nature biotechnology

Screening for peptide drugs from the natural repertoire of biodiverse protein folds

Paul M Watt

© 2006 Nature Publishing Group http://www.nature.com/naturebiotechnology

Although monoclonal antibody (mAb) drugs targeting protein interactions exist, these therapeutics cannot access intracellular proteins involved in disease complexes. Moreover, mAbs are more difficult to deliver and are frequently associated with a prohibitive 'royalty stack.' Outlined here is an alternative approach based on libraries of natural, highly structured peptides that offers new opportunities for identifying effective, specific inhibitors of protein-protein interactions. Libraries of such peptides (referred to hereafter as phylomers) comprise both random and structured peptides encoded by natural genes of diverse bacterial genomes. Because the number of protein subdomain structures found in nature is limited, diverse libraries containing millions of phylomers constitute virtually all of the available classes of protein fold structures, providing a rich source of peptides that interact specifically and with high affinity to human proteins. This approach may help not only in understanding the implications of each interaction identified within the interactome but also in the development of effective drugs targeted to particular protein functions. Although phylomers are active in animal models, the challenge remains to demonstrate efficacy and safety in a clinical setting.

The disappointingly low 'hit rate' of drugs targeting protein-protein interactions offered by traditional high-throughput screening necessitates the development of more efficient strategies[1,2]. Progress in finding small-molecule inhibitors of these targets has been fitful; most mAb molecules simply cannot access intracellular proteins involved in interactions, and approaches based on random combinatorial peptide libraries have thus far had only limited success. This perspective outlines an alternative approach developed in my laboratory based on libraries of natural, highly structured peptides that may offer new opportunities for designing effective and specific inhibitors of protein-protein interactions.

Protein-protein interactions as targets

Although small molecules continue to dominate the discovery of drugs for receptors, transporters, ion channels and enzymes, biological drugs (for example, antibodies, proteins and peptides) are emerging as powerful adjuncts for discovering inhibitors of targets, such as protein-protein interactions, that are not readily pharmaceutically treatable by conventional approaches. With the exhaustion of many drug discovery pipelines, it is becoming increasingly important to develop new means to access these nonclassical targets[3], particularly because conventional targets represent a small minority of the disease-associated proteome.

Efforts by the drug industry to find more efficient ways of obtaining blockers of specific protein interaction targets are currently hindered by three major obstacles: (i) limited structural knowledge of the binding surfaces involved in interactions; (ii) formidable thermodynamic barriers encountered by small molecules in binding protein interaction interfaces, which are often large and relatively featureless[4]; (iii) difficulty of finding suitable compounds in chemical libraries capable of overcoming these thermodynamic challenges.

Surmounting these obstacles has proved very challenging. For example, the hit rate in conventional high-throughput screens for small-molecule blockers of protein-protein interactions has been disappointing. The few notable exceptions can be viewed as anomalous cases in which such blocking is feasible, such as the recently described dissociation of the p53-MDM2 complex[5,6], in which most of the binding energy of the 1,500-$Å^2$ interface is localized to a smaller area of the interface than can be effectively bound by the molecule Nutlin 3 (for a review, see ref. 7).

Such 'hot spots' accessible to small-molecule disruptors have proven elusive in many other protein complexes. However, in this context we can be encouraged that nature has already evolved potent delivery vectors and molecules to interact efficiently with protein surfaces—most frequently using other proteinaceous interfaces.

Despite these challenges, we now have several blockbuster mAb drugs that target extracellular protein interactions (for example, infliximab (Remicade), which blocks the interaction of tumor necrosis factor α with its cognate receptor on B cells) very effectively, because of the large surface area and structural complexity of this class of therapeutic. It is feasible, however, that particular members of another class of biological drugs, namely peptides (in this article, peptides are defined as molecules containing as many as ~50 amino acids), may combine comparable efficacy in blocking protein interactions with more suitable characteristics for synthesis and delivery than mAbs.

Natural interfaces as inhibitors of protein-protein interactions

Several examples have been reported of competitive inhibitors of protein dimerization that exploit the structure inherent in the interfaces of the interacting proteins themselves. For example, interfacial peptides have been shown to efficiently inhibit homodimerization of HIV-1 integrase[8,9]. Interfacial peptides have also been shown to be effective inhibitors of heterodimerization in animal models. My

Phylogica Ltd., 105 Roberts Road, Subiaco, Perth, Western Australia 6008. Correspondence should be addressed to P.M.W. (paulw@phylogica.com).
Published online 7 February 2006; doi:10.1038/nbt1190

Box 1 Enfuvirtide, a peptide therapeutic derived from a protein subdomain

In March 2003, the 36-mer anti-HIV peptide enfuvirtide (Fuzeon, Trimeris/Roche) received US FDA approval. This drug is a natural interfacial sequence taken from the gp41 moiety of the HIV precursor protein gp160 and represents the first case of a peptide drug derived from a protein fragment. Enfuvirtide was very challenging to make on a large scale synthetically and thus represents a landmark in industrial peptide chemistry[56].

Enfuvirtide inhibits viral fusion by interacting with transient conformational intermediate forms of both the gp41 and gp120 target proteins, which are themselves derived from proteolysis of the gp160 precursor[57]. The allosteric change that allows gp41 to penetrate the plasma membrane of human host cells requires noncovalent binding of two segments of its chain in the context of a trimer bundle. Enfuvirtide was initially thought to be a simple competitive inhibitor of this complex formation (for a review, see ref. 58). However, recent studies show that enfuvirtide resistance mutations from patients map to both the gp41 and the interacting gp120 proteins, suggesting that the drug may in fact bind to multiple sites in both proteins and potentially act by an allosteric mechanism[57]. The skin sensitivity reaction side effects of enfuvirtide that some patients encounter may be in part due to characteristics of its pharmacokinetic profile, which normally requires twice-daily injection[59]. This profile may be improved through enhancements in the stability or the affinity of the drug or by the use of other interfaces with better affinity or pharmacokinetics.

© 2006 Nature Publishing Group http://www.nature.com/naturebiotechnology

group and collaborators[10,11] have shown how a 20-mer competitive inhibitor peptide derived from the JNK (c-Jun N-terminal kinase) interacting scaffold protein JIP competitively inhibits the access of the JNK enzyme to its substrate c-Jun. A cell-permeable version of this peptide has been constructed by fusing it to the 10–amino acid protein transduction domain derived from HIV-Tat. Injection of this Tat-JIP peptide decreases insulin resistance and improves glucose tolerance in a mouse model of type 2 diabetes[12]. An analogous but larger interfacial Tat fusion peptide derived from the same region of JIP has also been demonstrated to be sufficient to protect rats from ischemic neuronal damage when delivered by intraperitoneal injection[13].

Another dissociative peptide has been discovered that can disrupt the interaction between the tumor suppressor protein pRb (retinoblastoma protein) and Raf-1. Delivery of this peptide by the carrier peptide penetratin led to a 79% reduction in tumor volume and a 57% reduction in microvessel formation in nude mice[14]. Another recently described dominant-negative protein fragment, a portion of DSCR1, is a potent competitive inhibitor of calcineurin phosphatase with a K_i in the low nanomolar range[15]. A 13–amino acid peptide derived from the p21 activated kinase, when fused to the Tat protein transduction domain, has also been shown to block vascular leak and angiogenesis[16,17]. A prototypical example is the peptide therapeutic drug Trimeris/Roche's enfuvirtide (Fuzeon), which was approved by the US Food and Drug Administration (FDA) in 2003 and was derived from the interface of a target protein, although this may act by an allosteric mechanism (see **Box 1**).

In contrast, screens of randomly encoded peptide libraries for dissociators of protein interactions have yielded disappointing hit rates. For example, a screen of random 9-mer peptide sequences for dissociators of HIV protease gave a hit rate of less than one per million clones screened[18]. In another approach, random peptides were expressed within a scaffold sequence such as the active site loop of thioredoxin to maximize the degree of conformational constraint[19]. Screening libraries of such peptide aptamers for binders to the targets cdk2 (cyclin-dependent kinase 2) and Ras yielded hit rates of 0.0002% (ref. 19) and 0.001% (ref. 20), respectively. These low hit rates may reflect the discrepancy between numerical complexity of peptide libraries and true structural diversity. Thus, many of the peptides in randomly encoded libraries may be incapable of adopting a stable conformation unless artificially constrained in a manner that limits the potential for structural diversity.

Random versus naturally encoded peptides

The majority of peptide-based drugs that have achieved marketplace status have come from the 'stacked' deck of nature, rather than from the 'straight' deck of random combinatorial peptide libraries (**Box 2**). Whereas this could be attributed to the relatively recent development of the screening of random peptide libraries, it is intriguing to note that many of the peptide drugs awaiting FDA approval continue to be derived from natural sequences. One plausible explanation for this is that natural subdomains have been selected for stability *in vivo*. In principle, it should be possible to derive similar structures from random libraries through *in vitro* evolution approaches. In practice, however, this might present a daunting task, even in the case of very large random libraries of peptide sequences selected through phage or ribosome display screens. Given the number of possible combinations of the 20 amino acids for each residue of a peptide sequence of >12 residues, the largest libraries ever to have been constructed do not have the complexity to cover even a small fraction of the possible variants of such peptides. Indeed, it is possible that these randomly encoded libraries may not even comprehensively cover enough of the required specificities to bind particular targets, because the vast majority of random sequences do not fold into any stable structure. This may account for the poor hit rates of these libraries already described.

In contrast, earlier work has shown that subdomains of natural proteins can fold into stable structures. For example, a 30-mer peptide corresponding to the N-terminal subdomain of carp granulin folds into a structure analogous to that found in the native protein[21]. Similarly, a 19-residue fragment from the epidermal growth factor 4-like domain of thrombomodulin folds into a stable structure resembling known structures of epidermal growth factor-like proteins[22].

Additional evidence for the intrinsic stability of protein subdomains comes from domain shuffling experiments. For example, Reichmann and Winter[23] fused DNA encoding the N-terminal half of a β-barrel domain from the cold shock protein A (CspA) to fragmented *Escherichia coli* DNA. After selection for resistance to limited proteolysis, most of the chimeric polypeptides in the screen contained structurally unrelated natural open reading frames of ~40 residues fused to the CspA fold. Indeed, one of the chimeric peptides was more stable than CspA itself. In another experiment that directly compared screens, a library of peptides derived from fragments of natural proteins was found to be a much richer source of immunologically fit epitopes than libraries of random peptides, regardless of whether these peptides were conformationally constrained or unconstrained[24]. Part of this improvement in hit rate is due to the greater structural complexity of natural fold libraries, consistent with the large average size of peptide epitopes that were selected from the screen[24].

Evolution of a limited repertoire of protein folds

Compared with random peptide libraries, natural peptides provide drug developers with a shortcut to the subset of peptide conformations that have been optimized by evolution for protein interactions.

Box 2 Naturally derived peptides and protein fragments as drugs

Several of the top 100 best-selling drugs approved by the FDA are relatively unmodified peptides, many of which are derived from natural sequences. In fact, the majority of naturally effective peptide hormones are large and highly structured, accounting for their enhanced stability in the bloodstream. For example, calcitonin, somatostatin, adrenocorticotropic and beta-melanocyte-stimulating hormones are 32, 14/28, 39 and 18 amino acids in length, respectively. The peptide drug teriparatide (Forteo; Eli Lilly) is a recombinant 34-residue N-terminal fragment of human parathyroid hormone.

A major advance in recent years is the ability to make at industrially economic scale peptides of ~30 amino acids using synthetic chemistry. For example, growth hormone–releasing factor (residues 1–29), salmon calcitonin (32 amino acids) and corticotropin-releasing factor (41 amino acids) can all be made by solid-phase synthesis[60]. The noninjectable delivery of large peptides is also improving. For example, intranasally delivered calcitonin has been marketed for some time, and the FDA has recently recommended approval of an inhaled form of insulin (Exubera; Pfizer/Sanofi-Aventis/Nektar Therapeutics).

Until now, only a fraction of the true diversity of the peptides found in nature has been sampled in the search for new drug candidates. Even so, efforts are continuing to develop insect peptides as antimicrobials (for example, pyrrhocoricin[61–66]) and the venoms of spiders, scorpions and cone snails as analgesics (for example, ziconotide (Prialt); for a review, see ref. 67).

© 2006 Nature Publishing Group http://www.nature.com/naturebiotechnology

Estimates based on sequenced microbial genomes suggest that soluble globular proteins from all natural species may be derived from the assembly of a limited repertoire of only ~1,000 'folds' (structural prototype motifs)[25]. A computer database of protein folds has been constructed[26], the current version of which (http://www.cathdb.info) contains ~900 of these predicted fold classes, with ~50 new folds being added each year[27]. According to another distinct modeling approach, which takes into account many rare folds not widely represented in proteins, the total number of distinct folds that could exist is estimated to be closer to 10,000 (ref. 28).

From this defined repertoire, only a limited set of functional protein domains have evolved[29], with current estimates suggesting that there may be as few as 4,000–8,000 distinct protein families in the entire natural world[25,30]. In light of the significant redundancy of these basic structural motifs, and the presence of >100,000 distinct protein isoforms, it is clear that very similar protein subdomain structures have been used for different purposes in distinct proteins across species[31,32].

Amino acids are not randomly distributed in natural proteins, as opposed to the random composition of most artificially constructed peptide libraries[33,34]. Moreover, the distribution of amino acid residues in different structures is not random when amino acids are considered within their chemical groups[35]. One of the themes emerging from the convergence of the various genome sequencing projects, bioinformatics and crystallography is that similar structures can be encoded by highly divergent sequences because biological molecules normally seem to recognize shape and charge rather than merely primary sequence[36]. A good example of structural homology can be found in the nuclear hormone receptor superfamily. These proteins possess a DNA-binding domain, activation domain(s) as well as a ligand-binding domain (LBD) required for binding their cognate ligands. This family of receptors is involved in signaling by hydrophobic molecules as diverse as steroid and thyroid hormones, retinoids, fatty acids, prostaglandins, leukotrienes, oxysterols, bile acids and xenobiotics[32]. The distinct LBD regions in this superfamily are bound by structurally diverse ligands but retain a structurally related fold. For example, the estrogen receptor β binds estrogen, and peroxisome proliferator-activated receptor γ binds rosiglitazone (Avandia), a drug used in the treatment of type 2 diabetes. However, both of these receptors bind the phytoestrogen genistein and even share a common class of structural fold with the farnesoid X receptor, which is activated by bile acids[32].

Yet although some cross-reactivity of drugs and ligands of nuclear receptors within each basic structural fold group is sometimes observed, a considerable degree of specificity of ligand recognition remains available. Frequently such protein subdomains have evolved to interact with other protein subdomains, a feature that can be exploited in drug discovery. For example, the predominant interface by which nuclear receptors interact with cofactors is a shallow hydrophobic groove on the surface of the LBD. Upon ligand binding, a conformational shift permits docking of coactivators such as SRC-1 via their characteristic LXXLL motif[37].

If the basic shapes of protein subdomains are redundant, then we might expect that surfaces from some proteins could fortuitously interact with high affinity with surfaces from other proteins, which are not their natural partners. This phenomenon is probably the basis for the spurious background binding observed in a wide array of interaction and screening studies, ranging from immunological cross-reactivity in western blotting of proteins to background interactions observed in yeast two-hybrid, phage display or affinity chromatography assays. What is less widely appreciated is how specific such spurious interactions can occasionally be. For example, a rare class of nonbiological interactions identified from yeast two-hybrid assays exist that are highly specific unlike the common 'sticky' nonspecific interactors, which can account for much of the background obtained in such screens (http://www.fccc.edu/research/labs/golemis/InteractionTrapInWork.html). These cryptic interactions, however, could not possibly be physiological, because the two interacting protein surfaces never coexist in the same place simultaneously.

Natural dominant-negative peptides

Some groups have isolated bioactive peptides using genomic gene fragment libraries from a single species that is related to the target organism. For example, a library of such domains was used to clone new transdominant inhibitors of a pheromone response pathway in yeast[38]. Similarly, random cDNA fragments may encode coding or antisense fragments of genes that inhibit activity of the same genes or their products as transdominant suppressors[39]. Although these approaches have proved useful for cloning the natural partner domains of known proteins involved in a given pathway, intriguingly they also identified a few dominant-negative inhibitory sequences with no apparent biological link to that pathway. It is probable that such sequences are not able to exert inhibitory effects *in vivo*, being kept apart from their targets through expression in distinct tissues, subcellular compartments or through different stages of the cell cycle. Without such adaptation it seems probable that the fortuitously interacting sequences would be counterselective in evolution if the genome encoding them also encoded the target.

Thus looking for sequences encoding 'blocking' interactors in the same genome that encodes the target protein may be against the odds if the intention is to isolate high-affinity binding motifs without regard for their original function. Consistent with this proposal is the

observation that a mammalian screen for short dominant-negative cDNA molecules derived from a mammalian library does not yield a very high hit rate of natural open reading frames[40]. Nevertheless, the hit rate observed for dominant-negative agents is better than those commonly encountered for random peptide screens, suggesting that using naturally encoded sequences may still be fruitful.

Prospecting for biodiverse natural peptides

Recent data have confirmed that unrelated species can contain proteinaceous sequences that are capable of interacting with high affinity[41]. For example, a yeast two-hybrid screen of a genomic library conducted against the HIV-1 integrase identified a clone expressing a 33-mer peptide that binds tightly to the integrase target and inhibits integrase activity with an IC_{50} of 9 μM (ref. 42). Presumably, in evolution such fortuitous interactions are exploited for new purposes as new selective pressures arise. Thus, for example, peptides derived from hemoglobin and from the N terminus of human lactoferrin have been found to have potent antimicrobial activity against *Staphylococcus aureus*[43,44], and fragments of ubiquitin have known antifungal activity[45], all of which may have evolved as part of innate immunity.

To fully exploit the rich diversity of peptides in the natural kingdom with the potential to inhibit a specific human protein-protein interaction, it is possible to create novel libraries that comprise combinations of sequenced genomes that are distinct from the target genome. Unlike small random peptides, the peptides encoded by these alternative, natural sequence libraries would not require extensive modification for stability, because their larger size and natural selection have allowed enrichment for inherently more stable subdomain structures. This approach also removes evolutionary constraints on the target genome that could have selected against high-affinity irrelevant interactions with co-expressed interfaces that do not form functional complexes, thereby interfering with biological activities.

An approach adopted by some genetic bioprospectors for identifying genomic sequences encoding novel enzymatic activities involves isolating

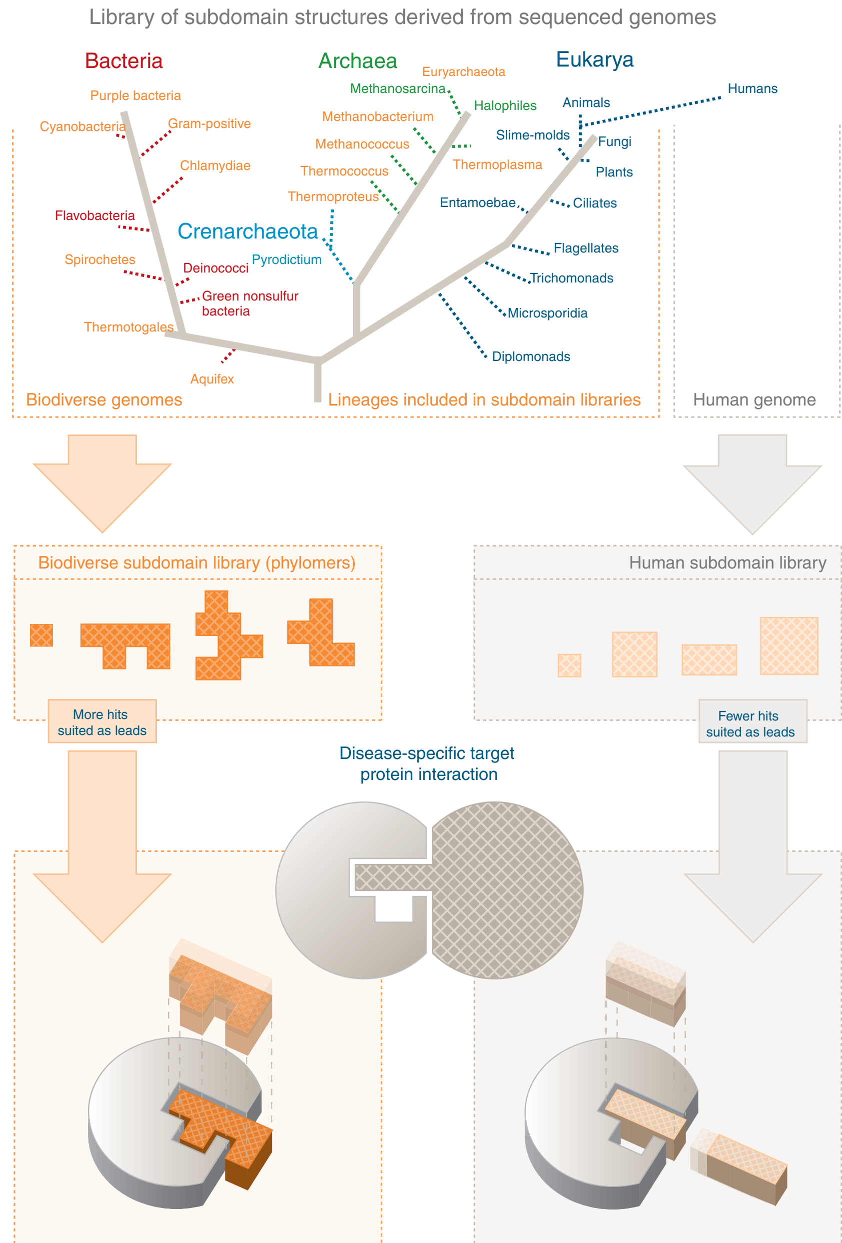

Figure 1 Constructing libraries from subdomains encoded by biodiverse genomes. The center of the diagram shows the interaction between two target proteins (shaded in gray). In this example the target proteins are depicted as human in origin. However, the target complex could also be from a pathogen source, such as a virus or bacterium. The right-hand portion of the diagram depicts the construction of a subdomain library from a set of expressed human proteins. Such a library has limited structural diversity (structures shown as lightly shaded rectangles), because it is derived only from a single source of genomic or cDNA. In addition, in the evolution of such subdomains it is expected that there may have been selection against several subdomains with very high affinities to human targets. A subdomain derived from the natural target interface is shown in the lower right portion of the diagram interacting with its cognate target. Such interactions are often selected to be reversible (depicted by ghosting) for biological reasons. The left-hand portion shows the genomic sources of a library of subdomains derived from a biodiverse array of genomes, which are distinct from the origin of the target proteins (in this example, human). The actual genomic sources used to make several such 'phylomer' libraries are shown in orange, as are the subdomain structures encoded by such libraries that are expected to be more structurally diverse than those subdomains from a single genomic source. For example, the sequence diversity between many of these archaebacterial and eubacterial genomes is vast, presumably as a result of the length of time that has elapsed for evolutionary divergence and the large differences in environments in which these organisms evolved. In contrast, there is less sequence similarity among the genera *Aquifex*, *Methanococcus* and *Cyanobacteria* than among humans, worms and yeast.

© 2006 Nature Publishing Group http://www.nature.com/naturebiotechnology

© 2006 Nature Publishing Group http://www.nature.com/naturebiotechnology

large DNA fragments capable of encoding whole genes or gene clusters, which are cloned from amplified biodiverse genome pools taken directly from the environment[46]. In contrast, the approach outlined in **Figure 1** is not designed to clone enzyme activities and thus focuses libraries on small gene fragments that generally encode peptides of ~15–80 amino acids in size. Some of those peptides that are larger than ~25 amino acids might be capable of encoding at least one structural fold, and some of the smaller peptides in this size range could be expected to have retained some secondary structure. It should be noted that some peptide sequences may be capable of adopting an induced fold on the target even if they do not fold spontaneously in solution. Moreover, such inhibitory peptides, which disrupt complexes, may include allosteric inhibitors as well as straightforward competitive inhibitors of target proteins.

By basing such libraries solely on fully sequenced microbial genomes, a protein engineer can attain maximal control of the representation of the breadth of their biodiversity, eliminating biases that might arise from those genomes of dominant species found in libraries amplified directly from uncharacterized genome pools. It should also be possible to exercise choice over the range and quantity of input DNA to account for genome size and biodiversity. The greatest sequence divergence in the biota is found between bacterial species, particularly the most ancient life forms, the archaebacteria. In constructing libraries of protein folds for screens of dissociators of protein interactions, my group has therefore chosen a diverse selection of such sequenced microbial genomes to maximize the structural diversity of such libraries. We have termed these biodiverse subdomains in our libraries 'phylomers' in recognition of their relationship to phylogeny (**Fig. 1**).

Artificially and naturally constrained scaffolds

Considerable commercial interest is now focusing on the use of small scaffolds as alternatives to antibodies, such as affibodies (Affibody), affilins (Scil Proteins), avidins (Avida), anticalins (Pieris), adNectins (Compound Therapeutics) and Kunitz domains (Dyax)[47,48]. These companies have all turned to such scaffolds to identify reagents that have comparable affinities and specificities to conventional mAbs, with the added advantages of relative ease of synthesis (compared with mAbs), suitability for noninjectable delivery approaches and the avoidance of royalty stacking associated with conventional mAb-based drugs.

The phylomer approach proposed here differs fundamentally from these approaches of screening libraries of random peptide sequences artificially constrained within one particular class of scaffold; instead, one screens libraries of many different scaffold structures in their native form. In this respect, phylomers are most analogous to 'crypteins' (Cryptome Pharmaceuticals), which are unmodified natural peptides derived from proteolytic fragments of protein extracts with (sometimes unexpected) biological activities. Phylomers differ from crypteins, however, in being derived from biodiverse sources, normally nonhuman in origin.

Using this approach, my group has obtained phylomer hits against protein targets with high affinities, comparable to those of target complexes and therapeutic antibodies. Moreover, lower affinity phylomer hits could be optimized for affinity and/or specificity using standard mutagenesis and *in vitro* evolution techniques used for refining the sequence of protein therapeutic hits (for a review, see ref. 49). The avidity of phylomers may also be optimized through multivalence engineering approaches, such as those recently described for avimers[50].

One potential drawback of screening several structural fold classes in parallel (as with phylomers or crypteins) is that the structure of each individual lead may have to be obtained empirically, rather than rapidly modeled *in silico,* as is possible with peptides based on known scaffold structures. This normally is not a significant limitation in practice, however, because most phylomer or cryptome peptide leads are <40 amino acids and are generally soluble in water, allowing the structure to be solved with relative ease by nuclear magnetic resonance.

Phylomer immunogenicity

An obvious risk of the phylomer approach is that because the peptides are derived from foreign sources, they might be expected to have a higher chance of immunogenicity than human sequences. This risk, however, may be mitigated by the small size of phylomers, which frequently lack the major histocompatibility complex (MHC)-II epitopes required for significant immunogenicity. For example, when 28 biologically active phylomer sequences were subjected to an *in silico* prediction of potential immunogenicity, most of the sequences were found to have fewer predicted MHC-II epitopes than approved peptide drugs, which were derived from foreign sources (that is, enfuvirtide, porcine insulin and salmon calcitonin; see **Supplementary Results** online). As would be expected stochastically from this analysis, the predicted probability of immunogenicity was generally proportional to the size of the phylomer peptides, with the average size of these peptides in the lowest, medium and highest immunogenicity risk groups being 22, 31 and 69 amino acids, respectively. This analysis suggests that in practice, humanization of smaller phylomers may not always be necessary.

Nevertheless, candidate phylomer-derived therapeutics would need to be individually assessed for human immunogenicity empirically. Several marketed protein therapeutics have been successfully humanized by various approaches such as complementarity-determining region grafting and chimerization, which ultimately results in the removal of MHC-II epitopes likely to cause antigenicity[51,52]. These standard humanization techniques or more concerted removal of putative epitopes by mutagenesis can be applied to immunogenic phylomers, for which the task should be simpler than with protein-based biologicals, because these peptides contain, on average, fewer MHC epitopes than large proteins, such as mAbs. Moreover, standard strategies for reducing immunogenicity and increasing half-life of biologicals, such as pegylation or the introduction of unnatural amino acids (reviewed in ref. 49), would also be appropriate to naturally encoded peptides such as phylomers and crypteins.

Protein subdomains as a source of leads in drug discovery

Naturally constrained peptide libraries could be used both as leads for developing peptidomimetic drugs, as specific blocking reagents for target validation or even as a starting point for creating a new class of small-molecule scaffolds. Although small-molecule scaffolds traditionally have been unsuitable for mimicking peptides >15 amino acids, one study has reported a terphenyl small molecule that mimics 33- and 45-residue peptide sequences[53]. Similarly, another study reported that a calixarine scaffold derivatized with several synthetic peptide loops is capable of blocking a protein-protein interaction covering a surface area of hundreds of square angstroms[54].

At the same time, our knowledge of the structural motifs of protein subdomains is rapidly increasing, together with empirical evidence for their efficacy for interfering in specific target protein-protein interactions. The increasing availability of fully sequenced genomes will thus provide protein engineers the opportunity to use algorithms to search genome sequence and model structures, ultimately allowing the adoption of virtual screening approaches.

Much progress has been made in this field of virtual structural screening and the assigning of functions of proteins (see ref. 55, for example). This has been aided by the rapid increase in solved protein structures. The Protein Structure Initiative consortium of the US National Institutes of Health is expected to have solved ~5,000 protein structures by 2010 (ref. 31).

At least in the small-molecule field, the limited repertoire of protein folds is already being exploited in drug discovery. For example, the discovery that the same basic structural fold was present in three diverse enzymes that share negligible primary sequence homology led Koch and Waldmann[32] to propose that inhibitors of one member of such a structural grouping of enzymes might be effective against other members of the cluster sharing the same fold. Encouragingly, ~2–3% of small molecules in a combinatorial library based on a natural inhibitor of CDC25 (cell division cycle 25 phosphatase) were also found to be inhibitors of other enzymes, namely acetylcholinesterase and 11β–hydroxysteroid dehydrogenases I and II. These all share a common structural fold with the CDC25 enzyme[32]. This finding confirms the importance of protein structure similarity in predicting ligand binding, despite poor sequence homology within the enzyme cluster[32].

An advantage of this approach is that it can allow the discovery of new candidate indications treatable with approved drugs, potentially minimizing the probability of major problems with safety, manufacturability and delivery. Whether this approach can also be applied to peptide inhibitors of a particular fold class remains to be seen. However, even for small molecules, this approach can only be applied to known groups of solved structures. Moreover, because the best modeling algorithms are based on and judged by empirical data, the approaches suggested here would be expected to enhance, rather than compete with, the progress of the virtual screening field, because they would allow a rapid expansion in the mapping of protein folds, which in practice are known to be capable of interaction with particular targets or drugs.

Although natural peptide interfaces clearly have the capacity to efficiently block protein-protein interactions *in vivo*, key challenges common to all peptide therapeutics will apply to these inhibitors, such as the potential for immunogenicity, stability and delivery of the dissociating agent. However, significant progress has been made in this area as the number of marketed peptide drugs accelerates.

Perhaps we risk losing the dissociative drug discovery game by continuing to gamble on long odds of random high-throughput screening of combinatorial peptide or small-molecule libraries. Peptides derived from natural protein folds offer a means of shortening our odds; presenting an opportunity to take a structural lead from nature, rather than working from scratch. It may be time to reshuffle the 'stacked' deck of functional protein folds selected through evolution to improve our chances.

Note: Supplementary information is available on the Nature Biotechnology website.

ACKNOWLEDGMENTS

I am grateful to Wayne Thomas, Prue Hart, John Finlay-Jones, Patrick Holt, Rob Aalberse, Vanessa Cull, Mark Fear, Richard Hopkins, Nadia Milech Graham Robertson, Frank Sotzik and laboratory colleagues for assisting with review of the manuscript. The assistance of Graham Carter (Accuro Biologics Ltd) in conducting the bioinformatics analysis of potential for immunogenicity of phylomer peptides is also gratefully acknowledged.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests (see the *Nature Biotechnology* website for details).

Published online at http://www.nature.com/naturebiotechnology/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

© 2006 Nature Publishing Group http://www.nature.com/naturebiotechnology

1. Pritchard, J.F. *et al.* Making better drugs: decision gates in nonclinical drug development. *Nat. Rev. Drug Discov.* **2**, 542–553 (2003).
2. Bleicher, K.H., Bohm, H.J., Muller, K. & Alanine, A.I. Hit and lead generation: beyond high-throughput screening. *Nat. Rev. Drug Discov.* **2**, 369–378 (2003).
3. Strosberg, A.D. Protein interaction mapping for target validation: the need for an integrated combinatory process involving complementary approaches. *Curr. Opin. Mol. Ther.* **4**, 594–600 (2002).
4. Archakov, A. *et al.* Protein-protein interactions as a target for drugs in proteomics. *Proteomics* **3**, 380–391 (2003).
5. Vassilev, L.T. Small-molecule antagonists of p53–MDM2 binding: research tools and potential therapeutics. *Cell Cycle* **3**, 419–421 (2004).
6. Vassilev, L.T. *et al. In vivo* activation of the p53 pathway by small-molecule antagonists of MDM2. *Science* **303**, 844–848 (2004).
7. Fischer, P. & Lane, D.P. Small-molecule inhibitors of the p53 suppressor HDM2: have protein-protein interactions come of age as drug targets? *Trends Pharmacol. Sci.* **25**, 343–346 (2004).
8. Maroun, R.G. *et al.* Peptide inhibitors of HIV-1 integrase dissociate the enzyme oligomers. *Biochemistry* **40**, 13840–13848 (2001).
9. Zhao, L., O'Reilly, M.K., Shultz, M.D. & Chmielewski, J. Interfacial peptide inhibitors of HIV-1 integrase activity and dimerization. *Bioorg. Med. Chem. Lett.* **13**, 1175–1177 (2003).
10. Barr, R.K., Hopkins, R.M., Watt, P.M. & Bogoyevitch, M.A. Reverse two-hybrid screening identifies residues of JNK required for interaction with the kinase interaction motif of JNK-interacting protein-1. *J. Biol. Chem.* **279**, 43178–43189 (2004).
11. Barr, R.K., Boehm, I., Attwood, P.V., Watt, P.M. & Bogoyevitch, M.A. The critical features and the mechanism of inhibition of a kinase interaction motif-based peptide inhibitor of JNK. *J. Biol. Chem.* **279**, 36327–36338 (2004).
12. Kaneto, H. *et al.* Possible novel therapy for diabetes with cell-permeable JNK-inhibitory peptide. *Nat. Med.* **10**, 1128–1132 (2004).
13. Borsello, T. *et al.* A peptide inhibitor of c-Jun N-terminal kinase protects against excitotoxicity and cerebral ischemia. *Nat. Med.* **9**, 1180–1186 (2003).
14. Dasgupta, P. *et al.* Disruption of the Rb-Raf-1 interaction inhibits tumor growth and angiogenesis. *Mol. Cell. Biol.* **24**, 9527–9541 (2004).
15. Chan, B., Greenan, G., McKeon, F. & Ellenberger, T. Identification of a peptide fragment of DSCR1 that competitively inhibits calcineurin activity *in vitro* and *in vivo*. *Proc. Natl. Acad. Sci. USA* **102**, 13075–13080 (2005).
16. Stockton, R.A., Schaefer, E. & Schwartz, M.A. p21-activated kinase regulates endothelial permeability through modulation of contractility. *J. Biol. Chem.* **279**, 46621–46630 (2004).
17. Kiosses, W.B. *et al.* A dominant-negative p65 PAK peptide inhibits angiogenesis. *Circ. Res.* **90**, 697–702 (2002).
18. Park, S. & Raines, R. Genetic selection for dissociative inhibitors of designated protein/protein interactions. *Nat. Biotechnol.* **18**, 847–851 (2000).
19. Colas, P. *et al.* Genetic selection of peptide aptamers that recognize and inhibit cyclin-dependent kinase 2. *Nature* **380**, 548–550 (1996).
20. Xu, C. & Luo, Z. Inactivation of Ras function by allele-specific peptide apatamers. *Oncogene* **21**, 5753–5757 (2002).
21. Vranken, W.F., James, S., Bennett, H.P. & Ni, F. Solution structures of a 30-residue amino-terminal domain of the carp granulin-1 protein and its amino-terminally truncated 3–30 subfragment: implications for the conformational stability of the stack of two beta-hairpins. *Proteins* **47**, 14–24 (2002).
22. Adler, M. *et al.* The structure of a 19-residue fragment from the C-loop of the fourth epidermal growth factor-like domain of thrombomodulin. *J. Biol. Chem.* **270**, 23366–23372 (1995).
23. Riechmann, L. & Winter, G. Novel folded protein domains generated by combinatorial shuffling of polypeptide segments. *Proc. Natl. Acad. Sci. USA* **97**, 10068–10073 (2000).
24. Matthews, L., Davis, R. & Smith, G. Immunogenetically fit subunit vaccine components via epitope discovery from natural peptide libraries. *J. Immunol.* **169**, 837–846 (2002).
25. Wolf, Y., Grishin, N. & Koonin, E. Estimating the number of protein folds and families from complete genome data. *J. Mol. Biol.* **299**, 897–905 (2000).
26. Orengo, C., Michie, A., Jones, S., Swindells, M. & Thornton, J. CATH—a hierarchic classification of protein domain structures. *Structure* **5**, 1093–1108 (1997).
27. Harrison, A. *et al.* Recognising the fold of a protein structure. *Bioinformatics* **19**, 1748–1759 (2003).
28. Coulson, A. & Moult, J. A unifold mesofold and superfold model of protein fold use. *Proteins* **46**, 61–71 (2002).
29. Chothia, C. Proteins. One thousand families for the molecular biologist. *Nature* **357**, 543–544 (1992).
30. Zhang, C.-T. Relations of the number of protein sequences families and folds. *Protein Eng.* **10**, 757–761 (1997).
31. Service, R. Structural genomics, round 2. *Science* **307**, 1554–1557 (2005).
32. Koch, M. & Waldmann, H. Protein structure similarity clustering and natural product structure as guiding principles in drug discovery. *Drug Discov. Today* **10**, 471–483 (2005).
33. Irback, A., Peterson, C. & Potthast, F. Evidence for nonrandom hydrophobicity structures in protein chains. *Proc. Natl. Acad. Sci. USA* **93**, 9533–9538 (1996).
34. Pande, V.S., Grosberg, A.Y. & Tanaka, T. Nonrandomness in protein sequences: evidence for a physically driven stage of evolution? *Proc. Natl. Acad. Sci. USA* **91**, 12972–12975 (1994).
35. Baud, F. & Karlin, S. Measures of residue density in protein structures. *Proc. Natl. Acad. Sci. USA* **96**, 12494–12499 (1999).
36. Yang, A.S. & Honig, B. An integrated approach to the analysis and modeling of protein sequences and structures. III. A comparative study of sequence conservation in protein structural families using multiple structural alignments. *J. Mol. Biol.* **301**, 691–711 (2000).
37. Nettles, K. & Greene, G. Ligand control of coregulator recruitment to nuclear receptors. *Annu. Rev. Physiol.* **67**, 309–333 (2005).
38. Caponigro, G. *et al.* Transdominant genetic analysis of a growth control pathway. *Proc. Natl. Acad. Sci. USA* **95**, 7508–7513 (1998).

39. Gudkov, A.V. *et al.* Cloning mammalian genes by expression selection of genetic suppressor elements: association of kinesin with drug resistance and cell immortalization. *Proc. Natl. Acad. Sci. USA* **91**, 3744–3748 (1994).
40. Schmelzl, B. & Geli, M. An efficient genetic screen in mammalian cultured cells. *EMBO Rep.* **3**, 682–687 (2002).
41. Parissi, V. *et al.* Functional interactions of human immunodeficiency virus type 1 integrase with human and yeast HSP60. *J. Virol.* **75**, 11344–11353 (2001).
42. de Soultrait, V. *et al.* A novel short peptide is a specific inhibitor of human immunodeficiency virus type 1 integrase. *J. Mol. Biol.* **318**, 45–58 (2002).
43. Parish, C.A. *et al.* Broad-spectrum antimicrobial activity of hemoglobin. *Bioorg. Med. Chem.* **9**, 377–382 (2001).
44. Nibbering, P.H. *et al.* Human lactoferrin and peptides derived from its N terminus are highly effective against infections with antibiotic-resistant bacteria. *Infect. Immun.* **69**, 1469–1476 (2001).
45. Kieffer, A. *et al.* The N- and C-terminal fragments of ubiquitin are important for the antimicrobial activities. *FASEB J.* **17**, 776–778 (2003).
46. Short, J. Recombinant approaches for accessing biodiversity. *Nat. Biotechnol.* **15**, 1322–1323 (1997).
47. Ladner, R.C. & Ley, A.C. Novel frameworks as a source of high-affinity ligands. *Curr. Opin. Biotechnol.* **12**, 406–410 (2001).
48. Binz, H., Amstutz, P. & Pluckthun, A. Engineering novel binding proteins from nonimmunoglobulin domains. *Nat. Biotechnol.* **23**, 1257–1268 (2005).
49. Szymkowski, D. Creating the next generation of protein therapeutics through rational drug design. *Curr. Opin. Drug Discov. Devel.* **8**, 590–600 (2005).
50. Silverman, J. *et al.* Multivalent avimer proteins evolved by exon shuffling of a family of human receptor domains. *Nat. Biotechnol.* **23**, 1556–1561 (2005).
51. Chirino, Ary, M.L. & Marshall, S. Minimizing the immunogenicity of protein therapeutics. *Drug Discov. Today* **9**, 82–90 (2004).
52. Gonzales, N.R., De Pascalis, R., Schlom, J. & Kashmiri, S.V. Minimizing the immunogenicity of antibodies for clinical application. *Tumour Biol.* **26**, 31–43 (2005).
53. Ernst, J.T., Becerril, J., Park, H.S., Yin, H., & Hamilton, A. Design and application of an α-Helix-mimetic scaffold based on an oligoamide-foldamer strategy: antagonism of the Bak BH3/Bcl-xL complex. *Angew. Chem. Int. Ed. Engl.* **42**, 535–539 (2003).
54. Park, H.S., Lin, Q., & Hamilton, A. Modulation of protein-protein interactions by synthetic receptors: design of molecules that disrupt serine protease-proteinaceous inhibitor interaction. *Proc. Natl. Acad. Sci. USA* **99**, 5105–5109 (2002).
55. Pazos, F. & Sternberg, M.J. Automated prediction of protein function and detection of functional sites from structure. *Proc. Natl. Acad. Sci. USA* **101**, 14754–14759 (2004).
56. Bray, B. Large-scale manufacture of peptide therapeutics by chemical synthesis. *Nat. Rev. Drug Discov.* **2**, 587–593 (2003).
57. Liu, S. *et al.* Different from the HIV fusion inhibitor C34, the anti-HIV drug Fuzeon (T-20) inhibits HIV1 entry by targeting multiple sites in gp41 and gp120. *J. Biol. Chem.* **280**, 11259–11273 (2005).
58. Matthews, T. *et al.* Enfuvirtide: the first therapy to inhibit the entry of HIV-1 into host CD4 lymphocytes. *Nat. Rev. Drug Discov.* **3**, 215–255 (2004).
59. Wheeler, D. *et al.* Safety, tolerability, and plasma pharmacokinetics of high-strength formulations of enfuvirtide (T-20) in treatment-experienced HIV-1-infected patients. *J. Clin. Virol.* **30**, 183–190 (2004).
60. Rappocciolo, E. Antimicrobial peptides as carriers of drugs. *Drug Discov. Today* **9**, 470 (2004).
61. Cudic, P. *et al.* Complexation of peptidoglycan intermediates by the lipoglycodepsipeptide antibiotic ramoplanin: minimal structural requirements for intermolecular complexation and fibril formation. *Proc. Natl. Acad. Sci. USA* **99**, 7384–7389 (2002).
62. Cudic, M. *et al.* Development of novel antibacterial peptides that kill resistant isolates. *Peptides* **23**, 2071–2083 (2002).
63. Cudic, M., Ertl, H.C. & Otvos, L., Jr. Synthesis, conformation and T-helper cell stimulation of an O-linked glycopeptide epitope containing extended carbohydrate side-chains. *Bioorg. Med. Chem.* **10**, 3859–3870 (2002).
64. Cudic, M., Lockatell, C.V., Johnson, D.E. & Otvos, L., Jr. *In vitro* and *in vivo* activity of an antibacterial peptide analog against uropathogens. *Peptides* **24**, 807–820 (2003).
65. Cudic, M. & Otvos, L., Jr. Intracellular targets of antibacterial peptides. *Curr. Drug Targets* **3**, 101–106 (2002).
66. Cudic, P. *et al.* Functional analysis of the lipoglycodepsipeptide antibiotic ramoplanin. *Chem. Biol.* **9**, 897–906 (2002).
67. Lewis, R. & Garcia, M. Therapeutic potential of venom peptides. *Nat. Rev. Drug Discov.* **2**, 790–802 (2003).

© 2006 Nature Publishing Group http://www.nature.com/naturebiotechnology