

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
6 February 2003 (06.02.2003)

PCT

(10) International Publication Number
WO 03/010176 A2

(51) International Patent Classification⁷: **C07H 1/00**

(21) International Application Number: PCT/GB02/03351

(22) International Filing Date: 23 July 2002 (23.07.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0118031.4 24 July 2001 (24.07.2001) GB

(71) Applicant (for all designated States except US): **OXFORD GLYCOSCIENCES (UK) LTD** [GB/GB]; Attn: Mary Gadsden, The Forum, 86 Milton Park, Abingdon, Oxfordshire OX14 4RY (GB).

(72) Inventor; and

(75) Inventor/Applicant (for US only): **SOLOVIEV, Mikhail** [GB/GB]; The Forum, 86 Milton Park, Abingdon, Oxfordshire OX14 4RY (GB).

(74) Agent: **BLAKEY, Alison, Jane**; Oxford GlycoSciences (UK) Ltd, The Forum, 86 Milton Park, Abingdon, Oxfordshire OX14 4RY (GB).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

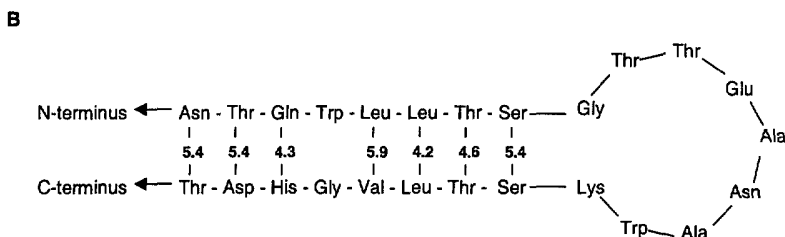
Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: PROTEIN-NUCLEIC ACID COMPLEXES

A
NTQWLLTSGTTEANAWKSTLVGHDT



(57) Abstract: Methods for the production of recombinant proteins that self-assemble into complexes with nucleic acids in solution or on supports, and protein-nucleic acid complexes produced according to the method e.g. in array format.

Protein-Nucleic Acid Complexes

The invention provides novel methods for the production of recombinant proteins or protein oligomers that self-assemble into complexes with nucleic acids in solution or, when the nucleic acid is presented, or is presentable as an array, that self-assemble to produce a protein array.

Large scale scanning of the human genome has become possible with the introduction of DNA microarray technology. The ability to survey the expression of up to 5,000 to 50,000 genes in a single experiment provides significant new opportunities as well as new challenges. It is however, evident that knowledge of the gene sequence or quantity of gene expression is not sufficient to predict the biological nature and function of a protein. Of particular interest is the development of protein-based chips that could be used to array protein substrates that could then be used for drug screening or diagnostic tests.

Protein microarrays are potentially powerful tools in proteomics. A protein function array can consist of thousands of native proteins immobilized on a substrate ready for parallel testing of protein function. Alternatively protein-binding agents may be arrayed, permitting expression profiling at the protein level. Different methods are known in the art for attaching proteins to solid supports but these can involve cross-linking and loss of protein function (see PCT Application No. WO 00/32823). Also, enough purified protein needs to be prepared in its native conformation, something that is difficult to ensure. Indeed, many proteins are not amenable to over-expression in, for example, *E. coli* or baculovirus systems. Therefore, proteins expressed using *in vitro* transcription/translation methods where correctly-folded protein can be produced, are more attractive.

Arrays displaying nascent peptides and proteins in an antibody-ribosome-mRNA (ARM) complex allowing selection of complexes carrying a particular peptide or protein by means of binding to a ligand, antibody or antigen and subsequently recovering the nucleic acid encoding the peptide or protein of interest from the complex, have been reported (PCT Application No. WO 98/54312; EP 0985032; Hanes *et al.*, 2000, Nat. Biotech. 18: 1287-1292). The production of protein-DNA and protein-RNA complexes has also been reported (US 6,143,503; PCT Application Nos. WO 99/51773, WO 01/68671 and WO 00/32823; Roberts, R. and Szostak, J. 1997, Proc. Natl. Acad. Sci. 94: 12297-12302; US 6,207,446). The latter references rely on the use of puromycin as a label or adaptor in order to produce either protein-RNA or protein-DNA hybrids (using reverse transcription) and suffer from the inherent problems associated with RNA instability, non-specific cross-linking to ribosomal proteins or multi-step processes that involve RNA handling.

The direct transfer of technologies used in nucleic acid-based screening approaches to protein-based screening approaches is not possible because of the fundamentally different nature of nucleic acids and proteins. For example, the advantages conferred by the use of the polymerase chain reaction (PCR) in nucleic acid-based approaches are irrelevant to protein-based approaches because that technology does not result in the amplification of protein, and there is no equivalent means for the amplification of proteomes, or even mixtures of proteins, *in vitro*.

The present invention provides a method for the production of self-assembling protein-nucleic acid complexes using transcription followed by translation or using coupled *in vitro* transcription/translation systems, thus providing an efficient cloning vehicle for the cost-effective production of quantities of proteins of interest.

Accordingly, the invention provides a method for the production of a protein-nucleic acid complex comprising:

- (a) providing a nucleic acid comprising code for a protein vector, said protein vector comprising an assembly sequence and a target protein sequence; and
- (b) performing transcription and translation, wherein upon expression of the protein vector, the encoded assembly sequence binds to its encoding nucleic acid at a cognate sequence present within, or attached to, said nucleic acid.

The term 'protein' includes any molecule that comprises a sequence of naturally-occurring or non-naturally-occurring amino acid residues. Amino acid residues that are post-translationally modified are also included, as are polypeptides, oligopeptides and peptides.

'Cognate sequence' or 'cognate biomolecular sequence' includes a sequence that is recognized by, or recognises a binding partner such that an interaction occurs. Such a sequence may comprise nucleic acid, amino acid, or other biomolecular or molecular sequences. Such binding or interactions, for example but without limitation, can be specific or non-specific, covalent or non-covalent, ionic or non-ionic.

Preferably, the complexes of the present invention are protein-DNA complexes but it is also apparent that the method of the invention also applies to the preparation of protein-RNA self-assembled complexes. The following text will now refer to protein-DNA complexes by way of explanation but not of limitation.

'Protein vectors' comprise (i) an assembly sequence that permits the binding of the protein vector to its encoding DNA and (ii) a target protein sequence, i.e. the protein sequence of interest. The 'assembly sequence' includes that part of a protein vector which binds to a cognate biomolecular sequence or site, for the purpose of immobilizing said protein vector directly or indirectly to its encoding DNA in a sequence specific or non-specific manner. In one embodiment, an assembly sequence can consist of, or comprise a target sequence. On direct binding of the assembly sequence of the protein vector to its encoding DNA, a protein-DNA complex is formed i.e. a 'self-assembled' protein-DNA complex. Alternatively, the assembly sequence binds indirectly to its encoding DNA by binding to a cognate sequence, e.g. a label, associated with the encoding, for example but without limitation, by labelling the encoding DNA with an appropriate molecule or ligand comprising said cognate sequence using methods known in the art. Labelling of the encoding DNA can be performed during transcription or translation or alternatively can be performed before or afterwards. Upon transcription and translation of the nucleic acids, the protein vector assembly sequence (which is a cognate sequence for the ligand used to label the DNA) binds to the label attached to its encoding DNA. Such labelled DNA and corresponding protein vectors comprise, for example but without limitation, interactors such as biotin and avidin, streptavidin or related proteins, derivatives or protein fragments that bind biotin,

where biotin is the DNA label and avidin, streptavidin etc., is the assembly sequence of the protein vector encoded by the labelled DNA.

Other means for the production of self-assembled complexes are envisaged and include, but are not limited to, an assembly sequence coding for stretches of positively charged amino acids such as lysine or arginine where, upon transcription and translation of the nucleic acids, the resultant positively charged assembly sequence will bind to the negatively charged DNA. In one aspect of the invention, sequence-specific pairs of interactions can be used to produce self-assembled protein-DNA complexes. Examples of these include, but are not limited to, providing a nucleic acid comprising code for a binding site of a nucleic acid-binding protein and providing a protein vector comprising the cognate nucleic acid-binding protein as the target protein. Proteins that bind to nucleic acids are known in the art (see for example US 5,270,163). The use of known transcription factors and their target DNA sequences is a preferred example of such sequence-specific interactions.

Preferably, binding of the assembly sequence to its encoding DNA occurs during protein synthesis whilst the peptide chain of the protein vector is still attached to a ribosome, i.e. the nascent protein binds to its encoding DNA. Alternatively, binding occurs upon the completion of protein synthesis after the protein vector has been released from the ribosome. Alternatively, complexes can be assembled step-wise, by firstly producing the protein vector using an appropriate protein expression system followed by addition of DNA encoding said protein vector, resulting in the association of the protein vector and its encoding DNA via the assembly sequence and hence the formation of a protein-DNA complex. In one embodiment, the protein vector of the protein-DNA complex may also bind to a suitable substrate directly or indirectly, or to a suitable molecule or ligand comprising a cognate sequence which can be attached to a suitable substrate. 'Suitable substrate' refers to any suitable solid or semi-solid support such as, but not limited to, glass, silicon, bead of material or other suitable materials.

In a preferred embodiment, an assembly sequence can comprise a target protein sequence wherein the target protein sequence is present within the assembly sequence i.e. interrupting the assembly sequence. In this embodiment, the target protein sequence preferably does not substantially alter the folding, if any, of the assembly sequence. In effect the target protein is "displayed" within the assembly sequence. Alternatively, the target protein sequence is not present within the assembly sequence but is present elsewhere within the protein vector, preferably C-terminal to the assembly sequence or alternatively, N-terminal to the assembly sequence.

In another embodiment, the protein vector additionally comprises an affinity tag. The affinity tag may be adapted to permit purification of the protein-DNA complex. The affinity tag may also permit identification and/or quantitation of a target protein/protein interaction or the interaction of the protein-DNA complex with another compound of interest. 'Compounds of interest' include biomolecules such as proteins, polypeptides, peptides, antibodies, antigens, nucleic acids and also other organic and non-organic molecules. The term 'affinity tag' is well known in the art and includes a region of protein sequence that recognises or is

recognised by a ligand or other molecule or biomolecule. Biomolecules include proteins, nucleic acids and other molecules of biological origin. The affinity tag supplies a means of isolating the protein-DNA complex from a solution or a mixture of protein-DNA complexes in solution, the protein-DNA complex may comprise at least two different affinity tags. In a particular embodiment, the affinity tag of the protein vector serves to immobilise the protein-DNA complex to a suitable substrate either directly or indirectly. Thus, the protein-DNA complexes of the invention can be presented in a spatial format as an array. The affinity tag is preferably present at the opposite end of the protein vector to the assembly sequence, e.g. the assembly sequence and the affinity tag are situated at, or near, opposite termini of the target protein sequence. However the affinity tag can, without limitation, be adjacent to the assembly sequence or separated from the assembly sequence by a linker (see below), which can be cleavable or non-cleavable.

In another embodiment, the DNA component of the protein-DNA complex can be used as a unique tag for characterising protein/protein interactions and hence can be used for separation, purification, identification and/or quantitation of a protein-DNA complex/protein interaction or the interaction of the protein-DNA complex of the invention with any other compounds of interest. The protein-DNA complexes can also be used to detect protein-nucleic acid interactions, for example but without limitation, transcription factor-nucleic acid interactions where the target sequence of the protein-DNA complex comprises a transcription factor or fragment of a transcription factor. The embodiments described above can be performed in a competitive assay format.

In a preferred embodiment, the protein vector also comprises a detectable protein sequence where 'detectable protein sequence' includes a protein sequence that is incorporated into the protein vector as a means of enabling the qualitative or quantitative measurement of the expression of the protein vector. For example, but without limitation, the detectable protein sequence may be a fluorescent protein such as autofluorescent protein, green fluorescent protein (GFP) or dsRed fluorescent protein. The detectable protein sequence can also be used as a marker for detecting the location and/or the quantity of a protein-DNA complex. Preferably, the detectable protein sequence is present at, or near, the C-terminus of the protein vector such that expression of the detectable protein sequence indicates successful expression of the preceding protein sequence or sequences.

'Linkers' comprise a nucleic acid sequence (or amino acid sequence) present between the nucleic acid code (or amino acid sequence) of, for example:

- (i) an assembly sequence or sequences and a target protein sequence; and/or
- (ii) an assembly sequence and an affinity tag; and/or
- (iii) an affinity tag or tags and a target protein sequence; and/or
- (iv) an assembly sequence and a detectable protein sequence; and/or
- (v) a detectable protein sequence and a target protein sequence.

A linker can also contain one or more consensus sites permitting cleavage using a known enzyme or other molecule. Cleavable linkers and their corresponding enzymes or other molecules are well known in the art, for example but without limitation, tobacco etch virus (TEV) cleavable sequences, factor X cleavable sequences and thrombin cleavable

sequences. A cleavable linker allows release of the target protein from its encoding DNA and assembly sequence by cleaving said linker. Where the protein vector additionally comprises an affinity tag that is separated from the target protein by a cleavable linker, the affinity tagged target protein which has been released from its encoding DNA can be isolated or purified using said affinity tag; the target protein can then be further released from its affinity tag by cleavage of the linker.

It is understood that any combination of assembly sequence, target protein sequence, affinity tag, detectable protein sequence and linker or linkers can be used.

In an alternative embodiment, the protein, polypeptide and/or peptide sequences comprising the protein vector are contiguous.

According to a further aspect the invention provides a protein-nucleic acid complex produced according to the method of the invention.

Most preferably, the protein vector of the protein-DNA complex is attached to its encoding DNA via an assembly sequence preferably present at, or near the N-terminus of the target protein. Alternatively, the protein vector of the protein-DNA complex is attached to its own DNA via an assembly sequence present at, or near the C-terminus of the target protein sequence.

As mentioned previously in a further one embodiment, the assembly sequence also comprises the target protein sequence or other sequence. For example but without limitation, the target protein component of the protein-DNA complex is present (displayed) within the assembly sequence which is attached to its encoding DNA either directly or indirectly, for example via a cognate molecule. One example of this provides streptavidin as assembly sequence within which a target protein sequence is present. Nucleic acid encoding streptavidin into which code for a target protein sequence has been inserted, is used to generate a protein vector that will subsequently assemble with its encoding DNA, said DNA being biotinylated. In one embodiment, one or more of the 8 amino acid pairs of the streptavidin core sequence forming anti-parallel strands immediately precede the side most loop of the molecule (see Figure 1) can be mutated to, for example, cysteine residues to limit any tertiary structural changes that may occur on insertion of a target protein sequence. This is possible because the distances between respective pairs of amino acids in these two anti-parallel strands (Figure 1 indicates these two anti-parallel 8 amino acid stretches and the distance between the pairs of residues) allow cysteine substitutions without major changes to the streptavidin folding pattern.

The complexes of the invention can be used to identify, without limitation, target protein/protein and target protein/nucleic acid interactions as well as interactions of a target protein with other compounds of interest. The detection and/or quantitation of such interactions can be achieved using conventional methods of analysing the nucleic acid component of the protein-DNA complex following their optional amplification and release from said complexes. For example, the target protein which interacts with a compound of interest can be identified by its association with its encoding DNA.

Thus according to a further aspect the invention provides a method for detecting the interaction of a target protein with a compound of interest comprising:

- (a) providing a protein-nucleic acid complex according to the method of the invention;
- (b) contacting the protein-nucleic acid complex with a compound of interest; and
- (c) detecting the interaction of the target protein with a compound of interest.

The complexes of the invention can also be presented attached to a substrate e.g. a solid or semi-solid support. The complexes may therefore be provided in a spatial format as an array of protein-DNA complexes on a substantially planar surface or in wells or on beads for screening, for example, compounds of interest for binding. In this manner the complexes are of particular utility for detecting the interaction of target proteins with a compound of interest.

Thus according to a further aspect the invention provides a spatial array of protein-nucleic acid complexes produced according to the method of the invention immobilised to a support.

Spatial arrays can be produced using means known in the art to directly attach nucleic acids directly to a suitable substrate. Alternatively, the protein-DNA complexes can be attached indirectly to a suitable substrate using standard means known in the art, for example but without limitation, by using nucleic acid hybridisation to immobilised complementary nucleic acids on a suitable substrate. An alternative means of spatial separation can be achieved by attaching the protein-DNA complexes of the invention to beads by, for example but without limitation, hybridisation to complementary nucleic acids attached to beads. Thus, the protein-DNA complexes of the invention can be used as an array to characterise antigens or other molecules or biomolecules. A target protein binding to such a compound of interest may then be identified by virtue of it being associated with the DNA encoding the protein vector comprising said target protein; the DNA may be amplified and sequenced or sequenced directly, thereby revealing the protein sequence it encodes.

In yet another embodiment, the protein-DNA complexes of the present invention can be presented in a spatial format as an array using the affinity tag of the protein vector, wherein the affinity tag binds to a cognate sequence present on a substrate, e.g. in a spatial format as an array. For example, the assembly sequence may be an amino acid sequence recognised by an antibody which is present as an array or alternatively on a bead or other suitable substrate. Alternatively, an array of protein-DNA complexes may be produced by binding of the affinity tag comprising, for example but without limitation, avidin, streptavidin or a derivative or fragment of such that binds to biotin where said biotin is presented in a spatial format as an array.

Preferably, the complexes of the invention are used to produce affinity reagents. In this embodiment, the target protein sequence may be a monoclonal, a divalent or polyvalent antibody sequence, an antibody fragment or antibody mimic sequence or a complementarity determining region (CDR) or other affinity reagent sequence. In one embodiment, the affinity reagent-DNA complexes can be presented in a spatial format as an array. Such an array of affinity reagent-DNA complexes can be screened for interacting molecules such as antigens or other biomolecules by, for example but without limitation, contacting an array of

affinity reagent-DNA complexes with a sample of interest and detecting the interaction, again by way of example but not of limitation, by detecting a fluorescent signal such as that of a detectable protein sequence present within the protein vector. Alternatively, the DNA component of the self-assembled protein-DNA complex is labelled for example, with a fluorescent molecule. Interactions are identified for example, but without limitation, by detecting and/or quantitating the fluorescently-labelled DNA of the self-assembled protein-DNA complex. The DNA of the complexes of the invention can be labelled to a high degree using well-established protocols known in the art, for example using fluorescent, radioactive or isotopic labels, or nucleic acid derivatives. Suitable labelling can be achieved using for example but without limitation, primer labelling, using labelled nucleic acid analogues and derivatives during initial synthesis and/or during amplification steps if any, prior to the transcription and translation stage, labelling using PNK, DNA and RNA polymerases, DNA and RNA ligases and by chemical modification. Nucleic acid labelling provides not only a means for labelling to a high specific activity, but also avoids the problem of compromising the folding, specificity and/or function of the protein component of the protein-DNA complex as protein labelling is avoided. The DNA of the protein-DNA complex can be characterised by means known in the art such as electrophoresis, sequencing or hybridisation after optional amplification and release from the protein-DNA complex. Hence, a protein-DNA complex where the protein component comprises an affinity reagent such as but not limited to an antibody, can be used as a cloning vehicle for said affinity reagent by performing transcription and translation to produce a quantity of protein. An affinity reagent binding to such an antigen or other molecule or biomolecule can also be identified by virtue of being associated with the DNA encoding the protein vector comprising said affinity reagent; the DNA may be amplified and sequenced or sequenced directly, thereby revealing the protein sequence it encodes.

An alternative to presenting protein-DNA complexes (for example, where the affinity reagent is an antibody) in a spatial format as an array is to use them in solution to screen a spatial array of biomolecules, for example but without limitation, antibody-antigen interactions, wherein an addressable array of antigens is contacted with a sample of antibody-DNA complexes. The interaction can be detected as described above and the affinity reagent-DNA complex isolated and used as a cloning vehicle for the production of affinity reagent cognate for the detected target biomolecule. In this manner, a library of, for example but without limitation, antibody-DNA complexes in solution can be characterised and/or biomolecular targets recognised by said antibodies identified by virtue of the array being addressable. In another embodiment the array of target biomolecules may be non-addressable. Additionally, the antibody-DNA complex can be amplified *in situ* and, by virtue of being complexed with its encoding DNA, the DNA of the affinity reagent-DNA complex can be characterised by means known in the art such as electrophoresis, sequencing or hybridisation after optional amplification and release from the protein-DNA complex, thereby revealing the protein sequence it encodes. Preferably, said isolated antibody-DNA complex can be used as a cloning vehicle for said antibody by performing transcription and translation to produce a quantity of antibody.

It will be understood by one skilled in the art, that the above embodiments relating to affinity reagent-DNA complexes apply equally to a protein-DNA complex comprising any protein target. Thus, the self-assembled protein-DNA complexes of the invention can be used as cloning vehicles for the production in a solution, i.e. not attached to a substrate, of any recombinant peptide, polypeptide or protein of interest in its native form, complexed to the DNA encoding it i.e. its own DNA. The protein-DNA complexes of the invention can also be used to detect target protein/protein and target protein/nucleic acid interactions as well as interactions of a target protein with other compounds of interest if used in a spatial format as an array as described for affinity reagent-DNA complexes.

Where the protein-DNA complex of the invention is an antibody-DNA complex as described above, it is suitable for use in established immunochemical protocols, either replacing or in conjunction with ordinary polyclonal, monoclonal or recombinant antibodies (including single chain antibody Fab fragments), antibody-mimic molecules, CDRs and other protein affinity reagents. Such protocols include but are not limited to immunoblotting, immunoprecipitation, immunohistochemistry, immunocytochemistry, affinity chromatography, enzyme-linked immunosorbent assays (ELISA), immuno-electron microscopy etc. The protein-DNA complexes of the invention provide unique advantages for immunochemical applications as each complex has a unique nucleic acid sequence associated with the protein vector. Specific precipitation or affinity purification of a nucleic acid associated with its encoded and bound protein can be performed by virtue of the fact that each DNA molecule has a unique sequence. Each unique sequence can be distinguished and/or purified using nucleic acid hybridisation techniques (double helix formation by complementary nucleic acid strains or triple helix formation) or by restriction mapping, nucleic acid sequencing or other nucleic acid detection methods known in the art.

Additional amplification of signal can be achieved when using the self-assembled protein-DNA complexes of the invention for immunodetection applications such as ELISA, immunohistochemistry, immunocytochemistry, immunoblotting and similar techniques by virtue of the fact that nucleic acids can be amplified. Accordingly, the complexes of the invention can be amplified using standard polymerase chain reaction (PCR) techniques well known in the art or where the complex is a protein-RNA complex, using RNA polymerase to amplify the DNA component of the complex. Coupled transcription/translation is well known in the art and kits and reagents are available commercially. Examples of such systems are rabbit reticulocyte lysate, wheat germ lysate, SP6/T7 *in vitro* T&T and RTS 100 *E. Coli* HY transcription and translation kits (Roche Diagnostics Ltd., Lewes, UK) and the TNT Quick coupled Transcription/Translation System (Promega UK, Southampton, UK).

In the method of the invention *in vitro* transcription can be manipulated for the regulation of protein-DNA complex formation. For example, but without limitation, 3'-end modifications of DNA can pause or stall transcription, or the introduction of tertiary structure such as a triple helix can slow down transcription. The processivity of an *in vitro* translation reaction can be manipulated for example, but without limitation, by varying the concentration of tRNAs present in a reaction mixture or by omitting a translational stop codon. *In vitro* translation can be paused or stopped if the necessary tRNAs are not available. On addition

of the missing tRNAs to a reaction mixture protein synthesis will resume. This technique allows a user to manipulate the speed of protein synthesis and thus manipulate the speed of the folding of nascent protein chains. Additionally, it also permits the regulation of binding of the assembly sequence to its own DNA or suitable substrate, directly or indirectly.

Alternatively, it permits the regulation of binding of the target protein sequence to other cognate sequence, for example such as amino acid motifs involved in protein-protein interactions and those motifs involved in the formation of protein complexes. For example but without limitation, translation can be paused or slowed down after the protein vector assembly sequence is produced to allow it to bind to a nucleic acid or solid support or another protein, before the complete protein is translated and released from the ribosome. Translation can also be arrested by addition of a short complementary nucleic acid strand, a technique known in the art as the antisense approach (Blake, K, *et al.*, 1985, *Biochemistry* 24(22):6132-6138; Haeuptle, M., *et al.*, *Nucleic Acids Res.*, 1986 14(3):1427-1448; Nicole, L. and Tanguay, R. *Biosci. Rep.*, 1987, 7(3):239-246; Marcus-Sekura C. *et al.*, *Nucleic Acids Res.*, 1987, 15(14):5749-5763).

The advantages provided by the present invention over methods known in the art include one or more of the following:

- (i) the method allows the production of self-assembled protein-DNA complexes in solution or alternatively, immobilized self-assembled protein-DNA complexes (preferably in an array format) in a single enzymatic reaction. The method avoids the issue of RNA stability associated with the amount of RNA handling generally required in existing methods for the production of nucleic acid-protein complexes and hence also avoids the cross-linking problems associated with puromycin use, as described above.
- (ii) the binding of the protein portion of the complex to the DNA portion of same is reversible, permitting disassembly of the complex as necessary in downstream applications. Alternatively, cross-linking reagents can be used to form a covalently-linked complex.
- (iii) the complexes obtained can be used for the cloning of proteins, antibodies, antibody fragments, antibody mimics and peptide aptamers. Self-assembled protein-DNA complexes are especially useful in protein cloning as no additional reverse transcription step is required, unlike RNA-protein complexes.
- (iv) self-assembled protein-DNA complexes are especially advantageous over RNA-protein complexes or ARM complexes for the production of protein arrays as the absence of RNA and ribosomes permits more dense arrays. Also ARM arrays are subject to the inherent instability of RNA.
- (v) the sensitivity of immunodetection at a single protein molecule level can be achieved by directly labelling the DNA of the self-assembled protein-DNA complex where the protein component of said complex comprises an affinity reagent, for example an antibody.
- (vi) the present invention allows the production of self-assembled protein arrays using arrayed nucleic acids, a significant advantage over protein-spotting or in-situ protein synthesis arrays.
- (vii) while expression of recombinant proteins is often problematic with poor or absent expression or expression of inactive forms of the protein of interest due to mis-

folding, the protein-DNA complexes of the invention may be applied to the preparation of quantities of the encoded protein in a native, functionally correct form.

All publications, including, but not limited to, patents and patent applications, cited in this specification, are herein incorporated by reference as if each individual publication were specifically and individually indicated to be incorporated by reference herein as though fully set forth.

The invention will now be illustrated by reference to the following non-limiting examples, and corresponding figures, of the production of protein-DNA complexes. In the figures:

Figure 1 Panel A indicates the amino acid sequence of a fragment of Streptavidin comprising the side most loop (NTQWLLTSGTTEANAWKSTLVGHDT; SEQ ID No. 1). Panel B is a representation of the secondary structure of that fragment with its secondary structure shown. The side most loop links two antiparallel β -sheets. Stabilisation of the secondary structure can be achieved by substituting the amino acid pairs of the antiparallel sheet with cysteine residues. The seven pairs of amino acids indicated are especially suitable due to their proximity to the loop and molecular architecture; the distances ($\text{\AA}$) 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.

Figure 2 is the 466 base pair engineered nucleotide sequence (SEQ ID No. 2) used to make the Streptavidin-based protein vector (SAPV). Indicated are: a ribosome binding site (bold) immediately preceding the translational start codon and stop codons (underlined), the T7 terminator sequence (bold and underlined). The sequence (dotted underline) codes for the side most amino acid loop of Streptavidin selected for modification.

Figure 3 is the 1442 base pair engineered nucleotide sequence (SEQ ID No. 23) required to express streptavidin tagged with autofluorescent protein (AFP). Indicated are: two RNA polymerase binding sites (italic/bold), a ribosome binding site (bold) immediately preceding the translational start codon (bold/double underlined), stop codon (italic/underlined) and the linker region (underlined) separating the streptavidin and autofluorescent protein sequences.

Figure 4 is a histogram indicating the presence of untagged SAPV complexed with biotinylated DNA encoding tagged SAPV. PCR products were separated on an agarose gel as described in Example 3 and band intensities determined using GeneSnap and fluorescent imager from SynGene (Cambridge, UK). All values shown are normalised to the DNA sample from the 1st wash (which also contained a flow-through fraction of the total loaded DNA). Error bars represent standard deviation ($n=3$). Panel A shows assembly of SAPV with said biotinylated DNA: lanes (a) to (d) show the 1st to 4th washes, respectively; lane (e) shows the eluant containing SAPV-DNA complex eluted from a protein-binding filter. Panel B shows assembly of SAPV with non-biotinylated DNA encoding tagged SAPV: lanes (a) to (e) as for Panel A except that lane (e) indicates the absence of SAPV-DNA complex in the eluant.

Figure 5 is a histogram showing immunoprecipitation of SAPV-84. The filled bar indicates normalised signal corresponding to the precipitation of the SAPV-84 on immobilised anti-BCMP84 antibody. The open bar indicates relative signal strength corresponding to the eluate from anti-albumin antibodies. Error bar represents standard deviation (n=5). The scale is arbitrary.

Figure 6 is a histogram indicating the immunoprecipitation of self-assembled SAPV vectors. The captured samples from anti-albumin linked beads (Panel A) or anti-BCMP84 linked beads (Panel B) were assayed by PCR and equal amounts of the resultant products were run on a 2%(w/v) agarose gel containing ethidium bromide. 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. Error bars represent standard deviation (n=3). Lane (a), "empty" (unmodified) SAPV; lane (b), SAPV-alb5; lane (c), SAPV-84.

EXAMPLES

1. Generation of Streptavidin-based protein vector (SAPV)

The *Streptomyces avidinii* gene for streptavidin (Embl Accession No. X03591) was used as a scaffold for designing a SAPV. Full length nucleotide sequence coding for the SAPV (Fig. 2; SEQ ID No. 2) was produced using overlapping synthetic oligonucleotides (obtained from Sigma-Genosys, Cambridge, UK) and three rounds of PCR as follows:

PCR round 1, mixture of primers used:

Forward primer -5' aaattaatacgactcactatagggagaaggagatataccatggaggccggcatcaccggcacc tggtaaac 3' (SEQ ID No. 3);

Reverse primers - 5' ttccggtcagggcgccgtcgccgcccgcggtcacgatgaaggctgagccgagctggtgtac caggtgccggtgatg 3' (SEQ ID No. 4);

5' ggacgtagcggctctcgccgttgccgacggccgactcgttaggttccggtcagggcgccgctc 3' (SEQ ID No. 5);

5' gtggccggggcgctgtcgtgaacgaccgggtcaggacgtagcggctctcgg 3' (SEQ ID No. 6);

5' tcgaggagtggtcggttcggttagttatcttcaggccaccgtccaaccgagggcggtgccgtgccgtggtggccgggg cgctgt 3' (SEQ ID No. 7);

5' gccggaggtcagcagccactgggtgtgtatcctcgcccgccgacgtactggccgctccacgtggtcgcgagtg ggcgt 3' (SEQ ID No. 8);

5' ttccaccttggtgaagggtgctgtggccgaccagcgtggacttcaggcgttggcctcggtggtgccggaggtcagca 3' (SEQ ID No. 9); and

5' aaaaaacccctcaagaccggttagaggccccaaggggttatgctagtattcattcattcattccaccttggtg 3' (SEQ ID No. 10). Conditions: 20µl of a mixture of each primer at 1.25 pmol/µl in a 50µl total volume; 5' at 96°C followed by 20 cycles of [30 sec at 95°C, 30 sec at 72°C (with a 2°C decrement per cycle), 30 sec at 72°C (with 1 sec increment per cycle). Final incubation was for 1 min at 72°C.

PCR round 2, 5µl DNA obtained from PCR round 1 was used as a template mixture of primers used:

Forward primer - 5' aaattaatacgactcactatagggagaaggagatataccatggaggccggcatcaccggcacct ggtacaac 3' (SEQ ID No. 3).

The amplified SAPV DNA fragment (Q0225) was cloned into the multiple cloning site of TOPO-4-BLUNT vector. The sequence of the final construct is shown in Fig. 2. Expression of the SAPV protein vector was performed using a bacterial coupled transcription and translation kit obtained from Roche (RTS 100 *E.coli* HY). Transcription and translation reactions were performed according to manufacturer recommendations. Amounts of the SAPV DNA sufficient for transcription and translation were routinely obtained by 20 cycles of PCR amplification using round 3 conditions described above. Either biotinylated or non-biotinylated primers were used as stated.

(ii) a linear fragment of AFP was generated by PCR using pQBI25-fN1 vector DNA (QBiogene, Inc. CA) as a template. Conditions: 50µl total volume, 0.5µl of each of 100pmol/µl forward primer 5' tcacgcctgctcgccacgacccaattcgggaggcggaggtg 3' (SEQ ID No. 14) where the underlined nucleotides encode the amino acid sequence immediately C-terminal to the 8 amino acid loop identified in Fig. 1 and the remainder encode the 5' region of AFP within the pQBI25-fN1 vector, and reverse primer 5' ttattccaccttggtgcacgggggagggg

caaacaac 3' (SEQ ID No. 15); 7 min at 96°C, 30sec at 55°C, 1min at 72°C – followed by 15 cycles of [1 min at 96°C, 1.5 min at 72°C. Final incubation was 5 min at 96°C.

A transcription terminator sequence was added to the linear AFP fragment obtained above by re-amplification of the AFP fragment following by joining to untagged SAPV obtained as described in (i) above.

(iii) addition of transcription terminator sequence: 50µl total volume, 0.5µl 100pmol/µl forward primer 5' tccacgctggctcgccacgaccgaattcggggaggcggaggtg 3' (SEQ ID No. 14), 5µl 10pmol/µl reverse primer 5' aaaaaacccctcaagacccggttagaggccccaaggggttatgctagttatcattcat tcattccaccttggtg 3' (SEQ ID No. 10); 6 min at 96°C followed by 2 cycles of [1 min at 96°C, 30 sec at 40°C, 1 min at 72°C] then 15 cycles of [1 min at 96°C, 1.5 min at 72 °C]. Final incubation was 5 min at 72°C.

(iv) tagged SAPV was generated by joining the untagged SAPV fragment from (i) to the AFP with terminator sequence generated in (iii) as follows: 50µl total volume, 1µl SAPV fragment, 1µl AFP fragment with terminator sequence, 0.5µl each of 100pmol/µl forward primer 5' aaattaatacgaactcactat 3' (SEQ ID No. 11) and reverse primer 5' aaaaaacccctcaag accc 3' (SEQ ID No. 12); 5 min at 96 °C followed by 6 cycles of [30 sec at 72°C (with 5°C decrement per cycle), 1 min at 72°C (with 5 sec increment per cycle)] followed by 15 cycles of [1 min at 96°C, 30 sec at 40°C, 1.5 min at 72°C]. Final incubation was 5 min at 72°C. The amplified DNA was cloned into the TOPO4BLUNT vector.

A long 5'UTR was added to the AFP-tagged SAPV as follows:

(v) AFP-tagged SAPV DNA generated in part (iv) was used as a template to amplify a fragment of said DNA lacking the T7 polymerase binding site as following: 50µl total volume, 0.5µl 100pmol/µl forward primer 5' agaaggagatataccat 3' (SEQ ID No. 16) and 3µl 16pmol/µl reverse primer 5' attaacctcactaaaggga 3' (SEQ ID No. 17); 5 min at 96°C followed by [15 cycles of 1 min at 96°C, 30 sec at 40°C, 1.5 min at 72°C]. Final incubation was 5 min at 72°C.

(vi) A 5'UTR fragment was obtained from pIVEX vector (Roche Diagnostics Ltd., Lewes, UK) by firstly amplifying a fragment from said vector: 50µl total volume, 1µl of each of 100pmol/µl forward primer 5' aaattaatacgaactcactat 3' (SEQ ID No. 11) and reverse primer 5' aaaaaacccctcaagaccc 3' (SEQ ID No. 12); 5 min at 96 °C followed by 15 cycles of [1 min at 96°C, 30 sec at 40 °C, 1 min at 72°C]. Final incubation was 5 min at 72°C. Secondly the amplified fragment was cloned into TOPO-4-BLUNT vector and a long 5'UTR excised from said vector by PCR as follows: 50µl total volume, 0.5µl of 100pmol/µl forward primer 5' catggtatatctctct 3' (SEQ ID No. 18) and 3µl 16pmol/µl reverse primer 5' caggaaacagcta tgac 3' (SEQ ID No. 19); 5 min at 96°C followed by 15 cycles of [1 min at 96°C, 30 sec at 40°C, 30 sec at 72°C]. Final incubation was 5 min at 72°C.

(vii) The amplified fragment was joined with the tagged SAPV fragment generated in part (v) by PCR as follows: template DNA = a mixture of said tagged SAPV DNA and DNA containing the 5'UTR fragment generated in part (vi), 50 µl total volume, 0.5µl of each of 100pmol/µl forward primer 5' aaattaatacgaactcactat 3' (SEQ ID No. 11) and reverse primer 5' aaaaaacccctcaagaccc 3' (SEQ ID No. 12); 5 min at 96 °C followed by 10 cycles of [1 min at

96°C, 30 sec at 72°C (with a 3°C decrement per cycle)] followed by 10 cycles of [1 min at 96°C, 30 sec at 40 °C, 1.5 min at 72°C]. Final incubation was 5 min at 72°C.

The PCR product containing tagged SAPV DNA from part (vii) was purified by ethanol precipitation, prior to use in the transcription and translation reaction. A range of DNA concentrations and transcription and translation reaction temperatures were used in the generation of tagged Streptavidin. Reactions were typically run overnight. Detection of the tagged SAPV on Western blots was performed using anti-GFP Rabbit polyclonal antibody (AbCam Ltd, Cambridge, UK). Aliquots (5µl of 20µl) of each of the coupled transcription and translation reactions were loaded onto a precast 4-12% NuPAGE gel (Invitrogen). The proteins were resolved by SDS-polyacrylamide gel electrophoresis and transferred onto nitrocellulose membrane (0.2µM pore size, Invitrogen) using an Xcell SureLock Minicell and Blot Module according to the manufacturer's instructions. The membrane was blocked for 1 hr. in Tris buffered saline plus 0.1%(v/v) Tween-20 (TBST) with 2%(w/v) powdered milk and probed with a 1:3000 dilution of AbCam rabbit anti-GFP polyclonal (1hr. room temperature; RT) in TBST/milk. The membrane was washed in TBST (3 times for 5-10 min. each wash) and probed with 1:6000 HRP-labelled anti-rabbit IgG, (Amersham Pharmacia) in TBST/milk (1 hr. RT), washed as above and developed using enhanced chemiluminescence (ECL; Amersham Biosciences AB, Uppsala, Sweden) according to the manufacturer's instructions and exposed to ECL Hyperfilm (Amersham Biosciences AB, Uppsala, Sweden).

In subsequent examples, the optimal experimental conditions for coupled transcription and translation reactions selected from the results of the above experiment use 2µg DNA per 20µl reaction volume with the synthesis temperature maintained at 21°C.

3. Assembly of a SAPV protein-DNA complex

The following experiment shows one example of assembling a complex of untagged SAPV with its encoding DNA and is in no way intended to be limiting.

Untagged SAPV DNA lacking stop codons was obtained by PCR using tagged SAPV DNA generated as in part (vii) as a template as follows: 50µl volume, 3µl of 16pmol/µl forward primer 5' gtaaacgacgcccag 3' (SEQ ID No. 13) and 0.5µl of 100pmol/µl reverse primer 5' ttccaccttggtgaaggtgtcgtggccgaccagcgtggactccaggcgttgccctcggtgtgcccggaggatca gca 3' (SEQ ID No. 9); 6 min at 96°C followed by 15 cycles of [1 min at 96°C, 30 sec at 40°C, 1 min at 72 °C]. Final incubation was 5 min at 72°C.

Using this DNA untagged SAPV protein was prepared by coupled transcription and translation (see Example 2 last paragraph for details). After overnight incubation the coupled transcription and translation reaction was centrifuged for 3min at 15000rpm in a microfuge. Supernatant was transferred to a fresh tube.

Biotinylated and non-biotinylated SAPV DNA (sequence as in Figure 3) was generated by PCR using either 5'-biotinylated or non-biotinylated primers from Sigma-Genosys (Cambridge, UK): forward primer: 5' aaattaatacgactcactat 3' (SEQ ID No. 11); reverse primer: 5' aaaaaaccctcaagaccc 3' (SEQ ID No. 12). Conditions: 1µl of each primer at 100 pmol/µl in a 50ul total volume; 5' at 96°C followed by 25 cycles of [1 min at 96°C, 30 sec at 40°C, 1 min at 72°C]. Final incubation was 5' at 72°C.

The PCR product containing SAPV DNA was purified by ethanol precipitation. Assembly reactions containing 5µg DNA plus 10µl of the coupled transcription and translation reaction supernatant were incubated overnight at 4°C. Complexes were separated from uncomplexed DNA by filtration through protein-binding microfuge filters (Ultrafree-MC Probind Units modified PVDF, Millipore (UK) Ltd.). After 4 washes (by flow through filtration) retained materials were eluted by incubation with 50µl of 0.1xTAE for 30min with gentle agitation. Eluted DNAs were detected by PCR as follows: 10µl of each of the wash through and eluates from each assembly reaction was amplified in parallel using forward primer 5' aaattaatacagactcactat 3' (SEQ ID No. 11); reverse primer: 5' aaaaaacccctcaagaccc 3' (SEQ ID No. 12); 35 cycles of [1 min at 96°C, 30 sec at 40°C, 1.5 min at 72°C] were performed. Amplified products were separated on 2.5%(w/v) agarose gels containing EtBr. Equal amounts of PCR products were loaded onto each lane (see Figure 4). Absence of signal in the 4th wash (in both biotinylated and non-biotinylated DNAs assemblies) confirms absence of a non-specific background. The eluate from biotinylated DNA experiment (Figure 4A) contains large amount of the DNA, whilst the eluate from the non-biotinylated DNA assembly (Figure 4B) does not. This demonstrates that the designed SA-based tagged protein vector SAPV binds to its encoding biotinylated DNA.

4. Generation of SAPV comprising a target sequence

A display system using SAPV is described. A loop in the amino acid sequence of Streptavidin has been selected (Figure 1) as a useful site for insertion of a target protein sequence or for the insertion of other protein fragments, peptides or tags. The core protein sequence of the Streptavidin and the Streptavidin-based SAPV contains a 9 amino acid long loop (Figure 1) which was found to be mostly suitable for modifications, such as SAPV extension, modification or for expressing of other proteins, peptides and tags. This choice is based on the Streptavidin molecular architecture (Figure 1). To limit tertiary structure changes, the secondary structure of SAPV can be stabilised and positioned by one or more disulphide bonds (see Figure 1 for residues identified for mutation to cysteines; these have the distance between the identified pairs shown in Angstroms). SAPV nucleic acid sequence was modified to display peptide fragments of albumin and BCMP84 (see PCT Application No. WO 01/62914) proteins as follows:

(i) untagged SAPV DNA lacking STOP codons (as described in Example 3) was used as a template in the following PCR reactions to produce SAPV displaying either a BCMP 84 peptide (SAPV-84) or a BSA peptide (SAPV-Alb5): 50µl total volume, 1µl of 100pmol/ul forward primer (SEQ ID No. 13) and either 5µl of 10pmol/ul reverse primer 5' tcgtggcc gaccagcgtggaaactaaatctttaattcagaaggagttaaagtttcttaccacctcgccggaggtcagcagccactgg 3' (to make SAPV-84 construct; SEQ ID No. 20) or 5µl of 10pmol/ul reverse primer 5' tcgtgg ccgaccagcgtggatttagcaaaaaataataattcaggagcataaaaaataaggatggccggaggtcagcagccactgg 3' (to make SAPV-alb5 construct; SEQ ID No. 21); 5 min at 96°C followed by 30 cycles of [30 sec at 96°C, 30 sec at 54°C for 30", 30 sec at 72°C]. Final incubation was 5min at 72°C.

(ii) STOP codons were added to both constructs as follows: SAPV-84 or SAPV-alb5 DNA fragments (as described above) were used as a template in PCR amplifications.

Conditions: 50 µl total volume, 1 µl of 100pmol/µl forward primer 5' gtaaaacgacggccag 3' (SEQ ID No. 13) and 5 µl of the 10pmol/µl reverse primer 5' aaaaaacccctcaagaccggttagag gccccaaaggggtatgctagttatcattcattcattccaccttggtgaagggtgctggccgaccagcgtgga 3' (SEQ ID No. 22); 5min at 96°C followed by 30 cycles of [30 sec at 96°C, 30 sec at 54°C, 30 sec at 72°C]. Final incubation was 5min at 72°C.

A quantity of DNAs encoding each of SAPV-84 and SAPV-alb5 (both biotinylated and non-biotinylated using biotinylated or non-biotinylated primers) were produced for transcription and translation by PCR: forward primer 5' aaattaatacgaactcactat 3' (SEQ ID No. 11); reverse primer: 5' aaaaaacccctcaagaccc 3' (SEQ ID No. 12). Conditions: 1 µl of each primer at 100 pmol/µl in a 50ul total volume; 5' at 96°C followed by 25 cycles of [1 min at 96°C, 30 sec at 40°C, 1 min at 72°C]. Final incubation was 5' at 72°C.

5. Affinity purification of a SAPV modified to display BCMP84 peptide (SAPV-84) protein-DNA complex

A coupled transcription and translation reaction (50µl volume) was run as described above (see Examples 2 and 3 for details) using DNA coding for untagged SAPV-84. Assembly was achieved by overnight incubation at 4°C of biotinylated SAPV-84 DNA construct with the coupled transcription and translation reaction supernatant as before. The SAPV-DNA complex was immunoprecipitated with either anti-BCMP84 or anti-albumin antibodies as follows: 3-mm squares of wetted nitrocellulose membrane were incubated with either 20µl of anti-albumin antibody or 20µl of anti-BCMP84 antibody. The squares were washed twice in 1ml PBS then blocked by incubation in 2% (w/v) powdered milk plus 0.1 mg/ml tRNA in PBS for 30 min. Blocking buffer was removed and the nitrocellulose membrane squares were incubated for 60min. with the assembled SAPV-DNA complex. The squares were washed three times in wash buffer (1 ml PBS supplemented with 0.01 mg/ml tRNA and 2%(w/v) powdered milk). After a final wash in 50 µl wash buffer, the nitrocellulose membrane squares were incubated with 50 µl elution buffer (80mM glycine, pH 2.45) supplemented with 0.4%(w/v) powdered milk and 0.01µg/ml tRNA. The eluted samples were neutralised by the addition of 11µl 2M NaOH. The presence of the SAPV-84 in the eluted samples was detected by PCR amplification of the SAPV-84 DNA expression vector as follows: forward primer 5' aaattaatacgaactcactat 3' (SEQ ID No. 11) and reverse primer 5' aaaaaacccctcaagaccc 3' (SEQ ID No. 12). All reactions were run in parallel under identical conditions: 33 cycles of [30 sec at 96°C, 30 sec at 42°C, 30sec at 72°C] were performed. PCR products were run on a 2%(w/v) agarose gel and band intensities determined using GeneSnap and a fluorescent imager (SynGene, Cambridge, UK). The results of 5 measurements are presented in Figure 5 and indicate that immunoprecipitation of SAPV-84 using anti-BCMP84 is significantly higher than the control (anti-albumin antibody).

6. Self-assembly of SAPV-84 with its encoding DNA

The following example demonstrates self-assembly of SAPV-84 with its encoding DNA and of SAPV with an albumin peptide as a target sequence (SAPV-alb5) inserted into

the side most loop with its encoding DNA. An "empty" SAPV was included as a control. Coupled transcription and translation (using RTS 100 *E.coli* HY kit from Roche) and self-assembly of the SAPV-84, SAPV-Alb5 and an "empty" SAPV-only protein vectors were performed as before. Optimal reaction conditions were as used in Example 2, except that each reaction contained approximately 16µg of biotinylated SAPV DNA expression vectors (either SAPV-84, SAPV-alb5 or an "empty" SAPV-only DNA expression vector) in a total volume of 100µl each. Two 40µl aliquots of cleared supernatant from each self-assembly reaction were transferred to fresh tubes. The SAPV-84 and SAPV-alb5 were precipitated from these aliquots by incubation with either anti-BCMP84 or anti-albumin antibodies as follows:

(i) preparation of antibody linked glass beads: protein A-conjugated glass beads (120µl; PROSEP-A, Millipore (UK) Ltd) were prepared by washing 3 times in 1ml PBS followed by incubation with either 600µl of anti-albumin antibody (rabbit polyclonal, 2.8mg/ml) supplemented with 0.1mg/ml tRNA or 600µl of anti-BCMP84 antibody (rabbit polyclonal, 110 µg/ml) supplemented with 0.1 mg/ml tRNA for 90 min. All subsequent washes, incubations and elutions were carried out in the presence of 0.01mg/ml tRNA. Both sets of beads were washed 3 times and each split into separate tubes containing 20µl of beads.

(ii) Immunoprecipitation of SAPVs: each aliquot of beads (prepared as described above) was incubated with 40µl of supernatant from one of the self-assembled reactions for 90min. Each sample was transferred to a macroporous Wizard® Minicolumn filtration unit (Promega UK, Southampton, UK). The column with beads was washed with 80µl of PBS by centrifugation. Each column was then washed with 40ml of PBS delivered using a 50ml syringe. The filtration unit was transferred into a microcentrifuge tube and the beads were again washed with 50µl PBS by centrifugation. The beads were finally resuspended in 50µl elution buffer (100mM glycine, pH 2.45) and the eluates collected by centrifugation followed by neutralisation with 13µl of 2M NaOH.

(iii) Determination of presence of self-assembled complexes: Complexes were assayed by PCR following ethanol precipitation; 50µl eluate plus 80µl water, 20µl 3M NaOAc, 1µl of 10µg/ml carrier tRNA and 2.5 volumes of ethanol was incubated at -20°C overnight. Samples were centrifuged for 15min. at 15,000rpm in a microcentrifuge and following a wash with 70%(v/v) ethanol, were redissolved in 100µl deionised water. Solubilised precipitates (15µl) were assayed by PCR as described in Example 5. Equivalent amounts of the PCR reactions were run on a 2%(w/v) agarose gel containing EtBr (see Fig. 6). The figure clearly demonstrates that SAPV-84 is precipitated only on anti-BCMP84 beads and SAPV-alb5 is precipitated only on anti-albumin beads indicating that immunoprecipitated SAPV proteins are associated with their encoding DNA as a self-assembled protein-DNA complex.

Claims

1. A method for the production of a protein-nucleic acid complex comprising:
(a) providing a nucleic acid comprising code for a protein vector, said protein vector comprising an assembly sequence and a target protein sequence; and
(b) performing transcription and translation, wherein upon expression of the protein vector, the encoded assembly sequence binds to its encoding nucleic acid at a cognate sequence present within, or attached to said nucleic acid.
2. The method of claim 1, wherein the protein-nucleic acid complex is a protein-DNA complex.
3. The method according to claim 1 or 2, wherein the assembly sequence binds directly to its encoding nucleic acid.
4. The method according to claim 1 or 2, wherein the assembly sequence binds indirectly to its encoding nucleic acid by binding to a cognate sequence attached to the nucleic acid.
5. The method according to any one of the preceding claims, wherein binding of the assembly sequence to its encoding nucleic acid occurs whilst the protein vector is still attached to a ribosome.
6. The method according to any one of the preceding claims, wherein the assembly sequence is present at, or near, the N-terminus of the target protein sequence.
7. The method according to any one of claims 1 to 5, wherein the target protein sequence is displayed within the assembly sequence.
8. The method according to any one of the preceding claims, wherein the protein vector additionally comprises a detectable protein sequence.
9. The method according to claim 8, wherein the detectable protein sequence is present at, or near, the C-terminus of the protein vector.
10. The method according to any one of the preceding claims, wherein the protein vector comprises at least one affinity tag.
11. The method according to claim 10, wherein the affinity tag is adapted to immobilise the protein-nucleic acid complex to a substrate.

12. The method according to claim 10 or 11, wherein the assembly sequence and the affinity tag are situated at, or near, opposite termini of the target protein sequence.
13. The method according to any one of the preceding claims, wherein the target protein sequence is adapted for release from the assembly sequence by means of a cleavable linker.
14. The method according to any one of the preceding claims, wherein the target protein is a monoclonal, divalent or polyvalent antibody, an antibody fragment or antibody mimic, a complementarity determining region or other affinity reagent.
15. The method according to any one of the preceding claims wherein the nucleic acid is labelled with a fluorescent label.
16. A method for detecting the interaction of a target protein with a compound of interest comprising:
 - (a) providing a protein-nucleic acid complex according to the method of any one of the preceding claims;
 - (b) contacting the protein-nucleic acid complex with a compound of interest; and
 - (c) detecting the interaction of the target protein with a compound of interest.
17. The method according to claim 17 wherein the detection step comprises analysing the nucleic acid component of the protein-nucleic complex following its optional amplification and release from said complex.
18. The method according to claim 17 or 18 for detecting an antibody-antigen interaction.
19. A protein-nucleic acid complex produced according to the method of any one of claims 1 to 15.
20. A spatial array of protein-nucleic acid complexes produced according to the method of any one of claims 1-15 immobilised to a support.

A
NTQWLLTSGTTEANAWKSTLVGHDT

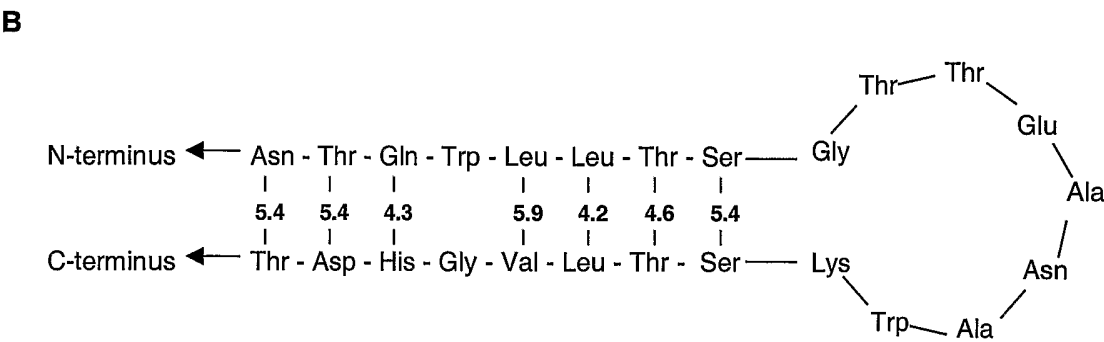


Fig. 1

AAATTAATAC GACTCACTAT AGGG**AGAAGG 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
GCCGAGGCCA GGATCAACAC CCAGTGGCTG CTGACCTCCG GCACCACCGA GGCCAACGCC
TGGAAGTCCA CGCTGGTCGG CCACGACACC TTCACCAAGG TGGAATGAAT GAATGATAAC
TAGCATAACC CTTGGGGCC TCTAAACGGG TCTTGAGGGG TTTTTT

Fig. 2

GTAAAACGAC GGCCAGTGAA TTG**TAATACG** **ACTCACTATA** **GGGCGAATTG** AATTTAGCGG
 CCGCGAATTC GCCCTTAAAT **TAATACGACT** **CACTATAGGG** AGACCACAAC GGTTCCTC
 TAGAAATAAT TTTGTTTAAC TTTA**AGAAGG** **AGATATACCA** TGGAGGCCGG CATCACCGGC
 ACCTGGTACA ACCAGCTCGG CTCGACCTTC ATCGTGACCG CGGGCGCCGA CGGCGCCCTG
 ACCGGAACCT ACGAGTCGGC CGTCGGCAAC GCCGAGAGCC GCTACGTCCT GACCGGTCTG
 TACGACAGCG CCCCCGCCAC CGACGGCAGC GGCACCGCCC TCGGTTGGAC GGTGGCCTGG
 AAGAATAACT ACCGCAACGC CCACTCCGCG ACCACGTGGA GCGGCCAGTA CGTCGGCGGC
 GCCGAGGCCA GGATCAACAC CCAGTGGCTG CTGACCTCCG GCACCACCGA GGCCAACGCC
 TGGAAGTCCA CGCTGGTCGG CCACGACCGA ATTTCGGGGAG 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
 TTCAAGAGTG CCATGCCCCA 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 ATAGTGTCAC CTAAATGCTA GAGCTCGCTG ATCAGCCTCG
 ACTGTGCCTT CTAGTTGCCA GCCATCTGTT GTTTGCCCTT CCCCCGTGCA CCAAGGTGGA
 ATGAATGAAT GATAACTAGC ATAACCCCTT GGGGCCTCTA AACGGGTCTT GAGGGGTTTT
 TT

Fig. 3

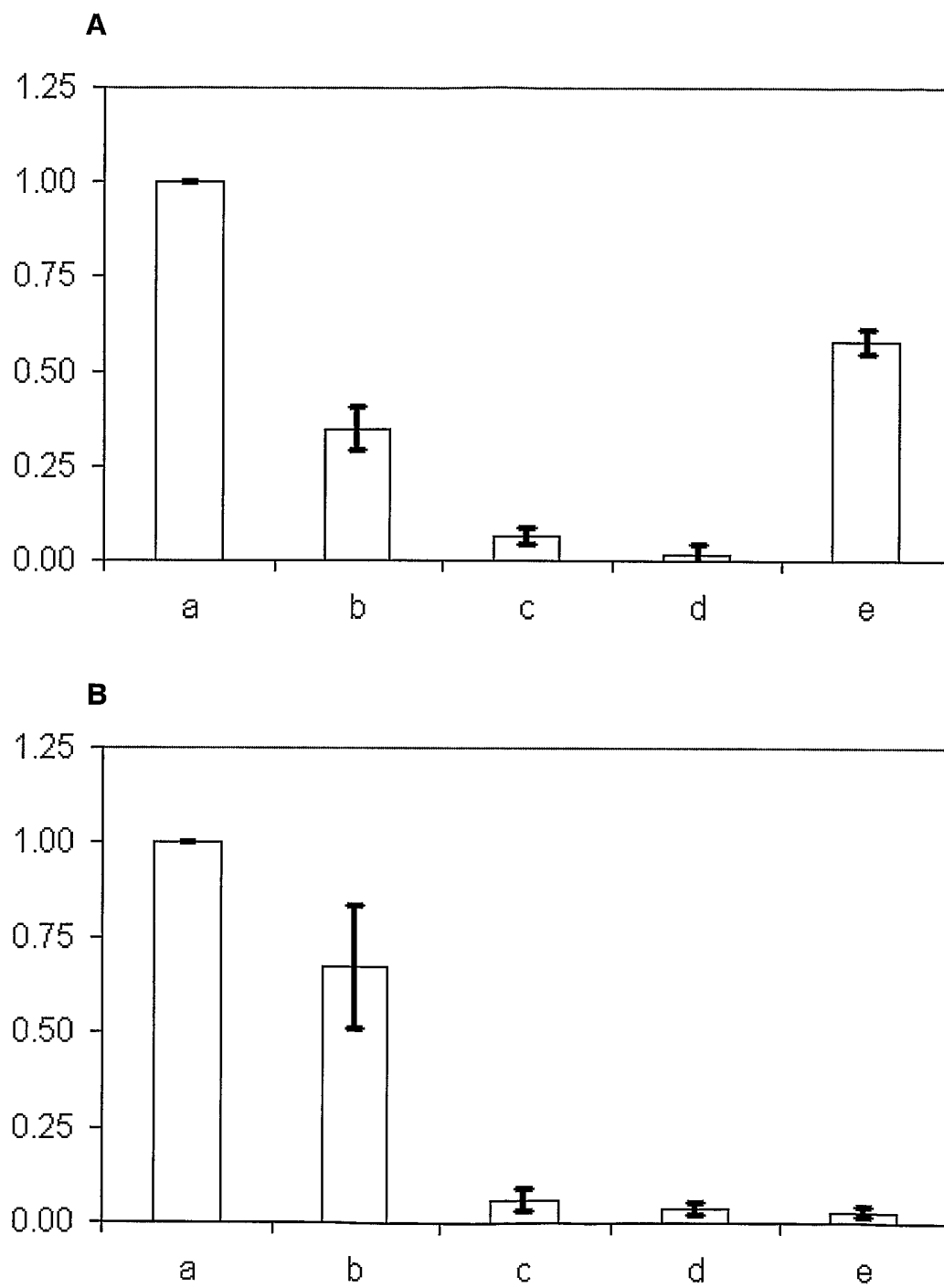


Fig. 4

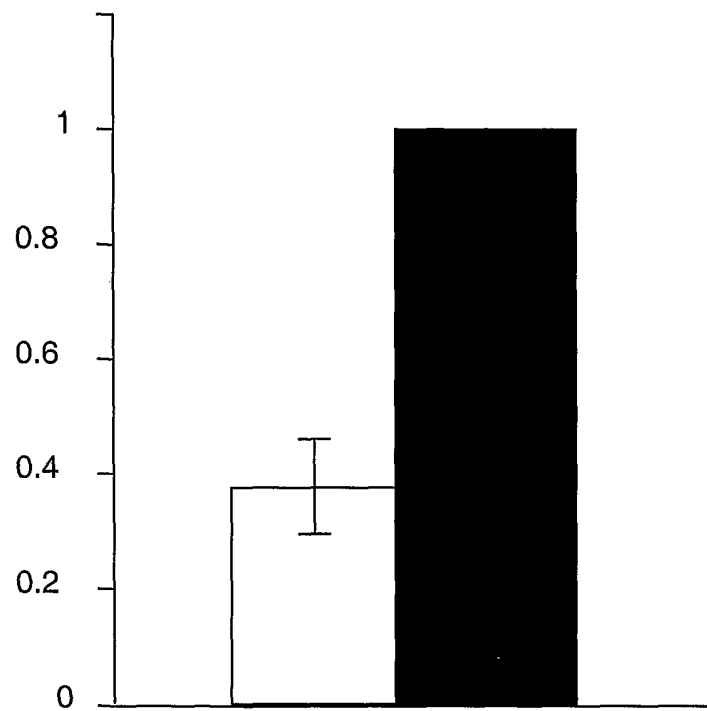


Fig. 5

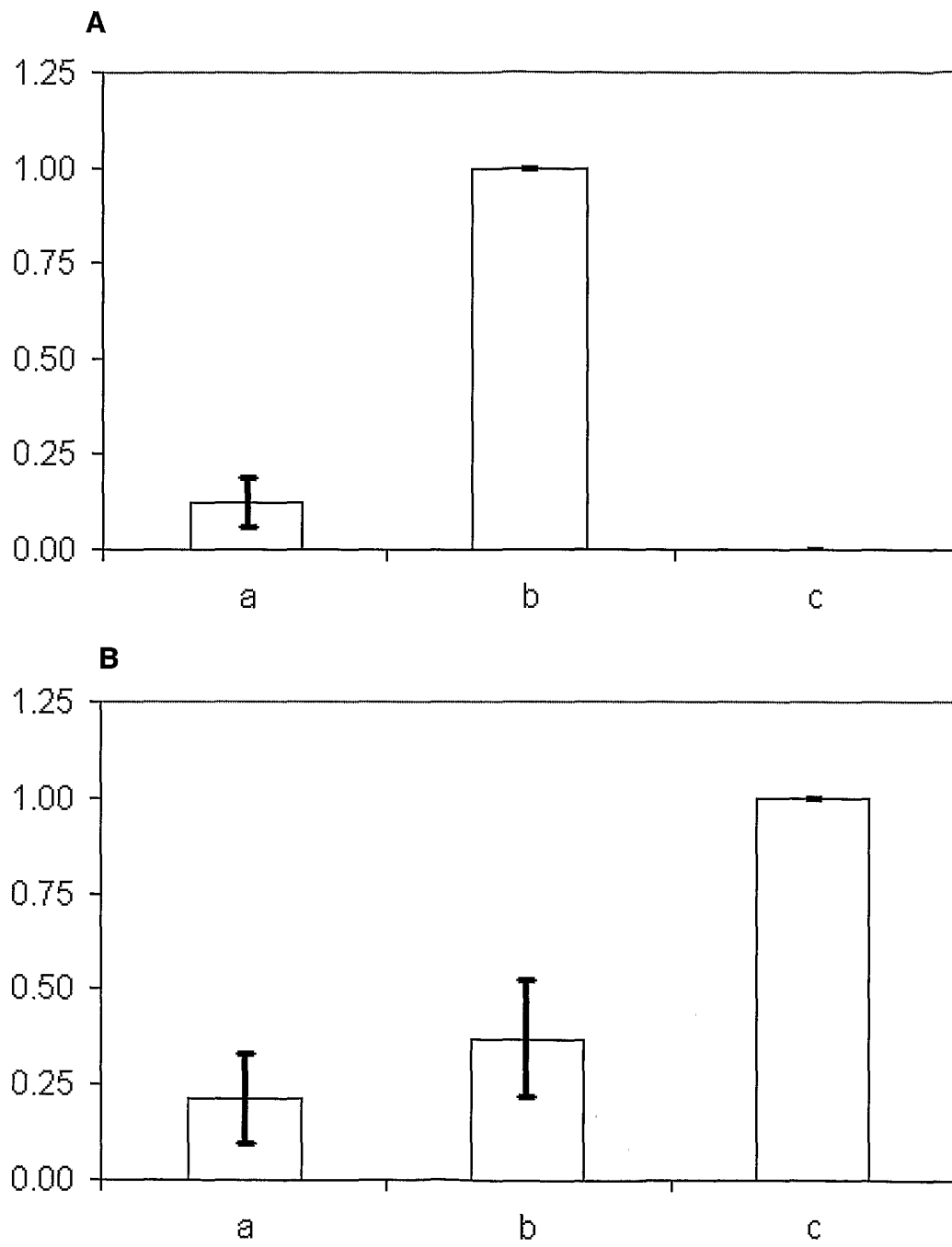


Fig. 6