

Exploring Biochemistry and Cellular Biology with Protein Libraries

Brett E. Howard*

Department of Molecular Biology and Biochemistry, University of California, Irvine Irvine, CA 92697-2025, USA

Abstract

Polypeptide libraries solid a broad web for outlining protein and binding supermolecule specificities. Additionally to uncovering rules for molecular recognition, the binding preferences and practical cluster tolerances from such libraries will reveal mechanisms underlying organic chemistry and cellular processes. Ligands obtained from supermolecule libraries can even provide pharmaceutical lead compounds and even reagents to more explore cell biology. Here, we review selected recent samples of supermolecule libraries demonstrating these principles. Specifically, we tend to specialise in combinatorial libraries composed of randomized peptides or variations of one super molecule. The characteristics of assorted techniques for library constructions and screening are in short surveyed.

Keywords: Mutagenesis; Phage display; Ribosome display; Photo-crosslinking; SPOT-synthesis

Corresponding author: Brett E. Howard

✉ grogandw@email.uc.edu

Department of Molecular Biology and Biochemistry, University of California, Irvine Irvine, CA 92697-2025, USA

Citation: Howard BE (2021) Exploring Biochemistry and Cellular Biology with Protein Libraries. Mol Enzy Drug Targ, Vol. 8 No. 4: 107.

Received: 27-Jul-2022, Manuscript No. IPMEDT-22-12913; **Editor assigned:** 29-Jul-2022, Preqc No. IPMEDT-22-12913; **Reviewed:** 12-Aug-2022, QC No. IPMEDT-22-12913 **Revised:** 17-Aug-2022, Manuscript No. IPMEDT-22-12913 (R); **Published:** 24-Aug-2022, DOI: 10.36648/2572-5475-08.04-107

Introduction

Biology involve some facet of molecular recognition – one molecule (often a protein) specifically binding to a different. The foundations for molecular recognition stay unpredictable. Current technologies, for example, sometimes cannot accurately predict binding partners from sequence data. Supermolecule libraries, diverse collections of polypeptides, supply a strong technique to fillin the gaps in our prophetical talents, like identification of binding partners, elucidation of key contributors to recognition or getting new binding interactions [1]. This review focuses on the primary aspects of supermolecule libraries – identification of binding partners and dissection of key residues at receptor-ligand interfaces. Additionally, we focus upon recent examples that illustrate the facility of supermolecule libraries to elucidate principles of molecular recognition and enzymatic contact action. Because of house limitations, some excellent contributions to the sphere are omitted, and that we apologize to those researchers. Such natural libraries illustrate two principles considered by every practitioner of protein library techniques. First, antibodies are quite large (150+ KD) and cannot be fully randomized at every position. Second, different antibody formats, such as the decavalent IgM or bivalent IgG, offer advantages and disadvantages that affect the outcome of the attempt to identify binding partners. For example, the decavalent format of IgM can

provide a Velcro or avidity effect to boost the apparent binding affinity of a weak initial binding partner; however, the absolute affinity of the binding partner obtained, when removed from the decavalent context of an IgM scaffold, can be quite weak [2]. Although many combinatorial protein library formats have been employed, libraries are constructed either by chemical or biological synthesis. Among the many examples of biologically derived library methods, one of the most common is phage display whereby members of a peptide library are fused to (“displayed” on) bacteriophage coat proteins. The key feature of these biological libraries is a direct link between the phenotype (the displayed peptide) and the genotype (DNA encoding the peptide) that allows for selection and amplification of desired peptides. Synthetic peptide libraries on the other hand do not require cellular components for construction, but instead rely on chemical techniques for the synthesis of peptides, through SPOT, parallel and split-pool synthesis [3].

Discussion

Phage Display For the display and screening of massive combinatorial peptide libraries with diversities up to 10¹² members in size, phage display has evolved into a potent technique. The well-known non-lytic M13 filamentous phage is used as a scaffold in the construction of most peptide libraries. The lytic phages, T4, and T7 have also been employed as display

scaffolds. An infected helper phage that is also co-infected packages the fusion protein into phage particles and supplies all of the other components required for viral assembly. A similar method of changing one copy of a coat protein gene fused to the desired protein can also include the rescue and assembly of the virions by a particular strain of *E. coli*, as will be discussed below. However, the idea of changing a small number of copies of the viral coat protein offers a more reliable framework for library development and screening, which is crucial for bigger polypeptides [4]. A filamentous bacteriophage is what M13 is. A viral coat made up of five viral coat proteins surrounds a single-stranded closed circular DNA genome. *E. coli* is infected with the lytic double-stranded DNA virus known as Bacteriophage T7, which has been used as a scaffold for phage display. The gene 10 viral capsid protein, which is organised into hexamers and pentamers, makes up the whole icosahedral particle in which T7 DNA is packaged. The gene for the 10 capsid protein produces a full-length 10B protein (397 residues), as well as a truncated 10A version (344 residues). Fusion proteins are seen as C-terminal fusions to the 10B main coat protein, with smaller polypeptides (50 residues) showing up in low copy numbers, typically between 0.1 and 1 copies per phage, while bigger polypeptides (1200 residues) are shown in high copy numbers (415 copies per phage). The variety of proteins that can be fused to and displayed on the M13 phage may be constrained by their requirement to tolerate extrusion through the bacterial inner membrane during viral assembly [5]. Recently, this T7 phage display strategy has been used to successfully screen and choose RNA-binding proteins.

Yeast two-hybrid

Unlike yeast two-hybrid libraries, which are completely *in vivo*, phage display, which physically attaches a peptide to the surface of the phage for *in vitro* selection and screening (usually *Saccharomyces cerevisiae* cells). The research of protein-protein interactions is best served by this type of library screening. Due to an enzyme alteration in the adenine biosynthesis pathway, ADE2 transcription has the potential to cause a hue shift [6]. The reverse two-hybrid and the three-hybrid (three binding partners) are variations of this approach.

Displayed ribosomes

Ribosome display depends on a stable ternary complex that forms a direct connection between the peptide being shown and the mRNA that encodes it. In this method, a T7 promoter and a ribosome binding region, respectively, start *in vitro* transcription and translation upstream of the library gene. The construct's absence of a stop codon stops the ribosome from separating from the mRNA template, which enables the newly generated protein to fold and stay connected to the ribosome via a C-terminal linker. The big ribosomal subunit's N-glycosidic bond is hydrolysed by the ricin A chain; glycoside hydrolysis hinders the separation of the developing protein and mRNA from the ribosome. Ribosome display relies on a direct linkage between the peptide being displayed and the mRNA that encodes it via a stable ternary complex. In this technique, *in vitro* transcription and translation are initiated upstream of the library gene by a T7 promoter and ribosome binding sequence, respectively [7]. The lack of

a stop codon terminating the construct prevents dissociation of the ribosome from the mRNA template, allowing the newly synthesized protein to fold and remain tethered to the ribosome via a C-terminal linker. The big ribosomal subunit's N-glycosidic bond is hydrolysed by the ricin A chain; glycoside hydrolysis hinders the separation of the developing protein and mRNA from the ribosome. The displayed polypeptide gene is linked to the 5' end of the RTA gene by a linker within the single ORF that makes up the ribosome-inactivation display construct. The RTA gene is followed by a C-terminal spacer that enables ricin A to depart the ribosome, fold correctly, bind, and inactivate the ribosome. Puromycin antibiotic is added to the mRNA's tail to create a more stable bond between the displayed polypeptide and its encoding mRNA. The construct's absence of a stop codon stops the ribosome from separating from the mRNA template, which enables the newly generated protein to fold and stay connected to the ribosome via a C-terminal linker. This ternary complex of the ribosome, the mRNA, and the visible protein is remarkably stable at 4 °C. The displayed polypeptide can be subjected to selections, and then the encoding mRNA can be reverse-transcribed for amplification, several rounds of selection, and ultimately sub-cloning into a vector for sequencing. Puromycin enters the ribosome (peptide transferase site) and forms a stable covalent amide bond to the developing protein as ribosomes start translating the mRNA transcripts but stall at the RNA-DNA junction. The mRNA-peptide fusion that is produced can then be used in selection experiments.

SPOT-synthesis

In a spatially addressable array, SPOT-synthesis is an effective approach for the synthesis and screening of potentially thousands of peptides and small compounds. With this method, chemical synthesis takes place at predetermined "spots" on a solid support, such as cellulose (e.g., filter paper). The chemical and mechanical stability of the connected molecule and the membrane has also been improved by using polypropylene membranes [8]. SPOT synthesis of peptide libraries involves dotting active amino acids onto a functionalized membrane at specified locations to create a "spot" where miniscale coupling reactions can take place. Each additional amino acid is supplied directly to the designated places in the traditional solid phase peptide synthesis process. To enable cleavage and recovery of the produced product, linkers may also be added to the functionalized membrane.

Protein Library Techniques for Mapping Molecular Recognition

Finding important residues that add energy to receptor-ligand interactions can be done effectively via single point, site-directed mutagenesis. The methods for site-directed mutagenesis, protein expression, and purification, however, can be time-consuming [9]. The identification of important residues at the receptor-ligand interaction can be accelerated by using additional techniques like homolog shotgun scanning and alanine analysis. These techniques simultaneously examine the roles that various amino acids play in receptor binding. It is possible to determine which residues are essential for molecular recognition by high throughput assessment of the energetic contributions made by

various side chain residues. Sidechain atoms past the carbon were terminated by changing the surface residues of hGH to alanine. Contributions to binding by atoms past the side chain-carbon were evaluated by comparing the binding of wild-type hGH with the hGH alanine mutant. It was discovered that a tiny group of hGH residues—dubbed the "functional epitope"—contributed the majority of the binding energy to the interaction with hGHbp. One million hGH mutants were created via phage display to better understand the structure of the hGH functional epitope. The consensus residues highlighted by alanine mutagenesis analysis were similar to the residues indicated in this protein library experiment. Maps of protein-small molecule interactions have also been created using shotgun alanine scanning. The streptavidin-biotin interaction is a paradigm for high affinity receptor-ligand interactions and has femtomolar dissociation constant. By using phage-displayed alanine shotgun scanning, streptavidin residues essential for biotin binding were identified. Shotgun scanning residues allowed researchers to expand their knowledge of the streptavidin-biotin interaction to include residues that weren't in direct contact with the biotin. The Fab2C4 antibody's antigen binding site, which binds to the extracellular domain of the ErbB2 oncogene, has been identified using phage display with shotgun alanine scanning and protein structural analysis. Aspartic acid and glutamic acid are two examples of homologous residues that share the same structure, charge, and binding-critical side chain geometries as the wild-type residue. A Fab2C4 functional epitope was discovered by homolog shotgun scanning, which adds specificity to the interaction with the antigen. However, alanine shotgun scanning discovered a wider functional epitope that encompassed the homolog shotgun scanning epitope. Both homolog and alanine shotgun scanning techniques were used to find side chains that indirectly contributed to the binding interaction as demonstrated by the identification of the Fab-antigen crystal structure. Protein libraries can reveal the function of side chains buried at receptor-ligand interactions as well as point out residues important for structure creation. For instance, the impact of certain residues on the binding affinity of a minimum binding protein motif has also been studied. A peptide library based on this turn sequence within a coiled coil stem loop miniprotein, which was previously discovered to be essential for IL-5 receptor binding, was displayed. Through the selection for inclusion of mutant P8 into a phage composed of wild-type coat proteins given by Tran's infection with helper phage, three functional epitopes were discovered. First, it is likely that a basic patch near the P8 C-terminus interacts with the negatively charged DNA that runs through the virus's core. Near the N- and C-termini of P8, two more hydrophobic functional epitopes were discovered that could lock into one another from nearby P8 molecules, resulting in a hard viral coat held together by spot weld. This work demonstrates the potential of robust library techniques to address topics in a comprehensive manner that conventional single point methods might find too challenging to take on, such as assessing relative contributions by each side chain in a protein to functionality. The previous section's methodologies for mapping receptor ligand interactions have likely found their most significant use in identifying the unique properties of protein domains, which can fold independently of a longer polypeptide chain. Three to nine

amino acid long peptide sequences. In eukaryotic cells, many protein-protein interactions necessitate such interaction modules and particular binding to cognate ligands. To find peptide-binding motifs for different protein domains, phage display and other protein library approaches have been applied. As an example of *in vitro/in vivo* "convergent evolution," peptide ligands obtained from *in vitro* selection of these libraries have frequently been used to find peptides that are identical to or similar in sequence to the *in vivo* interacting protein.

EH Domains

About 100 amino acids in length, the Eps homology (EH) domain regulates protein-protein interactions that regulate endocytosis, actin remodelling, and intracellular signal transduction. The epidermal growth factor receptor substrate Eps15 and the associated protein Eps15R both contain three copies of the EH domain as a motif in their N-termini. According to the results of phage display tests, the majority of EH domains bind to peptides of the amino acid sequence NPF. There are three distinct kinds of binding peptides. These findings show how polypeptide libraries can be used to find peptides that have the same affinity, specificity, and primary sequence as proteins with which they naturally interact. Particular membrane proteins' C-termini can be bound by PDZ domains [10]. Three to eight C-terminal residues provide selectivity, and the C-terminal carboxylate fits into a highly conserved hydrophobic pocket on the PDZ domain surface. Based on various C-terminal ligand binding sequences identified using polypeptide library approaches, PDZ domains have been divided into four classes. Class III ligands were identified using a yeast two-hybrid experiment, as shown by the PDZ Mint-1's affinity for the Ca²⁺ channel pore-forming 1b subunit. Phage display libraries, which typically display peptides as N-terminal coat fusions, have been of limited utility for identifying PDZ specificity because PDZ domains bind to particular C-terminal sequences of target proteins or dimerism to other PDZ domains.

Conclusion

Utilising library techniques (such as yeast two-hybrid, phage display, and identifying the protein's cellular binding partners (amongst others) domains. Frequently, the peptide motifs derived from these library procedures share affinities, specificities, and similar to the endogenous interacting protein's basic structure. Consequently, a useful method to understand the roles and Proteome interactions can be accomplished in two steps. First, target protein peptide ligands are extracted from extensive collections of possible ligands. In addition, database mining ligands found in peptide libraries can indicate proteins from a sequenced genome that may interact. Since the members of the peptide library are solely focused on besides receptor ligand affinity, local protein levels can also affect Concentration may be a key factor in determining whether a specific interaction occurs. Improved high throughput bioinformatics and structure determination methods, new tools to rationalize the results from enzyme engineering could be gained. Such efforts are expected to uncover new mechanistic understanding for how proteins function *in vitro* and in cells.

Acknowledgement

None

Conflict of interest

No conflict of interest

References

- 1 Panni S, Dente L, Cesarini G (2002) *In vitro* evolution of recognition specificity mediated by SH3 domains reveals target recognition rules. *J Biol Chem* 277: 21666-21674.
- 2 Paoluzi S, Castagnoli L, Lauro I, Salcini AE, Coda L, et al. (1998) Recognition specificity of individual EH domains of mammals and yeast. *EMBO J* 17: 6541-6550.
- 3 Quilliam LA, Lambert QT, Mickelson-Young LA, Westwick JK, Sparks AB et al. (1996) Isolation of a NCK-associated kinase, PRK2, an SH3-binding protein and potential effector of Rho protein signaling. *J Biol Chem* 271: 28772-28776.
- 4 Reineke U, Volkmer-Engert R, Schneider-Mergener J (2001) Applications of peptide arrays prepared by the SPOT-technology. *Curr Opin Biotechnol* 12: 59-64.
- 5 Rickles RJ, Botfield MC, Weng Z, Taylor JA, Green OM, et al. (1994) Identification of Src, Fyn, Lyn, PI3K, and Abl SH3 domain ligands using phage display libraries. *EMBO J* 13: 5598-5604.
- 6 Roberts RW (1999). Totally *in vitro* protein selection using mRNA-protein fusions and ribosome display. *Curr Opin Chem Biol* 3: 268-273.
- 7 Daugherty PS, Olsen MJ, Iverson BL, Georgiou G (1999) Development of an optimized expression system for the screening of antibody libraries displayed on the Escherichia coli surface. *Protein Eng* 12: 613-621.
- 8 Dente L, Vetriani C, Zucconi A, Pelicci G, Lanfrancone L, et al. (1997) Modified phage peptide libraries as a tool to study specificity of phosphorylation and recognition of tyrosine containing peptides. *J Mol Biol* 269: 694-703.
- 9 Di Fiore PP, Pelicci PG, Sorkin A (1997) A novel protein-protein interaction domain potentially involved in intracellular sorting. *Trends Biochem Sci* 22: 411-413.
- 10 Eck MJ, Shoelson SE, Harrison SC (1993) Recognition of a high-affinity phosphotyrosyl peptide by the Src homology-2 domain of p56lck. *Nature* 362: 87- 91.