FEBS 23899 FEBS Letters 480 (2000) 55–62

Minireview

# Convergent evolution with combinatorial peptides

Brian K. Kay*, Jeremy Kasanov, Stephen Knight, Alexei Kurakin

*Department of Pharmacology, University of Wisconsin-Madison, 1300 University Avenue, Madison, WI 53706-1532, USA*

Received 19 May 2000

Edited by Gianni Cesareni

**Abstract Once the sequence of a genome is in hand, understanding the function of its encoded proteins becomes a task of paramount importance. Much like the biochemists who first outlined different biochemical pathways, many genomic scientists are engaged in determining which proteins interact with which proteins, thereby establishing a protein interaction network. While these interactions have evolved in regard to their specificity, affinity and cellular function over billions of years, it is possible in the laboratory to isolate peptides from combinatorial libraries that bind to the same proteins with similar specificity, affinity and primary structures, which resemble those of the natural interacting proteins. We have termed this phenomenon 'convergent evolution'. In this review, we highlight various examples of convergent evolution that have been uncovered in experiments dissecting protein–protein interactions with combinatorial peptides. Thus, a fruitful approach for mapping protein–protein interactions is to isolate peptide ligands to a target protein and identify candidate interacting proteins in a sequenced genome by computer analysis.** 

*Key words:* Darwinian evolution; EH domain; Estrogen receptor; Molecular evolution; PDZ domain; Phage-display; Protein interaction module; Protein–protein interaction; SH3 domain; WW domain

## 1. Introduction

Natural evolution, as defined by Charles Darwin, consists of three essential phases (Fig. 1). First, a population of organisms is placed in an environment that selects for some individuals over others. Second, those individuals that are favored pass on their 'winning alleles' to progeny. Third, continued rounds of diversification and selection lead to optimized phenotypes and genotypes. Many aspects of this process can be found in experiments in molecular evolution (Fig. 1), in which a diverse population of molecules is generated through combinatorial chemistry, the molecules are then subjected to a selection step, and, if they are based on a biological expression system, additional rounds of selection are applied. Thus, from a library of millions to billions of different molecules generated in a combinatorial fashion, it now is possible to isolate optimized ligands to many protein targets.

One major difference between natural and molecular evolution is that typically the selected molecules are not permitted to undergo additional diversification through mutation or recombination. However, it is possible to take a sequence selected from a combinatorial library and then to generate a second combinatorial library that is varied around the selected sequence. These 'biased' or 'directed' libraries often yield even stronger binding entities [1,2]. Alternatively, it is possible to propagate selected phage in an error-prone bacterial strain [3] or to shuffle the insert DNA following each round of selection [4,5].

As will be discussed below, screening combinatorial peptide libraries by affinity selection often yields peptide ligands that resemble the primary structures within interacting proteins (Fig. 2). This seems to occur especially often when the interaction between two proteins involves a short contiguous stretch of peptide residues in one interacting protein and a peptide-binding pocket in the other protein. In this review, we highlight various examples where the peptides selected for binding to a protein resemble a region within a known or putative interacting protein(s). Thus, mapping protein–protein interactions with combinatorial peptides can be just as fruitful an approach in proteomics research as yeast two-hybrid screening of cDNAs [6].

## 2. Defining the specificity of protein interaction modules

A large number of protein–protein interactions in eukaryotic cells have been shown to involve protein interaction modules and their cognate ligands. These modules include the Src homology 2 (SH2), Src homology 3 (SH3), phosphotyrosine-binding, WW, Eps15 homology (EH), PSD-95/dlg/ZO1 (PDZ) and Ena/VASP homology domains, to name a few. The structure and function of these domains have been the subject of recent reviews elsewhere [7–11]. In general, the modules bind short continuous regions within interacting proteins. To define the specificity of these protein interaction modules, combinatorial peptide libraries have been used to isolate peptide ligands, and very often there is an excellent correspondence between the primary structures of the peptide ligands and regions within known interacting proteins.

### 2.1. SH3 domains

SH3 domains typically bind proline-rich peptides with the motif PxxP, where x may be any amino acid [12]. Phage-displayed and chemically synthesized combinatorial peptide libraries have been used to define the specificity of individual SH3 domains. In this manner, the specificity of SH3 domains of Abl [13], amphiphysin I [14], cortactin [13], Fyn, Grb2 [13], Lyn [15], phosphatidylinositol 3-kinase [16] and Yes [13] has been determined.

For example, the optimal ligand consensus defined for one

*Corresponding author. Fax: (1)-608-262 1257.
E-mail: bkkay@facstaff.wisc.edu

0014-5793/00/$20.00 
PII: S0014-5793(00)01778-6

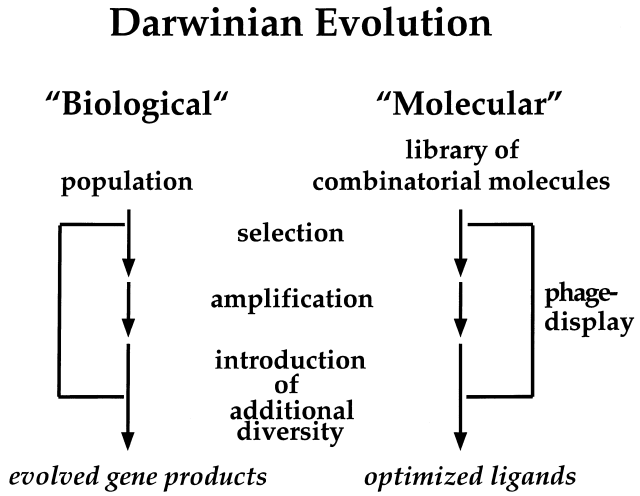


Fig. 1. Comparison of biological and molecular evolution. On the left, the steps of evolution are outlined as defined by Charles Darwin. Populations of organisms, which are selected based on the some trait, give rise to individuals who pass the selected alleles onto the next generation. This process is repeated each generation, eventually leading to the production of highly optimized genes, proteins and organisms. On the right, the steps of molecular evolution are shown for comparison. Populations of molecules, which have been generated by combinatorial chemistry, are selected by affinity selection. If the molecules are linked to a biological system, the selected molecules are propagated, ultimately yielding molecules that bind to the target.

of the two SH3 domains in Crk is PxLPxK, where x is any amino acid. As seen in Fig. 3A, this ligand sequence is present in Abl (PLLPTK), C3G (PALPPK), cDNA clone ST12 (PGLPSK), DOCK 180 (PPLPLK) and Eps15 (PALPPK), which were originally identified as Crk-interacting proteins by 'traditional means' (i.e. immunoprecipitation, far Western blotting, screening cDNA expression libraries) [17–20]. Thus, if the sequence of the entire mouse genome was known, scanning the mouse proteome for proteins bearing the PxLPxK motif could have identified the same five proteins as candidate Crk-interacting proteins.

The three-dimensional structure of the N-terminal SH3 domain of Crk complexed with a peptide segment of C3G has been solved [21]. In the structure (Fig. 3B), the central six residues (PALPPK) of the peptide adopt a left-handed polyproline type II helix that fits into the three shallow hydrophobic pockets of the SH3 domain. Interestingly, the four residues within the peptide that make contact with the surface of the SH3 domain (Fig. 3B) are the same ones conserved among the phage-displayed peptide ligands (PxLPxK).

### 2.2. WW domains

This domain was originally characterized as a short, 35 amino acid repeat found in human, mouse and chicken homologues of the Yes-associated protein (YAP). The specificity of the domain was first revealed when two proteins (WBP1 and 2) were identified by yeast two-hybrid analysis to bind to the YAP WW domain [22]. Alanine scanning experiments of a motif shared by both proteins, PPPPY, demonstrated that the PPxY sequence is the site of binding for the WW domain.

WW domains have been used to select peptides from combinatorial phage-displayed libraries and these analyses support previous findings that certain WW domains bind a PPxY motif [23]. Based on these phage-display-derived WW domain ligands, predictions have been made regarding the cellular ligands of the WW domain-containing proteins. When the C-terminal WW domain of the mouse Nedd-4 protein, an ubiquitin ligase, is used to affinity select peptides from a phage-display combinatorial peptide library, peptides containing the PPxY motif are isolated (Fig. 4). If the longer consensus sequence (i.e. PPxYES(L/M)) is used to search the Swiss-Prot protein database, one of the most interesting matches is the β-subunit of human, rat, rabbit and frog epithelial sodium channels (ENaC). This interaction is functionally significant, as truncations or substitutions within the PPxY motif [24–26] give rise to Liddle's syndrome, a form of hypertension that is caused by extending the half-life of the mutant ENaC subunits, which fail to be ubiquitinated and subsequently degraded by the ubiquitin/26S proteosome pathway [27].

### 2.3. EH domains

The EH domain mediates a number of important protein–protein interactions in endocytosis and molecular trafficking [11]. EH domains are about 100 amino acids in length, and bind short peptides in a hydrophobic pocket formed between two α-helices [28,29]. When a phage-displayed library of combinatorial peptides was screened with the EH domains of Eps15 [30] and intersectin [31], peptides containing the tripeptide motif, NPF, were isolated (Fig. 5). In support of this result, cDNA segments encoding several novel proteins with multiple NPF motifs [30,31] were isolated from a lambda-cDNA expression library, when the same EH domains were used as probes. The proteins have since been demonstrated to be cellular ligands for Eps15 and intersectin [32–36] and likely interact in a multivalent fashion [11,37]. Thus, analysis of the

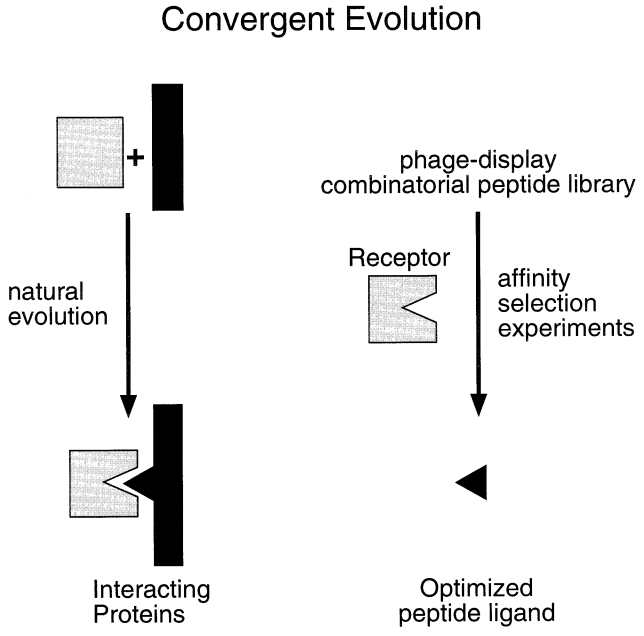


Fig. 2. Convergent evolution. In nature, two proteins evolve a highly specific interaction by adopting either a receptor or ligand mode of interaction (left). Conversely, a receptor can select peptides from phage-displayed combinatorial peptide libraries that represent optimal ligands for the receptor (right). When the peptides selected from the combinatorial libraries have the same primary structure as the site within the natural cellular ligand, we have termed this 'convergent evolution'.

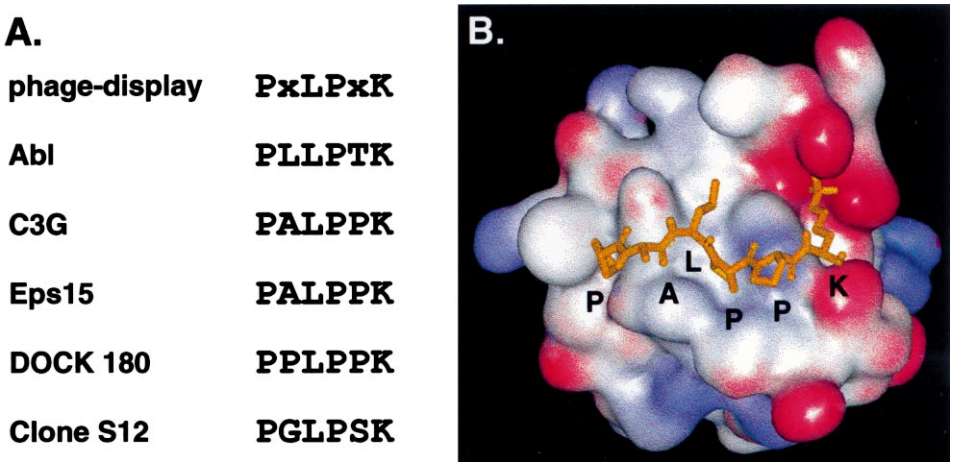


Fig. 3. Isolation of peptides and proteins that bind to the N-terminal SH3 domain of the oncoprotein Crk. Crk is a protein believed to act at the interface of the cytoskeleton and signal transduction, and contains two SH3 domains and a single SH2 domain. A: A phage-display combinatorial ($x_6PxxPx_6$) peptide library was screened for peptide ligands to the N-terminal SH3 domain of Crk and the consensus sequence is shown [13], where x is any amino acid. Also listed is the motif within C3G, DOCK 180, Eps15 and clone ST12 [19,20] identified to bind to the N-terminal SH3 domain of Crk. B: The three-dimensional structure of the N-terminal SH3 domain of Crk has been solved while complexed with a peptide segment, PPPALPPKKR, of C3G [21]. The surface of the SH3 domain is shown with blue and red denoting positive and negative electrostatic charges, respectively, along with the central six residues (PALPPK) of the C3G peptide. The PDB coordinates (1CKA) of the complex were modeled with the program WebLab ViewerLite (version 3.2) from Molecular Simulations Incorporated (http://www.msi.com).

molecular recognition properties of EH domains provides an excellent example of convergent evolution between combinatorial peptides and natural proteins.

### 2.4. PDZ domains

Another well-studied protein interaction module is the PDZ domain. PDZ domains were first described as 80–100 amino acid repeats within the post-synaptic density 95 protein [38], and since then, 685 examples have been identified in over 370 eukaryotic proteins [39]. In their 'classical' mode of recognition, PDZ domains bind to peptide sequences at the C-termini of certain membrane proteins, where the free carboxylate group docks into a highly conserved hydrophobic pocket on the PDZ domain surface and the side chains fit within a peptide-binding groove of the domain [40,41]. Specificity in binding is defined by the 3–8 amino acid residues preceding the C-terminus [41–43]. Ligand preferences for various PDZ domains have been defined with chemically synthesized combinatorial peptide libraries immobilized on cellulose membranes [44,45] or in solution [43].

Another example of the remarkable convergent evolution seen with combinatorial peptides was the independent prediction that the syntrophin PDZ domain can interact with voltage-gated sodium channels (VGSCs) [44]. The optimal ligand preference for the α-syntrophin PDZ domain was defined by screening a library of combinatorial peptides immobilized on membranes to be (K/R)E(S/T)xV-COOH. Interestingly, the sequence (K/R)ES(L/I)V-COOH is present at the C-terminus of the α-subunits of all known vertebrate VGSCs [44]. This match was biologically significant, since it has been independently shown that the VGSCs of skeletal and cardiac muscles copurify with syntrophin from extracts of skeletal and cardiac

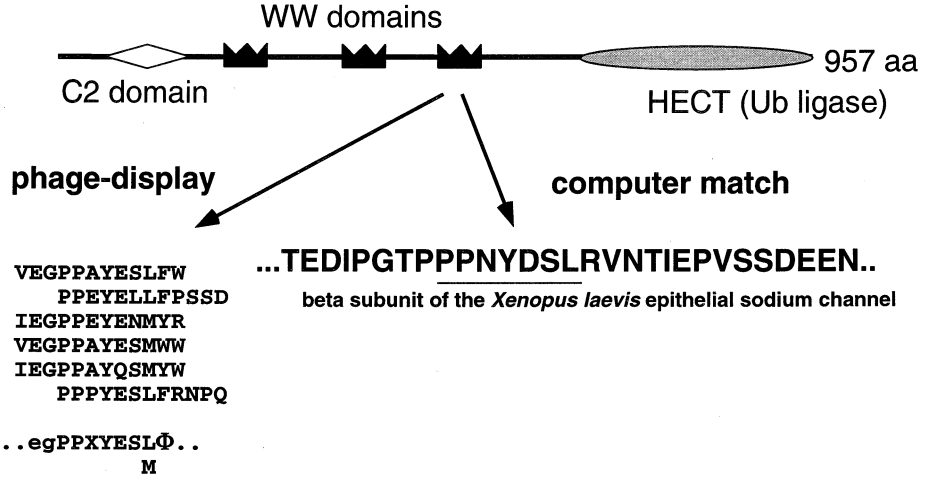


Fig. 4. Isolation of peptides and proteins that bind to the central WW domain of mouse Nedd-4. Nedd-4 is a signal transduction protein, which contains a C2 domain, three WW domains and an ubiquitin ligase catalytic domain [92]. A phage-display combinatorial ($X_{12}$) peptide library was screened for peptide ligands to the C-terminal WW domain of Nedd-4 and isolated ligands are shown on the left. The N-terminal peptide ligand for the domain contains the core motif, (L/P)PSY. Computer searches with this motif (boxed) identified the ENaC of *Xenopus laevis* as a potential interacting protein.

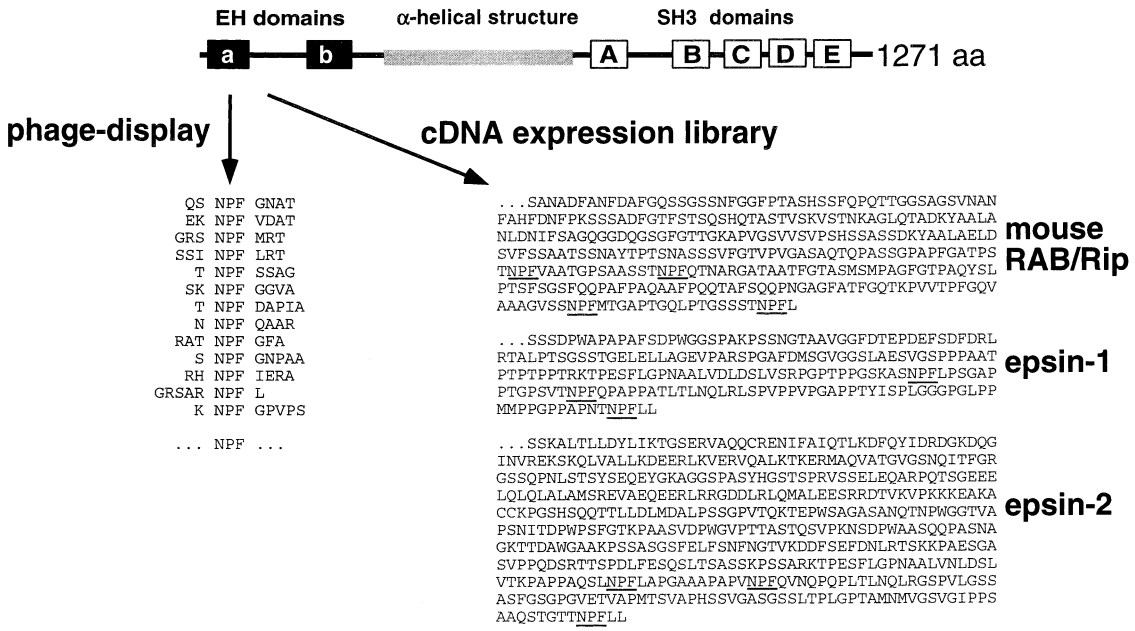


Fig. 5. Isolation of peptides and proteins that bind to the EH domains of intersectin. Intersectin is a protein component of the endocytic machinery, and contains two EH domains, a central region believed to adopt a coiled-coil structure, and five SH3 domains [11]. A phage-display combinatorial ($X_9$) peptide library was screened for peptide ligands to the N-terminal EH domain of intersectin; the isolated ligands are shown on the left [31]. The optimal peptide ligand for the domain contains the core motif, NPF. A mouse cDNA expression library [31], which was screened with the same EH domain, yielded three different sequences: Rev-associated binding proteins RAB [90], also known as the Rev-interacting protein (Rip) [91] or Hrb [34], and epsin-1 [32] and epsin-2. In the C-terminal region of the three proteins are 3–4 copies of the NPF motif (underlined).

muscles [46]. In addition, this interaction can be competed with peptides corresponding to the C-terminal 10 amino acids of the cardiac and skeletal muscle VGSC α-subunits, indicating that the syntrophin PDZ domain interacts with the channels through their C-terminal peptide sequences.

Songyang et al. [43] used a panel of nine different PDZ domains to fractionate chemically synthesized random peptide libraries in solution. The optimal peptide ligands predicted by the selected peptides were in good agreement with sequences present at the C-termini of known interacting proteins of the PDZ domains tested.

The requirement for a free carboxylate group has limited the use of phage-displayed libraries in defining the specificity of PDZ domains, since peptides are typically displayed at the N-terminus of proteins III or VIII of bacteriophage M13. However, it was possible to select peptide ligands from a

Table 1
List of recent experiments with combinatorial peptides exhibiting convergent evolution

| Protein | Peptide ligand motif | Matching sequence | Putative interacting protein | References |
|---|---|---|---|---|
| Cortactin SH3 domain | +PPΨPxK | KPPVPPK | CortBP1, Shank | [13,93,94] |
| Amphiphysin I SH3 domain | PxRPx(R/H)(R/H), #xRPxP | PSRPNR, PIRPSR | Dynamin, Synaptojanin | [14,95] |
| Endophilin 1 SH3 domain | Ψ+RPPPP | PKRPPPP | Synaptojanin | [95] |
| Eps8 SH3 domain | PxxDY | PxxDY | E3b1/Abi-1 | [96] |
| EH domains of Eps15 and intersectin | NPF | NPF | Epsins, Numb, Rev/Rip | [30,31] |
| YAP WW domain | PPxY | PPPY | YAP | [22,23] |
| Grb2 SH2 domain | PPYxNΨ | pYVNV | SHC | [97] |
| Cbl SH2 domain | D(N/D)xpY | SDGpY | ZAP-70 | [98] |
| PDZ domain | (K/R)E(S/T)C(L/M) | (K/R)ES(L/I)V | VGSCs | [44,99] |
| Estrogen receptor α | LxxLL | LxxLL | Transcriptional activators | [55,100] |
| DM2 | FxDxWxxL | FSDLWKLL | p53 | [89,101] |
| Caveolin | ΦxΦxxxxΦ | WSYGVTVW | Tyrosine kinases | [102,103] |
| HIV-1 Vpr | WxxF | WEQF | Uracil DNA glycosylase | [104] |
| Troponin C | (V/L)(D/E)xLKxxLxxLA | LKTLLLQIA | Troponin I | [105] |
| Vinculin | DVYDWARRVS | DVYTKKELIECARRVS | Talin | [106] |
| PP1c | Vx(F/W) | Vx(F/W) | Many PP1 regulatory proteins | [59] |
| Integrins: | | | | |
| $\alpha_v\beta_1$ | RGD | RGD | Fibronectin | [107,108] |
| $\alpha_v\beta_3$ | RGD | RGD | Fibronectin | [109] |
| $\alpha_v\beta_3$ | SFFARR | SLFAFR | Vitronectin | [110] |
| $\alpha_v\beta_6$ | DLxxL | DLxxL | Extracellular matrix components | [111] |

Ψ represents hydrophobic residues (i.e. I, L, P, V), Φ represents aromatic residues (i.e. W, F, Y), + represents positively charged residues (i.e. K, R), and x represents any residue. pY refers to a phosphorylated tyrosine residue.

library of cysteine-cyclized combinatorial peptides at the N-terminus of protein III with the PDZ domain of α-syntrophin. The consensus of the selected peptides, (K/R)ETC(L/M), was very similar to the peptide sequence, V(K/R)ES(L/I)V-COOH, which is present at the C-termini of the α-subunits of VGSCs [44].

The isolation of a cyclized peptide ligand for the PDZ domain of α-syntrophin suggested that this domain can recognize non-C-terminal peptide sequences. Recently, the three-dimensional structure of a heterodimer between the PDZ domains of α-syntrophin and neuronal nitric oxide synthase has confirmed this suggestion, by demonstrating that the PDZ domain of α-syntrophin interacts with internal peptide sequences present within a specific secondary structure [47].

### 2.5. Various protein targets

Over the last few years, a number of examples supporting the concept of convergent evolution have been published. Table 1 lists those instances where the peptide ligands to various proteins that have been isolated from combinatorial peptide libraries resemble or match peptide sequences present within interacting proteins. The text below focuses on selected entries in Table 1.

The half-life of the tumor suppressor protein p53 is regulated in part by its binding to the ubiquitin ligase, DM2 [48,49]. When the N-terminal 140 amino acids of DM2 were used to screen phage-displayed combinatorial peptide libraries [50], the selected peptides share the consensus xFxDxWxxLx, which is strikingly similar to a sequence at the N-terminus of p53, TFSDLWKLLP (amino acids 18–27). Interestingly, the same residues that are conserved also fit within the hydrophobic pocket of DM2 [51].

The conformational state of the estrogen receptor (ERα) is differentially affected by the binding of its ligands, such as estradiol and anti-estrogen drug tamoxifen [52,53]. Depending on the ligand bound, the receptor is hypothesized to expose different surfaces, thus engaging different or overlapping sets of protein components of the transcriptional machinery. Recently, phage-displayed combinatorial peptide libraries have been used to probe estradiol–ERα and tamoxifen–ERα complexes [54,55]. Intriguingly, several of the peptides isolated that bound to the estradiol-activated ERα contained the sequence LxxLL, a 'signature motif' of nuclear receptor co-activators [55]. On the other hand, screens with the tamoxifen-activated ERα yielded a very different set of peptides, with the consensus (S/M)RE(W/F)FxxxL. Searches of databases with this consensus revealed similar sequences in the yeast protein RSP5 and its human homologue, receptor potentiating factor 1. Although the biological significance of these matches remains to be clarified, both proteins are known to act as co-activators of progesterone and glucocorticoid receptors in mammalian and yeast cells [56].

Combinatorial peptides displayed on the surface of bacteria have been used to find peptide ligands to proteins. The entire coding sequence of *Escherichia coli* thioredoxin (trxA) was inserted into a dispensable region of the gene for flagellin (fliC), the major structural component of the *E. coli* flagellum, and the resulting recombinant protein (FLITRX) has been used to display combinatorial peptides on the surface of bacteria [57,58]. Not only has this system been used to define the epitopes of various monoclonal antibodies [57], but it has been used to isolate peptide ligands to the catalytic subunit of protein phosphatase 1 (PP1c) [59]. The PP1c-binding peptides share the motif Vx(F/W), where x is often H or R; in addition, the motif is commonly preceded by 2–5 basic residues and followed by one acidic residue. Interestingly, this motif occurs in a number of PP1c regulatory proteins [60] and structural studies have shown that it fits within a hydrophobic channel present in the C-terminal region of PP1c [61].

Yeast two-hybrid screening of combinatorial peptides has also been used to identify peptide ligands to a variety of protein targets. The retinoblastoma protein (Rb) has been used as 'bait', fused to the Gal4 DNA-binding domain, to search for interacting combinatorial peptides, fused to a transcriptional activation domain [62]. Out of an initial pool of three million transformants, seven peptides that bound Rb were discovered. Interestingly, these peptides shared the motif LxCxE, which is present in natural proteins such as adenovirus protein E1A, transcription factor E2F and cyclin D that have been previously shown to interact with Rb.

The heterodimerization domain of the E2F1 transcription factor has been used as bait to screen a library of 20-mer combinatorial peptides constrained within a loop of the *E. coli* thioredoxin protein [63]. One of the peptides isolated, RCVRCRFVVWIGLRVRCLV, contained a sequence similar to the KEKKEIWIGL in the heterodimerization domain of DP1, a physiological binding partner of E2F1. The biological significance of this match was supported with the observations that mutagenesis of the WIGL peptide sequence in DP1 markedly reduced its ability to heterodimerize with E2F1, and that introduction of the peptide into cells inhibited both the formation of functional E2F1/DP1 heterodimers and E2F1-regulated gene expression. These results suggest that the peptides selected from the combinatorial peptide library approximate the manner in which DP1 binds to E2F1.

However, not all peptides isolated from yeast two-hybrid screening contain sequences that match known interacting proteins. When peptides that bind to cyclin-dependent kinase 2 (Cdk2) were isolated from a library displayed in the *E. coli* thioredoxin protein [64], they failed to match any known protein, even though the peptides could inhibit interaction of Cdk2 with one of its substrates [65]. Specific inhibitors of this type represent a powerful tool for studying genetic networks inside cells. Recently, peptides were selected that block cell-cycle arrest in MATa *Saccharomyces cerevisiae* cells that have been exposed to α-factor [66]. From a collection of 38 peptides isolated in an initial selection experiment, a subset of 12 peptides was used as bait in two-hybrid screens of cDNAs. This screening revealed a gene previously not believed to be involved in pheromone response, Cbk1 (cell wall biosynthesis kinase 1), along with protein components of other cell-cycle pathways in yeast. Peptides that specifically block the spindle checkpoint, inhibit transcriptional silencing or block pheromone signaling have also been identified in an analogous manner, although they have not been extensively characterized [67].

Combinatorial peptide libraries displayed on phage have been isolated using selectively infective phage. In this novel method, combinatorial peptides are fused to the C-terminal domain of a soluble form of protein III, whereas the protein target is fused to the N-terminal domain of protein III attached to phage particles [68,69]. When there is an intermolecular interaction between the two 'halves' of the protein III molecule, infectivity is restored to the phage particles. This

technique has been used to select disulfide-constrained peptides that bind to the death domain of the p75 neurotrophin receptor [70]. Twenty-eight peptides that bound this target were recovered and sequenced, however database analysis failed to identify any sequence similarity between the recovered peptides and known proteins.

Finally, it should be mentioned that it is possible to isolate peptides that bind to non-proteinaceous targets [71]. For example, peptides have been isolated from combinatorial libraries that bind to biotin [72], plastic [73], single-stranded (ss) DNA [74], fluorescent dyes [75], taxol [76] or polyglutamine tracts within proteins involved in neurodegenerative diseases [77]. While many of these peptides do not resemble known interacting proteins, the motif shared by the ss DNA-binding peptides is present in the *E. coli* ss DNA-binding protein [74] and the taxol-binding peptides were used to predict successfully that Bcl-2 does interact with taxol inside cells [76].

## 3. Why do combinatorial peptide libraries often yield biologically relevant peptide ligands?

Data from mutagenesis studies of various protein–protein interfaces suggest that the binding energy is not evenly distributed over the protein interaction surfaces, but rather concentrated in energetic 'hot spots', in which only a small subset of the residues accounts for most of the change in free energy [78]. Elegant studies on the interaction between human growth hormone (HGH) and its receptor have revealed that of 30 or so residues comprising the interaction interface, only a small hydrophobic region of the receptor, dominated by two tryptophans, accounts for more than three quarters of the binding free energy. Similarly, only eight residues of HGH (217 amino acids) account for 85% of the binding energy [79,80]. In a study characterizing the epitopes of several monoclonal antibodies to HGH [81,82], it was discovered that typically less than three residues critically define each epitope. In another study, scanning mutagenesis of the interaction interface between the Shaker potassium channel and the 37 amino acid long charybdotoxin of scorpions revealed that only two residues contributed most of the binding, even though 15 residues of the toxin are involved in the interaction [83–85].

The highly uneven distribution of the binding energy throughout a protein–protein interface may help to explain why peptides selected from a combinatorial library by a target protein are so similar to the sequences of the physiologically relevant interacting proteins. The consensus residues in the selected peptides are the residues that are critical for binding, like the 'teeth' of the 'key' in the commonly used 'lock and key' metaphor. This is especially true for linear, unconstrained random peptides, where amino acid positions are relatively independent in their contribution to binding [86].

Examples of convergent evolution occurring at protein–protein interaction surfaces can be found in nature as well. Recently, the three-dimensional structure of the bacterial integrin-binding protein, invasin, has been solved [87]. Remarkably, the integrin-binding surfaces of invasin and fibronectin present the same key residues, two aspartic acids and an arginine, in the same relative spatial locations, but in the context of different folds and surface shapes. Thus, convergent evolution can occur in nature between different host and bacterial proteins.

## 4. Perspectives

The concept of convergent evolution has important implications in the analysis of the proteome in sequenced genomes. The genome of the yeast, *S. cerevisiae*, has been sequenced and the primary structures of its protein products have been cataloged in computer databases. The yeast genome contains 12 million bp and encodes 6027 potential proteins, for a total of 2.8 million amino acids [88]. Given that there are 3.2 million ($20^5$) mathematically possible different 5-mer peptide sequences, in a genome the size of yeast most 5-mer sequences are expected to occur approximately once in the genome, assuming randomness. Though the coding capacity of the genome is likely not random, due to functional and structural needs, once the ligand preferences for a protein interaction module or a target protein have been defined to five residues, a manageable number (i.e. $< 5$) of putative interacting candidates will result from database 'mining'. These candidates can then be tested for their ability to interact with the module or target protein by a variety of methods (i.e. co-immunoprecipitation, cross-linking, yeast two-hybrid screening, etc.).

In this review, we have highlighted several recent examples of convergent evolution of protein–protein interactions achieved with combinatorial peptides. While not all peptide ligands selected for binding to a target match the primary structure of the interacting protein, a surprising number do. Thus, a fruitful means of mapping protein–protein interactions is to isolate peptide ligands to a target protein and identify candidate interacting proteins in a sequenced genome by computer analysis. The identification of protein interaction networks should accelerate functional analysis of proteomes and drug discovery [89].

*Acknowledgements:* We thank Dr. Jeffery Rubin for comments on the manuscript. This work was supported by The Leukemia and Lymphoma Society.

## References

[1] Stemmer, W. (1995) Bio/Technology 13, 549–553.
[2] Kay, B.K., Adey, N.B. and Stemmer, W.P.C. (1996) in: Phage Display of Peptides and Proteins: a Laboratory Manual (Kay, B.K., Winter, J. and McCafferty, J., Eds.), Academic Press, San Diego, CA.
[3] Low, N.M., Holliger, P.H. and Winter, G. (1996) J. Mol. Biol. 260, 359–368.
[4] Crameri, A., Cwirla, S. and Stemmer, W.P. (1996) Nat. Med. 2, 100–102.
[5] Chang, C.C., Chen, T.T., Cox, B.W., Dawes, G.N., Stemmer, W.P., Punnonen, J. and Patten, P.A. (1999) Nat. Biotechnol. 17, 793–797.
[6] Phizicky, E. and Fields, S. (1995) Microbiol. Rev. 59, 94–123.
[7] Shoelson, S.E. (1997) Curr. Opin. Chem. Biol. 1, 227–234.
[8] Mayer, B.J. and Gupta, R. (1998) Curr. Top. Microbiol. Immunol. 228, 1–22.
[9] Rotin, D. (1998) Curr. Top. Microbiol. Immunol. 228, 115–133.
[10] Prehoda, K.E., Lee, D.J. and Lim, W.A. (1999) Cell 97, 471–480.
[11] Santolini, E., Salcini, A.E., Kay, B.K., Yamabhai, M. and Di Fiore, P.P. (1999) Exp. Cell Res. 253, 186–209.
[12] Kay, B., Williamson, M. and Sudol, M. (2000) FASEB J. 14, 231–241.
[13] Sparks, A., Rider, J., Hoffman, N., Fowlkes, D., Quilliam, L. and Kay, B. (1996) Proc. Natl. Acad. Sci. USA 93, 1540–1544.
[14] Grabs, D., Slepnev, V.I., Songyang, Z., David, C., Lynch, M., Cantley, L.C. and De Camilli, P. (1997) J. Biol. Chem. 272, 13419–13425.

[15] Stauffer, T.P., Martenson, C.H., Rider, J.E., Kay, B.K. and Meyer, T. (1997) Biochemistry 36, 9388–9394.
[16] Chen, J.K., Lane, W.S., Brauer, A.W., Tanaka, A. and Schreiber, S.L. (1993) J. Am. Chem. Soc. 115, 12591.
[17] Knudsen, B.S., Feller, S.M. and Hanafusa, H. (1994) J. Biol. Chem. 269, 32781–32787.
[18] Ren, R., Ye, Z.S. and Baltimore, D. (1994) Genes Dev. 8, 783–795.
[19] Matsuda, M., Ota, S., Tanimura, R., Nakamura, H., Matuoka, K., Takenawa, T., Nagashima, K. and Kurata, T. (1996) J. Biol. Chem. 271, 14468–14472.
[20] Schumacher, C., Knudsen, B.S., Ohuchi, T., Di Fiore, P.P., Glassman, R.H. and Hanafusa, H. (1995) J. Biol. Chem. 270, 15341–15347.
[21] Wu, X., Knudsen, B., Feller, S.M., Zheng, J., Sali, A., Cowburn, D., Hanafusa, H. and Kuriyan, J. (1995) Structure 3, 215–226.
[22] Chen, H.I. and Sudol, M. (1995) Proc. Natl. Acad. Sci. USA 92, 7819–7823.
[23] Linn, H., Ermekova, K.S., Rentschler, S., Sparks, A.B., Kay, B.K. and Sudol, M. (1997) Biol. Chem. 378, 531–537.
[24] Hansson, J.H. et al. (1995) Proc. Natl. Acad. Sci. USA 92, 11495–11499.
[25] Tamura, H., Schild, L., Enomoto, N., Matsui, N., Marumo, F. and Rossier, B.C. (1996) J. Clin. Invest. 97, 1780–1784.
[26] Inoue, J. et al. (1998) J. Clin. Endocrinol. Metab. 83, 2210–2213.
[27] Staub, O. et al. (2000) Kidney Int. 57, 809–815.
[28] de Beer, T., Carter, R.E., Lobel-Rice, K.E., Sorkin, A. and Overduin, M. (1998) Science 28, 1357–1360.
[29] de Beer, T., Yamabhai, M., Kay, B.K. and Overduin, M. (submitted).
[30] Salcini, A.E. et al. (1997) Genes Dev. 11, 2239–2249.
[31] Yamabhai, M., Hoffman, N.G., Hardison, N.L., McPherson, P.S., Castagnoli, L., Cesareni, G. and Kay, B.K. (1998) J. Biol. Chem. 273, 31401–31406.
[32] Chen, H., Fre, S., Slepnev, V., Capua, M., Takei, K., Butler, M., Di Fiore, P. and De Camilli, P. (1998) Nature (London) 394, 793–798.
[33] Chen, H., Slepnev, V.I., Di Fiore, P.P. and De Camilli, P. (1999) J. Biol. Chem. 274, 3257–3260.
[34] Doria, M., Salcini, A.E., Colombo, E., Parslow, T.G., Pelicci, P.G. and Di Fiore, P.P. (1999) J. Cell Biol. 147, 1379–1384.
[35] Hussain, N.K. et al. (1999) J. Biol. Chem. 274, 15671–15677.
[36] Sengar, A.S., Wang, W., Bishay, J., Cohen, S. and Egan, S.E. (1999) EMBO J. 18, 1159–1171.
[37] Wendland, B. and Emr, S.D. (1998) J. Cell Biol. 141, 71–84.
[38] Cho, K.O., Hunt, C.A. and Kennedy, M.B. (1992) Neuron 9, 929–942.
[39] Kornau, H.C. and Seeburg, P.H. (1997) Nat. Biotechnol. 15, 319.
[40] Cabral, J.H. et al. (1996) Nature (London) 382, 649–652.
[41] Doyle, D.A., Lee, A., Lewis, J., Kim, E., Sheng, M. and MacKinnon, R. (1996) Cell 85, 1067–1076.
[42] Stricker, N.L. et al. (1997) Nat. Biotechnol. 15, 336–342.
[43] Songyang, Z. et al. (1997) Science 275, 73–77.
[44] Schultz, J., Hoffmuller, U., Krause, G., Ashurst, J., Macias, M.J., Schmieder, P., Schneider-Mergener, J. and Oschkinat, H. (1998) Nat. Struct. Biol. 5, 19–24.
[45] Hoffmuller, U., Russwurm, M., Kleinjung, F., Ashurst, J., Oschkinat, H., Volkmer-Engert, R., Koesling, D. and Schneider-Mergener, J. (1999) Angew. Chem. Int. Ed. 38, 2000–2004.
[46] Gee, S.H., Madhavan, R., Levinson, S.R., Caldwell, J.H., Sealock, R. and Froehner, S.C. (1998) J. Neurosci. 18, 128–137.
[47] Hillier, B.J., Christopherson, K.S., Prehoda, K.E., Bredt, D.S. and Lim, W.A. (1999) Science 284, 812–815.
[48] Levine, A.J. (1997) Cell 88, 323–331.
[49] Piette, J., Neel, H. and Marechal, V. (1997) Oncogene 15, 1001–1010.
[50] Bottger, V., Bottger, A., Howard, S.F., Picksley, S.M., Chene, P., Garcia-Echeverria, C., Hochkeppel, H.K. and Lane, D.P. (1996) Oncogene 13, 2141–2147.
[51] Kussie, P.H., Gorina, S., Marechal, V., Elenbaas, B., Moreau, J., Levine, A.J. and Pavletich, N.P. (1996) Science 274, 948–953.
[52] Brzozowski, A.M. et al. (1997) Nature (London) 389, 753–758.
[53] Shiau, A.K., Barstad, D., Loria, P.M., Cheng, L., Kushner, P.J., Agard, D.A. and Greene, G.L. (1998) Cell 95, 927–937.
[54] Paige, L.A. et al. (1999) Proc. Natl. Acad. Sci. USA 96, 3999–4004.
[55] Norris, J.D. et al. (1999) Science 285, 744–746.
[56] Imhof, M.O. and McDonnell, D.P. (1996) Mol. Cell Biol. 16, 2594–2605.
[57] Lu, Z., Murray, K.S., Van Cleave, V., LaVallie, E.R., Stahl, M.L. and McCoy, J.M. (1995) Biotechnology 13, 366–372.
[58] Lu, Z., Tripp, B.C. and McCoy, J.M. (1998) Methods Mol. Biol. 87, 265–280.
[59] Zhao, S. and Lee, E.Y. (1997) J. Biol. Chem. 272, 28368–28372.
[60] Aggen, J.B., Nairn, A.C. and Chamberlin, R. (2000) Chem. Biol. 7, R13–R23.
[61] Egloff, M.P., Johnson, D.F., Moorhead, G., Cohen, P.T., Cohen, P. and Barford, D. (1997) EMBO J. 16, 1876–1887.
[62] Yang, M., Wu, Z. and Fields, S. (1995) Nucleic Acids Res. 23, 1152–1156.
[63] Fabbrizio, E., Le Cam, L., Polanowska, J., Kaczorek, M., Lamb, N., Brent, R. and Sardet, C. (1999) Oncogene 18, 4357–4363.
[64] Colas, P., Cohen, B., Jessen, T., Grishina, I., McCoy, J. and Brent, R. (1996) Nature (London) 380, 548–550.
[65] Cohen, B.A., Colas, P. and Brent, R. (1998) Proc. Natl. Acad. Sci. USA 95, 14272–14277.
[66] Geyer, C.R., Colman-Lerner, A. and Brent, R. (1999) Proc. Natl. Acad. Sci. USA 96, 8567–8572.
[67] Norman, T.C., Smith, D.L., Sorger, P.K., Drees, B.L., O'Rourke, S.M., Hughes, T.R., Roberts, C.J., Friend, S.H., Fields, S. and Murray, A.W. (1999) Science 285, 591–595.
[68] Krebber, C., Spada, S., Desplancq, D., Krebber, A., Ge, L. and Pluckthun, A. (1997) J. Mol. Biol. 268, 607–618.
[69] Spada, S. and Pluckthun, A. (1997) Nat. Med. 3, 694–696.
[70] Ilag, L.L., Rottenberger, C., Liepinsh, E., Wellnhofer, G., Rudert, F. and Otting, G. (1999) Biochem. Biophys. Res. Commun. 255, 104–109.
[71] Rodi, D.J. and Makowski, L. (1999) Curr. Opin. Biotechnol. 10, 87–93.
[72] Saggio, I. and Laufer, R. (1993) Biochem. J. 293, 613–616.
[73] Adey, N.B., Mataragnon, A.H., Rider, J.E., Carter, J.M. and Kay, B.K. (1995) Gene 156, 27–31.
[74] Cheng, X., Kay, B.K. and Juliano, R.L. (1996) Gene 171, 1–8.
[75] Rozinov, M.N. and Nolan, G.P. (1998) Chem. Biol. 5, 713–728.
[76] Rodi, D.J., Janes, R.W., Sanganee, H.J., Holton, R.A., Wallace, B.A. and Makowski, L. (1999) J. Mol. Biol. 285, 197–203.
[77] Nagai, Y., Tucker, T., Ren, H., Kenan, D.J., Henderson, B.S., Keene, J.D., Strittmatter, W.J. and Burke, J.R. (2000) J. Biol. Chem. 275, 10437–10442.
[78] Stites, W.E. (1997) Chem. Rev. 97, 1233–1250.
[79] Clackson, T. and Wells, J.A. (1995) Science 267, 383–386.
[80] Atwell, S., Ultsch, M., De Vos, A.M. and Wells, J.A. (1997) Science 278, 1125–1128.
[81] Jin, L., Fendly, B.M. and Wells, J.A. (1992) J. Mol. Biol. 226, 851–865.
[82] Jin, L. and Wells, J.A. (1994) Protein Sci. 3, 2351–2357.
[83] Hidalgo, P. and MacKinnon, R. (1995) Science 268, 307–310.
[84] Goldstein, S.A., Pheasant, D.J. and Miller, C. (1994) Neuron 12, 1377–1388.
[85] Ranganathan, R., Lewis, J.H. and MacKinnon, R. (1996) Neuron 16, 131–139.
[86] Houghten, R.A. (1993) Gene 137, 7–11.
[87] Hamburger, Z.A., Brown, M.S., Isberg, R.R. and Bjorkman, P.J. (1999) Science 286, 291–295.
[88] Mewes, H.W. et al. (1997) Nature (London) 387, 7–65.
[89] Kay, B.K., Kurakin, A. and Hyde-DeRuyscher, R. (1998) Drug Discov. Today 3, 370–378.
[90] Bogerd, H.P., Fridell, R.A., Madore, S. and Cullen, B.R. (1995) Cell 82, 485–494.
[91] Fritz, C.C., Zapp, M.L. and Green, M.R. (1995) Nature (London) 376, 530–533.
[92] Kumar, S., Harvey, K.F., Kinoshita, M., Copeland, N.G., Noda, M. and Jenkins, N.A. (1997) Genomics 40, 435–443.
[93] Du, Y., Weed, S.A., Xiong, W.C., Marshall, T.D. and Parsons, J.T. (1998) Mol. Cell Biol. 18, 5838–5851.
[94] Naisbitt, S., Kim, E., Tu, J.C., Xiao, B., Sala, C., Valtschanoff, J., Weinberg, R.J., Worley, P.F. and Sheng, M. (1999) Neuron 23, 569.
[95] Cestra, G. et al. (1999) J. Biol. Chem. 274, 32001–32007.

[96] Mongiov, A.M., Romano, P.R., Panni, S., Mendoza, M., Wong, W.T., Musacchio, A., Cesareni, G. and Di Fiore, P.P. (1999) EMBO J. 18, 5300–5309.
[97] Dente, L., Vetriani, C., Zucconi, A., Pelicci, G., Lanfrancone, L., Pelicci, P.G. and Cesareni, G. (1997) J. Mol. Biol. 269, 694–703.
[98] Lupher Jr., M.L., Songyang, Z., Shoelson, S.E., Cantley, L.C. and Band, H. (1997) J. Biol. Chem. 272, 33140–33144.
[99] Gee, S.H., Sekely, S.A., Lombardo, C., Kurakin, A., Froehner, S.C. and Kay, B.K. (1998) J. Biol. Chem. 273, 21980–21987.
[100] Heery, D.M., Kalkhoven, E., Hoare, S. and Parker, M.G. (1997) Nature (London) 387, 733–736.
[101] Bottger, A. et al. (1997) J. Mol. Biol. 269, 744–756.
[102] Couet, J., Li, S., Okamoto, T., Ikezu, T. and Lisanti, M.P. (1997) J. Biol. Chem. 272, 6525–6533.
[103] Couet, J., Sargiacomo, M. and Lisanti, M. (1997) J. Biol. Chem. 272, 30429–30438.
[104] BouHamdan, M., Xue, Y., Baudat, Y., Hu, B., Sire, J., Pomerantz, R.J. and Duan, L.X. (1998) J. Biol. Chem. 273, 8009–8016.
[105] Pierce, H.H., Schachat, F.H., Lombard, C.R. and Kay, B.K. (1998) J. Biol. Chem. 273, 23448–23453.
[106] Adey, N.B. and Kay, B.K. (1997) Biochem. J. 324, 523–528.
[107] Koivunen, E., Gay, D.A. and Ruoslahti, E. (1993) J. Biol. Chem. 268, 20205–20210.
[108] Koivunen, E., Wang, B. and Ruoslahti, E. (1994) J. Cell Biol. 124, 373–380.
[109] O'Neil, K.T., Hoess, R.H., Jackson, S.A., Ramachandran, N.S., Mousa, S.A. and DeGrado, W.F. (1992) Proteins Struct. Funct. Genet. 14, 509–515.
[110] Healy, J., Murayama, O., Maeda, T., Yoshino, K., Sekiguchi, K. and Kikuchi, M. (1995) Biochemistry 34, 3948–3955.
[111] Kraft, S., Diefenbach, B., Mehta, R., Jonczyk, A., Luckenbach, G.A. and Goodman, S.L. (1999) J. Biol. Chem. 274, 1979–1985.