEQ ID NO: 24), GPRRPRRPRRPRRGDDAPVEDLIRFYNDLQQYLNVVTRHRYC (SEQ ID NO: 25), GPRRPRRPRRPRRPRRGDDAPVEDLIRFYNDLQQYLNVVTRHRYC (SEQ ID NO:

26), GPRRPRRPRRPRRPRRPRRGDDAPVEDLIRFYNDLQQYLNVVTRHRYC (SEQ ID NO: 27), GPPRPPRPPRPPRGDD APVEDLIRF YNDLQQYLNVVTRHRYC (SEQ ID NO: 28), GPPRPPRPPRPPRPPRGDD APVEDLIRF YNDLQQYLNVVTRHRYC (SEQ ID NO: 29), and

5 GPPRPPRPPRPPRPPRPPRGDDAPVEDLIRFYNDLQQYLNVVTRHRYC (SEQ ID NO: 30).

According to one set of embodiments, the composition includes a miniature protein that is cell permeable and able to bind to Bcl2 with a  $K_d$  of at least about 7 nM. In another set of embodiments, the composition includes a miniature protein that is cell permeable and able to bind to hDM2 with a  $K_d$  of at least about 10 nM.

In one set of embodiments of the present invention, the composition includes the sequence GPRRPRRPGRRAPVEDLIRFVGRLLAYFGDTINRYC (SEQ ID NO: 31). In another set of embodiments of the present invention, the composition includes the sequence GPRRPRRPGRRAPVVDLIRFRGRWLA YLGDTINRYC (SEQ ID NO: 32).  
15 In yet another set of embodiments of the present invention, the composition includes the sequence GPRRPRRPGRRAPLGDLISFRGRFLAYFGDTINRYC (SEQ ID NO: 33).

In still another set of embodiments of the present invention, the composition includes the sequence GPRRPRRPGRRAPVEDLIRFKFLLQWYLLALSLRNYAAAC (SEQ ID NO: 34). In another set of embodiments of the present invention, the composition includes the sequence  
20 GPRQPRYPGRDAPVEDLIRFKFLLQWYLLALSLRNYAAAC (SEQ ID NO: 35).

In yet another set of embodiments of the present invention, the composition includes the sequence GPSRPTRPGDRAPVEDLIRFKFLLQWYLLALSLRNYAAAC (SEQ ID NO: 36). In still another set of embodiments of the present invention, the composition includes the sequence  
25 GPRRPRRPGRRAPVEDLIRFKFLLQWFLALTRHRYAAAC (SEQ ID NO: 37). The composition, in yet another set of embodiments of the present invention, the composition includes the sequence GPRRPRRPGRRAPVEDLIRFKFLLQ WYLLAL YHNNYAAAC (SEQ ID NO: 38). The composition, in still another set of embodiments of the present invention, includes the sequence  
30 GPRRPRRPGRRAPVEDLIRFKFLLQWYLLALSHYNYAAAC (SEQ ID NO: 39).

The composition, in one set of embodiments of the present invention, includes the sequence  $G(\text{PRR})_n\text{PGRRAPVEDLIRFVGRLLAYFGDTINRYC}$  (SEQ ID NO: 40),



where n is at least 2. The composition, in another set of embodiments of the present invention, includes the sequence

G(PRR)<sub>n</sub>PGRRAPVVDLIRFRGRWLAYLGDTINRYC (SEQ ID NO: 41), where n is at least 2. The composition, in still another set of embodiments of the present invention,  
5 includes the sequence G(PRR)<sub>n</sub>PGRRAPLGDLISFRGRFLAYFGDTINRYC (SEQ ID NO: 42), where n is at least 2.

In another set of embodiments of the present invention, the composition includes the sequence G(PRR)<sub>n</sub>PGRRAPVEDLIRFKFLLQWYLLALSLRNYAAAC (SEQ ID NO: 43), where n is at least 2. The composition, in yet another set of embodiments of  
10 the present invention, includes the sequence

G(PRR)<sub>n</sub>PGRRAPVEDLIRFKFLLQWFLALTRHRYAAAC (SEQ ID NO: 44), where n is at least 2. The composition, in still another set of embodiments, includes the

sequence G(PRR)<sub>n</sub>PGRRAPVEDLIRFKFLLQWYLLALYHNNYAAAC (SEQ ID NO: 45), where n is at least 2. According to yet another set of embodiments, the composition

15 includes the sequence G(PRR)<sub>n</sub>PGRRAPVEDLIRFKFLLQWYLLALSHYNYAAAC (SEQ ID NO: 46), where n is at least 2.

In one set of embodiments, the composition includes an avian pancreatic polypeptide (aPP) including a type II polyproline helix region, where the avian pancreatic polypeptide is modified by substitution of three to six amino acid residues,  
20 inclusively, with arginine residues. The composition, according to another set of embodiments, includes an avian pancreatic polypeptide (aPP) including a type II polyproline helix region, where the avian pancreatic polypeptide is modified by substitution of at least a portion of the type II polyproline helix region with the sequence (PPR)<sub>n</sub>, wherein n is at least 2.

25 In one set of embodiments of the present invention, the composition includes the sequence RRPRRPRRPRRPGRRAPVEDLIRFYNDLQQYLNVVTRHRYC (SEQ ID NO: 47). The composition, in another set of embodiments of the present invention, includes the sequence RRPRRPRRPGRRAPVEDLIRFYNDLQQYLNVVTRHRYC (SEQ ID NO: 48). In still another set of embodiments of the present invention, the

30 composition includes the sequence GPRRPRRPGRRAPVEDLIRFYNDLQQYLNVVTRHRYC (SEQ ID NO: 49).

In one set of embodiments, the composition comprises a protein modified by the addition of (PPR)<sub>n</sub>, wherein n is at least 2, wherein the protein, after modification,

equilibrates inside HeLa cells at a concentration that is at least 75 times greater than the equilibration concentration of an unmodified protein exposed to the HeLa cells under the same conditions.

In another set of embodiments, the composition includes a protein modified by the addition of  $(PPR)_n$ , wherein n is at least 2, wherein the protein, after modification, equilibrates inside HeLa cells at a concentration that is at least 75 times greater than the equilibration concentration of an unmodified protein exposed to the HeLa cells under the same conditions.

The composition, in yet another set of embodiments, is directed to a protein containing the sequence  $(PPR)_n$ , where n is at least 2. In still another set of embodiments, the composition includes a protein containing the sequence  $(PRR)_n$ , where n is at least 2.

In another aspect, the present invention is directed to a method of making one or more of the embodiments described herein, for example, a cell permeable miniature protein. In another aspect, the present invention is directed to a method of using one or more of the embodiments described herein, for example, a cell permeable miniature protein.

Other advantages and novel features of the present invention will become apparent from the following detailed description of various non-limiting embodiments of the invention when considered in conjunction with the accompanying figures. In cases where the present specification and a document incorporated by reference include conflicting and/or inconsistent disclosure, the present specification shall control. If two or more documents incorporated by reference include conflicting and/or inconsistent disclosure with respect to each other, then the document having the later effective date shall control.

### **BRIEF DESCRIPTION OF THE DRAWINGS**

Non-limiting embodiments of the present invention will be described by way of example with reference to the accompanying figures, which are schematic and are not intended to be drawn to scale. In the figures, each identical or nearly identical component illustrated is typically represented by a single numeral. For purposes of clarity, not every component is labeled in every figure, nor is every component of each

embodiment of the invention shown where illustration is not necessary to allow those of ordinary skill in the art to understand the invention. In the figures:

Figs. 1A-1B illustrate avian pancreatic polypeptide;

Fig. 2 illustrates an end-on view of part of the alpha helix portion of aPP;

5 Figs. 3A-3C illustrate the cell permeability of certain cell-penetrating peptides, in one embodiment of the invention;

Fig. 3D illustrates the CD spectra of the peptides shown in Figs. 3A-3D;

Fig. 4A illustrates the sequences of certain cell permeable miniproteins and controls, in accordance with one embodiment of the invention;

10 Figs. 4B, D and E illustrates cell permeability of the same;

Fig. 4C illustrates the CD spectra of molecules shown in Fig 4A

Figs. 5A-5B illustrate cellular viability, in one embodiment of the invention;

Fig. 6 illustrates the temperature dependence of PPII helicity of certain peptides, according to another embodiment of the invention;

15 Fig. 7 illustrates the thermostability of RR3-aPP, according to one embodiment of the invention;

Figs. 8A-8C illustrate various miniature proteins according to certain embodiments of the invention;

20 Figs. 9A-9E illustrate time-dependent and temperature and azide-dependent uptake of certain miniature proteins and controls by HeLa cells, in accordance with one embodiment of the invention;

Figs. 10A-10B illustrate uptake of certain labeled miniature proteins and controls, in accordance with another embodiment of the invention;

25 Figs. 11A-11B illustrate thermostability of certain miniature proteins in yet another embodiment of the invention;

Figs. 12A-12C illustrate thermostability of certain miniature proteins in still another embodiment of the invention;

Fig. 13 illustrates the effect of certain peptides and miniature proteins on cell viabilities, in one embodiment of the invention; and

30 Fig. 14 illustrates binding of certain miniature proteins in another embodiment of the invention.

**BRIEF DESCRIPTION OF THE SEQUENCES**

SEQ ID NO: 1 is aPP, having the sequence

GPSQPTYPGDDAPVEDLIRFYNDLQQYLNVVTRHRY;

SEQ ID NO: 2 is PPBH3-1, having the sequence

5 GPSQPTYPGDDAPVEDLIRFVGRLAYFGDTINRY;

SEQ ID NO: 3 is PPBH3-1, having the sequence

GPSQPTYPGDDAPVEDLIRFVGRLAYFGDTINRYC;

SEQ ID NO: 4 is PPBH3-5, having the sequence

GPSQPTYPGDDAPVVDLIRFRGRWLAYLGDTINRY;

10 SEQ ID NO: 5 is PPBH3-5, having the sequence

GPSQPTYPGDDAPVVDLIRFRGRWLAYLGDTINRYC;

SEQ ID NO: 6 is PPBH3-6, having the sequence

GPSQPTYPGDDAPLGDLISFRGRFLAYFGDTINRY;

SEQ ID NO: 7 is PPBH3-6, having the sequence

15 GPSQPTYPGDDAPLGDLISFRGRFLAYFGDTINRYC;

SEQ ID NO: 8 is 1.1, having the sequence

GPSQPTYPGDDAPVEDLIRFKFLLQWYLLALTRHRYAAA;

SEQ ID NO: 9 is 1.1, having the sequence

GPSQPTYPGDDAPVEDLIRFKFLLQWYLLALTRHRYAAAC;

20 SEQ ID NO: 10 is 2.1, having the sequence

GKSWMTVPGDDAPVEDLIRFKFLLQWYLLALTRHRYAAA;

SEQ ID NO: 11 is 2.1, having the sequence

GKSWMTVPGDDAPVEDLIRFKFLLQWYLLALTRHRYAAAC;

SEQ ID NO: 12 is (PPR)<sub>3</sub>-aPP, having the sequence

25 GPPRPPRPPRGDDAPVEDLIRFYNDLQQYLNVVTRHRYC;

SEQ ID NO: 13 is 3.1, having the sequence

GPSQPTYPGDDAPVEDLIRFKFLLQWYLLALYHNNYAAA;

SEQ ID NO: 14 is 3.1, having the sequence

GPSQPTYPGDDAPVEDLIRFKFLLQWYLLALYHNNYAAAC;

30 SEQ ID NO: 15 is 3.2, having the sequence

GPSQPTYPGDDAPVEDLIRFKFLLQWYLLALSHYNYAAA;

SEQ ID NO: 16 is 3.2, having the sequence

GPSQPTYPGDDAPVEDLIRPKFLLQWYLLALSHYNYAAAC;

SEQ ID NO: 17 is 3.3, having the sequence

GPSQPTYPGDDAPVEDLIRFKFLLQWYLLALSLRNYAAA;

5 SEQ ID NO: 18 is 3.3, having the sequence

GPSQPTYPGDDAPVEDLIRFKFLLQWYLLALSLRNYAAAC;

SEQ ID NO: 19 is 3.4, having the sequence

GPSQPTYPGDDAPVEDLIRFKFLLQWYLLALAADAYAAA;

SEQ ID NO: 20 is 3.4, having the sequence

10 GPSQPTYPGDDAPVEDLIRFKFLLQWYLLALAADAYAAAC;

SEQ ID NO: 21 is aPP<sup>6R1</sup>, having the sequence

GPSQPTYPGDDAPVRDLRRFYRDLRRYLNVVTRHRY;

SEQ ID NO: 22 is aPP<sup>5R1</sup>, having the sequence

GPSQPTYPGDDAPVRDLIRFYRDLRRYLNVVTRHRY;

15 SEQ ID NO: 23 is aPP<sup>4R1</sup>, having the sequence

GPSQPTYPGDDAPVRDLIRFYRDLQRYLNVVTRHRY;

SEQ ID NO: 24 is (PRR)<sub>3</sub>-aPP, having the sequence

GPRRPRRPRRGDDAPVEDLIRFYNDLQQYLNVVTRHRYC;

SEQ ID NO: 25 is (PRR)<sub>4</sub>-aPP, having the sequence

20 GPRRPRRPRRPRRGDDAPVEDLIRFYNDLQQYLNVVTRHRYC;

SEQ ID NO: 26 is (PRR)<sub>5</sub>-aPP, having the sequence

GPRRPRRPRRPRRPRRGDDAPVEDLIRFYNDLQQYLNVVTRHRYC;

SEQ ID NO: 27 is (PRR)<sub>6</sub>-aPP, having the sequence

GPRRPRRPRRPRRPRRPRRGDDAPVEDLIRFYNDLQQYLNVVTRHRYC;

25 SEQ ID NO: 28 is (PPR)<sub>4</sub>-aPP, having the sequence

GPPRPPRPPRPPRGDDAPVEDLIRFYNDLQQYLNVVTRHRYC;

SEQ ID NO: 29 is (PPR)<sub>5</sub>-aPP, having the sequence

GPPRPPRPPRPPRPPRGDDAPVEDLIRFYNDLQQYLNVVTRHRYC;

SEQ ID NO: 30 is (PPR)<sub>6</sub>-aPP, having the sequence

30 GPPRPPRPPRPPRPPRPPRGDDAPVEDLIRFYNDLQQYLNVVTRHRYC;

SEQ ID NO: 31 is CP-BcIXL-1, having the sequence  
GPRRPRRPGRRAPVEDLIRFVGRLAYFGDTINRYC;

SEQ ID NO: 32 is CP-Bcl2-1, having the sequence  
GPRRPRRPGRRAPVVDLIRFRGRWLAYLGDTINRYC;

5 SEQ ID NO: 33 is CP-Bcl2-2, having the sequence  
GPRRPRRPGRRAPLGDLISFRGRFLAYFGDTINRYC;

SEQ ID NO: 34 is CP-hDM2-1, having the sequence  
GPRRPRRPGRRAPVEDLIRFKFLLQWYLLALSLRNYAAAC;

10 SEQ ID NO: 35 is CP-hDM2-2, having the sequence  
GPRQPRYPGRDAPVEDLIRFKFLLQWYLLALSLRNYAAAC;

SEQ ID NO: 36 is CP-hDM2-3, having the sequence  
GPSRPTRPGRDAPVEDLIRFKFLLQWYLLALSLRNYAAAC;

SEQ ID NO: 37 is CP-hDM2-4, having the sequence  
GPRRPRRPGRRAPVEDLIRFKFLLQWYLLALTRHRYAAAC;

15 SEQ ID NO: 38 is CP-hDM2-5, having the sequence  
GPRRPRRPGRRAPVEDLIRFKFLLQWYLLALYHNNYAAAC;

SEQ ID NO: 39 is CP-hDM2-6, having the sequence  
GPRRPRRPGRRAPVEDLIRFKFLLQWYLLALSHYNYAAAC;

20 SEQ ID NO: 40 is CP-BcIXL-1, having the sequence  
G(PRR)<sub>n</sub>PGRRAPVEDLIRFVGRLAYFGDTINRYC;

SEQ ID NO: 41 is CP-Bcl2-1, having the sequence  
G(PRR)<sub>n</sub>PGRRAPVVDLIRFRGRWLAYLGDTINRYC;

SEQ ID NO: 42 is CP-Bcl2-2, having the sequence  
G(PRR)<sub>n</sub>PGRRAPLGDLISFRGRFLAYFGDTINRYC;

25 SEQ ID NO: 43 is CP-hDM2-1, having the sequence  
G(PRR)<sub>n</sub>PGRRAPVEDLIRFKFLLQWYLLALSLRNYAAAC;

SEQ ID NO: 44 is CP-hDM2-4, having the sequence  
G(PRR)<sub>n</sub>PGRRAPVEDLIRFKFLLQWYLLALTRHRYAAAC;

30 SEQ ID NO: 45 is CP-hDM2-5, having the sequence  
G(PRR)<sub>n</sub>PGRRAPVEDLIRFKFLLQWYLLALYHNNYAAAC;

SEQ ID NO: 46 is CP-hDM2-6, having the sequence  
G(PRR)<sub>n</sub>PGRRAPVEDLIRFKFLLQWYLLALSHYNYAAAC;

SEQ ID NO: 47 is RR5-aPP, having the sequence  
RRPRRPRRPRRPGRRAPVEDLIRFYNDLQQYLNVVTRHRYC;

5 SEQ ID NO: 48 is RR4-aPP, having the sequence  
RRPRRPRRPGRRAPVEDLIRFYNDLQQYLNVVTRHRYC;

SEQ ID NO: 49 is RR3-aPP, having the sequence  
GPRRPRRPGRRAPVEDLIRFYNDLQQYLNVVTRHRYC;

10 SEQ ID NO: 50 is YY2<sup>3R1</sup>, having the sequence  
APPLPPNRGEDASPEELSRYYRSLRHYNLVTRQRY;

SEQ ID NO: 51 is YY2<sup>4R1</sup>, having the sequence  
APPLPPNRGEDASPRELSRYYRSLRHYNLVTRQRY;

SEQ ID NO: 52 is YY2<sup>5R1</sup>, having the sequence  
APPLPPNRGEDASPRELRRYYRSLRHYNLVTRQRY;

15 SEQ ID NO: 53 is PYY, having the sequence  
YPAKPEAPGEDASPEELSRYYASLRHYNLVTRQRY;

SEQ ID NO: 54 is GPSQPTYPGDDAPVRDLRRFYRDLQYLNVVTRHRY;

SEQ ID NO: 55 is GPSQPTYPGDDAPVRDLRRFYNDLRQYLNVVTRHRY;

SEQ ID NO: 56 is GPSQPTYPGDDAPVRDLRRFYNDLQRYLNVVTRHRY;

20 SEQ ID NO: 57 is GPSQPTYPGDDAPVRDLIRFYRDLRQYLNVVTRHRY;

SEQ ID NO: 58 is GPSQPTYPGDDAPVRDLIRFYNDLRRYLNVVTRHRY;

SEQ ID NO: 59 is GPSQPTYPGDDAPVEDLRRFYRDLRQYLNVVTRHRY;

SEQ ID NO: 60 is GPSQPTYPGDDAPVEDLRRFYRDLQRYLNVVTRHRY;

SEQ ID NO: 61 is GPSQPTYPGDDAPVEDLRRFYNDLRRYLNVVTRHRY;

25 SEQ ID NO: 62 is GPSQPTYPGDDAPVEDLIRFYRDLRRYLNVVTRHRY;

SEQ ID NO: 63 is GPSQPTYPGDDAPVRDLRRFYRDLRQYLNVVTRHRY;

SEQ ID NO: 64 is GPSQPTYPGDDAPVRDLRRFYRDLQRYLNVVTRHRY;

SEQ ID NO: 65 is GPSQPTYPGDDAPVRDLRRFYNDLRRYLNVVTRHRY;

SEQ ID NO: 66 is GPSQPTYPGDDAPVEDLRRFYRDLRRYLNVVTRHRY;

30 SEQ ID NO: 67 is YY2, having the sequence  
APPLPPNRGEDASPEELSRYYASLRHYNLVTRQRY;

SEQ ID NO: 68 is (PRR)<sub>3</sub>, having the sequence PRRPRRPRR;

SEQ ID NO: 69 is (PRR)<sub>4</sub>, having the sequence PRRPRRPRRPRR;

SEQ ID NO: 70 is (PRR)<sub>5</sub>, having the sequence PRRPRRPRRPRRPRR;

SEQ ID NO: 71 is (PRR)<sub>6</sub>, having the sequence PRRPRRPRRPRRPRRPRR;

5 SEQ ID NO: 72 is R<sub>8</sub>, having the sequence RRRRRRRR;

SEQ ID NO: 73 is R<sub>12</sub>, having the sequence RRRRRRRRRR;

SEQ ID NO: 74 is R<sub>10</sub>, having the sequence RRRRRRRRRRRR;

SEQ ID NO: 75 is R<sub>15</sub>, having the sequence RRRRRRRRRRRRRR;

SEQ ID NO: 76 is SAP or Flu-SAP, having the sequence Flu-

10 VRLPPPVRLLPPPVRLLPPP;

SEQ ID NO: 77 is Bac or Flu-Bac, having the sequence Flu-GPRPLPFPRPG;

SEQ ID NO: 78 is Tat or Flu-Tat, having the sequence Flu-GRKKRRQRRRPPQ;

SEQ ID NO: 79 is Flu-R<sub>8</sub>, having the sequence Flu-GRRRRRRRRR;

SEQ ID NO: 80 is Flu-Ri<sub>0</sub>, having the sequence Flu-GRRRRRRRRRRR;

15 SEQ ID NO: 81 is Flu-R<sub>12</sub>, having the sequence Flu-GRRRRRRRRRRRRR;

SEQ ID NO: 82 is Flu-PPR<sub>3</sub>, having the sequence Flu-GPPRPPRPPR;

SEQ ID NO: 83 is Flu-PPR<sub>4</sub>, having the sequence Flu-GPPRPPRPPRPPR;

SEQ ID NO: 84 is Flu-PPR<sub>6</sub>, having the sequence Flu-

GPPRPPRPPRPPRPPRPPR;

20 SEQ ID NO: 85 is Flu-PPR<sub>3</sub> or Flu-(PRR)<sub>3</sub>, having the sequence Flu-GPRRPRRPRR;

SEQ ID NO: 86 is Flu-PPR<sub>4</sub> or Flu-(PRR)<sub>4</sub>, having the sequence Flu-GPRRPRRPRRPRR;

25 SEQ ID NO: 87 is Flu-PPR<sub>5</sub> or Flu-(PRR)<sub>5</sub>, having the sequence Flu-GPRRPRRPRRPRRPRR;

SEQ ID NO: 88 is Flu-PPR<sub>6</sub> or Flu-(PRR)<sub>6</sub>, having the sequence Flu-GPRRPRRPRRPRRPRRPRR;

SEQ ID NO: 89 is aPP-Flu, having the sequence

GPSQPTYPGDDAPVEDLIRF YNDLQQYLNVTTRHRYC-FIU;

30 SEQ ID NO: 90 is RR<sub>3</sub>-aPP-Flu, having the sequence

GPRRPRRPGRRAPVEDLIRF YNDLQQYLNVTTRHRYC-FIU;



SEQ ID NO: 91 is RR<sup>a</sup>PP-Flu, having the sequence  
RRPRRPRRPGRRAPVEDLIRFYNDLQQYLN<sup>V</sup>VTRHRYC-FIU;

SEQ ID NO: 92 is RR<sub>5</sub>-aPP-Flu, having the sequence  
RRPRRPRRPRRPGRRAPVEDLIRFYNDLQQYLN<sup>V</sup>VTRHRYC-FIU;

5 SEQ ID NO: 93 is PPR<sub>3</sub>Y or (PPR)<sub>3</sub>Y, having the sequence PPRPPRPPRY;

SEQ ID NO: 94 is PPR<sub>4</sub>Y or (PPR)<sub>4</sub>Y, having the sequence PPRPPRPPRPPRY;

SEQ ID NO: 95 is PPR<sub>5</sub>Y or (PPR)<sub>5</sub>Y, having the sequence  
PPRPPRPPRPPRPPRY;

10 SEQ ID NO: 96 is PPR<sub>6</sub>Y or (PPR)<sub>6</sub>Y, having the sequence  
PPRPPRPPRPPRPPRPPRY;

SEQ ID NO: 97 is PRR<sub>3</sub>Y or (PRR)<sub>3</sub>Y, having the sequence PRRPRRPRRY;

SEQ ID NO: 98 is PRR<sub>4</sub>Y or (PRR)<sub>4</sub>Y, having the sequence  
PRRPRRPRRPRRY;

15 SEQ ID NO: 99 is PRR<sub>5</sub>Y or (PRR)<sub>5</sub>Y, having the sequence  
PRRPRRPRRPRRPRRY;

SEQ ID NO: 100 is PRR<sub>6</sub>Y or (PRR)<sub>6</sub>Y, having the sequence  
PRRPRRPRRPRRPRRPRRY;

SEQ ID NO: 101 is aPP-Ac, having the sequence  
GPSQPTYPGDDAPVEDLIRFYNDLQQYLN<sup>V</sup>VTRHRYC-Ac;

20 SEQ ID NO: 102 is RR<sub>3</sub>-aPP-Ac, having the sequence  
GPRRPRRPGRRAPVEDLIRFYNDLQQYLN<sup>V</sup>VTRHRYC-Ac;

SEQ ID NO: 103 is RR<sup>a</sup>aPP-Ac, having the sequence  
RRPRRPRRPGRRAPVEDLIRFYNDLQQYLN<sup>V</sup>VTRHRYC-Ac;

25 SEQ ID NO: 104 is RR<sub>5</sub>-aPP-Ac, having the sequence  
RRPRRPRRPRRPGRRAPVEDLIRFYNDLQQYLN<sup>V</sup>VTRHRYCAc;

SEQ ID NO: 105 is aPP<sup>4RI</sup>-Flu, having the sequence  
GPSQPTYPGDDAPVRDLIRFYRDLQRYLN<sup>V</sup>VTRHRY-Flu;

SEQ ID NO: 106 is aPP<sup>5RI</sup>-Flu, having the sequence  
GPSQPTYPGDDAPVRDLIRFYRDLRRYLN<sup>V</sup>VTRHRY-FIU;

30 SEQ ID NO: 107 is aPP<sup>6RI</sup>-Flu, having the sequence  
GPSQPTYPGDDAPVRDLRRFYRDLRRYLN<sup>V</sup>VTRHRY-Flu;

SEQ ID NO: 108 is YY2-Flu, having the sequence  
APPLPPNRGEDASPEELSRYYASLRH YLNLVTRQRY-Flu;

SEQ ID NO: 109 is YY2<sup>3R1</sup>-Flu, having the sequence  
APPLPPNRGEDASPEELSRYYRSLRH YLNLVTRQRY-Flu;

5 SEQ ID NO: 110 is YY2<sup>4R1</sup>-Flu, having the sequence  
APPLPPNRGEDASPRELSRYRSLRH YLNL VTRQRY-Flu;

SEQ ID NO: 111 is YY2<sup>5R1</sup>-Flu, having the sequence  
APPLPPNRGEDASPRELRRYYRSLRH YLNL VTRQRY-Flu;

SEQ ID NO: 112 is R<sub>8</sub>Y-Flu, having the sequence RRRRRRRRY-Flu;

10 SEQ ID NO: 113 is Tat (48-60)-Flu, having the sequence  
GRKKRRQRRRPPQY-Flu;

SEQ ID NO: 114 is Penetratin-Flu, having the sequence  
RQIKIWFQNRRMKWKK-Flu;

15 SEQ ID NO: 115 is Transportan-Flu, having the sequence  
GWTLNSAGYLLKINLKALAALAKKIL-Flu;

SEQ ID NO: 116 is KLA-Flu, having the sequence  
KLALKLALKALKAAALKLA-Flu;

SEQ ID NO: 117 is R<sub>8</sub>Y, having the sequence RRRRRRRRY;

SEQ ID NO: 118 is Tat (48-60), having the sequence GRKKRRQRRRPPQY;

20 SEQ ID NO: 119 is Penetratin, having the sequence RQIKIWFQNRRMKWKK;

SEQ ID NO: 120 is Transportan, having the sequence  
GWTLNSAGYLLKINLKALAALAKKIL;

SEQ ID NO: 121 is KLA, having the sequence KLALKLALKALK AALKLA;

SEQ ID NO: 122 is aPP, having the sequence

25 GPSQPTYPGDDAPVEDLIRFYNDLQQYLN VVTRHRYC;

SEQ ID NO: 123 is YY1, having the sequence  
APPLPPNRPGEDASPEELSRYYASLRH YLNLVTRQRY;

SEQ ID NO: 124 is YY3, having the sequence  
APPLPPNRPGEDASPEELSRYYASLRH YLNLVTRQRY.

## DETAILED DESCRIPTION

The present invention generally relates to miniature proteins, including miniature proteins that are permeable to cells. Certain aspects of the invention are generally related to miniature proteins, such as avian pancreatic polypeptide (aPP), modified such that the miniature proteins are permeable to cells. For instance, a portion of the aPP, such as the alpha helix region and/or the type II polyproline helix region, may be modified to render the region substantially cationic. As an example, one or more residues may be substituted with cationic amino acid residues such as arginine. The miniature proteins may also have additional functions, such as the ability to bind to other proteins such as Bcl2 or hDM2. Another aspect of the invention is generally directed to sequences, such as PRR or PPR, that can be added to other proteins in order to increase their cell permeability. Still other aspects of the invention are generally directed to methods of making such proteins, methods of using such proteins, kits involving such proteins, and the like.

Various aspects of the invention are generally directed to various miniature proteins, such as aPP or PYY (modified pancreatic peptide YY), that has been modified such that the miniature proteins are permeable to cells. For example, the miniature proteins may be modified at one or more regions in a manner that causes the regions to become substantially cationic. One or more residues on a miniature protein may be substituted with, for example, cationic amino acid residues. Non-limiting examples of cationic amino acid residues include arginine or histidine. As used herein, the terms "miniature protein" or "miniprotein" refer to a relatively small protein containing at least a protein scaffold and one or more additional domains or regions that help to stabilize its tertiary structure. In some cases, the miniature protein may have a length of no more than 40 or 45 residues. For instance, in various embodiments, the miniature protein may have a length of 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, or 40 residues.

In some cases, a modified miniature protein may be identified as one that is cell permeable by exposing a cell, such as a HeLa cell, to a modified miniature protein and comparing the concentration of the modified miniature protein within the cell after equilibration (e.g., when a steady-state concentration is reached) to the concentration of an unmodified miniature protein exposed to the cell under the same conditions. A cell permeable miniature protein or other peptide has a greater ability to enter the cell, e.g., it can enter the cell at a greater concentration than an unmodified protein under the same

conditions. In some cases, the concentration of the modified miniature protein within the cell is greater than outside of the cell at equilibrium, for example, at least about 50 times greater than the concentration of the unmodified miniature protein. In some cases, the cell permeable miniature protein may reach higher concentrations, e.g., at least about 75,  
5 at least about 100, at least about 300, etc., times greater than the concentration of the unmodified miniature protein.

For example, for a modified aPP as used herein, cells such as a HeLa cells may be exposed to a concentration of 1 micromolar of miniature protein (modified and unmodified), and the concentrations of each within the cell may be determined in some  
10 fashion. As a specific example, the miniature proteins may be labeled with a fluorescent entity, such as fluorescein, and their relative concentrations determined using techniques such as flow cytometry, using routine techniques known to those of ordinary skill in the art. A non-limiting example of a flow cytometry technique is FACS (fluorescence-activated cell sorting).

15 The term "protein," as used herein, is given its ordinary definition in the art, e.g., any of a group of complex organic compounds which contain carbon, hydrogen, oxygen, nitrogen and usually sulfur, which are widely distributed in plants and animals. Twenty different amino acids are commonly found in proteins and each protein has a unique, genetically defined amino acid sequence which determines its specific shape and  
20 function. The term "protein" is generally used herein interchangeably with the terms "peptide" and "polypeptide." The term "protein scaffold" refers to a region or domain of a relatively small protein, such as a miniature protein, that has a conserved tertiary structural motif which can be modified to display one or more specific amino acid residues in a fixed conformation.

25 Non-limiting examples of miniature proteins include the PP fold protein scaffolds, which generally contain thirty-six amino acids and are the smallest known globular proteins. Despite their small size, PP fold proteins are stable and remain folded under physiological conditions. Some PP fold protein scaffolds of the invention comprise two anti-parallel helices, an N-terminal type II polyproline helix (PPII)  
30 between amino acid residues two and eight and an alpha helix between residues 14 and 31 and/or 32. The stability of the PP fold protein scaffolds of the invention derives predominantly from interactions between hydrophobic residues on the interior face of the

alpha helix at positions 17, 20, 24, 27, 28, 30, and 31 and the residues on the two edges of the polyproline helix at positions 2, 4, 5, 7, and 8.

Positions for grafting these binding site residues on the protein scaffold include, but are not limited to, positions on the solvent-exposed alpha-helical face of aPP.

5 Substitutions of binding site residues may be made, in some cases, for residues involved in stabilizing the tertiary structure of the miniature protein. As used herein, the term "exposed on the alpha helix domain" means that an amino acid substituted, for example, into aPP or PYY, is available for association or interaction with another molecule and is not otherwise bound to or associated with another amino acid residue on the aPP or  
10 PYY. This term is used interchangeably with the term "solvent-exposed alpha helical face" and similar terms (e.g., "solvent-exposed face of the aPP alpha helix").

Members of the PP fold family of protein scaffolds which are contemplated for use in the present invention include, but are not limited to, avian pancreatic polypeptide (aPP), Neuropeptide Y, lower intestinal hormone polypeptide and pancreatic peptide YY  
15 (PYY) (e.g., SEQ ID NO: 53). In one embodiment, the protein scaffold comprises the PP fold protein, avian pancreatic polypeptide (SEQ ID NO: 1 or 122). aPP is a PP fold polypeptide characterized by a short (eight residue) amino-terminal type II polyproline helix linked through a type I beta turn to an eighteen residue alpha helix. Because of its small size and stability, aPP is an excellent protein scaffold for, e.g., protein grafting of  
20 alpha-helical recognition epitopes.

It should be noted that the natural sequence of aPP is as shown in SEQ ID NO: 1, i.e., GPSQPTYPGDDAPVEDLIRFYNDLQQYLNVVTRHRY. However, in some cases, the aPP may be modified by the addition of a terminal C, as is shown in SEQ ID NO: 122, i.e., GPSQPTYPGDDAPVEDLIRFYNDLQQYLNVVTRHRYC. The C may  
25 be added to the end of aPP to facilitate the binding of other species to aPP, for example, a reporter species such as fluorescein, as is shown in SEQ ID NO: 89 for aPP-Flu. Similarly, any other miniature protein or peptide discussed herein may be modified by the addition of a terminal C or a terminal Y (if one is not already present), e.g., to facilitate binding, for example, PPBH3-1 (see SEQ ID NOs 2 and 3), (PRR)<sub>3</sub>-aPP (SEQ  
30 ID NO: 24), CP-BcIXL-1 (SEQ ID NO: 31), CP-hDM2-1 (SEQ ID NO: 34), etc.

One aspect of the invention is generally directed to miniature proteins, such as avian pancreatic polypeptide (aPP), that include an alpha helix region and/or a type II polyproline helix region that is modified by the substitution and/or addition of various

amino acid residues with cationic amino acid residues such as arginine or histidine. The modified aPP may also include other substitutions, e.g., for improving binding properties, such as is discussed below. Referring now to Fig. IA, the structure of avian pancreatic polypeptide (aPP) (SEQ ID NO: 1) is shown, including a type II polyproline helix region ("PPII helix"), a beta turn, and an alpha-helix region. Fig. IB illustrates the structure of the aPP when it is properly folded. However, it should be understood that the miniature proteins discussed herein are not limited to aPP (SEQ ID NO: 1), but include other suitable miniature proteins. A non-limiting example includes PYY (SEQ ID NO: 53).

Without wishing to bound by any theory, it is believed that the substitutions may be chosen to cause the alpha helix region and/or type II polyproline helix region to become substantially cationic, i.e., such that the alpha helix region and/or the type II polyproline helix region exhibits a net positive charge in a buffered aqueous solution having neutral pH, such as phosphate-buffered saline. In some cases, the substitutions may be chosen to cause only a portion of the alpha helix region and/or type II polyproline helix region to become substantially cationic, and it is believed that such substitutions may cause the miniature protein to become relatively cell permeable. Such cationic properties can be determined, for instance, by using charge electrophoresis or isoelectric focusing to determine the charge of the protein (or portions of the protein) under various conditions.

Any number of residues within the alpha helix region and/or the type II polyproline helix region may be substituted with arginine residues or the like, for instance, at least two residues, at least three residues, at least four residues, etc. In some cases, there may be between 3 and 6 or between 3 and 5 residues (inclusively) that are substituted with a cationic amino acid residue. The substitutions may be present on any location within the alpha helix region and/or the type II polyproline helix region, and may be consecutive or non-consecutive, in some cases. In some embodiments, the cationic amino acid residue substitutions are located in only one region of the miniature protein, e.g., within only the alpha helix region or within only the type II polyproline helix region (of course, in some cases, there may be other types of substitutions elsewhere, e.g., for altering binding properties of the molecule).

Thus, one set of embodiments is directed to a miniature protein, such as aPP, that includes a type II polyproline helix region modified by the substitution and/or addition of various amino acid residues with cationic amino acid residues, and optionally with other

substitutions as well. The residues chosen for substitution may be present anywhere within the type II polyproline helix region, and may be independently consecutive or non-consecutive.

In some cases, the residues chosen for substitution are chosen such that the type II polyproline helix is enriched in proline and/or arginine residues. For instance, residues may be substituted (and/or new residues may be added to the type II polyproline helix) such that one or more proline and/or arginine residues are located next to each other, and in some cases, such that one or more repeat units of PPR and/or PRR are present within the type II polyproline helix. For instance, in one set of embodiments, the type II polyproline helix may include at least one portion having a sequence PPR and/or PRR, and in some cases, the type II polyproline helix may include more than one PPR and/or PRR sequence, e.g.,  $(PPR)_n$  and/or  $(PRR)_n$ , where n is at least 3, e.g., 3, 4, 5, 6, etc. In certain cases, the type II polyproline helix may contain other residues as well, and in some cases, the residues may be located between or even within PPR and/or PRR repeat units within the type II polyproline helix. Specific, non-limiting examples of such sequences include SEQ ID NO: 47 is RR5-aPP (SEQ ID NO: 47), SEQ ID NO: 48 is RR4-aPP (SEQ ID NO: 48), or SEQ ID NO: 49 is RR3-aPP (SEQ ID NO: 49), which each contain varying numbers of the PRR repeat unit within the type II polyproline helix of the miniature protein. Other non-limiting examples include SEQ ID NOs: 12 and 24-30. It should also be noted, as discussed below, that PPR and/or PRR repeat units may be added to any suitable protein in order to increase its cell permeability.

Still other non-limiting examples include the modification of avian pancreatic polypeptide able to bind to (or otherwise interact with) other proteins such as Bcl2 or hDM2. It should be noted that such binding may be specific or non-specific, and involve various noncovalent interactions such as including hydrogen bonding, metal coordination, hydrophobic forces, van der Waals forces, pi-pi interactions, and/or electrostatic effects. For instance, miniature proteins such as PPBH3-1 (e.g., SEQ ID NO: 2), PPBH3-5 (e.g., SEQ ID NO: 4), or PPBH3-6 (e.g., SEQ ID NO: 6), 1.1 (e.g., SEQ ID NO: 9), 2.1 (SEQ ID NO: 11), 3.1 (SEQ ID NO: 14), 3.2 (SEQ ID NO: 16), 3.3 (SEQ ID NO: 18), 3.4 (SEQ ID NO: 20), etc. can be modified to include any number of PPR and/or PRR sequences, e.g., as shown in the sequences  
G(PRR)<sub>n</sub>PGRRAPVEDLIRFVGRLLAYFGDT<sub>r</sub>NRYC (SEQ IDNO: 40),  
G(PRR)<sub>n</sub>PGRRAPVVDLIRFRGRWLAYLGDTINRYC (SEQ ID NO: 41),

G(PRR)<sub>n</sub>PGRRAPLGDLISFRGRFLAYFGDTINRYC (SEQ ID NO: 42) for various Bel-binding miniature proteins, or

G(PRR)<sub>n</sub>PGRRAPVEDLIRFKFLLQWYLLALSLRNYAAAC (SEQ ID NO: 43),

G(PRR)<sub>n</sub>PGRRAPVEDLIRFKFLLQWFLLALTRHRYAAAC (SEQ ID NO: 44),

5 G(PRR)<sub>n</sub>PGRRAPVEDLIRFKFLLQWYLLALYHNNYAAAC (SEQ ID NO: 45), or

G(PRR)<sub>n</sub>PGRRAPVEDLIRFKFLLQWYLLALSHYNYAAAC (SEQ ID NO: 46) for

various hDM2-binding miniature proteins, where n is any positive integer, e.g., as

previously described. In addition, it should be noted that such modifications are not

limited to avian pancreatic polypeptides, but include other miniature proteins such as

10 PYY (SEQ ID NO: 53), YY2 (SEQ ID NO: 67), YY1 (SEQ ID NO: 123), or YY3 (SEQ

ID NO: 124). For instance, YY2 may be modified by using one or more PPR groups,

e.g., as in YY2<sup>3R1</sup> (SEQ ID NO: 50), YY2<sup>4R1</sup> (SEQ ID NO: 51), or YY2<sup>5R1</sup> (SEQ ID NO:

52 ).

Thus, in one set of embodiments, the avian pancreatic polypeptide may be

15 selected to be able to bind to Bcl2, and the avian pancreatic polypeptide may be modified

as discussed above in order to increase its cell permeability. For example, the avian

pancreatic polypeptide may include an alpha helix and/or a type II polyproline helix

region modified by the substitution and/or addition of various amino acid residues with

cationic amino acid residues. After such substitutions, the avian pancreatic polypeptide

20 may still exhibit Bcl2 binding activity, e.g., with a K<sub>d</sub> of at least about 7 nM, at least

about 10 nM, at least about 20 nM, at least about 30 nM, at least about 50 nM, at least

about 100 nM, etc. K<sub>d</sub> may be measured using any suitable technique known to those of

ordinary skill in the art. A non-limiting example are fluorescence polarization analysis

techniques such as those discussed in A.C. Gemperli, *et al*, "Paralog-selective Ligands

25 for Bcl-2 Proteins," *J. Am. Chem. Soc.*, 2005, 127, 1596-1597, or in J.W. Chin, *et al*,

"Design and Evolution of a Miniature Bcl-2 Binding Protein," *Angew. Chem. Int. Ed.*

*Eng.*, 2001, 40, 3806-3809. Non-limiting examples of such sequences include CP-

BclXL-1 (SEQ ID NO: 31), CP-Bcl2-1 (SEQ ID NO: 32), or CP-Bcl2-2 (SEQ ID NO:

33). See also Fig. 14. Additional examples include sequences that include one or more

30 repeat units of PPR and/or PRR are present within the type II polyproline helix, for

instance, (PPR)<sub>n</sub> and/or (PRR)<sub>n</sub>, where n is at least 3, e.g., 3, 4, 5, 6, etc., e.g., as

previously discussed. Specific non-limiting example include

G(PRR)<sub>n</sub>PGRRAPVEDLIRFVGRLLAYFGDTINRYC (SEQ ID NO: 40),



G(PRR)<sub>n</sub>PGRRAPVVDLIRFRGRWLAYLGDTINRYC (SEQ ID NO: 41), or  
 G(PRR)<sub>n</sub>PGRRAPLGDLISFRGRFLAYFGDTINRYC (SEQ ID NO: 42).

As another example, in another set of embodiments, the avian pancreatic polypeptide is selected to be able to bind to hDM2, and the avian pancreatic polypeptide  
 5 can be modified to increase its permeability into cells. As discussed above, the avian pancreatic polypeptide can include an alpha helix and/or a type II polyproline helix region modified by the substitution and/or addition of various amino acid residues with cationic amino acid residues, and after such substitutions, the avian pancreatic polypeptide may still exhibit hDM2 binding activity, e.g., with a  $K_d$  at least about 10 nM,  
 10 at least about 20 nM, at least about 30 nM, at least about 50 nM, at least about 100 nM, at least about 300 nM, etc. Specific examples include, but are not limited to, sequences having one or more repeat units of PPR and/or PRR are present within the type II polyproline helix, for instance, (PPR)<sub>n</sub> and/or (PRR)<sub>n</sub>, where n is at least 3, e.g., 3, 4, 5, 6, etc., e.g., as discussed above. Examples of these include CP-liDM2-1 (SEQ ID NO:  
 15 34), CP-hDM2-2 (SEQ ID NO: 35), CP-hDM2-3 (SEQ ID NO: 36), CP-hDM2-4 (SEQ ID NO: 37), CP-hDM2-5 (SEQ ID NO: 38), or CP-hDM2-6 (SEQ ID NO: 39). Other examples include, but are not limited to,

G(PRR)<sub>n</sub>PGRRAPVEDLIRFKFLLQWYLLALSLRNYAAAC (SEQ ID NO: 43),  
 G(PRR)<sub>n</sub>PGRRAPVEDLIRFKFLLQWYLLALTRHRYAAAC (SEQ ID NO: 44),  
 20 G(PRR)<sub>n</sub>PGRRAPVEDLIRFKFLLQWYLLALYHNNYAAAC (SEQ ID NO: 45), or  
 G(PRR)<sub>n</sub>PGRRAPVEDLIRFKFLLQWYLLALSHYNYAAAC (SEQ ID NO: 46), where n is at least 3, e.g., 3, 4, 5, 6, etc.

In some cases, the modified avian pancreatic polypeptide may be able to permeate cells such as Jurkat cells due to the modifications such that, when the modified  
 25 avian pancreatic polypeptide is exposed to a population of Jurkat cells, the modified avian pancreatic polypeptide is able to induce apoptosis (programmed cell death) in at least about 50% of the Jurkat cells at some concentration, such as about 10 micromolar. For instance, the modified avian pancreatic polypeptide may be selected to be able to bind to or otherwise interact with Bcl2 and/or hDM2, as discussed above.

30 Another set of embodiments is directed to a miniature protein, such as aPP, that includes an alpha helix region modified by the substitution and/or addition of various amino acid residues with cationic amino acid residues. A non-limiting example of a cationic amino acid residue is arginine. The residues chosen for substitution may be

present anywhere within the alpha helix region, and may be independently consecutive or non-consecutive.

In one embodiment, the residues chosen for substitution may be ones that are located on one portion or side of the alpha helix. For instance, referring now to Fig. 2, from an end-on view of the alpha helix portion of aPP (starting from A at residue 12, indicated by arrow 51, and moving clockwise), the residues to be substituted may be chosen from the ones located on one side of the alpha helix, e.g., on an exterior side or a solvent-exposed face of the alpha helix when the protein is properly folded, as indicated by portions 54. Without wishing to be bound by any theory, it is believed that such substitutions can cause the exterior side of the miniature protein to appear positively charged or cationic, which may facilitate permeation of the miniature protein into cells. Thus, for example, in aPP, the residues chosen for substitution may be selected from positions 12, 15, 18, 19, 22, 25, and/or 26. Non-limiting examples of such substitutions include aPP<sup>4R1</sup> (SEQ ID NO: 21), aPP<sup>5R1</sup> (SEQ ID NO: 22), aPP<sup>6R1</sup> (SEQ ID NO: 23), as well as SEQ ID NOs: 54-66.

In some aspects, the modified miniature proteins are able to associate with (or bind to) specific sequences of DNA or other proteins. These miniature proteins may be able to bind, for example, to DNA or other proteins with high affinity and selectivity. As used herein, the term "bind" or "binding" refers to the specific association or other specific interaction between two molecular species, such as, but not limited to, protein-DNA interactions and protein-protein interactions, for example, the specific association between proteins and their DNA targets, receptors and their ligands, enzymes and their substrates, etc. Such binding may be specific or non-specific, and can involve various noncovalent interactions such as including hydrogen bonding, metal coordination, hydrophobic forces, van der Waals forces, pi-pi interactions, and/or electrostatic effects. It is contemplated that such association may be mediated through specific sites on each of two (or more) interacting molecular species. Binding can be mediated by structural and/or energetic components. In some cases, the latter will comprise the interaction of molecules with opposite charges.

In one set of embodiments, the invention involves a technique known as protein grafting. Protein grafting has generally been described in U.S. Patent Application Serial No. 09/840,085, filed April 24, 2001, entitled "Modified Avian Pancreatic Polypeptide Miniature Binding Proteins," by A.S. Shrader, *et al*, now U.S. Patent No. 7,297,762,

issued November 20, 2007, incorporated herein by reference. Briefly, protein grafting identifies binding site residues from a globular protein that is able to participate in binding-type associations between that protein and its specific binding partners, then the residues are grafted onto a small but stable protein scaffold. As used herein, the term "binding site" refers to the reactive region or domain of a molecule that directly participate in its specific binding with another molecule. For example, when referring to the binding site on a protein or nucleic acid, binding occurs as a result of the presence of specific amino acids or nucleotide sequence, respectively, that interact with the other molecule. Examples of protein scaffolds of the invention comprise members of the pancreatic fold (PP fold) protein family, particularly avian pancreatic polypeptide (aPP) or pancreatic peptide YY (PYY).

Thus, in one aspect, a modified miniature protein may be able to associate with or bind to a specific sequence of DNA. In some embodiments, the DNA sequence may comprise sites for known proteins that bind to that specific DNA sequence (contemplated known proteins would be, e.g., a promotor or regulator). For example, in the design of a DNA-binding miniature protein, the amino acid residues of a known protein that participate in binding or other association of the protein to that particular DNA sequence are identified.

In some embodiments of the present invention, the relevant binding residues are identified using three-dimensional models of a protein or protein complex based on crystallographic studies while in other embodiments they are identified by studies of deletion or substitution mutants of the protein. The residues that participate in binding of the protein to the specific DNA sequence are then grafted onto those positions of the miniature protein that are not necessary to maintain the tertiary structure of the protein scaffold to form the DNA-binding miniature protein. The identification of such positions can readily be determined empirically by persons of ordinary skill in the art. Other embodiments of the present invention involve the screening of a library of modified miniproteins that contain peptide species capable of specific association or binding to that specific DNA (or, in other cases, protein) sequence or motif.

Generally, it is contemplated that any potential binding site on a DNA sequence can be targeted using the DNA binding miniature proteins of the invention. Certain embodiments include miniature proteins having helical structures which bind to a DNA binding site. In some embodiments, the binding involves a basic region leucine zipper

(bZIP) structure, while in other embodiments the structure involves a basic-helix-loop-helix (bHLH) structure. In another embodiment, the binding involves a structure like those found in homeodomain proteins. Example bZIP structures include, but are not limited to, those found in GCN4 and C/EBP-delta, and example bHLH structures  
5 include, but are not limited to, those found in Max, Myc and MyoD. Example homeodomain structures include, but are not limited to, those found in the Q50 engrailed variant protein.

As mentioned, one aspect of the present invention is generally directed to the addition of certain sequences containing repeat units such as PRR and/or PPR that may  
10 be added to any suitable protein in order to increase its cell permeability. Without wishing to be bound by any theory, it is believed that the presence of  $(\text{PRR})_n$  and/or  $(\text{PPR})_n$  repeat units (where n is at least 2, e.g., 2, 3, 4, 5, 6, 7, 8, etc.) may cause the formation of a helical structure within the modified protein, such as a type II polyproline helical structure, that appears to be positively charged or cationic (or at least substantially  
15 cationic), which may facilitate permeation of the protein into cells. Thus, a protein may be modified by adding a plurality of repeat units such as PRR and/or PPR at any suitable location, e.g., at a terminal end of the protein, within an internal sequence of the protein, or the like. In some cases, the modification may be chosen such that the repeat units are positioned on an exterior surface of the protein.

20 The cell permeability of the protein may be determined, for example, as previously described. Thus, for example, cells such as a HeLa cells, Jurkat cells, etc., may be exposed to a concentration of 1 micromolar of modified protein, and the concentrations of each within the cell may be determined in some fashion. For example, the protein may be labeled with a fluorescent entity, such as fluorescein, and the relative  
25 concentrations determined using techniques such as flow cytometry. In some aspects, a miniature protein of the present invention is produced and selected using a phage display method. In such a method, display of recombinant miniature proteins on the surface of viruses which infect bacteria (bacteriophage or phage) make it possible to produce soluble, recombinant miniature proteins having a wide range of affinities and kinetic  
30 characteristics. To display the miniature proteins on the surface of phage, a synthetic gene encoding the miniature protein is inserted into the gene encoding a phage surface protein (*pill*) and the recombinant fusion protein is expressed on the phage surface. Variability may be introduced into the phage display library to select for miniature

proteins which not only maintain their tertiary, helical structure but which also display increased affinity for a preselected target because the critical (or contributing but not critical) binding residues are optimally positioned on the helical structure.

Since the recombinant proteins on the surface of the phage are functional, phage bearing miniature proteins that bind with high-affinity to a particular target DNA or protein can be separated from non-binding or lower affinity phage by using techniques such as antigen affinity chromatography. Mixtures of phage are allowed to bind to the affinity matrix, non-binding or lower affinity phage are removed by washing, and bound phage are eluted by treatment with acid or alkali. Depending on the affinity of the miniature protein for its target, enrichment factors of twenty-fold to a million-fold are obtained by a single round of affinity selection. By infecting bacteria with the eluted phage, however, more phage can be grown and subjected to another round of selection. In this way, an enrichment of a thousand-fold in one round becomes a million-fold in two rounds of selection. Thus, even when enrichments in each round are low, multiple rounds of affinity selection leads to the isolation of rare phage and the genetic material contained within which encodes the sequence of the domain or motif of the recombinant miniature protein that binds or otherwise specifically associates with its binding target.

Accordingly, in various embodiments of the invention, the methods disclosed herein are used to produce a phage expression library encoding miniature proteins capable of binding to a DNA or to a protein that has already been selected using the protein grafting procedure described above. In such embodiments, phage display can be used to identify miniature proteins that display an even higher affinity for a particular target DNA or protein than that of the miniature proteins produced without the aid of phage display. In yet another embodiment, the invention encompasses a universal phage display library that can be designed to display a combinatorial set of epitopes or binding sequences to permit the recognition of nucleic acids, proteins or small molecules by a miniature protein without prior knowledge of the natural epitope or specific binding residues or motifs natively used for recognition and association.

Various structural modifications also are contemplated for the present invention that, for example, include the addition of restriction enzyme recognition sites into the polynucleotide sequence encoding the miniature protein that enable genetic manipulation of these gene sequences. Accordingly, the re-engineered miniature proteins can be ligated, for example, into an M13-derived bacteriophage cloning vector that permits

expression of a fusion protein on the phage surface. These methods allow for selecting phage clones encoding fusion proteins that bind a target ligand and can be completed in a rapid manner allowing for high-throughput screening of miniature proteins to identify the miniature protein with the highest affinity and selectivity for a particular target.

5           According to the methods of the invention, a library of phage displaying modified miniature proteins is incubated with the immobilized target DNA or proteins to select phage clones encoding miniature proteins that specifically bind to or otherwise specifically associate with the immobilized DNA or protein. This procedure involves immobilizing a oligonucleotide or polypeptide sample on a solid substrate. The bound  
10 phage are then dissociated from the immobilized oligonucleotide or polypeptide and amplified by growth in bacterial host cells. Individual viral plaques, each expressing a different recombinant miniature protein, are expanded to produce amounts of protein sufficient to perform a binding assay. The DNA encoding this recombinant binding protein can be subsequently modified for ligation into a eukaryotic protein expression  
15 vector. The modified miniature protein, adapted for expression in eukaryotic cells, is ligated into a eukaryotic protein expression vector.

          In another aspect, the invention encompasses miniature proteins that bind to other proteins and methods for making these miniature proteins. The binding of the miniature proteins modulates protein-protein and/or protein-ligand interactions. Thus, in some  
20 embodiments the binding blocks the association (or specific binding) of ligands and receptors. The ligand can be either another protein but also can be any other type of molecule such as a chemical substrate. In one embodiment of the present invention, making the protein-binding miniature protein of the invention involves determining the amino acid residues which are essential to binding of the ligand protein to its target  
25 receptor protein. In some embodiments, these essential residues are identified using three-dimensional models of a protein or protein complex which binds to or interacts with another protein based on crystallographic studies while in other embodiments they are identified by studies of deletion or substitution mutants of the protein. The residues that participate in binding of the protein to are then grafted onto those positions which  
30 are not necessary to maintain the tertiary structure of the protein scaffold to form the protein-binding miniature protein.

          The miniature proteins of the present invention further include conservative variants of the miniature proteins herein described, according to another aspect. As used

herein, a "conservative variant" refers to alterations in the amino acid sequence that do not substantially and adversely affect the binding or association capacity of the protein. A substitution, insertion or deletion is said to adversely affect the miniature protein when the altered sequence prevents, reduces, or disrupts a function or activity associated with the protein. For example, the overall charge, structure or hydrophobic-hydrophilic properties of the miniature protein can be altered without adversely affecting an activity. Accordingly, the amino acid sequence can be altered, for example to render the peptide more hydrophobic or hydrophilic, without adversely affecting the activities of the miniature protein.

These variants, though possessing a slightly different amino acid sequence than those recited above, will still have the same or similar properties associated with any of the miniature proteins discussed herein, for instance, SEQ ID NOs: 1-124. Ordinarily, the conservative substitution variants, will have an amino acid sequence having at least 75%, at least 80%, at least 85%, at least 90%, at least 95% amino acid, at least 98%, or at least 99% sequence identity with any of the miniature proteins discussed herein, for example, SEQ ID NOs: 1-124. Identity or homology with respect to such sequences is defined herein as the percentage of amino acid residues in the candidate sequence that are identical with the known peptides, after aligning the sequences and introducing gaps, if necessary, to achieve the maximum percent homology, and not considering any conservative substitutions as part of the sequence identity. N-terminal, C-terminal or internal extensions, deletions, or insertions into the peptide sequence shall not be construed as affecting homology.

Thus, the miniature proteins of the present invention include molecules comprising any of the amino acid sequences discussed herein, including SEQ ID NOs: 1-124; fragments thereof having a consecutive sequence of at least about 15, 20, 25, 30, 35 or more amino acid residues of the miniature proteins of the invention; amino acid sequence variants of such sequences wherein at least one amino acid residue has been inserted N- or C-terminal to, or within, the disclosed sequence; amino acid sequence variants of the disclosed sequences, or their fragments as defined above, that have been substituted by another residue. Contemplated variants further include those derivatives wherein the protein has been covalently modified by substitution, chemical, enzymatic, or other appropriate means with a moiety other than a naturally occurring amino acid (for example, a detectable moiety such as an enzyme or radioisotope).

The present invention further provides, in another aspect, nucleic acid molecules that encode any of the amino acid sequences discussed herein, including any of SEQ ID NOs: 1-124 or the related miniature proteins herein described, preferably in isolated form. As used herein, "nucleic acid" includes cDNA and mRNA, as well as nucleic acids based on alternative backbones or including alternative bases whether derived from natural sources or synthesized. Those of ordinary skill in the art, given an amino acid sequence, will be able to generate corresponding nucleic acid sequences that can be used to generate the amino acid sequence, using no more than routine skill. As used herein, a nucleic acid molecule is said to be "isolated" when the nucleic acid molecule is substantially separated from contaminant nucleic acid encoding other polypeptides from the source of nucleic acid.

The present invention also provides fragments of the encoding nucleic acid molecule. As used herein, a "fragment of an encoding nucleic acid molecule" refers to a portion of the entire protein encoding sequence of the miniature protein. The size of the fragment will be determined by the intended use. For example, if the fragment is chosen so as to encode an active portion of the protein, the fragment will need to be large enough to encode the functional region(s) of the protein. The appropriate size and extent of such fragments can be determined empirically by persons skilled in the art.

Modifications to the primary structure itself by deletion, addition, or alteration of the amino acids incorporated into the protein sequence during translation can be made without destroying the activity of the miniature protein. Such substitutions or other alterations result in miniature proteins having an amino acid sequence encoded by a nucleic acid falling within the contemplated scope of the present invention.

The present invention further provides, in some embodiments, recombinant DNA molecules that contain a coding sequence. As used herein, a "recombinant DNA molecule" is a DNA molecule that has been subjected to molecular manipulation. Methods for generating recombinant DNA molecules are well known in the art, for example, see Sambrook *et al.*, (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press. In some recombinant DNA molecules, a coding DNA sequence is operably linked to expression control sequences and vector sequences.

The choice of vector and expression control sequences to which one of the protein family encoding sequences of the present invention is operably linked depends directly, as is well known in the art, on the functional properties desired (e.g., protein



expression, and the host cell to be transformed). A vector of the present invention may be at least capable of directing the replication or insertion into the host chromosome, and preferably also expression, of the structural gene included in the recombinant DNA molecule.

5           Expression control elements that are used for regulating the expression of an operably linked miniature protein encoding sequence are known in the art and include, but are not limited to, inducible promoters, constitutive promoters, secretion signals, and other regulatory elements. Preferably, the inducible promoter is readily controlled, such as being responsive to a nutrient in the host cell's medium.

10           In one embodiment, the vector containing a coding nucleic acid molecule will include a prokaryotic replicon, i.e., a DNA sequence having the ability to direct autonomous replication and maintenance of the recombinant DNA molecule extra-chromosomal in a prokaryotic host cell, such as a bacterial host cell, transformed therewith. Such replicons are well known in the art. In addition, vectors that include a  
15           prokaryotic replicon may also include a gene whose expression confers a detectable marker such as a drug resistance. Typical of bacterial drug resistance genes are those that confer resistance to ampicillin or tetracycline.

          Vectors that include a prokaryotic replicon can further include a prokaryotic or  
20           bacteriophage promoter capable of directing the expression (transcription and translation) of the coding gene sequences in a bacterial host cell, such as *E. coli*. A promoter is an expression control element formed by a DNA sequence that permits binding of RNA polymerase and transcription to occur. Promoter sequences compatible with bacterial hosts are typically provided in plasmid vectors containing convenient  
25           restriction sites for insertion of a DNA segment of the present invention. Any suitable prokaryotic host can be used to express a recombinant DNA molecule encoding a protein of the invention.

          Expression vectors compatible with eukaryotic cells, preferably those compatible with vertebrate cells, can also be used to form recombinant DNA molecules that contains a coding sequence. Eukaryotic cell expression vectors are well known in the art and are  
30           available from several commercial sources. Typically, such vectors are provided containing convenient restriction sites for insertion of the desired DNA segment.

          Eukaryotic cell expression vectors used to construct the recombinant DNA molecules of the present invention may further include a selectable marker that is

effective in an eukaryotic cell, preferably a drug resistance selection marker. An example drug resistance marker is the gene whose expression results in neomycin resistance, i.e., the neomycin phosphotransferase (*neo*) gene. Alternatively, the selectable marker can be present on a separate plasmid, the two vectors introduced by co-  
5 transfection of the host cell, and transfectants selected by culturing in the appropriate drug for the selectable marker.

The present invention further provides, in yet another aspect, host cells transformed with a nucleic acid molecule that encodes a miniature protein of the present invention. The host cell can be either prokaryotic or eukaryotic. Eukaryotic cells useful  
10 for expression of a miniature protein of the invention are not limited, so long as the cell line is compatible with cell culture methods and compatible with the propagation of the expression vector and expression of the gene product.

Transformation of appropriate cell hosts with a recombinant DNA molecule encoding a miniature protein of the present invention is accomplished by well known  
15 methods that typically depend on the type of vector used and host system employed. With regard to transformation of prokaryotic host cells, electroporation and salt treatment methods can be employed (see, for example, Sambrook *et al.*, (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press). With regard to transformation of vertebrate cells with vectors containing recombinant DNA,  
20 electroporation, cationic lipid or salt treatment methods can be employed (see, for example, Graham *et al.*, (1973) *Virology* 52, 456-467; Wigler *et al.*, (1979) *Proc. Natl. Acad. Sci. USA* 76, 1373-1376).

Successfully transformed cells (cells that contain a recombinant DNA molecule of the present invention), can be identified by well known techniques including the  
25 selection for a selectable marker. For example, cells resulting from the introduction of a recombinant DNA of the present invention can be cloned to produce single colonies. Cells from those colonies can be harvested, lysed and their DNA content examined for the presence of the recombinant DNA using a method such as that described by Southern, (1975) *J. Mol. Biol.* 98, 503-517 or the proteins produced from the cell assayed  
30 via an immunological method.

The present invention further provides, in still another aspect, methods for producing a miniature protein of the invention using nucleic acid molecules herein described. In general terms, the production of a recombinant form of a protein typically

involves the following steps: a nucleic acid molecule is obtained that encodes a protein of the invention, such as the nucleic acid molecule encoding any of the miniature proteins described herein, including any of SEQ ID NOs: 1-124. The nucleic acid molecule may then be placed in operable linkage with suitable control sequences, as described above, to form an expression unit containing the protein open reading frame. The expression unit is used to transform a suitable host and the transformed host is cultured under conditions that allow the production of the recombinant miniature protein. Optionally the recombinant miniature protein is isolated from the medium or from the cells; recovery and purification of the protein may not be necessary in some instances where some impurities may be tolerated.

Each of the foregoing steps can be done in a variety of ways. The construction of expression vectors that are operable in a variety of hosts is accomplished using appropriate replicons and control sequences, as set forth above. The control sequences, expression vectors, and transformation methods are dependent on the type of host cell used to express the gene. Suitable restriction sites, if not normally available, can be added to the ends of the coding sequence so as to provide an excisable gene to insert into these vectors. An artisan of ordinary skill in the art can readily adapt any host/expression system known in the art for use with the nucleic acid molecules of the invention to produce a recombinant miniature protein.

In another aspect, the present invention provides methods for use in isolating and identifying binding partners of the miniature proteins of the invention. In some embodiments, a miniature protein of the invention is mixed with a potential binding partner or an extract or fraction of a cell under conditions that allow the association of potential binding partners with the protein of the invention. After mixing, peptides, polypeptides, proteins or other molecules that have become associated with a miniature protein of the invention are separated from the mixture. The binding partner bound to the protein of the invention can then be removed and further analyzed. To identify and isolate a binding partner, the entire miniature protein can be used. Alternatively, a fragment of the miniature protein which contains the binding domain can be used.

As used herein, a "cellular extract" refers to a preparation or fraction which is made from a lysed or disrupted cell. A variety of methods can be used to obtain an extract of a cell. Cells can be disrupted using either physical or chemical disruption methods. Examples of physical disruption methods include, but are not limited to,

sonication and mechanical shearing. Examples of chemical lysis methods include, but are not limited to, detergent lysis and enzyme lysis. A skilled artisan can readily adapt methods for preparing cellular extracts in order to obtain extracts for use in the present methods.

5           Once an extract of a cell is prepared, the extract is mixed with the a miniature protein of the invention under conditions in which association of the miniature protein with the binding partner can occur. A variety of conditions can be used, the most preferred being conditions that closely resemble conditions found in the cytoplasm of a human cell. Features such as osmolarity, pH, temperature, and the concentration of  
10 cellular extract used, can be varied to optimize the association of the protein with the binding partner.

          After mixing under appropriate conditions, the bound complex is separated from the mixture. A variety of techniques can be utilized to separate the mixture. For example, antibodies specific to a protein of the invention can be used to  
15 immunoprecipitate the binding partner complex. Alternatively, standard chemical separation techniques such as chromatography and density-sediment centrifugation can be used.

          After removal of non-associated cellular constituents found in the extract, the binding partner can be dissociated from the complex using conventional methods. For  
20 example, dissociation can be accomplished by altering the salt concentration or pH of the mixture.

          To aid in separating associated binding partner pairs from the mixed extract, the miniature protein of the invention can be immobilized on a solid support. For example, the miniature protein can be attached to a nitrocellulose matrix or acrylic beads.  
25 Attachment of the miniature protein to a solid support aids in separating peptide-binding partner pairs from other constituents found in the extract. The identified binding partners can be either a single DNA molecule or protein or a complex made up of two or more proteins. Alternatively, binding partners may be identified using the alkaline phosphatase fusion assay according to the procedures of *Flanagan & Vanderhaeghen*,  
30 (1998) *Annu. Rev. Neurosci.* 21, 309-345 or Takahashi *et al.*, (1999) *Cell* 99, 59-69; the Far-Western assay according to the procedures of Takayama *et al.*, (1997) *Methods Mol. Biol.* 69, 171-184, or Sauder *et al.*, *J. Gen. Virol.* (1996) 77, 991-996 or identified through the use of epitope tagged proteins or GST fusion proteins.

In another embodiment, the nucleic acid molecules encoding a miniature protein of the invention can be used in a yeast two-hybrid system. The yeast two-hybrid system has been used to identify other protein partner pairs and can readily be adapted to employ the nucleic acid molecules herein described (see, *e.g.*, Stratagene Hybrizap® two-hybrid system).

According to some aspects, the miniature proteins of the invention are useful for drug screening to identify agents capable of binding to the same binding site as the miniature proteins. The miniature proteins are also useful for diagnostic purposes to identify the presence and/or detect the levels of DNA or protein that binds to the miniature proteins of the invention. In one diagnostic embodiment, the miniature proteins of the invention are included in a kit used to detect the presence of a particular DNA or protein in a biological sample. The miniature proteins of the invention also have therapeutic uses in the treatment of disease associated with the presence of a particular DNA or protein. In one therapeutic embodiment, the miniature proteins can be used to bind to DNA to promote or inhibit transcription, while in another therapeutic embodiment, the miniature proteins bind to a protein resulting in inhibition or stimulation of the protein.

In some aspects of the invention, miniature proteins of the invention are administered to a subject in an effective amount to treat a cancer or a tumor. Non-limiting examples of such proteins include SEQ ID NOs. 40-42, which are targeted towards the BCL-2 pathway. Without wishing to be bound by any theory, it is believed that such proteins may inhibit the interaction of a protein containing a BH3 death domain (*e.g.*, Bak or Bid) with a multidomain BCL-2 family member.

As another example, a miniature protein may inhibit (completely or partially) migration of a tumor cell across a barrier, thereby forming an eclipse. The invasion and metastasis of cancer is a complex process which involves changes in cell adhesion properties which allow a transformed cell to invade and migrate through the extracellular matrix (ECM) and acquire anchorage-independent growth properties. Some of these changes occur at focal adhesions, which are cell/ECM contact points containing membrane-associated, cytoskeletal, and intracellular signaling molecules. Metastatic disease occurs when the disseminated foci of tumor cells seed a tissue which supports their growth and propagation, and this secondary spread of tumor cells is responsible for the morbidity and mortality associated with the majority of cancers. Thus the term

"metastasis" as used herein refers to the invasion and migration of tumor cells away from the primary tumor site.

Miniature proteins of the invention may be administered to cells of a subject to treat or prevent diseases (e.g., cancers) alone or in combination with the administration  
5 of other therapeutic compounds for the treatment or prevention of these disorders.

In certain aspects, miniature proteins of the invention are useful for diagnostic purposes to identify the presence and/or detect the levels of a target protein that binds to the miniature proteins of the invention. For example, miniature proteins of the invention can be used to detect the levels of Bcl2 or hDM2 due to its high affinity and high  
10 specificity. Miniature proteins of this method can be labeled with a detectable marker. A wide range of detectable markers can be used, including but not limited to biotin, a fluorogen, an enzyme, an epitope, a chromogen, or a radionuclide. The method for detecting the label will depend on the nature of the label and can be any known in the art, e.g., film to detect a radionuclide, an enzyme substrate that gives rise to a detectable  
15 signal to detect the presence of an enzyme, antibody to detect the presence of an epitope, etc.

In a specific diagnostic embodiment, miniature proteins of the invention are included in a kit used to detect the presence of a particular protein (e.g., Bcl2 or hDM2) in a biological sample.

20 In certain aspects, therapeutic compounds of the present invention (e.g., miniature proteins) are formulated with a pharmaceutically acceptable carrier. Miniature proteins of the present invention can be administered alone or as a component of a pharmaceutical formulation (composition). The compounds may be formulated for administration in any convenient way for use in human or veterinary medicine. Wetting  
25 agents, emulsifiers and lubricants, such as sodium lauryl sulfate and magnesium stearate, as well as coloring agents, release agents, coating agents, sweetening, flavoring and perfuming agents, preservatives and antioxidants can also be present in the compositions.

Formulations of the miniature proteins include those suitable for oral/nasal, topical, parenteral and/or intravaginal administration. The formulations may  
30 conveniently be presented in unit dosage form and may be prepared by any methods well known in the art of pharmacy. The amount of active ingredient which can be combined with a carrier material to produce a single dosage form will vary depending upon the host being treated, the particular mode of administration. The amount of active ingredient

which can be combined with a carrier material to produce a single dosage form will generally be that amount of the compound which produces a therapeutic effect.

Methods of preparing these formulations or compositions include combining one compound and a carrier and, optionally, one or more accessory ingredients. In general,  
5 the formulations are prepared by combining a compound with a liquid carrier, or a finely divided solid carrier, or both, and then, if necessary, shaping the product.

Formulations of the miniature proteins suitable for oral administration may be in the form of capsules, cachets, pills, tablets, lozenges (using a flavored basis, usually sucrose and acacia or tragacanth), powders, granules, or as a solution or a suspension in  
10 an aqueous or non-aqueous liquid, or as an oil-in-water or water-in-oil liquid emulsion, or as an elixir or syrup, or as pastilles (using an inert base, such as gelatin and glycerin, or sucrose and acacia) and/or as mouth washes and the like, each containing a predetermined amount of a compound as an active ingredient. A compound may also be administered as a bolus, electuary or paste.

15 In solid dosage forms for oral administration (capsules, tablets, pills, dragees, powders, granules, and the like), a miniature protein is mixed with one or more pharmaceutically acceptable carriers, such as sodium citrate or dicalcium phosphate, and/or any of the following: (1) fillers or extenders, such as starches, lactose, sucrose, glucose, mannitol, and/or silicic acid; (2) binders, such as, for example,  
20 carboxymethylcellulose, alginates, gelatin, polyvinyl pyrrolidone, sucrose, and/or acacia; (3) humectants, such as glycerol; (4) disintegrating agents, such as agar-agar, calcium carbonate, potato or tapioca starch, alginic acid, certain silicates, and sodium carbonate; (5) solution retarding agents, such as paraffin; (6) absorption accelerators, such as quaternary ammonium compounds; (7) wetting agents, such as, for example, cetyl  
25 alcohol and glycerol monostearate; (8) absorbents, such as kaolin and bentonite clay; (9) lubricants, such as talc, calcium stearate, magnesium stearate, solid polyethylene glycols, sodium lauryl sulfate, and mixtures thereof; and/or (10) coloring agents. In the case of capsules, tablets and pills, the pharmaceutical compositions may also comprise buffering agents. Solid compositions of a similar type may also be employed as fillers in soft and  
30 hard-filled gelatin capsules using such excipients as lactose or milk sugars, as well as high molecular weight polyethylene glycols and the like.

Liquid dosage forms for oral administration of a miniature protein include pharmaceutically acceptable emulsions, microemulsions, solutions, suspensions, syrups,

and elixirs. In addition to the active ingredient, the liquid dosage forms may contain inert diluents commonly used in the art, such as water or other solvents, solubilizing agents and emulsifiers, such as ethyl alcohol, isopropyl alcohol, ethyl carbonate, ethyl acetate, benzyl alcohol, benzyl benzoate, propylene glycol, 1,3-butylene glycol, oils (in particular, cottonseed, groundnut, corn, germ, olive, castor, and sesame oils), glycerol, tetrahydrofuryl alcohol, polyethylene glycols and fatty acid esters of sorbitan, and mixtures thereof. Besides inert diluents, the oral compositions can also include adjuvants such as wetting agents, emulsifying and suspending agents, sweetening, flavoring, coloring, perfuming, and preservative agents.

10           Suspensions, in addition to the active compounds (e.g., miniature proteins), may contain suspending agents such as ethoxylated isostearyl alcohols, polyoxyethylene sorbitol, and sorbitan esters, microcrystalline cellulose, aluminum metahydroxide, bentonite, agar-agar and tragacanth, and mixtures thereof.

15           Methods of the invention can be administered topically in some embodiments, either to skin or to mucosal membranes (e.g., those on the cervix and vagina). This offers the greatest opportunity for direct delivery to tumor with the lowest chance of inducing side effects. The topical formulations may further include one or more of the wide variety of agents known to be effective as skin or stratum corneum penetration enhancers. Examples of these are 2-pyrrolidone, *N*-methyl-2-pyrrolidone, 20 dimethylacetamide, dimethylformamide, propylene glycol, methyl or isopropyl alcohol, dimethyl sulfoxide, and azone. Additional agents may further be included to make the formulation cosmetically acceptable. Examples of these are fats, waxes, oils, dyes, fragrances, preservatives, stabilizers, and surface active agents. Keratolytic agents such as those known in the art may also be included. Examples are salicylic acid and sulfur.

25           Dosage forms for the topical or transdermal administration of a compound (e.g., a miniature protein) include powders, sprays, ointments, pastes, creams, lotions, gels, solutions, patches, and inhalants. The active compound may be mixed under sterile conditions with a pharmaceutically acceptable carrier, and with any preservatives, buffers, or propellants which may be required. The ointments, pastes, creams and gels 30 may contain, in addition to a therapeutic compound, excipients, such as animal and vegetable fats, oils, waxes, paraffins, starch, tragacanth, cellulose derivatives, polyethylene glycols, silicones, bentonites, silicic acid, talc and zinc oxide, or mixtures thereof.



Powders and sprays can contain, in addition to a compound, excipients such as lactose, talc, silicic acid, aluminum hydroxide, calcium silicates, and polyamide powder, or mixtures of these substances. Sprays can additionally contain customary propellants, such as chlorofluorohydrocarbons and volatile unsubstituted hydrocarbons, such as  
5 butane and propane.

Pharmaceutical compositions suitable for parenteral administration may comprise one or more compounds in combination with one or more pharmaceutically acceptable sterile isotonic aqueous or nonaqueous solutions, dispersions, suspensions or emulsions, or sterile powders which may be reconstituted into sterile injectable solutions or  
10 dispersions just prior to use, which may contain antioxidants, buffers, bacteriostats, solutes which render the formulation isotonic with the blood of the intended recipient or suspending or thickening agents. Examples of suitable aqueous and nonaqueous carriers which may be employed in the pharmaceutical compositions of the invention include  
15 water, ethanol, polyols (such as glycerol, propylene glycol, polyethylene glycol, and the like), and suitable mixtures thereof, vegetable oils, such as olive oil, and injectable organic esters, such as ethyl oleate. Proper fluidity can be maintained, for example, by the use of coating materials, such as lecithin, by the maintenance of the required particle size in the case of dispersions, and by the use of surfactants.

Injectable depot forms are made by forming microencapsule matrices of the  
20 compounds in biodegradable polymers such as polylactide-polyglycolide. Depending on the ratio of drug to polymer, and the nature of the particular polymer employed, the rate of drug release can be controlled. Examples of other biodegradable polymers include poly(orthoesters) and poly(anhydrides). Depot injectable formulations are also prepared by entrapping the drug in liposomes or microemulsions which are compatible with body  
25 tissue.

Formulations of the compounds for intravaginal administration may be presented as a suppository, which may be prepared by mixing one or more compounds of the invention with one or more suitable nonirritating excipients or carriers comprising, for example, cocoa butter, polyethylene glycol, a suppository wax or a salicylate, and which  
30 is solid at room temperature, but liquid at body temperature and, therefore, will melt in the rectum or vaginal cavity and release the active compound. Optionally, such formulations suitable for vaginal administration also include pessaries, tampons, creams,

gels, pastes, foams or spray formulations containing such carriers as are known in the art to be appropriate.

The following documents are incorporated herein by reference: U.S. Provisional Patent Application Serial No. 61/068,259, filed March 5, 2008, entitled "Cell Permeable  
5 Miniprotein Inhibitors of BCL2 Interactions," by A. Schepartz; U.S. Provisional Patent Application Serial No. 61/011,311, filed January 16, 2008, entitled "Cell-Permeable Miniature Proteins via  $\alpha$ -Helical Arginine Display," by A. Schepartz, *et al*; U.S. Provisional Patent Application Serial No. 61/009,905, filed January 3, 2008, entitled "Intrinsically Cell-Permeable Miniature Proteins Based on a Minimal Cationic PPII  
10 Motif," by A. Schepartz, *et al*; and U.S. Provisional Patent Application Serial No. 60/926,379, filed April 24, 2007, entitled "Cell Permeable Miniature Proteins," by A. Schepartz Shrader, *et al*; U.S. Provisional Patent Application Serial No. 60/963,744, filed August 6, 2007, entitled "Engineering a Monomeric Miniature Protein," by Hodges, *et al*., incorporated herein by reference; U.S. Patent Application Serial No. 09/840,085,  
15 filed April 24, 2001, entitled "Modified Avian Pancreatic Polypeptide Miniature Binding Proteins," by A.S. Shrader, *et al*, now U.S. Patent No. 7,297,762, issued November 20, 2007; U.S. Patent Application Serial No. 11/009,101, filed December 10, 2004, entitled "Protein Binding Miniature Proteins and Uses Thereof," by D. Golemi-Kotra, *et al*, published as U.S. Patent Application Publication No. 2005/0287643 on December 29,  
20 2005; U.S. Patent Application Serial No. 10/982,727, filed November 4, 2004, entitled "Protein Binding Miniature Proteins," by A.S. Shrader, *et al*, published as U.S. Patent Application Publication No. 2005/0287542 on December 29, 2005; and an International Patent Application filed on April 16, 2008, entitled "Modified Miniature Proteins," by A.S. Schepartz, *et al*

25 The following examples are intended to illustrate certain embodiments of the present invention, but do not exemplify the full scope of the invention.

#### EXAMPLE 1

This example illustrates encodable cell-penetrating peptides (CPPs) possessing  
30 type-II polyproline (PPII) helical structure that can be embedded within the primary sequence of a small structured protein. The marriage of arginine- and proline-rich sequences for the generation of structured CPPs is appealing for several reasons. PPII helices are stable in short, isolated sequence motifs. Moreover, due to its large 3.1

Angstrom rise per residue, the PPII helix readily tolerates charged side chains on successive turns. In fact, arginine has high PPII propensity in host-guest studies, facilitating the design of PPII helices possessing one or more cationic faces. In addition, natural proline-rich sequences such as Bactenecin-7 (Bac) and gamma-zein (SAP) permeate cells, albeit with low efficiency. Although non-natural amino acid modifications can substantially increase the uptake of PPII-based CPPs, these derivatives are not readily DNA-encodable.

To establish design rules for successfully wedding the uptake capacity of oligoarginines to the structural propensity of oligoprolines, two series of cationic PPII helices were synthesized. Exploiting the periodicity (3.0 residues per turn) of the left-handed PPII helix, these minimalist repeats of (PPR)<sub>n</sub> and (PRR)<sub>n</sub> created aligned faces of proline or arginine residues. All peptides were labeled at the N-terminus with fluorescein, and their uptake into live HeLa cells was evaluated by flow cytometry (Fig. 3A). Fig. 3A illustrates HeLa cell uptake of 1 micromolar fluorescein-labeled peptides after 1 h quantified by flow cytometry. This plot illustrates the mean cellular fluorescence  $\pm$  the standard error of three experiments.

Neither the previously reported proline-rich sequences, SAP and Bac, nor members of the proline-rich (PPR)<sub>n</sub> series showed significant uptake after 1 h at 1 micromolar. (PRR)<sub>3</sub> (SEQ ID: NO: 68) and (PRR)<sub>4</sub> (SEQ ID: NO: 69) were slightly more cell-permeable under these conditions, performing as well as Tat. In contrast, (PRR)<sub>5</sub> (SEQ ID NO: 70) and (PRR)<sub>6</sub> (SEQ ID NO: 71) were found to be extraordinarily cell-permeable, matching or surpassing the levels observed for simple oligoarginines R<sub>8</sub> (SEQ ID NO: 72), Ri<sub>0</sub> (SEQ ID NO: 73), and Ri<sub>2</sub> (SEQ ID NO: 74). The relative efficiencies of R<sub>8</sub>, Ri<sub>0</sub>, and Ri<sub>2</sub> were length-dependent, consistent with earlier reports that oligoarginine uptake increases with increasing charge but peaks between R<sub>8</sub> and Ri<sub>5</sub> (SEQ ID NO: 75). The (PRR)<sub>n</sub> series displayed an even stronger length dependence, such that (PRR)<sub>4</sub> significantly underperformed its "iso-ionic" counterpart R<sub>g</sub>, whereas (PRR)<sub>5</sub> matched Ri<sub>0</sub> and (PRR)<sub>6</sub> outperformed Ri<sub>2</sub>. Thus, at longer lengths, the uptake of simple (PRR)<sub>n</sub> peptides exceeded the maximal uptake of oligoarginine vectors, outperforming the commonly used sequences R<sub>8</sub> and Tat by 3-4 and 29-44 times, respectively.

While the flow cytometry protocol used here included trypsin treatment to remove any cell-surface bound peptide, live-cell confocal microscopy was also

performed to verify internalization and assess intracellular distribution (Fig. 3B). In this figure, HeLa cells were incubated with 1 micromolar (PRR)<sub>5</sub> for 20 min, and intracellular localization was monitored by confocal microscopy. The endosomes were visualized using 10 micromolar 10 kDa dextran labeled with AlexaFluor-647. The peptides generated diffuse cytoplasmic and nuclear staining, as well as strong nucleoli staining. The lack of co-localization with dextran, an endocytotic marker, revealed that little peptide was present in the endosomes. Another requisite feature of import tags is that they are not intrinsically cytotoxic at concentrations required for uptake. No significant reduction was found in cell viability after a 6 h incubation with 1-100 micromolar (PPR)<sub>n</sub> and (PRR)<sub>n</sub> peptides. See Table 1 and Figs. 5A and 5B, which shows HeLa cells were treated with varying concentrations of peptide for 6 hours, and cellular viability determined by resazurin reduction, as described below. The standard deviation of four independent experiments is reported.

Table 1

| Peptide          | % Cell Viability <sup>1</sup> | MKKMAX" | Wavelength (nm) | % PPII |
|------------------|-------------------------------|---------|-----------------|--------|
| PPR <sub>6</sub> | 102 ± 10                      | 396     | 228.0           | 47     |
| PPR <sub>4</sub> | 94 ± 10                       | 148     | 229.0           | 46     |
| PPR <sub>5</sub> | 104 ± 11                      | 39      | 229.0           | 45     |
| PPR <sub>6</sub> | 104 ± 3                       | -56     | 229.5           | 44     |
| PRR <sub>3</sub> | 96 ± 2                        | 192     | 228.0           | 46     |
| PRR <sub>4</sub> | 92 ± 8                        | -37     | 228.5           | 41     |
| PRR <sub>5</sub> | 99 ± 5                        | -232    | 227.5           | 43     |
| PRR <sub>6</sub> | 104 ± H                       | -376    | 229.0           | 42     |
| RR3-aPP          | 98 ± 6                        |         |                 |        |
| KR4-aPP          | 98 ± 5                        |         |                 |        |
| RRS-aPP          | 91 ± 8                        |         |                 |        |

<sup>1</sup>10 μM peptide, 6 hours, standard deviation of four experiments; "degH/mole (#eπ)\* at 5 X' in PBS. pli 7.4: '1τ(0)11 ·, PPII = ((MRJ)\_{MAX} † 6 100j/1 3.7W)•100

Consistent with their design, both the (PPR)<sub>n</sub> and (PRR)<sub>n</sub> series displayed characteristic PPII signatures, with weak mean residual ellipticity (MRE) maxima at ~228 nm and strong minima at ~204 nm (Fig. 3C). This figure illustrates two-color fluorescence with bright-field superposition. The bands observed for arginine-rich (PRR)<sub>n</sub> peptides were blue-shifted relative to proline-rich (PPR)<sub>n</sub> peptides, as expected from the lower tertiary amide content in the former. As the positions of the maxima for both series corresponded well to that of oligoproline peptides, the PPII content could be approximated using Creamer's method. Both (PPR)<sub>n</sub> and (PRR)<sub>n</sub> peptides showed extensive, temperature-dependent structure that was nearly 50% PPII helical at 5 °C and retained approximately 20% PPII helicity at 90 °C (see Fig. 6). This figure shows the temperature dependence of PPII helicity, with the maximum mean residue ellipticity for

each peptide (100 micromolar in PBS, pH 7.4) being plotted as a function of temperature. PPII helicity is estimated from the maximum MRE. In comparison, a heptaproline peptide (P<sub>7</sub>GY) was 67% PPII helical, with single amino acid substitutions (PsXP<sub>3</sub>GY) reducing helicity to 49-66%. Fig. 3D illustrates circular dichroism spectra of 100 micromolar (PPR)<sub>n</sub> and (PRR)<sub>n</sub> peptides at 5 °C in PBS (phosphate-buffered saline), pH 7.4.

To test the ability of the (PRR)<sub>n</sub> motif to deliver cargos both capable of supporting functionality and dependent on PPII structure, they were incorporated into the miniature protein avian pancreatic polypeptide (aPP) (Fig. 4A). Fig. 4A shows the installation of (PRR)<sub>5</sub> in a miniature protein by sequence alignment. Previous work has shown that aPP is a robust platform for the design of miniature proteins that bind DNA and proteins, inhibiting their interactions with high affinity and specificity in vitro and in cells. The PPII helix of aPP was replaced with (PRR)<sub>5</sub> to create RR5-aPP (SEQ ID NO: 47) (Fig. 4A). As measured by flow cytometry (Fig. 4B), RR5-aPP penetrated cells as efficiently as an isolated (PRR)<sub>5</sub> motif. RR4-aPP (SEQ ID NO: 48) and RR3-aPP (SEQ ID NO: 49), with successively truncated N-terminal extensions, demonstrated slightly greater cellular uptake than RR5-aPP. Notably, RR3-aPP did not add additional residues to the native aPP sequence and was significantly more cell-permeable than arginine homopolymers of comparable arginine content. Fig. 4B shows HeLa cell uptake of 1 micromolar fluorescein-labeled miniature proteins after 1 h quantified by flow cytometry. This plot illustrates the mean cellular fluorescence  $\pm$  the standard error of three experiments.

Confocal microscopy confirms that, like the uptake motifs in isolation, these cell-permeable miniature proteins were intracellular and not sequestered in endosomes (Fig. 4D). In these figures, HeLa cells were incubated with 1 micromolar RR3-aPP for 20 min, and intracellular localization was monitored by confocal microscopy. The endosomes were visualized using 10 micromolar 10 kDa dextran labeled with AlexaFluor-647. RR3-aPP and RR4-aPP did not show discernible toxicity at 10 micromolar, while cells incubated with RR5-aPP remained 91% viable. None of the three was cytotoxic at 1 micromolar. Circular dichroism demonstrated that, although installation of the PPII-based CPP results in some loss of alpha-helicity, the miniature proteins retained significant structure and thermostability (Figs. 4C and 7). RR3-aPP, RR4-aPP, and RR5-aPP exhibited melting transitions at 44, 45, and 38 °C, respectively,

comparable to those of monomeric aPP molecules. The CD spectra were taken at 50 micromolar concentration of miniature proteins at 5 °C in PBS, pH 7.4. Fig. 7 shows the thermostability of RR3-aPP, with the mean residue ellipticity at 222 nm of RR3-aPP (50 micromolar in PBS, pH 7.4) plotted as a function of temperature. Reported melting  
5 temperatures correspond to the maximum of the first derivative of the MRE with respect to temperature. Fig. 4E shows two-color fluorescence with bright-field superposition.

In summary, this example illustrates the design of encodable, cell-permeable peptides possessing polyproline type-II structure. These motifs penetrated eukaryotic cells with efficiencies higher than Tat and previous natural PPII peptides, and also higher  
10 than oligoarginine peptides. These PPII-based motifs are not cytotoxic at concentrations 10-100 times greater than that necessary for uptake. Moreover, encoding the motifs within miniature proteins endowed cell permeability without destroying structure or adding significant mass. Thus, these motifs can facilitate both the delivery of peptides and proteins for intracellular study, as well as expand the utility of functionalized  
15 miniature proteins.

Additional experimental details follow. Peptide synthesis, labeling and purification: Peptides were synthesized at the 25 micromolar scale using standard solid-phase Fmoc chemistry on a Symphony<sup>®</sup> automated solid-phase synthesizer (Protein Technologies, Inc., Tucson, AZ). Fmoc-protected amino acid monomers, NovaSyn TGR  
20 and Rink amide resins were purchased from Novabiochem (San Diego, CA) and *N,N*-Dimethylformamide (DMF), *N*-methylmorpholine, piperidine, and trifluoroacetic acid (TFA) from American Bioanalytical (Natick, MA). N-terminally fluorescein-labeled peptides (Table 2) were synthesized as C-terminal carboxamides on Rink resin. Following synthesis, the resin was washed thoroughly with DMF, N-terminally  
25 deprotected with 20% piperidine in DMF and washed again with DMF. Fluorescein-5-ex succinimidyl ester (5 mg, Invitrogen, Carlsbad, CA) in 1 mL DMF with 10 microliter *N,N*-diisopropylethylamine (Sigma-Aldrich, St. Louis, MO) was added to each peptide and coupled in the dark and under nitrogen for 2-4 hours before cleavage from the resin. Tyrosine-labeled peptides (Table 3) and miniature proteins were synthesized on 25  
30 micromoles of Tentagel TGR resin as C-terminal carboxamides and N-terminally acetylated.

All peptides were purified by reverse-phase HPLC on Vydac C8 (300 Angstrom silica, 10 micrometer particle diameter) or C18 (300 Angstrom silica, 5 micrometer

particle diameter) preparative or semi-preparative columns over water/acetonitrile gradients containing 0.1% TFA. Peptide purity was verified by reinjection on C8 or C18 reverse phase analytical columns. Molecular masses (Tables 2 and 3) were quantified on an Applied Biosystems Voyager-DE-Pro MALDI-TOF mass spectrometer (Foster City, CA).

Table 2

|    | Name    | Sequence  | Calc Mass (Da) | Obs Mass (Da) |
|----|---------|---|----------------|---------------|
| 10 | SAP     | Flu-VRLPPPVRLPPP (SEQ ID NO: 76)                                | 2473           | 2479          |
|    | Bac     | Flu-GPRPLPFPRPG (SEQ ID NO: 77)                                 | 1666           | 1670          |
|    | Tat     | Flu-GRKKRRQRRPPQ (SEQ ID NO: 78)                                | 2195           | 2195          |
|    | R8      | Flu-GRRRRRRRR (SEQ ID NO: 79)                                   | 1800           | 1799          |
| 15 | R10     | Flu-GRRRRRRRRRR (SEQ ID NO: 80)                                 | 2112           | 2112          |
|    | R12     | Flu-GRRRRRRRRRRRR (SEQ ID NO: 81)                               | 2425           | 2425          |
|    | PPR3    | Flu-GPPRPPRPPR (SEQ ID NO: 82)                                  | 1602           | 1605          |
|    | PPR4    | Flu-GPPRPPRPPRPPR (SEQ ID NO: 83)                               | 1952           | 1956          |
|    | PPR6    | Flu-GPPRPPRPPRPPRPPR (SEQ ID NO: 84)                            | 2653           | 2656          |
| 20 | PRR3    | Flu-GPRRPRRPRR (SEQ ID NO: 85)                                  | 1779           | 1783          |
|    | PRR4    | Flu-GPRRPRRPRRPRR (SEQ ID NO: 86)                               | 2189           | 2193          |
|    | PRR5    | Flu-GPRRPRRPRRPRRPRR (SEQ ID NO: 87)                            | 2598           | 2599          |
|    | PRR6    | Flu-GPRRPRRPRRPRRPRRPRR (SEQ ID NO: 88)                         | 3008           | 3011          |
| 25 | aPP     | GPSQPTYPGDD APVEDLIRFYNDLQQYLNVVTRHRYC-FIU<br>(SEQ ID NO: 89)   | 4771           | 4773          |
|    | RR3-aPP | GPRRPRRPGRRAPVEDLIRFYNDLQQYLNVVTRHRYC-Flu<br>(SEQ ID NO: 90)    | 4999           | 5002          |
|    | RR4-aPP | RRPRRPRRPGRRAPVEDLIRFYNDLQQYLNVVTRHRYC-Flu<br>(SEQ ID NO: 91)   | 5254           | 5256          |
| 30 | RR5-aPP | RRPRRPRRPRRPGRRAPVEDLIRFYNDLQQYLNWTRHRYC-Flu<br>(SEQ ID NO: 92) | 5663           | 5670          |



Table 3

|    | <u>Name</u>          | <u>Sequence</u>   | <u>Calc</u> | <u>Obs</u>  |
|----|----------------------|---|-------------|-------------|
|    |                      |   | <u>Mass</u> | <u>Mass</u> |
|    |                      |   | <u>(Da)</u> | <u>(Da)</u> |
| 5  | PPR <sub>3</sub>     | PPRPPRPPRY (SEQ ID NO: 93)                                      | 1275        | 1 273       |
|    | PPR <sub>4</sub>     | PPRPPRPPRPPRY (SEQ ID NO: 94)                                   | 1625        | 1624        |
|    | PPR <sub>5</sub>     | PPRPPRPPRPPRPPRY (SEQ ID NO: 95)                                | 1975        | 1974        |
|    | PPR <sub>6</sub>     | PPRPPRPPRPPRPPRPPRY (SEQ ID NO: 96)                             | 2326        | 2326        |
|    | PRR <sub>3</sub>     | PRRPRRPRRY (SEQ ID NO: 97)                                      | 1452        | 1450        |
| 10 | PRR <sub>4</sub>     | PRRPRRPRRPRRY (SEQ ID NO: 98)                                   | 1861        | 1861        |
|    | PRR <sub>5</sub>     | PRRPRRPRRPRRPRRY (SEQ ID NO: 99)                                | 2271        | 2271        |
|    | PRR <sub>6</sub>     | PRRPRRPRRPRRPRRPRRY (SEQ ID NO: 100)                            | 2680        | 2682        |
|    | aPP                  | GPSQPTYPGDD APVEDLIRF YNDLQQ YLNVVTRHRYC-<br>(SEQ ID NO: 101)   | AC4441      | 4437        |
| 15 | RR <sub>3</sub> -aPP | GPRRPRRPGRRAP VEDLIRF YNDLQQ YLNV VTRHRYC-<br>(SEQ ID NO: 102)  | Ac4668      | 4665        |
|    | RRi-aPP              | RRPRRPRRPGRRAPVEDLIRFYNDLQQYLNVVTRHRYC-Ac<br>(SEQ ID NO: 103)   | 4924        | 4923        |
|    | RR <sub>5</sub> -aPP | RRPRRPRRPRRPGRRAPVEDLIRFYNDLQQYLNWTRHRYC-Ac<br>(SEQ ID NO: 104) | 5333        | 5334        |
| 20 |                      |   |             |             |

Following purification, the C-terminal cysteine of aPP-based peptides was either fluorescein-labeled or capped with iodoacetamide to prevent disulfide formation. Fluorescein labeling was carried out by dissolving peptide to approximately 0.3 mM in 1:5 DMF:H<sub>2</sub>O with 10 mM phosphate buffer, pH 7.25. Iodoacetamidofluorescein (10 eq., 50 mM) in DMF was then added and allowed to react for 1-2 hours at room temperature. Acetamide labeling was conducted by reacting 20 equivalents of iodoacetamide (10 micromoles) with peptide (500 nmol, 0.2-0.5 mM) in PBS buffer, pH 7.4 for 1 hour at room temperature. Both reactions were stopped by flash freezing to -80 °C and stored frozen until purified by reverse-phase HPLC. Non-fluorescent peptide concentrations were calculated using tyrosine absorption in 6 M guanidinium hydrochloride, 20 mM sodium phosphate, pH 6.5, while fluorescent peptide

concentrations were determined using fluorescein absorption in 8 M urea, 100 mM Tris-HCl, pH 9.0.

Flow cytometry: The uptake of fluorescently labeled peptides by live HeLa cells was quantified by flow cytometry. HeLa cells (American Type Culture Collection, Manassas, VA) were grown in T-75 culture flasks containing High-Glucose Dulbecco's Modified Eagle Medium (DMEM) (Invitrogen # 11995) supplemented with 10% fetal bovine serum (Invitrogen) (DMEM<sup>+</sup>) to ~80% confluency, washed twice with 37 °C PBS (140 mM sodium chloride, 3 mM potassium chloride, 10 mM phosphate, pH 7.4), and incubated with 10 mL of PBS-based non-enzymatic cell dissociation solution (Chemicon International, Temecula, CA) for 15 minutes. Cells were then collected at 500g, resuspended in DMEM<sup>+</sup>, counted by hemocytometer and diluted to 2083 cells/microliter with DMEM<sup>+</sup>. Aliquots of cells (240 microliters) were then added to fluorescein-labeled peptides (10 microliters, 25 micromolar in water) to give 2-5 x 10<sup>5</sup> cells in DMEM<sup>+</sup> containing 1 micromolar peptide. Cells were incubated with peptide for 60 minutes at 37 °C before washing twice with 500 microliters 37 °C PBS to remove extracellular peptide. To ensure removal of any surface-bound peptide, the cells were then incubated with 0.05% trypsin at 37 °C for 10 minutes, washed with DMEM<sup>+</sup> and then 4 °C PBS. Finally, the cells were resuspended in 500 microliters PBS containing 1 microgram/mL propidium iodide, and analyzed on a BD FACScan (BD Biosciences, San Jose, CA) equipped with a 488 nm Argon laser.

A total of 30,000 events were collected, monitoring fluorescein and propidium iodide with 530/30 bandpass and 650 longpass filters, respectively. Events corresponding to cellular debris were removed by gating on forward and side scatter, while dead cells are removed by propidium iodide staining. Geometric means were then calculated from the histogram of fluorescence intensity, and corrected for background cellular fluorescence by subtracting the geometric mean of mock-treated cells. The standard error for the average of geometric means for three separate experiments is reported.

Confocal microscopy: To verify the internalization of fluorescently labeled peptides and determine their intracellular localization, live cell confocal microscopy was employed. Approximately 10<sup>5</sup> HeLa cells were seeded in 2 mL DMEM<sup>+</sup> into 6-well plates containing cover glasses. After allowing the cells to adhere for at least 24 hours, media was removed by aspiration and the cells washed twice with 37 °C PBS.

Incubation was performed by floating inverted cover glasses on 200 microliters of DMEM<sup>+</sup> containing 1 micromolar peptide and/or 10 micromolar 10 kDa dextran labeled with AlexaFluor 647 (Invitrogen) for 20-30 minutes at 37 °C. Cover glasses were then washed with 37 °C DMEM<sup>+</sup>, PBS, and PBS containing 1 microgram/mL anti-fluorescein rabbit polyclonal IgG antibody (Invitrogen #A889) to quench extracellular fluorescence. After mounting the cover glasses on microscope slides, cells were imaged on an LSM 510 Meta (Carl Zeiss Microimaging, Thornwood, NY), using a 488 nm Ar laser line with a 525/25 nm filter or 633 nm HeNe laser line with a 680/30 filter for visualizing fluorescein and AlexaFluor 647, respectively.

Cellular viability assays: To determine the potential toxicity of the cell-penetrating peptides, the reduction of resazurin to resorufin was monitored by fluorescence using the CellTiter-Blue<sup>®</sup> Cell Viability Assay (Promega, Madison, WI). HeLa cells (5000 cells/well) in DMEM<sup>+</sup> were seeded into 96-well plates and allowed to adhere overnight. Non-fluorescent peptides (12.5 microliters) were then added at a concentration appropriate to bring the well to 1, 10 or 100 micromolar peptide. After 6 hours at 37 °C, CellTiter-Blue<sup>®</sup> reagent (20 microliters in 40 microliters DMEM<sup>+</sup>) was added and cells incubated for another 2 hours. Fluorescence was monitored on an Analyst<sup>™</sup> AD 96-384 fluorescence plate reader (LJL Biosystems) using 530/25 excitation and 580/10 emission filters. Fluorescence values were then corrected for CellTiter-Blue<sup>®</sup> background fluorescence by subtracting the reading of parallel-processed cell-free wells. Cell viability was calculated as the percentage of peptide-treated cells to buffer-treated cells. As a control, standard curves (created by seeding 0-5000 cells/well) were used to corroborate the linearity of this assay. The mean viability with standard deviation of four independent experiments, each containing at least three replicates, was reported (Table 1, Fig. 5).

Circular dichroism: Structural characterization of non-fluorescent peptides was accomplished by acquiring circular dichroism (CD) spectra on a Jasco J-810 Spectropolarimeter (Jasco, Tokyo, Japan) equipped with a Peltier temperature-control module. All peptides were characterized in PBS, pH 7.4 using 100 micromolar model peptides or 50 micromolar miniature proteins in a 2 mm cuvette. Wavelength-dependent spectra were acquired as the average of five scans at 5 °C from 260 to 190 nm at 100 nm/min, sampling every 0.5 nm with a 1 nm bandwidth. Mean residue ellipticity (MRE) was calculated as  $MRE = (\Theta_{peptide} - \Theta_{buffer}) / (\ell \times c \times r)$ , where  $\Theta$  is the observed signal in

degrees,  $L$  is the path length in cm,  $c$  is the concentration of peptide in  $\text{dmol}/\text{cm}^3$  and  $n$  is the number of residues.

To chart the temperature-dependence of PPII structure, abbreviated wavelength scans were acquired as the average of scans from 235-210 nm at 5 °C and 10° intervals  
5 from 10-90 °C. Data collection parameters were otherwise identical to those described above. MRE values at the wavelength corresponding to maximal ellipticity at 5 °C were plotted as a function of temperature. The positions of the peak closely match those reported elsewhere, and the percentage of PPII helicity (%PPII) was estimated as follows: %PPII =  $((\text{MREMAX} + 6100)/13,700) \times 100$ , where MREMAX is the  
10 maximum mean residue ellipticity of the PPII peak in this region (Table 1).

To determine the temperature dependence and melting temperature ( $T_m$ ) of aPP-based molecules, the MRE was monitored at 222 nm ( $\text{MRE}_{222}$ ) from 5 °C to 90 °C. Data were collected with a 1.0 °C data pitch, 10 s delay time, 60 °C/hour temperature slope and 1 nm bandwidth. Melting temperatures were determined from the maximum of the  
15 first derivative of the mean residue ellipticity ( $\delta\text{MRE}_{222}/\delta T$ ), and were 44, 45 and 38 °C for RR3-aPP, RH4-aPP, and RR5-aPP, respectively.

## EXAMPLE 2

This example illustrates that a smaller cationic motif can be embedded within the  
20 alpha-helix of a small, folded protein to generate molecules that penetrate cells significantly more efficiently than oligoarginine-rich sequences or Tat. These results suggest that the function of cell permeability can be encoded by judicious placement of as few as 2-3 additional arginine residues on a protein alpha-helix.

The avian pancreatic polypeptide (aPP) is a 36-residue polypeptide composed of  
25 an N-terminal PPII helix and a C-terminal alpha-helix. See Fig. 8A; residues that contribute to aPP/YY2 folding are indicated by brackets, and arginines located on the alpha-helices are indicated by underlining. This well-packed, thermostable structure provides a starting point for the presentation of well-folded PPII and alpha-helical epitopes that bind protein targets with high affinity and specificity *in vitro* and *in vivo*. It  
30 was reasoned that substitution of arginines for residues located on the aPP alpha-helix would impart cell permeability, but perhaps only at the expense of structural stability due to charge-charge repulsion. To determine the minimum number of arginine residues that would facilitate cell permeability and retain a stable fold, aPP variants aPP<sup>4R1</sup> (SEQ ID

NO: 23), aPP<sup>5R1</sup> (SEQ ID NO: 22), and aPP<sup>6R1</sup> (SEQ ID NO: 21) were synthesized (Fig. 8). These molecules contained four, five, or six arginines, respectively, on the solvent-exposed face of the aPP alpha-helix in place of side chains that contribute minimally, if at all, to protein stability. See Table 4.

5

Table 4

|                       |   |
|-----------------------|---|
| aPP                   | GPSQPTYPGDDAPVEDLIRFYNDLQQYLNVVTRHRY<br>(SEQ ID NO: 1)  |
| aPP <sup>4Ri</sup>    | GPSQPTYPGDDAPVRDLIRFYRDLQRYLNVVTRHRY<br>(SEQ ID NO: 23) |
| 10 aPP <sup>5Ri</sup> | GPSQPTYPGDDAPVRDLIRFYRDLRRYLNVVTRHRY<br>(SEQ ID NO: 22) |
| aPP <sup>6Ri</sup>    | GPSQPTYPGDDAPVRDLRRFYRDLRRYLNVVTRHRY<br>(SEQ ID NO: 21) |
| 15 YY2                | APPLPPNRGEDASPEELSRYYASLRHYLNLVTRQRY<br>(SEQ ID NO: 67) |
| YY2 <sup>3R1</sup>    | APPLPPNRGEDASPEELSRYYRSLRHYLNLVTRQRY<br>(SEQ ID NO: 50) |
| YY2 <sup>4R1</sup>    | APPLPPNRGEDASPRELSRYYRSLRHYLNLVTRQRY<br>(SEQ ID NO: 51) |
| 20 YY2 <sup>5R1</sup> | APPLPPNRGEDASPRELRRYYRSLRHYLNLVTRQRY<br>(SEQ ID NO: 52) |
| R <sub>8</sub> Y      | RRRRRRRRY<br>(SEQ ID NO: 117)                           |
| 25 Tat (48-60)        | GRKKRRQRRRPPQY<br>(SEQ ID NO: 118)                      |
| Penetratin            | RQIKIWFQNRRMKWKK<br>(SEQ ID NO: 119)                    |
| Transportan           | GWTLNSAGYLLKINLKALAALAKKIL<br>(SEQ ID NO: 120)          |
| 30 KLA                | KLALKLALKALKAALKLA<br>(SEQ ID NO: 121)                  |

The secondary structure of each variant was characterized using circular dichroism (CD) spectroscopy (Fig. 8B). This figure shows circular dichroism spectra of aPP, aPP<sup>4RI</sup>, aPP<sup>5RI</sup>, and aPP<sup>6RI</sup> (10 micromolar) in PBS (pH 7.4) at 37 °C. The CD spectra of aPP<sup>4RI</sup> and aPP<sup>5RI</sup> were virtually identical to that of aPP, with characteristic negative ellipticity at 208 and 222 nm. By contrast, the spectrum of aPP<sup>6RI</sup> indicated significantly less alpha-helix structure. Temperature-dependent CD analysis demonstrated that each peptide, like aPP, underwent a cooperative melting transition. Each arginine substitution decreases thermostability slightly, with aPP<sup>4RI</sup> and aPP<sup>5RI</sup> characterized by  $T_m$ 's of 54 and 47 °C, near that of aPP; the  $T_m$  of aPP<sup>6RI</sup> is significantly depressed (35 °C) (Fig. 11). Referring now to Fig. 11, which shows circular dichroism analysis of miniature protein thermostability, Fig. 11A illustrates the temperature dependence of the negative ellipticity at 222 nm of the indicated peptides (10 micromolar) in PBS buffer (pH 7.4). Fig. 11B shows the first derivative of MRE<sub>222 nm</sub> with respect to temperature (units are deg cm<sup>2</sup> dmol<sup>-1</sup> K<sup>-1</sup>).

Next, flow cytometry was used to determine whether aPP<sup>4RI</sup>, aPP<sup>5RI</sup>, and/or aPP<sup>6RI</sup> penetrated eukaryotic cells. HeLa cells were incubated with 1 micromolar fluorescein-labeled peptide for times between 5 min and 90 min and treated with trypsin to remove peptide from the cell surface. Cellular penetration of Flu<sub>a</sub>PP<sup>5RI</sup> (SEQ ID NO: 106), Flu<sub>a</sub>PP<sup>6RI</sup> (SEQ ID NO: 107), and the well-studied oligoarginine Flu<sub>R</sub><sub>8</sub>Y (SEQ ID NO: 112) were similar at early times ( $t < 20$  min) (Fig. 9A). Fig. 9A shows the time-dependent uptake of 1 micromolar fluorescein-labeled peptides by HeLa cells as quantified by flow cytometry. This plot illustrates the mean cellular fluorescence  $\pm$  the standard error of three experiments.

Surprisingly, uptake of Flu<sub>R</sub><sub>8</sub>Y remained constant at longer times whereas uptake of Flu<sub>a</sub>PP<sup>5RI</sup> and Flu<sub>a</sub>PP<sup>6RI</sup> increased significantly. At an incubation time of 30 min, Flu<sub>a</sub>PP<sup>4RI</sup> (SEQ ID NO: 105) entered cells more efficiently than Flu<sub>a</sub>Tat (SEQ ID NO: 113) but less efficiently than Flu<sub>R</sub><sub>8</sub>Y, Flu<sub>p</sub>Antp (SEQ ID NO: 114), Flu<sub>T</sub>Transportan (SEQ ID NO: 115), and Flu<sub>K</sub>LA (SEQ ID NO: 116) (also known as MAP) (Fig. 8C). This figure shows the mean cellular fluorescence of HeLa cells incubated for 30 min with 1 micromolar fluorescently labeled peptide in DMEM containing 10% FBS. By contrast, Flu<sub>a</sub>PP<sup>5RI</sup> and Flu<sub>a</sub>PP<sup>6RI</sup> penetrated cells significantly more effectively than any cell penetrating peptide tested, generating cellular fluorescence 3-4 times that of Flu<sub>R</sub><sub>8</sub>Y and 35-45 times that of Flu<sub>T</sub>Tat. Confocal microscopy confirmed that Flu<sub>a</sub>PP<sup>6RI</sup> was

internalized to the cytosol and endosomes and was not limited to the cellular membrane (Fig. 9C).

The mechanism by which cationic peptides enter cells is not well understood and may vary depending on sequence, conditions, and cargo. Recent results obtained with  
5 live cells imply that endocytosis is often a major pathway. Indeed, cellular uptake of  $\text{Flu}_{\text{a}}\text{P}^{\text{SRi}}$  and  $\text{Flu}_{\text{a}}\text{P}^{\text{6Ri}}$  was both temperature- and ATP-dependent (Fig. 9B). This figure shows HeLa cells were incubated with 1 micromolar peptide for 30 min at 37 °C, 4 °C, and at 37 °C in glucose-free media supplemented with  $\text{NaN}_3$ , and intracellular localization was quantified by flow cytometry.

10 Moreover, confocal microscopy of cells treated with  $\text{Flu}_{\text{a}}\text{PP}^{\text{6R1}}$  revealed a punctate pattern of fluorescence in the cytosol (Fig. 9C) that colocalizes in part with 10 kDa dextran, a marker for endocytosis (Figs. 9D and 9E). In particular, this figure shows confocal microscopy of HeLa cells co-incubated with 5 micromolar  $\text{Flu}_{\text{a}}\text{PP}^{\text{6R1}}$  (Green) and 10 kDa dextran labeled with Alexa Fluor 647 (red) for 30 min at 37 °C. Fig. 9C  
15 illustrates the green channel, while Fig. 9D illustrates the red channel and Fig. 9E is the superposition of Figs. 9C, 9D, and transmitted light. Together these data suggest a significant contribution from an energy-dependent mechanism such as endocytosis. However,  $\text{Flu}_{\text{a}}\text{pp}^{\text{6R1}}$   $\text{i}_{\text{so}}$  shows diffuse cytosolic staining that is not colocalized with dextran, suggesting either endosomal escape or an alternative entry pathway.

20 Oligoarginine tags can increase cytotoxicity, and thus the toxicity of all peptides was examined using the CellTiter-Blue cell viability assay (Promega). Incubation of HeLa cells with any of the aPP- or YY2-based peptides at concentrations as high as 50 micromolar (50-fold higher than required for penetration) for 6 h (70-fold longer than necessary for uptake) led to less than 10% loss in viability (Fig. 12). In this figure, Fig.  
25 12A is a circular dichroism analysis of YY2 and arginine-rich variants (10 micromolar, 37 °C) in PBS buffer (pH 7.4). Fig. 12B illustrates the temperature dependence of  $\text{MRE}_{222}$ , while Fig. 12C illustrates the first derivative of  $\text{MRE}_{222}$  with respect to temperature (units are  $\text{deg cm}^2 \text{ dmol}^{-1} \text{ K}^{-1}$ ).

Example 1 describes a series of miniature proteins that made use of their PPII  
30 helix to bind SH3 domains selectively *in vitro*; one such molecule, YY2, activated Hck kinase in cell extracts. To evaluate whether a minimal cationic motif could be transferred from aPP to an alternative protein context,  $\text{YY2}^{\text{3R1}}$ ,  $\text{YY2}^{\text{4R1}}$ , and  $\text{YY2}^{\text{5R1}}$  were synthesized (Fig. 8A). Although  $\text{Flu}_{\gamma} \gamma^{\text{23R1}}$  (SEQ ID NO: 109) did not penetrate

cells appreciably at 1 micromolar,  $^{\text{Flu}}\gamma\gamma 2^{4\text{R}1}$  (SEQ ID NO: 110) and  $^{\text{Flu}}\gamma\gamma 2^{5\text{R}1}$  (SEQ ID NO: 111) did so at levels similar to  $^{\text{Flu}}\alpha\text{PP}^{6\text{R}1}$  (SEQ ID NO: 107), the most cell-permeant aPP derivative studied (Fig. 8C). Unexpectedly, even  $^{\text{Flu}}\text{YY}2$  (SEQ ID NO: 108), whose sequence contains only two arginine residues within the alpha-helix, penetrates cells, albeit at higher concentration. At 6 micromolar,  $^{\text{Flu}}\text{YY}2$  shows intracellular fluorescence twice that of  $^{\text{Flu}}\text{R}_8\text{Y}$  (Fig. 10A). This figure shows HeLa cell uptake of fluorescein-labeled peptides at the indicated concentration after 30 min at 37 °C as quantified by flow cytometry, and this observation highlights the expanding utility of miniature proteins, which, like certain peptoids, may not require additional engineering to offer cell permeability. In Fig. 10A, the figure is a plot illustrating the mean cellular fluorescence (the standard error of three experiments. Fig. 10B is a confocal microscopy image of HeLa cells incubated with 5 micromolar  $^{\text{Flu}}\text{YY}2$  for 30 min at 37 °C.

In summary, this example illustrates the creation of a second family of minimally cationic miniature proteins that effectively cross the plasma membrane of eukaryotic cells. Although introduction of multiple arginine residues decreases the value of  $Tu$ , judicious placement and empirical charge minimization afford miniature proteins that are both well-folded and cell permeable. Thus, these scaffolds are capable of supporting high affinity and specificity interactions *in vitro* and *in vivo*.

Additional experimental details follow. Peptide synthesis: All peptides were synthesized using standard solid-phase Fmoc chemistry on a 25 micromole scale with a Symphony<sup>®</sup> multi-channel solid phase synthesizer (Protein Technologies, Inc., Tuscon, AZ). All alpha-amino acids and resins were purchased from Novabiochem (San Diego, CA) and solvents were purchased from American Bioanalytical (Natick, MA). All peptides were synthesized to carry free amines at their N-termini and carboxamides at their C-termini. Peptides were labeled on their N-termini on resin for at least one hour with fluorescein-5-EX, succinimidyl ester (Invitrogen, Carlsbad, CA, Cat. # F-6130), which was dissolved in 1 mL DMF and added to the reaction vessel with 20 microliters *N,N*-diisopropylethylamine (EDIPA).

Purification: The peptides were purified by reverse-phase HPLC using Grace Vydac C8 preparative or semi-preparative scale columns (300 Angstrom silica, 10 micrometer particle size, 22 mm x 250 mm) and water/acetonitrile gradients containing 0.1% TFA. Peptide identity was confirmed by mass spectrometry on an Applied



Biosystems Voyager-DE Pro MALDI-TOF mass spectrometer (Foster City, CA). Once purified, peptides were lyophilized to dryness, dissolved in water and stored at -20 °C. Labeled peptides were protected from light.

Concentration determination: The concentration of each fluorescently labeled peptide or miniature protein stock solution was determined by measuring the fluorescein absorption at 500 nm in 8 M urea, 100 mM Tris-HCl (pH 9.0). Concentration was determined using the following formula:  $\text{Concentration (M)} = A_{500} / (C_{500} L)$ , where  $A_{500}$  is the absorbance at 500 nm,  $C_{500}$  is the molar extinction coefficient at 500 nm (86000 M<sup>-1</sup>cm<sup>-1</sup> (Invitrogen)), and  $L$  is the path length (1 cm). The concentration of the unlabeled KLA peptide was calculated by the mass to volume ratio. The concentrations of the other non-labeled peptides were calculated by monitoring the absorbance at 280 nm in 8 M urea, 100 mM Tris-HCl (pH 7.4). Concentration was determined using the following formula:  $\text{Concentration (M)} = A_{280} / (G_{280} L)$ , where  $A_{280}$  is the absorbance at 280 nm,  $G_{280}$  is the molar extinction coefficient at 280 nm, calculated for each peptide sequence), and  $z$ , is the path length (1 cm). Table 5 shows calculated masses of labeled and unlabeled peptides and miniature proteins and masses observed by MALDI-TOF mass spectrometry.

Table 5

|    | Unlabeled          |               | Labeled         |          |        |
|----|--------------------|---------------|-----------------|----------|--------|
|    | Calculated (Da)    | Observed (Da) | Calculated (Da) | Observed |        |
|    | (Da)               |               |                 |          |        |
|    | aPP                | 4238.6        | 4244.1          | 4712.1   | 4718.7 |
|    | aPP <sup>4R1</sup> | 4335.8        | 4340.2          | 4811.3   | 4819.0 |
| 25 | aPP <sup>5R1</sup> | 4363.9        | 4357.7          | 4839.4   | 4847.0 |
|    | aPP <sup>6R</sup>  | 4406.9        | 4410.5          | 4882.4   | 4889.3 |
|    | YY2                | 4386.9        | 4385.4          | 4862.4   | 4855.2 |
|    | YY23R1             | 4472.0        | 4476.5          | 4947.5   | 4950.6 |
|    | γγ <sup>24R1</sup> | 4491.1        | 4496.4          | 4974.6   | 4979.7 |
| 30 | YY2 <sup>5R1</sup> | 4568.2        | 4563.6          | 5043.7   | 5051.4 |
|    | R <sub>8</sub> Y   | 1430.7        | 1429.2          | 1906.2   | 1910.0 |
|    | Tat                | 1882.2        | 1884.9          | 2357.7   | 2363.2 |
|    | Penetratin         | 2246.7        | 2250.9          | 2722.2   | 2726.9 |

|             |        |        |        |        |
|-------------|--------|--------|--------|--------|
| Transportan | 2784.4 | 2787.7 | 3259.9 | 3265.2 |
| KLA         | 1876.5 | 1877.7 | 2352.0 | 2351.8 |

Circular dichroism: Circular dichroism spectra were measured on a Jasco J- 810-  
 5 150S Spectropolarimeter (Jasco, Tokyo, Japan) equipped with a Peltier temperature-  
 control module and analyzed using Spectra Manager software v.1.53.01 . Spectra were  
 acquired at peptide concentrations of 10 micromolar in PBS (pH 7.4) in a 2 mm cuvette.  
 Wavelength-dependent spectra were acquired from 260 nm to 190 nm (data pitch 0.5 nm,  
 scan speed 50 nm/min, 4 sec, 1 nm bandwidth and 3 accumulations). Thermal  
 10 denaturation was monitored at 222 nm as the temperature was raised from 5 °C to 85 °C  
 using the variable temperature model (data pitch 1 °C, 5 second delay, temperature slope  
 1 °C/min, 4 sec response, 1 nm bandwidth, continuous scan mode). Renaturation from  
 85 °C to 5 °C was then performed to verify that the melting transitions were reversible.  
 Mean residue ellipticity values were calculated from the equation  $MRE = (\Theta_{\text{sample}} -$   
 15  $\Theta_{\text{buffer}})/(-\epsilon c l)$  where  $\Theta$  is observed signal in millidegrees,  $L$  is the length of the cuvette  
 in cm,  $c$  is the concentration of peptide in dmol/cm<sup>3</sup>, and  $n$  is the number of amino acid  
 residues in the peptide. The melting temperature ( $T_m$ ) was estimated by the inflection  
 point of the melt, as determined by the peak of the graph of the first derivative of the  
 curve (calculated by Spectra Manager software). Melting temperatures for aPP, aPP<sup>4R1</sup>,  
 20 aPP<sup>5R1</sup>, aPP<sup>6R1</sup>, YY2, and YY2<sup>3R1</sup> are presented in Table 6, showing observed melting  
 temperature as determined by circular dichroism at 10 micromolar in PBS, pH 7.4.  
 Cooperative melting transitions were not found for YY2<sup>4R1</sup> or YY2<sup>5R1</sup> (Fig. 12).

Table 6

| 25 | Peptide            | $T_m$ (°C) |
|----|--------------------|------------|
|    | aPP                | 62         |
|    | aPP <sup>4R1</sup> | 54         |
|    | aPP <sup>5R1</sup> | 49         |
|    | aPP <sup>6R1</sup> | 35         |
| 30 | YY2                | 21         |
|    | YY2 <sup>3R1</sup> | 20.5       |

Flow cytometry: HeLa cells (American Type Culture Collection, Manassas, VA) were grown at 37 °C in an atmosphere containing 5% CO<sub>2</sub> in an IR Autoflow CO<sub>2</sub> water incubator (Nuair, Plymouth, MN). Cells were grown in T-75 culture flasks (BD Falcon, Cat. #353134) containing High-Glucose Dulbecco's Modified Eagle Medium (Gibco Cat. # 11995-065) containing 10% Fetal Bovine Serum (Gibco Cat. # 26140-079), hereinafter called DMEM\*. Cells were passaged every 5-7 days before reaching confluency, as determined by visual inspection. For flow cytometry experiments, four flasks of cells were grown to ~90% confluency, the media was removed, and the cells were washed twice with 10 mL/flask 37 °C PBS (140 mM NaCl, 3 mM KCl, 10 mM phosphate, pH 7.4). The cells were dissociated from the flask surface by incubation for 15 min with 5 mL/flask Non-Enzymatic Cell Dissociation Solution (Sigma, Cat. # C5914) at 37 °C. Cells were collected at 500g, resuspended in 5 mL DMEM\*, and counted by hemocytometer. Cells at a maximum concentration of 2000 cells/microliter were added in aliquots of 240 microliters to aliquots of 10 microliters peptides in water and incubated 30 min at 37 °C. Cells were collected at 500g, washed twice with 500 microliters 37 °C PBS and incubated at 37 °C with 500 microliters of 0.05% (vv/v) trypsin to remove any surface-bound peptide. Cells were collected at 500g, washed with 500 microliters 4 °C PBS, and were resuspended in 500 microliters 4 °C PBS containing 1 microgram/mL propidium iodide (Sigma-Aldrich Cat. # P4864). Flow cytometry was performed on a BD FACScan (Becton Dickinson, Franklin Lakes, NJ, settings: FSC E-I 7 Lin, SSC 300 1.8 Lin, F1 520 Log, F2 288 Log, F3 500 Log), collecting 10,000 events and monitoring fluorescein and propidium iodide fluorescence with 530/30 bandpass and 650 longpass filters, respectively. Analysis was performed using FlowJo software (v6.4.7).

Cell debris and dead cells were gated out first by forward- and side-scatter and then by propidium iodide fluorescence. The geometric mean of the fluorescein fluorescence of the remaining cells was calculated. The standard error, as calculated from at least three independent trials on different days, is reported. Experiments examining the effect of temperature and energy depletion were conducted in media that did not contain FBS. Cells (2000 cells/microliters) were incubated in 4 °C DMEM, 37 °C DMEM, or 37 °C DMEM containing 10 mM NaN<sub>3</sub> and 50 mM 2-deoxy-D-glucose, for 30 minutes prior to the addition of peptides.

Confocal microscopy: Live cell fluorescence microscopy was employed to confirm the internalization of fluorescently labeled peptides and determine their intracellular localization. HeLa cells were grown as described above. The day prior to experiments,  $10^5$  cells/well were seeded into 6-well plates containing microscope cover slips and 3 mL DMEM\*. After allowing the cells to adhere for at least 24 hours, the media was replaced and to the cells was added 5 micromolar of a fluorescently labeled peptide/miniature protein and/or 5 micromolar 10 kDa dextran labeled with AlexaFluor 647 (Invitrogen) in DMEM\*. The cells were incubated for 30 min, washed for 15 min in DMEM\*, and rinsed with PBS. The cover slips were mounted onto microscope slides immediately before imaging on a LSM 510 Meta (Carl Zeiss Microimaging, Thornwood, NY), visualizing fluorescein with a 488 nm Ar laser with a 525/25 nm bandpass filter and AlexaFluor 647 with a 633 nm HeNe laser with a 680/30 bandpass filter.

Cell viability: HeLa cells were grown as described above. The day prior to initiating experiments, cells were removed with 0.05% (w/v) trypsin (Gibco Cat. # 25200-056), and 5000 cells/well were seeded into 96-well, black surface/clear flat bottom plates (Corning Cat. # 3603) in 45 microliters DMEM\*. Control wells containing between 0 and 5000 cells/well were also seeded. After allowing the cells to adhere for at least 24 h, non-fluorescent peptides were added (5 microliters) to a final volume of 50 microliters and final concentrations of 10 or 50 micromolar. After 6 h at 37 °C, 70 microliters of a 5:2 mixture of DMEM\* to CellTiter-Blue<sup>®</sup> (Promega Cat. # 6808A) was added to each well (final volume = 120 microliters). The cells were incubated for an additional 2 h at 37 °C and the fluorescence measured on an Analyst<sup>™</sup> AD 96-384 spectrofluorimeter (Molecular Devices, Sunnyvale, CA) using 530/25 excitation and 580/10 emission filters. Each reading was corrected to eliminate the contribution from CellTiter-Blue<sup>®</sup> background fluorescence by subtracting the fluorescence of cell-free wells from the value recorded. All experiments were performed in triplicate. The percent of viable cells (%V) was calculated as the percentage of peptide treated-cells to that of parallel-processed water-treated cells according to the equation:  $\% V = (F_{\text{peptide}} - F_{\text{blank}}) / (F_{\text{water}} - F_{\text{blank}})$  where  $F_{\text{peptide}}$  is the fluorescence of peptide-treated cells,  $F_{\text{water}}$  is the fluorescence of water-treated cells, and  $F_{\text{blank}}$  is the fluorescence of CellTiter-Blue<sup>®</sup>-treated wells that lacked cells (cell-free wells). Each plate was read three times and the readings averaged. Standard error, as calculated from three separate trials on three different days, is reported in Fig. 13, showing the percent of viable cells

(% V) after 6 h incubation at 37 °C with the indicated peptide or miniature protein at either 10 micromolar (lighter bars) or 50 micromolar (darker bars). Error bars represent standard error.

5           While several embodiments of the present invention have been described and illustrated herein, those of ordinary skill in the art will readily envision a variety of other means and/or structures for performing the functions and/or obtaining the results and/or one or more of the advantages described herein, and each of such variations and/or modifications is deemed to be within the scope of the present invention. More generally,  
10 those skilled in the art will readily appreciate that all parameters, dimensions, materials, and configurations described herein are meant to be exemplary and that the actual parameters, dimensions, materials, and/or configurations will depend upon the specific application or applications for which the teachings of the present invention is/are used. Those skilled in the art will recognize, or be able to ascertain using no more than routine  
15 experimentation, many equivalents to the specific embodiments of the invention described herein. It is, therefore, to be understood that the foregoing embodiments are presented by way of example only and that, within the scope of the appended claims and equivalents thereto, the invention may be practiced otherwise than as specifically described and claimed. The present invention is directed to each individual feature,  
20 system, article, material, kit, and/or method described herein. In addition, any combination of two or more such features, systems, articles, materials, kits, and/or methods, if such features, systems, articles, materials, kits, and/or methods are not mutually inconsistent, is included within the scope of the present invention.

25           All definitions, as defined and used herein, should be understood to control over dictionary definitions, definitions in documents incorporated by reference, and/or ordinary meanings of the defined terms.

          The indefinite articles "a" and "an," as used herein in the specification and in the claims, unless clearly indicated to the contrary, should be understood to mean "at least one."

30           The phrase "and/or," as used herein in the specification and in the claims, should be understood to mean "either or both" of the elements so conjoined, i.e., elements that are conjunctively present in some cases and disjunctively present in other cases. Multiple elements listed with "and/or" should be construed in the same fashion, i.e., "one

or more" of the elements so conjoined. Other elements may optionally be present other than the elements specifically identified by the "and/or" clause, whether related or unrelated to those elements specifically identified. Thus, as a non-limiting example, a reference to "A and/or B", when used in conjunction with open-ended language such as "comprising" can refer, in one embodiment, to A only (optionally including elements other than B); in another embodiment, to B only (optionally including elements other than A); in yet another embodiment, to both A and B (optionally including other elements); etc.

As used herein in the specification and in the claims, "or" should be understood to have the same meaning as "and/or" as defined above. For example, when separating items in a list, "or" or "and/or" shall be interpreted as being inclusive, i.e., the inclusion of at least one, but also including more than one, of a number or list of elements, and, optionally, additional unlisted items. Only terms clearly indicated to the contrary, such as "only one of" or "exactly one of," or, when used in the claims, "consisting of," will refer to the inclusion of exactly one element of a number or list of elements. In general, the term "or" as used herein shall only be interpreted as indicating exclusive alternatives (i.e. "one or the other but not both") when preceded by terms of exclusivity, such as "either," "one of," "only one of," or "exactly one of." "Consisting essentially of," when used in the claims, shall have its ordinary meaning as used in the field of patent law.

As used herein in the specification and in the claims, the phrase "at least one," in reference to a list of one or more elements, should be understood to mean at least one element selected from any one or more of the elements in the list of elements, but not necessarily including at least one of each and every element specifically listed within the list of elements and not excluding any combinations of elements in the list of elements. This definition also allows that elements may optionally be present other than the elements specifically identified within the list of elements to which the phrase "at least one" refers, whether related or unrelated to those elements specifically identified. Thus, as a non-limiting example, "at least one of A and B" (or, equivalently, "at least one of A or B," or, equivalently "at least one of A and/or B") can refer, in one embodiment, to at least one, optionally including more than one, A, with no B present (and optionally including elements other than B); in another embodiment, to at least one, optionally including more than one, B, with no A present (and optionally including elements other than A); in yet another embodiment, to at least one, optionally including more than one,

A, and at least one, optionally including more than one, B (and optionally including other elements); etc.

It should also be understood that, unless clearly indicated to the contrary, in any methods claimed herein that include more than one step or act, the order of the steps or  
5 acts of the method is not necessarily limited to the order in which the steps or acts of the method are recited.

In the claims, as well as in the specification above, all transitional phrases such as "comprising," "including," "carrying," "having," "containing," "involving," "holding," "composed of," and the like are to be understood to be open-ended, i.e., to mean  
10 including but not limited to. Only the transitional phrases "consisting of" and "consisting essentially of" shall be closed or semi-closed transitional phrases, respectively, as set forth in the United States Patent Office Manual of Patent Examining Procedures, Section 2111.03.

15 What is claimed is:

## CLAIMS

1. A composition, comprising:  
a miniature protein including an alpha helix region, the miniature protein  
5 being modified by substitution of three to six amino acid residues, inclusively,  
with arginine residues, the substitutions being at non-sequential positions within  
the alpha helix region.
- 10 2. The composition of claim 1, wherein the miniature protein, prior to substitution,  
is PYY.
3. The composition of claim 1, wherein the miniature protein, prior to substitution,  
is avian pancreatic polypeptide.
- 15 4. The composition of claim 1, wherein the miniature protein, prior to substitution,  
is YY2.
5. The composition of claim 1, wherein the miniature protein is modified by the  
substitution of three amino acid residues with arginine residues.  
20
6. The composition of claim 1, wherein the miniature protein is modified by the  
substitution of four amino acid residues with arginine residues.
7. The composition of claim 1, wherein the miniature protein is modified by the  
substitution of five amino acid residues with arginine residues.  
25
8. The composition of claim 1, wherein the three to six amino acid substitutions are  
each chosen from the group consisting of positions 15, 18, 19, 22, 25, and 26 of  
the miniature protein.  
30
9. A composition, comprising:  
a miniature protein including a type II polyproline helix region, the  
miniature protein being modified by substitution of three to six amino acid



residues, inclusively, with arginine residues, the substitutions being at non-sequential positions within the type II polyproline helix region.

- 5 10. The composition of claim 9, wherein the miniature protein, prior to substitution, is PYY.
11. The composition of claim 9, wherein the miniature protein, prior to substitution, is avian pancreatic polypeptide.
- 10 12. The composition of claim 9, wherein the miniature protein, prior to substitution, is YY2.
13. The composition of claim 9, wherein n is between 3 and 6, inclusively.
- 15 14. The composition of claim 9, wherein the miniature protein is modified by the substitution of three amino acid residues with arginine residues.
15. The composition of claim 9, wherein the miniature protein is modified by the substitution of four amino acid residues with arginine residues.
- 20 16. The composition of claim 9, wherein the miniature protein is modified by the substitution of five amino acid residues with arginine residues.
17. The composition of claim 9, wherein the miniature protein is modified by the substitution of six amino acid residues with arginine residues.
- 25 18. The composition of claim 9, wherein the type II polyproline helix region is modified such that the type II polyproline helix region includes at least one portion having the sequence PPR.
- 30 19. The composition of claim 9, wherein the type II polyproline helix region is modified such that the type II polyproline helix region includes at least one portion having the sequence  $(PPR)_n$ , wherein n is at least 3.

20. A composition, comprising:  
a PYY including a type II polyproline helix region, the PYY being modified by substitution of three to six amino acid residues, inclusively, with arginine residues.
- 5
21. A composition, comprising:  
a PYY including a type II polyproline helix region, the PYY being modified by substitution of at least a portion of the type II polyproline helix region with the sequence  $(PPR)_n$ , wherein n is at least 2.
- 10
22. A composition, comprising:  
APPLPPNRGEDASPEELSRYYRSLRHLYLNLVTRQRY (SEQ ID NO: 50).
- 15
23. A composition, comprising:  
APPLPPNRGEDASPRELSRYYRSLRHLYLNLVTRQRY (SEQ ID NO: 51).
- 20
24. A composition, comprising:  
APPLPPNRGEDASPRELRRYYRSLRHLYLNLVTRQRY (SEQ ID NO: 52).
- 25
25. A composition, comprising:  
a miniature protein including a type II polyproline helix region, the miniature protein being modified by substitution of at least a portion of the type II polyproline helix region with the sequence  $(PPR)_n$ , wherein n is at least 4.
26. The composition of claim 25, wherein the miniature protein, prior to substitution, is PYY.
- 30
27. The composition of claim 25, wherein the miniature protein, prior to substitution, is YY2.

28. The composition of claim 25, wherein the miniature protein, prior to substitution, is avian pancreatic polypeptide.
- 5 29. The composition of claim 25, wherein n is between 4 and 6, inclusively.
30. A composition, comprising:  
a miniature protein including a type II polyproline helix region, the miniature protein being modified by substitution of at least a portion of the type  
10 II polyproline helix region with the sequence (PRR)<sub>n</sub>, wherein n is at least 3.
31. The composition of claim 30, wherein the miniature protein, prior to substitution, is PYY.
- 15 32. The composition of claim 30, wherein the miniature protein, prior to substitution, is YY2.
33. The composition of claim 30, wherein the miniature protein, prior to substitution, is avian pancreatic polypeptide.
- 20 34. The composition of claim 30, wherein n is between 3 and 6, inclusively.
35. A composition, comprising:  
a miniature protein modified by substitution of at least three amino acid  
25 residues with arginine residues, wherein the substitutions are chosen from the group consisting of positions 15, 18, 19, 22, 25, and 26 of the avian pancreatic polypeptide.
36. The composition of claim 35, wherein the miniature protein, prior to substitution,  
30 is PYY.
37. The composition of claim 35, wherein the miniature protein, prior to substitution, is YY2.

38. The composition of claim 35, wherein the miniature protein, prior to substitution, is avian pancreatic polypeptide .
- 5 39. The composition of claim 35, wherein the miniature protein contains arginine substitutions in positions 15, 19, 22, and 26.
40. The composition of claim 35, wherein the miniature protein contains arginine substitutions in positions 15, 19, 22, 25, and 26.
- 10 41. The composition of claim 35, wherein the miniature protein contains arginine substitutions in positions 15, 18, 19, 22, 25, and 26.
42. A composition, comprising:
- 15 an avian pancreatic polypeptide (aPP) including an alpha helix region, the avian pancreatic polypeptide being modified by substitution of three to six amino acid residues, inclusively, with arginine residues, the substitutions being at non-sequential positions within the alpha helix region.
- 20 43. The composition of claim 42, wherein the avian pancreatic polypeptide has a sequence, before modification, of  
GPSQPTYPGDDAPVEDLIRFYNDLQQYLNVVTRHRY (SEQ ID NO: 1).
44. The composition of claim 42, wherein the avian pancreatic polypeptide is  
25 modified by substitution of three to six amino acid residues, inclusively.
45. The composition of claim 42, wherein the avian pancreatic polypeptide is modified by the substitution of three amino acid residues with arginine residues.
- 30 46. The composition of claim 42, wherein the avian pancreatic polypeptide is modified by the substitution of four amino acid residues with arginine residues.

47. The composition of claim 42, wherein the avian pancreatic polypeptide is modified by the substitution of five amino acid residues with arginine residues.
48. The composition of claim 42, wherein the three to six amino acid substitutions  
5 are each chosen from the group consisting of positions 15, 18, 19, 22, 25, and 26 of the avian pancreatic polypeptide.
49. The composition of claim 42, wherein the avian pancreatic polypeptide has a sequence, before modification, of  
10 GPSQPTYPGDDAPVEDLIRFVGRLLAYFGDTINRY (SEQ ID NO: 2) or  
GPSQPTYPGDDAPVEDLIRFVGRLLAYFGDTINRYC (SEQ ID NO: 3).
50. The composition of claim 42, wherein the avian pancreatic polypeptide has a sequence, before modification, of  
15 GPSQPTYPGDDAPVVDLIRFRGRWL AYLGDITINRY (SEQ ID NO: 4) or  
GPSQPTYPGDDAPVVDLIRFRGRWLA YLGDITINRYC (SEQ ID NO: 5).
51. The composition of claim 42, wherein the avian pancreatic polypeptide has a sequence, before modification, of  
20 GPSQPTYPGDDAPLGLDISFRGRFLAYFGDTINRY (SEQ ID NO: 6) or  
GPSQPTYPGDDAPLGLDISFRGRFLAYFGDTINRYC (SEQ ID NO: 7).
52. The composition of claim 42, wherein the avian pancreatic polypeptide has a sequence, before modification, of  
25 GPSQPTYPGDDAPVEDLIRFKFLLQWYLLALTRHRYAAA (SEQ ID NO: 8)  
or GPSQPTYPGDDAPVEDLIRFKFLLQWYLLALTRHRYAAAC (SEQ ID  
NO: 9).
53. The composition of claim 42, wherein the avian pancreatic polypeptide has a  
30 sequence, before modification, of  
GKSWMTVPGDDAPVEDLIRFKFLLQWYLLALTRHRYAAA (SEQ ID NO:  
10) or GKSWMTVPGDDAPVEDLIRFKFLLQWYLLALTRHRYAAAC (SEQ  
ID NO: 11).

54. The composition of claim 42, wherein the avian pancreatic polypeptide has a sequence, before modification, of  
GPSQPTYPGDDAPVEDLIRFKFLLQWYLLALYHNNYAAA (SEQ ID NO:  
5 13) or GPSQPTYPGDDAPVEDLIRFKFLLQWYLLALYHNNYAAAC (SEQ  
ID NO: 14).
55. The composition of claim 42, wherein the avian pancreatic polypeptide has a sequence, before modification, of  
10 GPSQPTYPGDDAPVEDLIRFKFLLQWYLLALSHYNYAAA (SEQ ID NO:  
15) or GPSQPTYPGDDAPVEDLIRFKFLLQWYLLALSHYNYAAAC (SEQ  
ID NO: 16).
56. The composition of claim 42, wherein the avian pancreatic polypeptide has a sequence, before modification, of  
15 GPSQPTYPGDDAPVEDLIRFKFLLQWYLLALSLRNYAAA (SEQ ID NO:  
17) or GPSQPTYPGDDAPVEDLIRFKFLLQWYLLALSLRNYAAAC (SEQ ID  
NO: 18).
- 20 57. The composition of claim 42, wherein the avian pancreatic polypeptide has a sequence, before modification, of  
GPSQPTYPGDDAPVEDLIRFKFLLQWYLLALAADAYAAA (SEQ ID NO:  
19) or GPSQPTYPGDDAPVEDLIRFKFLLQWYLLALAADAYAAAC (SEQ  
ID NO: 20).
- 25 58. A composition, comprising:  
an avian pancreatic polypeptide (aPP) including a type II polyproline  
helix region, the avian pancreatic polypeptide being modified by substitution of  
three to six amino acid residues, inclusively, with arginine residues, the  
30 substitutions being at non-sequential positions within the type II polyproline helix  
region.

59. The composition of claim 58, wherein the avian pancreatic polypeptide has a sequence, before modification, of  
GPSQPTYPGDDAPVEDLIRFYNDLQQYLNVVTRHRY (SEQ ID NO: 1).
- 5 60. The composition of claim 58, wherein the avian pancreatic polypeptide is modified by the substitution of three amino acid residues with arginine residues.
61. The composition of claim 58, wherein the avian pancreatic polypeptide is modified by the substitution of four amino acid residues with arginine residues.
- 10 62. The composition of claim 58, wherein the avian pancreatic polypeptide is modified by the substitution of five amino acid residues with arginine residues.
63. The composition of claim 58, wherein the avian pancreatic polypeptide is modified by the substitution of six amino acid residues with arginine residues.
- 15 64. The composition of claim 58, wherein the type II polyproline helix region is modified such that the type II polyproline helix region includes at least one portion having the sequence PRR.
- 20 65. The composition of claim 58, wherein the type II polyproline helix region is modified such that the type II polyproline helix region includes at least one portion having the sequence (PRR)<sub>n</sub>, wherein n is at least 3.
- 25 66. A composition, comprising:  
an avian pancreatic polypeptide (aPP) including a type II polyproline helix region, the avian pancreatic polypeptide being modified by substitution of at least a portion of the type II polyproline helix region with the sequence (PPR)<sub>n</sub>, wherein n is at least 4.
- 30 67. The composition of claim 66, wherein the avian pancreatic polypeptide has a sequence, before modification, of  
GPSQPTYPGDDAPVEDLIRFYNDLQQYLNVVTRHRY (SEQ ID NO: 1).

68. The composition of claim 66, wherein n is between 4 and 6, inclusively.
69. A composition, comprising:
- 5 an avian pancreatic polypeptide (aPP) including a type II polyproline helix region, the avian pancreatic polypeptide being modified by substitution of at least a portion of the type II polyproline helix region with the sequence (PRR)<sub>n</sub>, wherein n is at least 3.
- 10 70. The composition of claim 69, wherein the avian pancreatic polypeptide has a sequence, before modification, of  
GPSQPTYPGDDAPVEDLIRFYNDLQQYLNVVTRHRY (SEQ ID NO: 1).
71. The composition of claim 69, wherein n is between 3 and 6, inclusively.
- 15 72. A composition, comprising:
- an avian pancreatic polypeptide modified by substitution of at least three amino acid residues with arginine residues, wherein the substitutions are chosen from the group consisting of positions 15, 18, 19, 22, 25, and 26 of the avian pancreatic polypeptide.
- 20 73. The composition of claim 72, wherein the avian pancreatic polypeptide has a sequence, before modification, of  
GPSQPTYPGDDAPVEDLIRFYNDLQQYLNVVTRHRY (SEQ ID NO: 1).
- 25 74. The composition of claim 72, wherein the avian pancreatic polypeptide contains arginine substitutions in positions 15, 19, 22, and 26.
75. The composition of claim 72, wherein the avian pancreatic polypeptide contains arginine substitutions in positions 15, 19, 22, 25, and 26.
- 30 76. The composition of claim 72, wherein the avian pancreatic polypeptide contains arginine substitutions in positions 15, 18, 19, 22, 25, and 26.



77. A composition, comprising:  
GPSQPTYPGDDAPVRDLRRFYRDLRRYLNVVTRHRY (SEQ ID NO:  
21).
- 5
78. A composition, comprising:  
GPSQPTYPGDDAPVRDLIRFYRDLRRYLNVVTRHRY (SEQ ID NO:  
22).
- 10
79. A composition, comprising:  
GPSQPTYPGDDAPVRDLIRFYRDLQRYLNVVTRHRY (SEQ ID NO:  
23).
80. A composition, comprising:  
15 an avian pancreatic polypeptide (aPP) including an alpha helix region that  
is substantially cationic.
81. A composition, comprising:  
20 an avian pancreatic polypeptide (aPP) including a type II polyproline  
helix region that is substantially cationic.
82. A composition, comprising:  
a cell-permeable avian pancreatic polypeptide.
- 25
83. A composition, comprising:  
a modified avian pancreatic polypeptide that, when exposed to HeLa cells  
at a concentration of 1 micromolar, equilibrates inside the HeLa cells at a  
concentration that is at least 75 times greater than the equilibration concentration  
of an unmodified avian pancreatic polypeptide exposed to the HeLa cells under  
30 the same conditions.
84. The composition of claim 83, wherein the concentration of the modified avian  
pancreatic polypeptide inside the HeLa cells is at least 100 times greater than the

concentration of the unmodified avian pancreatic polypeptide inside the HeLa cells under the same conditions.

85. A composition, comprising:

5 a protein selected from the group consisting of  
GPRRPRRPRRGDDAPVEDLIRFYNDLQQYLNVVTRHRYC (SEQ ID NO:  
24), GPRRPRRPRRPRRGDDAPVEDLIRFYNDLQQYLNVVTRHRYC (SEQ  
ID NO: 25),  
GPRRPRRPRRPRRPRRGDDAPVEDLIRFYNDLQQYLNVVTRHRYC (SEQ  
10 ID NO: 26),  
GPRRPRRPRRPRRPRRPRRGDDAPVEDLIRFYNDLQQYLNVVTRHRYC  
(SEQ ID NO: 27),  
GPRRPRRPRRPRRPRRPRRGDDAPVEDLIRFYNDLQQYLNVVTRHRYC (SEQ ID  
NO: 28),  
15 GPRRPRRPRRPRRPRRPRRGDDAPVEDLIRFYNDLQQYLNVVTRHRYC (SEQ  
ID NO: 29), and  
GPRRPRRPRRPRRPRRPRRGDDAPVEDLIRFYNDLQQYLNVVTRHRYC  
(SEQ ID NO: 30).

20 86. A composition, comprising:

a miniature protein that is cell permeable and able to bind to Bcl2 with a  $K_d$  of at least about 7 nM.

87. The composition of claim 86, wherein  $K_d$  is at least about 50 nM.

25

88. The composition of claim 86, wherein the miniature protein is an avian pancreatic polypeptide.

89. The composition of claim 86, wherein the avian pancreatic polypeptide has a  
30 sequence, before modification, of

GPSQPTYPGDDAPVEDLIRFYNDLQQYLNVVTRHRY (SEQ ID NO: 1).

90. A composition, comprising:  
a miniature protein that is cell permeable and able to bind to hDM2 with a  $K_d$  of at least about 10 nM.
- 5 91. The composition of claim 90, wherein the miniature protein is an avian pancreatic polypeptide.
92. The composition of claim 91, wherein the avian pancreatic polypeptide has a sequence, before modification, of  
10 GPSQPTYPGDDAPVEDLIRFYNDLQQYLNVVTRHRY (SEQ ID NO: 1).
93. The composition of claim 90, wherein the miniature protein is cell permeable and able to bind to hDM2 with a  $K_d$  of at least about 30 nM.
- 15 94. The composition of claim 90, wherein the miniature protein is cell permeable and able to bind to hDM2 with a  $K_d$  of at least about 100 nM.
95. The composition of claim 90, wherein the miniature protein is cell permeable and able to bind to hDM2 with a  $K_d$  of at least about 300 nM.  
20
96. A composition, comprising:  
GPRRPRRPGRRAPVEDLIRFVGRLLA YFGDT INRYC (SEQ ID NO:  
31).
- 25 97. A composition, comprising:  
GPRRPRRPGRRAPVVDLIRFRGRWL AYLGDTINRYC (SEQ ID NO:  
32).
98. A composition, comprising:  
30 GPRRPRRPGRRAPLGDLISFRGRFLA YFGDTINRYC (SEQ ID NO:  
33).

99. A composition, comprising:  
G(PRR)<sub>n</sub>PGRRAPVEDLIRFVGRLLAYFGDTINRYC (SEQ ID NO: 40),  
wherein n is at least 2.
- 5 100. A composition, comprising:  
G(PRR)<sub>n</sub>PGRRAPVVDLIRFRGRWLAYLGDTINRYC (SEQ ID NO:  
41), wherein n is at least 2.
101. A composition, comprising :  
10 G(PRR)<sub>n</sub>PGRRAPLGDLSFRGRFLAYFGDTINRYC (SEQ ID NO: 42),  
wherein n is at least 2.
102. A composition, comprising:  
GPRRPRRPGRRAPVEDLIRFKFLLQWYLLALSLRNYAAAC (SEQ  
15 ID NO: 34).
103. A composition, comprising:  
GPRQPRYPGRDAPVEDLIRFKFLLQWYLLALSLRNYAAAC (SEQ  
ID NO: 35).  
20
104. A composition, comprising:  
GPSRPTRPGDRAPVEDLIRFKFLLQWYLLALSLRNYAAAC (SEQ  
ID NO: 36).
- 25 105. A composition, comprising:  
GPRRPRRPGRRAPVEDLIRFKFLLQWFLALTRHRYAAAC (SEQ  
ID NO: 37).
106. A composition, comprising:  
30 GPRRPRRPGRRAPVEDLIRFKFLLQWYLLALYHNNYAAAC (SEQ  
ID NO: 38).

107. A composition, comprising:  
GPRRPRRPGRRAPVEDLIRFKFLLQWYLLALSHYNYAAAC (SEQ  
ID NO: 39).
- 5 108. A composition, comprising:  
G(PRR)<sub>n</sub>PGRRAPVEDLIRFKFLLQWYLLALSLRNYAAAC (SEQ ID  
NO: 43), wherein n is at least 2.
109. A composition, comprising:  
10 G(PRR)<sub>n</sub>PGRRAPVEDLIRFKFLLQWFLLALTRHRYAAAC (SEQ ID  
NO: 44), wherein n is at least 2.
110. A composition, comprising:  
15 G(PRR)<sub>n</sub>PGRRAPVEDLIRFKFLLQWYLLALYHNNYAAAC (SEQ ID  
NO: 45), wherein n is at least 2.
111. A composition, comprising:  
20 G(PRR)<sub>n</sub>PGRRAPVEDLIRFKFLLQWYLLALSHYNYAAAC (SEQ ID  
NO: 46), wherein n is at least 2.
112. A composition, comprising:  
a miniature protein that, when exposed to a population of Jurkat cells at a  
concentration of 10 micromolar, induces apoptosis in at least about 50% of the  
Jurkat cells.
- 25 113. A composition, comprising :  
an avian pancreatic polypeptide (aPP) including a type II polyproline  
helix region, the avian pancreatic polypeptide being modified by substitution of  
three to six amino acid residues, inclusively, with arginine residues.
- 30 114. A composition, comprising:  
an avian pancreatic polypeptide (aPP) including an alpha helix region, the

avian pancreatic polypeptide being modified by substitution of three to six amino acid residues, inclusively, with arginine residues.

115. A composition, comprising:

5 an avian pancreatic polypeptide (aPP) including a type II polyproline helix region, the avian pancreatic polypeptide being modified by substitution of at least a portion of the type II polyproline helix region with the sequence  $(PPR)_n$ , wherein n is at least 2.

10 116. A composition, comprising:

RRPRRPRRPRRPGRRAPVEDLIRFYNDLQQYLNVVTRHRYC (SEQ  
ID NO: 47).

117. A composition, comprising:

15 RRPRRPRRPGRRAPVEDLIRFYNDLQQYLNVVTRHRYC (SEQ ID  
NO: 48).

118. A composition, comprising:

20 GPRRPRRPGRRAPVEDLIRFYNDLQQYLNVVTRHRYC (SEQ ID  
NO: 49).

119. A composition, comprising:

a cell-permeable miniature protein.

25 120. A composition, comprising:

a miniature protein including an alpha helix region that is substantially cationic.

121. A composition, comprising:

30 a miniature protein including a type II polyproline helix region that is substantially cationic.

122. A composition, comprising:  
a modified miniature protein that, when exposed to HeLa cells at a concentration of 1 micromolar, equilibrates inside the HeLa cells at a concentration that is at least 75 times greater than the equilibration concentration of an unmodified miniature protein exposed to the HeLa cells under the same conditions.
123. A composition, comprising:  
a protein modified by the addition of  $(\text{PPR})_n$ , wherein  $n$  is at least 2, wherein the protein, after modification, equilibrates inside HeLa cells at a concentration that is at least 75 times greater than the equilibration concentration of an unmodified protein exposed to the HeLa cells under the same conditions.
124. A composition, comprising:  
a protein modified by the addition of  $(\text{PPR})_n$ , wherein  $n$  is at least 2, wherein the protein, after modification, equilibrates inside HeLa cells at a concentration that is at least 75 times greater than the equilibration concentration of an unmodified protein exposed to the HeLa cells under the same conditions.
125. A composition, comprising:  
a protein containing the sequence  $(\text{PPR})_n$ , wherein  $n$  is at least 2.
126. The composition of claim 125, wherein  $n$  is at least 3.
127. A composition, comprising:  
a protein containing the sequence  $(\text{PPR})_n$ , wherein  $n$  is at least 2.
128. The composition of claim 127, wherein  $n$  is at least 3.

Fig. 1A

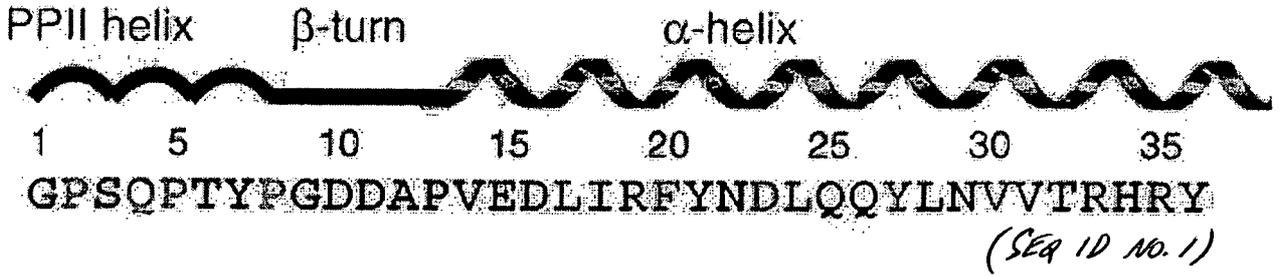
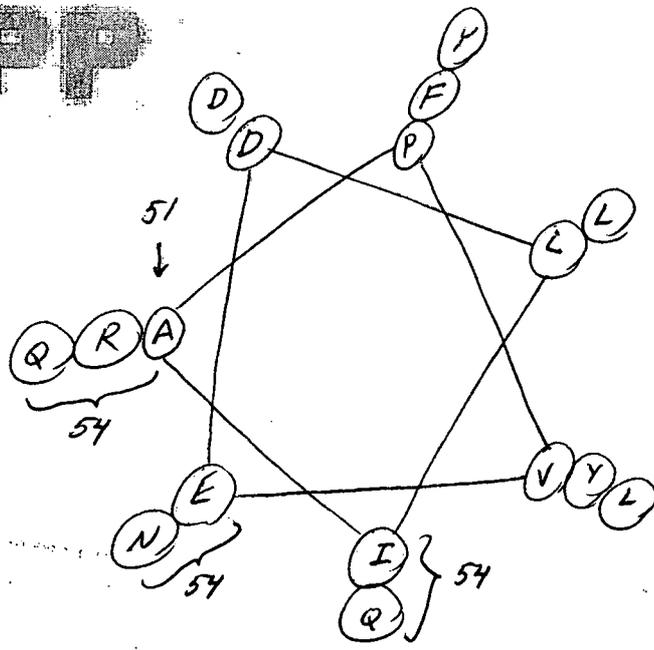


Fig. 1B



FIG. 2





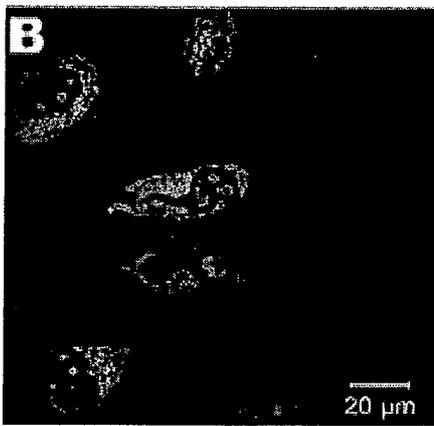
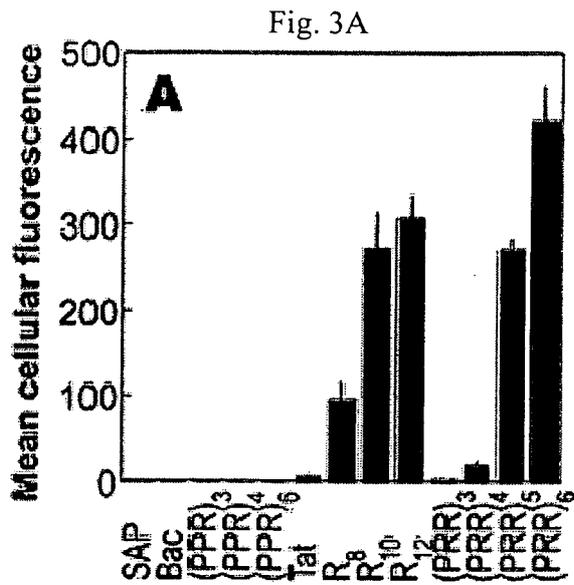


Fig. 3B

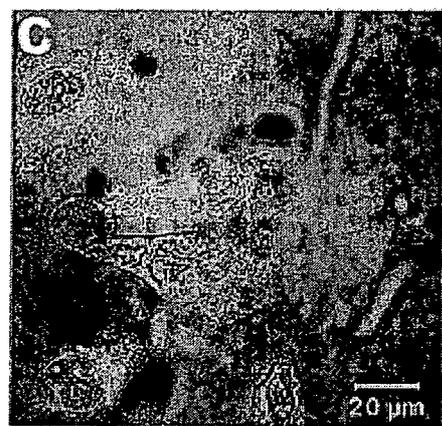
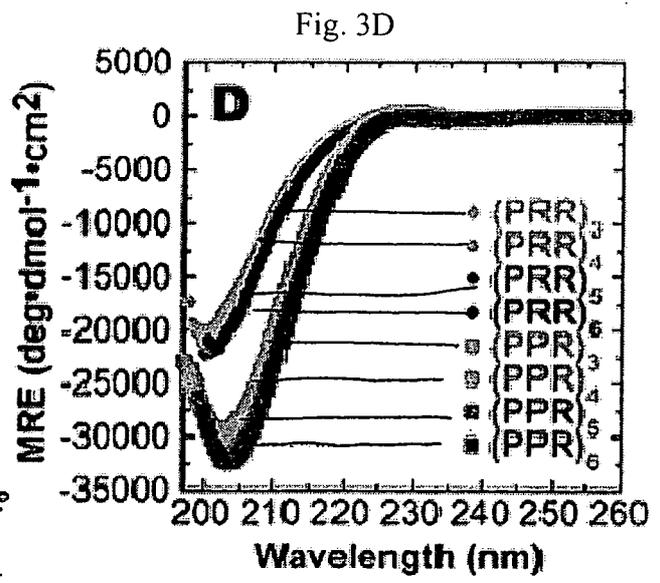
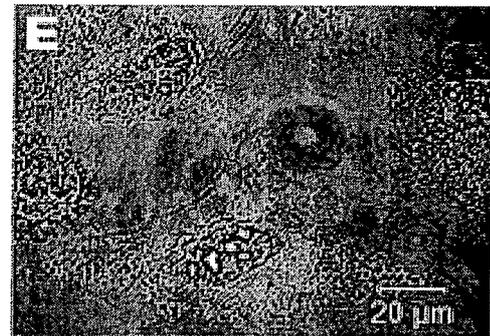
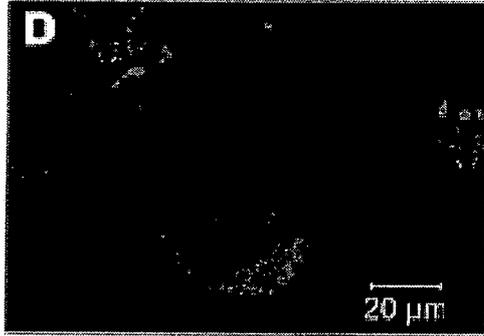
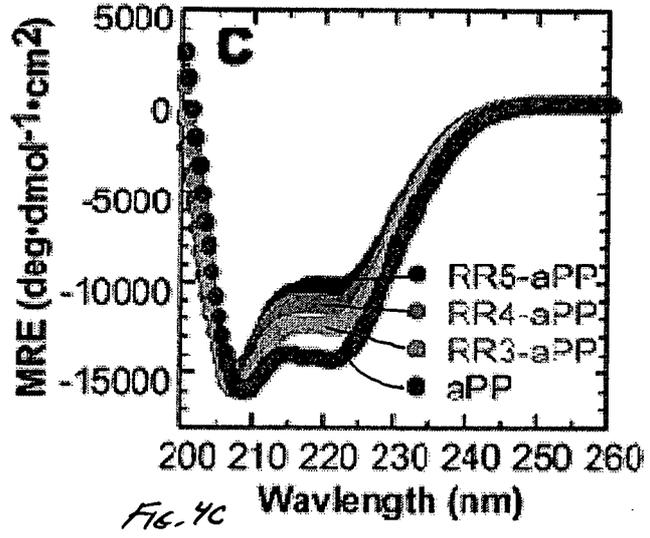
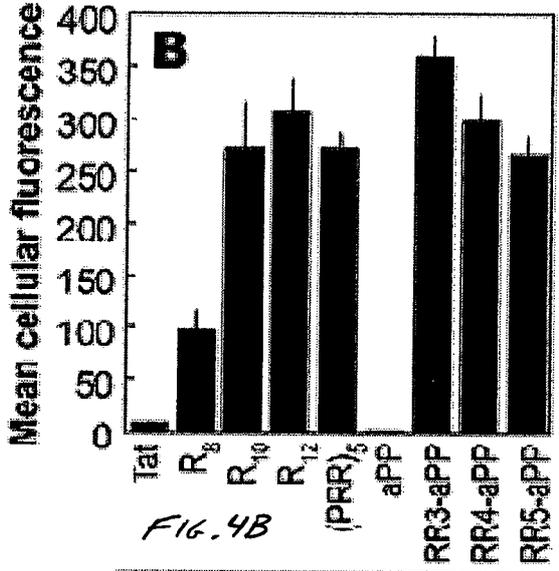
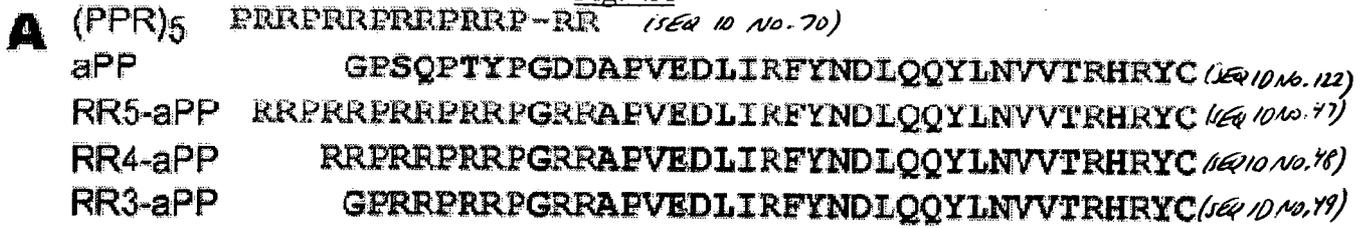
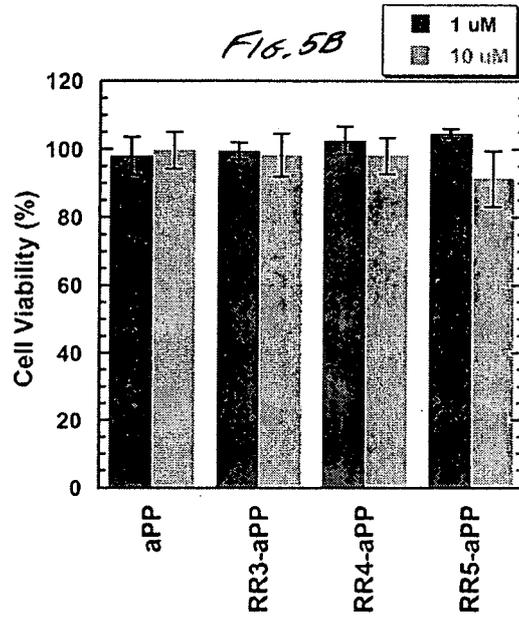
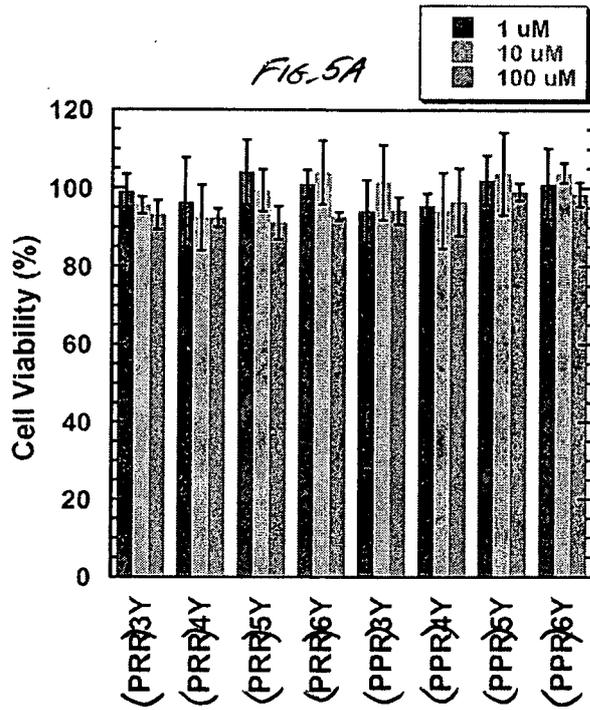


Fig. 3C

Fig. 4A





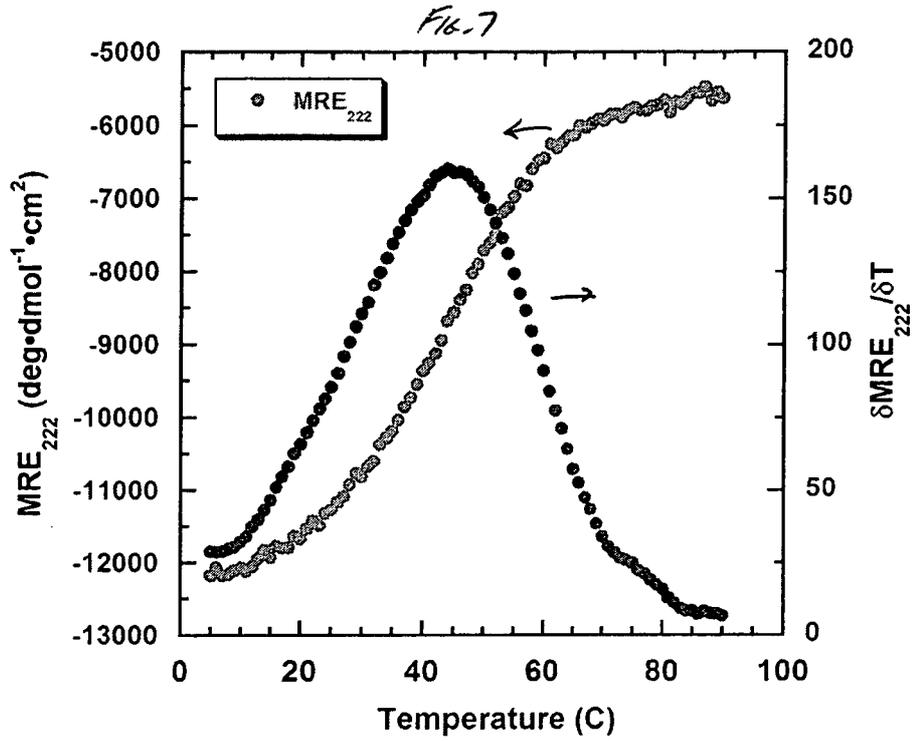


Fig. 6

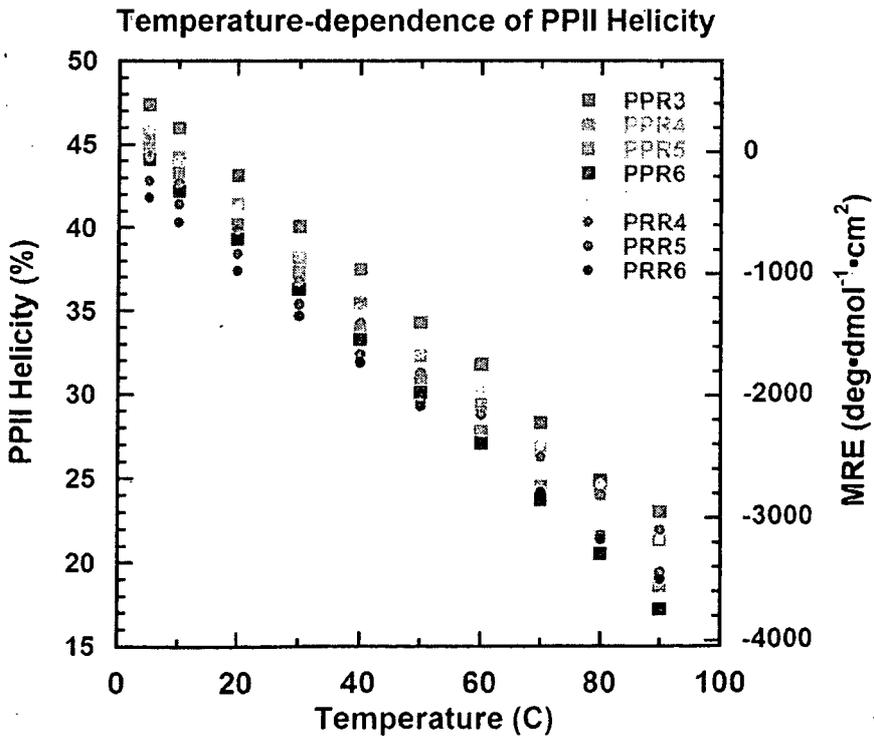
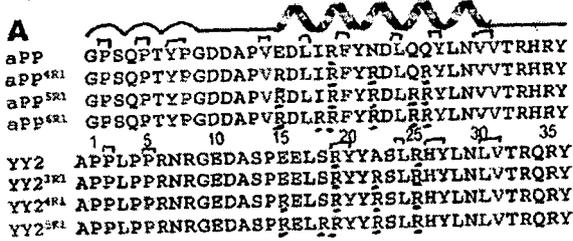


FIG. 8A



(SEQ ID NO: 1)  
 (SEQ ID NO: 29)  
 (SEQ ID NO: 32)  
 (SEQ ID NO: 21)  
 (SEQ ID NO: 67)  
 (SEQ ID NO: 53)  
 (SEQ ID NO: 51)  
 (SEQ ID NO: 52)

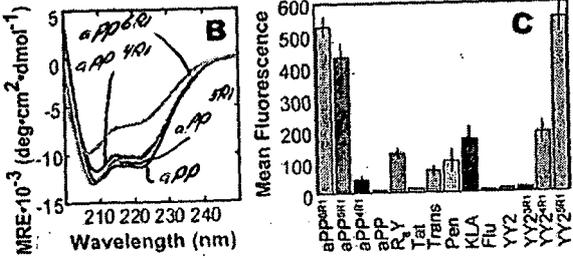


FIG. 8C

FIG. 8B

FIG. 9A

FIG. 9B

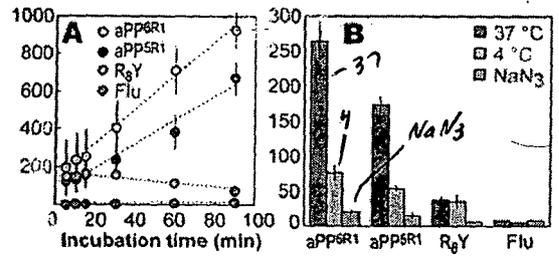


FIG. 9C

FIG. 9D

FIG. 9E

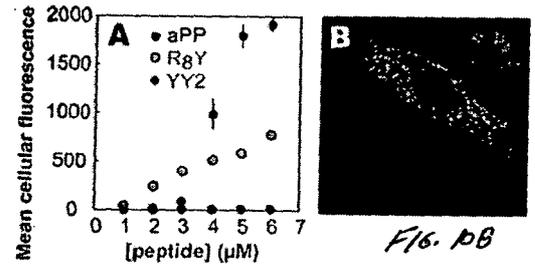


FIG. 10B

FIG. 10A

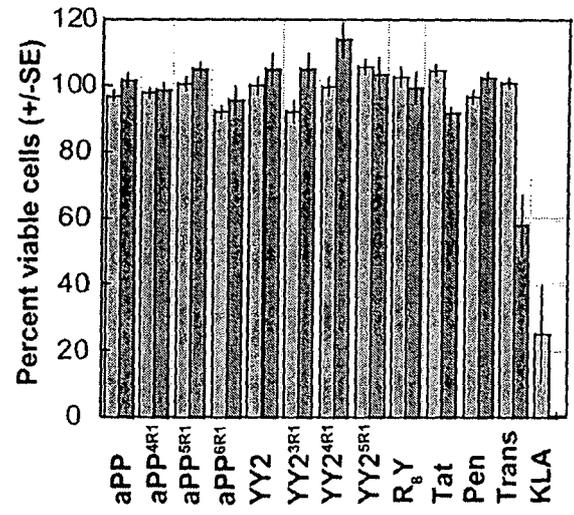


FIG. 13

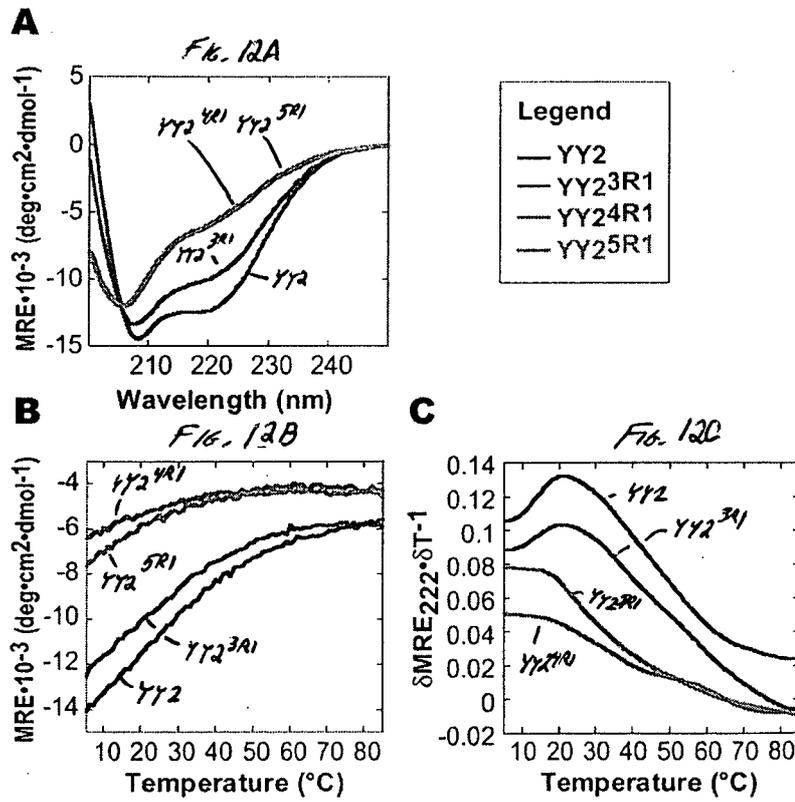
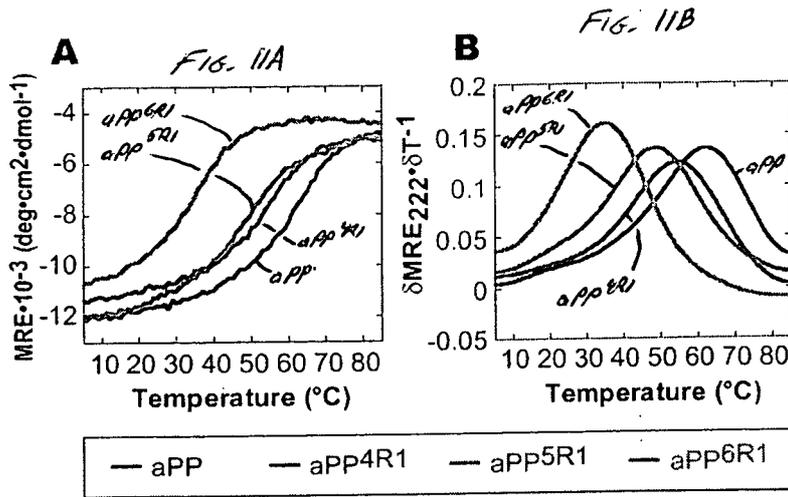


Fig. 14

