

Design and Production of Specifically and with High Affinity Reacting Peptides (SHARP®-s)

Jan C. Biro

Homulus Foundation, 88 Howard, #1205, 94 105 CA, USA

Abstract: A partially random target selection method was developed to design and produce affinity reagents (target) to any protein query. It is based on the recent concept of Proteomic Code which suggests that significant number of amino acids in specifically interacting proteins are coded by partially complementary codons. It means that the 1st and 3rd residues of codons coding many co-locating amino acids are complementary but the 2nd may but not necessarily complementary: like 5'-AXG-3'/3'-CXT-5' codon pair, where X is any nucleotide. A mixture of 45 residue long, reverse, partially complementary oligonucleotide sequences (target pool) were synthesized to selected epitopes of query mRNA sequences. The 2nd codon residues were randomized. The target oligonucleotide pool was inserted into vectors, expressed and the protein products were screened for affinity to the query in Bacterial Two-Hybrid System. The best clones were used for larger-scale protein syntheses and characterization. It was possible to design and produce specific and with high affinity reacting (Kd ~100 nM) oligopeptide reagents to GAL4 query oligopeptides.

Key words: Combinatorial engineering, protein interaction, proteomic code, peptide design, methods

INTRODUCTION

According to the Human Genome Project, the estimate for the total number of genes in the human genome has been revised down from 30,000-35,000 to 20,000-25,000 (http://www.ornl.gov/sci/techresources/Human_Genome/project/20to25K.shtml).

Despite the fairly low number of genes, the best estimates for the total number of proteins encoded by the human genome (the proteome) remain anywhere from 300,000 to 1 million. Ongoing efforts to study the proteome have kept the research antibody industry flourishing. Also reaping financial benefits are motivating companies that provide reagents and devices required for antibody-related protocols. With basic science and drug developing researchers creating a tide of demand, the revenue stream from antibody-led protein hunts won't be drying up anytime soon (Antibodies, 2005).

Revenue from antibodies for therapeutic and diagnostic uses is expected to grow at an average annual growth rate of 11.5%, according to a 2005 report, "Dynamic Antibody Industry," published by the Business Communications Company (Dynamic Antibody Industry, 2005). With an estimated market of \$15 billion in 2005, revenues should reach \$26 billion by 2010.

The original procedure to induce Antibodies (Ab) is the *in vivo* immunization. The success of this polyclonal antibody production in animals is variable, often not

predictable, takes several weeks, require the use of well purified antigens (Harlow and Lane, 1988). Monoclonal antibody production uses animals only as the source of biological machinery which is responsible for the immune response (B lymphocytes). The Ab production itself takes place *in vitro* (Borrebaeck *et al.*, 1998). It makes the monoclonal Ab production cheaper, faster and more reliable. In fact, monoclonal antibodies produced by animal immunization remain the 'gold standard' of affinity reagents. They are relatively renewable, can usually be made with high specificity and affinity for their target and can be used in common biochemical assays such as Western blotting, ELISA and immunochemistry. But the traditional monoclonal antibody has its drawbacks. Its production can be challenging, time-consuming and costly.

Additional concerns associated with Ab-s which are seriously limiting their therapeutic applications are their size and their origin. Antibodies are large and complex proteins (>150 KDaltons) which are much larger than necessary for antigen recognition and binding. They are antigenic themselves and are carrying the species characteristics of their origin. Size reduction, like single-chain variable (scFv) antibody fragments (Pelat *et al.*, 2007) and "humanization" (Pichla *et al.*, 1997) help to solve these problems.

A moderate discomfort, even if not a ban, of Ab production is that we still not know exactly how

antibodies are made by the immune system and therefore, it is not possible to reproduce it *de novo*, without “borrowing” the technique of a living organism.

So there is a lot of interest in identifying novel affinity reagents that would be less expensive and quicker to produce.

Affibodies (Affibody AB of Bromma in Sweden) were among the first non-immunoglobulin-based affinity reagents. These small molecules are based on a bacterial receptor (Staphylococcus aureus protein A) and use combinatorial protein engineering to introduce random mutations in the affinity region (Nord *et al.*, 1997). Protein Z is a 58-residue three-helix bundle domain derived from Staphylococcal Protein A (SPA), which binds to the Fc portion of IgG from different species. By simultaneously randomizing 13 amino acid positions located at the two helices making up the Fc-binding face of protein Z, binding proteins (affibodies) capable of binding to desired targets have been selected by using phage display technology.

Another non-immunoglobulin-based affinity reagent that is becoming more widely used is the aptamer. Made of DNA, RNA or modified nucleic acids and typically 15-40 bases in length, aptamers have a stable tertiary structure that permits protein binding through van der Waals forces, hydrogen bonding and electrostatic interactions. Early studies showed that aptamers can be highly specific for target proteins, with the ability to distinguish between related members of a protein family (Seiwert *et al.*, 2000).

No one of the above mentioned methods is able to satisfy the emerging need to produce these reagents at a truly high-throughput scale. The estimated number of affinity reagents (antibodies or else) needed to monitor the human proteome is probably not less than the number of different proteins (including splicing variations, but not the unknown number of folding variations), which is around 0.3-1.0 millions.

A completely different approach to understand and utilize the nature of specific protein-protein interactions is the concept of the Proteomic Code.

The Proteomic Code is a set of rules by which information in genetic material is transferred into the physico-chemical properties of amino acids. It determines how individual amino acids interact with each other during folding and in specific protein-protein interactions. The Proteomic Code is part of the redundant Genetic Code (Biro, 2007). This, 25-years-old theory states, that significant number of amino acids co-located on the specifically interacting protein-protein interfaces are coded by complementary codons.

The original concept expected perfect complementarity coding of co-locating amino acids. There are hundreds of experiments from reliable laboratories which are suggesting the validity of this expectation (Biro, 2007; Heal *et al.*, 2002). However, there are many (to many) exceptions.

Recently, bioinformatical studies suggested that co-locating amino acids are coded by partially complementary codons, where the 1st and 3rd codon residues are complementary in reverse orientation, but the 2nd codon residue is not necessarily complementary. This second generation Proteomic Code may be described by the $3'-NXN-5'/5'-nXn-3'$ formula, where N and n denote complementary base pairs, while X indicates any nucleotide. A method (The Method in this text) to Design and Production of Specifically and with High Affinity Reacting Peptides (SHARP®-s) is built on this formula (Biro, 2007).

MATERIALS AND METHODS

The protein which is used to generate affinity peptides is called Query (Q) in this description and it is analog to the terms ligand, antigen or bait. Query is one protein sequence that the target protein, designed and produced with the Method, will specifically interact with. The generated affinity peptide is called Target (T) and it is analog to the names receptor, antibody or hit. Target proteins are protein sequences which are designed by the Method to specifically interact with the query protein sequence.

The query and target are expected to react with each other specifically (able to distinguish between related but not identical peptide sequences) and with high affinity (Kd is at least in μM range).

Target Oligo-Nucleotide Pool (TONP) is designed by using a Target Oligo-Nucleotide Template (TONT) which is a nucleic acid sequence containing 2/3 defined and 1/3 undefined (any) nucleotides (X). A TONT, which contains 15 undefined nucleic acid residues, (defining a pool of $4^{15} = 10^9$ different oligo-nucleotides, TONP) will be translated into the corresponding number of oligopeptides.

Expression of TONT will result in the syntheses of a large number of different oligo-peptides, called Target Oligo-Peptide Pool (TOPP). Those oligopeptides which satisfy the criteria for specific, high affinity reactions with the query protein are called SHARP®-s.

The flow of SHARP® production may comprise the following main steps (Fig. 1).

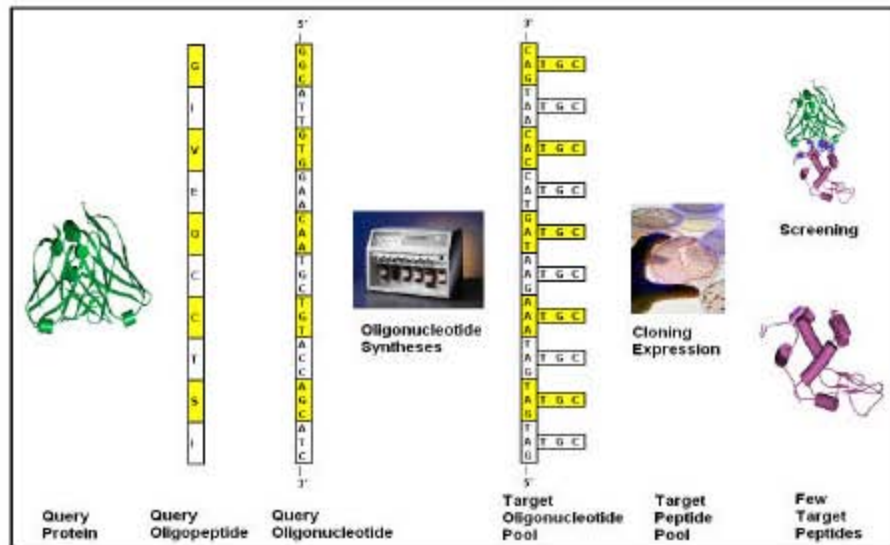


Fig. 1: Schematic illustration of design and production of specifically interacting proteins SHARP@s

1. Query selection:

- The length of query peptide is limited to about 15 amino acids, depending on the available screening method.
- Query may be part of a much larger protein. In these cases physicochemical, structural considerations might guide to select the most “promising” query sequences.
- It is necessary to know the correct coding sequence of the query. A nucleic acid sequence derived by back-translation, using any Codon Usage Frequency Table, is not satisfactory.

2. Design the target CDS:

- The CDS of target oligopeptides (TONP) is derived from the CDS of the query, using the 3'-NXN-5'/5'-nXn-3' formula or

3'-_NNN_NNN_NNN_-5' QUERY
 5'-_nXn_nXn_nXn_-3' TARGET OLIGO-NUCLEOTIDE TEMPLAT (TONT)

N and n are complementary bases and X is any nucleotide:

- X should always be the central codon residue, therefore knowing the correct translation frame is essential. Frame-shifts are not permitted.

- The limitation of CDS length is around 45 nucleotides, depending on the capacity of the expression system.

3. Additions to the target CDS:

The target CDSs (TONP) will be inserted into vectors for expression, therefore sequence additions are necessary

- Add restriction enzyme sites at both ends (choose restrictions enzymes which have no cut site in the designed target sequences).
- Add known sequences (after the start signal) which will identify the expressed mRNA and proteins and able to confirm the correct translation frame, if necessary.

4. Synthesize the target CDSs (TONP): There are 3 different nucleic acid synthesizing methods, which might be suitable for the synthesis of TONP:

- Routine nucleic acid synthesis, using synthesizers. A mixture containing equal concentration of A, T, G, C residues should be used at the X residues.
- Site directed mutagenesis by Error-Prone PCR.

This method is based on the incorporation of mutagenic dNTP analogs, such as 8-oxo-dGTP and dPTP, into an amplified DNA fragment at the X positions by PCR. The mutagenic dNTPs are eliminated by a second

PCR step in the presence of the four natural dNTPs only, resulting in a rate of mutagenesis of up to 20%. (Mutagenesis Kits are commercially available).

- Codon-varied oligonucleotide synthesis for synthetic shuffling. This method uses codons instead of single nucleotides for CDS synthesis (Welch *et al.*, 2002). This method can make it easier to position the mutations (X) always into the central codon residue.

The result of these syntheses will be a mixture of more or less different oligonucleotides which differ from each other at one or more central codon residue, but they are identical at the 1st and 3rd codon residues. The number of expected target oligonucleotide sequences is 4^c , where c is the number of codons (amino acids) in the designed target sequences. This mixture of different target sequences is called the Target Oligonucleotide Template Pool (TONP).

It is necessary to validate the correctness of nucleic acid synthesis to make sure that the sequences in TONP contain the TONT pattern.

5. Insert the TONP into vectors: Using restriction enzymes. This is a routine procedure.

6. Insert of TOPN: Insert the TONP containing vectors into the chosen screening system (phage, yeast or bacteria). This is a routine procedure.

7. Test of TONP: Test that the expression of TONP works properly, i.e.,

- TONP mRNAs are present;
- the TONP mRNAs contain the TONT pattern (correct translation frame is used, orientation is correct);
- Detect the signal protein (if it's CDS was added to the TONT). This is a further indication that translation is correct, no frame shift occur and the Target Oligo-Peptides (TOPP) are expressed.

8. Prepare the query protein: Some screening systems (two-hybrid yeast or bacterial systems) request the expression of the query too. In these cases prepare the query accordingly to the systems requirements and test the correctness of query expression.

9. Expression/cloning library construction: The TOPP might be expressed and multiplied in yeast, bacteria (using plasmid or phage vectors) which are providing the TOPP expression library. This libraries are expected to contain

and express about 10^6 - 10^9 more or less different, oligopeptides. Target Oligopeptide Library is a partially (33%) random library, because the central residues of the codons are randomly selected.

10. Screening of the TOPP expression library: There are at least 2 different methods which are suitable for screening the clones

- Protein-fragment Complementation Assays (PCA) (Michnick *et al.*, 2000).

This assay is utilized by the bacterial and yeast two hybrid methods (Fig. 2).

- Phage display technique (Smith and Petrenko, 1997), (Fig. 3).

Figure 2 is general description of a PCA (Michnick *et al.*, 2001). The gene for a protein or enzyme is rationally dissected into two or more fragments. Using molecular biology techniques, the chosen fragments are sub cloned and to the ends of each, proteins that either are known or thought to interact are fused. Co-transfection or transformation of these DNA constructs into cells is then carried out. Reassembly of the probe protein or enzyme from its fragments is catalyzed by the binding of the test proteins to each other and reconstitution is observed with some assay.

11. Select the clones: Select the best clones and continue the cloning until monoclonal stadium. Save the best clones.

12. Testing for physicochemical properties: The best clones should be further propagated to obtain larger quantity of the desired SHARP®-s for purification and testing for physicochemical properties (binding specificity and K_d).

- Large number of methods is known for K_d determination. Label-free real-time interaction analyses using surface plasmon resonance and a sensor chip technology, is one of the latest technologies (<http://www.biacore.com/lifesciences/introduction/index.html>; http://www.fortebio.com/pages/system_overview.html). These methods work on crude bacterial extracts, target protein purification is not necessary.

13. Re-sequence the target CDSs: Extract and re-sequence the target CDSs producing the best targets.

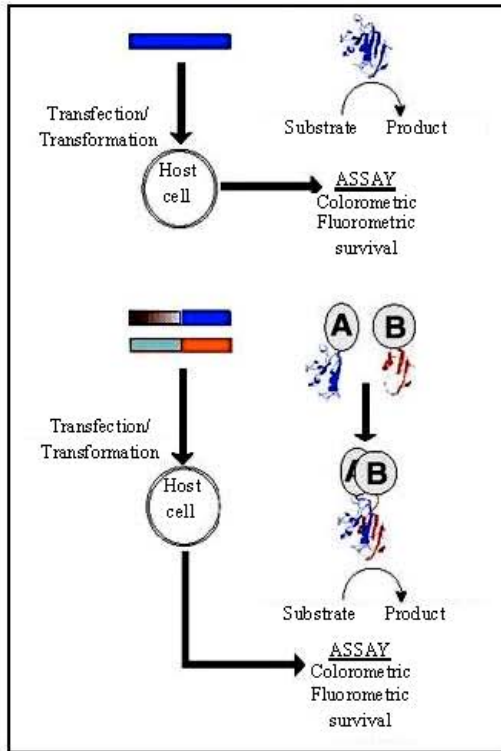


Fig. 2: Protein-fragment Complementation Assays (PCA)

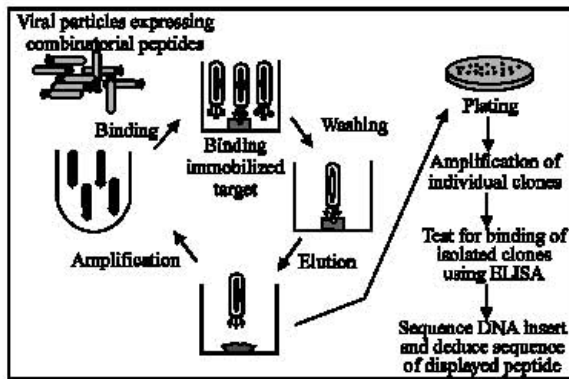


Fig. 3: Scheme of phage display (Smith, 1985)

14. Re-sequence the best target oligo-peptides and their mRNA: This step is not absolutely necessary. Finding the proper SHARP® might be satisfactory. However, re-sequencing is the source of important information for further experiments. For example

- Re-sequencing is the final confirmation, that the target peptides are found by design and not only by chance.

- Re-sequencing and analyzing the query/target sequences might be the source of valuable knowledge about the further rules of protein-protein interactions.

Therefore, re-sequencing is highly recommended.

15. Visualization of query/target complex: Visualization of query/target complex (NMR) may be an additional step to gain further knowledge about the protein folding and interactions.

Iterate the procedure: Iterate the procedure 1-15 using another query.

RESULTS

Bacteriomatch II two-hybrid system: The system chosen was the BacterioMatch II two-hybrid system commercially available from Stratagene (<http://www.stratagene.com>). This system utilizes a Bait vector (pBT) and a Target vector (pTRG). The bait vector contains a portion of the λ cI gene and the target vector contains a portion of the alpha subunit of RNA polymerase. The bait and target peptides (or proteins) are genetically fused to the cI gene in pBT and the RNA polymerase subunit in pTRG, respectively. Inside a cotransfected bacterial cell, the λ cI gene product binds to an λ operator on a reporter plasmid. The bait portion of the construct is available to interact with the target portion of the target vector. When an interaction occurs (the bait and target bind to each other), the RNA polymerase subunit is in close proximity to a weak RNA polymerase binding site (from the *lacZ* promoter) on the reporter plasmid. This binding allows the RNA polymerase to transcribe a pair of reporter genes, *HIS3* and *aadA*. The *HIS3* gene allows the bacterial cells to grow on medium lacking histidine (or more accurately containing the histidine antagonist 3-amino-1, 2, 4-triazole or 3AT). If the binding interaction is strong enough, the *aadA* gene is also transcribed conferring streptomycin resistance on the cells. Double selection on 3AT and streptomycin containing plates reduces the number of false positives. The BacterioMatch II kit is supplied with a pair of control vectors containing portions of the yeast *GAL4* and *GAL11* genes. The interaction of these gene products is both highly specific and tight. Cells from a transfection using both of these controls produce many colonies on doubly selective plates (Fig. 4).

Outline of experimental design: We designed a series of target peptide gene sequences that could potentially bind to the control vector pBT-LGF2 (containing the λ cI gene).

These target oligonucleotides (the Target Oligo Nucleotide Pool) were synthesized by Retrogen [21], digested with appropriate restriction enzymes and ligated into similarly digested pTRG. The pBT-LGF2 and pTRG-TONP were cotransfected into *E. coli*.

The sequence of the GAL4 portion of the pBT-GAL4 obtained from the Strategene website was inconsistent with the size of the GAL4 protein as described in the Strategene literature. Forward and reverse oligonucleotide primers for sequencing pBT were synthesized (Retrogen) and the nucleotide sequence of the GAL4 protein was determined. This sequence was then matched by to the published sequence to establish the exact portion of the GAL4 protein contained in the pBT-LGF2 vector.

TONTs were designed by to match the protein sequence predicted from the newly determined nucleic acid sequence of GAL4. The sequences in TONP included the restriction sites for *Bam*HI and *Not*I (Fig. 5). One strand of the sequences in TONPs was chemically synthesized using multiple bases at various positions

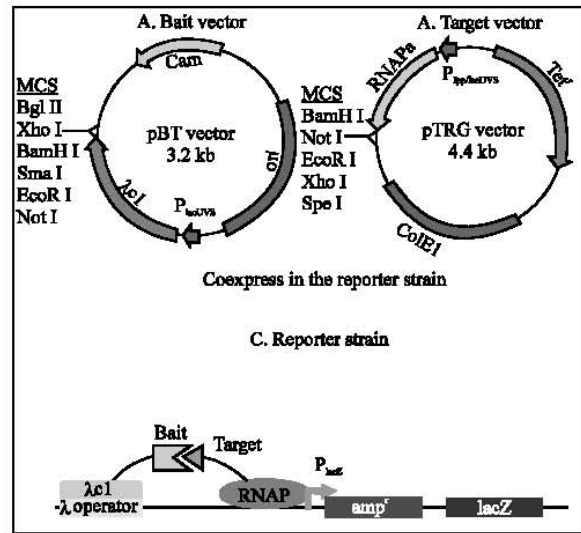


Fig. 4: The BacterioMatch™. Two-hybrid system (<http://www.stratagene.com>)

Design of SHARP@s

A. Query Selection

```
>1HBW:A REGULATORY PROTEIN GAL4
1      9      17     23     31                               57
.      .      .      .      .
TRAHLTEVESRLERLEQLFLLIFPREDLDMILKMDSLRDIEALLTGLFVQDNVNKDA
HHHHHHHHHHHHHHH SSSSSSS HHHHHHTTHHHHHHHHHHH S SS
```

```
>K01486 SCGAL4 DIMDOM | ND, 171 bases, 1D18 checksum.
5'-- ACTAGGCGACRCTGTGACAGAACTGGAATCAAGGCTAGAAGACTGGAACAGCTATTTCTACTGATTTTCCTCGAGAAGACCTTGACATGATTTT
H2N- T R A H L T E V E S R L E R L E Q L F L L I F P R E D L D M I L
```

```
>Query 1_ESRLERLEQLFLLIF (GAL4 09-23AA)
5'--GAATCAAGGCTAGAAAGACTGGAACAGCTATTTCTACTGATTTT
H2N- E S R L E R L E Q L F L L I F
```

```
>Query 2_QLFLLIFPREDLDMI (GAL4 17-31AA)
5'--CAGCTATTTCTACTGATTTTCCTCGAGAAGACCTTGACATGATTT
H2N- Q L F L L I F P R E D L D M I
```

B. Target Oligo-Nucleotide Template (TONT) Design

```
>TARGET TEMPLATE to Query 1_ESRLERLEQLFLLIF (GAL4 09-23AA)-ssDNA-sense
5'-GGATCC-[AxAxTCxGTxGAxATxGCxGTxCxGTxTTxCTxGCxTTxATTxC]-CGGCCGC-3'
|--BamHI--[ VARIABLE SEQUENCE, TONT 45 NA ]-NotI----
```

```
>TARGET TEMPLATE to Query 2_QLFLLIFPREDLDMI (GAL4 17-31AA)-ssDNA-sense
5'-GGATCC-[AxTCxTGxCAxGGxCTxCTxGAxGAxAAxTCxGTxGAxATTxGCxG]-CGGCCGC-3'
|--BamHI--[ VARIABLE SEQUENCE, TONT 45 NA ]-NotI----
```

Fig. 5: Design of SHARP@s

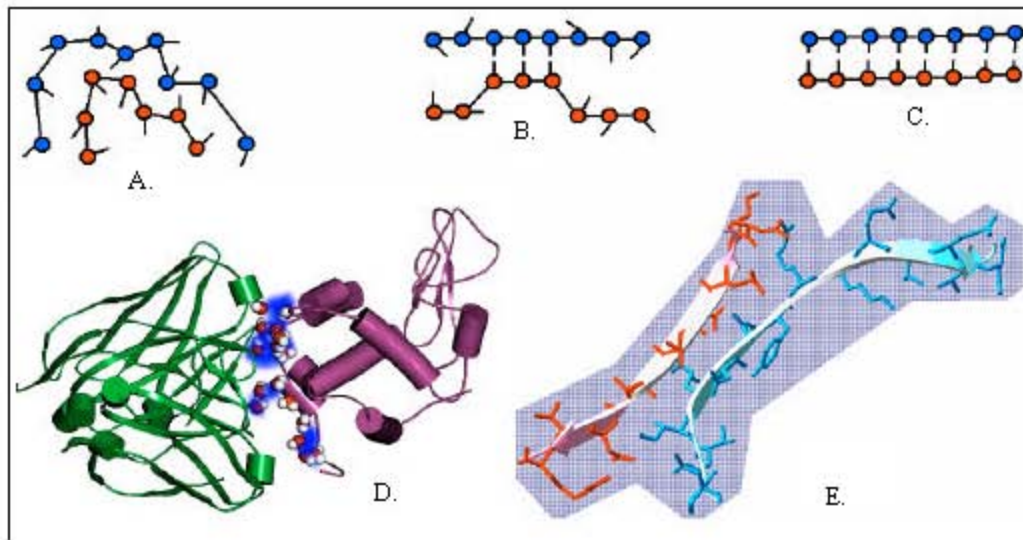


Fig. 7: Forms of peptide to peptide interactions

original concept of the Proteomic Code (perfect complementary coding of interacting peptides) were developed in many laboratories and were remarkably successful in many (but not all) cases. These methods never became widely accepted because of the poor predictability of the results (Biro, 2007).

However, new bioinformatics tools (Biro, 2005) and a large Protein structure Database (PDB) became available. Statistical analyzes of amino-acid co-locations in real protein structures revealed that the redundant Genetic Code contains even information necessary to protein folding and interactions (Biro *et al.*, 2003; Biro, 2006).

Recent observations suggest, that naturally co-locating amino acids (in- and between specifically interacting proteins) are preferentially coded by partially complementary codons. The method above is to replicate this biological phenomenon and try to design specifically interacting oligo-peptides. It is assumed, that protein-protein interactions are determined already on the amino acid level (Fig 7).

The specificity of interactions between two peptides might be explained in two ways. First, many amino acids collectively form larger configurations (protrusions and cavities, charge and hydrophathy fields) which fit each other (A and D). Second, the physico-chemical properties (size, charge and hydrophathy) of individual amino acids fit each other like "lock and key" (C and E). There are even intermediate forms (B).

Is it really plausible that there is physicochemical compatibility between individual amino acids on the interacting surfaces? Bioinformatical studies clearly

indicate that co-locating (interacting) amino acids are preferentially compatible (complementary) to each other regarding their size, charge and hydrophathy (Biro, 2006).

This small-scale compatibility suggests that large, complex structure-forming is not an absolute structural requirement of specific protein interactions, but relatively strong and specific interactions might be formed already between a few (Seiwert *et al.*, 2000; Heal *et al.*, 2002; Biro, 2007; Welch *et al.*, 2002; Michnick *et al.*, 2000; Smith and Petrenko, 1997) amino acid long oligo-peptides.

What about if the concept of Proteomic Code is wrong? The SHARP® design Method still remains a novel variant of combinatorial protein engineering an approach proved to be successful for affinity peptide design and production.

CONCLUSION

Second codon residue randomization is a promising method to design and produce affinity peptides to any protein sequences. The method has the potential to be a rapid, inexpensive, high throughput, non-immunoglobulin based alternative to recent *in vivo* antibody generating procedures.

REFERENCES

Antibodies, 2005. Keys to Success In a Competitive Market Biocompare Surveys and Reports Published http://www.biocompare.com/research/surveys_files/ExecSumm/antibodies09162005_execsumm.pdf.

- Biro, J.C. and G. Fordos, 2005. SeqX: A tool to detect, analyze and visualize residue collocations in protein and nucleic acid structures. *BMC Bioinform.*, 6: 170.
- Biro, J.C., 2006. Amino acid size, charge, hydropathy indices and matrices for protein structure analysis. *Theor. Biol. Med. Model*, 3: 15.
- Biro, J.C., 2006. Protein folding information in nucleic acids which is not present in the genetic code. *Ann. N.Y. Acad. Sci.*, 1091: 399-411.
- Biro, J.C., 2007. System and method to obtain oligopeptides with specific high affinity to query proteins. US Patent Application Publication.
- Biro, J.C., 2007. The Proteomic Code: Molecular recognition code for proteins. *Review Journal of Theor Biological Medicine Modelling*.
- Biro, J.C., B. Benyo, C. Sansom, A. Szlavecz, G. Fordos, T. Micsika and Z. Benyo, 2003. A common periodic table of codons and amino acids. *Biochem. Biophys. Res. Commun.*, 306: 408-415.
- Borrebaeck, C.A., L. Danielsson and S.A. Möller, 1988. Human monoclonal antibodies produced by primary *In vitro* immunization of peripheral blood lymphocytes. *Proc. Natl. Acad. Sci. U.S.A.*, 85: 3995-3999.
- Dynamic Antibody Industry, 2005. Published by the Business Communications Company.
- Harlow, E. and D. Lane, 1988. Protocols for preparing immunogens, immunization of animals and collection of antiserum may be found in *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, pp: 55-120.
- Heal, J.R., G.W. Roberts, J.G. Raynes, A. Bhakoo and A.D. Miller, 2002. Specific interactions between sense and complementary peptides: The basis for the proteomic code. *Chembiochemistry*, 3: 136-151 (Review. Erratum in: *Chembiochemistry*, 3: 271).
- Michnick, S.W.W., J.N. Pelletier and I. Remy, 2001. Protein fragment complementation assays for the detection of biological or drug interactions. US Patent, 6: 270, 964.
- Michnick, S.W., I. Remy, F.X. Campbell-Valois, A. Vallee-Belisle and J.N. Pelletier, 2000. Detection of protein-protein interactions by protein fragment complementation strategies. *Methods Enzymol.*, 328: 208-230.
- Nord, K., E. Gunneriusson, J. Ringdahl, S. Stahl, M. Uhlen and P.A. Nygren, 1997. Binding proteins selected from combinatorial libraries of an alpha-helical bacterial receptor domain. *Natl. Biotechnol.*, 15: 772-777.
- Pelat, T., M. Hust, E. Laffly, F. Condemine, C. Bottex, D. Vidal, M.P. Lefranc, S. Dübel and P. Thullier, 2007. High-Affinity, Human Antibody-Like Antibody Fragment (Single-Chain Variable Fragment) Neutralizing the Lethal Factor (LF) of *Bacillus anthracis* by Inhibiting Protective Antigen-LF Complex Formation. *Antimicrobial Agents and Chemother.*, 51: 2758-2764.
- Pichla, S.L., R. Murali and R.M. Burnett, 1997. The crystal structure of a Fab fragment to the melanoma-associated GD2 ganglioside. *J. Struct. Biol.*, 119: 6-16.
- Seiwert, S.D., T. Stines Nahreini, S. Aigner, N.G. Ahn and O.C. Uhlenbeck, 2000. RNA aptamers as pathway-specific MAP kinase inhibitors. *Chem. Biol.*, 7: 833-843.
- Smith, G.P. and V.A. Petrenko, 1997. Phage Display. *Chem. Rev.*, 97: 391-410.
- Smith, G.P., 1985. Filamentous fusion phage: Novel expression vectors that display cloned antigens on the virion surface. *Science*, 228: 1315-1317.
- Welch, M., J. Ness, C. Gustafsson, W.P.C. Stemmer and J. Minshull, 2002. Use of codon-varied oligonucleotide synthesis for synthetic shuffling. US Patent Issued on United States Patent, 6436675.