# BMC Biotechnology

Research article

# A surrogate-based approach for post-genomic partner identification

Renuka C Pillutla[1], Ku-chuan Hsiao[1], Renee Brissette[1], Paul S Eder[2], Tony Giordano[2], Paul W Fletcher[1], Michael Lennick[1], Arthur J Blume[1] and Neil I Goldstein*[1]

Address: [1]DGI BioTechnologies, Inc., Edison NJ, USA and [2]Message Pharmaceuticals, Inc., Malvern, USA

E-mail: Renuka C Pillutla - pillutla@dgibt.com; Ku-chuan Hsiao - hsiaok@dgibt.com; Renee Brissette - reab@dgibt.com; Paul S Eder - eder@messagepharm.com; Tony Giordano - giordano@messagepharm.com; Paul W Fletcher - pwfletcher@dgibt.com; Michael Lennick - mlennick@dgibt.com; Arthur J Blume - ablume@dgibt.com; Neil I Goldstein* - goldstein@dgibt.com
*Corresponding author

## Abstract

**Background:** Modern drug discovery is concerned with identification and validation of novel protein targets from among the 30,000 genes or more postulated to be present in the human genome. While protein-protein interactions may be central to many disease indications, it has been difficult to identify new chemical entities capable of regulating these interactions as either agonists or antagonists.

**Results:** In this paper, we show that peptide complements (or surrogates) derived from highly diverse random phage display libraries can be used for the identification of the expected natural biological partners for protein and non-protein targets. Our examples include surrogates isolated against both an extracellular secreted protein (TNFβ) and intracellular disease related mRNAs. In each case, surrogates binding to these targets were obtained and found to contain partner information embedded in their amino acid sequences. Furthermore, this information was able to identify the correct biological partners from large human genome databases by rapid and integrated computer based searches.

**Conclusions:** Modified versions of these surrogates should provide agents capable of modifying the activity of these targets and enable one to study their involvement in specific biological processes as a means of target validation for downstream drug discovery.

## Background

Modern drug discovery is concerned with identification and validation of novel protein targets from among the >30,000 genes postulated to be present in the human genome [1]. In understanding the importance of any new gene and its connection to a given phenotype, there is the need to know the immediate "neighborhood" of partners for each gene product since they are most likely involved in the action of the gene product. In this regard, there are few if any new chemical entities (NCEs) capable of regulating protein:protein interactions as either agonists or antagonists. In the past, peptides have sometimes been

used to obtain information about protein:protein interactions as well as regulate their activity [2,3]. This has most often been accomplished with libraries consisting of peptides between <15 amino acids in length. Using this approach, peptides have been identified which act as agonists and antagonists though, in most cases, these peptides have not shown any sequence homology to the natural ligand [4–6]. Clearly these peptides did not use any of the natural amino acid contacts required for binding of the growth factors to their receptors. As these peptides were functional but not compositional mimics, they lacked primary sequence information useful for identifying (by motif, sequence identity or similarity) the true biological partner. These results are not surprising since the putative contact domains between receptors and hormones are expected to be conformational and short peptides were probably unable to mimic a large three-dimensional shape. In this report, we describe a novel post-genomic approach (called Phenogenix®) that involves the use of complex and random peptide libraries of large size (up to 40 amino acids) and diversity (>$10^{11}$ independent clones per library) in combination with computational analysis for garnering information on the natural biological partners and pathways [7]. We also show that the specificity of these peptides can be improved by mutation at the DNA level that has implications for phenotyping and the development of Site Directed Assays. Overall, the data indicates that the surrogate peptides, derived in this manner, can contain sequence information regarding the natural contact domains for both protein:protein and protein:non-protein interactions.

## Results and Discussion
### Criteria for a Partner Hit
The first stage in the computational analysis of our surrogates requires the alignment of the peptides into groups based on motifs or consensus regions. In addition, we examine each peptide for significant differences in the expected frequency of amino acids and the number of times a specific peptide sequence has been repeated. After defining our query strategy (e.g., entire surrogate sequence, motifs, etc.), we simultaneously search several public databases using programs such as Blastp, MAST (Motif Alignment and Search Tool) and Patternfind (see Materials and Methods). The output from each search is further analyzed based on criteria described in Table 1. Homology between the partner and surrogate oftentimes ranges over a long stretch (15–20 amino acids) or may be found in a perfect match within a short sequence of 5–8 amino acids (unpublished data). Other positive indicators include: 1. the appearance of the partner in at least 50% of the top cohort (i.e., first 10 matches) of any one search; 2. the appearance of the same or related hits occurring in several different searches; 3. the identification

of the same partner for multiple peptides from the same or related pans. Criterion 2 addresses the biological relevance of a hit (e.g., distribution, disease indication, etc.) and criterion 3 relates to the biological activity of the surrogate and its ability to cause a phenotypic change in the appropriate test system (phenotyping).

**Table 1: Criteria for a Partner Hit**

| |
|---|
| 1. Search gives an exact match of ≥ 5–7 amino acids or appearance of the partner in at least 50% of the top cohort of any one search, and/or the appearance of the same or related hits occurring in multiple searches. |
| 2. Search matches an expected class of protein partners based on function, cellular location or tissue/disease distribution. |
| 3. Candidate produces a phenotype change when added into the appropriate model system |

### Panning of mRNA targets
In one series of experiments, biotinylated oligonucleotides comprising the UTRs (untranslated regions) of four mRNAs were synthesized (Table 2). The oligonucleotides were heat denatured and allowed to anneal at room temperature to allow the appropriate re-folding. All of the mRNAs were subjected to 4 rounds of panning using both 40 mer and 20 mer random libraries under similar but not equivalent conditions. Individual phage clones from rounds three and four were amplified, tested for binding to the specific and a non-specific mRNA and sequenced. Table 3 shows the overall results that were obtained from each of the pannings. On average, about 8% of the surrogates were found to be specific for each target when compared to a control RNA (RRE). For each RNA target, the predicted amino acid sequences of the peptide binders were analyzed in terms of both overall amino acid content and the occurrence of known RNA-binding motifs and consensus domains. Two motifs were observed for the APP and HCV RNAs (Figure 1 and see below). RNA binding proteins are known to have an overall abundance of certain amino acid residues [8,9]. Table 4 shows a comparison of the specific amino acid composition of peptide binders with regard to their average frequency of occurrence seen within the original unpanned library. All of the peptide binders showed enrichment of arginine residues, as would be expected for RNA binding proteins. Also, tryptophan, serine, and glycine residues were enriched. In addition, several peptide binders showed the presence of the **RGG** box (Fig-

**A**

```
M1-3-B7    RGLFTEWFRGGSWSNYRVTS
M1-3-E8    TDGGRSVISDNVRGGSRLWLWIRHGSWSQAWGPQDAWSSK
M1-3-H6    RVSSAQPGCTSRVRFRCPRGGLLFNGVTSTNPKTGLSNAQ
M1-4-H1    VVYVGVLSYWPHLSGGGRLQVRCLIGRGGFGCRGG
M2-3-C1    WPPGRTLSDLIRGGAGARGM
M2-3-C9    SSGGLHRWSALRGGHGHGLA
M2-3-E2    AMRLKPIAFKGPRAGAGWVEVQPCFAAFRAACTRGGSHHH
M2-3-E3    LHAGWDVTAPRRACKGAQGPGLHGRFYCHRGGLCSGLGRC
M2-3-E9    DEQSSLKGKLRGALVRLGMGHAMPHRGGVWPSTGRPSKQG
M2-3-H12   WTPRHGPMRCWRHQSVFPVGAGPHWALWPIKGPRGGRTAC
M2-NG-C7   RKTGSNIWLPLYHKVCPASTRAGNGRGGSRFLWGSMQTNC
M3-3-B9    RLQRRGGGAVAVVWVGFGVGLLWGRLLLIILGWVLMWFLS
M3-3-C2    QHSEHGGTEWRKRGGMAFAASFLCMRDSYRTTRLRSLLG
M3-3-C7    GTRHVINRVRDSSGVPCKRFGGLQFSQMGKCTIPRGGA
M4-NG-A4   VLRGGSVGKGSLMWCQEVDWRTGGPRSNLWGLWNGRQPPK
```

**B**

```
KH Motif   VIGxxGxxF
M1-3-C6    GVIGGRGLLFPLSGFLHQHR
```

**Figure 1**
Peptide surrogates with RGG Box Sequences. Random peptide libraries were panned on four different mRNA targets. Isolated phage binders from rounds three and four of each pan, were sequenced. Several peptides from each pan showed the presence of the **RGG** box, a well-defined RNA-binding motif [8,9]. **RGG** sequences in each surrogate is in bold and underlined. Peptide surrogate with KH Domain. Panning of the 20-mer random peptide library on target M1 isolated a phage clone containing the sequence **VIG**xx**G**xx**F** which is similar to an RNA-binding motif, the KH motif [8,9]. The surrogate motif corresponding to the KH domain is in bold and underlined.

```
Rev Peptide                TRQARRNRRRRWRERQR

SURROGATES                 SEQUENCE

CONSENSUS:                 (K/R)LRRRP
RRE-NG-C9    ATNFARSQGTQVRGMDVGRLRRRPLEAQGTSGCWWSRVSC
RRE-NG-C2    PAFRQVHWRILAGRDWPIADRFRRQPGGSSSGRSHF
RRE-NG-B5    QSHSLFWHGFTDMRCRWQTGSISKLRRRPAVL-GVSRSPCT
```

**Figure 2**
RRE Binding Motif. Alignment of Rev peptide with surrogate peptides containing the (K/R)LRRRP motif. surrogates were obtained by panning a portion of the RRE mRNA using the 40 mer random peptide library. Consensus motifs are in bold and underlined.

ure 2A) and one sequence was found that contained the **KH** motif (Figure 2B), both of which are known RNA-binding motifs [8,9]. The isolation of surrogates containing generic RNA binding motifs is not unexpected and probably results from enhanced binding and concomitant enrichment of these peptides during the panning process. In addition, an additional consensus motif was

```
APP    VARVRLNRTSPHCGRAPPKAWSFLLVRRQPSPCALVKNPA
HCV    RVLDKRRSAQGRQNAYYIVPWRACRPLIGALSGRPRLGCG
IGF    QYWKARPYQSAPVVRNWFRCQVGVPGRQTGSSRADPGCP
```

**Figure 3**
surrogate Maturation. One clone for each of the three targets – APP, HCV, and IGF was identified for generation of secondary libraries. Residues that were selected for after four rounds of panning are indicated in bold and underlined.

**Table 2: Description of targets used for the mRNA pans**

| MRNA | Length (nucleotides) | Binding Protein |
|---|---|---|
| APP (Amyloid Precursor Protein) | 48 | Unknown |
| HCV (Hepatitis C Virus) | 51 | EIF3 [10–13] |
| IGF-1 (Insulin-like Growth Factor-1) | 61 | Unknown |
| RRE (Rev Response Element | 47 | Rev [10] |

identified among peptides isolated by panning on RRE RNA [10]. This motif [**K/R**] **LRRR,** aligns with a region on the expected natural partner, the Rev peptide (Figure 3).

One peptide was chosen from each of the APP, HCV and IGF pans (based on the highest specific binding vs. RRE RNA) as templates for the construction of secondary libraries. Each of these libraries contained >$10^{10}$ independent clones and was used for panning the appropriate target RNA. Results are shown in Figure 4 and Table 5. Since library construction was based on peptides previously selected for binding and selectivity to specific RNAs, the number of target-specific clones increased dramatically following *in vitro* maturation. In addition, motifs were observed suggesting the critical nature of these specific residues in terms of binding to target. Preliminary studies have shown that the secondary surrogates have higher relative affinities when compared to the original clones (unpublished data).

In the case of the HCV RNA, we panned a sequence (AA UUG CCA GGA CGA CCG GGU CCU UUC UUG GAU CAA CCC GCU CAA UGC CUG GAG AUU) predicted to bind to at least one of the proteins comprising the translation complex eukaryotic Initiation Factor 3 (eIF3; [11–14]). Published reports identify the p120 subunit of eIF3 as the one binding to the apical loop of the domain III of

**Table 3: Results of panning experiments on selected mRNA targets using the primary phage displayed peptide libraries**

| RNA Target | # Picked | # Binders (%) | # Specific Binder (%) | Signal Ratios/SA | Signal Ratios/RNA |
|---|---|---|---|---|---|
| APP | 240 | 186 (78) | 21 (11) | 19.6 | 7.4 |
| HCV | 240 | 207 (86) | 31 (15) | 16.7 | 7.9 |
| IGF-1 | 240 | 184 (77) | 9 (5) | 18.7 | 3.1 |
| RRE | 48 | 25 (42) | 4 (16) | 25.3 | 2.3 |

Individual clones were picked after three to four rounds of panning of random peptide displayed phage libraries for four different RNA targets. Phage rescued from these clones were tested for binding to the specific RNA target and an irrelevant RNA target. Clones that had a binding ratio of ≥ 2.0 over streptavidin (SA) background were categorized as binders. Binders were then sub-categorized as specific binders if they had a binding ratio of target:irrelevant ≥ 2.0.

**Table 4: Comparison of Amino Acid Frequency in surrogates with expected frequencies.**

|  | Arg | Gly | Trp | Ser |
|---|---|---|---|---|
| Expected Frequency | 9.4 | 6.3 | 3.1 | 9.4 |
| Library | 9.4 | 11.6 | 3.1 | 7.7 |
| M1 | 13.9 | 12.6 | 5.2 | 10 |
| M2 | 13 | 13 | 4.3 | 8.3 |
| M3 | 12.4 | 12.3 | 5.1 | 9.8 |

The expected frequency of each amino acid within the library was calculated based on the probability of occurrence for each codon in the library. These data were compared to the actual frequency of occurrence in the library before and after panning on the various mRNA targets denoted as M1, M2, and M3. All numbers are expressed as a percentage of the expected frequency.

```
eIF3:        EDLDNIQTPE-SVLLSAVSGEDTQDRTDRLLLTPWVKFLWESY

CONSENSUS:                             TxRLL

HCV-NG-D9              TSGESSGDRTRRVLTSSSARTLPN
HCV-3-F5      LLVTGQFP--SQLLLGGAVCGP--STPRLRTGLCRLSGT
HCV-3-H8      RRTCGDPAAMLERLSCRAGDYRGASHTGRLLNLRGMHQYP
HCV-3-C3                      FTTPRHLSGRTVQMMRDSTS
```

**Figure 4**
HCV-eIF3 Binding Motif. Alignment of eIF3 with surrogate peptides containing the **TxRLL** motif. surrogate peptides were obtained by panning a portion of the 5'UTR of HCV mRNA using both the 20 mer and 40 mer random libraries. Peptides HCV-3-F5 and HCV-3-H8 were obtained from the 40-mer library from the same pan. Peptide HCV-NG-D9 was obtained from the 40 mer library using modified experimental conditions. Peptide HCV-3-C3 was obtained from the 20 mer library. Consensus sequences are in bold and underlined. Sequences outside the motif that are conserved between the surrogates and eIF3 are in Italics and underlined.

HCV 5'UTR, and p170 (also called p160) as binding the stem portion of the domain III. The oligonucleotide used for these studies contains the apical portion and part of the stem portion of domain III and, therefore, might bind one or both of these subunits. Interestingly, the binding of p170 (p160) to the stem of domain III is position-independent (i.e., not constrained by other structural elements on the mRNA [12,13]).

Sequence analysis of surrogate peptide binders to HCV using MEME (Motif Elicitation Program) and other peptide sequence alignment programs identified a consensus sequence **TxRLL** found in four peptides binding to the HCV mRNA (Figure 5). This motif was unique to HCV and not found in peptides derived from any of the other mRNA pans. Interestingly, the **TxRLL**-containing surrogates were found by two investigators in the laboratory, using the 20 mer or 40 mer random libraries under different experimental conditions. Database searches using Patternfind at the ISREC server were performed using parameters appropriate for short protein queries and were successful in identifying a human gene product, subunit p170 of eIF3 (Figure 3). Searches using the con-

**Table 5: Results of panning experiments on RNA targets using secondary libraries**

| Target | Primary Binder (%) | Secondary Binder (%) | Target Specific Binders: Primary (%) | Target Specific Binders: Secondary (%) |
|--------|--------------------|----------------------|--------------------------------------|----------------------------------------|
| APP | 78 | 95 | 9 | 40 |
| HCV | 86 | 98 | 13 | 17 |
| IGF-1 | 77 | 99 | 4 | 18 |

Clones were categorized as specific binders if they had a binding ratio of target:irrelevant ≥ 2.0, as described in Table 2 above. Comparison of percentage specific binders from pannings of primary libraries (expressing random peptides) vs. secondary libraries (based on target-specific primary surrogates) is shown in this Table.
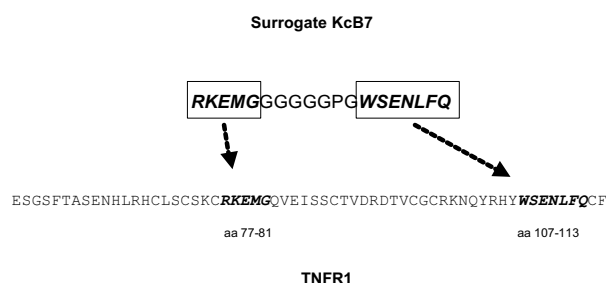
**Surrogate KcB7**



**Figure 5**
Alignment of the TNFβ surrogate peptide KcB7 with its cognate receptor TNFR1 (p55). Peptide KcB7 was isolated by panning TNFβ. The peptide has the sequence **RKEMG-GGGGPGWSENLFQ** and a BLASTp search revealed identity to amino acids 77–81 (**RKEMG**) and amino acids 107–113 (**WSENLFQ**) on TNFR1 (p55). **RKEMG** is involved in binding of TNFR1 to the A subunit of TNFβ while **WSENL** binds to several amino acids in the C subunit of TNFβ

sensus region as the query likewise identified sequence homology with the large subunit p170 of eIF3. These results fit our partner criteria 1–3 (Table 1). Interestingly, other amino acids in these surrogates were found to be identical to residues flanking the **TxRLL** motif in p170 on both the amino and carboxy sides (Figure 5). Thus, the library contained peptide surrogates binding to the HCV target mRNA as well as containing sequence information identifying the natural interacting partner and predicting the putative contact amino acids on the p170 subunit of eIF3. The fact that multiple surrogates had the **TxRLL** motif suggests that this region of p170 is necessary and critical for binding of HCV mRNA to eIF3.

### Panning of TNF-β
In a separate study, we panned the immune cytokine TNF-β (lymphotoxin α, Lt α; [15–17]) using a highly stringent protocol involving both positive (vs. TNF-β)

and negative (vs. TNF-α, TNFR1 and TNFR2) selection. One of these peptides, designated KcB7, had the amino acid sequence **RKEMGGGGGPGWSENLFQ**. A Blastp search, using several different queries revealed TNFR1, which is the natural biological partner of TNFβ (Figure 5). Interestingly, the other cognate partner (i.e., receptor TNFR2) was not identified. Closer examination of the complementary sequences revealed that the short N-terminal sequence **RKEMG** and the C-terminal sequence **WSENLFQ** were identical to regions on TNFR1 (amino acids 77–81 and 107–113 respectively). Although not complete, these segments corresponded to amino acids within two critical ligand:receptor contact domains [18]. In the case of the N-terminal grouping, the surrogate contained 5 of the 15 amino acids of the 77–81 contact domain whereas in the C-terminal grouping, the surrogate contained 6 of the 9 amino acids identified within the 107–113 contact domain. It is also worth noting that, on the receptor, these two domains are not contiguous but are separated by 25 amino acids whereas only a 6 amino acid hydrophobic stretch separates the two noncontiguous contact-mimicking domains in the surrogate. Therefore, the seven amino acid linker may provide the appropriate molecular distances needed for TNFβ binding to the receptor.

### Robustness of the System
The successful isolation of a useful surrogate may seem an improbable task especially in cases where nothing is known about the hotspot surface(s) on a target. In reality, biopanning using the RAPIDLIB® library has allowed, (in >90% of cases) the identification of partner-specific peptides among the >10^10-11 independent clones that are expressed within this library [[7,19,20] and unpublished data]. More often than not, between 10 and 100 different surrogates have been found for any one target panned and >75% of the targets gave rise to surrogates that bound to regulatory hotspots on the target [[7] and unpublished data]. Several important facts are critical to the success of the process. In the first place, there

must be a high degree of diversity within the library so that one surrogate can be found for each protein target. Additionally, It appears that panning seems to favor the enrichment of regulatory surrogates versus irrelevant peptides. In our experience, the large number of random peptides includes a broad sampling of linear and conformational protein surfaces sufficient to present at least one low affinity "complement' for any target's surface [19,20]. The immune system's antibody repertoire arises through a process [21–24] similar to in some aspects to what occurs during the panning process. Initially, immunoglobulins are synthesized containing multiple small contact domains (i.e., CDR or Complementarity Defining Regions) to provide a "rough" complement of an antigen's surface. Subsequently, antibody/antigen binding is improved via mutational events within the antibody genes to produce a complement with higher affinity and selectivity. Biopanning works in an analogous manner by initially enriching for peptides with low affinity for a specific target. However, surrogates may only need to contain some amino acids identical to those on the partner to bind with sufficient "avidity" rather than affinity.

Other methods that have been used extensively for high throughput partner identification include the yeast two-hybrid system [25]. While the two hybrid approach is popular, it has a number of inherent problems including a high potential for false positives, the inability to use non-protein targets such as mRNA or membrane bound/extracellular proteins and the inability to address post-translational modifications on a target. The generation of surrogates, on the other hands, is target independent and their utility for partner identification resides in the computational analysis of the peptide's sequence. Since our libraries contain totally random peptides ranging from 20 and up to 40 amino acids in length, there are no known restrictions on the amino acids that can be selected to create the surrogate's 'complementary' surface [7]. Thus, the examples described in this report relate to the utility of the surrogate approach for finding the cognate receptor for both protein and non-protein targets. In the case of the surrogates for both HCV-mRNA and TNFβ, it is clear that the large diversity and size of the original library was, in fact, critical to their successful isolation since libraries of <20 amino acids would not have contained either the KcB7 peptide or the HCV-specific surrogates.

In addition to the data presented in this paper, we have screened other targets using this approach. While the expected natural partners were found for many of the proteins, there were instances where surrogates were generated but lacked partner information (e.g., IGF-1R, growth hormone receptor, insulin receptor [[7] and manuscripts in preparation]). There are several possible explanations for these results. While our libraries are large and diverse, it is probable that identification of a surrogate peptide with partner information is a rare event. With that in mind, it may require the isolation and sequencing of large numbers of clones (perhaps >500/target) in order to find the appropriate surrogate for partner identification. On the other hand, some targets may have complex or unusual protein:protein contact sites that preclude generation of a surrogate with partner information. Ongoing experiments will address both of these possibilities. Surrogates have also been found to have the minimal structural content necessary to induce a pharmacological effect on any target in addition to their use in partner identification. Most of our surrogates have been shown to have either agonist or antagonist activity in the appropriate biochemical and/or biological models ([7]manuscripts in preparation). The surrogates were also able to subdivide large contact surfaces into smaller contact domains through which target activity can be modified [7]. These attributes point to surrogate use in phenotyping and validating novel genes whose functions are unknown and for which no known partners exist. Finally, it is worthwhile to note that surrogates have been used to develop competitive Site Directed Assays (SDAs) for essential sub-domains thereby allowing their use in high throughput screening of large combinatorial libraries of small molecules (unpublished data). In our experience, almost all of the peptide surrogates isolated from these complex libraries by our various panning procedures bind to regulatory hot spots on varied targets. This non-random association between a surrogate and a target's pharmacologically active site assures a high degree of probability that, once found, surrogates will have utility for the rapid development of SDAs capable of identifying small molecules of pharmacological importance.

## Conclusions

The results of these experiments support the use of *in vitro* panning procedures with our highly complex and random 20–40 mer peptide libraries as a method to obtain information on large numbers of protein partners and enable the elucidation of biologically relevant protein networks. This post-genomic approach can be automated to increase the number of known and unknown genes and gene products that can be used as targets for partner identification as the first step in the drug discovery process. Additionally, the surrogates isolated in such studies would be prime candidates for phenotyping and target validation through their ability to regulate target activity and for identifying small molecule drug leads through their use in Site Directed Assays. Overall, the surrogates can be seen as providing a discovery continuum by bridging the gap between functional genomics and proteomics and modern drug discovery.

## Materials and Methods

### Targets

Oligonucleotides for the mRNA pans were synthesized by Dharmacon Research (Lafayette, CO) and used for the experiments after heat treatment and re-annealing (65° for 10' and slow cooling at room temperature). The APP and IGF oligo sequences are proprietary to Message Pharmaceuticals. The following sequence was used to obtain the HCV surrogates: 5'-biotin'AA UUG CCA GGA CGA CCG GGU CCU UUC UUG GAU CAA CCC GCU CAA UGC CUG GAG AUU-3'. The sequence for RRE (Rev-response element) has been previously published [10]. Streptavidin coated plates were obtained from Pierce (Rockford, IL). TNFR1, TNFR2, TNF-α and TNF-β were obtained from R&D Systems (Minneapolis, MN) and re-constituted according to the manufacturer's instructions. The *E. coli*, strain TG1 (genotype = K12Δ*(lac-pro), supE, thi, hsdΔ5/F' [traD36, proAB, lacI* q, *lacZΔM15*]) was obtained from Pharmacia (Piscataway NJ).

### Design of the Primary Peptide Libraries

DNA fragments coding for peptides containing 40 random amino acids were generated by a PCR approach using synthetic oligonucleotides as previously described [7,26]. Peptides are expressed on the capsid protein pIII of the phage at low copy number (1–2 peptides/phage).

### Construction of secondary cell libraries

Amino acid mutations were introduced at the oligonucleotide level under controlled conditions [7]. This oligonucleotide was used as the template in a PCR reaction with two shorter 5' biotinylated oligonucleotide primers contributing the restriction sites. The library was then produced essentially as previously described (above). Cell transformants were pooled and an aliquot was plated to determine the total number of transformants. The diversity of the secondary cell libraries >$10^{10}$ independent clones per library.

### Biopanning and ELISA protocols

For the TNF experiments, 6 wells of a 96-well microtiter plate (Maxisorb Immuno plates, Nunc) are coated with target at concentrations ranging from 5–50 ng/well and are incubated overnight at 4°C. At the same time, an aliquot of *E. coli* TG1 cells is inoculated into 2x YT media and grown overnight at 37°C. The next day, unbound antigen is removed and the coated wells are blocked with 300 μl of 2% non-fat milk in PBS (NFM-PBS) for one hour at room temperature. The phage libraries (in NFM-PBS) are thawed and 100 μl of each library is added to the antigen-coated wells and the plates are incubated for 3 hours at room temperature. Each well is washed 13 times with PBS and the phage eluted with 100 μl of 50 mM glycine-HCl containing 0.1% BSA (pH2.2) following a five minute incubation. The eluted phage are pooled, neu-

tralized with 100 μl of 1 M Tris-HCl (pH 8.0). The eluted phage were then panned vs. other members of the TNF family as a way to obtain peptides specific to the original target (Subtractive panning). For example, phage, initially panned vs. TNF-α would be subsequently panned vs. TNF-β, TNFR1 and TNFR2 to remove peptides cross-reacting with the other members. After the final panning, the phage are added to 10 ml of log phase *E coli* TG1 (OD$_{600}$ = 1.0), and amplified in 2x YT-glucose medium for one hour at 37°C. Helper phage (M13K07) and ampicillin are then added and the cells are incubated for an additional hour at 37°C. The cells are pelleted at 3500 RPM for 10 minutes, resuspended in 2x YT-AK medium (YT medium containing ampicillin and kanamycin) and incubated overnight at 37°C. The next day, the infected bacterial cells are centrifuged at 3500 RPM at 4°C for 10 minutes and the pellet discarded. The supernatant contains the phage and is precipitated with $^1/_4$ volume of 20% PEG-8000 in 1.6 M NaCl by incubating on ice for 1 hour. The precipitant is centrifuged at 10,000 RPM at 4°C for 10 minutes and the precipitated phage pellet resupended in about 1 ml of NFM-PBS. The phage is then used for the next round of panning. Normally, 3–4 rounds of panning are done for secondary libraries. Usually, 96 random clones are picked from rounds 3 and 4 and grown in 96 well cluster plates as a master stock.

For the mRNA pans, all solutions and surfaces are pre-treated with DEPC or RNaseZap (Ambion, Austin TX), respectively, to eliminate RNase contamination that may compromise the integrity of the RNA. The biotinylated – RNA target diluted to 1 mg/ml in binding buffer (PBS containing 1 mM MgCl$_2$), denatured at 65°C for 5 min and reannealed by slow cooling to room temperature. Re-annealed mRNAs are stored in small aliquots (10–25 μl/tube) at -20°C. Microtiter wells are treated with RNaseZap (Ambion) before use. One hundred microliters of RNA solution diluted to 2.5 ng/ μl was added to an appropriate number of wells in a 96-well microtiter plate precoated with Streptavidin (Pierce) and incubated for 1 hr at room temperature. Unbound streptavidin is then blocked with 50 μl of 2 mM biotin at room temperature for 1 hr. Panning, with slight modifications, proceeded as described for TNF-β.

For phage rescue prior to ELISA analysis, 40 μl of master stock is transferred from each master to another set of cluster tubes containing 400 μl of 2x YT-AG (ampicillin and glucose) and helper phage (final concentration of 5 × $10^{10}$/ml). The tubes are incubated at 37°C with constant shaking for two hours. The cultures are centrifuged at 2500 × g at 4°C for 20 minutes, the supernatant is discarded, and the bacterial pellet is resuspended in 400 μl of 2x YT-AK (ampicillin and kanamycin) medium and is incubated overnight at 37°C. At that time, the cells are re-

moved by centrifugation at 2500 × g and the supernatants used for ELISAs (see below).

### Enzyme Linked Immunosorbent Assays (ELISA)

For the TNF-β surrogate ELISAs, each well of a MaxiSorp Immuno plate (Nunc) is coated with 50 μl of target (1 μg/ml) overnight at 4°C. In all cases, the wells are blocked with NFM-PBS for 1 hour at room temperature. Phage is added at 100 μl/well and the plates incubated for 2–3 hours at room temperature. After washing 3x with PBS-Tween (DPTC-treated PBS for the mRNA), plates are probed with an anti-M13 antibody conjugated to horseradish peroxidase (1:3000 in PBS-NFM) for 1 hour at room temperature followed by addition of 100 μl of ABTS for 15–30 minutes at room temperature. The OD is measured using a SpectraMax Microplate Spectrophotometer (Molecular Devices) at 405 nm after a 30 minute incubation at room temperature.

The mRNA ELISAs are performed in a similar manner except streptavidin-coated microtiter plates are blocked with PBS containing 2% non fat milk for 1 hr at room temperature, treated with RNaseZap, then coated with biotinylated RNA target (100 ng/well) by incubation for 1 hr at room temperature. Superasin was added to the wells prior to addition of 100 μl/well of phage from isolated clones and incubated at room temperature for 2 hr.

### Steps in computational analysis to identify natural partner

Once a surrogate peptide binder has been identified and shown to bind specifically to its target, it is subjected to partner analysis using several different database search programs. In the initial step, the entire peptide sequence and consensus motifs (if found) are entered into an Advanced BLAST search [http://www.ncbi.nlm.nih.gov/blast/blast.cgi?Jform=1] : using the following parameters:

i. Programs: blastp, tblastn

ii. Databases: protein and nucleotide databases including dbest (ESTs), dsts (STSs) and htgs (unfinished high throughput genomic sequences)

iii. Expect value: 1000 to 20000

iv. Matrix: PAM30 or PAM70

v. Query: Consensus motif alone and varying combinations of sequence at the N- and C-terminal ends

In subsequent steps, motifs identified by sequence alignment programs like MEME (Multiple EM for Motif Elicitation), [http://meme.sdsc.edu/meme/website/intro.html] can also be used to search the available databases using MAST (Motif Alignment and Search Tool, [http://meme.sdsc.edu/mem/website/mast-intro.html] ). Motifs and consensus domains can also be used as query patterns to search the protein databases using Patternfind [http://www.isrec.isb-sib.ch/software/PATFND_form.html] .

For Patternfind, the following parameters are used:

i. Databases: Nonredundant, Swissprot, TREST and TRGEN

ii. Limit: Between 10 and 5000

iii. Query: Consensus motif alone and varying combinations of sequence at N- and C-terminal ends

Data obtained from the various searches are analyzed under the following conditions:

i. Results of different searches should be analyzed independently and then together to look for similar classes of proteins (e.g., Nucleic acid binding proteins, kinases) that may emerge.

ii. Pick some of the best matches that show up in more than one kind of search (e.g. Same protein/ORF picked up by BLAST searches using different parameters, or by both BLAST and Patternfind) and compare sequence of protein in this region with other peptide surrogates containing this motif.

iii. Look at potential significance of protein interaction in the context of the cellular function of target

The criteria for a partner hit are listed in Table 1. Unless there is an exact match (Criterion No. 1), the potential hit has to match at least two of the criteria described in the table to be considered a partner hit.

## References

1. Drews J: **A historical perspective.** *Science* 2000, **297**:1960-1967
2. Scott JK, Smith G: **Searching for peptide ligands with an epitope library.** *Science* 1990, **249**:386-390
3. Gallop MP, Barrett RW, Dower WJ, Fodor SP, Gordon EM: **Applications of combinatorial technologies to drug discovery. I. Background and peptide combinatorial libraries.** *J Med Chem* 1994, **37**:1233-1251
4. Yanofsky SD, Baldwin DN, Butler JH, Holden FR, Jacobs JW, Balasubramanian P, Chinn JP, Cwirla SE, Peters-Bhatt E, Whitehorn EA, Tate EH, Akeson A, Bowlin TL, Dower WJ, Barrett RW: **High affinity type I interleukin 1 receptor antagonists discovered by screening recombinant peptide libraries.** *Proc Natl Acad Sci USA* 1996, **93**:7381-73

5.  Wrighton NC, Farrell FX, Chang R, Kashyap AK, Barbone FP, Mulcahy LS, Johnson DL, Barrett RW, Jolliffe LK, Dower WJ: **Small peptides as potent mimetics of the protein hormone erythropoietin.** *Science* 1996, **273**:458-463

6.  Cwirla SE, Balasubramanian P, Duffin DJ, Wagstrom CR, Gates CM, Singer SC, Davis AM, Tansik RL, Matteakis LC, Boytos CM, Schatz PJ, Baccanari DP, Wrighton NC, Barrett RW, Dower WJ: **Peptide agonist of the thrombopoietin receptor as potent as the natural cytokine.** *Science* 1997, **276**:1696-1699

7.  Blume AJ, Beasley J, Goldstein NI: **The use of peptides in Diogenesis: A novel approach to drug discovery and phenomics.** *Biopolymers* 2000, **55**:347-356

8.  Burd CG, Dreyfuss G: **Conserved structures and diversity of functions of RNA-binding proteins.** *Science* 1994, **265**:615-621

9.  Siomi H, Dreyfuss G: **RNA-binding proteins as regulators of gene expression.** *Curr Opin Genet Dev* 1997, **7**:345-353

10. Zemmel RW, Kelley AC, Karn J, Butler PJG: **Flexible regions of RNA structure facilitate co-operative Rev assembly on the Rev-response element.** *J Mol Biol* 1996, **258**:763-777

11. Buratti E, Tisminetzky S, Zotti M, Baralle FE: **Functional analysis of the interactions between HCV 5' UTR and putative subunits of eukaryotic translation initiation factor eIF3.** *Nuc Acid Res* 1998, **26**:3179-3187

12. Odreman-Macchioli FE, Tisiminetzky SG, Zotti M, Baralle FE, Buratti E: **Influence of correct secondary and tertiary RNA folding on the binding of cellular factors to the HCV IRES.** *Nuc Acids Res* 2000, **28**:875-885

13. Kruger M, Berger C, Li QX, Welch PJ, Trinz R, Leavitt M, Barber JR, Wong-Staal F: **Identification of eIF2Bγ and eIF2γ as cofactors of hepatitis C virus internal ribosome entry site-mediated translation using a functional genomics approach.** *Proc Natl Acad Sci USA* 2000, **97**:8566-8571

14. Sizova DV, Kolupaeva VG, Pestova TV, Shatsky IN, Hellen CUT: **Specific interaction of eukaryotic translation initiation factor 3 with the 5' nontranslated regions of hepatitis C virus and classical swine fever virus RNAs.** *J Virol* 1998, **72**:4775-4782

15. Bazzoni F, Beutler B: **How do tumor necrosis factor receptors work?** *J Inflamm* 1995, **45**:221-238

16. Tartaglia LA, Goeddel DV: **Two TNF receptors.** *Immunol Today* 1992, **13**:151-153

17. Heller RA, Kronke M: **Tumor necrosis factor receptor-mediated signaling pathways.** *J Cell Biol* 1994, **126**:5-9

18. Banner D, D'Arcy A, Janes W, Gentz R, Schoenfeld H-J, Broger C, Loetscher H, Lesslauer W: **Crystal structure of the soluble human 55kd TNF receptor-human TNF-β complex: Implications for TNF receptor activation.** *Cell* 1993, **73**:431-445

19. Hexham MJ, White KD, Carayannopoulos LN, Mandecki W, Brissette R, Yang YS, Capra J: **A human immunoglobulin (Ig) alpha3 domain motif directs polymeric Ig receptor-mediated secretion.** *J Exp Med* 1999, **189**:747-752

20. Ravera MW, Carcamo J, Brissette R, Alam-Moghe A, Dedova O, Cheng W, Hsiao KC, Shen KI, Tang P, Blume A, Mandecki W: **Identification of an allosteric binding site on the transcription factor p53 using a phage-displayed peptide library.** *Oncogene* 1998, **16**:1993-1999

21. Tonegawa S: **Somatic generation of antibody diversity.** *Nature* 1983, **302**:575-581

22. Kocks C, Rajewsky K: **Stepwise intraclonal maturation of antibody affinity through somatic hypermutation.** *Proc Natl Acad Sci* 1988, **85**:8206-8210

23. Wagner SD, Neuberger MS: **Somatic hypermutation of immunoglobulin genes.** *Annu Rev Immunol* 1996, **14**:441-457

24. Weiss U, Rajewsky K: **The repertoire of somatic antibody mutants accumulating in the memory compartment after primary immunization is restricted through affinity maturation and mirrors that expressed in the secondary response.** *J Exp Med* 1990, **172**:1681-1689

25. Cagney G, Uetz P, Fields S: **High-throughput screening for protein-protein interactions using two-hybrid assay.** *Methods Enzymol* 2000, **328**:3-14

26. Carcamo J, Ravera MW, Brissette R, Dedova O, Beasley JR, Alam-Moghe A, Wan C, Blume A, Mandecki W: **Unexpected frameshifts from gene to expressed protein in a phage-displayed peptide library.** *Proc Natl Acad Sci USA* 1998, **95**:11146-11151