CIS display, a DNA-based *in vitro* selection technology for therapeutic peptides

PASCALE MATHONET[1], AVGOUSTA IOANNOU[2], JASON BETLEY[2], CHRIS ULLMAN[1]*

*Corresponding author
1. Isogenica Ltd Chesterford Research Park Little Chesterford, CB10 1XL, United Kingdom
2. Illumina Cambridge Ltd Chesterford Research Park Little Chesterford, CB10 1XL, United Kingdom

KEYWORDS

Keywords: CIS display, peptide discovery, next generation sequencing.

ABSTRACT

CIS display is a DNA-based in vitro display technology that enables the display and selection of peptides and proteins from extremely large libraries. The system uses a natural process to couple the phenotype and genotype and allow selection or maturation of high affinity peptide binders against targets. In this article, we review the CIS display technique and its integration with next-generation sequencing and bioinformatics to provide a data-rich output to facilitate rational design of biologics and small molecules and also for intellectual property protection.

CIS display is a recombinant DNA-based technique that links an expressed peptide or protein library to its own DNA sequence without the need for cloning. The activity of a bacterial replication initiator protein, RepA, is core to the technology. This protein is an *E. coli* plasmid replication initiator protein that has the unusual characteristic of exclusively binding to the same DNA template from which it was derived – "*cis* activity". CIS display has distilled the essential components of this natural system so that RepA and its genetic control elements are carried on a short linear DNA sequence that can be readily generated by polymerase chain reaction (PCR). These controls elements are the CIS element and the *ori* region which terminate the transcription complex so that the nascently expressed RepA protein can be loaded onto the *ori* region on its own template. By encoding peptide or protein libraries fused to RepA, the expressed library peptide is attached to its coding DNA (Figure 1). The DNA code can be subsequently sequenced to reveal the peptide sequence (1).

CIS display is a recombinant procedure that requires components of the bacterial transcription and translation machinery for operation; however the process can be performed *ex cellulo*, unlike other technologies such as phage display, which requires replication inside bacteria (1-2). Therefore, CIS display can use bacterial cell lysates in a purely acellular fashion which overcomes the limitations of other techniques that need DNA to be transferred into cells, which is an inefficient process and sets a ceiling on library sizes. In practice, this means that CIS display is simple to operate and larger libraries can be rapidly generated and screened, thereby shortening the time from library design to hit identification. It is feasible to generate library sizes in excess of 10^{13} different peptides linked to their own code within a couple of days and screened within a couple of weeks. Peptides fused to RepA are exclusively and efficiently linked to their own DNA: in tests using a peptide tag, over 40 percent of the DNA templates can be recovered using an anti-peptide antibody (unpublished data). CIS display is also adaptable for the display of different peptide topologies and proteins up to 90kDa, including antibody fragments and protein scaffolds such as the Centyrin scaffold (3). Other acellular technologies use a more labile RNA template or require compartmentalisation of the DNA into water in oil emulsions (4-5).

For such large libraries of sequences, conventional screening is impossible, therefore cycles of "panning" or selection against the target of interest are necessary in order to enrich the binding

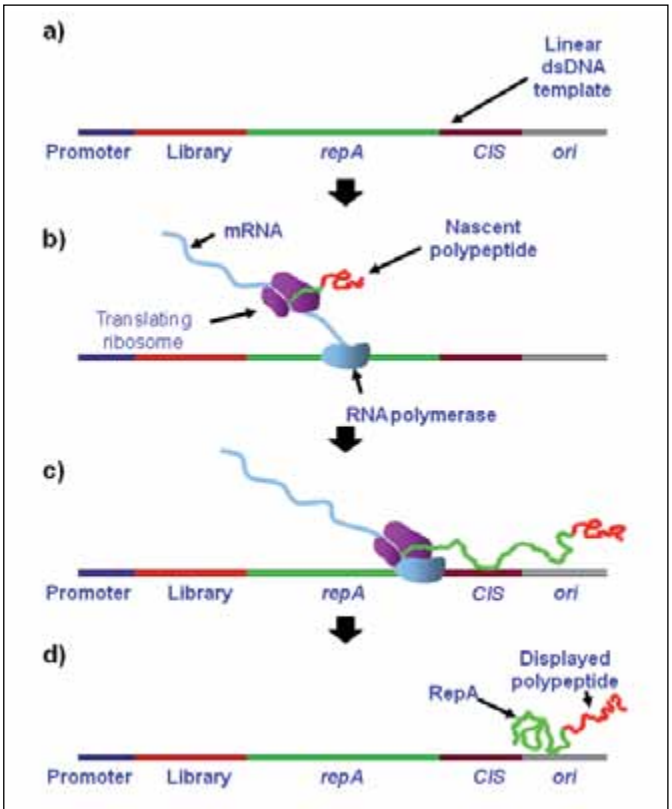

Figure 1. The CIS display technique. a) A linear double stranded DNA template is generated by PCR to encode a peptide/protein library fused to the 5' region of RepA; b) The DNA is transcribed and translated in an E. coli lysate preparation; c) The transcription process is stalled and terminated at the transcription termination sequence at CIS. This allows RepA to locate its own binding site on the same template from which it was transcribed; d) Peptides are attached to their own DNA sequence through RepA binding. The peptide sequence can be determined by DNA sequencing of the template.

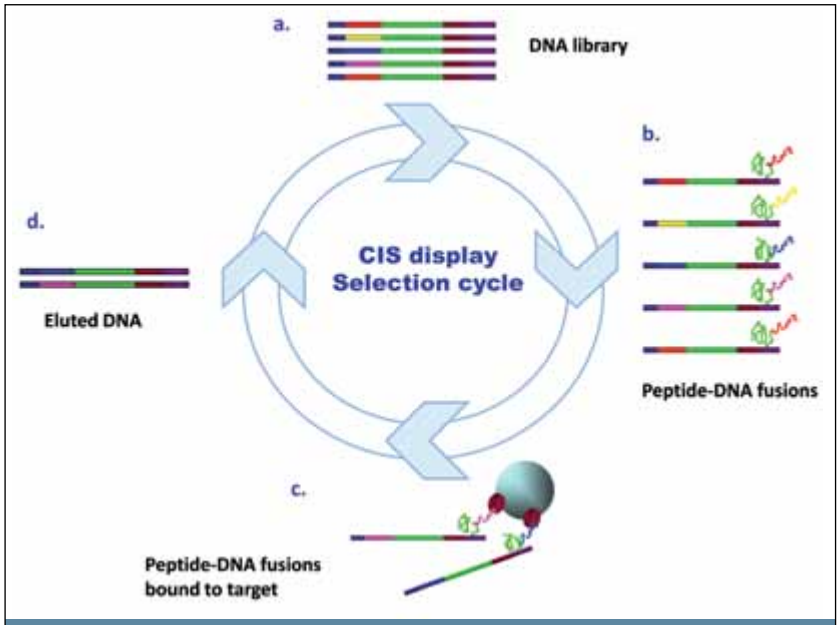

Figure 2. Ligand discovery using CIS display. (a) A DNA library is created by PCR which is then converted to a CIS display library of peptide-DNA fusions by transcription and translation in vitro; (b), (c) The CIS display library is exposed to a target, usually immobilized on magnetic beads, and high affinity peptides (carrying their coding DNA) bind. The bound DNA is eluted (d) and amplified for the next cycle. Typically four sequential cycles of enrichment are required.

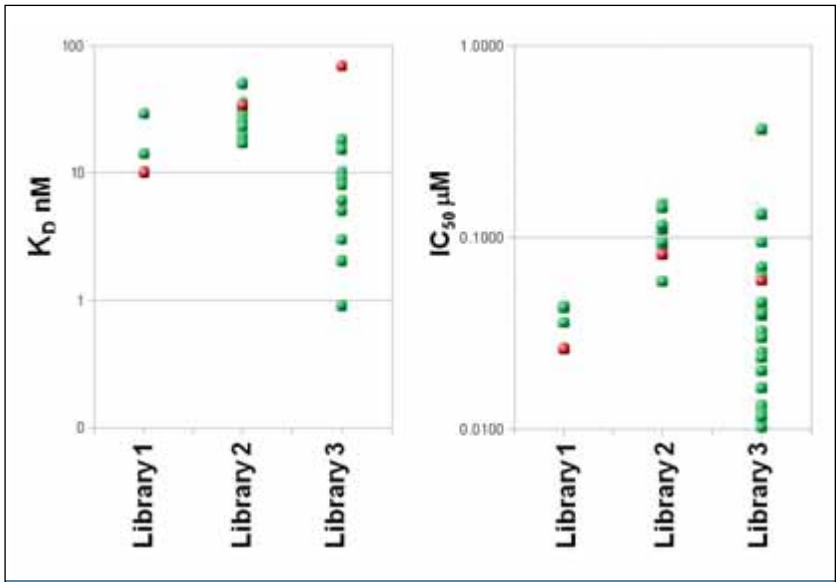

Figure 3. An example of a peptide maturation using CIS display. The red dots represent primary hits following four panning cycles against a protease. Each red dot represents a single clone. Maturation libraries were designed by limited randomisation of the sequences represented by each red dot and 4 rounds of further selection were performed under more stringent conditions ("maturation"). The green dots represent the clones selected in the maturation. In Libray 1 no improvements over the original sequence were found for both affinity or activity. However, maturation of the original clone in Library 3 produced the best candidates.

population (Figure 2). For CIS display, it is usual for four rounds of panning to be performed prior to screening.

Using 16-mer and 18-mer peptide libraries of random sequences, hits with single digit nanomolar affinities have been identified against pharmaceutically relevant targets. This supports the correlation between increased library size and the interrogation of greater chemical space leading to improved binders against a target (6).

However, even a technology that displays greater than 10^{13} peptides cannot fully capture the diversity of a random 16-mer random library which will contain in excess of 10^{20} different sequences. As a result, primary hits may be identified that have the potential to be further matured for improved activity.

These peptide sequences serve as frameworks for subsequent, focused, maturation libraries which can be designed on a fixed consensus sequence or random mutations of the original sequence to search a more restricted chemical space (7). Such maturation libraries often lead to improvements in activity by identifying closely related sequences that were not isolated in the original screen. The combination of library design and stringent selection conditions can often lead to vast improvements in affinity (8-9). Occasionally, no improvements can be made on a particular sequence and further candidates have to be optimised. These may appear to be outliers in the original screen but can be improved using maturation beyond the limits of the best clone from the primary selection (Figure 3).

This phenomenon can be explained by the stochastic nature of the sampling of the original library in which clones may be selected that are the best examples of a particular sequence motif, whereas others may benefit from further exploration of defined sequence space to produce better candidates that were not represented in the original sampling of the library.

Simple peptides with low or subnanomolar affinities have been isolated using CIS display to proteins such as human NGF, thrombin and HIV-1 gp41 (Isogenica Ltd., unpublished data) and low picomolar affinities for larger folded protein domains are achievable using CIS display.

However, despite their potential for high affinity and activity, peptides are notoriously labile *in vivo* and this presents a major barrier for the use of peptides as drugs. Knowledge of the protease sensitive regions enables stabilisation by chemical methods using unnatural amino acids or peptide bond surrogates (10). However, there are advantages in finding natural alternatives and methodologies.

CIS display, in conjunction with an protease challenge, has selected more robust peptides. Interestingly, these peptides were not only more resistant to the enzyme used in the selection but had a wider spectrum of resistance and improved *in vivo* stability (11).

Recently, the entire CIS display selection output has been determined using Illumina's next-generation sequencing technology (12). This provides a data rich output which, using bioinformatics, can provide a detailed view of the dynamics of the clonal populations in the enrichment process of CIS display. It is possible to use Illumina methods to sequence millions, even hundreds of millions of clones, each representing individual peptides, and cluster these into hundreds of families of related sequences thereby providing an "*in silico* screen" for enriched binding populations (Figure 4). These enriched populations may have different modes of binding to a target and peptide candidates from each of these clusters can be chemically synthesised and screened *in vitro* to activity. Using this approach, binding peptides that may have been lost due to poor expression or excluded, by chance, during conventional screening can be identified, synthesised and evaluated.

This improves the identification of consensus motifs that can be

converted into small molecule pharmacophores.
To illustrate this process, a selection was performed, using a peptide library of 16 random residues, against an antibody (anti-FLAG M2) that binds a short peptide epitope (DYKDDDDK), known as the FLAG tag (13).
After 4 successive rounds of selection, the DNA output was sequenced using an Illumina Genome Analyzer IIx. Greater than six million sequences were obtained, translated into peptide sequences and filtered into clones appearing more than fifty times. Three million sequences remained, which represented 450 different peptides in six clusters.
The HMM profile of the most abundant clusters, highlighted the known consensus features of the FLAG motif YKxxD (1, 13-15) but also showed that Trp (W) is favoured within the epitope as well as Asp, Tyr, Phe (Figure 5). This validates and extends previous observations by phage display, CIS display and ribosome display in which positions 4 and 5 are known to be variable. However, Trp has not previously been considered a major component of binding, despite the fact that the residue is enriched at protein-protein interfaces (16).

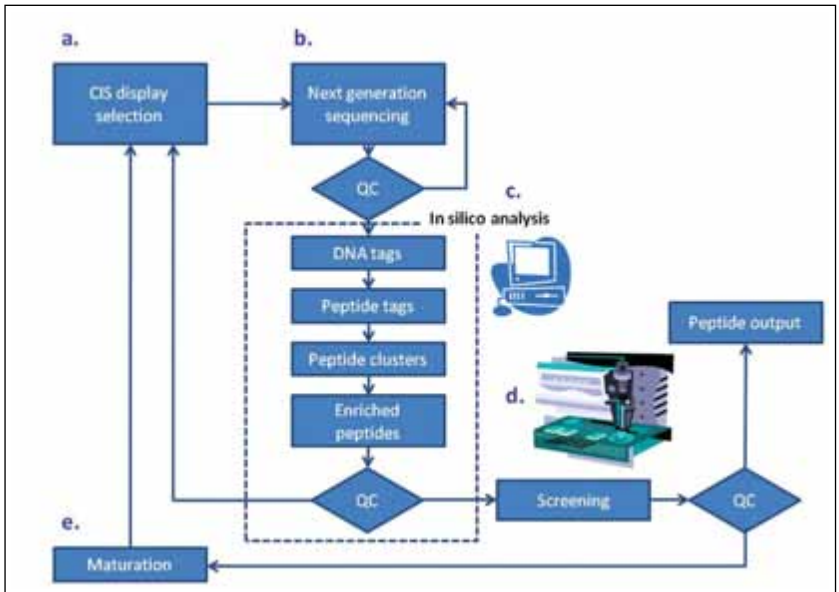

Figure 4. Integration of CIS display into a "next generation" approach. The DNA output from CIS display (a) can be decoded using next generation sequencing to identify peptide sequences (b); These can be clustered into families in silico (c) and representative members synthesised chemically for screening (d); If the desired level of activity is not achieved, the sequence space of primary hits can be quickly searched for improved performance in maturation (e) using CIS display.

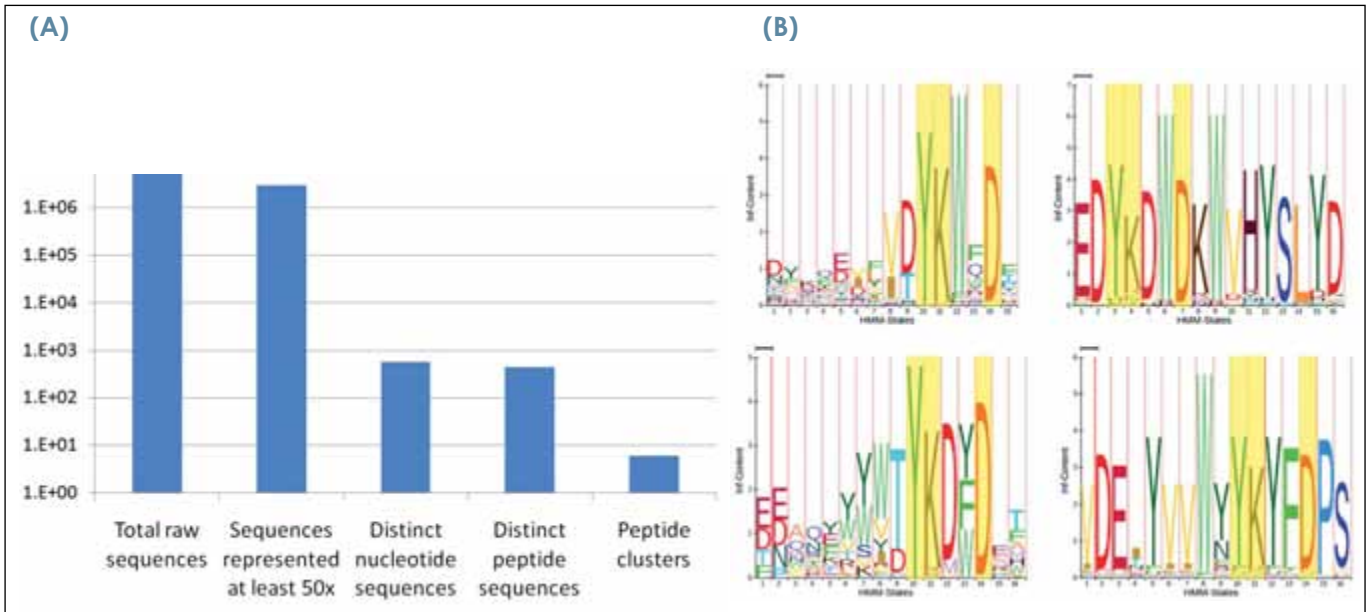

Figure 5. Results obtained from the high-throughput sequencing of peptides selected for anti-FLAG affinity. (A) The number of sequences remaining after each analysis step (B). Four top-ranked peptide clusters identified highlight the known FLAG-epitope consensus motif.

In summary, CIS display is a powerful peptide discovery and maturation platform that can be used to isolate peptides possessing high affinity, activity and stability against therapeutically relevant targets.
Recent advances have applied CIS display with Illumina next generation sequencing methods to provide valuable information for epitope mapping and for the rational design of biologics and small molecule pharmacophores.

ACKNOWLEDGMENTS

The authors thank Dr. Simon Andrews (The Babraham Institute, Cambridge, United Kingdom) for his help with the bioinformatics, and all current and past staff members at Isogenica for their contribution.

REFERENCES AND NOTES

1. R. Odegrip et al., *Proc Natl Acad Sci U S A*, **101**, p. 2806 (2004).
2. G.P. Smith, V.A. Petrenko, *Chem Rev.*, **97**, p. 391 (1997).
3. J.M. Reichert, N. Jacob et al., *MAbs*, **1**, p. 190 (2009).
4. A. Rothe, R.J. Hosse et al., *European Biopharmaceutical Review*, **53**, p. 72 (2010).
6. M.M. Ling, *Comb Chem High Throughput Screen*, **6**, p. 421 (2003).
7. S.W. Millward, S. Fiacco et al., *ACS Chem Biol.*, **2**, p. 625 (2007).
8. C. Zahnd et al., *J Biol Chem.*, **279**, p. 18870 (2004).
9. A.M. Levin, G.A. Weiss, *Mol Biosyst*, **2**, p. 49 (2006).
10. J. Vagner, H. Qu et al., *Curr Opin Chem Biol.*, **12**, p. 292 (2008).
11. B. Eldridge et al., *Protein Eng Des Sel.*, **22**, p. 691 (2009).
12. D.R. Bentley et al., *Nature*, **456**, p. 53 (2008).
13. J.W. Slootstra, D. Kuperus et al., *Mol Divers.*, **2**, p. 156 (1997).
14. E. Osada, Y. Shimizu et al., *J Biochem.*, **145**, p. 693 (2009).
15. R.M. Miceli, M.E. DeGraaf et al., *J Immunol Methods*, **167**, p. 279 (1994).
16. S. Gong et al., *Biochem Soc Trans*, **37**, p. 727 (2009).