

Research

Self-assembly of proteins and their nucleic acids

Graham Fletcher, Sean Mason, Jon Terrett and Mikhail Soloviev*

Address: Oxford GlycoSciences (UK) Ltd, Abingdon, Oxon OX14 3YS, United Kingdom

Email: Graham Fletcher - Graham.Fletcher@ogs.co.uk; Sean Mason - Sean.Mason@ogs.co.uk; Jon Terrett - Jon.Terrett@ogs.co.uk; Mikhail Soloviev* - Mikhail.Soloviev@ogs.co.uk

* Corresponding author

Published: 28 January 2003

Received: 25 November 2002

Journal of Nanobiotechnology 2003, 1:1

Accepted: 28 January 2003

This article is available from: <http://www.jnanobiotechnology.com/content/1/1/1>

© 2003 Fletcher et al; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

Abstract

We have developed an artificial protein scaffold, herewith called a protein vector, which allows linking of an in-vitro synthesised protein to the nucleic acid which encodes it through the process of self-assembly. This protein vector enables the direct physical linkage between a functional protein and its genetic code. The principle is demonstrated using a streptavidin-based protein vector (SAPV) as both a nucleic acid binding pocket and a protein display system. We have shown that functional proteins or protein domains can be produced in vitro and physically linked to their DNA in a single enzymatic reaction. Such self-assembled protein-DNA complexes can be used for protein cloning, the cloning of protein affinity reagents or for the production of proteins which self-assemble on a variety of solid supports. Self-assembly can be utilised for making libraries of protein-DNA complexes or for labelling the protein part of such a complex to a high specific activity by labelling the nucleic acid associated with the protein. In summary, self-assembly offers an opportunity to quickly generate cheap protein affinity reagents, which can also be efficiently labelled, for use in traditional affinity assays or for protein arrays instead of conventional antibodies.

Background

The 20th century has witnessed the birth of molecular biology and an explosion in cloning applications, the numbers of which exceeds hundreds of thousands. Traditional molecular cloning approaches are dependant on the ability of cells to both synthesise proteins from DNA and to replicate themselves and any exogenous DNA. This enables the linkage, within an individual cell, of the information-carrying DNA to the encoded protein or the cellular phenotype. Viruses and phages are also used in molecular biology and provide another means of "linking" protein (or protein function) to corresponding DNA but they are entirely dependent upon a host cell to replicate. Using cell- or phage-based cloning systems resolves a number of important problems. It allows the creation of a "one DNA vector per cell" system, which following a physical separa-

tion (by plating on a dish or through dilution) can be amplified (through self-replication) into a macroscopic colony which could then be catalogued, stored or grown further for preparative applications. However, the use of living cell-based systems has a number of disadvantages. Performing such experiments not only requires proper facilities, but they are also lengthy processes. Bacterial or phage cloning takes about a day to go from a single bacteria to a clone; yeast takes days to grow; and mammalian cells take weeks to form a clone. An adequate amplification of DNA can be achieved by other means. For the last decade PCR has been widely used instead of cloning for the production of large amounts of DNAs. However, no adequate system has so far been developed for linking the DNA, an information carrier, to its protein, a function carrier.

Direct linking of proteins to their DNAs or RNAs to bypass the limitation of cellular systems has been attempted before. One strategy has been to utilise components of the cellular protein synthesis machinery to transiently or permanently link mRNAs and proteins. Protein synthesis in living cells is a two-step process involving transcription, which is followed by translation. During transcription of DNA, an mRNA is made and processed by RNA polymerases and spliceosome complexes. Translation involves protein synthesis on ribosomes using mRNA as a template molecule. If transcription termination is blocked, the mRNA will remain in the complex with its DNA (and with the enzymes responsible for the RNA synthesis and splicing). Similarly, if translation termination is prevented the ribosome will remain associated with both the mRNA and the nascent protein chain. The discovery that the processes of transcription and translation could be performed outside the cell [1-3] has encouraged attempts to "link" such in vitro synthesised proteins to their nucleic acid. Taussig and He have employed the transcription-translation termination blockade to create transient {mRNA-ribosome-protein} complexes which physically crosslink the RNA with the associated proteins [4,5]. Such a "ribosome display" approach has a number of disadvantages, including the fact that the complexes obtained also include all elements of the protein synthesis machinery, i.e. ribosomes with all their associated RNAs and proteins. This not only depletes the translation reaction but also results in a very high background and large number of unrelated proteins linked to the mRNA. Xu et al [6] have produced intermediate {mRNA-DNA-adaptor-ribosome-Protein} complexes where a puromycin-labelled DNA adaptor, separately ligated to RNA molecules, covalently links to a nascent protein chain in a sequence-independent manner (an "mRNA display" approach, [6]). Such a modification results in covalent {mRNA-protein} complexes, which lack bulky ribosomes, but involve a high degree of non-specific crosslinking of the RNA to ribosomal proteins. Ligation of a puromycin-modified DNA to mRNA requires an additional step, which makes the whole procedure significantly longer especially if a few rounds of subsequent amplification and selection are required. A variation of RNA-protein complex production using puromycin was also reported by Roberts and Szostak, and by Liu et al [7,8] respectively. All the methods reported so far result in the production of covalently crosslinked protein-RNA hybrids and/or complexes containing bulky ribosomes or requiring multi-step processes and excessive RNA handling in order to make protein-DNA complexes. The use of mRNA in the techniques described above is disadvantageous because of the instability of RNA and its fast degradation compared to the more stable DNA molecules. Another disadvantage is the requirement for the two additional enzymatic steps, namely reverse transcription and cDNA

amplification, before sequence information can be extracted.

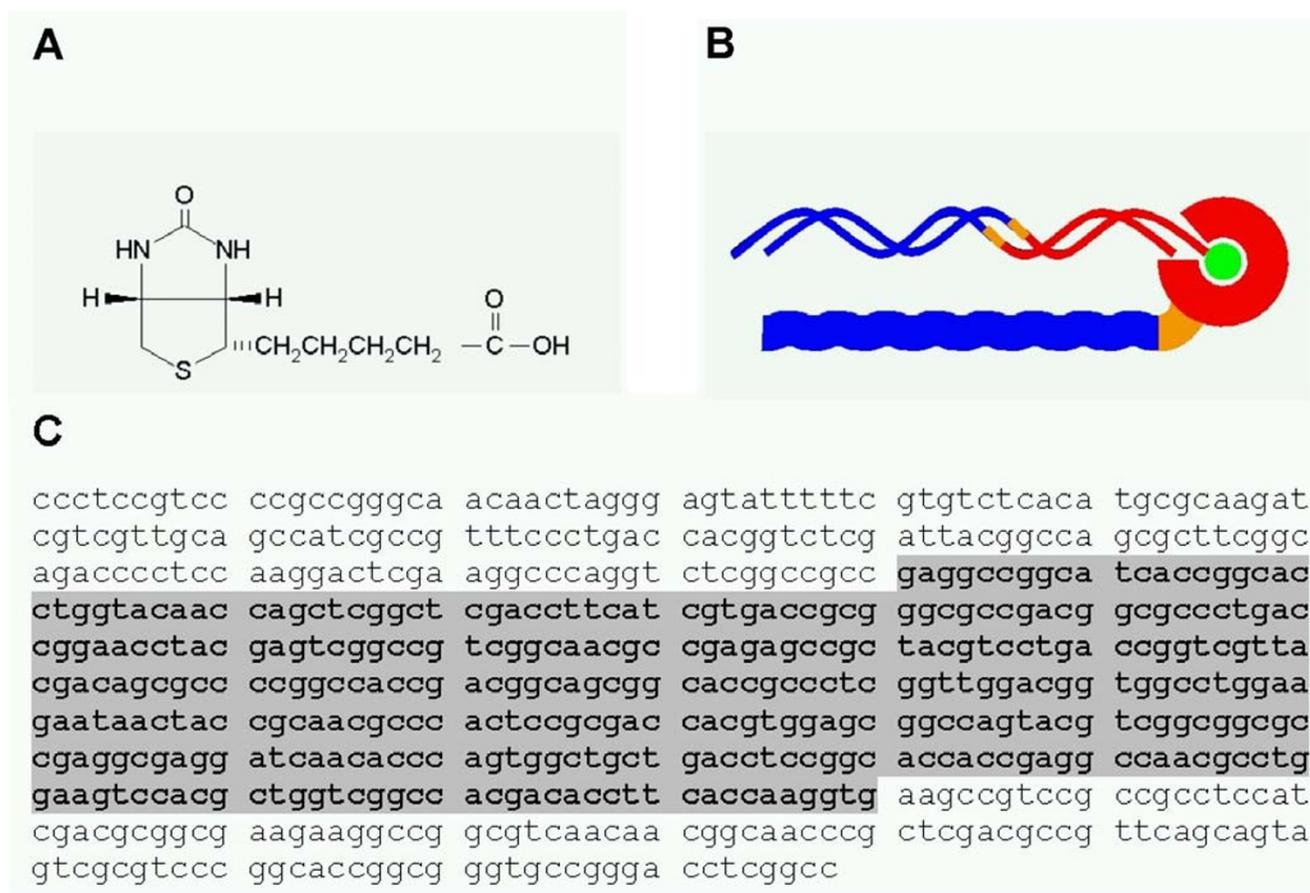
Using a molecular scaffold of a streptavidin protein we have designed a protein vector – an interface synthesised in vitro, which contains a nucleic acid assembly module and a protein sequence of interest, thus providing a direct physical link between the expressed protein feature and its encoding DNA.

Results

Design of a protein vector based on the core protein sequence of streptavidin (SA)

Streptavidin (from *Streptomyces avidinii*) is a naturally occurring protein, which is able to bind biotin (Figure 1A) with high affinity. The nucleotide sequence of the streptavidin gene was reported in 1986 by Argarana et al [9]. We have used the *Streptomyces avidinii* gene for streptavidin (X03591, Figure 1C) as a scaffold for designing a streptavidin based protein vector (SAPV, Figure 1B). Full length nucleotide sequence coding for the SAPV (Figure 2) was produced using overlapping synthetic oligonucleotides (obtained from Sigma-Genosys) and several rounds of PCR (for oligonucleotide primers and details of the synthesis see Materials and Methods). For efficient transcription by bacterial T7 polymerase, two T7 RNA polymerase binding sites and a T7 terminator sequence were inserted into the engineered SAPV DNA. It also contained a ribosome-binding site (RBS) – a signal necessary for efficient translation, see Figure 2. SAPV DNAs for use in the in vitro Transcription/Translation (T&T) were routinely obtained by PCR (see Methods). To confirm efficient expression of the SAPV at the protein level, the SAPV was designed with a protein tag (autofluorescent protein AFP). The engineered nucleotide sequence of the tagged SAPV is shown in Figure 3. Tagged SAPV DNA was generated in the same way as the untagged SAPV DNA. Tagged SAPV was detected on Western blots with anti-GFP Rabbit polyclonal antibody, see Figure 4. The strong staining confirmed efficient synthesis of the SAPV-AFP. Based on the results of this experiment, the optimal experimental conditions for all subsequent T&T reactions included the use of 2 ug DNA per 20 ul of the in vitro T&T reaction, the synthesis temperature was maintained at 21 °C.

To control whether SAPV protein vector is able to bind biotinylated DNA, a completed T&T reaction was incubated with either biotinylated or non-biotinylated DNA. The longer DNAs were chosen for assembly reactions to avoid non-specific background due to the SAPV DNA used in the in vitro T&T reaction. Protein-DNA complexes were separated from free DNAs by filtration through a protein-binding filter and the retained DNAs were detected by PCR. The amplified products were separated on agarose gels. Equal amounts of each PCR reaction were loaded

**Figure 1**

Design of the SAPV (streptavidin based protein vector). Biotin (panel **A**) can routinely and cheaply be included in oligonucleotide primers and thus be easily introduced (in a fully controllable manner) into nucleic acids used for self-assembly. Schematic diagram showing a principle behind the SAPV (panel **B**). Part of the SAPV DNA (a "double spiral") encodes for a streptavidin protein domain (marked in red) which can bind its own DNA through binding to the biotin molecule (marked green). Protein fragments (and a corresponding DNA fragment) marked in blue – a protein of interest (e.g. displayed peptides or affinity reagents or cloned proteins etc.). Yellow denotes a linker region (both protein and DNA). *Streptomyces avidinii* gene for streptavidin (X03591) mRNA sequence (panel **C**). The corresponding deduced amino acid sequence of the streptavidin protein is available from the SwissProt database (P22629). Fragment of the coding region used in the design of the SAPV protein vector is shaded grey.

onto each lane (Figure 5). The absence of a signal in the 4th wash (in both the biotinylated and non-biotinylated DNA assemblies) confirms the absence of a non-specific background. The eluates from the biotinylated DNA experiments (Figure 5A,5C) contained large amounts of amplified DNA, whilst the eluates from the non-biotinylated DNA assemblies (Figure 5B,5D) did not. This clearly demonstrates that the designed SA-based tagged protein vector is able to bind biotinylated DNAs.

Assembly and affinity precipitation of SAPVs displaying a BCMP84 peptide

The core protein sequence of streptavidin and the streptavidin-based SAPV contains a 9 amino acid long loop (GT-TEANAWK, Figures 6 and 7), which we predicted to be most suitable for modifications, such as SAPV extension, modification, or for expressing other protein fragments, peptides and tags. This choice is based on the molecular architecture of streptavidin (Figure 6B). To illustrate the "display" capabilities of the SAPV, we have engineered SAPV-Alb5 and SAPV-84 which display peptide fragments of Albumin and BCMP84 proteins, respectively (Table 1).

```

AAAT TAATAC GACTCACTAT AGGGAGAAGG AGATATACCA TGGAGGCGCG CATCACCGGC ACCTGGTACA
T7-F
ACCAGCTCGG CTCGACCTTC ATCGTGACCG CGGGCGCCGA CGGCGCCCTG ACCGGAACCT ACGAGTCGGC
CGTCGGCAAC GCCGAGAGCC GCTACGTCCT GACCGGTCGT TACGACAGCG CCCC GGCCAC CGACGGCAGC
GGCACCGCCC TCGGTTGGAC GGTGGCCTGG AAGAATAACT ACCGCAACGC CCACTCCGCG ACCACGTGGA
GCGGCCAGTA CGTCGGCGGC GCCGAGGCGA GGATCAACAC CCAGTGGCTG CTGACCTCCG GCACCACCGA
GGCCAACGCC TGGAATCCA CGCTGGTCGG CCACGACACC TTCACCAAGG TGGAA TGAAT GAATGATAAAC
SA-7R
TAGCAT AACC CCTTGGGGCC TCTAAACGGG TCTTGAGGGG TTTTTT
T7TER-R

```

Figure 2

Full length engineered nucleotide sequence (466 b.p.) coding for the SAPV protein vector. Oligonucleotide primer sequences used to amplify the SAPV DNA are underlined (T7-F forward and T7TER-R reverse primers). The reverse oligonucleotide primer SA-7R was used to amplify SAPV lacking stop codons (to facilitate self-assembly by slowing down transcription and translation). Turquoise highlighting denotes T7 RNA polymerase binding sites, red highlighting – a ribosome binding site, preceding the ATG start codon (light green). Sequence fragment within the SA-7R oligonucleotide highlighted in yellow codes for the amino acid loop within the Streptavidin sequence, which is suitable for modifications (see also Figures 6 and 7). Stop codons are highlighted in blue, the transcription termination site in pink.

The choice of the peptides was determined by the antibodies available (polyclonal anti-albumin antibody, which recognise the Albumin peptide, and polyclonal anti-BCMP84 anti-peptide antibody). DNAs encoding the modified SAPV (SAPV-Alb5 or SAPV-84) were obtained by PCR. A co-immunoprecipitation system was designed to quickly separate different SAPVs. The protocol was tested using a recombinant BCMP84 protein. We separately tested glass bead-based and nitrocellulose-based systems. Comparable amounts of BCMP84 protein were present in the eluates from both the beads and the nitrocellulose, indicating that the protein was selectively retained (Figure 8).

Assembled SAPV-84 protein-DNA complexes were immunoprecipitated using either anti-BCMP84 or anti-albumin antibodies bound to nitrocellulose. Following a number of washes, the SAPVs were eluted and the eluates assayed by PCR amplification of the SAPV-84 DNA. The results of 5 independent measurements are presented in Figure 9. The results indicate that immunoprecipitation of SAPV-84 on the anti-BCMP84 nitrocellulose is significantly higher than on the control anti-Albumin nitrocellulose. The approximately 2.5x fold difference cannot be taken as a fully quantitative measurement as this assay employed an end point PCR detection, which may have gone out of the logarithmic amplification phase. However, the clear predominance of the assembled SAPV-84 in the eluate from the

anti-BCMP84 nitrocellulose confirms that the BCMP84 peptide was adequately displayed on the SAPV-84 protein vector, which was assembled with the biotinylated SAPV-84 DNA and precipitated by anti-BCMP84 antibody.

Self-assembly of protein vectors with their DNAs and affinity separation

Co-transcriptional and co-translational self-assembly of SAPV protein vectors with their encoding DNAs is demonstrated using SAPV-84, SAPV-Alb5 and "empty" SAPV-only (unmodified) protein vectors. The in vitro synthesised and assembled SAPVs were incubated with either anti-BCMP84 or anti-Albumin antibodies, which were immobilised on beads. Following incubation and washings, the co-immunoprecipitated SAPVs were eluted and assayed by PCR. Equal amounts of each PCR reaction were analysed by electrophoresis (see Figure 10). The figure clearly demonstrates that only correct self-assembled SAPVs are precipitated, i.e. SAPV-84 DNA is co-precipitated on anti-BCMP84 beads and SAPV-Alb5 DNA is co-precipitated on anti-Albumin beads.

Discussion

Protein vectors

We have reported the design of protein vectors that are capable of self-assembly with nucleic acids. The key principle behind our design of the protein vectors is the use of nucleic acids which encode proteins that contain, as part

```

GTAAAACGAC GGCCAGTGAA TTGTAATACG ACTCACTATA GGGCGAATTG AATTTAGCGG CCGCGAATTC
M13F-20
GCCCTTAAAT TAATACGACT CACTATAGGG AGACCACAAC GGTTCCTC TAGAAATAAT TTTGTTAAC
TTTAAGAAGG AGATATACCA TGGAGGCCGG CATCACCGGC ACCTGGTACA ACCAGCTCGG CTCGACCTTC
ATCGTGACCG CGGGCGCCGA CGGCGCCCTG ACCGGAACCT ACGAGTCGGC CGTCGGCAAC GCCGAGAGCC
GCTACGTCCT GACCGGTCGT TACGACAGCG CCCC GGCCAC CGACGGCAGC GGCACCGCCC TCGGTTGGAC
GGTGGCCTGG AAGAATAACT ACCGCAACGC CCACTCCGCG ACCACGTGGA GCGGCCAGTA CGTCGGCGGC
GCCGAGGCGA GGATCAACAC CCAGTGGCTG CTGACCTCCG GCACCACCGA GGCCAACGCC TGGAAAGTCCA
CGCTGGTCGG CCACGACCGA ATTCGGGGAG GCGGAGGTGG AGCTAGCAAA GGAGAAGAAC TCTTCACTGG
AGTTGTCCCA ATTCTTGTTG AATTAGATGG TGATGTTAAC GGCCACAAGT TCTCTGTCAG TGGAGAGGGT
GAAGGTGATG CAACATACGG AAAACTTACC CTGAAGTTCA TCTGCACTAC TGGCAAACCTG CCTGTTCCAT
GGCCAACACT AGTCACTACT CTGTGCTATG GTGTTCAATG CTTTTCAAGA TACCCGGATC ATATGAAACG
GCATGACTTT TTCAGAGTG CCATGCCCGA AGGTTATGTA CAGGAAAGGA CCATCTTCTT CAAAGATGAC
GGCAACTACA AGACACGTGC TGAAGTCAAG TTTGAAGGTG ATACCCTTGT TAATAGAATC GAGTTAAAAG
GTATTGACTT CAAGGAAGAT GGCAACATTC TGGGACACAA ATTGGAATAC AACTATAACT CACACAATGT
ATACATCATG GCAGACAAAC AAAAGAATGG AATCAAAGTG AACTTCAAGA CCCGCCACAA CATTGAAGAT
GGAAGCGTTC AACTAGCAGA CCATTATCAA CAAAATACTC CAATTGGCGA TGGCCCTGTC CTTTTACCAG
ACAACCATTA CCTGTCCACA CAATCTGCCC TTTCGAAAGA TCCCAACGAA AAGAGAGACC ACATGGTCCT
TCTTGAGTTT GTAACAGCTG CTGGGATTAC ACATGGCATG GATGAACTGT ACAACATCGA TGACTAACT
AGACGCGTTG GCCCTATTCT ATAGTGTAC CTAATGCTA GAGCTCGCTG ATCAGCCTCG ACTGTGCCTT
CTAGTTGCCA GCCATCTGTT GTTTGCCCTT CCCCCTGCA CCAAGGTGGA ATGAATGAAT GATAACTAGC
ATAAACCCCTT GGGGCCTCTA AACGGGTCTT GAGGGTTTTT TT
T7-TER-R

```

Figure 3

Nucleotide sequence of the tagged SAPV (1442 b.p.). A sequence coding for the autofluorescent protein (AFP, shaded grey) was fused C-terminal to SAPV coding sequence. The linker sequence is highlighted in dark green. See legend to Figure 2 for other details.

of their protein sequence (or structure), a fragment (or fragments) which upon synthesis are able to bind the nucleic acids in either a sequence-specific or non-specific manner. This self-assembly is achieved by labelling the nucleic acid with a ligand, which is then bound by the synthesised protein vector, or may alternatively be achieved by utilising nucleotide sequence-specific interactors. Sequence independent recognition pairs can be exemplified by the following pairs of interactors: (i) biotin as nucleic acid label and avidin, streptavidin, related proteins or derivatives which bind biotin as part of a protein vector which is encoded by the labelled nucleic acids; (ii) a small molecule ligand or ligands (for example glutathione), as a nucleic acid label, and an appropriate receptor or protein fragment which binds the ligand as part of the protein vector and which is encoded by the labelled nucleic acid (i.e. GST protein or fragments); (iii) nucleic acids, which additionally encode stretches of Lysine or Arginine which are inherently positively charged, and which upon synthesis of protein vector will bind the nucleic acid (which is inherently negatively charged). If sequence-specific recognition is sought, then nucleic acids should in-

clude binding sites (i.e. specific sequences) for nucleic acid-binding proteins and should also encode corresponding nucleic acid-binding proteins. The use of known protein transcription factors and their target DNA sequences is a possibility. Sequence-specific interaction may seem preferable to sequence-independent recognition. However, the low affinity of known DNA sequence-specific recognition pairs and the limited number of such pairs available are clear disadvantages. On the other hand, sequence-independent recognition, if performed co-transcriptionally and co-translationally whilst DNA, RNA and the nascent protein are present in a single transient complex may be as effective in linking DNA with the encoded protein as using the sequence specific interactors. Moreover, it is possible to extend the life time of such DNA-mRNA-protein complexes or even to transiently block their disassembly and thus to increase the chances of formation of the correct protein-DNA or protein-RNA pairs. We have designed our protein vectors using streptavidin as a scaffold due to its high affinity to biotin, which could be routinely and cheaply incorporated into nucleic acids and primers.

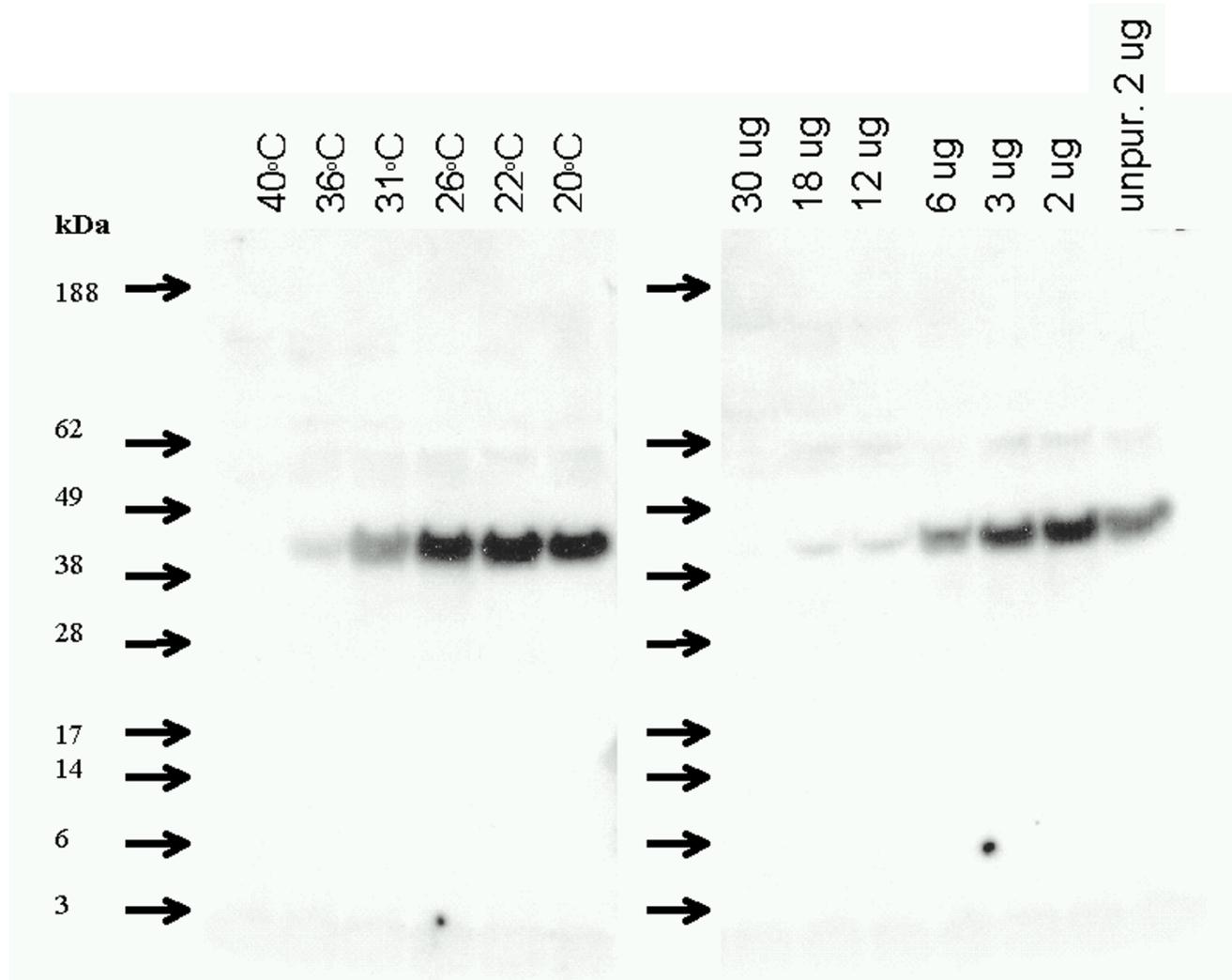


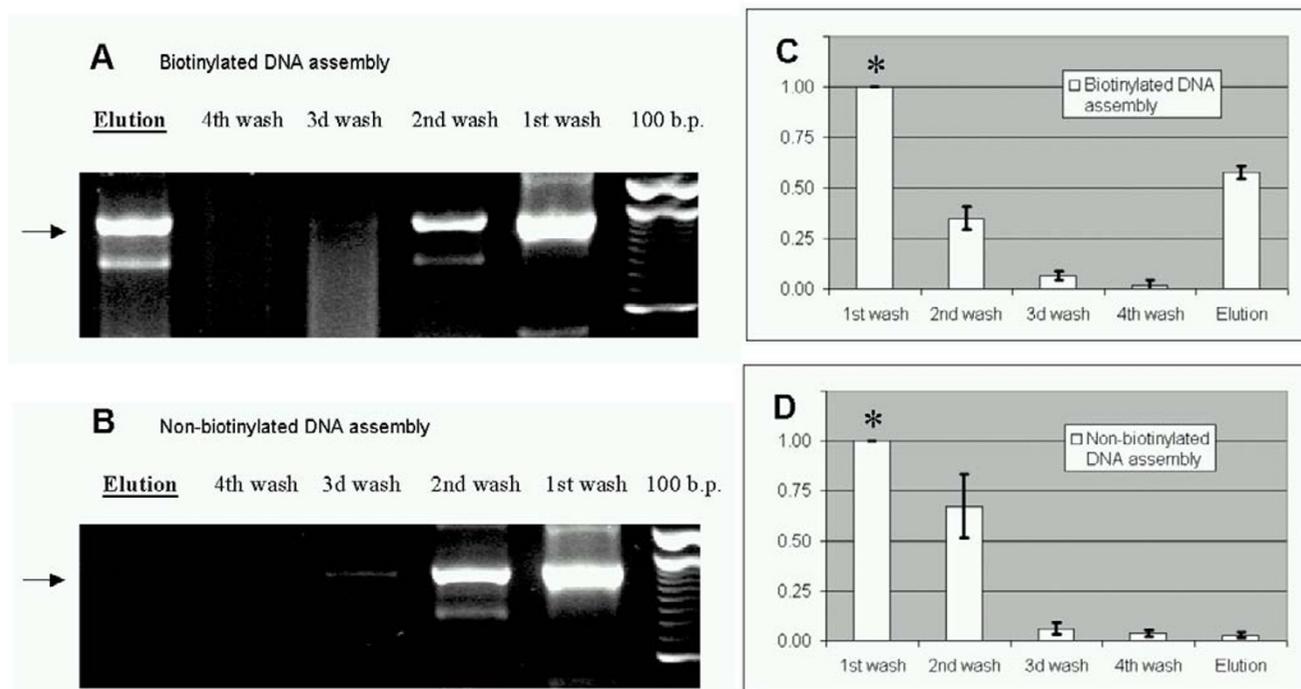
Figure 4

Detection of the tagged SAPV on Western blots. SAPV protein vector was tagged with AFP sequence. In vitro T&T reactions were run either at different temperatures (left panel) or different amounts of DNA was added to the reactions (right panel). Detection of the tagged SAPV was done using anti-GFP Rabbit polyclonal antibody (from AbCam). The right most lane (right panel) represents T&T reaction containing 2 ug of unpurified PCR products. Second lane from the right – 2 ug of DNA was ethanol-precipitated prior to T&T, following lanes – 3 ug, 6 ug, 12 ug, 18 ug and 30 ug DNA, all were ethanol-precipitated prior to T&T.

Display system based on the SAPV protein vector

We have identified a loop in the amino acid sequence of streptavidin (Figures 6 and 7) which can be used as a site for SAPV extension, modification or for expressing other protein fragments, peptides or tags. We have illustrated how our system can be used for displaying proteins and protein fragments (Figures 9 and 10). Generally speaking,

however, "displaying" artificial sequences may change the folding of the SAPV. To avoid this, the secondary structure elements of the SAPV could be additionally stabilised and positioned by one or more disulphide bonds. In particular, one (or more) of the 8 amino acids of the streptavidin core sequence, immediately preceding the loop (NTQWLLTS) and one (or more) of the respective 8

**Figure 5**

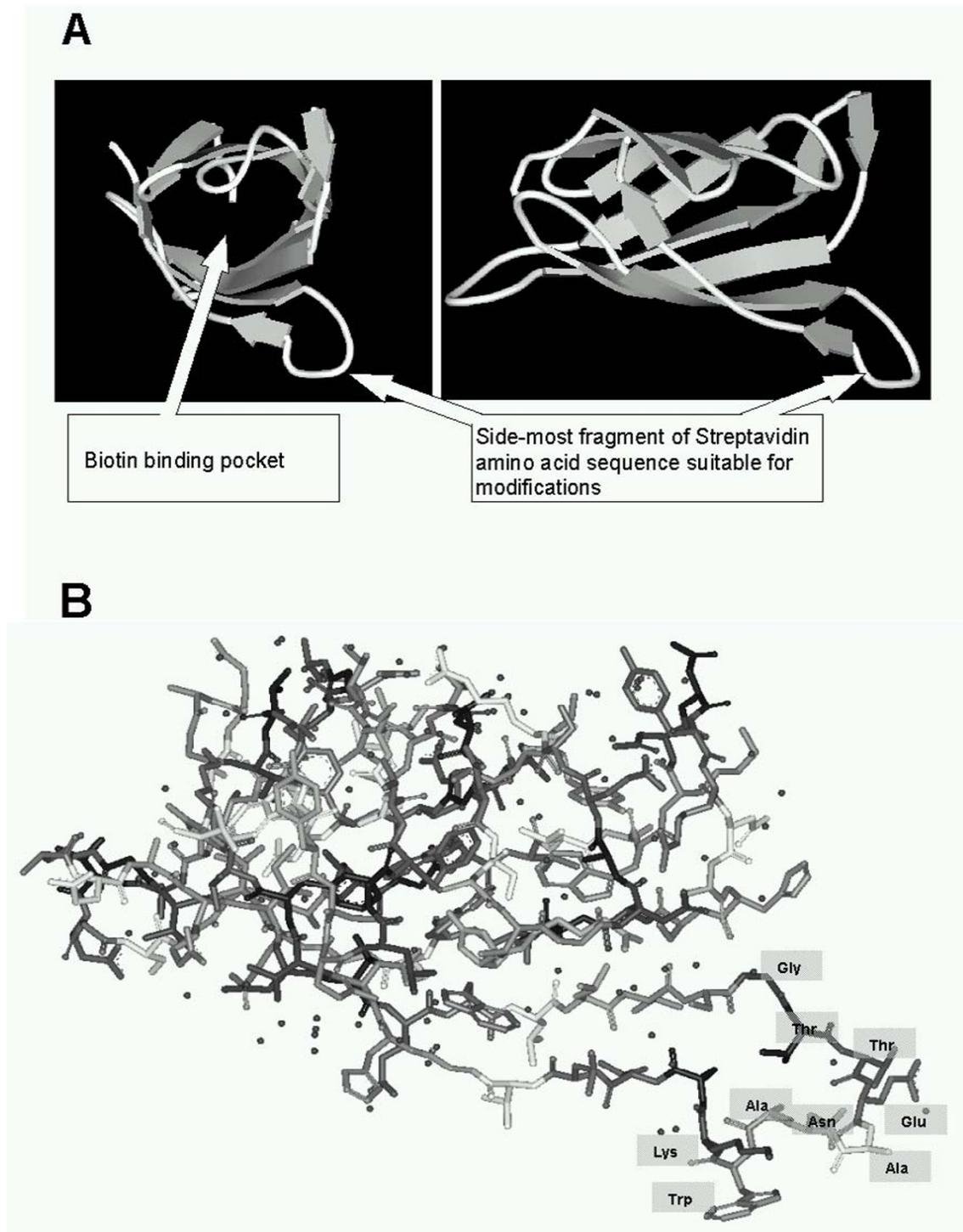
Assembly of the SAPV protein vector with biotinylated DNA. Panel **A** – biotinylated DNA was added to the SAPV vector. The assembled complexes were separated from the rest of the reaction components by filtration through protein-binding filters. The four washes and the eluate were tested by PCR. Large amount of the DNA was retained indicating that biotinylated DNA was retained by the SAPV vector. Panel **B** – same as in panel **A**, except that non-biotinylated DNA was added to the SAPV. Arrows on the left of both gels indicate the expected size (position) of the amplified products corresponding to the assembled DNAs. Panel **C**, same as panel **A**, but data pooled from three experiments. The band intensities were determined using GeneSnap and fluorescent imager from SynGene (Cambridge, UK). All values shown were normalised to the DNA sample from the 1st wash (which also contained a flow-through fraction of the total loaded DNA, marked by asterisk). Error bars represent standard deviation ($n = 3$). Large amounts of the DNA were eluted in all three experiments (the right most bar) confirming that biotinylated DNA was retained by the SAPV vector. Panel **D** – same as panel **B**, but data pooled from three experiments. No biotinylated DNA was co-precipitated (the right most bar).

amino acids C-terminal to the loop (STLVGHDT) could be substituted with Cysteine residues (Figure 7). This is possible because the distances between respective pairs of amino acids in these two antiparallel strands (the two 8 amino acid stretches) and their orientation should allow pairwise Cysteine substitutions without major changes to the streptavidin folding pattern (Figure 7).

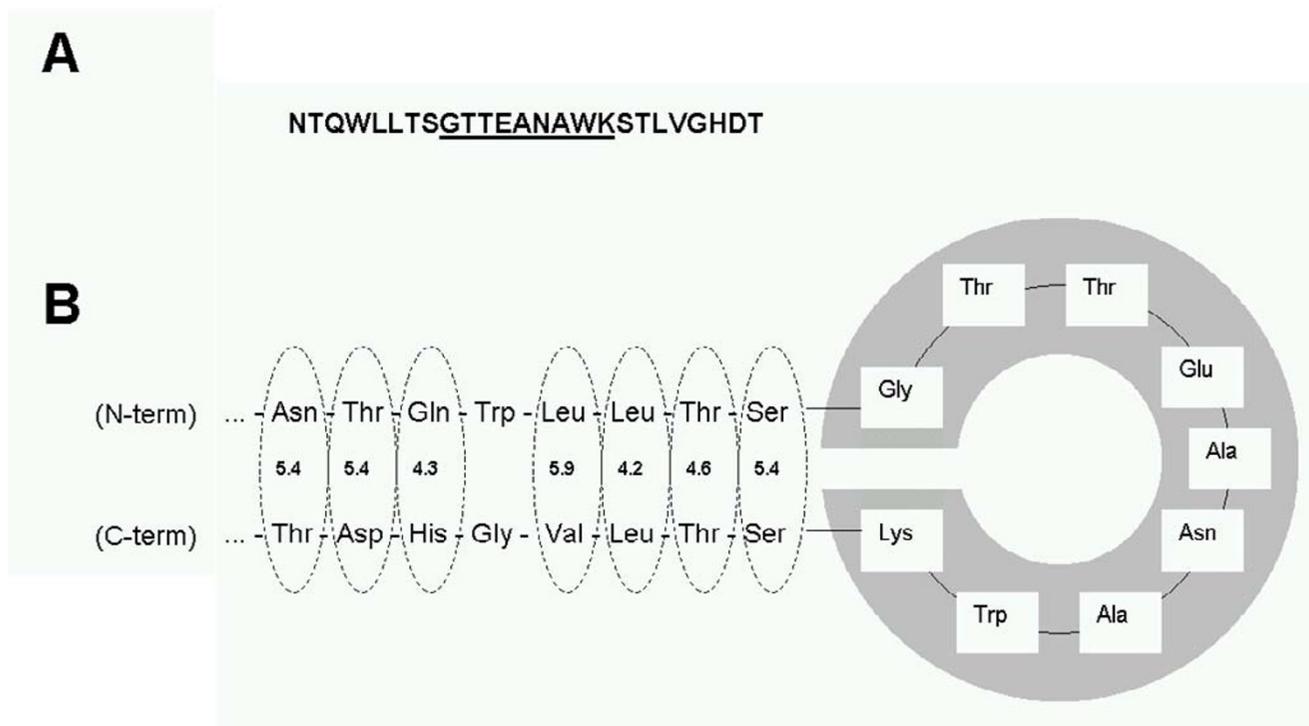
Self-assembly

If required, the efficiency of the self-assembly process could be manipulated by regulating a processivity of the transcription and/or translation reactions. This could be achieved by varying the concentration of the tRNAs present in the reaction mixture and the use of respective codons in RNA (or DNA) sequences coding for the proteins processed. The translation reaction can be paused or

stopped if required tRNA(s) is not available. Protein synthesis will continue after the missing tRNAs are added to the translation reaction. This could allow a user to manipulate the speed of synthesis and folding of the nascent protein chains and also to regulate protein vector binding to the nucleic acid molecules as well as protein-protein interactions (in protein complex formation). For example translation can be paused or slowed down after the assembly domain of the protein vector is produced, to allow binding to a nucleic acid or solid support or another protein, before the complete protein is translated and released from the ribosome. In vitro translation could also be slowed down by addition of a short complementary nucleic acid strand, the technique used in vivo and known as the antisense approach [10–13].

**Figure 6**

Structure of the streptavidin protein. Panel **A** – A 3D structure of the streptavidin protein (PDB Acc.No. 1STP) showing a biotin binding site (left) and the side-most amino acid chain loop (both panels). Panel **B** – Visualisations of the crystal structure of the Streptavidin protein, obtained from the PDB Protein Data Bank <http://www.pdb.org/>. Nine amino acids forming the loop, which can be modified or substituted are identified (bottom right corner). The loop (see also Figure 7) was used in the design of a display system based on the SAPV.

**Figure 7**

Fragment of the core streptavidin amino acid sequence. Panel **A** – the amino acid sequence of the fragment (see also Figure 6). The amino acid sequence loop (GTTEANAWK) links two antiparallel β -sheets (fragments underlined). Panel **B** – same amino acid sequence fragment with its secondary structure shown. The 9 amino acid loop could be modified and other protein fragments, peptides or tags could be inserted without destabilising the secondary structure of the core streptavidin sequence. Stabilisation of the secondary structure could be achieved by substituting the circled pairs of amino acids (dashed lines) with Cysteines. The seven pairs of amino acids are especially suitable due to their proximity to the loop and molecular architecture. The distances between corresponding C_{β} atoms in amino acid pairs (indicated on the panel **B** in Angstroms) are sufficient to accommodate two sulfhydryl groups and the resulting disulphide bond without major disturbances of the SAPV folding. The (Trp + Gly) pair is less suitable for (Cys + Cys) substitution due to Trp involvement in biotin binding.

We have shown that both post-translational and co-translational assemblies are achievable (Figures 9 and 10). Post-translational assembly is most useful if a large amount of one protein-nucleic acid complex is sought (e.g. for immunoprecipitation studies, for use instead of ordinary affinity reagents etc.). A co-translational assembly is necessary for a protein vector to assemble with its own DNA and should therefore be employed if protein vectors displaying different features are produced. There is another major difference between these two modes. Co-translational (as well as co-transcriptional) assembly depletes the available pool of DNAs (or mRNAs respectively), which would otherwise be transcribed or translated a number of times, which in turn reduces the efficiency of transcription and translation. It is therefore important to provide enough biotinylated DNA if co-transcriptional and co-translational assembly is attempted. Our approach

is nevertheless preferable to the "ribosome display" protocol [4,5], because in "ribosome display" both mRNAs and the components of the translational machinery (including ribosomes) are being depleted, resulting in extremely low efficiency of the protein synthesis. In the puromycin approach ("mRNA display") [6–8], the labelled mRNAs are also likely to crosslink in a non-specific manner with ribosomal proteins, thus reducing the overall efficiency of the reaction. The use of assembly sequences (as part of protein vectors) and their corresponding cognate regions or ligands results in non-covalent bonds between nucleic acid and its encoded expressed protein circumvents the need for cross-linking protein with its encoding nucleic acid or with a substrate. If required, however, the DNA and protein component of the self-assembled complex can be cross-linked to each other or to a substrate using known techniques [14].

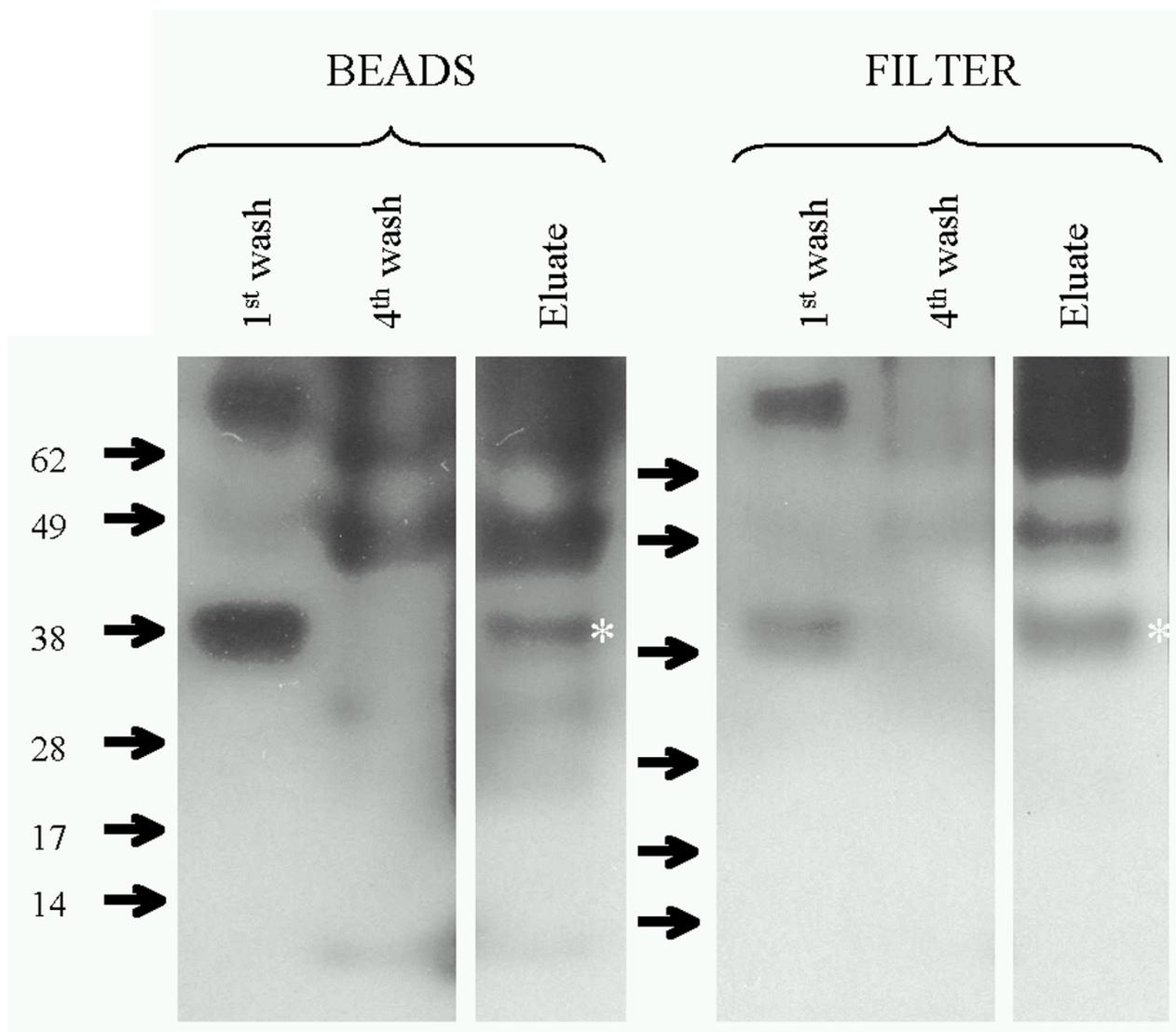


Figure 8
BCMP84 immunoprecipitation on protein A-conjugated glass beads and on nitrocellulose membrane. Recombinant BCMP84 was incubated with beads or nitrocellulose that had BCMP84 antibody bound to them. Samples from the first wash, 4th wash and the eluate from these incubations were run as indicated. The washes and eluates from the beads are on the left and the washes and eluates from the filter paper are on the right. White asterisks denote immunoprecipitated and eluted BCMP84 protein, which is not detected in the last (4th) wash prior to elution (both blots). The 1st wash, as expected, includes recombinant BCMP84 as indicated by the band at approximately 40 kDa. This band is not present in the 4th wash. Comparable amounts of the BCMP84 protein are present in the eluate of both the beads and nitrocellulose (marked with asterisks), indicating that it was selectively retained.

Conclusions

Protein vectors and the principle of self-assembly described here provide new exciting possibilities in molecular biology research. Because proteins can be directly linked to their nucleic acids, such self-assembled com-

plexes can be used for cloning proteins or protein affinity reagents (antibody, their fragments or antibody mimics, etc.). The ability to quickly generate thousands of affinity reagents may be a crucial factor in the development of protein affinity arrays [15–17]. Also, the ability to quickly

Table 1: Engineered SAPV displaying Albumin and BCMP84 peptides.

Construct:	Sequence fragment*
unmodified SAPV*	...NTQWLLTSG <u>TTEANAWK</u> STLVGHDT...
SAPV-Alb5	...NTQWLLTSG <u>HPYFYAPELLFFAK</u> STLVGHDT...
SAPV-84	...NTQWLLTSG <u>EGGKETLTPSELRLDV</u> STLVGHDT...

* Underlined in the SAPV sequence (only) is the fragment, which was modified.

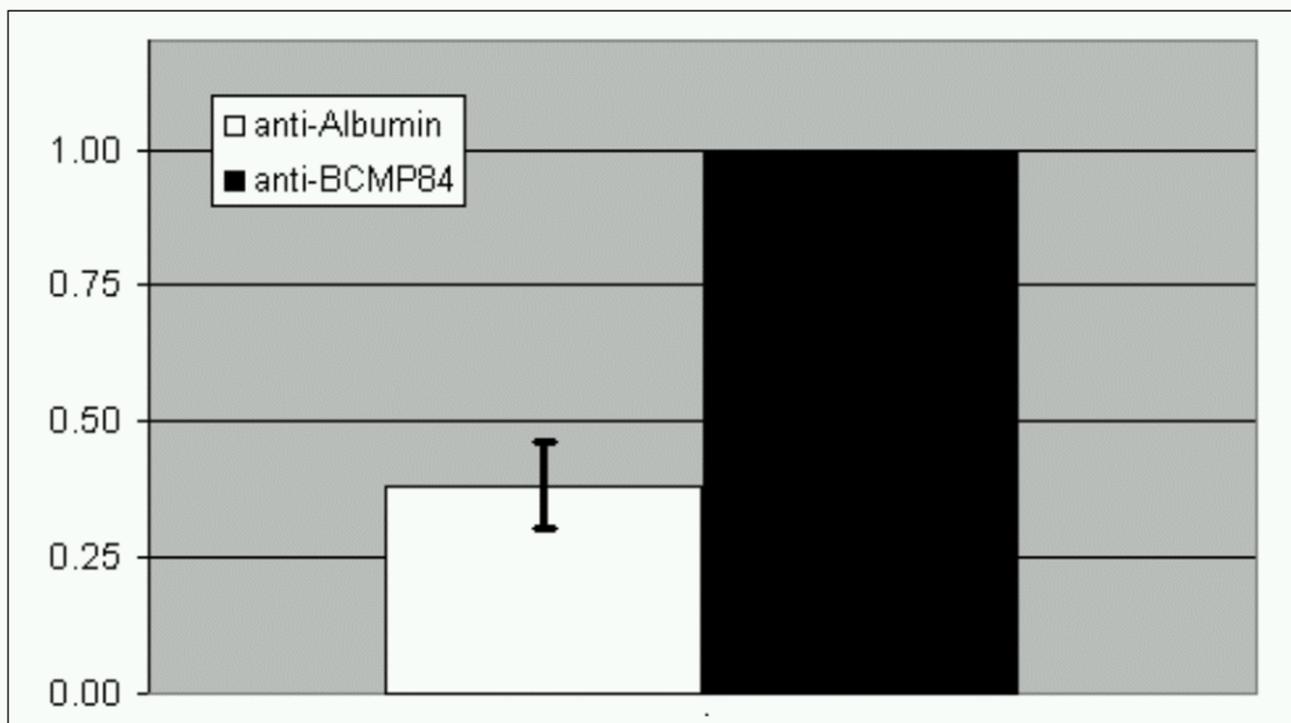
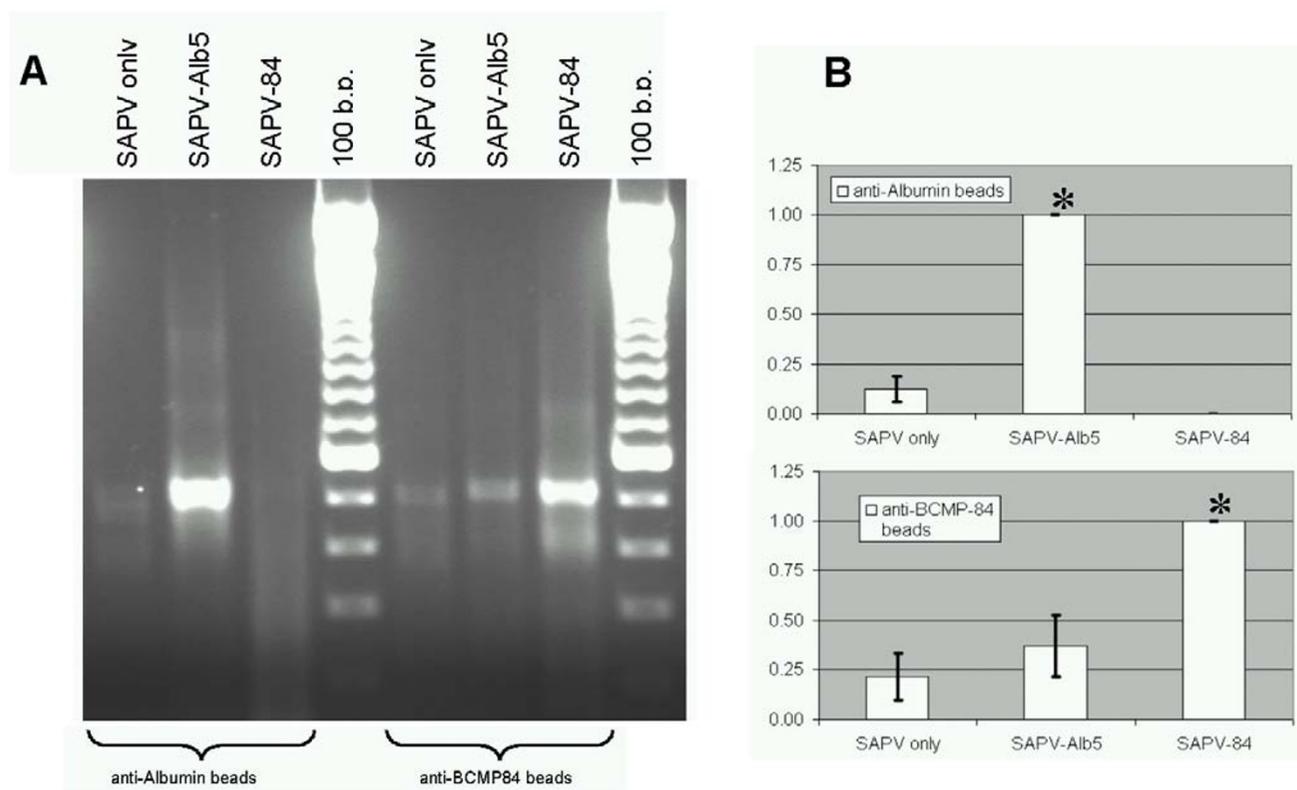


Figure 9
Co-immunoprecipitation of the assembled SAPV-84 Co-immunoprecipitation of the assembled SAPV-84 assayed by PCR amplification of the co-eluted SAPV-84 DNAs. Filled bar – normalised signal corresponding to the precipitation of the SAPV-84 on immobilised anti-BCMP84 antibody. Open bar – relative signal strength corresponding to the eluate from anti-Albumin antibodies. Error bar represents standard deviation (n = 5).

determine nucleic acid sequence and therefore to identify the associated proteins could be extremely helpful for protein affinity selection or directed protein evolution. Alternatively, proteins can be assembled with labelled nucleic acids, which can be done either co-translationally or post-translationally; the proteins thus become labelled to a high specific activity through their association with their nucleic acids, without being chemically modified. This could result in not only a higher specific activity of labelling but would also avoid the chemical modification that

takes place when proteins are labelled directly. Nucleic acid molecules associated with proteins could also improve sensitivity of detection of such proteins down to a single molecule level, by enabling detection using PCR. This could surpass all other known protein detection techniques, of which only immunogold detection in combination with electron microscopy is capable of detecting individual protein molecules [18–20].

**Figure 10**

Immunoprecipitation of the self-assembled SAPV vectors Immunoprecipitation of the self-assembled SAPV vectors (SAPV-84, SAPV-Alb5 and unmodified SAPV). Panel **A** – The eluates from anti-Albumin beads (left) or from anti-BCMP84 beads (right) were assayed by PCR and equal amounts of the resultant products were run on 2% agarose gel containing ethidium bromide. The gel indicates that anti-Albumin beads retain exclusively SAPV-Alb5 construct, whilst anti-BCMP84 beads precipitate SAPV-84 construct. The unmodified "empty" SAPV vector was retained by neither anti-Albumin nor anti-BCMP84 beads. Panel **B** – Immunoprecipitation of the self-assembled SAPV vectors (SAPV-84, SAPV-Alb5 and unmodified SAPV), data pooled from 3 experiments. The eluates from anti-Albumin beads (top panel) or from anti-BCMP84 beads (bottom panel) were assayed by PCR and quantified by agarose gel electrophoresis. Band intensities were determined using GeneSnap and fluorescent imager from SynGene (Cambridge, UK). All values shown are normalised to the strongest DNA sample in each case (marked by asterisks). Error bars represent standard deviation ($n = 3$). Largest eluted amounts of DNA (and therefore the corresponding SAPVs) were: SAPV-Alb5 (eluted from anti-Albumin beads) and SAPV-84 (eluted from anti-BCMP84 beads). The results confirm that anti-Albumin beads precipitate SAPV-Alb5 self-assembled construct, whilst anti-BCMP84 beads retain SAPV-84 construct.

Materials and Methods

Design of the SAPV protein vector

Full length nucleotide sequence coding for the SAPV was produced using overlapping synthetic oligonucleotides (Sigma-Genosys) and 3 rounds of PCR amplification. First round: a mixture of oligonucleotides (here and later see Table 2 for sequences) SA-1F + SA-2R + SA-3R + SA-4R + SA-5R + SA-6R + SA-7R + SA-8R was used (20 μ l of a mixture, each oligo at 1.25 μ mol/ μ l) plus 5 μ l of 10x Pfx buffer, plus 2 μ l 50 mM MgSO₄, plus 0.5 μ l 20 mM dNTPs, plus 0.5 μ l Pfx polymerase and H₂O to a total of

50 μ l. Both here and later, only proofreading DNA polymerase (Pfx polymerase, Invitrogen) was used. Cycling: following 5' at 96 °C, 20 cycles were done as follows 95 °C for 30", 72 °C (with a 2 °C decrement per cycle) for 30", 72 °C for 30" (with 1" increment per cycle). Final incubation was for 1' at 72 °C. Second round: the DNAs were re-amplified (5 μ l of original PCR product per reaction was used as a template) using a mixture of oligonucleotides SA-1F + SA-5R + SA-6R + SA-7R + SA-8R (1 μ l each at 10 μ mol/ μ l each). Other PCR conditions were as described above. Cycling: following 5' at 96 °C, 15 cycles

72°C for 1'30". Final incubation was 5' at 72°C. Following that, a fragment of the pIVEX vector (ROCHE) was used as a template for PCR amplification using T7-F and T7TER-R primers. Cycling was as follows: 5' at 96°C, 15 cycles of 96°C for 1', 40°C for 30" and 72°C for 1'. Final incubation was 5' at 72°C. Amplified fragment of the pIVEX vector was cloned into TOPO4BLUNT vector and a long 5' UTR was excised by PCR using primers RBS-2R and M13-R oligo. Cycling was as follows: 5' at 96°C, 15 cycles of 96°C for 1', 40°C for 30" and 72°C for 30". Final incubation was 5' at 72°C. The amplified UTR fragment was joined with the linearised tagged SAPV DNA (RBS-1F/T3 fragment, see above) by PCR using primers T7-F and T7TER-R and a mixture of the linearised tagged SAPV DNA (RBS-1F/T3 fragment) and the cloned re-amplified pIVEX (RBS-2R/M13-R) fragment as a template. Cycling was as follows: 5' at 96°C, 10 cycles of 96°C for 1', 72°C (with decrement of 3°C per cycle) for 30", 72°C for 1'30" and 10 cycles of 96°C for 1', 40°C for 30" and 72°C for 1'30". Final incubation was 5' at 72°C. The amplified DNA was cloned into the TOPO4BLUNT vector. Final sequence of the engineered tagged SAPV is shown on Figure 3.

In vitro coupled transcription and translation reactions

For temperature optimisation the T&T reactions (20 ul final volume each) were assembled according to manufacturer recommendations, but synthesis was done at a range of temperatures (2 ug of the SAPV-AFP DNA was added to each tube, Figure 4, left panel). To optimise the amount of DNA used for T&T reactions, different amounts of DNA were added (by Ethanol precipitating the precalculated amount of the PCR product prior to assembling the in vitro T&T reaction, Figure 4, right panel). Reactions were run overnight. 5 ul aliquots of each of the T&T reactions were loaded onto a precast 4–12% NuPAGE gel (Invitrogen). The proteins were resolved by SDS-polyacrylimide gel electrophoresis and transferred onto nitrocellulose (0.2 uM pore size, Invitrogen) using an Xcell SureLock Minicell and Blot Module according to the manufacturer's instructions. The blot was then blocked for 1 hour in TBST (TBS plus 0.1% tween-20) with 2% powdered milk and probed with a 1:3000 dilution of AbCam anti-GFP Rabbit polyclonal (1 hr room temp) in TBST/milk. After washing in TBST (3x, 5–10' each wash) the blot was probed with 1:6000 HRP-labelled Anti-Rabbit IgG, (Amersham Pharmacia) in TBST/milk (1 hr room temp), washed again and then developed using ECL (Amersham) according to the manufacturer's instructions and exposed to ECL Hyperfilm (Amersham).

Assembly of the SAPV protein vector with biotinylated DNA

Untagged SAPV was obtained by means of the in vitro T&T as described above and using SAPV DNA lacking STOP codons (see legend to Figure 2). This DNA was obtained by

PCR using M13F and SA-7R primers and tagged SAPV DNA (Figure 3) as a template. Cycling was as follows: 6' at 96°C, and 15 cycles of 96°C for 1', 40°C for 30", 72°C for 1'. Final incubation was 5' at 72°C. SAPV DNA was used for the T&T reaction. Following an overnight incubation, the T&T reaction was spun for 3 min at 15,000 RPM in a microcentrifuge to precipitate insoluble components of the in vitro reaction mixture. Clear supernatant was transferred to fresh tube prior to adding DNAs for assembly. Biotinylated and non-biotinylated DNAs for assembly were generated by PCR using T7 forward primer (biotinylated or non-biotinylated, respectively) and non-biotinylated T7TER-R primers and the long DNA coding the tagged SAPV (Figure 3) as a template (all other conditions were as described previously). The longer DNAs were chosen for assembly reactions to avoid non-specific background due to SAPV DNA used for in vitro T&T. DNAs were ethanol-precipitated prior to assembly and redissolved in water at 1 ug/ul. Cleared T&T supernatants were aliquoted (10 ul per tube) and DNAs (biotinylated/non-biotinylated) were added (5 ug per tube). Assembly reactions were allowed to run overnight at +4°C. Protein-DNA complexes were separated from free DNAs by filtration through protein-binding microcentrifuge filters ("Ultrafree-MC Probind Units" modified PVDF, Millipore). After 4 washes (by flow through filtration) the retained materials were eluted by incubation for 30' with gentle agitation in 50 ul volume 0.1 x TAE. Eluted DNAs were detected by PCR as follows: 10 ul of each of the wash through and eluates from each assembly reaction were amplified in parallel using primers T7-F and T7TER-R. 35 cycles of amplification of 1' at 96°C, 30" at 40°C and 1'30" at 72°C were carried out. Amplified products were separated on 2.5% agarose gels containing Ethidium Bromide. Equal amounts of each PCR reaction were loaded onto each lane (Figure 5A,5B).

Display system based on the SAPV protein vector

To illustrate the "display" capabilities of the SAPV, we engineered SAPV displaying peptide fragments of Albumin and BCMP84 proteins (Table 1). The DNA coding for the modified SAPV were obtained by PCR using SAPV DNA as a template and synthetic oligonucleotide primers M13F plus loop-84-1R (to make SAPV-84 construct) or M13F plus loop-Alb5-R (to make SAPV-Alb5 construct), see Table 1. Stop codons were added to both constructs by PCR using M13F and SA-10R primers. Cycling was as follows: 5' at 96°C, and 30 cycles of 96°C for 30", 54°C for 30", 72°C for 30". Final incubation was 5' at 72°C. Large amounts of the full length DNAs coding for all SAPV variants (both biotinylated and non-biotinylated) were produced for in vitro T&T by PCR using T7-F forward and T7TER-R reverse primers as described earlier for SAPV vector.

Co-immunoprecipitation system for affinity separations

A co-immunoprecipitation system for affinity separations was designed to quickly separate different SAPVs. To test the system a recombinant BCMP84 was used. We tested glass bead-based and nitrocellulose-based systems separately as follows. Twenty microlitres of protein A-conjugated glass beads (PROSEP-A, Millipore) were washed 3 times in 1 ml PBS then incubated with 20 ul of anti-BCMP84 antibody (rabbit polyclonal, 110 ug/ml). Nitrocellulose was wetted in deionised water for 5' then cut into 3-mm squares and incubated with 20 ul of anti-BCMP84 antibody. The beads and nitrocellulose squares were washed twice in 1 ml PBS then blocked by incubation in 3% powdered milk in PBS for 30'. Blocking buffer was removed and the beads and nitrocellulose squares were incubated for 90' with 2.5 ug of recombinant BCMP84 in PBS with 0.5% powdered milk. Beads and protein solution were transferred to Vectaspin microcentrifuge tubes containing a 0.2 um pore Anapore membrane (Millipore) and the beads were washed 3 times by resuspension in 600 ul PBS followed by centrifugation. After a final wash in 50 ul PBS, the beads were resuspended in 50 ul elution buffer (100 mM glycine, pH 2.45), shaken periodically over 10' and then spun. The eluted sample was neutralised by the addition of 13 ul 2 M NaOH. One millilitre PBS was added to the incubations of the nitrocellulose squares and protein. The nitrocellulose squares were washed 2 times by transferral to 1 ml fresh PBS followed by brief shaking. After a final wash in 50 ul PBS, 50 ul elution buffer was added to the nitrocellulose which was then shaken periodically over 10' before the eluate was removed and then neutralised by the addition of 13 ul 2 M NaOH. 10 ul of the eluate, the final wash and the first wash were run on a 4–12% NuPage 1D polyacrylamide gel (Invitrogen) under non-reducing conditions and transferred to a nitrocellulose membrane (0.2 um pore size, Invitrogen) by western blotting. After blocking the nitrocellulose by incubation in 2% powdered milk in TBST (TBS with 0.1% tween-20) the blots were probed using 0.5 ug/ml anti-BCMP84 in TBST plus 2% powdered milk. The blots were washed extensively in TBST and probed with an HRP-conjugated, anti-rabbit secondary antibody (1:6000 dilution, Amersham) washed extensively and then developed using ECL (Amersham Pharmacia Biotech) according to the manufacturer's instructions (see Figure 8).

Affinity precipitation of the SAPVs displaying BCMP84 peptide

In vitro T&T (50 ul volume) was run as described above using in-vitro coupled Transcription/Translation (T&T) kit from Roche (RTS 100 *E. coli* HY) and DNA coding for the SAPV-84 (see Table 1). Following an overnight incubation at 21 °C the reaction was spun down for 3' to separate insoluble components of the T&T from the

synthesised SAPV. The cleared T&T supernatant was kept at 4 °C until assembly reaction. 20 ul of the PCR-amplified unpurified biotinylated SAPV-84 DNA was added to 20 ul of the cleared supernatant and the assembly mixture was incubated overnight at 4 °C. The assembled SAPV-DNA complex was immunoprecipitated with either anti-BCMP84 or anti-albumin (control) antibodies as follows. Nitrocellulose was wetted in deionised water for 5' then cut into 3-mm squares and incubated with either 20 ul of anti-albumin antibody or 20 ul of anti-BCMP84 antibody. The nitrocellulose squares were washed twice in 1 ml PBS then blocked by incubation in 2% powdered milk and 0.1 mg/ml tRNA in PBS for 30 minutes. Blocking buffer was removed and the nitrocellulose squares were incubated for 60' with the assembled SAPV-DNA complex. The nitrocellulose was washed three times in wash buffer (1 ml PBS supplemented with 0.01 mg/ml tRNA and 2% powdered milk). After a final wash in 50 ul wash buffer, the nitrocellulose squares were incubated with in 50 ul elution buffer (80 mM glycine, pH 2.45) supplemented with 0.4% powdered milk and 0.01 ug/ml tRNA. The eluted samples were neutralised by the addition of 11 ul 2 M NaOH. The presence of the SAPV-84 in the eluted samples was detected by PCR amplification of the SAPV-84 DNAs using primers T7-F and T7TER-R. All reactions were run in parallel under identical conditions. 33 cycles of amplification of 30" at 96 °C, 30" at 42 °C and 30" at 72 °C were carried out. PCR products were run on a 2% gel and band intensities determined using GeneSnap and fluorescent imager from SynGene. The results of 5 measurements are presented on Figure 9.

Self-assembly of protein vectors with their DNAs and affinity separation

In vitro synthesis and self-assembly of the SAPV-84, SAPV-Alb5 and "empty" SAPV-only (unmodified) protein vectors were done using RTS 100 *E. coli* HY kit from Roche. Optimal reaction conditions were used (as determined earlier, see Figure 4), except that each reaction contained ~16 ug of biotinylated SAPV DNAs (SAPV-84, SAPV-Alb5 and SAPV-only coding DNAs) and a total volume of each reaction was 100 ul. The combined in vitro [T&T self-assembly] reactions were allowed to run at 21 °C. Following an overnight incubation the reactions were cleared by spinning for 3' at 15,000 RPM in a microcentrifuge and two 40 ul aliquots from each reaction were transferred to fresh tubes. The SAPVs were precipitated from these aliquots by incubation with either anti-BCMP84 or anti-Albumin antibodies using protein A-conjugated glass beads (120 ul, PROSEP-A, Millipore). Beads were prepared by washing 3 times in 1 ml PBS. They were then incubated with either 600 ul of anti-albumin antibody (rabbit polyclonal, 2.8 mg/ml) supplemented with 0.1 mg/ml tRNA or 600 ul of anti-BCMP84 antibody (rabbit polyclonal, 110 ug/ml) supplemented with 0.1 mg/ml tRNA for 90

minutes. All subsequent washes, incubations and elutions were carried out in the presence of 0.01 mg/ml tRNA. Both sets of beads were washed 3 times and then split into separate tubes each containing 20 ul of beads. Each aliquot of beads was then incubated with 40 ul of supernatant from one of the self-assembled reactions for 90 minutes. Each sample was then transferred to a macroporous Wizard[®] Minicolumn filtration unit (Promega). The column with beads was washed with 80 ul of PBS by centrifugation, following which each column was washed with 40 ml of PBS using a 50 ml syringe. The filtration unit was transferred into a microcentrifuge tube and the beads were washed with 50 ul PBS by centrifugation. Finally the beads were resuspended in 50 ul elution buffer (100 mM glycine, pH 2.45) and then the eluates were collected by centrifugation and neutralised with 13 ul of 2 M NaOH. The eluted SAPV-DNA complexes were ethanol precipitated and assayed by PCR using T7-F and T7TER-R primers. Equal amounts of the PCR reactions were run on 2% agarose gel (see Figure 10).

References

- Emel'ianov VV, Kalinin VN, Girdo BM and Tikhonenko TI **Colicin EI synthesis in a cell-free system of paired transcription – translation.** *Biull Eksp Biol Med* 1976, **82**:1324-1326
- Spirin AS **Cell-free systems of polypeptide biosynthesis and approaches to the evolution of translation apparatus.** *Orig Life* 1976, **7**:109-118
- Erickson AH and Blobel G **Cell-free translation of messenger RNA in a wheat germ system.** *Methods Enzymol* 1983, **96**:38-50
- He M and Taussig MJ **Antibody-ribosome-mRNA (ARM) complexes as efficient selection particles for in vitro display and evolution of antibody combining sites.** *Nucleic Acids Res* 1997, **25**:5132-5134
- He M, Menges M, Groves MA, Corps E, Liu H, Bruggemann M and Taussig MJ **Selection of a human anti-progesterone antibody fragment from a transgenic mouse library by ARM ribosome display.** *J Immunol Methods* 1999, **231**:105-117
- Xu L, Aha P, Gu K, Kuimelis R, Kurz M, Lam T, Lim A, Liu H, Lohse P, Sun L, Weng S, Wagner R and Lipovsek D **Directed evolution of high-affinity antibody mimics using mRNA display.** *Chem Biol* 2002, **9**:933
- Roberts RW and Szostak JW **RNA-peptide fusions for the in vitro selection of peptides and proteins.** *Proc Natl Acad Sci U S A* 1997, **94**:12297-12302
- Liu R, Barrick JE, Szostak JW and Roberts RW **Optimized synthesis of RNA-protein fusions for in vitro protein selection.** *Methods Enzymol* 2000, **318**:268-293
- Argarana CE, Kuntz ID, Birken S, Axel R and Cantor CR **Molecular cloning and nucleotide sequence of the streptavidin gene.** *Nucleic Acids Res* 1986, **14**:1871-1882
- Blake KR, Murakami A and Miller PS **Inhibition of rabbit globin mRNA translation by sequence-specific oligodeoxyribonucleotides.** *Biochemistry* 1985, **24**:6132-6138
- Haeuptle MT, Frank R and Dobberstein B **Translation arrest by oligodeoxynucleotides complementary to mRNA coding sequences yields polypeptides of predetermined length.** *Nucleic Acids Res* 1986, **14**:1427-1448
- Marcus-Sekura CJ, Woerner AM, Shinozuka K, Zon G and Quinnan GV Jr **Comparative inhibition of chloramphenicol acetyltransferase gene expression by antisense oligonucleotide analogues having alkyl phosphotriester, methylphosphonate and phosphorothioate linkages.** *Nucleic Acids Res* 1987, **15**:5749-5763
- Vlasov VV and Lurchenko LV **Mechanism for suppression mRNA translation with antisense oligonucleotides.** *Mol Biol (Mosk)* 1990, **24**:1157-1161
- Wong SS **A Chemistry of Protein Conjugation and Cross-Linking.** CRC Press 1993,
- Soloviev M **EuroBiochips: spot the difference.** *Drug Discov Today* 2001, **6**:775-777
- Barry R, Scrivener E, Soloviev M and Terrett J **Chip-Based Proteomics Technologies.** *Int Genomic Proteomic Technology* 2002, **Feb**:14-22
- Scrivener E, Barry R, Platt A, Calvert R, Masih G, Hextall P, Soloviev M and Terrett J **Peptidomics: a new approach to affinity protein microarrays.** *Proteomics* 2003,
- Nusser Z, Roberts JD, Baude A, Richards JG and Somogyi P **Relative densities of synaptic and extrasynaptic GABAA receptors on cerebellar granule cells as determined by a quantitative immunogold method.** *J Neurosci* 1995, **15**:2948-2960
- Baude A, Nusser Z, Molnar E, McIlhinney RA and Somogyi P **High-resolution immunogold localization of AMPA type glutamate receptor subunits at synaptic and non-synaptic sites in rat hippocampus.** *Neuroscience* 1995, **69**:1031-1055
- Bernard V, Somogyi P and Bolam JP **Cellular, subcellular, and sub-synaptic distribution of AMPA-type glutamate receptor subunits in the neostriatum of the rat.** *J Neurosci* 1997, **17**:819-833

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

