AD_____

Award Number: DAMD17-97-1-7231

TITLE: The Identification of Novel Ligands for Cell Surface Receptors

PRINCIPAL INVESTIGATOR: Pamela A. Silver, Ph.D.

CONTRACTING ORGANIZATION: Dana Farber Cancer Institute
Boston, Massachusetts 02115

REPORT DATE: October 2000

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	October 2000	Final (1 Oct 97 - 30 Sep 00)

4. TITLE AND SUBTITLE
The Identification of Novel Ligands for Cell Surface Receptors

5. FUNDING NUMBERS
DAMD17-97-1-7231

6. AUTHOR(S)
Pamela A. Silver, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Dana Farber Cancer Institute
Boston, Massachusetts 02115

E-MAIL:
pamela_silver@dfci.harvard.edu

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
Report contains color graphics.

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; Distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT ***(Maximum 200 Words)***

A general problem in metazoan biology is the identification of the specific ligands for transmembrane receptors. One goal of the original proposal was to selectively modify the physiology of the budding yeast *Saccharomyces cerevisiae* to speed the identification and study of ligand/receptor interactions - in particular that of erB2 given its involvement in breast cancer. Significant progress was made on development of a system that allows a rapid readout for interaction of a receptor with its corresponding ligand. A second approach involves a novel search for small interacting peptide aptamers that could be used as ligand mimics or inhibitors. Peptide aptamer libraries have been characterized and in doing so, a novel peptide effector of the cell cycle regulator Cdk4 was characterized.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Breast Cancer	72
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

Table of Contents

FINAL REPORT (con.)

(4) INTRODUCTION

A general problem in metazoan biology is the identification of the specific ligands for transmembrane receptors. One goal of the original proposal was to selectively modify the physiology of the budding yeast *Saccharomyces cerevisiae* to speed the identification and study of ligand/receptor interactions - in particular that of erB2 given its involvement in breast cancer. In brief, we exploited the biological process of protein folding in the ER in a completely innovative way so as to achieve these goals. Accumulation of unfolded proteins in the ER induces transcription of factors that help refold proteins (1) . This effect is termed the Unfolded Protein Response (UPR). The Ire1 protein of yeast plays a major role in transducing the signal from the ER lumen to the nucleus (2). Ire1 is embedded in the ER membrane and has a cytoplasmic protein kinase domain. In many respects, Ire1 behaves like a cell surface receptor; Ire1 signaling in yeast is very similar to the ligand-receptor interaction seen in higher organisms. Thus, we took advantage of this to design a system to search for novel ligands of mammalian receptors, in particular for erbB2. A second approach involves a novel search for small interacting peptide aptamers that could be used as ligand mimics or inhibitors (3). Peptide aptamer libraries have been characterized and in doing so, a novel peptide effector of Cdk4 was characterized. Progress was made on all fronts and is summarized below.

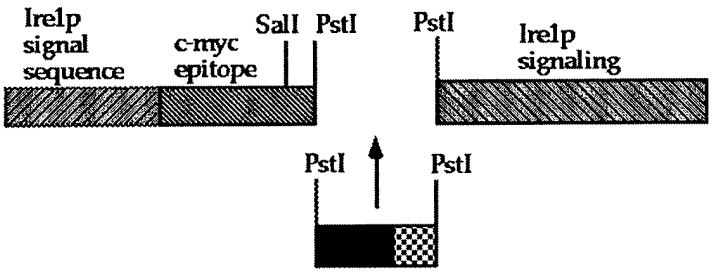
(5) BODY

Objective 1: To develop a system to identify ligand-receptor interactions in yeast

I. A plasmid encoding the Ire1 signal sequence fused to the extracellular domain of EGFR followed by the transmembrane, signaling, and 3' UTR domains of Ire1 has been constructed as illustrated below.

This schematic represents the EGFR/Ire1p chimeric receptor that is produced from a yeast centromeric plasmid. Expression of this chimera in one construct is driven by the endogenous IRE1 promoter and in another construct by the inducible GAL1 promoter. Both constructs contain the IRE1 3' untranslated region. A feature of this construct is that the EGFR extracellular domain can be replaced
by the extracellular domain of any other receptor using the SalI and EcoRI cloning sites.

II. A more general plasmid has been constructed into which any receptor extracellular and transmembrane domains can conveniently be fused to the IRE1 signaling domain as illustrated below.



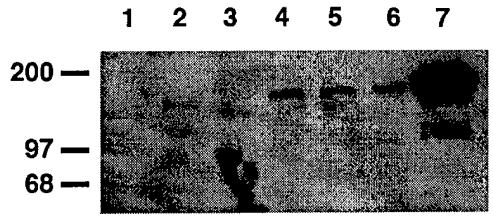
An additional feature of this general vector is the presence of a region encoding the c-myc epitope so that production of the chimeric proteins can be confirmed by immunoblotting with anti-c-myc monoclonal antibodies as demonstrated in the next section.

III. Ligand expressing plasmids were constructed as follows. A plasmid consisting of the yeast ADH promoter followed by the yeast pre-pro alpha factor signal sequence fused to the sequence of mature EGF and the ADH terminator sequences has been constructed. A similar plasmid which has the endoplasmic reticulum retention signal HDEL at the carboxy terminus of EGF has also been constructed. The highlight of this plasmid is that EGF portion can be removed by cutting with the restriction enzymes HincII and XhoI and subsequently replaced by another ligand of choice. In addition, we have constructed a plasmid encoding EGF where the codon usage has been optimized for expression in yeast. Others have found this to aid in the expression of human EGF in yeast.

IV. Yeast strains were constructed to assay the interactions between ligand and receptor via theunfolded protein response. To accomplish this, we constructed yeast strains deleted for the *IRE1* gene so that there will be no interference from the endogenous Ire1 protein in our experiments. In addition, we have constructed a reporter plasmid containing the UPR response element controlling the expression of the *lacZ* gene. When wild-type cells (*IRE1*$^+$) are grown under conditions that induce the concentration of unfolded proteins in the ER lumen, e.g. in the presence of tunicamycin, the UPR-LacZ reporter is induced as evidenced by the production of active β-galactosidase. However, in cells deleted for *IRE1*, there is no expression from the UPR-LacZ reporter as would be predicted. Thus, any signal we detect upon introduction of the receptor chimeras and their respective ligands should be attributable only to their interaction with each other.

V. In order to ascertain whether or not the EGF-Ire1 receptor chimera was indeed made when expressed in yeast, immunoblots were performed with anti-EGF receptor antibody as the probe. As shown below, the yeast cells are capable of producing the receptor-Ire1

chimera at a detectable level as evidenced by the presence of a protein of the correct predicted size that specifically reacts with the anti-EGF receptor antibody when cells are grown on galactose but not on glucose - the receptor chimera is under control of the inducible *GAL1* promoter.



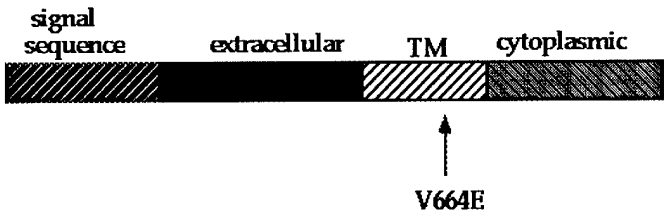
Whole cell extracts (from *ire1Δ* cells) expressing the indicated proteins were subjected to gel electrophoresis and proteins transferred to nitrocellulose and probed with anti-EGF receptor antibody. Lane 1: EGFR/Ire1p (glu); Lane 2: EGF (gal); Lane 3; vector alone (gal); Lane 4: EGFR-Ire1p + EGF (gal); Lane 5 EGFR-Ire1p + EGF/HDEL (gal); Lane 6: EGFR-Ire1p (gal); Lane 7: human A431 cell lysate.

VI. The chimeric receptor plasmid and each of the EGF encoding plasmids have been expressed alone or together in the unfolded protein response reporter yeast strains (see above). The level of response is determined by the amount of measurable β-galactosidase present in each strain. Data from a typical experiment is presented in the Table below and indicate that the system is working (see Conclusions).

STRAIN	Tunicamycin	Sugar	β-galactosidase units
Wildtype (IRE1^{+})	+	Galactose	237
Wildtype (IRE!$^{+)}$	-	Galactose	16
Δ*ire1*	+	Galactose	14
Δ*ire1*	-	Galactose	11
Δ*ire1* + EGFR-Ire1p	-	Galactose	226
Δ*ire1* + EGFR-Ire1p	-	Glucose	10
Δ*ire1* + vector	-	Galactose	6
Δ*ire1* + EGFR-Ire1p +vector	-	Galactose	35
Δ*ire1* + EGFR-Ire1p + EGF	-	Galactose	60
Δ*ire1* +EGFR-Ire1p +EGF-HDEL	-	Galactose	65

Objective 2: Identification of ligands for orphan receptors

ErbB2 chimeric receptor plasmids for expression of erbB2-Ire1 in yeast as illustrated below were generated - the major difference from those described above being that we have incorporated the c-myc epitope after the signal sequence so as to confirm production of full-length chimeric proteins..



In addition, we took advantage of a special mutant erbB2 that contains a single amino acid change in the transmembrane domain (V664E). This single mutation yields a ligand-independently activated erbB2 receptor. Thus, when incorporated into our reporter system, we would expect to get a ligand independent signal when this protein is expressed thus serving as a positive control. A second constitutively active receptor construct was constructed and will serve as another ligand-independent proof-of-concept controls. This construct is made up of the extracellular and transmembrane domains of constitutively active human Fibroblast Growth Factor Receptor 2 mutant C342Y found in Crouzon syndrome fused to the signaling domain of Ire1. Unfortunately, when introduced into yeast strains, very poor expression of these fusion proteins was obtained making further progress on this objective difficult in the time frame of the funding period.

Objective 3: Identification of novel interacting and/or inhibitory peptides

We proposed to apply a novel approach to identify small interacting peptides that may affect receptor function. This approach takes advantage of the bipartite nature of transcription factors. In such a screen, the yeast serve as a vessel in which the interaction occurs and generate the signal that alerts one to the potential interaction. In its most commonly used form, a protein from any organism can be expressed in yeast (the so-called 'bait') in conjunction with a cDNA library from any organism (the 'fish'). The novel methodology to be used here is designed to allow the rapid examination of the interaction of proteins of interest with a large number of random peptides expressed as 'aptamers' (from *aptos* - "to fit"). The aptamers are synthesized from a library of at least 10^8 plasmids that direct the synthesis of randomly encoded 20-mer peptides within *E. coli* thioredoxin, such that the peptides are displayed as loops that protrude from the surface at the thioredoxin active site; the chimeric peptide-proteins have no thioredoxin activity. The gene encoding each aptamer is fused to an activation domain and a nuclear targeting sequence, and the screen for aptamer binding to the protein of interest is carried out in a manner similar to the standard two-hybrid approach (6,14). Thus, the thioredoxin-aptamer is the 'prey', and the 'bait' will be the extracellular domain of erbB-2. We do not expect that the aptamers will induce receptor dimerization, because they will have been isolated using a monomeric target. Instead, in these experiments we will seek peptides that simply bind to the surface of the ligand-binding domain of erbB-2. Some of these may compete for ligand binding. And it

is possible to genetically engineer a dimeric thioredoxin-aptamer, which could cause receptor activation.

Once an initial set of aptamers has been identified, an optimized set can be constructed from an initial sequence by mutating each amino acid in the 20-mer peptide to all other possible amino acids. These experiments will allow us to determine the optimal peptide sequence that will, for example, fit into the ligand binding site of erbB-2. Aptamers identified in an initial screen usually have a binding constant of about 10^{-8}M. Knowledge of the peptide sequence will provide information about the ligand binding site that may be used in rational drug design. Finally, an understanding of the basis for the effect of the peptide on protein function will allow efficient design of anti-cancer strategies.

Much of our work on this aim concerns the characterization of the aptamer libraries and a shift in the original proposal which was accepted in last year's report. To validate the approach, we used as 'bait' the well-characterized Cdk4. Cdk4 is a cyclin-dependent protein kinase that functions immediately upstream of Rb and, as such, is a popular target for anti-tumor drug design. Inhibition of Cdk4 activity could impact on cancer cell growth and/or treatment. Therefore, we sought aptamers that specifically bind to Cdk4.

RESULTS

An exhaustive screen of the available aptamer library (Colas et al, 1996) identified a single aptamer, named 10T3, that possessed the desired characteristics: interaction with CDK4 but not with closely related kinases such as CDK2 or other protein kinases, such as the MAP kinase homolog Hog1 (Figure 1). Like the physiological

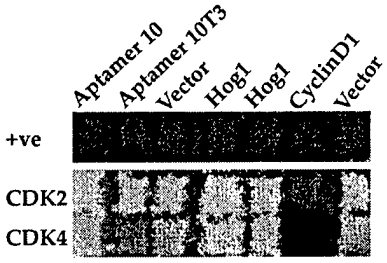


Figure 1. *Comparison of two hybrid interactions, assessed by the activation of β-galactosidase expression in yeast cells. Aptamer 10T3 interacts weakly with CDK4 (first blue patch, bottom row) but not with CDK2 (corresponding white patch on the top row). A control aptamer (Aptamer 10) interacts neither with CDK2 nor with CDK4. Neither aptamer 10 nor 10T3 interact with other protein kinases, such as Hog1.*

CDK4 interacting protein cyclin D1, 10T3 interacts with CDK6 as well as CDK4, although in both cases these interactions are much weaker with CDK6 than with CDK4 (not shown). In contrast to cyclin D1, 10T3 appears not to interact with CDK2, even when overexpressed in yeast cells, suggesting that it is a highly specific CDK4 interactor.

Sequence analysis of aptamer 10T3 gave a surprising result. The library from which 10T3 was isolated encodes for as many as 10^{26} different 20 residue peptides, each with a unique sequence. This is a much larger set than the number of unique 20-mer peptides contained within the human proteome (10^8 possible unique peptide 20-mers). Therefore, it is thought to be very unlikely that the sequence of an aptamer will resemble the sequence of a human protein. Indeed, BLAST searches of the available sequences within GENBANK yield no homologous matches. However, closer inspection of the sequences

of proteins known to interact with CDK4 revealed that the peptide sequence of 10T3 closely resembles that of cyclinD1. Furthermore, the region of cyclinD1 that resembles 10T3 corresponds to the region of the cyclinD1 homolog cyclin A that has been shown to directly contact CDK2, the corresponding CDK partner of cyclin A (Figure 2).

```
Cyclin D1:      RPEELLQMELLLVNKLKWNLA
Apt.10T3: GPQGLVLGELLTSLGMRWQNPQ
Cyclin A: TKKQVLRMEHLVLKVLTFDLA
```

Figure 2. *An aptamer (10T3) that binds to human CDK4 shows homology to a known CDK4 partner namely Cyclin D. The crystal structure of Cyclin A/CDK2 suggests that aptamer 10T3 may compete with cyclin D for binding to CDK4. Residues in bold are conserved between 10T3 and cyclinD1. The cyclin A sequence shown is that which is conserved with cyclinD1. The EHL motif makes critical contacts with CDK2 in the published critical structure (Jeffrey PD, Russo AA, Polyak K, Gibbs E, Hurwitz J, Massague J, Pavletich NP: Mechanism of CDK activation revealed by the structure of a cyclinA-CDK2 complex. Nature. 1995 Jul 27;376(6538):313-20).*

On the basis of these results, we undertook further characterization of 10T3. We wished to ask whether 10T3 could be used as a biochemical probe for CDK4 and whether expression of 10T3 within cells would give rise to predictable cell cycle phenotypes.

We fused aptamer 10T3 to glutathione S-transferase (GST) and expressed the fusion protein in *Escherichia coli* cells. The recombinant protein was expressd at high levels and was readily recovered on glutathione agarose beads. Unfortunately, we were unable to recover CDK4 from yeast cell extracts, even when GST-fusions of an aptamer that recognizes the human huntingtin protein can be shown to isolate its target from yeast extract (Figure 3).

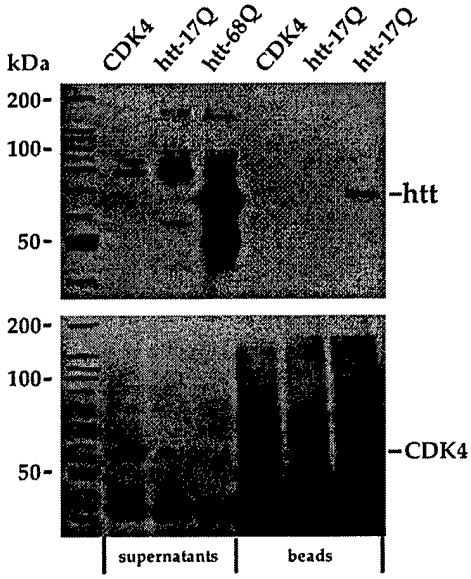


Figure 3. *A fusion of GST to aptamer 10T3 fails to affinity purify its target, CDK4, from yeast cell extracts (bottom panel). CDK4 is apparent as a ~57 kDa protein in the supernatants (left hand side) but is not visibly associated with glutathione-agarose beads that are used to isolate the GST-10T3 fusion. In contrast, a fusion of GST to aptamer 10 is able to isolate at least a fraction of its target, exon 1 of the human huntingtin protein from yeast cell extracts (the 55 kDa bandat the right hand side in the top panel).*

When aptamer 10T3 was transiently over-expressed in a range of mammalian cells (human U2OS osteosarcoma cells, HeLa cells and the myoblast precursor cell line C2C12), no robust phenotypes were observed. Expected phenotypes included an early (G1/S) cell cycle arrest and, in the case of the myoblast precursor cell line, cell cycle arrest followed by differentiation into myotubes. A reproducible but weak phenotype was the observation of diverse cellular phenotypes in cells transfected with constructs expressing 10T3, but not with the empty thioredoxin scaffold. These included elongated spindle shaped cells (not shown) and cells that had swollen to 10-20 times the size of untransfected cells (Figure 4).

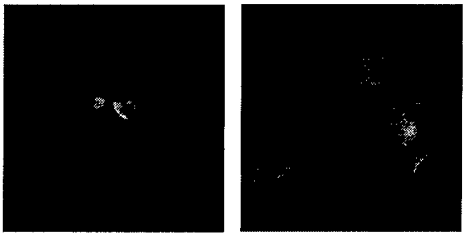


Figure 4. *U2OS osteosarcome cells were transfected with empty vector (left hand panel) or with the same vector carrying the gene for aptamer 10T3 (right hand panel). Cells that took up either vector express GFP and are shown in green. The total number of cells visible in the field is revealed by their DAPI- stained nuclei, in blue. 10T3 transfected cells often adopt a swollen morphology (right hand panel).*

Although this morphology superficially resembles that of differentiated osteoclasts , the U2OS osteosarcoma has not previously been shown to differentiate and the significance of this observation is therefore unclear. Given how poorly reproducible it was, we have decided to further optimize aptamer 10T3 rather than attempt to characterize what may be partial effect. Our strategy has been to use 10T3 as the starting point in a search for aptamers that will interact with CDK4. We have performed random PCR-mutagenesis of the 60 nucleotide sequence that encodes 10T3 under conditions were approximately 1 mutation will occur every 100 nucleotides, in an attempt to ensure that each PCR product will contain one or less base change. We placed the resulting oligonucleotides into the thioredoxin scaffold construct and screened 4000 of these new, 10T3-derived aptamers for their strength of interaction with CDK4 compared to the original 10T3 (Figure 5).

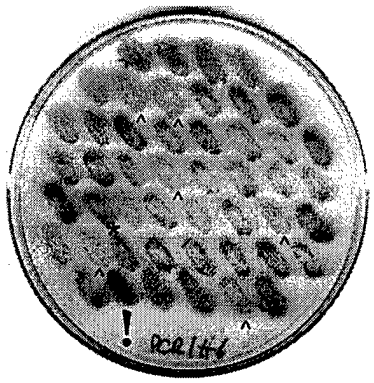


Figure 5. *The plate shows ~100 of the 4000 mutagenized 10T3 constructs. The degree of interaction of each aptamer with CDK4 is measured by the intensity of the color blue- aptamer 10T3 itself is marked by an asterisk (*). Aptamers that have lost interaction with CDK appear white and are marked by an arrowhead (^). An example of an aptamer that appears to interact more tightly than 10T3 with CDK is marked by a !.*

Most mutations are neutral, as would be expected given that less than one-third of base substitutions will lead to an amino acid change. Of those aptamers whose affinity for CDK is altered, most show a decrease, suggesting that many residues in the peptide portion of the aptamer are likely to be important for the interaction. Aptamers that interact more strongly with CDK4, and a subset of those that have lost interaction, are being sequenced.

(6) KEY RESEARCH ACCOMPLISHMENTS

- A system was developed whereby ligand-receptor interactions could be detected using a novel in vivo readout in yeast.

- Peptide aptamer libraries were characterized and a novel Cdk4-interacting peptide was identified and further characterized.

(7) REPORTABLE OUTCOMES

-Manuscripts, abstracts and presentations

Colas, P., Cohen, B., Ferrigno, P., Silver, P. and Brent, R. Targeted modification and transportation of cellular proteins. 2000, Proc. Natl. Acad. Sci., In press.

Morehouse, H., Ferrigno, P., Way, JC and Silver, PA. Using Yeast to model disease processes in cancer. 2000. Era of Hope Proceedings, DOD Breast Cancer Research Meeting, Atlanta, GA.

-Patents filed

"Genetic selection method for identifying ligands for transmembrane proteins"

-Degrees obtained

Heather Morehouse who did the work concerning the ligand-receptor system was a graduate student in the BBS graduate program at Harvard Medical School and will receive her PhD 12/00.

-Employment applied for and training based on this award

Dr. Paul Ferrigno who was supported, in part, by this award and who developed the aptamer technology has applied for both academic and industrial research positions. Thus far, he is deciding between offers from Astra and the MRC, Cambridge, UK.

Tweeney Kau participated in aspects of this research during her first year in the Harvard Medical School BBS graduate program. This resulted in co-authorship on an Abstract

and her decision to join my laboratory to complete her PhD and to extend aspects of this work.

(8) CONCLUSIONS

I. With regard to the first part of the proposal concerning development of a novel approach for screening for novel receptor ligands, the following conclusions were reached.

1) We showed that we could successfully measure the unfolded protein response in wild-type yeast cells with our reporter system.

2) Cells deleted for *IRE1* show no response to tunicamycin as would be expected since Ire1p is essential for cells to sense the presence of misfolded proteins in the ER.

3) When cells are induced to express the EGFR-Ire1p chimera, there is some induction of the UPR reporter. This is independent of the addition of tunicamycin. One possibility is that this indicates that at high concentrations there is some auto-dimerization of the receptor in the ER via its ligand binding domain. Since dimerization is necessary for Ire1p signaling activity, this then elicits a positive response. This means that the chimeric receptor may be correctly oriented in the ER membrane.

4) Finally, there is some response to the presence of EGF ligand together with the EGFR-Ire1p receptor chimera. In order to understand these data, first we observe that under conditions where the receptor chimera is induced in triple dropout media (which is different from the conditions discussed above), the amount of presumed autodimerization is reduced. We think this to be because the level of induced EGFR-Ire1p is less under these conditions. However, when we co-express the EGF ligand there is a reproducible small amount of increase in the readout of the UPR response suggesting a ligand-dependent increase in the amount of receptor-Ire1p dimerization.

Taken together, these data provide proof of concept that this unique reporter system can be used to detect ligand receptor interactions. Our long term plan is to apply this to recpetors for which ligands are not known.

II. With regard to the second part of the proposal, we have identified a peptide aptamer that specifically binds to the medically important cdk4 cell cycle regulator. We have characterized this aptamer in some detail with regard to its specificity for cdk4. Interestingly, we found that the aptamer is similar to cyclin and therefore may be interacting with the same surface on cdk4. We have mutagenized this aptamer with the hope of obtaining a higher affinity binder. Preliminary results indicate that the presence of the aptamer can affect cell differentiation. Once the optimal, 10T3 derived aptamer has been identified and constructed, we will use it for the biochemical and in vivo studies outlined above. Should the aptamer studies suggest that CDK is a viable drug target, we will use it as a basis for identification of small molecules that affect cdk4 function in mammalian cells.

(9) REFERENCES

1. Kozutsumi, Y., Segal, M., Normington, K., Gething, MJ, and Sambrook, J. (1988). The presence of unfolded proteins in the endoplasmic reticulum signals the induction of glucose regulated proteins. Nature 332 462-464.

2. Mori K, Ma W, Gething MJ, Sambrook J (1993) A transmembrane protein with a cdc2+/CDC28-related kinase activity is required for signaling from the ER to the nucleus. Cell 743-56.

3. Colas, P., Cohen, B., Jessen, T., Grishna, I., McCoy, J., and Brent, R. (1996) Genetic selection of peptide aptamers that recognize and inhibit cyclin dependent kinase 2. Nature 380 548-550.

(10) APPENDICES

Reprints of abstracts and manuscripts

(12) BIBLIOGRAPHY

Colas, P., Cohen, B., Ferrigno, P., Silver, P. and Brent, R. Targeted modification and transportation of cellular proteins. 2000, Proc. Natl. Acad. Sci., In press.

Morehouse, H., Ferrigno, P., Way, JC and Silver, PA. Using Yeast to model disease processes in cancer. 2000. Era of Hope Proceedings, DOD Breast Cancer Research Meeting, Atlanta, GA.

PERSONNEL RECEIVING PAY FROM THE RESEARCH EFFORT

Heather Morehouse, Graduate Student

Paul Ferrigno, Postdoctoral Fellow

Elbert Chiang, Technician

Docket No. 157/47097

UNITED STATES PATENT APPLICATION

for

Title: GENETIC SELECTION METHOD FOR IDENTIFYING LIGANDS FOR TRANSMEMBRANE PROTEINS

Inventor(s): Pamela Silver
Tom Roberts
Charles Stiles

Counsel: Ronald I. Eisenstein
DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
130 Water Street
Boston, Massachusetts 02109-4280
Tel. (617) 523-3400

GENETIC SELECTION METHOD FOR IDENTIFYING LIGANDS FOR TRANSMEMBRANE PROTEINS

The present invention is directed to an improved method for identifying "orphan receptors" which involves a genetic selection for ligand-receptor interaction using a recombinant eukaryotic cell, preferably yeast, as a selection system.

BACKGROUND

Advances in molecular, cellular and viral biology have resulted in the identification of numerous transmembrane receptors. These advances have also made it possible to obtain transcripts and DNA encoding a range of proteins including putative transmembrane receptors. The identification of these receptors and putative receptors makes it possible to identify ligands that interact with these receptors permitting one to better understand the biology of those receptors and/or screen for compounds that modulate the effect of such receptors. However, an increasing problem is finding simple and accurate methods for identifying the specific ligands for each of these transmembrane receptors and putative transmembrane receptors. Those receptors for which a ligand has not yet been identified are referred to as "orphan receptors". Such orphan receptors are becoming more numerous as more DNA sequences, including DNA sequences encoding putative receptors, become available.

Identifying the actual ligand that interacts with such receptors can be extremely important as many of these transmembrane receptors are associated with important cellular functions. For example, many transmembrane receptors have kinase activity and are growth factor receptors and some have been associated with malignant transformation of cells. For instance, growth factor independence in cancer cells has been correlated with overexpression of growth factor receptors such as erbB2 in breast cancer [19]. The overexpression of erbB2 has been shown to activate the ras/MAP kinase pathway and inhibition of the activation of this pathway has been shown to correlate with decreased cellular

proliferation [3,9]. Potential links between tumor-associated overexpression of the erbB2 receptor and reduced survival of primary breast cancer patients with metastatic auxiliary lymph node involvement exists [1,2,23]. However, despite the considerable interest in erbB2, the specific ligand that interacts with it has not yet been identified making it an orphan receptor.

Being able to identify the actual ligand that interacts with a receptor such as the erbB2 receptor permits not only a better understanding of the complex physiological interactions involved, but facilitates the development of better drug assays. Thus, it would be desirable to have a better means for assaying and selecting ligands for these orphan receptors.

Another difficulty that currently exists in rational drug development is being able to identify when a ligand-receptor interaction occurs.

Compounds including small polypeptides that interact with not only orphan receptors, but other transmembrane receptors are currently screened by a wide variety of different assays. However, it would be desirable to develop new and simple assays to determine where an interaction is occurring not only to select compounds that modulate receptor activity, either positively or negatively, but to have a simple means to determine optimal peptide sequences that will, for example, fit into the ligand binding site. Identifying such compounds permits more efficient design of compounds that can be used in, for example, anti-cancer or anti-viral strategies. Thus, it would be useful to have a simple method for selecting only those cells where such ligand-receptor occurs.

SUMMARY OF INVENTION

We have now discovered a simple method for identifying ligands and ligand-receptor interactions involving transmembrane proteins that form dimers, preferably homodimers, for activation. Preferably, the transmembrane protein is a transmembrane receptor having protein kinase activity, such as a transmembrane tyrosine kinase receptor.

This method uses the unfolded protein response (UPR) pathway that is present in all eukaryotic cells, and conserved through evolution in organisms as

divergent as mammals and yeast. The accumulation of unfolded proteins in the endoplasmic reticulum triggers a signal that is transmitted to the nucleus and results in increased transcription of chaperone proteins and enzymes that function to induce the correct protein folding. In the yeast, *Saccharomyces cerevisiae*, two of the essential components of the UPR pathway have been identified. In the presence of unfolded proteins, the transmembrane kinase IRE1p transmits a downstream signal that activates transcription of chaperone proteins and enzymes. This signal is manifested by the binding of nuclear factors to the unfolded response element (UPRE), a 22 bp upstream activating element having the sequence:

5'-GATCTGTCGACAGGAACTGGACAGCGTGTCGAAAAAGC-3' (SEQ ID NO:1)

3'-ACAGCTGTCCTTGACCTGTCGCACAGCTTTTTCGAGCT-5' (SEQ ID NO:2)

The UPRE is necessary and sufficient to activate transcription of a linked promoter in response to the accumulation of unfolded proteins in the endoplasmic reticulum (ER) (Mori *et al., EMBO J* 11:2583-2593, 1992).

While the present method can be used in any eukaryotic cell that has a UPR pathway by using the IRE1/UPRE interaction or analog thereof, a preferred embodiment of the present invention involves using recombinant *S. cerevisiae* cell that contains a DNA segment encoding a IRE1/ERNI1 kinase domain fused to the extracellular domain of the receptor of interest, referred to as a chimeric receptor. Receptors of interest include, but are not necessarily limited to eukaryotic, viral, insect and mammalian receptors. Preferably, the receptor is a mammalian receptor. More preferably, the mammal is a human and the receptor is an orphan receptor. Like mammalian growth factors, IRE1 oligomerizes and is phosphorylated in trans in response to an accumulation of unfolded proteins in the endoplasmic reticulum (ER) [22]. When the appropriate ligand is secreted into the ER lumen, the chimeric receptor will oligomerize and activate the unfolded protein response (UPR) signaling pathway. Those cells where the pathway has been activated can readily be identified and selected by using a reporter system activated by the UPR signalling pathway. For example, one can use a responsive element such as the unfolded protein responsive element (UPRE) containing

promoter fused to a marker gene such as LacZ [Cox, 1993]. Although the induction by the UPRE in an unfolded protein response is normally only two to four-fold because the protein is made at a high basal level, one can increase the level of induction by generating constructs containing multiple copies of the UPRE, and wherein the constitutive promoter elements of the marker genes are absent. Thus, the induction ratio would be much higher. The yeast colonies that have a UPR signal would turn blue on the Xgal indicator plates containing tunicamycin. Numerous other reporters can readily be used. For example, fusing the UPRE-containing promoter to the gene encoding the naturally green fluorescent protein (GFP) so that induction can be measured in living cells by fluorescence, thereby permitting the use of cell sorters. By this means, one can readily identify cells wherein a ligand has bound to a receptor, and induced the UPR signal. Thereafter, the cDNA encoding the putative ligand can be readily selected. In the instances where more than a single cDNA is selected, those cDNAs which actually encode a ligand which binds to the receptor can be readily resolved by transfecting individual cell lines containing the receptor with a vector containing one of the selected cDNAs encoding the putative ligand. One can then screen each of the transfected cell lines using standard methods (e.g., receptor binding assays) to identify those cells in which a ligand-receptor interaction occurs. The transfected cell can be a yeast cell of the present invention or any other cell that expresses the receptor protein.

DETAILED DESCRIPTION OF THE INVENTION

The present invention uses eukaryotic cells having a UPR pathway. In preferred embodiments, the present invention permits one to take advantage of the unique nature of single-celled eukaryotic organisms such as yeast in a method for readily identifying (a) ligands for orphan receptors, (b) compounds such as small molecules that specifically bind to transmembrane proteins that dimerize for activation and (c) compounds that will modulate the effect of transmembrane receptors that dimerize.

S. cerevisiae is the simplest eukaryote to possess all the characteristic features of mammalian cells from its highly conserved cell cycle machinery to hormone dependent differentiation pathways. However, this system also combines many characteristics similar to that of prokaryotes, for example, a relatively simple genome, rapid growth, etc. These features in combination with its unusually well understood biology and the recent complete sequencing of its genome make it an excellent system to study protein interactions.

We have discovered that one can take advantage of one of the signalling pathways of eukaryotic cells to identify and select ligands and other compounds that interact with transmembrane proteins. Specifically, the transmembrane protein is one that oligomerizes upon interaction with another compound on one surface of the membrane and becomes activated at the other side of the membrane to transmit a detectable response as a result of the oligomerization. Any transmembrane protein whose function involves the binding of a ligand, protein oligomerization and consequent signal transmission can be subjected to the methods of the present invention. The transmembrane protein can be derived from any organism including eukaryotes and viruses. In preferred embodiments, the transmembrane protein is a receptor, a receptor with kinase activity, and more preferably, a class I growth factor receptor. Preferably the protein is a mammalian protein. Still more preferably, the protein is a human protein.

We have discovered that one can construct chimeric transmembrane proteins containing the extracellular domain of, for example, a receptor fused to the cytoplasmic kinase domain of a yeast receptor, namely the IRE1/ERN1 (hereinafter IRE1) receptor.

IRE1 is involved in the Unfolded Protein Response (UPR) pathway. Accumulation of unfolded proteins in the ER induces transcription of proteins, including chaperones and enzymes that function to properly fold the proteins [12]. The IRE1 protein in yeast plays a major role in transducing the signal from the ER lumen to the nucleus [5,14].

The IRE1 gene encodes a transmembrane serine/threonine protein kinase that is located in the ER membrane with its kinase domain in the cytoplasm (or

the nuclear interior) [5, 16]. IRE1 is believed to act analogously to plasma membrane receptors by transmitting a signal from the ER lumen to the cytoplasm after interaction with an appropriate ligand. Like many transmembrane receptors, for example growth factor receptors, IRE1 oligomerizes and is phosphorylated in trans in response to accumulation of unfolded proteins in the ER [22]. Analogously to the case with other receptors oligomerization results in a signalling cascade causing the activation of transcription factors in the nucleus.

The synthesis of the ER-resident proteins such as the chaperone BiP which in yeast is also known as Kar2, and protein disulfide isomerase (PDI) involved in protein folding and cellular reactions is regulated in response to cellular requirements. For example, when cells are exposed to reagents such as tunicamycin, that inhibit glycosylation, to reducing agents, or to calcium ionophores that deplete ER-calcium stores, induction of several ER-resident proteins occurs at the transcriptional level [8,12,18,20]. All of these treatments are thought to cause improper protein folding in the ER, the aforementioned UPR. A signal from the ER lumen is transmitted to the nucleus by activated IRE1 where transcription is then activated. Potential unfolded protein response elements (UPREs) have now been identified in promoters of at least six genes encoding ER-based enzymes that are induced in response to unfolded proteins [11,14,21]. The UPRE is a 22 base pair sequence. Moreover, it has been found that a single UPRE is sufficient to activate transcription in response to the accumulation of unfolded proteins when it is inserted into a heterologous promoter [Kohn et al, 1993, Mori et al, 1992].

We have discovered that one can take advantage of the signal transmitted by IRE1 to a UPRE to readily identify and select (a) ligands that bind with a wide range of orphan receptors, (b) other compounds that interact with such receptors and/or (c) compounds that modulate the response of such receptors.

This can be done by removing the ligand binding domain of IRE1 and substituting therefore the extracellular domain of the transmembrane protein under study. Preferably, the transmembrane protein is a receptor protein. Preferably, the mammalian protein is a human protein. Transmembrane

receptors are well known in the art. For example, they include both receptors and oncogenes. For example, many oncogenes show some homology to genes involved in cell growth. For example, see the table below.

TABLE[1]

CATEGORY	ONCOGENE	HOMOLOGUS CELLULAR GENE
Growth Factors	sis int-2	PDGF-/2 FGF-like
Transmembrane growth factors	erbB erbB-2 (neu, HER-2) fms ros, kit, and others	EGF receptor M-CSF receptor
Membrane-associated tyrosine kinases	abl	
Membrane associated guanine nucleotide binding proteins	src family[2] fes, fps[3] K-, N- and H-ras	
Cytoplasmic serine-threonine kinases	raf/mil mos	
Cytoplasmic hormone receptors	erbA	Thyroid hormone receptor

Putative receptors that share certain of the analogous domains have been identified. By using known techniques the DNA encoding the extracellular portion of a transmembrane protein such as a receptor protein can be substituted for the DNA encoding the ER-luminal portion of the IRE1 gene. For example, using at least the ligand binding portion from the extracellular domain of a

[1] Adapted from Druker, B.J., et al., *N. Eng. J. of Mol.* 321:1383-1392 (1982). PDGF denotes platelet-derived growth factor, FGF fibroblast growth factor, EGF epidermal growth factor, and M-CSF mononuclear-phagocyte growth factor.

[2] The family includes src, fgr, yes, lck, hck, fyn, lyn, and tkl.

[3] The subcellular location of these oncogene products is uncertain.

mammalian growth factor receptor such as EGF or erbB2. In preferred embodiments, one would fuse the entire cDNA portion encoding the extracellular domain of the transmembrane protein to the cDNA encoding transmembrane and cytoplasmic kinase domains of IRE1. In alternative embodiments, one can create a chimeric gene encoding the chimeric protein wherein at least the cytoplasmic domain of the receptor is deleted, and replaced by the cytoplasmic kinase domain of IRE1. These constructs can readily be made by those skilled in the art using known techniques based upon the present disclosure. For example, the sequence of IRE1 is well known. See for example, Mori, et al. 1993. Similarly, the general structure of most receptors and domains is well known. Convenient restriction sites are known in IRE1 and can readily be identified in the transmembrane receptors. Additionally, unique restriction sites in these genes can also be created by standard techniques. Thereafter using standard techniques one can fuse the portions together. Prokaryotic hosts such as *E. coli* are a convenient source for preparation on large amounts of chimeric genes. These chimeric receptors can be inserted in a yeast expression vector and used to either transiently transfect or constitutively transfect the yeast. In preferred embodiments, cassettes containing the desired portion of nucleic acid from IRE1 can be made wherein the extracellular portion from a transmembrane protein can readily be inserted.

A number of other promoters have also been shown to be useful for expression of various genes in yeast. For example, promoters naturally associated with the *Saccharomyces cerevisiae* genes TP11 (triose phosphate isomerase), PGK1 (phosphoglycerate kinase), PYK1 (pyruvate kinase) TKH1, TDH2, and TDH3 (glyceraldehyde phosphate dehydrogenase or triose phosphate dehydrogenase), and ENO1 (enolase 1) have been described as useful for expression of genes in yeast (Kawasaki, U.S. Patent No. 4,599,311; Kingsman and Kingsman, U.S. Patent No. 4,615,974; Burke *et al.*, EPO Patent Application NO. 84300091.0) as well as the native IRE1 promoter. These promoters can be used to control expression of the chimeric protein. Typically, it is preferable to use promoters that do not result in high levels of expression of the chimeric protein,

particularly when looking at identifying ligands. The reason for this is that if too high a level of chimeric protein is expressed autooligomerization may occur in the absence of ligand-receptor interaction.

In contrast, when looking for compounds that modulate--either positively or negative-- the signal by the chimeric protein, then high levels of expression are preferable.

In addition to the particular examples described herein, cassettes containing cDNA encoding the extracellular domain of the receptor of interest fused to cDNA encoding the transmembrane and cytoplasmic domains of IRE1 can be constructed using a variety of methods well known to those skilled in the art. As stated above, generally, where it is desirable to identify the naturally occurring ligand of the receptor, or to screen for compounds which enhance the function of the receptor low copy number or integrating vectors will be used, and the expression of the receptor-IRE1 fusion protein will be placed under the control of the native IRE1 promoter or one which promotes the expression of the fusion protein at levels comparable to that of the IRE1 promoter. In contrast, where it is desirable to identify compounds which inhibit the action of the receptor higher levels of the receptor-Ire1 fusion can be achieved by using a high copy plasmid, by placing the cDNA under the control of a strong promoter (e.gs, GAL1, GAL10, ENO1, ENO2, ADH2, Met3, or both.

In preferred embodiments, one would use yeast cell lines wherein the native IRE1 gene has been inactivated. This can also readily be accomplished by standard techniques by those skilled in the art. (See, for example, Guthrie et al, *Methods in Enzymology*, Vol. 194, Academy Press, Inc., 1991.)

Similarly the yeast cells are transfected with nucleic acid encoding a reporter system that would respond to the activation of the chimeric protein. This can readily be done by inserting at least one UPRE sequence in the promoter of a reporter gene of interest. Procedures for the construction of suitable vectors, the stable transfection of cells and the analysis of the transfected cells for gene expression are well known in the art. See for example, Kaiser et al, Methods in Yeast Genetics, Cold Spring Harbor Press, 1994; Ausubel et al., Current Protocols

in Molecular Biology, John Wiley & Sons, New York, 1989)

Generally, the reporter construct contains a reporter gene whose expression is under the control of a promoter containing at least one, and preferably at least between 2-4 copies of the UPRE sequence. The reporter gene can be selected from any gene whose protein product is readily detected, including but not limited to, those detectable by enzyme assays, fluorescence, immunoassay, drug resistance or auxotrophic selection. Examples of the numerous useful reporter genes available include lacZ, CAT, GFP, URA3, TRP2, LYS2 and HIS3.

Promoters which naturally contain one or more UPRE's are preferably be used in the reporter construction. Such promoters include the promoters that regulate KAR2, PD1, EUG1, and FKB2 (Kohn, 1993, Meri 1993; Schlenstedt, 1995). Alternatively, one or more UPRE's can be inserted into any other desirable promoter such as those described *supra*. The UPRE is a 22 bp sequence present in the promoters of genes that are activated by the UPR. It has been demonstrated that a single UPRE is sufficient to activate transcription in response to the accumulation of unfolded proteins when it is inserted into a heterologous promoter (Kohn et al. 1993, Mori et al. 1992). Mutational analysis has defined the nucleotides within the UPRE that are essential for its ability to activate transcription (Mori et al., supra). When one copy of UPRE is present in the promoter of the reporter gene there is typically a 4 to 8-fold increase in expression of the reporter protein upon addition of tunicamycin. However, increases of between 40-43 fold have been observed when the promoter contain 2-4 tandem copies. Reporter constructs containing UPRE elements can be prepared as described by Cox, et al, 1993; Cox and Walter, Cell 87:391-394, 1996.

Both the chimeric gene cassettes and the reporter constructs can be introduced into the yeast cell using any suitable vector. One of the advantages of the yeast based system is that a number of bacteria/yeast shuttle vectors are readily available (e.g., New England BioLabs, American Tissue Culture Corporation) which allow introduction of different copy numbers of the cDNA of interest into the cell. For example, when high levels of gene expression are

desirable, yeast episomal plasmids, such as Yep24, based on the yeast two micron circle and which are replicated in the cell at high copy numbers can be used. Plasmids, such as YRp17, which contain a yeast chromosomal derived autonomous replicating sequence (ARS) can be used when intermediate copy numbers of the DNA are preferred. Alternatively, vectors (e.g., YCp50) which contain yeast chromosomal centromere sequences are also available when it is desirable to maintain the DNA of interest at a level of only one to two copies per cell. In addition, a number of yeast integrating plasmids such as YIp5 can be used for stable introduction of the cDNA into the yeast strains containing a reporter construct because integration of the reporter construct into a yeast chromosome would minimize the need for maintaining multiple selection criteria.

Using the aforementioned vectors for the chimeric receptor and the reporter one can create transient cells or preferably, stable cell lines expressing the chimeric receptor/reporter. In either event, one can then readily use those cell lines to identify the ligand or compound of interest by a simple selection based upon looking at the reporter. Typically, when trying to identify a ligand one does so by selecting cells where the receptor has been activated. With respect to the other compounds, one can select those cells where there has been a change in the reporter as opposed to control cells. For example, one can use yeast cells where the reporter indicates the system has been activated, e.g., by high level expression of the chimeric reporter and autoactivation. Then one can identify compounds that inhibit the receptor by looking at a diminishing of the reporter.

In a preferred embodiment, the present invention is used to identify ligands for receptors. In that case the chimeric receptor has at least the ligand binding domain of the orphan receptor as part of the chimeric receptor. Preferably one uses stable cell lines expressing the chimeric receptor under the control of the IRE1 promoter and a reporter where the promoter contains at least one UPRE. One then needs to transfect the stable cells with nucleic acid segment encoding a putative ligand. This can readily be done by standard techniques. For example, using a library of cDNAs encoding proteins that could be expected to contain the ligand contained in an expression plasmid that replicates and produces mRNA

extrachromosomally when transfected into yeast cells. For example, the orphan receptor erbB2 is associated with breast tissue. Thus, one would preferably use a library containing DNA encoding breast tissue associated proteins. Similarly, if the orphan receptor of interest had been identified from a brain tissue library one would expect to find the ligand for that receptor using a brain tissue library. Preparation of nucleic acid libraries such as cDNA libraries is known in the art and can readily be accomplished. Alternatively, one can purchase commercial libraries. For example, a number of libraries have been designed to express mammalian proteins, preferably human, proteins, in systems such as yeast. One would tranfect a population of yeast cells containing the chimeric receptor/reporter system with the cDNA library by standard means. For example, using plasmids containing cDNA encoding a putative ligand operably linked to a yeast promoter. As a result of the use of yeast and the reporter system disclosed herein, one can readily identify from the population of thousands of potential choices of cDNA only those cells that show activation of the UPRE. Those cells wherein UPRE has been activated demonstrate binding by the ligand to chimeric receptor are selected and plasmid DNA isolated by standard technique. DNA encoding the ligand can be isolated from the plasmid using standard techniques including PCR. If desired, selected cells can be cultivated by standard culturing techniques to large numbers. Additionally, if desired, the plasmid can then be amplified using, for example, *E. coli* and used to transfect yeast cells for a second round of screening.

In most instances this permits one to precisely identify the particular cDNA encoding the ligand. In some instances, there may be a few choices. This can be handled readily by a second selection system wherein a population of yeast cells containing the chimeric receptor UPRE reporter system are transfected by an expression vector containing cDNA encoding only a specific sequence. Alternatively, one can use other cells expressing the actual receptor of interest. By this means, ligands for orphan receptors can readily be identified.

In some preferred embodiments, passage of the ligands through the ER can be slowed by use of an ER retention sequence such as a KDEL sequence or its

analog. For example, in yeast an HDEL sequence [15]. Additionally, in some preferred instances, it will be useful to have the plasmids encoding the DNAs in these libraries also contain a marker gene such as a nutrient specific marker to further assist selective growth of cells containing the cDNA of interest. This can be done by standard techniques using the information in the present disclosure.

The receptor ligand interactions that occur in the ER lumen are essentially analogous to the extracellular milieu where such interactions normally occur. Moreover, the present system isolates the receptor ligand interactions from other receptors and possible cross-talk which can confuse ligand identification. For example, we can construct a fusion gene encoding a chimeric protein containing the extracellular domain of a mammalian tyrosine kinase receptor such as an erbB2 or EGF receptor fused to the transmembrane and cytoplasmic kinase domain for IRE1. As previously mentioned, the over expression of erbB2 is associated with severity of certain breast cancers, while EGF is the canonical growth factor receptor which undergoes ligand induced oligomerization and activation. Activity is monitored by looking at induction of UPRE reporters. A positive response indicates that ligand dependent dimerization of the IRE1 cytoplasmic kinase domain has occurred. As a result of this, when trying to identify a putative ligand one must take care to ensure that low numbers of chimeric receptors are expressed. Thus, it is typically preferable in this instance to use a yeast promoter such as the IRE1 promoter in a cell where nature IRE1 expression has been knocked out. Thereafter, one looks for ligand-receptor interaction. The possibility of this interaction occurring can be increased by a variety of means. For example, by having the ligand contain an ER retention sequence such as HDEL. In one embodiment *sec* mutants in which transport of proteins from the ER to the golgi apparatus can be blocked can be used. These mutants are temperature sensitive and die at non-permissive temperatures. Thus, one would select an intermediate temperature at which exit of the protein from the ER is slowed, but not stopped. This effectively would increase the length of exposure of the putative ligand to the ligand binding portion of the transmembrane protein of interest.

Any of a wide range of assays can be used to identify activation of the UPRE. For example, a reporter plasmid containing the UPRE-containing promotor fused to *lacZ* [5] or having a UPRE-containing promoter fused to the gene encoding HIS3. When *lacZ* is used, colonies turn blue on Xgal indicator plates containing tunicamycin. When GFP is used, one can use cell sorters to take advantage of the fluorescence to identify positive yeast transformants. When HIS3 is used, the cell line should be a HIS3-yeast mutant, so that cells containing the reporter would be unable to grow on medium lacking histidine, except in the presence of productive IRE1 activation.

Thereafter, one would use an appropriate library such as with erbB2 and look for cells that have activated the UPRE system. As aforesaid, one can then select those cells, grow them up and then identify the putative ligand by standard techniques.

The DNA sequence of the clones that pass the above test can then be determined and compared to known sequences in various data banks. These putative ligand can then be expressed in a wide range of cells. For example, *E. coli*, yeast, baculovirus cells and radiolabeled. These labeled ligands can then be tested directly for binding to the surface of breast cancer cells such as 21MT-1 and 21MT-2 [19] known to express varying levels of erbB2 and/or to cells engineered to express only erbB2 using iodinated and/or biotinylated ligand in order to measure binding. One can use competition by the same ligand, only unlabeled, versus competition by an unrelated growth factor such as PDGF as a control. Alternatively, as mentioned above, one can express only the ligand in a yeast cell to determine if that specific ligand activates the UPRE response. Thereafter, one can confirm that identification with the appropriate cell which expresses the intact naturally occurring or recombinant transmembrane protein. In this manner, one can readily identify ligands for orphan receptors.

One can also use this system to identify compounds (aptamers) that bind to the ligand binding domain of the transmembrane protein. In such an assay, the yeast serves as a vessel in which the interaction occurs and generates the signal that alerts one to the potential interaction. For example, a protein from

any organism can be expressed in yeast (the so called "bait") in conjunction with a cDNA library from any organism ("the fish"). The present method permits the rapid examination of the interaction of proteins of interest with a large number of random proteins expressed as aptamers. These aptamers can be prepared by known techniques or bought commercially as a library. For example, there are libraries of 10^8 plasmids available (MGH & Genetics Institute) that direct the synthesis of randomly encoded 20-mer polypeptides within *E. coli* thioredoxin. The peptides are displayed as loops that protrude from the surface at the thioredoxin active sites, whereas chimeric peptide-proteins have no thioredoxin activity. Genes encoding each aptamer would be fused to an activation domain and a nuclear targeting domain, inserted into a plasmid and used to transfect the yeast cell and then screened for aptamer binding to the protein of interest by standard techniques such as those used in the standard two-way hybrid approach [4,10]. Thus, the thioredoxin-aptamer is the "prey" and the "bait" will be the extracellular domain of the protein of interest, e.g. erbB-2. While these aptamers are unlikely to induce receptor dimerization, they will identify peptides that bind to the surface of the ligand binding domain. Thus, they will readily identify ligands that can compete for ligand binding. Moreover it is possible to take the aptamer and modify such ligands, e.g. by adding different groups or altering the sequence of the aptamer to modulate receptor activation.

For example, an insert containing the cassette encoding each aptamer plasmid can be placed into a similar vector in which the thioredoxin protein scaffold is expressed in yeast with an N-terminal signal sequence and a C-terminal ER retention sequence. The protein can be expressed by an inducible promoter such as the GAL promoter. Expression of at least some of the peptides can cause an observable phenotype such as induction of UPR, slow growth, lethality and/or inhibition of UPR as opposed to the situation when a normal ligand is present. Those plasmids that create an observable phenotype can then be produced in bacteria and used to test whether the aptamer binds the transmembrane receptor directly or competes in the binding of ligand or some other interaction. Thioredoxin fusion proteins always appear to fold correctly,

even when grossly overproduced in *E. coli*; inclusion bodies do not appear to form even when the protein is 40% of total *E. coli* [13].

After identifying an initial group of aptamers, an optimized set can be constructed by standard techniques such as mutagenesis so that the peptide expresses all other possible amino acids. Moreover, by appropriate screening, one can insure that the aptamer identified have a specific binding content. For example, of at least about 10^{-8}M.

In an alternative embodiment of the above technique, one can introduce the cDNA encoding compounds or aptamers by encoding an ER signal with the compounds or aptamers directly into the ER and test the transfected cells directly for activation or inhibition of activation. This makes it possible to bypass an initial screening step and immediately obtain peptide ligands that mimic, inhibit or enhance receptor function. In these tests one would preferably use higher numbers of chimeric receptors for each observation.

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, numerous equivalents to the specific methods, nucleic acids, assays and reagents described herein. For example, using the information disclosed herein, one could readily identify the analogous *IRE* genes and UPRE elements from other yeast species such as *S. pombe* or *K. lactis*, as well as other eukaryotic systems including, but not limited to, insect and mammalian culture systems. Accordingly, such equivalents are considered to be within the scope of this invention.

REFERENCES

1. Berger, M.S., et al., *Cancer Res.* 48:1238-1243 (1988).
2. Borg, A., et al., *Cancer Res.* 50:4332-4337 (1990).
3. Clark, J.W., et al., *Int. J. Cancer* 65:186-191 (1996).
4. Colas, P., et al., *Nature* 380:548-550 (1996).
5. Cox, J.S., et al., *Cell* 73:1197-1206 (1993).
6. Cox, J.S., et al., *Cell* 87:391-394 (1996).
7. Fields, S., et al., *Nature* 340:245-246 (1989).
8. Filmus, J., et al., *Oncogene* 9:3627-3633 (1994).
9. Graus-Porta, D., et al., *Mol. Cell. Biol.* 15:1182-1191 (1995).
10. Gyuris J., et al., *Cell* 75:791-803 (1993).
11. Kohno, K., et al., *Mol. cell. Biol.* 13:877-890 (1993).
12. Kozutsumi, Y., et al., *Nature* 332:462-464 (1988).
13. LaVallie, E.R., et al., *Biotech.* 11:187-193 (1993).
14. Mori, K., et al., *Cell* 74:743-756 (1993).
15. Munro, S., et al., *Cell* 48:899-907 (1987).
16. Nikawa, J., et al., *Mol. Microbiol.* 6:441-1446 (1992).
17. Nojima, H., et al., *Nucl. Acids Res.* 22:5279-5288 (1994).
18. Normington, K., et al., *Cell* 57:1223-1236 (1989).
19. Ram, T.G., et al., *Mol. Carcinogenesis* 15:227-238 (1996).
20. Rose, M.D., et al., *Cell* 57:1211-1221 (1989).
21. Schlenstedt, G., et al., *J. Cell Biol.* 129:979-988 (1995).
22. Shamu, C.E., et al., *EMBO J.* 15:3028-3039 (1996).
23. Slamon, D.J., et al., *Science* 235:177-182 (1987).

All publications and patents mentioned herein are incorporated by reference.

IN THE CLAIMS

We claim:

1. A chimeric protein comprising

(a) a cytoplasmic kinase domain of an IRE1 protein or analog thereof,

(b) a transmembrane domain and,

(c) a ligand binding domain of a transmembrane protein other than IRE1.

2. The chimeric protein of claim 1, wherein the IRE1 protein is a yeast IRE1 protein.

3. The chimeric protein of claim 1, wherein the transmembrane protein is a viral protein or a mammalian protein.

4. The chimeric protein of claim 2, wherein the transmembrane protein is a mammalian receptor protein.

5. The chimeric protein of claims 1, 2, 3 or 4, which further contains an endoplasmic reticulum retention sequence.

6. A nucleic acid sequence encoding the protein of claims 1, 2, 3 or 4.

7. A nucleic acid sequence encoding the protein of claim 5.

8. A method of identifying compounds that interact with a transmembrane protein comprising:

(a) transfecting a eukaryotic cell having an unfolded protein response (UPR) pathway with the nucleic acid sequence of claim 6;

(b) transfecting said eukaryotic cell with a nucleic acid sequence having a reporter gene operably linked to a promoter that responds to a signal

from said chimeric protein upon activation of said chimeric protein by oligomerization and ligand binding;

(c) adding a compound to be tested; and

(d) identifying cells wherein there is a change in the reporter, wherein such change is indicative of interaction between the compound to be tested and the chimeric protein thereby identifying said compound as interacting with a transmembrane protein.

9. The method of claim 8, wherein the eukaryotic cell is a yeast cell.

10. The method of claim 9, wherein the gene encoding the native IRE1 protein has been inactivated.

11. The method of claim 10, wherein the compound is added by transfecting the yeast cell with a DNA sequence encoding the compound.

12. The method of claim 11, wherein the compound is a ligand.

13. The method of claim 10, wherein the cell has been transiently transfected by the nucleic acid sequence encoding the reporter gene and the nucleic acid sequence of claim 6.

14. The method of claim 10, wherein the cell has been constitutively transfected by the nucleic acid sequence encoding the reporter gene and the nucleic acid sequence of claim 6.

15. The method of claim 10, wherein the promoter operably linked to the reporter gene responds to the signal from the chimeric protein by having an unfolded protein responsive element (UPRE) in it.

ABSTRACT

A chimeric protein having an (a) IRE1 or analog cytoplasmic kinase domain, (b) a transmembrane domain, and (c) a ligand binding domain of a transmembrane protein other than IRE1 is described. This protein can be used to identify and/or screen for ligands and other molecules that interact with the ligand binding domain.

Classification : Genetics

Targeted modification and transportation of cellular proteins

Peptide aptamers/ combinatorial peptide libraries/ protein interactions/ cellular nanotechnology/ protein design/

Pierre Colas, Barak Cohen, Paul Ko Ferrigno, Pamela A. Silver and Roger Brent

Laboratoire de Biologie Moléculaire et Cellulaire, Ecole Normale Supérieure de Lyon, 46 Allée d'Italie, 69364 Lyon Cedex 07, France

Department of Genetics, Harvard Medical School, Boston, MA 02115, USA

Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School and The Dana Farber Cancer Insitute, 44 Binney Street, Boston, MA 02115

The Molecular Sciences Institute, 2168 Shattuck Avenue, Berkeley, CA 94704, USA

510 647 0695 / 510 647 0699 (fax) / brent@molsci.org

Abstract

Peptide aptamers are proteins selected from combinatorial libraries that display conformationally constrained variable regions. Peptide aptamers can disrupt specific protein interactions, and thus represent a useful method for manipulating protein function *in vivo*. Here, we describe aptamer derivatives that extend the range of such functional manipulations. We isolated an aptamer with increased affinity for its Cdk2 target by mutagenizing an existing aptamer and identifying tighter binding mutants with calibrated two-hybrid reporter genes. We used this and other anti-Cdk2 aptamers as recognition domains in chimeric proteins that contained other functional moieties. Aptamers fused to the catalytic domain of an ubiquitin ligase specifically decorated LexA-Cdk2 with ubiquitin moieties *in vivo*. Aptamers against Cdk2 and another protein, Ste5, that carried a nuclear localization sequence transported their target proteins into the nucleus. These experiments indicate that fusion proteins containing aptameric recognition moieties will be useful for specific modification of protein function *in vivo*.

Introduction

Peptide aptamers are recognition reagents that embody some features of antibodies (1). They consist of a conformatiomally constrained variable region (here 20 amino acids) displayed by a platform protein, (here, *E.coli* Thioredoxin A). We currently select peptide aptamers from combinatorial libraries by two-hybrid methods, using aptamer derivatives that also bear acidic activation domains, and using LexA derivatives of the desired target proteins as baits; such selection ensures that the aptamers function *in vivo*. Peptide aptamers typically exhibit Kds for their target of about 10^{-7}-10^{-8}M (1). These molecules can discriminate between closely related members of protein families (1), and even between different allelic forms of proteins (2, this work). Anti-Cdk2 aptamers competitively inhibit the interaction of Cdk2 with one of its substrates and, when expressed in human cells, delay progress through the cell cycle (3). Similarly, anti-Ras and anti-E2F aptamers disrupt the function of their protein targets in mammalian cells (4, Xu et al. in preparation) and anti-*Drosophila* Cdc2 and Cdc2c aptamers inhibit the function of their targets in imaginal disks (5). Finally, aptamers can be used as dominant genetic agents to cause a phenotype and then to identify, in subsequent two hybrid assays, the proteins and interactions they target (6). These results demonstrate

that peptide aptamers can disrupt specific protein interactions *in vivo*, and thus allow the manipulation of regulatory networks with a high level of precision (reviewed in ref. 7).

Here, we describe derivative peptide aptamers that covalently modify or change the subcellular localization of their target proteins. We first describe selection of a peptide aptamer with an improved affinity for its target. We use this improved aptamer with others to construct two types of chimeric proteins: "modifiers", which ubiquitinate their target proteins, and "transporters", which translocate their target proteins to the cell nucleus.

Identification of a higher affinity variant aptamer

We amplified the variable region of anti-Cdk2 aptamer 10 from the original library vector (1) following a mutagenic PCR protocol as described (8). We ligated the purified amplified products into the RsrII-cut library vector, pJM-1 (1), which directs their conditional expression under the control of the Gal1 promoter, and introduced the ligation mix into *E.coli* DH5α. We prepared plasmid DNA from a pool of 15,000 independent colonies. We transformed (9) this pool into EGY48 that contained LexA-Cdk2 (1) and pRB1840 (10) to obtain 40,000 transformants on $Ura^-His^-Trp^-$ glucose plates. We replica plated these transformants onto $Ura^-His^-Trp^-$ glucose /Xgal and $Ura^-His^-Trp^-$ galactose/Xgal plates. After two days at 30°C, we transferred the 16 colonies that showed a blue color onto $Ura^-His^-Trp^-$ glucose plates. We replica plated these master plates onto $Ura^-His^-Trp^-$ glucose/Xgal and $Ura^-His^-Trp^-$ galactose/Xgal plates, and confirmed that 12 of the initial 16 colonies again displayed galactose-dependent blue color. We rescued the aptamer-encoding plasmids from these strains (11) and reintroduced the plasmids into EGY48. 7 out of these 12 plasmids conferred an interaction phenotype on galactose- but not on glucose-containing medium.

Construction of fusion proteins

Modifiers. We isolated DNA encoding the *hect* domain of yeast RSP5 by PCR using the oligonucleotides 5'-

ATATCTCGAGATTAAAGTACGTAGAAAGAAC-3' and 5'-ATATGTCGACGGATCCTCATTCTTGACCAAACCCTATG-3', which respectively contained an XhoI site, and BamHI and SalI sites. We subcloned this amplified product into XhoI-cut pJG4-4, a plasmid containing a Trp1 marker, a 2μ replication origin, and that directed the expression of native proteins under the control of the GAL1 promoter, to create pJG4-4(*hect*). We PCR amplified TrxA and peptide aptamers using the oligonucleotides 5'-GGAGGCGAATTCGCCGCCACCCATGGCCGATAAAATTATTCACCTGACTGACG-3' and 5'-ATATCTCGAGCGCCAGGTTAGCGTCGAGGAAC-3', which contained respectively an EcoRI site followed by an initiator codon in a Kozak context, and an XhoI site, and inserted these amplified products into EcoRI/XhoI-cut pJG4-4(*hect*). In order to express the *hect* domain only, we used the following 5' oligonucleotide containing an EcoRI site and an initiator codon in a Kozak context with the above-described 3' oligonucleotide: 5'-ATATGAATTCGCCGCCACCATGGCCATTAAAGTACGTAGAAAGAACATTTTTGAG-3' to PCR amplify the domain from *Saccharomyces cerevisiae* gene RSP5, and introduced this amplified *hect* domain into EcoRI/XhoI-cut pJG4-4. We constructed the mutant *hect* domain using the Transformer site-directed mutagenesis kit from Clontech, according to the manufacturer's instructions, using the mutagenic oligonucleotide 5'-GCCAAAATCTCACACAGCTTTTAACAGAGTTG-3' to change the *hect* active site cysteine to alanine, and the selection oligonucleotide

5'-CGCTAACCTGGCGCCTAGGATTAAAGTACGTAG-3' to change to the XhoI site on the vector into an AvrII site.

For the experiments featuring Myc-tagged ubiquitin, we expressed the modifiers from another vector. To this end, we PCR-amplified 5-, 8-, and 10M-*hect* fusions described above using the oligonucleotides 5'-ATATGTCGACGGATCCTCATTCTTGACCAAACCCTATG-3'and 5'-GGAGGCGAATTCGCCGCCACCCATGGCCGATAAAATTATTCACCTGACTGACG-3'. We introduced the amplified products into EcoRI/XhoI cut pBC103, a plasmid that carries a Ura3 marker, a 2μ replication origin, and that directed the expression of aptamer-*hect* fusions under the control of the GAL1 promoter.

Myc-ubiquitin. Yep105, contains a Trp1 marker, a 2μ replication origin and directs the expression of a Myc-tagged synthetic yeast ubiquitin gene under the control of the *CUP1* promoter, inducible by copper (12).

LexA-7Lys-Cdk2. We began with the bait plasmid encoding LexA-Cdk2 (1). We constructed a DNA sequence encoding a stretch of 7 lysines by annealing the oligonucleotides 5'-AATTGAAGAAGAAAAAAAAGAAAAAGC-3' and 5'-AATTGCTTTTTCTTTTTTTTCTTCTTC-3' and introduced this duplex into the EcoRI site of the bait plasmid. We treated the ligation mixture with EcoRI, introduced it into *E. coli* XL-1 blue, and identified by PCR cells that bore plasmids containing the insert.

Transporters. TrxA and anti-Cdk2 peptide aptamers were PCR amplified as described in "*modifiers*". We introduced the

amplified products into EcoRI/XhoI cut pBC103 and pBC104, two plasmids containing a Ura3 marker, a 2μ replication origin, and that directed the expression of aptamers or NLS-aptamers respectively, under the control of the GAL1 promoter. pJM-C1, -C6, -N1, -N2 and -N3, the plasmids that direct the synthesis of non-NLS- and NLS-tagged anti-Ste5 aptamers, are described elsewhere (6).

Yeast manipulations

We performed interaction mating assays as described (13) using the EGY48 strain for the preys and the EGY42 strain for the baits and reporter plasmids (14). We used pSH18-34, pJK103, and pRB1840 (carrying respectively 8, 2, and 1 LexA operators upstream of a Gal1-*lacZ* fusion gene) in Fig. 1*a*, the pSH18-34-derived LexAop-GFP reporter (Display Systems Biotech, Copenhagen) in Fig. 1*c* and pSH18-34, in Fig. 2.

To measure the interaction phenotypes using the LexAop-GFP reporter gene, we grew overnight liquid cultures from diploid exconjugants in Ura-His-Trp- galactose liquid medium. We measured fluorescence with a FACStar plus (Becton-Dickinson) illuminated with two argon lasers tuned to 488nm and to multiline UV. We recorded with a 530-/+15nm filter to measure yeast fluorescence. We set the FL3-2 voltage (the background) using yeast that did not show an interaction phenotype. We analyzed 30,000 cells for each interaction, and determined mean fluorescence of the yeast population above background using the CellQuest software (Becton-Dickinson).

To perform the modifier assays, we transformed the EGY48 yeast strain with different combinations of targets and aptamer-*hect* fusions (9). We plated transformants onto His^- Trp^- glucose medium. We grew colonies overnight in 4ml of His^-Trp^- galactose liquid medium. For the experiments that used Myc-ubiquitin, we transformed EGY48 with aptamer-hect fusions and Yep105, and EGY42 was transformed with LexA-Cdk2 and pSH18-34. The strains obtained were mated and diploid exconjugants were selected on Ura-His-Trp-Leu- glucose medium. Liquid cultures were inoculated into Ura-His-Trp-Leu- galactose liquid medium, in which 100μM CuSO4 was added or not.

For western analysis (below), we pelleted equal amounts of yeast in logarithmic growth phase and treated the pelleted yeast with zymolase (Seikagaku corp.) at 1mg/ml in 50μl of (1M sorbitol, 0.5M Sodium citrate, 0.5m EDTA, 1M DTT, 1M Potassium phosphate, 0.1M PMSF) for 1hr at 30°C and lysed them with SDS PAGE sample buffer.

Immunoassays

We performed western blots following SDS-PAGE as described (15) using a rabbit anti-LexA serum (16) or a rabbit anti-TrxA serum (17) and secondary antibodies coupled to alkaline phosphatase together with NBT and BCIP as substrates. We scanned the membranes with an optical scanner. For the Myc-Ubiquitin experiments, we used an ECF substrate (Amersham Pharmacia Biotech) and scanned the blot using a phosphoimager (Molecular Dynamics).

For immunofluorescence, we induced aptamer expression by growth in galactose for 3 hr and fixed the cells by adding formaldehyde (3.7% final concentration) to the culture medium for 90 min. We probed samples with a polyclonal rabbit anti-LexA antibody (UBI, New York), followed by affinity purified Texas Red- or fluorescein-conjugated secondary antibodies (Jackson ImmunoResearch Laboratories, PA), essentially as described (18).

Generation and identification of a higher affinity anti-Cdk2 aptamer

We mutagenized an existing aptamer to generate a variant with a higher affinity for Cdk2. The starting molecule, aptamer 10, has a measured Kd of 1.05 X 10^{-7}M (1). We performed a PCR-based random mutagenesis on the variable region of aptamer 10, and we reintroduced the PCR products into the library vector, pJM-1. This vector directs the expression of aptamers fused to the SV40 nuclear localization sequence, the B112 activation domain, and the HA epitope tag under the control of P_{GAL1}, a galactose-inducible yeast promoter (1). We created a pool of 15,000 mutants. To obtain a measure of the efficiency of the PCR mutagenesis, we sequenced the variable regions of two clones from this pool; sequencing revealed that these carried 3 and 4 mutations that resulted in 1 and 3 amino acid changes respectively. To identify tighter-binding variants from this pool, we took advantage of the existence of different LexA operator reporter genes with different sensitivities. We began with a strain that expressed the LexA-Cdk2 bait and that carried pRB1840, a relatively insensitive lacZ reporter (10) that contains a single synthetic LexA operator. This reporter gene was not activated by aptamer 10 (Fig. 1*a*). We transformed into this strain the pool of PCR mutagenized plasmids, and we

screened the transformants for those that gave blue colonies. The sequence of the variable regions of the 7 plasmids rescued after the screening revealed that all carried identical nucleotide changes that caused two amino acid substitutions: Leu to Ser at residue 5 of the variable region and Asn to Gly at residue 19 (Fig. 1*d*). We termed this mutant aptamer 10M.

We compared the affinity of aptamers 10M and 10 for Cdk2 by interaction mating. The three reporters, pSH18-34, pJK103, and pRB1840, contain respectively 8 high affinity operators, 2 high affinity operators, and one lower affinity LexA operator (Fig. 1*a*). While, as judged by blue color from the pSH18-34 reporter, aptamer 10M had only a slightly greater affinity than aptamer 10, blue color from pJK103 and pRB1840 clearly indicated that 10M bound LexA-Cdk2 more strongly. To verify that the apparent higher affinity was not due to an increased expression level and/or stability of aptamer 10M, we performed Western blotting experiments and showed that the steady-state levels of aptamer 10, 10M, and a control aptamer were identical (Fig. 1*b*). Aptamer 10M did not interact with proteins unrelated to the Cdk family (data not shown).

We measured the gain in affinity by two different means. First,we used a LexAop-GFP reporter gene to quantify the interactions between the anti-Cdk2 aptamers and LexA-Cdk2.

Mean fluorescence obtained from each interaction was plotted against the Kds measured in evanescent wave experiments (1), and the plot was shown to follow a logarithmic equation (Table 1, Fig.1c). We then used this equation to calculate the Kd of the interaction between LexA-Cdk2 and aptamer 10M from the fluorescence it conferred (Table 1). We also performed evanescent wave experiments with purified aptamer 10M and His6-Cdk2 (9). The measured Kd was 5nM (data not shown), quite close to the 2nM Kd calculated from the GFP data.

To evaluate the respective contributions of the two amino acid changes in the variable region to the increased affinity, we mutated both residues individually to wild type and analyzed the single mutants by interaction mating (data not shown). These experiments showed that the mutation of the 5th residue in the variable region (Leu to Ser) contributed to the gain of affinity, but that the mutation in the 19th residue (Asn to Gly) did not.

We then sought to determine whether the gain in affinity in 10M was caused by changes in its contact(s) with Cdk2. To this end, we analyzed the binding of all existing anti-Cdk2 aptamers, including 10M, to a collection of Cdk2 mutants and related members of this protein family. Figure 2 shows that, as previously observed, some of these aptamers recognize different epitopes conserved among Cdk proteins (1).

However, by contrast with the aptamers isolated in our previous work, aptamer 10M showed distinct crossreactivity to other Cdk proteins. This interaction with Cdk family members is consistent with three ideas. First, the Leu to Ser change might create contact(s) between the variable region and residue(s) conserved among the Cdk proteins tested. Second, the change might indirectly create such contacts by changing the conformation of the variable region. Third, even though Leu is more dihedrally constrained than Ser, the change to Ser might create new contacts within the variable region which overall reduces the conformational entropy of this loop.

Targeted intracellular protein modifiers

We used aptamer 10M and others to build protein derivatives that ubiquitinated target proteins *in vivo*. We based our design on the structural organization of the *hect* domain-containing ubiquitin ligases, which conjugate ubiquitin received from a E2 enzyme and transfer it to a protein substrate (19). The amino-terminal substrate-recognizing region of these enzymes varies greatly, whereas the carboxy-terminal region, the *hect* domain, which carries the catalytic activity, is conserved between family members and throughout evolution (20). This modular structure suggested that we could fuse a *hect* domain to peptide

aptamers and create ubiquitin ligases with engineered specificities (Fig.3*a*).

Accordingly, we used a hect domain native to the yeast protein Rsp5 to construct various aptamer-*hect* fusions. We expressed these in yeast together with their putative LexA-Cdk2 targets, and examined the fates of these fusion proteins. Figure 3*b* shows that while the control TrxA-*hect* fusion that lacked a variable region was detectable by Western analysis, none of the anti-Cdk2 aptamer-*hect* fusions could be detected. This result indicates that attachment of a TrxA aptamer to a *hect* domain destabilizes the aptamer. However, expression of anti-Cdk2 aptamer-*hect* fusions resulted in the appearance of a ladder of higher molecular weight forms of LexA-Cdk2, suggesting that these chimeric proteins, even expressed at low levels, still directed *hect*-mediated ubiquitination of the target. The ladder of higher molecular weight LexA-Cdk2 forms was most apparent when the 10M-*hect* fusion was expressed, suggesting that the affinity of the modifier for the target affected the degree of target modification. Modification did not result in destruction: as determined from Western (Fig.3*b*) and pulse chase experiments (data not shown), expression of these modifiers had no effect on LexA-Cdk2 stability or half life.

To confirm that this ladder of higher molecular weights corresponded to LexA-Cdk2-ubiquitin conjugates, we performed

two different experiments. First, we used a vector that directed the conditional expression of Myc-tagged ubiquitin, placed under the control of a copper-inducible promoter (12), and repeated the above experiments. Upon CuSO4 addition in the culture medium, the ladder of higher molecular weights showed an upper shift as compared to the ladder observed without inducing expression of Myc-ubiquitin (Fig.3*c*). It has already been shown that the apparent molecular weight of Myc-ubiquitin protein conjugates is higher than that of ubiquitin protein conjugates (12). Second, we generated a loss of function anti-Cdk2 aptamer-*hect* fusion by changing the cysteine residue that forms the thioester bond with ubiquitin to alanine (20). In cells that expressed the mutant fusion, we did not observe the LexA-Cdk2 ladder, but now did observe, for the first time, the enzymatically dead aptamer-ubiquitin ligase derivative (Fig. 3*b*). This fact suggests that the lack of detectable *hect*-linked aptamer might be due to its proteolysis after self-ubiquitination. This idea is supported by two lines of evidence. First, ubiquitination occurs on lysine side-chains (21), and, by contrast with aptamer 5-, 8-, 10-, and 10M-*hect* fusions, aptamer 2- and 11-*hect* fusions, which contain lysines in their variable regions, did not ubiquitinate LexA-Cdk2 (not shown). Since these aptamers bind Cdk2 as tightly as the others (Fig. 2), this fact is consistent with the idea that ubiquitination of their variable regions blocks their binding. Second, variants of the aptamer-*hect* fusions in which we changed different

combinations of the 5 solvent exposed lysine residues (K19, 37, 53, 70, 83) (22) on TrxA to arginine did result in proteins whose intracellular expression was detectable by anti-TrxA antiserum (data not shown). These observations suggest that the aptamer-*hect* fusions are vulnerable to ubiquitination by their own *hect* moieties and subsequently proteolyzed, but are able to ubiquitinate their targets even though expressed at very low steady states.

Finally, we tested whether we could sensitize LexA-Cdk2 to ubiquitin-dependent proteolysis by introducing into it extra lysines as ubiquitin acceptors. To this end, we constructed and expressed in yeast a LexA-Cdk2 derivative that carried 7 lysines between the LexA and the Cdk2 moieties. We coexpressed this LexA-Lys7-Cdk2 putative target together with TrxA-*hect*, 8-*hect*, and 10M-*hect* modifiers and visualized it with anti-LexA antibody as above. Fig. 3*d* shows that, as compared with the "native" LexA-Cdk2 target, the additional lysines on the experimental target improved its ubiquitination, both increasing the amount of higher molecular weight LexA-Cdk2 derivatives, and causing the appearance of at least one new still higher molecular weight conjugate. However, as judged by its steady-state level, the increased ubiquitination of this LexA-Lys7-Cdk2 target did not destabilize it (Fig. 3*d*).

Targeted intracellular protein transporters

We then investigated the possibility of using derivatized peptide aptamers to change the localization of their protein targets *in vivo*. In the original library vector, aptamers are expressed in yeast fused to an SV40 nuclear localization sequence (NLS). LexA-fusion proteins that lack nuclear localization signals are uniformly distributed within the yeast cell (23, this study). We tested whether anti-Cdk2 aptamers addressed to the nucleus would concentrate also LexA-Cdk2 in the nucleus. As shown by immunofluorescence experiments, LexA-Cdk2 is uniformly distributed in cells in which the chimeras B112-NLS-TrxA and B112-NLS-aptamer are not expressed. Similarly, LexA-Cdk2 is evenly distributed inside cells in which the control chimera B112-NLS-TrxA is expressed. However, in cells in which B112-NLS-anti-Cdk2 aptamers are expressed, LexA-Cdk2 is concentrated in the nucleus (Fig. 4*a*). All the tested fusions triggered substantial nuclear localization of their LexA-Cdk2 target.

To confirm that the concentration of LexA-Cdk2 in the nucleus was due to the nuclear translocation of peptide aptamers, rather than any other aspect of the binding of aptamers to their target, we expressed LexA-Cdk2 together with either aptamers or NLS-aptamers and determined the percentage of yeast in which LexA-Cdk2 was clearly nuclear. The results show that the nuclear localization of LexA-Cdk2 depends on the expression of peptide aptamers that contain nuclear localization sequences (Fig. 4*b*). Finally, we used

NLS-containing aptamers made against another protein, Ste5 (6) to determine whether these could cause the nuclear localization of a protein that is thought to be to be predominantly cytoplasmic (24). When no aptamer was expressed or when a non-nuclear localized aptamer was expressed, LexA-Ste5 showed a predominant cytoplasmic localization. However, when NLS-aptamers were expressed, LexA-Ste5 showed a distinct concentration in the nucleus (Fig. 4*c,d*)

Discussion

We have described peptide aptamer derivatives that covalently modify and change the localization of target proteins *in vivo*. To make them, we first generated a higher-affinity mutant anti-Cdk2 aptamer by mutagenizing the variable region of an existing aptamer and screening for tighter binding mutants using relatively insensitive two-hybrid reporter genes. This variant had a significant increase in affinity and exhibited a Kd comprised between 2 and 5nM. We imagine that use of still-less sensitive reporter genes (25) in these schemes or of LexA mutants with a decreased affinity for their operators (26) and perhaps substitution of weaker activation domains on the aptamer library, should allow us to identify mutant aptamers with sub-nanomolar affinity in one step.

To construct modifiers, proteins that modify target proteins *in vivo*, we exploited ubiquitination, the covalent coupling of ubiquitin molecules to lysine residues on proteins. We showed that these modifiers ubiquitinated the LexA-Cdk2 target. Moreover, our results indicate that the modifiers destroyed themselves, by self-ubiquitination on lysine residues exposed at the surface of the thioredoxin platform. This observation is consistent with the fact that the native Rsp5 protein ubiquitinates itself *in vitro*,

probably on lysine residues lying in the amino-terminal part of the protein, outside of its *hect* domain (20,27).

To our knowledge, our work provides the first *in vivo* demonstration of targeted protein modifications by enzymes of redirected specificity. However, these aptamer-*hect* fusions did not destroy their Cdk2 targets, even those that contained extra lysine residues. It is possible that, for Cdk2, ubiquitination mediated by a more active effector domain would result in destabilization. Consistent with this idea, Gosink et al. redirected the specificity of two plant E2 ubiquitin-conjugating enzymes, Ubc1 and Ubc4, *in vitro* by fusing different protein-binding peptides, including the Ig binding domains of *Staphylococcus aureus* A protein, to Ubc carboxy-termini (28). *In vitro*, these authors observed ubiquitination of the cognate substrates, and a partial degradation of the targeted IgG. However, we believe the most likely explanation for the stability of ubiquitinated LexA-Cdk2 fusions is that Cdk2 is simply refractory to ubiquitin-mediated proteolysis. In fact, although some of its binding partners are degraded by this means (29) no Cdk is known to undergo ubiquitin-dependant proteolysis. Moreover, a number of other proteins are also ubiquitinated without being degraded, including H2A (21), cyclins in certain cell cycle phases (30) and some membrane receptors whose ubiquitination signals endocytosis without involving the proteasome (31).

We also used peptide aptamers fused to a nuclear localization sequence as "transporters". We found that anti-Cdk2 and anti-Ste5 aptamers that carried nuclear localization signals caused their targets to accumulate in the nucleus. Our results suggest that we should be able to build transporters that readdress their protein targets to other subcellular compartments: the endoplasmic reticulum (32), the mitochondrial membrane (33) or the plasma membrane (34). Moreover, very recently, Schneider et al. have shown that chimeric proteins consisting of PDZ domains selected to bind different target peptides, fused to nuclear localization sequences, direct the peptides they target to the nucleus of mammalian 283T cells (35). Because transporters should allow mislocalization of targeted proteins that are expressed under the control of their own promoters, the perturbations they induce in cell function should be less severe than those resulting from ectopic overexpression of target proteins fused to various addressing sequences.

The construction of new proteins by assembling functional modules taken from other ones has been reported for many types of proteins (25,36), and in fact the success of these methods can be taken to support the current picture in which exons are shuffled among proteins during evolution (37). We imagine that peptide aptamers can function as general purpose recognition moieties in intracellular chimeras with other functions that those described here.

From this work, the ability to control the spatial and temporal expression of such "modifiers" and "transporters" in cells and whole organisms should facilitate a finer control of protein modification, inactivation, and localization. Furthermore, the ability to select aptamers that distinguish among allelic variants of proteins should allow selective modification of the activities of individual alleles. The design and use of peptide aptamer derivatives may take part in an emerging intracellular nanotechnology that should permit a high-resolution study of regulatory pathways, and could possibly inspire new therapeutic strategies

Acknowledgments

We are grateful to Sandrine Mouradian for the flow cytometry experiments, to Ron Geyer and Alejandro Colman-Lerner for the anti-Ste5 aptamers, to Fred Winston for RSP5, to John McCoy for rabbit anti-TrxA antiserum, to Rosine Haguenauer-Tsapis for Yep105, to Stan Tabor for advice about TrxA mutants, to Mark Stahl for advice with the evanescent wave experiments. We thank Ron Geyer, Alejandro Colman-Lerner and Jeffrey C. Way for comments on the manuscript. These experiments were supported by a grant to RB from the National Institute of General Medical Sciences, to PAS from the Department of Defense Breast Cancer Research program and by grants to PC from the Association pour la Recherche sur le

Cancer, the Ligue Nationale contre le Cancer, and the Fondation pour la Recherche Médicale.

Figure Legends

Table 1. GFP interaction phenotypes and Kds.
Average green fluorescence of yeast above background was measured for interactions between LexA-Cdk2 and aptamers 2,3,5,8,10 and 11 in 4 independent experiments. Previously measured Kds (1) were plotted against measured fluorescence (Fig.1*c*) and the plot was shown to follow the following equation : Kd = 10exp-(fluo+123.7 / 21.9). We calculated the Kd for aptamer 10M by interpolation using this equation.

Fig. 1. A peptide aptamer with an enhanced affinity for its target. (*a*) Interaction mating assay between LexA-Cdk2 and aptamer 10, two strains carrying aptamer 10M, and a non-interacting aptamer C4, using three different sensitivity LexAop-lacZ reporters.
(*b*) Western blot assay using an anti-TrxA antibody. Diploid exconjugates were grown in galactose containing medium, and proteins were extracted and subjected to a Western blot analysis with anti-TrxA antibody. (*c*) Strength of interaction phenotypes as determined by fluorescence from a LexAop-GFP reporter plotted against Kds measured in evanescent wave experiments. Fluorescence values are in arbitrary units (a.u.).
(*d*) Sequence of the variable regions of aptamers 10 and 10M.

Fig. 2. Mapping of the Cdk2 binding sites of the original aptamers and of aptamer 10M. We collected the Cdk2 mutant bait proteins, described elsewhere (3), Cdk3, and *Drosophila* Cdc2 and Cdc2c (13). In this experiment, we expressed TrxA, the 14 different aptamers originally selected (1), and aptamer 10M as preys. We mated yeast to generate an interaction matrix (13).

Fig. 3. Targeted ubiquitination of LexA-Cdk2 by aptamer-*hect* fusions. (*a*) Design of a "modifier", inspired by the structure of *hect* domain containing ubiquitin ligases. (*b*) Western blot analysis of LexA-Cdk2 (upper panel) and TrxA-*hect* or aptamer-*hect* fusions (lower panel) using anti-LexA and anti-TrxA antibodies respectively. (*c*) Western blot analysis of LexA-Cdk2 when aptamer-hect fusions are expressed by growth overnight in a medium that does or does not contain CuSO4 and that does or does not express Myc-tagged ubiquitin. (*d*) Western blot analysis of LexA-Cdk2 and LexA-7Lys-Cdk2 when TrxA-, 8- and 10M-*hect* fusions are expressed, using the anti-LexA antibody.

Fig. 4. Nuclear translocation of LexA-Cdk2 and -Ste5 by interacting aptamers addressed to the nucleus. (*a*) Yeast microphotographs. Left panels: Indirect immunofluorescence using the anti-LexA antibody. Right panels: DNA staining with DAPI. GAL/TrxA: TrxA is expressed. GAL/14: aptamer 14 is expressed. GLU/TrxA: aptamer 14 is not expressed.

(*b*) Percentage of yeast that displayed clear nuclear immunofluorescence, in presence of aptamers addressed to the nucleus or not. Dark bars: nuclear and cytoplasmic staining. White bars: Nuclear staining. At least 50 cells were observed for each assay

(c) Yeast microphotographs. Left panels: Indirect immunofluorescence using anti-LexA antibody. Center panels: DNA staining with DAPI. Right pannels; Yeast observed with Nomarski optics. GAL/C1: non-NLS aptamer C1 is expressed. GLU/C1: non-NLS aptamer C1 is not expressed. GAL/N1: NLS-aptamer N1 is expressed. GLU/N1: NLS-aptamer N1 is not expressed. (d) Percentage of yeast that displayed cytoplasmic or nuclear and cytoplasmic staining, in presence of various aptamers addressed to the nucleus or not. Dark bars: cytoplasmic staining. White bars: cytoplasmic and nuclear staining. At least 100 cells were observed for each assay.

Aptamer	Kd (nM)	Mean fluorescence (arbitrary units)	Standard deviation
8	38	40.8	4.3
5	52	36.8	3.7
2	64	30.9	3.0
11	87	33.0	1.4
10	105	29.0	2.8
3	112	30.3	1.7
10M	*2*	*67.5*	*3.8*

Table I

References

1. Colas, P., Cohen, B., Jessen, T., Grishina, I., McCoy, J. & Brent, R. (1996) *Nature* **380,** 548-550.
2. Xu, C. W., Mendelsohn, A. & Brent, R. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 12473-12478.
3. Cohen, B. A., Colas, P. & Brent, R. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 14272-14277.
4. Fabrizzio, E., Le Cam, J., Polanowska, J., Kakzorek, M., Lamb, N., Brent, R. & Sardet, C. (1999) *Oncogene* **18,** 4357-4363.
5. Kolonin, M. G. & Finley, R. L., Jr. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 14266-14271.
6. Geyer, C.R., Colman-Lerner, A. & Brent, R. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 8567-8572.
7. Colas, P. (2000) *Curr. Opin. Chem. Biol.* **4**, 54-59.
8. Cadwell, R. C. & Joyce, G. F. (1992) *PCR Methods Appl.* **2,** 28-33.
9. Gietz, D., St. Jean, A., Woods, R. A. & Schiestl, R. H. (1992) *Nuc. Acids Res.* **20,** 1425.
10. Estojak, J., Brent, R. & Golemis, E. A. (1995) *Mol. Cell. Biol.* **15,** 5820-5829.
11. Finley, R. L., Jr. & Brent, R. (1995) in *DNA Cloning, Expression Systems: A Practical Approach*, eds. Hames, B. D. & Glover, D. M. (Oxford University Press, Oxford), pp.169-203.

12. Ellison, M.J. & Hochstrasser, M. (1991) *J. Biol. Chem.* **266**, 21150-21157.

13. Finley, R. L., Jr. & Brent, R. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 2980-12984.

14. Golemis, E. A. & Brent, R. (1992) *Mol. Cell. Biol.* **12,** 3006-3014.

15. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D., Seidman, J. G. & Struhl, K. (1987-1997) (Greene and Wiley-interscience, New York).

16. Brent, R. & Ptashne, M. (1984) *Nature* **312,** 612-615.

17. LaVallie, E. R., DiBlasio, E. A., Kovacic, S., Grant, K. L., Schendel, P. F. & McCoy, J. M. (1993) *Biotechnology (N Y)* **11,** 187-193.

18. Ferrigno, P. & Silver, P. A. (1999) *Methods Cell. Biol.* **58,** 107-122.

19. Scheffner, M., Nuber, U. & Huibregtse, J. M. (1995) *Nature* **373,** 81-83.

20. Huibregtse, J. M., Scheffner, M., Beaudenon, S. & Howley, P. M. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 2563-2567.

21. Jennissen, H. P. (1995) *Eur. J. Biochem.* **231,** 1-30.

22. Katti, S. K., LeMaster, D. M. & Eklund, H. (1990) *J. Mol. Biol.* **212,** 167-184.

23. Silver, P. A., Brent, R. & Ptashne, M. (1986) *Mol. Cell. Biol.* **6,** 4763-4766.

24. Mahanty, S.K., Wang, Y., Farley, F.W., Elion, E.A. (1999) *Cell* **98**, 501-512.

25. Brent, R. & Ptashne, M. (1985) *Cell* **43,** 729-736.

26. Oertel-Buchheit, P., Lamerichs, R.M., Schnarr, M. & Granger-Schnarr, M. (1990) *Mol. Gen. Genet.* **1**, 40-48.

27. Wang, G., Yang, J. & Huibregtse, J. M. (1999) *Mol. Cell. Biol.* **19,** 342-352.

28. Gosink, M. M. & Vierstra, R. D. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 9117-9121.

29. Hershko, A. (1997) *Curr. Opin. Cell. Biol.* **9,** 788-799.

30. Mahaffey, D. T., Yoo, Y. & Rechsteiner, M. (1995) *FEBS Lett.* **370,** 109-112.

31. Terrell, J., Shih, S., Dunn, R. & Hicke, L. (1998) *Mol. Cell* **1,** 193-202.

32. Pelham, H. R. (1996) *Cell. Struct. Funct.* **21,** 413-419.

33. Haucke, V. & Lithgow, T. (1997) *J. Bioenerg. Biomembr.* **29,**11-17.

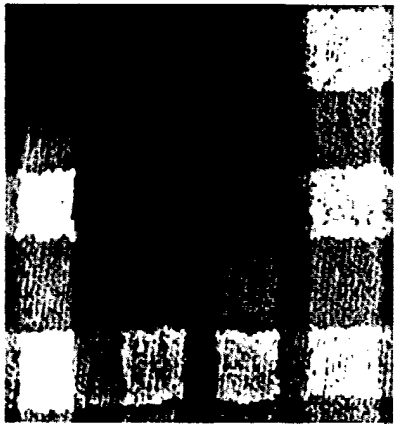
34. Boutin, J. A. (1997) *Cell Signal.* **9,** 15-35.

35. Schneider, S., Buchert, M., Georgiev, O., Catimel, B., Halford, M., Stacker, S.A., Baechi, T., Moelling, K. & Hovens, C.M. (1999) *Nature Biotech.* **17**, 170-175.

36. Vita, C. (1997) *Curr. Opin. Biotechnol.* **8,** 429-434.

37. Gilbert, W., de Souza, S. J. & Long, M. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 7698-7703.

A 10 10M 10M C4



8 ops LacZ

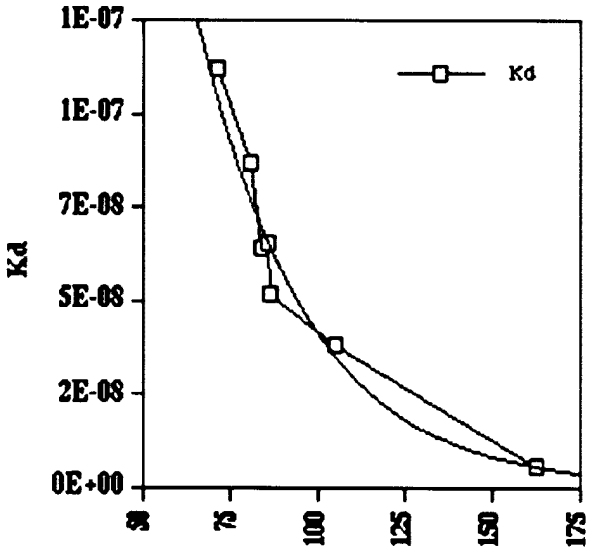
2 ops LacZ

1 op LacZ

B

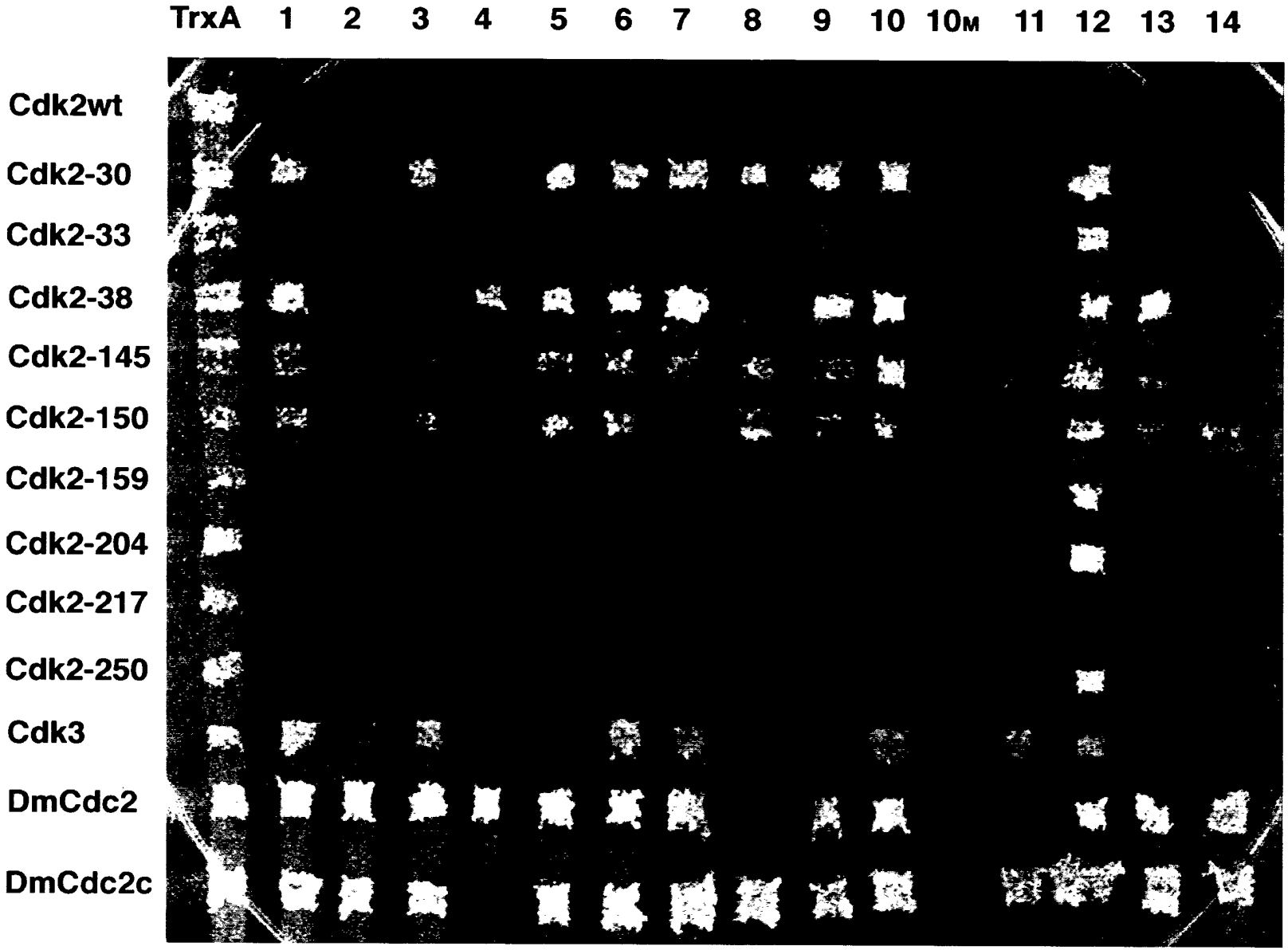
10 10M 10M C4

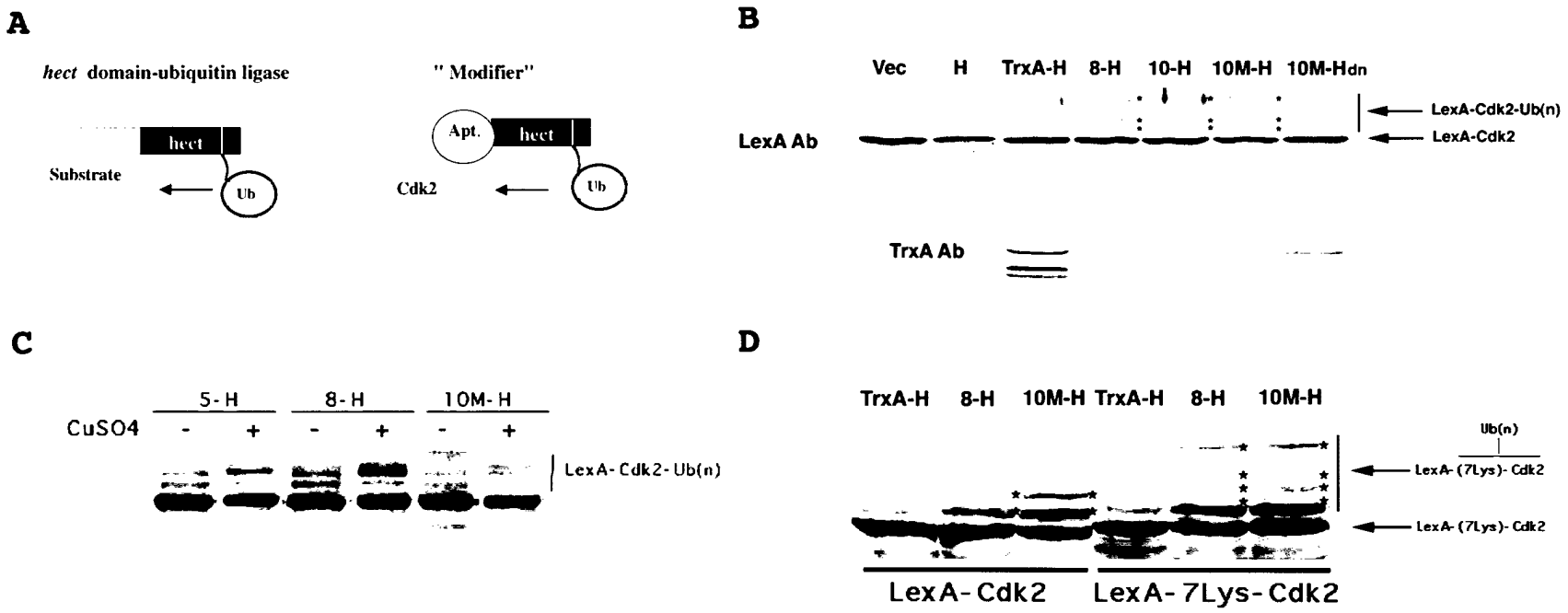
C

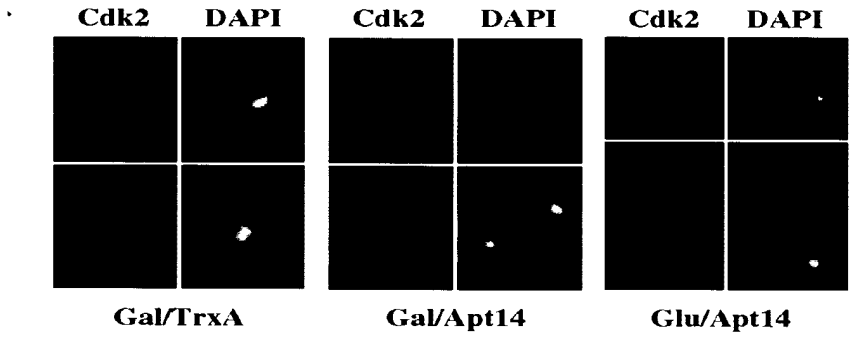


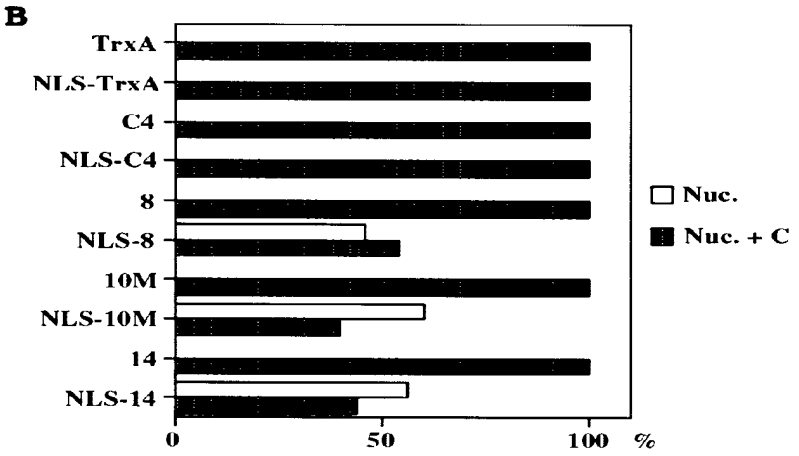
D

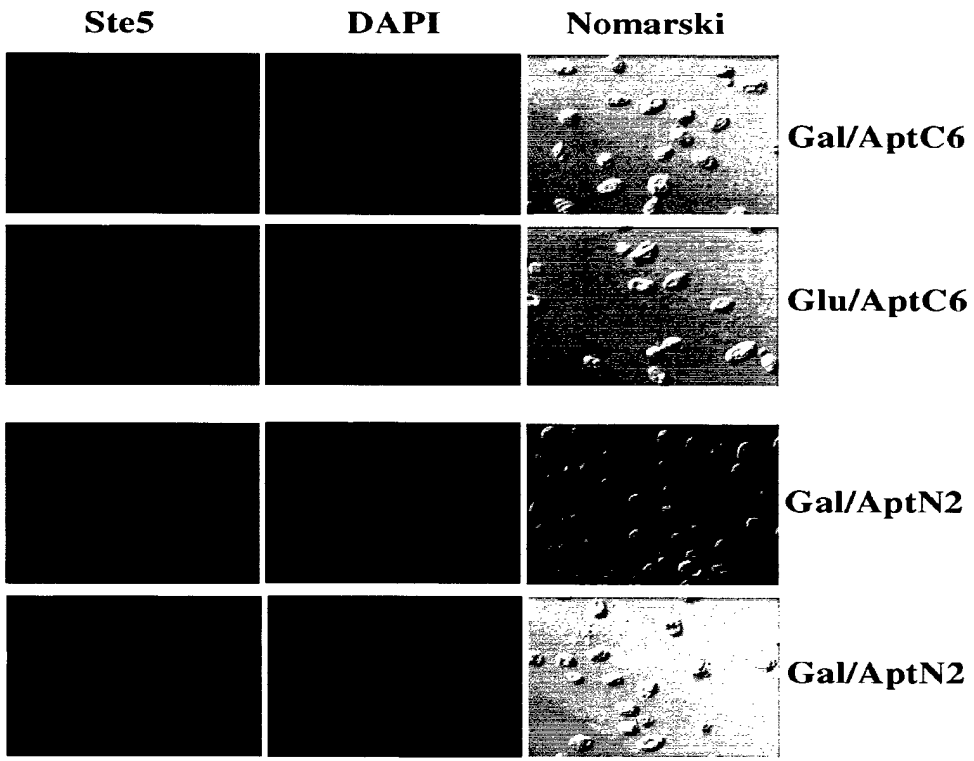
```
10 :  QVWSLWALGWRWLRRYGWNM
      :::: :::::::::::::: :
10M : QVWSSWALGWRWLRRYGWGM
```

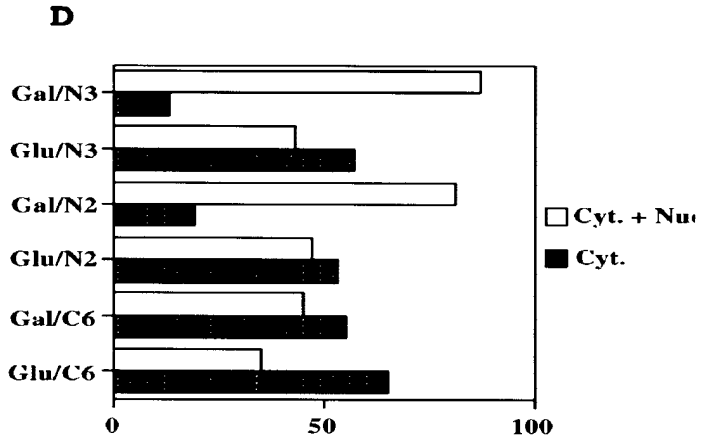

TrxA
1
2
3
4
5
6
7
8
9
10
10M
11
12
13
14
Cdk2wt
Cdk2-30
Cdk2-33
Cdk2-38
Cdk2-145
Cdk2-150
Cdk2-159
Cdk2-204
Cdk2-217
Cdk2-250
Cdk3
DmCdc2
DmCdc2c


A
hect domain-ubiquitin ligase
" Modifier"
hect
Substrate
Ub
Apt.
hect
Cdk2
Ub
B
Vec
H
TrxA-H
8-H
10-H
10M-H
10M-Hdn
LexA Ab
LexA-Cdk2-Ub(n)
LexA-Cdk2
TrxA Ab
C
5-H
8-H
10M-H
CuSO4
-
+
-
+
-
+
LexA-Cdk2-Ub(n)
D
TrxA-H
8-H
10M-H
TrxA-H
8-H
10M-H
Ub(n)
LexA-(7Lys)-Cdk2
LexA-(7Lys)-Cdk2
LexA-Cdk2
LexA-7Lys-Cdk2


Cdk2
DAPI
Cdk2
DAPI
Cdk2
DAPI
Gal/TrxA
Gal/Apt14
Glu/Apt14


B
TrxA
NLS-TrxA
C4
NLS-C4
8
NLS-8
10M
NLS-10M
14
NLS-14
Nuc.
Nuc. + C
0
50
100
%


Ste5
DAPI
Nomarski
Gal/AptC6
Glu/AptC6
Gal/AptN2
Gal/AptN2


D
Gal/N3
Glu/N3
Gal/N2
Glu/N2
Gal/C6
Glu/C6
Cyt. + Nu
Cyt.
0
50
100