(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
10 April 2003 (10.04.2003)

PCT

(10) International Publication Number
WO 03/029821 A1

(51) International Patent Classification[7]: G01N 33/53, C12Q 1/70, C12P 21/04, C07K 16/00

(21) International Application Number: PCT/US02/31349

(22) International Filing Date: 1 October 2002 (01.10.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/326,566 1 October 2001 (01.10.2001) US

(71) Applicant *(for all designated States except US)*: **BUCK INSTITUTE** [US/US]; 8001 Redwood Boulevard, Novato, CA 94945 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **KOURAKINE, Alexei** [RU/US]; 1721 Novato Boulevard, Apt. 13, Novato, CA 94947 (US). **BREDESEN, Dale** [US/US]; 19 Racoon Drive, Novato, CA 94949 (US).

(74) Agents: **QUINE, Jonathan, Alan** et al.; Quine Intellectual Property Law Group, P.C., P.O. Box 458, Alameda, CA 94501 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 03/029821 A1

(54) Title: TARGET ASSISTED ITERATIVE SCREENING (TAIS) : A NOVEL SCREENING FORMAT FOR LARGE MOLECULAR REPERTOIRES

(57) Abstract: This invention provides a new in vitro screening method for the detection of protein-protein and other interactions. The method has been developed and applied to a commercial cDNA library to search for novel protein-protein interactions. PDZ, WW and SH3 domains from PSD95, Nedd4, Src, Abl and Crk proteins were used as test targets. 12 novel putative and 2 previously reported interactions were identified for 6 protein interaction modules in test screens. The novel screening format, dubbed TAIS (target-assisted iterative screening), provides an alternative platform to existing technologies for a pair-wise characterization of protein-protein, and other, interactions.

TARGET ASSISTED ITERATIVE SCREENING (TAIS): A NOVEL SCREENING FORMAT FOR LARGE MOLECULAR REPERTOIRES

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority to and benefit of USSN 60/326,566, filed on October 1, 2001, which is incorporated herein by reference in its entirety for all purposes.

STATEMENT AS TO RIGHTS TO INVENTIONS MADE UNDER FEDERALLY SPONSORED RESEARCH AND DEVELOPMENT

[0002] This invention was made, in part, with Government Support under Grant No: NS33376 awarded by the National Institutes of Health. The Government of the United States of America may have certain rights in this invention.

FIELD OF THE INVENTION

[0003] This invention pertains to the field of proteomics. In particular, this invention pertains to a dual screening method for determining interactions between members of a library and various targets that allows simultaneous screening for large numbers of interactions (*e.g.* protein-protein interactions) between library members and the target(s).

BACKGROUND OF THE INVENTION

[0004] Understanding the cell at a system level involves a comprehensive analysis of both the structure and the dynamics of cellular protein interaction networks. A large-scale analysis of protein-protein interactions has been attempted in lower eukaryotes, providing a first glimpse of the astounding structural complexity of the protein interaction webs (Walhout *et al.* (2000) *Science* 287: 116-122; Uetz *et al.* (2000) *Nature* 403: 623-627; Ito *et al.* (2001) *Proc. Natl. Acad. Sci., USA,* 98: 4569-4574).

[0005] Concurrently, a completed draft of the human genome has now delineated the dimensions of the human proteome (Venter *et al.* (2001) *Science* 291: 1304-1351; Lander *et al.* (2001) *Nature* 409: 860-921). Assembling of the estimated 30,000 to 50,000 human gene products into a comprehensive protein interaction map would provide a view of

the cell as a molecular system or molecular network and provide a system in which the timing and dynamics of protein-protein and other interaction events, could be examined.

[0006] Currently, the only practical method for a pair-wise characterization of protein-protein interactions with relatively high throughput is the yeast two hybrid system (Fields and Song (1989) *Nature* 340: 245-246). However, a high rate of false positives, poor performance in case of transcription factors, membrane bound, mistargeted and toxic proteins limit applicability of the two-hybrid system.

[0007] The limitations of the two-hybrid system have been recently highlighted by results of independent large scale protein interaction experiments performed on the yeast proteome ((Ito *et al.* (2001) *Proc Natl Acad Sci U S A* 98: 4569-74; Uetz et al., (2000) *Nature* 403: 623-627). The comparison revealed unexpectedly low overlap between the results of two groups (about 20%). Moreover, analysis of protein-protein interactions deposited in the Yeast Proteome Database showed that systematic two-hybrid projects failed to reproduce as much as approximately 90% of the interactions identified in conventional two-hybrid screens (Ito *et al.* (2001) *Proc Natl Acad Sci U S A* 98: 4569-4574).

[0008] The absence of a positive control in two-hybrid systems is particularly problematic as this approach is known for its abundance of false positives. In addition, it is known that the two-hybrid system is poorly designed for the identification of proteins interacting with transcription factors, and with toxic, membrane-bound, mistargeted or large proteins.

[0009] Therefore, the development of new methods with high throughput potential to characterize protein-protein interactions is of paramount importance, and increasingly so with the increasing availability of the human, and other, genome sequences.

SUMMARY OF THE INVENTION

[0010] The present invention pertains to a novel, rapid *in vitro* screening method for the identification and characterization of protein-protein interactions (*e.g.* interactions mediated by specialized protein modules such as SH3, PDZ and WW domains). The method is well suited to large-scale functional genomics approaches. In essence the present method combines the advantages of phage display technology and cDNA expression libraries.

[0011] In one embodiment, this invention provides a method of identifying interacting proteins from a plurality of potentially interacting proteins. The method typically involves i) contacting one or more targets (*e.g.* target proteins) with a protein display library comprising a plurality of potential binding proteins for the one or more target proteins; ii) selecting members of the protein display library that bind to the one or more target proteins to provide a preselected set of potential binding proteins; iii) separating the members of the preselected set of potential binding proteins from the bound target protein and localizing and/or immobilizing the members on a solid support such that the members are spatially addressable; and iv) contacting members of the preselected set of potential binding proteins with one or more target proteins; and v) detecting binding of members of the preselected set of potential binding proteins with the one or more target proteins whereby binding of a member of said set of potential binding partners with a target protein indicates that the member and the target protein are interacting proteins.

[0012] In certain preferred embodiments, the target proteins are attached to a solid support during the first contacting step. The protein display library can be any convenient display library. Preferred display libraries include, but are not limited to phage display, bacterial display, yeast display, eukaryotic virus display library, direct plasmid display library, and so forth. In certain embodiments, the library is an *in vitro* display library (*e.g.* covalent display technology (CDT), polysome display, eukaryotic in vitro transcription/translation systems, RNA-peptide fusions, and the like). Such libraries typically comprise at least 100 different members, preferably at least 1000 different members, more preferably at least 10,000 and most preferably at least 10^6, 10^7, 10^8, 10^9 or 10^{10} different members. In particularly preferred embodiments, the library displays a cDNA library (*e.g.* from a particular organism, tissue, cell type, *etc.*).

[0013] In certain embodiments, amplification of preselected subset of potential interactors of the target(s) is often performed, and can be performed in a spatially addressable manner. Thus, in certain embodiments, the "separating" comprises amplifying members of the protein display library that bind to said one or more target proteins and/or the separating and/or immobilizing comprises amplifying members of the protein display library that bind to said one or more target proteins. The amplifying can comprise amplification of the members when they are spatially separated and addressable.

[0014] In certain embodiments, the selecting comprises removing unbound members of the display library from the solid support. The selecting can comprise capturing one or more target proteins and/or bound library members (*i.e.* in a bound complex) using an affinity matrix. In certain embodiments, contacting *members* of the preselected set of potential binding partners with one or more target proteins comprises adsorbing members of the preselected set of potential binding partners to a solid support (*e.g.* a membrane). The detecting can be by means of a label attached to the target protein(s). Preferred labels include, but are not limited to a fluorescent label, a radioactive label, an enzymatic label, a colorimetric label, and a magnetic label.

[0015] In certain preferred embodiments, the contacting of step (i) comprises contacting the one or more target proteins with a protein display library where said one or more target proteins are attached to a solid support; the contacting of step (iv) comprises attaching members of the preselected set of potential binding proteins to a solid support to provide a set of attached preselected potential binding proteins and contacting the attached preselected potential binding proteins with the one or more target(s) (*e.g.* target proteins). The target proteins used in the contacting of step (iv) can be labeled with a detectable label before, during, or after the target proteins are contacted to the preselected potential binding proteins. In certain embodiments, the method further comprises sequencing the nucleic acid encoding the displayed protein on a member of the preselected display library that binds to the target protein. In certain embodiments, the contacting of step (i) comprises contacting one or more target proteins with a protein display library where said one or more target proteins and the protein display library are in solution. The selecting step can comprise capturing target proteins bound to members of the protein display library using an affinity matrix that specifically binds the target proteins or a tag attached to the target proteins. The contacting of step (iv) can comprise attaching members of said preselected set of potential binding proteins to a solid support to provide a set of attached preselected potential binding proteins and contacting the attached preselected potential binding proteins with the one or more target proteins. In certain preferred embodiments, the detecting comprises determining the amino acid sequence of a member of the set of potential binding partners (*e.g.*, binding proteins) that binds a target protein. The method can further involve

recording the amino acid sequence or identity of a member of the set of potential binding partners that binds a target protein in a database of proteins that interact with the target.

[0016] The methods described herein are not limited simply to target protein(s). Essentially any target moiety can be used. Such moieties include, but are not limited to various natural or synthetic chemical compounds including, but not limited to drugs, small organic molecules, nucleic acids, proteins, glycoproteins, carbohydrates, and the like. Similarly, the display library need not be limited to proteins. Virtually any moiety that can be displayed in a library is suitable. Particularly preferred display libraries include, but are not limited to protein or nucleic acid display libraries.

[0017] In one particularly preferred embodiment, this invention provides a method of identifying proteins or nucleic acids that interact with target moieties from a nucleic acid or protein library comprising a plurality of nucleic acids or proteins. The method typically comprises, i) contacting one or more target moieties with the library; ii) selecting members of the library that bind to the one or more target moieties to provide a preselected set of potential binding partners; iii) separating the members of the preselected set of potential binding partners from the bound target and immobilizing the members on a solid support such that the members are spatially addressable; iv) contacting members of the preselected set of potential binding partners with one or more target moieties; and v) detecting binding of members of the set of potential binding partners with said one or more target moieties whereby binding of a member of the set of potential binding partners with a target binding moiety indicates that said member is a binding partner that interacts with the target moiety. Preferred libraries include, but are not limited to a phage display library, a bacterial display library, a yeast display library, a eukaryotic virus library, a direct encoded plasmid library, and the like. In certain preferred embodiments, the library is an *in vitro* display library (*e.g.* a covalent display technology (CDT) library, a polysome display library, an RNA-peptide fusion library, *etc.*). In certain embodiments, the target moiety is a nucleic acid (*e.g.* a DNA, an RNA), a lipid, a carbohydrate, a glycoprotein, or a small organic molecule.

[0018] This invention also provides a kit practicing any of the methods described herein. In one embodiment, the kit comprises a protein display library; and instructional materials providing protocols for the methods described herein.

[0041] Unlike traditional panning approaches that select for the best binders, TAIS eliminates the loss of weaker binders and propagation biases, that result from competition between individual phage during repetitive selection-amplification cycles. In addition, the method permits screening of significantly larger libraries than the ones routinely used in cDNA expression library screening. For example, if a practical limit of the cDNA expression library screening assay is 10^6-10^7 phage, the upper limit on the size of the library used in TAIS is defined by existing technologies of phage display library preparation, i.e., on the order of 10^8-10^{12} or more phage.

[0042] TAIS provides a number of advantages: The method does not require costly and sophisticated equipment, and can be used with commercially available reagents. The method involves only simple biochemical and microbiological manipulations, and, additionally because of the low cost is easily attainable for almost any lab, with minimal investment for setup. The method has a short turnaround time: normally within 24 hours an investigator will know whether or not a particular screen has been successful, and often, in 48 to 72 hours an investigator has DNA ready for sequencing to analyze the cDNAs selected in the screen. The screening is performed *in vitro*, i.e., under defined and manipulatable conditions; the readout is direct, and is easily accurately quantitated. The method provides a powerful tool to characterize ligand preferences of peptide recognition domains. In this application, cDNA libraries (*e.g.* phage-displayed cDNA libraries) have unique features when compared to traditional combinatorial peptide libraries. The lengths of the peptides in the library are not fixed. The libraries can feature natural peptide ligands of the target that provide internal references for physiologically relevant affinities and specificities of the interaction in question.

[0043] Since it is not usually known *a priori* within what length of the peptide ligand all determinants of a specific interaction reside and what are physiologically relevant interaction affinities, the features described above make displayed cDNA libraries an invaluable complement to traditional peptide libraries in the characterization of molecular recognition properties of peptide interaction modules.

[0044] Furthermore, TAIS allows the analysis of relatively weak and/or poorly propagating binders that are typically lost during the standard phage display panning procedure. Propagation biases and disparity in stabilities between different phages are of

special issue in the case of cDNA libraries, since the size and composition of displayed polypeptides in such libraries vary greatly in comparison to more traditional peptide or antibody libraries.

[0045] We believe that the application of the screening format described here to cDNA libraries provides a powerful platform complementing existing technologies for a pair-wise characterization of protein-protein interactions. The relatively high efficiency and technical simplicity of the proposed screening method, as well as its readily standardized output, will allow TAIS to be utilized as a high throughput tool for mapping of protein-protein interactions.

[0046] Finally, it is noted that, in essence, the TAIS format allows efficient, target affinity-driven reduction of enormous molecular diversity in liquid phase to a manageable size sub-library immobilized in a spatially addressable form that can be processed robotically or manually. As such the screening method can be applied to a number of other large molecular diversities such as phage-displayed peptide and recombinant antibody libraries, cell displayed polypeptide libraries, *etc.* Iterative presentation of the target in two different molecular contexts facilitates minimization of non-specific interactions.

[0047] As indicated above, in preferred embodiments, the methods of this invention involve two screening steps. Generally the methods comprise: i) contacting one or more target proteins with a molecular library (*e.g.* a protein display library, nucleic acid display library) comprising a plurality of potential binding partners for the one or more targets (*e.g.* target proteins); ii) selecting members of the display library that bind to the one or more targets to provide a preselected set of potential binding partners; iii) separating said members of said preselected set of potential binding partners from the bound target and immobilizing said members on a solid support such that said members are spatially addressable; and iv) contacting members of the preselected (and optionally amplified) set of potential binding partners with one or more targets again; and v) detecting binding of members of the set of potential binding proteins with the one or more targets whereby binding of a member of the set of potential binding partners with a target indicates that the member and the target interact.

Contacting one or more target moieties with a display library.

[0048] In preferred embodiments, the methods of this invention typically involve an initial screen that entails contacting one or more target moieties with a library of potential binding partners (*e.g.* preferably nucleic acids or proteins). The library is preferably a display library, more preferably a protein display library (*e.g.* phage display, bacterial display, yeast display, eukaryotic virus display library, direct plasmid display library, *etc.*).

[0049] The target moieties can include any moiety that is expect to be bound or is capable of being bound by a protein. Such moieties include, but are not limited to proteins, nucleic acids, lipids, glycoproteins, carbohydrates, polysaccharides, and the like. The target moieties need not be limited to individual molecules. Thus, for example, it is possible to use cell surfaces, receptors, tissues, and the like as targets.

[0050] The target moieties are typically contacted with a library of potential binding partners (*e.g.* proteins that might be capable of binding to the target(s)). Such libraries typically comprise at least 100 different members, preferably at least 1000 different members, more preferably at least 10,000 and most preferably at least 10^6, 10^7, 10^8, 10^9 or 10^{10} different members. In certain embodiments, the libraries are cDNA libraries derived from a particular cell type/line, and/or a particular tissue, and/or a particular organism. The libraries, however, need not be limited to cDNA libraries. Other libraries include, but are not limited to antibody libraries (*e.g.* single chain antibody libraries), libraries of proteins randomized in one or more domains, libraries comprising shuffled polypeptides, and the like.

[0051] In preferred embodiments, the libraries of potential binding partners are provided on a "display vector". Such display vectors include, but are not limited to phage-display vectors, bacterial display vectors (Fuchs *et al.* (1991) *Biotechnology* 9, 1369–1372), yeast display libraries (Boder and Wittrup (1997) *Nat. Biotechnol.* 15: 553–557), eukaryotic virus libraries (Kasahara *et al.* (1994) *Science* 266: 1373–1376), and direct plasmid display libraries (Cull *et al.* (1992) *Proc. Natl. Acad. Sci. U. S. A.* 89: 1865–1869), and the like. Suitable libraries also include *in vitro* display technologies (*e.g.* covalent display technology (CDT), polysome display, eukaryotic in vitro transcription/translation systems, RNA-peptide fusions, and the like (*see, e.g.,* Fitzgerald (2000) *Drug Discovery Today* 5(6): 253-258, and references cited therein).

[0052] The ability to express polypeptides on the surface of bacteria or of viruses that infect bacteria (bacteriophage or phage) makes it possible to screen and one or more binding polypeptide or a libraries of greater than 10^{10} clones. To express polypeptides on the surface of phage (phage display), a nucleic acid encoding the polypeptide is inserted into the gene encoding a phage surface protein (*e.g.*, pIII) and the polypeptide-surface fusion protein is displayed on the phage surface (McCafferty *et al.* (1990) *Nature*, 348: 552-554; Hoogenboom *et al.* (1991) *Nucleic Acids Res.* 19: 4133-4137). Since the polypeptides on the surface of the phage are functional, phage bearing binding polypeptides can be separated from non-binding phage by binding to a target (*e.g.* via antigen affinity chromatography) (*see, e.g.*, McCafferty *et al.* (1990) *Nature*, 348: 552-554).

[0053] Phage display has been successfully applied to a wide range of peptides and proteins, including antibodies McCafferty *et al.* (1990) *Nature*, 348: 552-554), growth hormone (Bass *et al.* (1990) *Proteins: Struct. Funct. Genet.* 8(4): 309-314), DNA binding proteins (Jamieson *et al.* (1994) *Biochem.*, 33(19): 5689-5695), enzymes (McCaffety *et al.* (1991) *Protein Eng.*, 4(8): 955-961); Corey *et al.* (1993) *Gene,* 128(1): 129-134); Soumillion *et al.* (1994) *J. Mol. biol.*, 237(4): 415-422), and macromolecular protease inhibitors (Roberts *et al.* (1992) *Proc. Natl. Acad. Sci. USA,* 89(6): 2429-2433); Pannekoek *et al.* (1993) *Gene*, 128(1): 135-140; Wang *et al.* (1995) *J. Biol. Chem.*, 270(20): 12250-12256); Markland *et al.* (1996) *Biochem.*, 35: 8058-8067; Markland *et al.* (1996) *Biochem.*, 35: 8045-8057).

[0054] In certain embodiments, a phage display library utilizes so called "hyperphage". In hyperphage, the number of single-chain antibody fragments (scFv) or other proteins, presented on filamentous phage particles can be increased by more than two orders of magnitude by using a newly developed helper phage (hyperphage). Hyperphage have a wild-type pIII phenotype and are therefore able to infect F+ *Escherichia coli* cells with high efficiency; however, their lack of a functional pIII gene means that the phagemid-encoded pIII–antibody fusion is the sole source of pIII in phage assembly. This results in a considerable increase in the fraction of phage particles carrying an the inserted protein on their surface (*see, e.g.*, Rondot *et al.* (2001) *Nature Biotechnology,* 19(1): 75-78).

[0055] Similar to phage-display systems, methods are known to display heterologous proteins on the surface of bacteria. Thus, for example, U.S. patent 6,190,662

provides methods and vectors for obtaining surface expression of a desired protein or polypeptide in Gram-positive host organisms (*e.g.* a Lactococcus host). Similarly U.S. Patent 5,348,867 teaches the expression of heterologous proteins on the surface of gram negative bacteria (*e.g. E. coli, Pseudomonas aeruginosa, Haemophilus influenza, etc.*).

[0056] Generally bacterial systems comprise tripartite chimeric genes. One segment of the tripartite gene is a targeting DNA sequence encoding a polypeptide capable of targeting and anchoring the fusion polypeptide to a host cell outer membrane. Targeting sequences are well known and have been identified in several of membrane proteins including Lpp. Generally, as in the case of Lpp, the protein domains serving as localization signals are relatively short. The Lpp targeting sequence includes the signal sequence and the first 9 amino acids of the mature protein. These amino acids are found at the amino terminus of Lpp. *E. coli* outer membrane lipoproteins from which targeting sequences may be derived include TraT, OsmB, NlpB and BlaZ. Lipoprotein 1 from *Pseudomonas aeruginosa* or the PA1 and PCN proteins from *Haemophilus influenza* as well as the 17 kDa lipoprotein from *Rickettsia rickettsii* and the H.8 protein from *Neisseria gonorrhea* and the like can be used.

[0057] A second component of the tripartite chimeric gene is a DNA segment encoding a membrane-transversing amino acid sequence. Transversing is intended to denote an amino acid sequence capable of transporting a heterologous or homologous polypeptide through the outer membrane. In preferred embodiments, the membrane transversing sequence will direct the fusion polypeptide to the external surface. As with targeting DNA segments, transmembrane segments are typically found in outer membrane proteins of all species of gram-negative bacteria. Transmembrane proteins, however, serve a different function from that of targeting sequences and generally include amino acids sequences longer than the polypeptide sequences effective in targeting proteins to the bacterial outer membrane. For example, amino acids 46-159 of the *E. coli* outer membrane protein OmpA effectively localize a fused polypeptide to the external surface of the outer membrane when also fused to a membrane targeting sequence. These surface exposed polypeptides are not limited to relatively short amino acid sequences as when they are incorporated into the loop regions of a complete transmembrane lipoprotein.

[0058] The third gene segment comprising the tripartite chimeric gene fusion is a DNA segment that encodes any one of a variety of desired heterologous polypeptides.

[0059] Other suitable display systems include, but are not limited to various *in vitro* display technologies such as covalent display technology (CDT), polysome display, eukaryotic in vitro transcription/translation systems, RNA-peptide fusions, and the like (*see, e.g.,* Fitzgerald (2000) *Drug Discovery Today* 5(6): 253-258, and references cited therein).

[0060] CDT exploits the properties of a replication initiator protein from the *E. coli* bacteriophage P2. The protein is the product of the viral Agene (P2A) and is an endonuclease that initiates a rolling circle replication process by binding to the viral origin (*ori*) and introducing a single strand discontinuity (nick) in the DNA. The 3'-OH group that is exposed by the action of P2A is used to prime progeny DNA synthesis using the host replication machinery (Schnos and Inman (1971) *J. Mol. Biol.* 55: 31–38; Geisselsoder (1976) *J. Mol. Biol.* 100: 13–22; Chattoraj (1978) *Proc. Natl. Acad. Sci., USA,* 75: 1685–1689). The nicking event also exposes a 5' phosphate and this becomes covalently attached to a tyrosine residue in the active site of P2A (Lindahl (1970) *Virology* 42: 522–533; Liu *et al.* (1994) *Nucleic Acids Res.* 22: 5204–5210).

[0061] One further property of P2A that is exploited in the CDT system is that P2A exclusively attaches to the same molecule of DNA from which it has been expressed. The high fidelity of the *cis* activity and the fact that the recognition sequence for the covalent attachment, *ori*, occurs within P2A's own coding sequence (Schnos and Inman (1971) *J. Mol. Biol.* 55: 31–38; Geisselsoder (1976) *J. Mol. Biol.* 100: 13–22; Chattoraj (1978) *Proc. Natl. Acad. Sci., USA,* 75: 1685–1689; Lindahl (1970) *Virology* 42: 522–533; Liu *et al.* (1994) *Nucleic Acids Res.* 22: 5204–5210; Liu *et al.* (1993) *J. Mol. Biol.* 231: 361–374) enables pools of polypeptides that are genetically fused to P2A to be synthesized *in vitro* such that they also become covalently attached to their own coding sequences.

[0062] To operate CDT, a pool of DNA molecules is prepared, each containing the coding sequence of P2A fused to the coding sequence for one of a diverse population of potential binding moieties (linear peptides or protein domains). The DNA pool is transcribed and translated concurrently *in vitro* using an *E. coli* S30 lysate and, because of the *cis*activity of P2A, each DNA molecule becomes covalently tagged with its own

expressed gene product. The protein–DNA complexes are then subjected to various screening/ selection strategies.

[0063] Polysome display systems work by transcribing and translating DNA templates *in vitro* under conditions that enable the isolation of stable mRNA–ribosome–nascent polypeptide complexes (Schaffitzel *et al.* (1999) *J. Immunol. Methods* 231: 119–135). This is achieved by controlling the concentration of magnesium ions (to stabilize the ribosome particle) and by either terminating polypeptide elongation by the addition of chloramphenicol or cooling down the translation products of mRNA templates that lack stop codons. Target-specific polysome complexes are retained on an appropriately derivatized solid surface and the co-selected mRNAs released by dissociation of ribosomes using ethylene diamine tetraacetate (EDTA). These are then recovered by reverse transcription (RT) and PCR for further manipulation.

[0064] Another *in vitro* display system uses a puromycin molecule to provide a covalent linkage between mRNA molecules and their encoded polypeptides (Roberts and Szostak (1997) *Proc. Natl. Acad. Sci., USA,* 94: 12297–12302). Puromycin is an antibiotic that mimics the aminoacyl end of tRNA and functions by entering the ribosomal A-site and forming an amide linkage with nascent polypeptide through the peptidyl transferase activity of the ribosome.

[0065] In the RNA-peptide fusion system, the puromycin is attached to the 3' end of a single-stranded DNA linker that is in turn ligated to the 3' end of the library-encoding mRNA. When the mRNA is translated *in vitro*, a ribosome reaches the junction between the mRNA and the DNA linker and stalls. The puromycin can then enter the ribosomal A-site and form a stable amide linkage with the encoded peptide. A library pool of mRNA-DNA-puromycin molecules can therefore be translated *in vitro* and purified RNA–peptide complexes incubated with a target molecule for screening. As with the polysome display system, retained complexes are recovered for further manipulation by RT-PCR.

[0066] These embodiments of display libraries are illustrative and not intended to be limiting. Other suitable display library formats will be known to those of skill in the art.

[0067] In a particularly preferred embodiment, display libraries are created that express a library of cDNAs, or other potential binding proteins as described herein. Nucleic

acids cDNAs encoding all the desired potential binding proteins can be prepared and inserted into the "vehicle(s) comprising the display library.

[0068] The inserted nucleic acids are made according to methods well known to those of skill in the art. For example, in one approach, the nucleic acids can be chemically synthesized using nucleotide reagents. However, in a particularly preferred embodiment, however, the nucleic acids are created using standard cloning techniques, *e.g.,* amplification (*e.g.,* PCR) cloning with appropriate primers. Detailed protocols for the production of libraries using phage display technology are provided in Example 1.

Selecting bound members of the phage- or bacterial-display library.

[0069] In preferred methods, members of the display library that bind to said one or more target proteins are selected to provide a preselected set of potential binding proteins. Methods of selecting bound phage-display or bacterial display members or other display library members are well known to those of skill in the art.

[0070] In a particularly preferred embodiment the target moiety (*e.g.* protein, DNA, etc.) is provided attached to a solid support/substrate. In such instances, after the phage-or bacterial-display library is contacted with the target(s), the unbound phage can be washed away and/or the substrate bearing the target(s) bound by phage can be separated from the solution containing the library. Repetitive wash steps will eliminate unbound library members.

[0071] Suitable supports for the attachment of target moieties include, but are not limited to the surfaces of wells, capillaries, planar surfaces, particulate materials (beads, etc), slurries, gels, and the like. Preferred materials include, but are not limited to magnetic beads, glass, plastic, ceramics, metals, various resins, membranes, and the like. The target moiety is coupled to the surface according to standard methods well known to those of skill in the art.

[0072] The target moieties can be directly coupled to the substrate or can be joined to the substrate through a linker. The procedure for attaching a target moiety to the substrate will vary according to the chemical structure of the moiety. Proteins contain a variety of functional groups (*e.g.*, -OH, -COOH, -SH, or $-NH_2$) groups, that are available for reaction with a suitable functional group on a surface or a linker to bind the target thereto.

Alternatively, the target moiety can be derivatized to expose or attach additional reactive functional groups. The derivatization may involve attachment of any of a number of linker molecules such as those available from Pierce Chemical Company, Rockford Illinois. A bifunctional linker having one functional group reactive with a group on a particular target moiety and another group reactive with a group on the substrate can be used to anchor the target moiety.

[0073] In certain embodiments, the target moieties can be attached to the surface by simple adsorption.

[0074] In other embodiments, the target moieties can be provided in solution and contacted to the members of the phage- or bacterial display library also in solution. In such instances, the target moiety can comprise a domain (tag) that can be specifically captured/bound by an affinity reagent (*e.g.* an antibody, ligand, etc.). Alternatively, the target moiety can be attached to a tag (*e.g.* an affinity tag) that can be captured by an affinity reagent.

[0075] Affinity tags are well known to those of skill in the art. Such tags include, but are not limited to biotin with avidin/streptavidin, ligands and their cognate receptors, particularly haptens and antibodies, polyhistidine with Ni-NTA, glutathione S-transferase (GST) and glutathione, epitopes and cognate antibodies, and the like.

[0076] Certain affinity tags include epitope tags. Epitope tags are well known to those of skill in the art. Moreover, antibodies (intact and single chain) specific to a wide variety of epitope tags are commercially available. These include but are not limited to antibodies against the DYKDDDDK (SEQ ID NO:5) epitope, c-myc antibodies (available from Sigma, St. Louis), the HNK-1 carbohydrate epitope, the HA epitope, the HSV epitope, the His_4, His_5, and His_6 epitopes that are recognized by the His epitope specific antibodies (*see, e.g.,* Qiagen), and the like.

[0077] In certain preferred embodiments, the target moiety is tagged with a hexahistidine (His_6) epitope tag that is bound by a Cu, Ni, or Co complex. One particularly preferred complex for binding His_6 tags is Ni-NTA (Ni- nitrilotriacetic acid). In certain particularly preferred embodiments, the affinity tag is a biotin which can then be captured by avidin, streptavidin, or variants thereof.

[0078] The affinity tagged target moiety is contacted with the phage- or bacterial display library, *e.g.*, in solution. Where suitable binding polypeptides exist in the library the target moieties are bound thereby forming a target moiety/binding polypeptide complex. The bound complexes can be recovered from solution phase by the use of an affinity matrix (*e.g.* a resin or other substrate attached to a ligand that binds to the affinity tag on the target moieties). Once isolated, the assay proceeds as with the target moieties provided attached to a substrate.

[0079] The target moieties binding polypeptides are isolated thereby providing a preselected set of potential binding proteins. The bound library members can then be separated (*e.g.* eluted) from the target moieties by the use of standard methods well known to those of skill in the art (*e.g.* using denaturing reagents, high salt, chaotropic reagents, and the like).

Contacting members of the preselected set of potential binding partners with one or more target proteins.

[0080] In preferred embodiments, the methods of this invention involve a second screening assay. In this assay, the preselected set of potential binding partners is again probed with the one or more target moieties to identify which members of the potential binding partners bind (*e.g.* specifically bind) to particular target moieties.

[0081] In preferred embodiments, the second assay is a different format from the first assay. In particularly preferred embodiment, however, the preselected members of the display library (preselected set of potential binding partners) is provided in a "spatially addressable" format. This permits individual members of the library that screen positive (for specific target binding) in the second screen to be detected and discriminated from each other. Such assays are thus preferably "inclusive" selecting for all binding partners rather than "exclusive" screening for a single one or few optimal binding partners.

[0082] Numerous assays are suitable. In one particular preferred embodiment, the second screen is a conventional cDNA expression library screening method. In this instance, the expressed cDNA library is immobilized on a solid substrate (*e.g.* blotted onto a membrane) and then probed with the one or more targets. Targets that specifically bind to the library members are identified and the binding members are optionally sequenced.

[0083] In preferred embodiments, the target moieties are labeled with a detectable label. Detectable labels suitable for use in the present invention include any composition detectable by spectroscopic, photochemical, biochemical, immunochemical, electrical, optical or chemical means. Useful labels in the present invention include biotin for staining with labeled streptavidin conjugate, magnetic beads (*e.g.*, Dynabeads^{TM}), fluorescent dyes (*e.g.*, fluorescein, texas red, rhodamine, green fluorescent protein, and the like, *see, e.g.*, Molecular Probes, Eugene, Oregon, USA), radiolabels (*e.g.*, ^{3}H, ^{125}I, ^{35}S, ^{14}C, or ^{32}P), enzymes (*e.g.*, horse radish peroxidase, alkaline phosphatase and others commonly used in an ELISA), and colorimetric labels such as colloidal gold (*e.g.*, gold particles in the 40 -80 nm diameter size range scatter green light with high efficiency) or colored glass or plastic (*e.g.*, polystyrene, polypropylene, latex, etc.) beads. Patents teaching the use of such labels include U.S. Patent Nos. 3,817,837; 3,850,752; 3,939,350; 3,996,345; 4,277,437; 4,275,149; and 4,366,241.

[0084] A fluorescent label is preferred because it provides a very strong signal with low background. It is also optically detectable at high resolution and sensitivity through a quick scanning procedure.

[0085] The label can be coupled to the target moiety prior to, during, or after the binding assay. So called "direct labels" are detectable labels that are directly attached to or incorporated into the target moiety prior to the binding assay. In contrast, so called "indirect labels" are joined to the target moiety/binding protein complex after binding. Often, the indirect label is attached to a second binding moiety that specifically binds to the target moiety or to a tag attached thereto. Thus, for example, the target moiety can be biotinylated before the screening assay. After hybridization, an avidin-conjugated fluorophore will bind the biotin bearing complexes providing a label that is easily detected. For a detailed review of methods of labeling nucleic acids and detecting labeled hybridized nucleic acids see *Laboratory Techniques in Biochemistry and Molecular Biology, Vol. 24: Hybridization With Nucleic Acid Probes*, P. Tijssen, ed. Elsevier, N.Y., (1993)).

[0086] It will be recognized that fluorescent labels are not to be limited to single species of organic molecules, but include inorganic molecules, multi-molecular mixtures of organic and/or inorganic molecules, crystals, heteropolymers, and the like. Thus, for example, CdSe-CdS core-shell nanocrystals enclosed in a silica shell can be easily

derivatized for coupling to a biological molecule (Bruchez *et al.* (1998) *Science,* 281: 2013-2016). Similarly, highly fluorescent quantum dots (zinc sulfide-capped cadmium selenide) have been covalently coupled to biomolecules for use in ultrasensitive biological detection (Warren and Nie (1998) *Science*, 281: 2016-2018).

Kits.

[0087] In still another embodiment, this invention provides kits for the practice of the methods described herein. Preferred kits include one or more components of a display library (*e.g.* phage display, bacterial display, yeast display, eukaryotic virus display library, direct plasmid display library, *etc.*) and instructional materials providing protocols for the assays disclosed herein.

[0088] While the instructional materials typically comprise written or printed materials they are not limited to such. Any medium capable of storing such instructions and communicating them to an end user is contemplated by this invention. Such media include, but are not limited to electronic storage media (e.g., magnetic discs, tapes, cartridges, chips), optical media (*e.g.,* CD ROM), and the like. Such media may include addresses to internet sites that provide such instructional materials.

TAIS Database.

[0089] In certain embodiments, this invention contemplates the use of a database to permit storage, retrieval, and management of TAIS data. Thus, for example, such a database can records showing amino acid sequence or identity of a member of a set of potential binding partners or proteins that interact with a one or more particular targets.

[0090] An illustration of an entry in such a database is provided in Figure 4. The term database refers to a means for recording and retrieving information. In preferred embodiments the database also provides means for sorting and/or searching the stored information. The database can comprise any convenient media including, but not limited to, paper systems, card systems, mechanical systems, electronic systems, optical systems, magnetic systems or combinations thereof. Preferred databases include electronic (e.g. computer-based) databases. Computer systems for use in storage and manipulation of databases are well known to those of skill in the art and include, but are not limited to

"personal computer systems", mainframe systems, distributed nodes on an inter- or intra-net, data or databases stored in specialized hardware (e.g. in microchips), and the like.

EXAMPLES

[0091] The following examples are offered to illustrate, but not to limit the claimed invention.

Example 1

Use Of TAIS To Study Protein-Protein Interactions

[0092] Results from screening of a T7 cDNA library derived from the normal human brain (NOVAGEN. Cat. #70637-3. (2001)) are presented and discussed below to demonstrate the potential of TAIS in mapping of protein-protein interactions. SH3, PDZ and WW domains of the Abl, Src, Crk, PSD95 and Nedd4 proteins have been used as test targets. In total, 12 novel putative and 2 previously described interactions have been identified by TAIS for these well studied protein interaction modules.

[0093] Combinatorial peptide libraries displayed on the phage or synthesized chemically have proved to be an excellent tool to define ligand preferences of peptide interaction modules (Cheadle *et al.* (1994) *J Biol Chem* 269: 24034-24039; Rickles *et al.* (1994) *Embo J* 13: 5598-5604; Sparks *et al.* (1996) *Proc. Natl. Acad. Sci., USA,* 93: 1540-1544; Kay *et al.* (2000) *FEBS Lett* 480, 55-62). The recognition consensus of an individual domain can be inferred by analyzing amino acid sequences of peptides selected from a random peptide library by the domain in question (Sparks *et al.* (1996) *Proc. Natl. Acad. Sci., USA,* 93: 1540-1544; Kay *et al.* (2000) *FEBS Lett* 480, 55-62). Defining the recognition consensus facilitates identification of potential interacting partners of the domain in protein databases (Kurakin *et al.* (1998) *J Pept Res* 52: 331-337) and/or mapping of its interaction sites within known partners (*Id.*). However, since combinatorial peptide repertoires are artificial, it is not clear how accurate the inferred consensus reflects natural interacting sequences and, often, the consensus defined in this way is too broad to limit the number of potential interactors in databases to a manageable quantity. The advent of cDNA libraries displayed on phage provides an opportunity to search natural peptide repertoires in order to map interacting partners and to refine recognition consensuses.

[0094] We demonstrate here that TAIS when applied to cDNA libraries allows rapid and simultaneous exploration of combinatorial and natural peptide repertoires with protein interaction modules as targets. This feature makes TAIS an efficient tool for both direct mapping of protein-protein interactions and studies aiming to characterize molecular recognition properties of protein interaction modules.

Results and discussion

The method.

[0095] A cDNA library derived from normal human brain was used in all presented screens (NOVAGEN. Cat. #70637-3. (2001), Novagen, Inc.). The library was generated using purified poly(A)$^+$ mRNA from the brain tissue as a template to create first strand cDNAs, which in turn served as templates for the synthesis of double stranded cDNA fragments. In both cases priming was random, thus the size and composition of resultant cDNA inserts vary greatly. The cDNA fragments longer than 300 base pair were directionally ligated to the C-terminus of gene product 10 of the lytic bacteriophage T7. Therefore, upon phage assembly a fragmented tissue-specific proteome is displayed on the surface of T7 phage as a C-terminal fusion to the major phage coat protein (NOVAGEN. OrientExpress cDNA Manual, TB247. (1999)). The reported diversities of tissue specific cDNA libraries from this source are in the order of $5x10^7$ primary recombinants, suggesting that even rare mRNA sequences are represented in these libraries with high probability (Soares *et al.* (1994) *Proc. Natl. Acad. Sci., USA,* 91: 9228-9232; Maniatis, *et al.* (1982) Molecular cloning. A Laboratory Manual. p. 225. (Cold Spring Harbor)). An important point to keep in mind is that theoretically, due to random priming, only one-third of all cDNA inserts result in the display of peptide sequences from the proteome. Two-thirds can be considered as "random" peptides originating from frameshifts upon ligation. In reality, the proportion of proteome sequences in the library is even less, due to priming from untranslated regions of mRNA. This structure of the library, however, is of great advantage when it is used to characterize ligand preferences of peptide interaction domains, for it allows parallel exploration of natural and artificial peptide repertoires.

[0096] To evaluate the new screening method, representatives of three families of peptide interaction modules, PDZ, SH3 and WW, were chosen as test targets. The domains

were derived from well-known proteins, such as PSD95, Src, Abl, Crk and Nedd4, for the following reasons: all five proteins have been the subjects of extensive protein interaction studies for a number of years performed by different groups and by different methods. In fact, PDZ and SH3 domains were first described in PSD95 and Src proteins, respectively, about a decade ago (Cho *et al.* (1992) *Neuron* 9: 929-942; Koch *et al.* (1991) *Science* 252: 668-674). A number of protein interactions mediated by these domains have been reported in the literature (Barfod *et al.* (1993) *J Biol Chem* 268: 26059-26062; Weng *et al.* (1994) *Mol Cell Biol* 14: 4509-4521; Kapeller *et al.* (1994) *J Biol Chem* 269: 1927-1933; Gout *et al.* (1993) *Cell* 75: 25-36; Weng *et al.* (1993) *J Biol Chem* 268: 14956-14963; Ren *et al.* (1993) *Science* 259: 1157-1161; Gertler *et al.* (1995) *Genes Dev* 9: 521-533; Ren *et al.* (1994) *Genes Dev* 8: 783-795; Knudsen *et al.* (1994) *J Biol Chem* 269: 32781-32787; Hasegawa *et al.* (1996) *Mol Cell Biol* 16: 1770-1776). In addition, ligand preferences of the tested domains have been characterized by screening of artificial peptide repertoires (Cheadle *et al.* (1994) *J Biol Chem* 269: 24034-24039; Rickles *et al.* (1994) *Embo J* 13: 5598-5604; Sparks *et al.* (1996) *Proc. Natl. Acad. Sci., USA,* 93: 1540-1544Sparks *et al.* (1994) *J Biol Chem* 269: 23853-23856; Rickles *et al.* (1995) *Proc. Natl. Acad. Sci., USA,* 92: 10909-10913; Yu *et al.* (1994) *Cell* 76: 933-945; Feng *et al.* (1994) *Science* 266: 1241-1247; Musacchio *et al.* (1994) *Nat Struct Biol* 1: 546-551; Wu *et al.* (1995) *Structure* 3: 215-226). The reported interactions were meant to serve as a positive control while known recognition consensuses of tested domains were expected to match sequences in peptides selected by TAIS from a cDNA library.

PDZ domains of PSD95.

[**0097**] PDZ domains were originally described as 80-100 amino acid conserved repeats within the post-synaptic density 95 protein (PSD95) (Cho *et al.* (1992) *Neuron* 9: 929-942; Kornau *et al.* (1997) *Curr Opin Neurobiol* 7: 368-373). The prototypical PDZ domain protein PSD95 comprises three PDZ domains at the N-terminus followed by an SH3 domain and an inactive guanylate kinase domain (Cho *et al.* (1992) *Neuron* 9: 929-942). By providing an architectural and functional scaffold via its multiple protein interaction modules, it is thought to orchestrate assembly and function of molecular complexes responsible for neurotransmission and synaptic plasticity at the post-synaptic

membranes (Kennedy (2000) *Science* 290: 750-754; El-Husseini *et al.* (2000) *Science* 290: 1364-1368).

[0098] In their classical mode, PDZ domains recognize and bind to the extreme C-terminal sequences of interacting partners with reported affinities from high nanomole to low micromole range (Niethammer *et al.* (1998) *Neuron* 20: 693-707; Songyang *et al.* (1997) *Science* 275: 73-77). Specificity of binding within the PDZ family is thought to be defined by 3-5 amino acids preceding the C-terminal residue (Songyang *et al.* (1997) *Science* 275: 73-77; Stricker *et al.* (1997) *Nat Biotechnol* 15: 336-342; Doyle *et al.* (1996) *Cell* 85: 1067-1076) Ligand preferences of different PDZ domains have been studied mostly with chemically synthesized, rather than displayed peptide libraries, due to historical difficulties in displaying free carboxy-termini on the filamentous phage (Songyang *et al.* (1997) *Science* 275: 73-77; Stricker *et al.* (1997) *Nat Biotechnol* 15: 336-342; Doyle *et al.* (1996) *Cell* 85: 1067-1076; Hoffmuller *et al.* (1999) *Angew. Chem. Int. Ed.* 38: 2000-2004). Analysis of the ligand preferences of several PDZ domains resulted in inferred recognition consensus sequences, which, though fitted well when compared to natural binding sites discovered by other methods, were of limited predictive power due to a too broadly defined consensus.

[0099] A cDNA human brain library displayed on the T7 phage was TAISed with the N-terminal fragment of PSD95 comprising three PDZ domains as a target (PSD95-PDZ(1+2+3)). The pre-selected cDNA library formed about 1500 plaques on a bacterial lawn, when plated on two 150 mm Petri dishes. 11 clones gave positive signals on the membranes after plaque lift and screening of membranes with biotinylated PSD95-PDZ(1+2) complexed to streptavidin-alkaline phophatase (AP) conjugate (see Fig. 2).

[0100] Sequences of the peptides displayed on the phages that gave positive plaques are numbered PD1 through PD11 and shown in Table 1, together with their relative affinity ranks towards PSD95-PDZ(1+2+3).

[0101] Table 1. Results of screening of a phage-displayed human brain cDNA library with an N-terminal fragment of PSD95 comprising its three PDZ domains. Sequences of polypeptides displayed by phages from positive plaques along with their relative affinity ranks towards the target and identities of the respective cDNA inserts. FS – frameshift, ? – undefined, DGKζ - diacylglycerol kinase zeta, UTR - untranslated region, ">" – denotes free carboxylate group.

Phage Clone	Displayed Peptide	Binding	cDNA	SEQ ID NO
PD1, PD4, PD5	SRSTWATWQSPIYT**K**KPKTSQV>	++++++++	?	6
PD2	SKI**K**YFR**E**SII>	++++++++	?	7
PD3	SSRQHYQMIQ**R**EDQ**E**TAV>	++++++++	DGK ζ	8
PD6	SSLRLETGV>	+	?	9
PD7	LRNGRRECHIHLWKQRGQMRISAV>	+++	?	10
PD8, PD9	PASAQPAAGDPVPAPAVLLGWTLV>	++	FS	11
PD10	SSRKCRQCFHKSKCTVI>	+	UTR	12
PD11	SSLV>	+/-	FS	13
Minimum Consensus	xRxSxV> K T I			14
Refined Consensus (PD1-PD5)	KxxRESxV> R K T I			15

[0102] The minimum consensus sequence of peptides that bound PSD95-PDZ(1+2) can be readily defined as (R/K)-x-(S/T)-x-(V/I)-COOH (SEQ ID NO:16). This consensus matches well with C-terminal sequences of known interacting partners of PSD95, such as inward rectifier K^+ channel (Kir2.3: NISYRRESAI-COOH, SEQ ID NO:17) (Cohen *et al.* (1996) *Neuron* 17: 759-767), embryonic skeletal muscle sodium channel (SkM2: SPDRDRESIV-COOH, SEQ ID NO:18) (Gee *et al.* (1998) *J Neurosci* 18: 128-137) and Shaker-type potassium channel (Kv1.4: SNAKAVETDV-COOH, SEQ ID NO:19) (Kim *et al.* (1995) *Nature* 378: 85-88). It is also notably similar to the consensus previously reported for syntrophin PDZ domains, (R/K)-E-(S/T)-x-V-COOH (SEQ ID NO:20, Gee *et al.* (1998) *J Neurosci* 18: 128-137) (see below). Significantly, 2 out of the 3 strongest binders have a conserved glutamate at ligand position –3 and all of the strongest binders (PD1, PD2, PD3) have a positively charged residue at the position – 7*, lysine or arginine.

(Conventionally, residues of a peptide ligand for PDZ domains are numbered so that the extreme C-terminal residue position is designated as 0 and positions of preceding residues towards the N-terminus are –1, -2, -3, -4 and so on). Therefore, a refined binding consensus of PSD95-PDZ(1+2) can be described as (K/R)-x-x-(R/K)-E-(S/T)-x-(V/I)-COOH (SEQ ID NO:21). It should be noted that residues of PDZ ligands distant from the C-terminus, such as -7 or -8 positions, have been implicated previously as contributing to the binding specificity, at least in the cases of some PDZ domains (Niethammer *et al.* (1998) *Neuron* 20: 693-707; Songyang *et al.* (1997) *Science* 275: 73-77). Collectively, our data and that of others suggest that the recognition mechanism of PDZ domains may be more complex than currently believed, and may involve additional specificity determinants proximal to the C-terminal five amino acids.

[0103] The cDNA library can be viewed as a combinatorial library that is highly enriched in natural peptide sequences. The latter provide a unique internal reference about physiologically relevant affinities and specificities when the library is assayed for the interaction with a target protein. Taking into account these considerations, we believe that PD1 and PD2 peptides, that bound strongly to PSD95-PDZ(1+2+3), may represent novel proteins that interact with PSD95. The nucleotide sequences of PD1 and PD2 inserts match a number of human ESTs and genomic sequences with no assigned open reading frame (not shown). The biochemical characterization of corresponding full-length cDNA products can substantiate this putative activity/function.

[0104] When used in pattern searches of the SWISS-PROT database, (K/R)-x-x-(R/K)-E-(S/T)-x-(V/I)-COOH (SEQ ID NO:22) consensus matches sequences in about 30 proteins, a reasonable number to assess experimentally. Therefore, the potential interacting partners that are missed in a physical screen due to their absence, low abundance or sensitivity to proteolysis can be retrieved by bioinformatic tools using the recognition consensus of the target refined by TAIS.

[0105] We have used the PSD95-PDZ(1+2+3) recognition consensus defined by TAIS, [KR]-x-x-[QRK]-E-[ST]-x-[VI]-COOH (SEQ ID NO:23), in homology searches of SWISS and TrEMBL databases. Proteins with the C-termini conforming to the query consensus are grouped below according to their functionality or their host (*see,* e.g. Table 2 and Table 3).

[0106] Table 2. Proteins with the C-termini conforming to the query consensus (TAIS, [KR]-x-x-[QRK]-E-[ST]-x-[VI]-COOH, SEQ ID NO:24) grouped according to their functionality or their host.

Protein	SEQUENCE	SEQ ID NO
Receptors:	RNLRETDI	25
A1AD Rabbit 002666	rabbit	
Alpha-1D adrenergic receptor)	LADYSNLRETDI	26
Oryctolagus cuniculus (Rabbit) 569-576	human	
Microtubule Associated Motor		
KF1B_HUMAN O60333	KAGRETTV	27
Kinesin-like protein KIF1B (Klp)		
Homo sapiens (Human)		
SPLICE ISOFORM 3 1146-1153		
Q60575 (KF1B_MOUSE)	KAGRETTV	28
Kinesin-like protein KIF1B [Mus musculus (Mouse)]		
SPLICE ISOFORM 3 OF Q60575 1143-1150		
Q9H8Z3 CDNA FLJ13122 fis,	KAGRETTV	29
clone NT2RP3002688 weakly similar to mouse kinesin-like protein (Kif1b) [Homo sapiens (Human)].		
122-129 KAGRETTV		
AAK33008	KGSRETAV	30
Kinesin-like protein Kif1b alpha [Brachydanio rerio (Zebrafish) (Danio rerio)]. 1154-1161 KGSRETAV		
HUMAN VIRAL PROTEINS.		
TAT_HTL1A P03409	KHFRETEV	31
Trans-activating transcriptional regulatory protein (X-LOR protein) (PX protein).		
Human T-cell leukemia virus type I (strain ATK & Caribbean isolate) (HTLV-I) 351-358		
VE6_HPV45 P21735	RRRRETQV	32
E6 protein.		
Human papillomavirus type 45 (conforms for types 56, 68, 70, ME180, 151-158		
O73280	KRPRESDI	33
GAG polyprotein [Contains: core protein(s) P24] (Fragment).		
Human immunodeficiency virus type 2118-125		
US32_HCMVA P09708	RRHRETYV	34
Hypothetical protein HHRF7.		
Human cytomegalovirus (strain AD169) 176-183		

HUMAN BACTERIAL PARASITES'S PROTEINS

Y3C2_MYCTU O53600 Hypothetical 13.3 kDa protein Rv3922c. [Mycobacterium tuberculosis] 113-120	RGERESFV	35
Q9R886 Hypothetical 3.6 kDa protein (Fragment). Chlamydia trachomatis 23-30	RQNKETKI	36
O84715 ROD SHAPE PROTEIN-SUGAR KINASE. Chlamydia trachomatis 359-366	KKRKESLV	37
Q9PLL7 Cell shape-determining protein MreB. Chlamydia muridarum 359-366	KKRKESLV	38

SIGNALING (EXOCYTOSIS - RAL FAMILY BINDING PROTEIN)

Q62796 RalBP1 Rattus norvegicus (Rat) 640-647	KDRKETPI	39
Q15311 RLIP76 protein (Similar to ralA binding protein 1). Homo sapiens (Human) 648-655	RDRKETSI	40
Q62172 RIP1 protein. Mus musculus (Mouse) 641-648	KDRKETPI	41
Q9DDA3 RalB-binding protein (Fragment). Xenopus laevis (African clawed frog) 604-611	KDWKETLI	42

SIGNALING (SECOND MESSAGER METABOLISM)

Q13574 (KDGZ_HUMAN) Diacylglycerol kinase, zeta (EC 2.7.1.107) Diglyceride kinase) (DGK-zeta) (DAG kinase zeta) [Homo sapiens (Human)]. 1110-1117	REDQETAV	43
SPLICE SHORT ISOFORM OF Q13574 921-928	REDQETAV	44
Q08560 (KDGZ_RAT) Diacylglycerol kinase, zeta (EC 2.7.1.107) (Diglyceride kinase) (DGK-zeta) (DAG kinase zeta) (DGK-IV) (104 kDa diacylglycerol kinase) [Rattus norvegicus (Rat)]. 922-929	REDQETAV	41
Q91YS0 Similar to diacylglycerol kinase (Fragment) [Mus musculus (Mouse)]. 451-458	REDQETAV	45

[0107] Table 3. Other proteins with the C-termini conforming to the query consensus (TAIS, [KR]-x-x-[QRK]-E-[ST]-x-[VI]-COOH, SEQ ID NO:46).

Accession No	Description
Q920A7	(AF31_MOUSE) AFG3-like protein 1 (EC 3.4.24.-) [Mus musculus (Mouse)].
P51464	(ARLY_RANCA) Argininosuccinate lyase (EC 4.3.2.1) (Arginosuccinase) (ASAL) [Rana catesbeiana (Bull frog)].
Q9P280	KIAA1448 protein (Fragment) [Homo sapiens (Human)].
Q9UIZ9	Cellular DNA/human papillomavirus proviral DNA [Homo sapiens (Human)].
Q9VHT6	CG9626 protein [Drosophila melanogaster (Fruit fly)].
Q9TR85	DNA ligase II (Fragment) [Bos taurus (Bovine)].
Q9LVM3	Genomic DNA, chromosome 5, P1 clone:MCK7 [Arabidopsis thaliana (Mouse-ear cress)].
Q90YA3	6-phosphofructokinase [Gallus gallus (Chicken)].
AAM32072	Conserved protein [Methanosarcina mazei Goe1].
YC11_AQUAE	Hypothetical protein AQ_1211. [Aquifex
O67264	aeolicus]
O29148	DNA-DIRECTED RNA POLYMERASE, SUBUNIT E' (RPOE1) [Archaeoglobus fulgidus].
Q08300	CHROMOSOME XV READING FRAME ORF YOL159C.[Saccharomyces cerevisiae] (Baker's yeast)
O80591T27I1.2	protein. [Arabidopsis thaliana] (Mouse-ear cress)

[0108] The interspecies conservation of the TAIS-defined PSD95-PDZ(1+2+3) recognition consensus at the C-termini of diacylglycerol kinase zeta (DGKζ), kinesin-like protein KIF1B and Ral-binding protein makes them strong candidates for being physiological interacting partners of PSD95. Notice that the C-terminus of human DGKζ interacted *in vitro* with PSD95-PDZ(1+2+3). The presence of PSD95-binding sequences at the C-termini of proteins from different Chlamydia strains may indicate on interesting and unexpected molecular connections exploited by this intracellular parasite, which is implicated in a host of human ailments such as trachoma, arthritis, Alzheimer's disease among others.

[0109] Figure 3 illustrates another example of PDZ domain profiling. The x-axis shows an array of individual phages selected to bind a number of different PDZ domains, while the y-axis shows the relative affinities of individual phages to the 2nd PDZ domains

from SAP97 and SAP90 in an ELISA-type assay. Table 4 illustrates PDZ2 domain best binders.

[0110] Table 4. SAP97_PDZ2 domain best binders and SAP90_PDZ2 domain best binders.

SAP97_PDZ2 domain best binders		SEQ ID NO
#1	PGQHGESPSLLKTHKKISWV>	47
#45	EKCHQSYSHSIYERKKWTDV>	48
#21	SQPQEPVPVALQGVRRETRV>	49
#48	GLGKSSRSLWGGEWHLETYV>	50
#32	WAGPRKAGPLGAAPGRATLV>	51
#30	NCCVNEPDTLLNLSPRWTMV>	52
consensus	WTxV E I A	53
SAP90_PDZ2 domain best binders		
#38	PARPTWGNSISTKNTKISWV>	54
#45	EKCHQSYSHSIYERKKWTDV>	55
#1	PGQHGESPSLLKTHKKISWV>	56
#30	NCCVNEPDTLLNLSPRWTMV>	57
#32	WAGPRKAGPLGAAPGRATLV>	58
#46	RVPRRGQDFCSGFPGCWTQV>	59
consensus	WTxV> IS A	60
Peptides that bound strongly to SAP97_PDZ2, but only weakly to SAP90_PDZ2 share glutamic acid (**E**) at position "-3" (shown in bold)		
#21	VSQPQEPVPVALQGVRR**E**TRV>	61
#67	ARAGGGFEDASLGFGGR**E**TAV>	62
#48	GLGKSSRSLWGGEWHL**E**TYV>	63

> indicates carboxy terminus.

[0111] Thus, despite the high degree of similarity between PDZ2 domains of SAP90 and SAP97 (84% of identity and 92% of similarity) their binding specificities are overlapping, but not identical.

[0112] The accumulation and arraying of peptides (on phages) that have been preselected to bind PDZ domains allows the rapid cross-comparison of PDZ domain specificities to reveal their unique binding characteristics. A rrays of PDZ-binding phages are easily propagated in multi-well formats and can be used for the rapid characterization of novel PDZ domains omitting library screening.

DGK ζ.

[0113] Diacylglycerol kinase zeta (DGK ζ) was identified in the screen as a novel putative interacting partner of PSD95. DGKs metabolize a lipid second messenger diacylglycerol (DAG), thus negatively regulating DAG-induced cell responses (Topham *et al.* (1999) *J Biol Chem* 274: 11447-11450; Sanjuan (2001) *J Cell Biol* 153:207-220). DAG is generated by phosphoinositide-specific phospholipase C (PLC) isoforms and accumulates locally and transiently upon activation of a large number of growth factor and other cell surface receptors (Bishop and Bell (1986) *J Biol Chem* 261: 12513-12519; Rhee (2001) *Annu Rev Biochem* 70: 281-312). We speculate that PSD95 by interacting with the C-terminus of DGK ζ maintains a diacylglycerol kinase activity as a component of signal-processing machinery at the postsynaptic membranes of glutamatergic synapses, where group I metabotropic glutamate receptors (mGluRs) (Skeberdis *et al.* (2001) *Neuropharmacology* 40: 856-865; Hannan *et al.* (2001) *Nat Neurosci* 4: 282-288; Reyes-Harde *et al.* (1998) *Neurosci Lett* 252: 155-158) and, conceivably, tyrosine kinases such as ErbB4 (Huang *et al.* (2000) *Neuron* 26: 443-455; Huang *et al.* (2001) *J Biol Chem* 276: 19318-19326) are coupled to the PLC cascade. Localization of DGK in close proximity to its substrate, rather than its shuttling between the cytosol and membrane, would allow higher frequencies of signal relay dependent on DAG generation.

[0114] Interestingly, DGK ζ has been recently reported by Gee and colleagues to bind via its C-terminus to PDZ domains of syntrophins (Hogan *et alo.* (2001) *J Biol Chem* 276: 26526-26533). Based on the similarities in critical residues between syntrophin PDZ domains and the second PDZ domain of PSD95, as well as their cross-reactivity to a number of targets, the same authors earlier suggested that these domains may compete for similar ligands (Gee *et al.* (1998) *J Biol Chem* 273: 21980-21987). Their suggestion is compatible with our findings as well as with the recently reported solution structure of the PSD95-PDZ2 domain, which most closely resembles that of α1-syntrophin (an rmsd value

of 1.36 angstrom for the entire PDZ domains) (Tochio *et al.* (2000) *J Mol Biol* 295: 225-237).

WW3 domain of Nedd4.

[0115] WW domains, named after two tryptophan residues highly conserved in the family, are protein interaction modules recognizing short proline-rich sequences (Bork and Sudol (1994) *Trends Biochem Sci* 19: 531-533). They are found in proteins with functions as diverse as cell cycle control, pre-mRNA 3' end formation and targeted protein degradation (Sudol and Hunter (2000) *Cell* 103: 1001-1004; Lu *et al.* (1999) *Science* 283: 1325-1328; Morris *et al.* (1999) *J Biol Chem* 274: 31583-31587; Morris and Greenleaf (2000) *J Biol Chem* 275: 39935-39943; Verdecia *et al.* (2000) *Nat Struct Biol* 7: 639-643). On the basis of ligand preferences, WW domains are segregated into at least five classes (Kasanov *et al.* (2001) *Chem Biol* 8: 231-241): Class I prefers peptide ligands with a core motif PPxY (Chen and Sudol (1995) *Proc. Natl. Acad. Sci., USA,* 92: 7819-7823); Class II – PPLP (Bedford *et al.* (1997) *Embo J* 16: 2376-2383); Class III – PxxGMxxPP (Bedford *et al. Proc. Natl. Acad. Sci., USA,* 95: 10602-10607); Class IV – (pS/pT)P (Lu *et al.* (1999) *Science* 283: 1325-1328); and Class V – RxPPGPPPxR (Komuro *et al.* (1999) *J Biol Chem* 274: 36513-36519).

[0116] The third WW domain of the mouse Nedd4 ubiquitin protein ligase (Nedd4-WW3) (Kumar *et al.* (1997) Genomics 40: 435-443) has been used as a target to screen a human brain cDNA library by TAIS. The peptides selected by the Nedd4-WW3 from the cDNA library, together with the names of the proteins from which they are derived, are shown in Table 5. The Nedd4-WW3 belongs to the Class I WW domains and a characteristic Class I core recognition motif PPxY is readily discernible in all selected peptide sequences (underlined in Table 5). In fact, if the selected peptides are subjected to unbiased analysis by software that is "unaware" of WW domain family ligand preferences and simply identifies homologous stretches in unrelated peptide sequences, the only common motif between four selected peptides is PPPY(E/D)EV (SEQ ID NO:64, Table 7).

[0117] Table 5. Results of screening of a human brain cDNA library with the third WW domain of Nedd4 ubiquitin ligase as a target. Sequences and identities of polypeptides selected by the Nedd4-WW3 domain. The PPxY core recognition motif of WW domain family is underlined.

Protein	Sequence	SEQ ID NO
>AF327246.1 /gene="**SCN2A**" /product="voltage-gated sodium channel type II alpha subunit" /protein_id="AAG53413.1"	PPXYESL-WW3	65
	STPEKTDMTPSTTS**PPSYDSV**TKPEKEKFEKDKSEKEDKGKDIRESKK	66
>XM_001374 /gene="**LAPTM5**" /product="Lysosomal-associated multispanning membrane protein-5" /protein_id="XP_001374.2"	P LPxYxEA-WW2 ?	67
	SSYRLIKCMNSVEEKRNSKMLQKVVLPSYEEALSLPSK-	68
	PPxYESL-WW3	69
	-TPEGGPA**PPPYSEV**	68 cont'd
>AF320999 /gene="**Nogo-A**" /product="Nogo-A protein short form" /note="alternatively spliced" /protein_id="AAG40878.1"	PPxYESL-WW3	70
	390_SAVPSAGASVIQPSSSPLEASSVNYESIKHEPEN**PPPYEEA**MSVSLKK	71
	VSGIKEEIKEPENINAALQETEAPYISIACDLIKETKLSAEPAPDFSDYSEM	
	AK-491	
>AL137579.1/gene="DKFZp434A1010" /note="**N-chimaerin homolog** F25965_3, alternative spliced" /protein_id="CAB70821.1"	GPRTPHRVPGPWGPPEPLLLYRAAPPAYGRGGELHRGSLYRNGGQRGEGAGP	72
	PPPYPTPSWSLHSEGQTRSYC>	

This motif is in good agreement with a recognition consensus for Nedd4-WW3, PPxYES(L/M) (SEQ ID NO:73), defined independently by artificial peptide repertoire analysis (Kay *et al.* (2000) *FEBS Lett* 480, 55-62). A contribution of peptide ligand residues C-terminal to the PPxY core to binding energy and specificity of interaction mediated by the Nedd4-WW3 domain has been demonstrated convincingly by the recently published solution structure of the Nedd4-WW3 domain complexed with the peptide derived from the β subunit of the epithelial sodium channel (EnaC), TLPIPGTPPPNYDSL (SEQ ID NO:74,

Kanelis *et al.* (2001) *Nat Struct Biol* 8: 407-412). It should be noted that two PPxY motifs in the chimaerin homologue peptide, PPAYGRG (SEQ ID NO:75) and PPPYPTP (SEQ ID NO:73), do not conform well to the extended recognition consensus of Nedd4-WW3, PPxYES(L/M) (SEQ ID NO:76). A conceivable explanation is that they, or one of them, represent secondary recognition motif(s) for Nedd4-WW3 domain. Alternatively, the chimaerin homologue may be a false positive picked up due to avidity provided by two closely situated PPxY core motifs.

[0118] Nedd4 has been proposed to control stability and/or turnover of ENaC at the cell surface, presumably by directing its ubiquitination, which is followed by endocytosis and degradation of the channel (Staub *et al.* (1996) *Embo J* 15: 2371-2380; Staub *et al.* (1997) *Embo J* 16: 6325-6336; Abriel *et al.* (1999) *J Clin Invest* 103: 667-673). WW domains of Nedd4 are thought to function in this system as targeting modules, since they specifically bind subunits of ENaC. Deletions or point mutations in the PPxY motif on β or γ subunits of ENaC are associated with a hereditary form of hypertension, Liddle's syndrome, which is characterized by deregulated activity of ENaCs (Shimkets *et al.* (1994) *Cell* 79: 407-414. A number of authors have proposed that Nedd4 and Nedd4-like proteins, due to their unique structure comprising a membrane targeting C2 domain, two to four WW domains and a C-terminal HECT -type ubiquitin protein ligase domain, are strong candidates for regulators of ubiquitin-mediated turnover of many membrane proteins (Jolliffe *et al.* (2000) *Biochem J* 351 Pt 3, 557-565; Abriel *et al.* (2000) *FEBS Lett* 466: 377-380; Rotin *et al.* (2000) *J Membr Biol* 176: 1-17). Indeed, the yeast ubiquitin-protein ligase Rsp5p, a homologue of mammalian Nedd4 and Itch, is required for the ubiquitination and subsequent internalization of several plasma membrane proteins, including the alpha-factor receptor (Ste2p) (Hicke *et al.* (11996) *Cell* 84: 277-287; Dunn and Hicke (2001) *Mol Biol Cell* 12: 421-435), uracil permease (Galan *et al.* (1996) *J Biol Chem* 271: 10946-10952), general amino acid permease (Springael *et al.* (1998) *Mol Biol Cell* 9: 1253-1263) and others (Hicke (1997) *Faseb J* 11: 1215-1226). Therefore, it is reasonable to assume an existence of multiple Nedd4 targets in the cell.

[0119] Nogo-A, lysosomal-associated multispanning membrane protein 5 (LAPTM5), type II α subunit of voltage gated sodium channel (SCN2A) and a novel human protein with homology to chimaerin have been identified by TAIS as novel putative

interaction partners of Nedd4 (Table 5). Notably, all but chimaerin homolog are membrane proteins.

Nogo-A

[0120] *Nogo-A* has been recently cloned independently by three different teams as a long sought myelin inhibitor of regenerating axons, and is the subject of intensive studies assessing the contribution of Nogo to the failure of axonal regeneration in the adult CNS (Prinjha *et al.* (2000) *Nature* 403: 383-384; GrandPre *et al.* (2000) *Nature* 403: 439-444; Chen *et al.* (2000) *Nature* 403: 434-439). A possible regulation of Nogo-A through ubiquitin-mediated degradation pathways may provide a fruitful framework for studies aiming to understand the molecular basis of CNS regeneration and plasticity.

LAPTM5

[0121] *LAPTM5* was originally cloned as a lysosomal membrane associated protein that interacts with ubiquitin, developmentally downregulated and preferentially expressed in adult tissues with high cell turnover (Adra *et al.* (1996) *Genomics* 35: 328-337). The function of the protein is unknown. The rat homologue of mouse LAPTM5, Granule Cell Death – 10 protein (GCD-10), is up-regulated in microglia in response to degeneration and cell death of neurons *in vitro* and *in vivo* and is involved in the dynamics of lysosomal membranes of activated microglia (Origasa *et al.* (2001) *Brain Res Mol Brain Res* 88: 1-13). To our knowledge, the present report is a first link that connects ubiquitin-dependent endocytic machinery to the integral lysosomal membrane protein, thus shedding light on the receiving end of this degradation pathway. Indeed, several authors have suggested a function for the Nedd4 yeast orthologue Rsp5p and its WW domains downstream of plasma membrane protein ubiquitination (Rotin *et al.* (2000) *J Membr Biol* 176: 1-17; Dunn and Hicke (2001) *Mol Biol Cell* 12: 421-435; Beck *et al.* (1999) *J Cell Biol* 146: 1227-1238). Recent report on localization of Rsp5p at multiple sites within endocytic pathways, such as plasma membrane invaginations, late endosomes and perivacuolar sites, supports the notion of a direct role for Rsp5p and ubiquitin in protein sorting and trafficking (Wang *et al.* (2001) *Mol Cell Biol* 21: 3564-3575). The ability of LAPTM5 to interact with both ubiquitin and Nedd4 suggests a potential role for LAPTM5 as a lysosomal receptor for ubiquitinated cargo destined for destruction.

SCN2A

[0122] The ability of neurons to communicate by generation and propagation of action potentials along their axons is crucially dependent on activity of voltage-gated sodium channels (VGSC) (Armstrong and Hille (1998) *Neuron* 20: 371-380). Identification of SCN2A as a putative interaction partner of Nedd4 ubiquitin ligase is indicative of a possible role of ubiquitin-mediated degradation pathways in the control of neuronal VGSC stability and/or turnover. In fact, a conservation of a PPxY motif, a presumptive WW domain binding site, within the C-termini of a number of sodium channels, was noticed as early as 1996 by Einbond, and Sudol (1996) *FEBS Lett* 384: 1-8. The functional significance of this conservation has been confirmed by experimental data indicating that both ENaC and the cardiac voltage-gated Na+ channel H1 (SCN5A) are regulated by Nedd4 ubiquitin-protein ligase in a WW domain dependent manner (Abriel *et al.* (2000) *FEBS Lett* 466: 377-380). Table 6 shows results of screening of a human brain cDNA library with the third WW domain of Nedd4 ubiquitin ligase as a target.

[0123] Table 6 shows results of screening of a human brain cDNA library with the third WW domain of Nedd4 ubiquitin ligase as a target. Homologous sequences shared by polypeptides selected with Nedd4-WW3 domain as defined by the BLOCK MAKER algorithm (see, e.g., http://www.blocks.fhcrc.org/blockmkr/make_blocks.html).

	Sequence	ID	SEQ ID NO
15	PPSYDSV	SCN2A	77
53	PPPYPTP	N-chimaerin homolog	78
46	PPPYSEV	LAPTM5	79
35	PPPYEEA	Nogo-A	80
	PPPYEEV D	Consensus	81
	PPxYESL	Kay *et al.*(2000)	82

In Table 7 we show the C-termini of all proteins from the SWISS-PROT and TrEMBL databanks that share Nedd4 recognition site on the cardiac voltage-gated Na+ channel H1, PPSYDSV (SEQ ID NO:83).

[0124] Table 7 shows results of screening of a human brain cDNA library with the third WW domain of Nedd4 ubiquitin ligase as a target. C-terminal sequences of all proteins from SWISS-PROT and TrEMBL databases that share a PPSYDSV (SEQ ID NO:84), sequence (bold). Underlined are putative PEST sequences as defined by PESTfinder algorithm (http://www.at.embnet.org/embnet/tools/bio/PESTfind/). PPxYESL (SEQ ID NO:85, Kay et al. (2000) *FEBS Lett* 480: 55-62) and (P/L)PxYxEA (SEQ ID NO:86, Kasanov et al. (2001) *Chem Biol* 8: 231-241) recognition consensuses for Nedd4-WW3 and Nedd4-WW2 domains, respectively, as well as Nedd4-WW3 domain binding site on □EnaC (Kanelis et al. (2001) Nat Struct Biol 8: 407-412) are shown for comparison.VGSC – voltage gated sodium channel; CNS – central nervous system; PNS – peripheral nervous system. ">" – denotes carboxylate group.

Gene Name Accession Origin	Seq ID NO: No
VGSCs from heart:	
PLGPPSSSSISSTSF**PPSYDSV**TRATSDNLQVRGSDYSHSEDLADFPPSPDRDRESIV *Q14524* SCN5A human	87
RRSAPLSSSSISSTSF**PPSYDSV**TRATSDNLPVRASDYSRSEDLADFPPSPDRDRESIV *P15389* **SCN5A** rat	88
RRSGPLSSSSISSTSF**PPSYDSV**TRATSDNLPVRASDYSRSEDLADFPPSPDRDRESIV *Q9JJV9* SCN5A mouse	89
VGSCs from CNS:	
KLNENSTPEKTDMTPSTTS**PPSYDSV**TKPEKEKFEKDKSEKEDKGKDIRESKK *Q99250* **SCN2A** human	90
KLNENSTPEKTDVTPSTTS**PPSYDSV**TKPEKEKFEKDKSEKEDKGKDIRESKK *P04775* SCN2A rat	91
KLNGNSTPEKTDGSSSTTS**PPSYDSV**TKPDKEKFEKDKPEKESKGKEVRENQK *Q9NY46* SCN3A human	92
KLNGNSTPEKTDGSSSTTS**PPSYDSV**TKPDKEKFEKDKPEKEIKGKEVRENQK *P08104* SCN3A rat	93
VGSCs from PNS:	
DNVNSSSPEKTDATASTIS**PPSYDSV**TKPDKEKYEKDKTEKEDKGKDGKETKK *Q28644* none rabbit	94
VNENCALPDKSETASAASF**PPSYDSV**TRGLSDQINMSTSSSMQNEDEGTSKKVTAPGP *O46669* none dog	95
FMANSGLPDKSETASATSF**PPSYDSV**TRGLSDRANINPSSSMQNEDEVAAKEGNSPGPQ *Q63554* SNS rat	96
NVNENSSPEKTDVTASTIS**PPSYDSV**TKPDQEKYETDKTEKEDKEKDESRK *O08562* none rat	97
RLNGNSTTEKMDMTPSTAS**PPSYDSV**TKPSKEKHEKDKSEREDKGKDVRHNRK *Q9YGN7*	98

none	newt	
Other VGSCs:		
NVNENSSPEKTDATASTIS**PPSYDSV**TKPDQEKYETDKTEKEDKEKDESRK		99
Q62205		
SCN9A	mouse	
NVNENSSPEKTDATSSTTS**PPSYDSV**TKPDKEKYEQDRTEKEDKGKDSKESKK		100
Q15858		
HNE-NA	human	
ANDNGGLPDKSETASATSF**PPSYDSV**TRGLSDRANISTSSSMQNEDEVTAKEGKSPGPQ		101
Q62243		
none	mouse	

[0125] As one can see, the PPSYDSV (SEQ ID NO:102) sequence: i) is strictly conserved across species and between different alpha subunit isoforms of cardiac and neuronal VGSCs; ii) is embedded in sequences shown to be prerequisite for proteins degraded through ubiquitin-directed endocytosis, such as PEST sequences, multiple serines and threonines (phosphorylation acceptors) and lysines (ubiquitination acceptors); and iii) conforms well to recognition consensus of the Nedd4 WW3 domain, PPxYES(L/M) (SEQ ID NO:103), defined recently by a combinatorial peptide library approach (Kay *et al.* (2000) *FEBS Lett* 480, 55-62). Remarkable parallels in the control of ENaC and cardiac sodium channel by Nedd4 ubiquitin-protein ligase (Abriel *et al.* (2000) *FEBS Lett* 466: 377-380), strict conservation of the Nedd4-WW3 recognition sequence within C-termini of cardiac and neuronal voltage gated sodium channels and an *in vitro* interaction of Nedd4-WW3 with a C-terminus of alpha subunit of neuronal VGSC (as noted in the present paper) strongly suggest a role of the Nedd4 ubiquitin-mediated endocytotic pathway in the regulation of stability and/or turnover of neuronal VGSC. It is relevant that high expression of Nedd4 was demonstrated in the heart and nervous tissues (Staub *et al.* (1996) *Embo J* 15: 2371-2380).

Chimaerin homolog.

[0126] A novel protein homologous to human chimaerins has been identified by TAIS as a putative interaction partner of Nedd4. Homology to chimaerins is restricted to the first 85 out of 862 amino acids of the protein, which constitute a domain conserved in GTPase activators for Rho-like GTPases (RhoGAP domain). A role for Rho family GTPases has been demonstrated convincingly at different steps of endocytosis, intracellular sorting and trafficking, although the molecular mechanisms involved remain unknown (Ellis and Mellor (2000) *Trends Cell Biol* 10: 85-88; Chavrier and Goud (1999) *Curr Opin Cell*

Biol 11: 466-475; Hall (1998) *Science* 279: 509-514; Ridley (1996) *Curr Biol* 6: 1256-1264). Interaction between the WW domain of Nedd4 and a chimaerin homolog may shed light on the mechanism of recruitment of Rho family GTPase machinery to the protein ligase complexes controlling ubiquitin-mediated endocytosis.

SH3 domains.

[0127] The Src homology 3 (SH3) domain has become a prototype of protein interaction modules since it was first described as a conserved repeat in the N-terminus of Src family tyrosine kinases (Koch *et al.* (1991) *Science* 252: 668-674). Small, about 50-70 amino acids long, with a compact fold, SH3 domains recognize and bind peptide sequences with the core PxxP motif. The specificity of interaction within the SH3 family is determined by additional contacts formed between amino acids adjacent to the PxxP core of peptide ligand and variable amino acids within SH3 domain specificity pocket (Rickles *et al.* (1995) *Proc. Natl. Acad. Sci., USA,* 92: 10909-10913; Feng *et al.* (1995) *Proc. Natl. Acad. Sci., USA,* 92, 12408-12415). Peptide ligands can bind SH3 domains in two pseudosymmetrical (with respect to the PxxP core motif) orientations – the Class I orientation, ZxxPxxP, and the Class II orientation, PxxPxZ, where Z denotes the ligand residue(s) responsible for discrimination between individual SH3 domains (Feng *et al.* (1994) *Science* 266: 1241-1247).

[0128] The function of SH3 domains within the Src and Abl tyrosine kinases is believed to be two-fold. On one hand, through intramolecular interaction, SH3 domains of Src and Abl participate in the autoinhibitory control of the respective kinases (Sicheri and Kuriyan (1997) *Curr Opin Struct Biol* 7: 777-785; Barila and Superti-Furga (1988) *Nat Genet* 18: 280-282). On the other, they serve as targeting modules by binding to a specific subset of proteins containing polyproline sequences (Koch *et al.* (1991) *Science* 252: 668-674; Pawson and Nash (2000) *Genes Dev* 14: 1027-1047). Therefore, identification of binding partners of SH3 domains of the tyrosine kinases either directly suggests physiological targets of their activity or may indicate the multiprotein complexes to which they are targeted.

[0129] Crk is an adaptor protein composed of an SH2 domain and one or two (depending on the isoform) SH3 domains (Feller *et al.* (1998) *J Cell Physiol* 177: 535-552). By interacting with specific sets of proteins via their interaction modules, adaptor proteins

function to provide a molecular connection between signal transduction pathways. Identification of interaction partners of an adaptor protein facilitates the unraveling of interconnections and possible cross-talk between different signaling cascades.

[0130] c-Src and c-Abl tyrosine kinases and the adaptor protein Crk are cellular counterparts of classical viral oncogenes, v-Src (Radke *et al.* (1980) *Cell* 21: 821-828), v-Abl (Rosenberg and Witte (1988) *Adv Virus Res* 35: 39-81) and v-Crk (Mayer *et al.* (1988) *Nature* 332: 272-275). The pathways affected by these oncogenes have been the subjects of extensive studies with a number proteins identified as interacting partners of the respective SH3 domains (Barfod *et al.* (1993) *J Biol Chem* 268: 26059-26062; Weng *et al.* (1994) *Mol Cell Biol* 14: 4509-4521; Kapeller *et al.* (1994) *J Biol Chem* 269: 1927-1933; Gout *et al.* (1993) *Cell* 75: 25-36; Weng *et al.* (1993) *J Biol Chem* 268: 14956-14963; Ren *et al.* (1993) *Science* 259: 1157-1161; Gertler *et al.* (1995) *Genes Dev* 9: 521-533.; Ren *et al.* (1994) *Genes Dev* 8: 783-795; Knudsen *et al.* (1994) *J Biol Chem* 269: 32781-32787; Hasegawa *et al.* (1996) *Mol Cell Biol* 16: 1770-1776). Ligand preferences as well as the molecular basis of recognition specificity of Src-SH3, Abl-SH3 and Crk-SH3 domains have been recurrently addressed by screening of combinatorial peptide libraries and structural studies (Cheadle *et al.* (1994) *J Biol Chem* 269: 24034-24039; Rickles *et al.* (1994) *Embo J* 13: 5598-5604; Sparks *et al.* (1996) *Proc. Natl. Acad. Sci., USA,* 93: 1540-1544; Sparks *et al.* (1994) *J Biol Chem* 269: 23853-23856; Rickles *et al.* (1995) *Proc. Natl. Acad. Sci., USA,* 92: 10909-10913; Yu *et al.* (1994) *Cell* 76: 933-945; Feng *et al.* (1994) *Science* 266: 1241-1247; Musacchio *et al.* (1994) *Nat Struct Biol* 1: 546-551; Wu *et al.* (1995) *Structure* 3: 215-226).

[0131] We have identified by TAIS in non-exhaustive screens a number of previously described as well as novel putative interacting partners for Src, Abl and Crk SH3 domains (see Table 8).

[0132] Table 8. Summary of TAIS performed on a phage-displayed human brain cDNA library with the indicated targets.

Target	Hits	Accession (GenBank)	Individual Hit Frequency	Novelty	Statistics
rPSD95-PDZ(1+2+3)	DGKζ	U51477	1	Novel	From 11 clones analyzed: Hits 1 Frameshifts 3 Untranslated region 1 Undefined* 6
mNedd4-WW3	Chimaerin homolog VGSC type II α LAPTM5 Nogo-A	AL137579 AF327246 XM_001374 AF320999	1 2 (siblings) 1 3 (2 siblings + 1)	Novel Novel Novel Novel	From 7 clones analyzed: Hits 7
hSrc-SH3	WIP dynamin	NM_003387 XM_011757	2 siblings 3 (2 siblings + 1)	Novel **	From 12 clones analyzed: Hits 5 Frameshifts 5 Untranslated region 2
hAbl-SH3	SNRPC ZNF162 Aczonin/Piccolo MEA11/MGEA6	XM_004292 XM_006534 HSY19188 HSU73682	1 1 1 1	Novel Novel Novel Novel	From 27 clones analyzed: Hits 4 Frameshifts 12 Undefined 11
hCrk-SH3N	KIAA0716 DKFZp434KO31 DOCK1	XM_004923 AL137317 NM_001380	2 (siblings) 11 (3 independent sibling groups: 7+2+2) 1	Novel Novel ***	From 20 clones analyzed: Hits 14 Frameshifts 2 Undefined 4

*Undefined – see explanation in the text

** Gout *et al.* (1993) *Cell* 75: 25-36.

*** Hasegawa *et al.* (1996). *Mol. Cell Biol.* 16: 1770-1776.

[0133] In total, 77 clones that gave positive plaques on the membranes were analyzed by sequencing. 75 of them contained amino acid sequences that conformed to known recognition motifs of the respective target domains, thereby highlighting the performance of TAIS in the deliniation of target recognition preferences. The information about binding preferences such as recognition consensus can be used then for "in silico" identification of putative interactors of the respective target from protein databases (see example below).

[0134] In the screening experiments summarized above 40% of all positives clones displayed polypeptides that belong to known proteins demonstrating thus a high rate of true positives for direct in vitro identification of putative target interacting partners from cDNA libraries. Nucleotide sequences of 21 positive clones (27% of all analyzed) did not match any known protein coding sequences in NCBI database, though matches were found in the human EST database for all of them. Since a definite conclusion as to whether these sequences represent polypeptides from the human proteome or random peptides cannot be drawn at present, they have been designated as "undefined". Given the statistics we expect that a significant fraction of undefined sequences represent novel uncharacterized proteins.

[0135] All peptides, except two, which were selected by tested SH3 domains, contained sequences that conformed to the described recognition consensuses of the respective SH3 domains (*see, e.g.,* Table 9).

[0136] Table 9. Alignments of polypeptides selected from a phage-displayed human brain cDNA library by the indicated SH3 domains in comparison to previously reported recognition consensuses of corresponding SH3 domains. Underlined residues in previously reported consensuses for Src and Crk SH3 domains are position that have been fixed in biased peptide libraries used to define the respective consensuses. ψ denotes aliphatic residues. Note the additional specificity determinants uncovered by TAIS for the Crk SH3 domain at +4 and +5 positions (in respect to the PxxP core) of the selected peptide ligands.

CRK-SH3	ID	SEQ ID NO
VTSEP**P**A**LP**P**K**PLAARSSH	KIAA0716	104
SETIS**P**LR**P**Q**R**P**K**SQVMN	DOCK1	105
APTSP**P**IV**P**L**K**S**R**HLVAAA	DKFZp434K031	106
NLRGA**P**A**LP**G**R**SL**R**PPVDAP	ICR1	107
ELARS**P**S**LP**R**K**L**RR**LNEYYP	IICRK1	108
SSQ**P**R**LP**P**K**Q**R**GNARAH	IICRK4	109
MEK**P**C**LP**E**KK**K**K**KISQMW	IICRK6	110
HGET**P**S**LP**K**K**KY**K**N	IICRK7	111
PVIR**PP**L**P**P**K**VLGLQA	ICR8	112
Px**LP**x**K**x+	TAIS consensus	113
•P•LP•K*		114

Src-SH3

PRPIQS**S**LHN**R**GS**P**PV**P**G	WIP	115
RKRR**P**LPS**PR**LP**P**FP**P**SATREF	6TAK	116
PPSP**P**TLAR**RTLP**LS**P**AALKKNNN	2BS	117
GPPPQV**PS**RPN**R**AP**P**GV**P**SRSGQA	dynamin	118
S**S**PP**PR**S**LP**TP**P**PRSLPTPP	1BS	119
SG**P**RRA**PR**G**LP**PI**P**LRWGSERS	ITAK	120
PSxxP**R**xL**P**xx**P**	TAIS consensus	121
SLxxRPLPPLPP*	Other consensus	122
LxxRPLPx•P**		123
RxLPPLP***		124

Abl-SH3

MPMMPG**PP**MMR**PP**AR**P**MMVPTR	SNRPC	125
QHNPNGPP**PP**WMQ**PP**P**PP**MNQGPHPP	ZNF162	126
GASRDYF**P**PRDF**P**G**PP**PAPFAMR	MEA11	127
IQAGGSRG**P**VRA**PP**TR**P**CPGASGTG	1AS13	128
RQSCEPWAG**P**RVA**PP**R**PP**GHQGSEGE	2AA	129
WGRIYRGA**PP**TFAA**P**QA**P**KPFRQLLPM	2AS13	130
REGSCLQ**P**LPP**PP**P**PP**RLRPVR	3ABR	131
TKKPQRE**PP**ALP**PP**P**PP**LIKFL	3AS13	132
GGHRD**PP**KAR**PP**R**PP**SAPKP	4AB	133
EPLLP**PP**LPA**PP**A**PP**PVPA	5AS	134
HRSSTMNP**PP**HTQ**PP**SQ**P**QPRPPIYS	9AB	135
PPxxx**PP**x**PP**	TAIS Consensus	136
PPxΘxPPPΨP*		137
PPPYPPPPIP		138

Θ is aromatic residue.

* Sparks *et al.* (1996) *Proc. Natl. Acad. Sci., USA,* 93: 1540-1544

** Rickles *et al.* (1995) *Proc. Natl. Acad. Sci., USA,* 92: 10909-10913

*** Yu *et al.* (1994) *Cell* 76: 933-945

[0137] The case of the three SH3 domains is of a special interest, for it addresses a question of cross-reactivity between domains within the same family. Surprisingly enough, the analysis of 59 clones positive for interaction with the tested SH3 domains showed that SH3 domains from Crk, Src and Abl selected non-overlapping sets of polypeptides from the same library.

[0138] Previous studies of the Src SH3 domain family molecular recognition mechanism showed that specific amino acids of the peptide ligands that lie outside the SH3 core recognition motif play a critical role in ligand discrimination by related SH3 domains

and contribute significantly to the affinity of interaction (Rickles *et al.* (1995) *Proc. Natl. Acad. Sci., USA,* 92: 10909-10913; Feng *et al.* (1995) *Proc. Natl. Acad. Sci., USA,* 92, 12408-12415). Careful inspection of the amino acid sequences of peptides selected by TAIS with SH3 domains of Src, Abl and Crk revealed that the vast majority of selected peptides contained at least one continuous stretch featuring additional specificity determinants outside the SH3 core recognition sequence. Some of these determinants have been described previously, whereas others appear to be novel (see Table 9).

[0139] As in the case of the Nedd4-WW3 domain, it was possible to built up an extended recognition consensus for Crk SH3 domain without *a priori* knowledge about SH3 domain recognition preferences. This fact suggests that the Crk-SH3 domain has a strong preference for one of the two possible pseudosymmetrical orientations described for SH3 ligands, namely the Class II orientation, and exhibits a strong affinity for its cognate ligands. The majority of peptides selected by Crk-SH3 domain contained positively charged residues at position +4 and/or +5 following their PxxP cores. These residues may represent additional specificity determinant(s) not previously reported. The presence of multiple SH3 core motifs in both orientations within the same selected polypeptides prevented unambigous mapping of Src and Abl SH3 domain binding sites without knowledge of their recognition motifs.

[0140] Collectively, the results of screens performed with PDZ, WW and SH3 domains suggest that the TAIS format allows detection of interactions in a physiologically relevant range of affinities and is well suited for the characterization of ligand preferences of protein interaction modules.

Conclusions

[0141] A significant fraction of all specific protein-protein associations in the cell may be mediated by specialized peptide recognition domains such as PDZ, SH3, WW, EH, SH2, etc. Indeed,, 3300 proteins out of 6148 predicted ORFs in the yeast proteome have been reported to contain the SH3 domain recognition core PxxP (Zucconi *et al.* (2000) *FEBS Lett* 480: 49-54; Cherry *et al.* (1998) *Nucleic Acids Res* 26: 73-79). Similarly, SH3 and PDZ domains were ranked as 14th and 19th, respectively, among the most populous domain families in the human proteome (Lander *et al.* (2001) *Nature* 409: 860-921). On the qualitative side, protein interaction modules, in the context of proteins with enzymatic,

scaffolding or adaptor activity, are often constituents of a node of a protein interaction network, mediating multiple connections that diverge from or converge onto the node. Therefore, the identification of interacting partners of peptide interaction modules would contribute significantly to assembly of a comprehensive protein interaction map.

[0142] We have developed a new *in vitro* method, TAIS, that allows rapid screening of cDNA libraries for binding partners of peptide interaction modules. PDZ, WW and SH3 domains from PSD95, Nedd4, Abl, Crk and Src proteins were tested as targets. Summaries and statistics of test screens are compiled in Table 1. Two known and 12 novel potential interacting partners of these well studied domains were identified from a human brain cDNA library. All novel putative interacting partners contained recognition sequences of the respective target domains. Moreover, the absence of cross-reactivity between domains from the same family (SH3) and the presence of conserved ligand residues outside the family cores in all tested cases indicate high selectivity of the novel screening format. Most of the interactions make good sense in terms of biological relevance in the context of the known functions of PSD95, Nedd4, Src and Abl proteins, and allow generation of testable hypotheses about the functionality of detected interactions.

[0143] Deciphering rules that dictate binding specificity of protein interaction modules, or "protein recognition code," (Cherry *et al.* (1998) *Nucleic Acids Res* 26: 73-79; Sudol (1998) *Oncogene* 17: 1469-1474) would greatly facilitate mapping of protein-protein interactions on a genomic scale by bioinformatic tools. In this regard, TAIS of cDNA libraries is a powerful complement to traditional random peptide library analysis. Indeed, we have confirmed known recognition consensuses for all protein interaction modules tested, defined a recognition consensus for the tandem of the first two PDZ domains of PSD95, and identified additional putative specificity determinants for the Crk-SH3 domain.

Experimental protocol

GST fusions.

[0144] GST fusion constructs of , PDZ domains from rat PSD95 protein, human Src, Abl and Crk SH3 domains were kindly provided by Brian Kay, University of Wisconsin-Madison. The third WW domain from mouse Nedd4 was amplified by PCR from Nedd4 cDNA supplied by Sharad Kumar, Hanson Center for Cancer Research,

Adelaide, and cloned into the pGEX-2TK expression vector. All constructs were verified by sequencing.

Target protein preparation.

[0145] Immobilized GST fusions of target proteins were purified according to the supplier's instructions (Pharmacia Biotech.). To prepare biotinylated target complexes with streptavidin-alkaline phosphatase (STRAP) conjugate in solution, target domains were released from Glutathione Sepharose 4B beads by thrombin cleavage and mixed with freshly prepared water solution of EZ-link™ Sulfo-NHS-LC-LC-biotin (Pierce) at a molar ratio of 1:5. Biotinylation reaction was incubated for 30 minutes at room temperature followed by purification on MicroSpin G-25 column (Pharmacia Biotech.). The extent of biotinylation was kept at 1 to 2 moieties of biotin per target molecule. For detection of positive plaques on membranes, 5 μg of biotinylated target per membrane were pre-mixed with STRAP conjugate at a molar ratio of 4:1 to ensure multivalent target presentation and incubated for 10 minutes at RT before use in Tris-buffered saline, pH7.4 + 0.1% Tween 20 (TBS-T).

TAIS protocol.

[0146] 30 μg of target GST fusion immobilized on sepharose beads was blocked in 1 ml of 0.5% bovine serum albumin (BSA) in TBS-T for 1 hour at RT on a tumbler. After 3x1 ml washes with TBS-T, beads were mixed with an aliquot of cDNA library (10^8 pfu) (Novagen) in 1 ml of 0.5% BSA in TBS-T and incubated at RT for 90 minutes on a tumbler. After 5x1 ml washes with TBS-T, the phages bound to the target were eluted by incubation of washed beads in 200 μl of 1% SDS for 15 minutes at RT. 2 equal parts of eluate were plated on two 150 mm agar plates with BLT5615 host (Novagen). Plates were incubated at 37°C to develop plaques, usually for 2 to 3 hours. Plates with developed plaques were pre-cooled for 45 minutes at 4°C and overlaid with 132 mm nitrocellulose membranes (Schleicher&Schuell) for 10 minutes. While on plates, membranes were punctured on periphery asymmetrically with red hot needles to introduce a coordinate system. After plaque lift membranes were blocked with 1% BSA in TBS for 1 hour at RT and left overnight at 4°C on a rocker with 25 ml of 0.5% BSA in TBS-T containing 5 μg of biotinylated target complexed to STRAP. After extensive washing with TBS-T, positive

plaques on membranes were developed with insoluble AP substrate, BCIP/NBT (Sigma). Individual positive plaques were identified on plates and picked up for sequencing. If density of plaques was too high to pick up individual phage, agar stubs containing positive plaques were excised and phages from stubs eluted in PBS. Eluted phages were plated for a secondary screening on membranes. T7 phage DNA was prepared for sequencing with lambda DNA Wizard kit from Promega.

[0147] It is understood that the examples and embodiments described herein are for illustrative purposes only and that various modifications or changes in light thereof will be suggested to persons skilled in the art and are to be included within the spirit and purview of this application and scope of the appended claims. All publications, patents, and patent applications cited herein are hereby incorporated by reference in their entirety for all purposes.

CLAIMS

What is claimed is:

1. A method of identifying interacting proteins from a plurality of potentially interacting proteins, said method comprising:

i) contacting one or more target proteins with a protein display library comprising a plurality of potential binding proteins for said one or more target proteins;

ii) selecting members of said protein display library that bind to said one or more target proteins to provide a preselected set of potential binding proteins;

iii) separating said members of said preselected set of potential binding proteins from the bound target protein and immobilizing said members on a solid support such that said members are spatially addressable; and

iv) contacting members of said preselected set of potential binding proteins with one or more target proteins; and

v) detecting specific binding of members of said preselected set of potential binding proteins with said one or more target proteins whereby binding of a member of said set of potential binding partners with a target protein indicates that said member and said target protein are interacting proteins.

2. The method of claim 1, wherein said one or more target proteins are attached to a solid support.

3. The method of claim 1, wherein said protein display library is a phage- or bacterial-display library.

4. The method of claim 3, wherein said phage- or bacterial-display library is a phage display library.

5. The method of claim 4, wherein said phage display library is a lytic phage library.

6. The method of claim 1, wherein said separating comprises amplifying members of said protein display library that bind to said one or more target proteins.

7. The method of claim 1, wherein said separating and/or immobilizing comprises amplifying members of said protein display library that bind to said one or more target proteins.

8. The method of claim 7, wherein said amplifying comprises amplification of said members when they are spatially separated and addressable.

9. The method of claim 3, wherein said phage- or bacterial-displayed library comprises a cDNA library.

10. The method of claim 1, wherein said protein display library comprises at least 100 different members.

11. The method of claim 10, wherein said protein display library comprises at least 1000 different members.

12. The method of claim 2, wherein said selecting comprises removing unbound members of said protein display library from said solid support.

13. The method of claim 1, wherein said selecting comprises capturing said one or more target proteins using an affinity matrix.

14. The method of claim 1, wherein contacting members of said preselected set of potential binding partners with one or more target proteins comprises adsorbing members of said preselected set of potential binding partners to a solid support.

15. The method of claim 14, wherein said solid support is a membrane.

16. The method of claim 1, wherein said detecting comprises detecting a label attached to said target protein.

17. The method of claim 16, wherein said label is selected from the group consisting of a fluorescent label, a radioactive label, an enzymatic label, a colorimetric label, and a magnetic label.

18. The method of claim 1, wherein:

said contacting of step (i) comprises contacting said one or more target proteins with a protein display library where said one or more target proteins are attached to a solid support;

said contacting of step (iv) comprises attaching members of said preselected set of potential binding proteins to a solid support to provide a set of attached preselected potential binding proteins and contacting the attached preselected potential binding proteins with the one or more target proteins.

19. The method of claim 18, where the one or more target proteins used in the contacting of step (iv) are labeled with a detectable label before the target proteins are contacted to the preselected potential binding proteins.

20. The method of claim 18, where the one or more target proteins used in the contacting of step (iv) are labeled with a detectable label simultaneous with or after the target proteins are contacted to the preselected potential binding proteins.

21. The method of claim 18, further comprising sequencing the nucleic acid encoding the displayed protein on a member of the preselected display library that binds to the target protein.

22. The method of claim 1, wherein:

said contacting of step (i) comprises contacting said one or more target proteins with a protein display library where said one or more target proteins and said protein display library are in solution.

23. The method of claim 22, wherein said selecting comprises capturing target proteins bound to members of said protein display library using an affinity matrix that specifically binds the target proteins or a tag attached to the target proteins.

24. The method of claim 23, wherein said contacting of step (iv) comprises attaching members of said preselected set of potential binding proteins to a solid support to provide a set of attached preselected potential binding proteins and contacting the attached preselected potential binding proteins with the one or more target proteins.

25. The method of claim 24, where the one or more target proteins used in the contacting of step (iv) are labeled with a detectable label before the target proteins are contacted to the preselected potential binding proteins.

26. The method of claim 24, where the one or more target proteins used in the contacting of step (iv) are labeled with a detectable label simultaneous with or after the target proteins are contacted to the preselected potential binding proteins.

27. The method of claim 1, wherein, said detecting comprises determining the amino acid sequence of a member of said set of potential binding partners that binds a target protein.

28. The method of claim 1, further comprising recording the amino acid sequence or identity of a member of said set of potential binding partners that binds a target protein in a database of proteins that interact with the target.

29. A method of identifying proteins or nucleic acids that interact with target moieties from a nucleic acid or protein library comprising a plurality of nucleic acids or proteins, said method comprising:

i) contacting one or more target moieties with said library;

ii) selecting members of said library that bind to said one or more target moieties to provide a preselected set of potential binding partners;

iii) separating said members of said preselected set of potential binding partners from the bound target and immobilizing said members on a solid support such that said members are spatially addressable;

iv) contacting members of said preselected set of potential binding partners with one or more target moieties; and

v) detecting binding of members of said set of potential binding partners with said one or more target moieties whereby binding of a member of said set of potential binding partners with a target binding moiety indicates that said member is a binding partner that interacts with the target moiety.

30. The method of claim 26, wherein said library is selected from the group consisting of a phage display library, a bacterial display library, a yeast display library, a eukaryotic virus library, a direct encoded plasmid library.

31. The method of claim 26, wherein said library is an in vitro display library selected from the group consisting of a covalent display technology (CDT) library, a polysome display library, and an RNA-peptide fusion library.

32. A method of identifying proteins that interact with target moieties from a plurality of potentially interacting proteins, said method comprising:

i) contacting one or more target moieties with a protein display library comprising a plurality of potential binding partners for said target moieties;

ii) selecting members of said protein display library that bind to said one or more target moieties to provide a preselected set of potential binding partners;

iii) separating said members of said preselected set of potential binding proteins from the bound target protein and immobilizing said members on a solid support such that said members are spatially addressable; and

iv) contacting members of said preselected set of potential binding partners with one or more target moieties; and

v) detecting binding of members of said set of potential binding partners with said one or more target moieties whereby binding of a member of said set of potential binding partners with a target binding moiety indicates that said member is a protein that interacts with the target moiety.

33. The method of claim 32, wherein said target moiety is selected from the group consisting of a nucleic acid, a lipid, a carbohydrate, a glycoprotein, a small organic molecule, and an inorganic molecule.

34. The method of claim 32, wherein said target moiety is a DNA or an RNA.

35. A kit for identifying interacting proteins from a plurality of potentially interacting proteins, said kit comprising:

a protein display library; and

instructional materials providing protocols for the method of claim 1.

36. The kit of claim 35, wherein said protein display library is a bacterial or phage display library.

37. The kit of claim 36, wherein said bacterial or phage display library comprises a cDNA library.

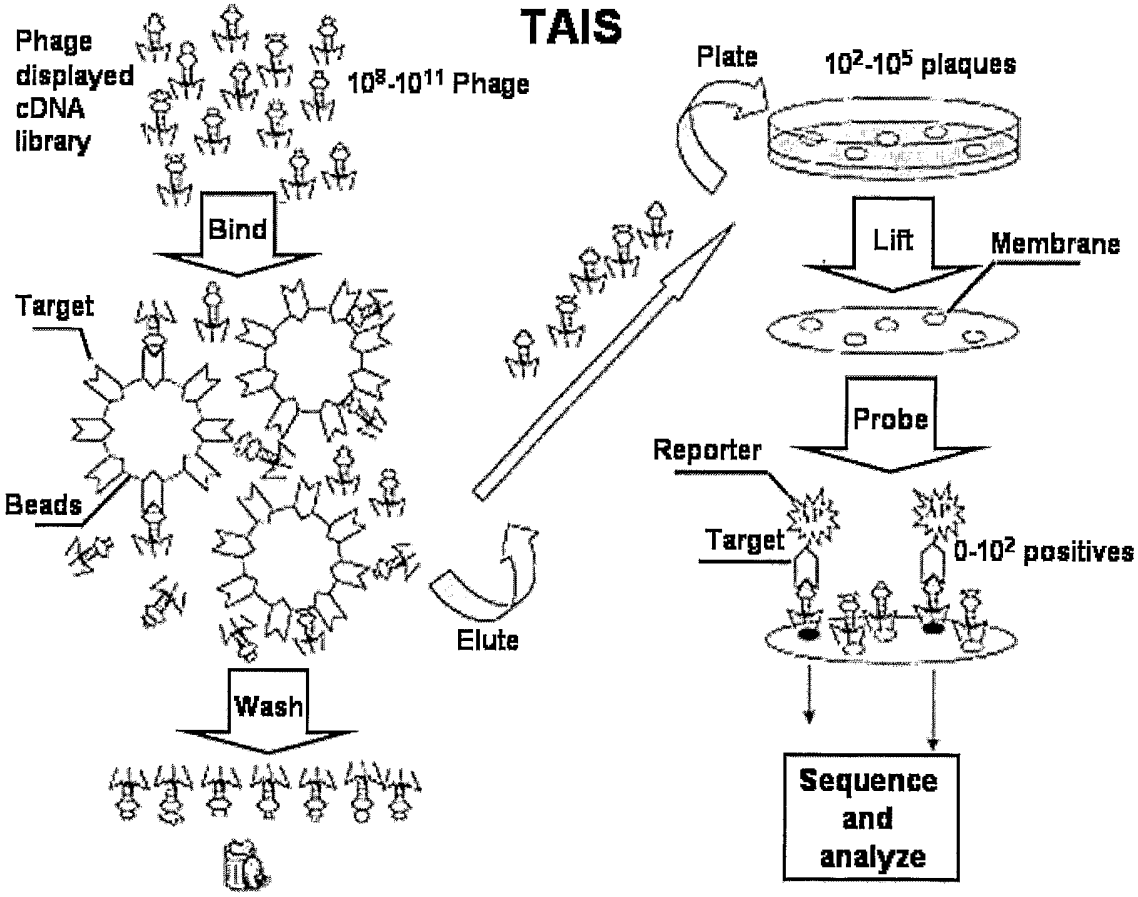
TAIS
Phage displayed cDNA library
10^8-10^{11} Phage
Bind
Target
Beads
Wash
Elute
Plate
10^2-10^5 plaques
Lift
Membrane
Probe
Reporter
Target
0-10^2 positives
Sequence and analyze

Fig. 1

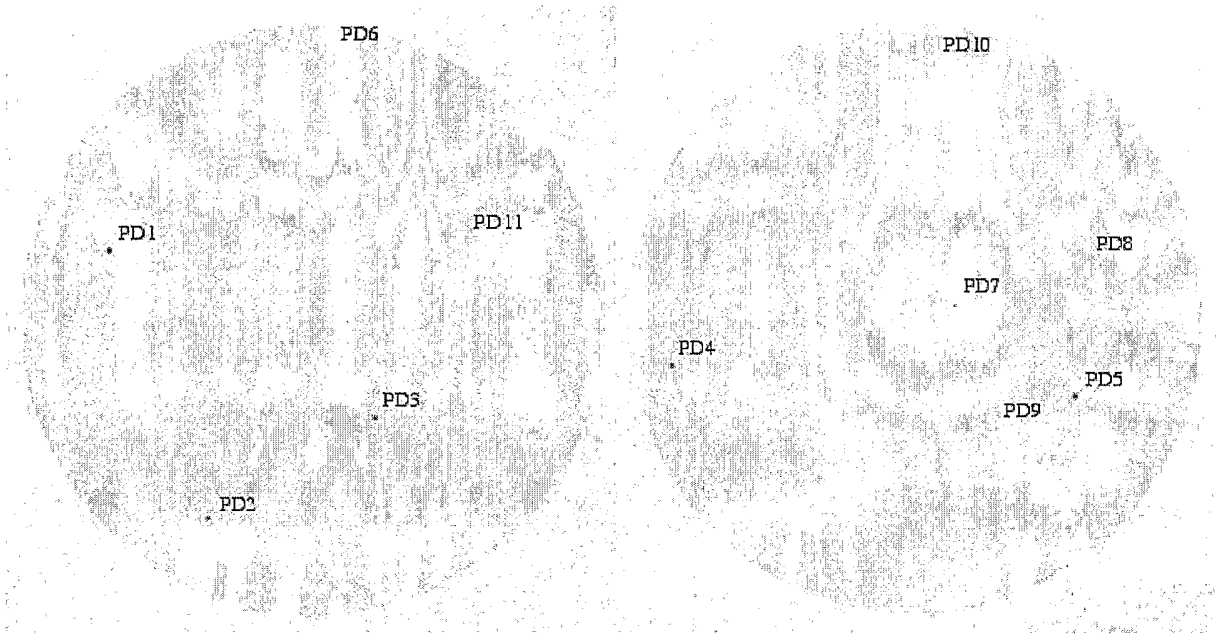

Fig. 2

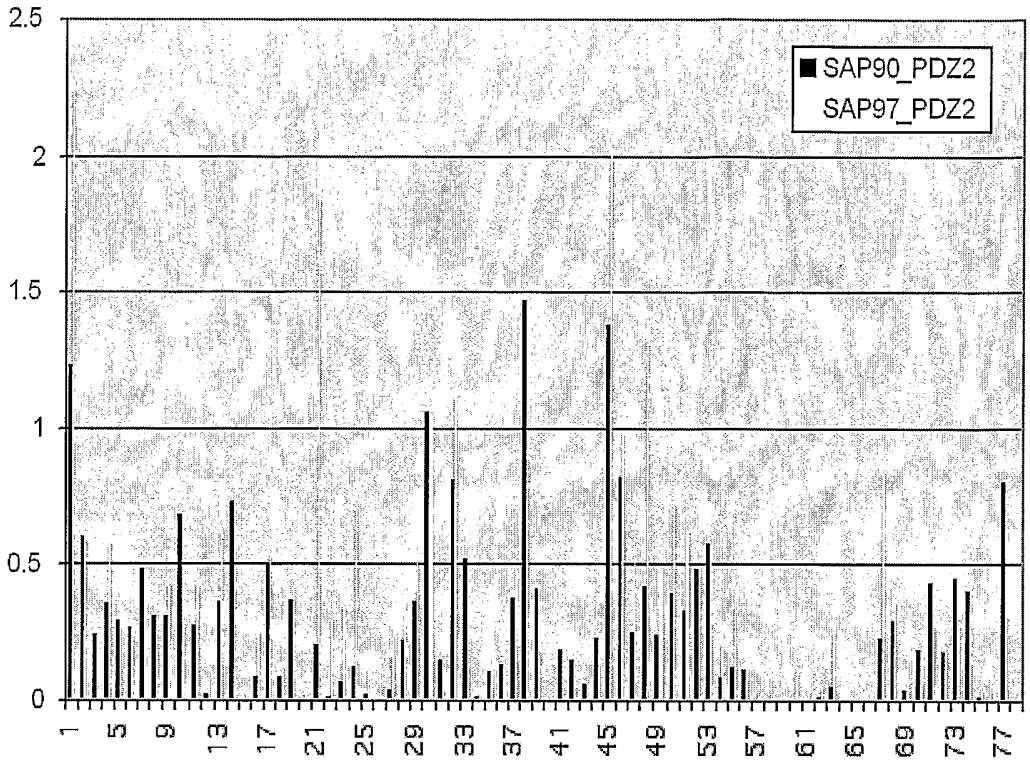
2.5
2
1.5
1
0.5
0
SAP90_PDZ2
SAP97_PDZ2
1
5
9
13
17
21
25
29
33
37
41
45
49
53
57
61
65
69
73
77

Fig. 3

Target-Assisted Iterative Screening™

Target: Crk-SH3N Genbank Acc#: undefined Posted date: June 2002

Library: normal human brain cDNA, T7select10-3 (Novagen, Cat. #70637-3)

Reference: Kurakin A, Bredesen D: Target Assisted Iterative Screening reveals novel interactors for PSD95, Nedd4, Src, Abl and Crk proteins. J Biomol Struct Dyn. 2002, 19:1-15.

Author: A.Kurakin

Target sequence:

RQGSGVILRQREAKYVRALFDFNGNDEEDLPFKKGDILRIRDKPEEQWWNAEDSRGKRGM
IPVPYVEKYRPASASVSALIGGH>

Hits

1. DOCK1

   Genbank Acc#: NM_001380

   Sequence:

   SS DEKKDKKKEKRNSKHQEIFEKEFKFTDISLQQSEAVILSETISPLRPQRPKSQVHNV
   IGSERRFSVSPSSPSSQQTPPPVTPRAKLSFSHQSSLELNGKTGADVADVPPPLPLKGSV
   ADYGNLNENQDLLGSPIPPPPPPHQRHLPPPLPSKTPPPPPPKT SLRPHSSN>

2. DKFZp434K031

   Genbank Acc#: AL137317

   Sequence:

   SS PSTPQPASPDVSSSHSLSEYSHGQSPRSPISPELHSAPLTPVARDTPRHLASEDTRC
   STPELGLDEQSVQPWERRTPPLAHSPQAECEDQLDAQERAARCTNRTSGSKTGRETEAAP
   TSPPIVPLKSRHLVAAATAQRFTHR>

3. KIAA0716

   Genbank Acc#: XM_004923

   Sequence:

   SS PARHTTSVSPSPAGRSPLKGSVQSFTPSFVEYHSPGLISNSPVLSGSYSSGISSLSR
   CSTSETSGFENQVNEQSAPLPVPVPVPVPSYGSEEPVRKESKTPPPYSVYERTLRRPVPL
   PHSLSIPVTSEPPALPPKPLAARSSHLEKGARRTDPGPRPRPLPRKVSQL>

<< Back to the domain family list

Fig. 4

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US02/31349

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : G01N 33/53; C12Q 1/70; C12P 21/04; C07K 16/00
US CL : 435/7.1, 7.31, 255.1, 320.1; 530/350, 413

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 435/7.1, 7.31, 255.1, 320.1; 530/350, 413

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y,P	US 6,472,147 B1 (JANDA et al.) 29 October 2002 (29.10.2002), entire document.	1-37
Y,P	US 6,410,246 B1 (ZHU et al.) 25 June 2002 (25.06.2002), entire document.	1-37
Y	US 6,031,071 A (MANDEVILLE et al.) 29 February 2000 (29.02.2000), entire document.	1-37
Y	US 6,156,511 A (SCHATZ et al.) 05 December 2000 (05.12.2000), entire document.	1-37
Y	US 5,733,731 A (SCHATZ et al.) 31 March 1998 (31.03.1998), entire document.	1-37
Y	US 5,702,892 A (MULLIGAN-KEHOE) 30 December 1997 (30.12.1997), entire document.	1-37

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search
16 December 2002 (16.12.2002)

Date of mailing of the international search report
13 JAN 2003

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231
Facsimile No. (703)305-3230

Authorized officer
My-Chau T. Tran

Telephone No. 703-308-1123

Form PCT/ISA/210 (second sheet) (July 1998)