(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
26 January 2017 (26.01.2017)

WIPO | PCT

(10) International Publication Number
WO 2017/013400 A1

(51) International Patent Classification:
C07K 14/00 (2006.01) *G01N 33/533* (2006.01)
G01N 33/53 (2006.01)

(21) International Application Number:
PCT/GB2016/052124

(22) International Filing Date:
14 July 2016 (14.07.2016)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
1512617.0 17 July 2015 (17.07.2015) GB

(71) Applicant: **THE FRANCIS CRICK INSTITUTE LIMITED** [GB/GB]; Gibbs Building, 215 Euston Road, London NW1 2BE (GB).

(72) Inventors: **WEBB, Martin Ronald**; The Francis Crick Institute, Mill Hill Laboratory, The Ridgway, Mill Hill, London NW7 1AA (GB). **CHISTY, Liisa Tellervo**; The Francis Crick Institute, Mill Hill Laboratory, The Ridgway, Mill Hill, London NW7 1AA (GB).

(74) Agents: **SCRIPT IP LIMITED** et al.; 18 Bridge Street, Frome, Somerset BA11 1BB (GB).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

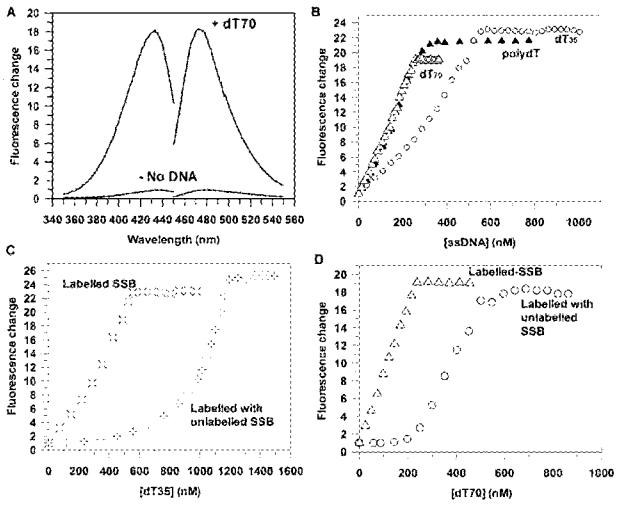
— *with international search report (Art. 21(3))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

WO 2017/013400 A1

(54) Title: SINGLE-STRANDED DNA SENSOR

FIGURE 1



(57) Abstract: The invention relates to a protein of the single stranded DNA binding domain (SSB) family, wherein said protein is a modified *Plasmodium* protein, said protein comprising at least one detectable label attached to an amino acid of said protein, wherein said amino acid is located on the (L1-1') loop of residues joining the two beta sheets (β1) and (β1'), wherein the characteristics of the detectable label alter on binding single stranded DNA. The invention also relates to nucleic acids, uses and methods.

SINGLE-STRANDED DNA SENSOR

FIELD OF THE INVENTION

The invention relates to sensor proteins for detection of single stranded DNA, for example in biological systems or solutions. Suitably the sensor proteins are based on *Plasmodium spp.* single-stranded DNA binding protein (SSB).

BACKGROUND TO THE INVENTION

Single-stranded DNA is a product of many cellular processes that include processing of dsDNA such as DNA replication, transcription, translation and repair. Single-stranded DNA may bind to multiple targets and form secondary structures, which may lead to undesirable events such as DNA mutation if left free in the cell. In nature, a ubiquitous protein has evolved to protect the single-stranded DNA, namely single-stranded DNA binding protein (SSB) and its structure and function is highly conserved between different bacterial species.

Methods for assaying important biomolecules play an important role in quantifying them to understanding biological processes, but also in drug discovery. Fluorescent reagentless biosensors are single species that produce a fluorescent signal and can be used *in situ* to measure biomolecules[1]. One way to develop reagentless biosensors is by using a protein scaffold to bind the molecule of interest specifically but also to be an adduct with the reporter fluorophore. Strengths of such reagentless biosensors are their simplicity, sensitivity and rapid response. In addition, there is only one process involved in the fluorescent reporting, namely interaction between analyte and biosensor. This limits the difficulties that can be found in multicomponent assays, where several stages may be involved (e.g. in coupled enzyme assays): each component can be affected by the assay additions and conditions and each component adds to costs.

An existing reagentless biosensor for single-stranded DNA is an adduct of the tetrameric single-stranded DNA-binding protein (SSB) of *Escherichia coli* and a diethylaminocoumarin fluorophore[2]. This gives approximately a 6-fold increase in fluorescence when fully bound by ssDNA, which could be used to measure ssDNA. For example, it has been used to assay in real time the ssDNA product of helicase unwinding of double-stranded DNA (dsDNA)[3, 4]. DNA unwinding by helicases occurs during most DNA manipulations within the cell, including replication, but also repair

and recombination. Thus such an assay for ssDNA could be used to monitor aspects of DNA repair or stalled replication by polymerases. These are important targets for intervention in cell genetic machinery. Mutations in several helicases are directly implicated in human diseases (e.g. Bloom's syndrome).

WO2008/152379 describes labelled single-strand DNA-binding protein (SSB) and its use in a biosensor for detection and visualisation of single-stranded DNA. This document describes bacterial SSBs, in particular nine different bacterial SSBs in an alignment presented in Figure 8 of WO2008/152379.

However, the multiple binding modes of ssDNA to this known *E coli* protein[5, 6] have proved problematic for simple application of this probe, due to the binding mode of the DNA to SSB being dependent on experimental conditions.

Prior art sensor proteins such as labelled *E.coli* SSB proteins show limited fluorescence response size. It is a problem to provide sensor proteins with enhanced or increased fluorescence response size.

Prior art sensor proteins such as labelled *E.coli* SSB proteins show non-linear responses, especially in low salt conditions. It is a problem to provide sensor proteins with linear responses, especially in low salt conditions.

Prior art sensor proteins such as labelled *E.coli* SSB proteins show multiple binding modes, and this results in slow fluorescence decreases giving lower fluorescence as it changes between binding modes. It has not been possible in the art to provide sensor proteins addressing these problems.

Prior art sensor proteins such as labelled *E.coli* SSB proteins suffer from limited signal and/or sensitivity. It is a problem to provide sensor proteins with enhanced or increased signal and/or sensitivity.

Prior art sensor proteins such as labelled *E.coli* SSB proteins have the drawback of decreasing fluorescence in real time assays. It is a problem to provide sensor proteins which do not suffer from this drawback.

A *Plasmodium* ssDNA binding protein is known. The *Plasmodium* ssDNA binding protein is very different to the *E.coli* ssDNA binding protein, including in terms of the amino acid sequence.

Prior art studies of the *Plasmodium* ssDNA binding protein have not attempted any labelling. Prior art studies have been carried out using intrinsic tryptophan residues only. These have been purely abstract studies not directed at the creation of any biosensor molecule.

The present invention seeks to overcome problem(s) associated with the prior art.

SUMMARY OF THE INVENTION

Plasmodium SSB, especially the exemplary *Plasmodium falciparum* SSB, a tetramer, has been specifically labelled on the surface with diethylammoniumcoumarin fluorophore to produce a biosensor that reports of concentration of single-stranded DNA. In contrast to the prior art *E.coli* sensors, the sensors of the invention have specific valuable properties which are absent from the art. For example, there is 20-fold fluorescence on DNA binding with PfSSB labelled at cysteine 93. Binding is rapid and may be diffusion controlled and 65-70 bases of DNA bind to each tetramer. The sensor protein of the invention has only small changes in binding on changing salt conditions, which appears to be evidence that the 65-70–base binding mode predominates under a wide range of conditions. Thus the invention delivers a range of significant technical benefits over the known *E.coli* sensors.

In one aspect, the invention relates to a protein of the single stranded DNA binding domain (SSB) family, wherein said protein is a modified *Plasmodium* protein, said protein comprising at least one detectable label attached to an amino acid of said protein, wherein said amino acid is located on the (L1-1') loop of residues joining the two beta sheets (β1) and (β1'), wherein the characteristics of the detectable label alter on binding single stranded DNA.

Suitably the detectable label is attached to a region of the protein surface.

Suitably the detectable label is attached via a cysteine residue in the protein.

Suitably the cysteine residue is a naturally occurring cysteine residue. Suitably the cysteine residue is, or corresponds to, C93 (Cys 93) of SEQ ID NO: 1.

Suitably said cysteine residue is a cysteine residue engineered into the protein at a position which is, or corresponds to, one selected from the group consisting of: G92, E94 and K151 of SEQ ID NO: 1.

Suitably said modified *Plasmodium* protein is a modified *Plasmodium falciparum* protein.

Suitably said protein comprises amino acid sequence corresponding to at least amino acids 78 to 198 of SEQ ID NO:1,
said amino acid sequence having at least 50% sequence identity to amino acids 78 to 198 of SEQ ID NO:1.

Suitably said protein comprises amino acid sequence corresponding to at least amino acids 78 to 250 of SEQ ID NO:1,
said amino acid sequence having at least 50% sequence identity to amino acids 78 to 250 of SEQ ID NO:1.

Suitably said protein comprises amino acid sequence corresponding to at least amino acids 78 to 277 of SEQ ID NO:1,
said amino acid sequence having at least 50% sequence identity to amino acids 78 to 277 of SEQ ID NO:1.

Suitably said protein comprises amino acid sequence corresponding to at least amino acids 78 to 285 of SEQ ID NO:1,
said amino acid sequence having at least 50% sequence identity to amino acids 78 to 285 of SEQ ID NO:1.

In one embodiment the polypeptide part of said protein consists of the amino acid sequence of SEQ ID NO:1.

Suitably the detectable label is a fluorescent label.

Suitably the detectable label is a coumarin.

In one aspect, the invention relates to a protein as described above wherein the label is selected from the group consisting of *N*-[2-(1-maleimidyl)ethyl]-7-diethylaminocoumarin-3-carboxamide and *N*-[2-(iodoacetamido)ethyl]-7-diethylaminocoumarin-3-carboxamide (IDCC).

In one aspect, the invention relates to a protein as described above wherein the label is IDCC.

In one aspect, the invention relates to a protein as described above which further comprises a mutation compared to SEQ ID NO: 1 at a position selected from the group consisting of K84, Y156, K187 and D189, preferably Y156. Suitably said mutation is selected from the group consisting of K84D, Y156R, K187D and D189K, preferably Y156R.

In one aspect, the invention relates to a method for detecting single stranded DNA in a sample comprising the steps of:
(i) mixing the sample with the protein as described above and
(ii) detecting a change in the mixture arising from the interaction between the single stranded DNA and the protein.

In one aspect, the invention relates to a method for monitoring changes in ssDNA concentration in a sample comprising contacting said sample with a protein as described above and determining changes in the characteristics of the detectable label, wherein changes in the characteristics of the detectable label indicate changes in the concentration of ssDNA in said sample.
Suitably the characteristics of the detectable label are monitored by measurement of changes in fluorescence of a fluorophore comprised by said protein.

In one aspect, the invention relates to a method of screening for inhibitors of DNA processing enzymes which
comprises assaying single stranded DNA levels *in vitro* using a protein as described above in the presence and absence of the inhibitors and assaying
for an alteration in the single stranded DNA levels.

In one aspect, the invention relates to use of a protein as described above in the determination of ssDNA concentration in a sample.

In one aspect, the invention relates to a nucleic acid having a nucleotide sequence encoding the polypeptide portion of the protein as described above.

LOCATION OF LABEL

Suitably proteins of the invention such as PfSSB are labelled on a short loop of residues joining the two beta sheets (β1 and β1').

Suitably proteins of the invention such as PfSSB are labelled on a short loop of residues (L1-1'), joining two beta sheets (β1 and β1').

Suitably proteins of the invention such as PfSSB are labelled on a short loop comprising residues 91 – 96 (L1-1'), joining two beta sheets (β1 and β1').

Suitably proteins of the invention such as PfSSB are labelled on a short loop comprising residues 91 – 96 (L1-1'), joining two beta sheets (β1 and β1'), most suitably on amino acid residue 92, 93 or 94.

Suitably proteins of the invention such as PfSSB are labelled on a short loop comprising residues 91 – 96 (L1-1'), joining two beta sheets (β1 and β1'), most suitably on amino acid residue C93.

Residues for labelling are suitably wild type cysteine or are suitably mutated to cysteine as necessary. For example suitably when labelling position 92 of protein of the invention such as PfSSB, a G92C mutation is made; suitably when labelling position 94 an E94C mutation is made.

Suitably a protein of the invention has only one cysteine; suitably when sites other than that corresponding to C93 of PfSSB are labelled, C93 is mutated to an amino acid other than cysteine, such as C93A. The same is true for any cysteines present at positions other than C93 - suitably any such cysteine(s) are mutated to residues other than cysteine, such as mutated to alanine, so that the proteins of the invention suitably comprise only one cysteine.

Labelling at the corresponding three-dimensional site may be carried out such as at position 151 of protein of the invention such as PfSSB. Suitably when labelling position 151 a K151C mutation is made.

Amino acids adjacent to C93 may be labelled. In this context, 'adjacent' means adjacent in three dimensional space. This may mean near neighbour residues such as position 92 or position 94 which are both near neighbours of C93 i.e. they are the amino acid before or the amino acid after the one specified.

Amino acids close in three-dimensional space to those exemplified for labelling might also be used, such as amino acids close in three-dimensional space to C93.
Amino acids close in three-dimensional space to C93 include:
G92
E94
K151
Suitably such amino acids are substituted for Cysteine and the resulting Cysteine is then labelled.

PfSSB C93 neighbouring in space locations can be expected to give similar signal if mutated to cysteine and labelled with an environmentally sensitive fluorophore. C93 and neighbouring residues are shown in Figure 21: ssDNA is shown in orange (as a ribbon/ladder); two SSB subunits are shown. When adjacent residues are labelled the expectation is that a similar signal to that achieved with the preferred C93 site is observed.
The C93 labelling surface accessibility (ASA) value was over 30%, in higher end of intermediately accessible residues. The immediately adjacent residues to C93 are the G92 and E94, which have ASA values of 35.3% and 55.8%, respectively. K151 on another loop is very close in three dimensional space to the C93 residue and is also expected to give valuable signal when labelled. The ASA value for this residue is 68.8%.

DETAILED DESCRIPTION OF THE INVENTION

The SSB from Pla*smodium falciparum* (PfSSB) has been characterised and its structure determined[7-9], which overall is very similar to bacterial SSBs. The crystal structure of *Pf*SSB (24.5 kDa) DNA-binding domain showed it be structurally highly similar with the *Ec*SSB (19.5 kDa) with major structural differences can be expected to be at the intrinsically disordered region, which is 57 amino acids long in *Ec*SSB and over 80 amino acids long in *Pf*SSB [7, 8, 10, 11]. This region is unstructured and hence not visible in the crystal structures of the both species SSBs. The inventors hypothesised that ssDNA binding is simpler than with *E. coli,* in that the 65-70-base binding mode, that is 65-70

bases of ssDNA wrap around the SSB tetramer, predominates over the 35-base binding mode. So here, *Plasmodium* SSB such as PfSSB has been used to develop a probe for ssDNA by forming an adduct with a fluorophore. The best combination of fluorophore and labelling position is believed to be a diethylammoniumcoumarin at C93. This adduct showed tight and rapid binding of ssDNA with a maximum fluorescence change of 20-fold. This fluorescent signal could then be used to characterise both the mode and kinetics of binding. The ability of this biosensor to monitor ssDNA formation in a real-time assay is also shown.

Thus the invention provides a modified *Plasmodium* SSB comprising at least one detectable label attached to an amino acid of the protein. Suitably the detectable label is on a short loop of residues joining the two beta sheets (β1 and β1').

In one aspect, the invention relates to a ssDNA binding molecule, such as a protein, comprising a polypeptide wherein said polypeptide comprises a cysteine residue for attachment of at least one reporter moiety, and comprises at least one reporter moiety attached thereto.
Suitably said cysteine residue for attachment of at least one reporter moiety is positioned such that said reporter moiety undergoes a change in fluorescence upon ssDNA binding.

It is an advantage of the invention that simpler signals are provided. By simpler signals it is meant a response that approximates to, or is, a single mode binding.

Prior art biosensors such as those based on *E.coli* sequence (Kunzelmann et al 2010 Biochemistry vol. 49 pages 843-852) exhibit multi-mode binding. In more detail, those sensors exhibit at least two modes of binding. These alternate binding modes affect the fluorescent response of the sensor proteins. This can present a problem in use of the prior art biosensors. It is an advantage of the sensors of the invention that they are strongly shifted or strongly biased towards single mode binding. Suitably, the sensors are used for single mode binding. Suitably measurements using the biosensors of the invention are interpreted as single mode binding.

It is an advantage of the invention that a far larger signal is generated than the prior art sensors. Prior art sensors generate, at best, six to seven times signal on ssDNA binding compared to unbound (*E.coli* based ssDNA biosensors). In contrast, sensors of the present invention offer signals of at least twelve times, more typically approximately

twenty times signal in the bound compared to unbound state. This is a dramatic improvement on the performance of prior art biosensors.

LABEL ATTACHMENT

The *Plasmodium* protein has a very similar overall three dimensional structure to the *E.coli* protein. However, the arrangement of DNA binding to the plasmodium proteins is very different. One difference is that the DNA binds in the opposite direction to the *Plasmodium* protein compared to the *E.coli* protein. This is an extremely surprising observation. It is especially surprising considering the overall three dimensional structural similarity between the two proteins.

The inventors have labelled the ssDNA binding protein of *Plasmodium* at the position of cysteine 93. This is a naturally occurring cysteine present in the wild-type protein. This site is near to the DNA binding groove of the ssDNA binding protein. The sequence location of the dye attachment site is completely different to any of the prior art biosensors. Moreover, the three dimensional structure of the dye attachment site is completely different to any prior art biosensor. For example, dye attachment to the prior art *E.coli* ssDNA binding protein based biosensors has been at the very end of a loop present on the protein surface of the *E.coli* polypeptide. By contrast, the inventors have attached the dye to an entirely different structural part of the *Plasmodium* protein. The attachment is not on the same loop as the *E.coli* prior art sensors. The dye is attached to a surface of the *Plasmodium* protein, but on a very different secondary structure.

This clear departure from the proteins known in the art is illustrated in non-limiting Example 1 below.

Thus it is clear that there is no structural relationship between the prior art label locations and the locations of labels taught herein. The labelling taught herein is to a completely different part of the three dimensional structure which was not, and could not have been, predicted from an understanding of the prior art.

It is a special advantage of the invention that the new site for attachment of dye has been selected. Prior art biosensors attached dyes to various locations in the *E.coli* ssDNA binding proteins. The equivalent positions have been tried by the inventors in the context of the *Plasmodium* scaffold/backbone. However, "transplanting" those

prior art attachment points from the *E.coli* polypeptide into the *Plasmodium* polypeptide has not produced useful biosensors. We present comparative data showing that the sensors for the invention have significant advantages over merely trying to use prior art attachment points from the *E.coli* protein in the *Plasmodium* protein.

Suitably the label or dye is attached to the C93 residue of the *Plasmodium falciparum* single stranded DNA binding protein.

It is an advantage of the invention that the signal produced is quite possibly the highest ever recorded compared to the fluorophores used in any prior art setting.

The corresponding amino acid residues in other SSBs can be identified based on sequence homology for example using the alignment shown herein.

DYES/LABELS

The SSB of the invention is modified in order to include a detectable label attached to an amino acid of the protein and whose detectable characteristics alter on binding single stranded DNA. This alteration may, but not necessarily, result from a change in protein conformation. Whether the alteration is due to a change in the protein confirmation or not, the change in the detectable characteristics is due to an alteration in the environment of the label, which is bound to the single stranded DNA binding protein.

Labels used with the invention can give various signals, but preferred labels are luminescent labels. Luminescent labels include both fluorescent and phosphorescent labels. However, the use of other labels is envisaged. For example, electrochemical labels could be used wherein the alteration in the environment of the labels will give rise to a change in redox state. Such a change may be detected using an electrode.

The detectable label is preferably a fluorescent label. The use of fluorescent labels, which may be excited to fluoresce by exposure to certain wavelengths of light, is preferred.

Attachment of the detectable label to the SSB may sometimes reduce the affinity of the protein for single stranded DNA. However, this does not prevent an alteration of the detectable characteristics on binding single stranded DNA which can be quantitatively measured.

Preferably the fluorescent label is a coumarin or a rhodamine.

Shorter excitation wavelengths, such as those of the coumarin labels, are useful in providing a biosensor suitable for measurement of rapid kinetics of single stranded DNA formation, for example in real time assays of DNA helicase activity *in vitro*. On the other hand, longer excitation wavelengths such as those of the rhodamine fluorophores, which also demonstrate greater photostability, make such a labelled biosenor more suitable for applications such as single molecule studies and high throughput assays. Therefore, labels of the invention may have shorter excitation wavelengths of between about 400-460nm or longer excitation wavelengths of between about 540-600nm.

Thus it is possible to attach Rhodamine dyes to the *Plasmodium* ssDNA binding protein, for example at cysteine 93. This produces a signal change, but the sensor requires careful handling in titration type experiments. The reason is that Rhodamine type dyes have tended to produce nonlinear signal changes when attached at C93. Thus, in some embodiments it is possible to use Rhodamine type dyes. However, Rhodamine type dyes attached to *Plasmodium* proteins at C93 may produce complex responses requiring careful interpretation. For these reasons, Coumarin type dyes are preferred. Coumarin type dyes are especially preferred for attachment at C93 of the *Plasmodium* ssDNA binding protein.

Suitably the dyes/labels used in the invention are Coumarin type dyes/labels.

Suitably any coumarin may be used, including the iodoacetamide- and maleimide-linked diethylaminocoumarins. Examples of diethylaminocoumarins include *N*-[2-(1-maleimidyl)ethyl]-7-diethylaminocoumarin-3-carboxamide and *N*-[2-(iodoacetamido)ethyl]-7-diethylaminocoumarin-3-carboxamide (sometimes referred to as referred to as 7-diethylamino-3-((((2-iodoacetomido)ethyl)amino)carbonyl)coumarin) (IDCC).
Suitably the coumarin used is IDCC.

In addition other fluorophore types, known to have fluorescence intensity that depends on physical environment such as interactions with protein surfaces, may be used including:

CPM (7-(diethylamino)-3-[4'-(1-maleimidyl)phenyl]-4-methylcoumarin)

Alexafluor 350 – C5 maleimide
Acrylodan (6-acryloyl-2-dimethylaminonaphthalene)
Badan (6-bromoacetyl-2-dimethylaminonaphthalene)
MIANS (2-(4′-maleimidylanilino)naphthalene-6-sulfonic acid)
IAEDANS (5-[2-[(2-Iodo-1-oxoethyl)amino]ethylamino]-1-naphthalenesulfonic acid)
Alexa Fluor 488 maleimide
Fluorescein-5-maleimide
Cy3-maleimide (1-(6-{[2-(2,5-dioxo-2,5-dihydro-1H-pyrrol-1-yl)ethyl]amino}-6-oxohexyl)-2-[(1E,3E)-3-(1-ethyl-3,3-dimethyl-5-sulfo-1,3-dihydro-2H-indol-2-ylidene)prop-1-enyl]-3,3-dimethyl-3H-indolium)
Pyrene-maleimide

ATTACHMENT OF LABELS

The SSBs of the present invention have labels attached to certain amino acid(s).

The present invention also provides a method for making a modified SSB which involves modification of a SSB to include a detectable label attached to an amino acid of the protein.

The label may be attached to the SSB by any conventional means known in the art. For example, the label may be attached via amines or carboxyl residues on the protein. However, especially preferred is linkage via thiol groups on cysteine residues.

In a further aspect of the invention, the modified SSB may comprise more than one detectable label. In this case, the labels are preferably attached to separate amino acid residues.

In an embodiment of the invention, each monomer of the SSB has the same number of labels attached to it. There may be one label per monomer or alternatively two, three or four labels attached to each monomer. Preferably there is one label per monomer.

In a further aspect of the invention, where more than one monomer of the SSB tetramer is modified to include a detectable label, such labels may stack.

If appropriate, natural cysteine residues in the sequence of the SSB may be used for the attachment of the label. However, where no natural cysteine residues are available for label attachment, or where a non-cysteine site is chosen for label attachment, cysteine residues may be engineered into the sequence of the SSB, suitably by site-directed mutagenesis.

The invention provides a SSB wherein a wild-type non-cysteine residue is replaced by a cysteine residue.

Site-directed mutagenesis will be performed by methods well known in the art for this purpose. Briefly, however, the gene encoding the SSB is isolated, and oligonucleotide probes are constructed to alter by recombination, or more suitably by amplification, the codon encoding the amino acid which it is desired to change into a codon encoding cysteine. The mutated gene is subsequently expressed, typically in a bacterial expression system, to produce the mutated protein.

The invention also provides a SSB wherein the label is attached to a region of the protein surface. Regions of the protein such as a subunit-subunit interaction surface are not exposed and are thus unsuitable for labelling purposes. Surface located residues are more easily accessible for labelling purposes and are less likely to disrupt the overall shape of the protein when labelled.

Preferably, residues chosen for label attachment are located in a region of the protein surface, which is above a binding channel and on loops of the protein structure. 3D structures of several SSBs are known in the art. The residues chosen are not in a region of the protein which would directly interact with the DNA binding surface per se but are close enough to the DNA surface such that the attached label might be affected by the presence of the DNA. This strategy is a compromise between disrupting DNA binding by, for example, sterically hindering DNA binding (by being "too close") and getting no signal change (by being "too far away"). However, an alteration in the environment of the label may result from a conformational change in a region of the protein to which the label is not directly bound.

Fluorophores will rarely be attached to an amino acid directly, but will instead be attached via a linker. The choice of linker can also have an effect on the way the labelled SSB functions, as the size, shape and flexibility of the linker can change the ability of a linker to come into proximity with other groups.

Labels are preferably attached to the SSB in a manner that does not introduce a new chiral centre. Thus the label-protein adduct does not exist in diastereoisomeric form thus allowing a substantially homogenous labelled SSB to be obtained. This can be achieved by the use of linkers such as the haloacetamides (preferably iodoacetamides).

When using labels such as rhodamines it is also preferable to minimise any non-covalently bound rhodamine since this is likely to have high fluorescence and increase the background levels.

After attachment of the label, labelled protein will usually be purified to separate it from free label and from any mis-labelled protein. The mis-labelled protein may be unlabelled protein with which label did not react or protein where label has attached in the wrong position (either in place of or in addition to the desired label). During

purification of the labelled protein, treatment with a thiol reagent may be included, such as β-mercaptoethanol, dithiothreitol or sodium 2-mercaptoethanesulfonate as this can improve the fluorescence response of the protein.

Where more than one label can be attached, it is preferred to use the protein in homogenous form. A homogenous form, e.g. pure single-labelled species, may be purified (for example by ion exchange and/or hydrophobic interaction chromatography) to obtain homogenous, double-labelled species. Single and double labelled SSBs can be distinguished by methods such as electrospray mass spectrometry.

The detectable label preferably shows an increase in its detectable characteristics upon binding single stranded DNA. Advantageously, this is at least 12-fold, more preferably 20-fold.

NUCLEOTIDE BINDING

Prior art biosensors such as the *E.coli* based ssDNA sensor have tended to bind 35 nucleotide segments. The inventors have observed that the 35 nucleotide long ssDNA is insufficient to wrap completely around tetramer. Thus, when binding a 35 nucleotide ssDNA, at most only approximately 2 out of the 4 available fluorescent dye molecules are likely to be affected by binding. This contributes to a lower signal which is a problem of the prior art sensors. In practice, such a short run of nucleotides may wrap so that they only affect 1 of the 4 possible fluorophores – with one end of the ssDNA associating with a protein a certain distance from a first fluorescent dye, continuing around the ssDNA binding protein and affecting one such fluorophore, and ending before reaching the fluorophore located on the third such molecule of the tetrameric sensor, which would produce an even lower signal. In contrast, it is an advantage that the homo-tetramers of the present invention bind in a 65-70 nucleotide mode, which means that each of the tetramers is likely to be affected by ssDNA binding leading to an advantageously larger signal.

The *Plasmodium* ssDNA binding protein has a large "unstructured region" exceeding 80 amino acids in length. By contrast, the *E.coli* protein has a much shorter 57 amino acid long unstructured region. Without wishing to be bound by theory, the inventors believe that differences in the unstructured region lead to a decreased cooperativity in the *Plasmodium* protein. It is preferred that the unstructured region is retained in the biosensor molecules of the invention, since deletion of the unstructured region may affect DNA binding.

It is an advantage of the invention that certain assay conditions are enabled in which prior art sensors such as *E.coli* sensor do not work. One example of such conditions is low salt conditions. Prior art *E.coli* based sensors tend to give peculiar shaped binding curves when used in low salt conditions. By contrast, *Plasmodium* based sensors of the invention give good results even in low salt conditions. Suitably "low salt" conditions means 20mM.

It is an advantage of the invention that DNA/Sensor ratios are less likely to adversely affect signal. For example the sensor of the invention performs well with DNA/sensor ratios in the range sensor to DNA (take in 65 base units): From 2 to >20 fold.

THE SINGLE STRAND DNA-BINDING DOMAIN PROTEIN FAMILY

The single strand DNA-binding domain protein family (SSB family) is an example of a family of proteins that interact with single stranded DNA and unwound double stranded DNA. This family of proteins is defined in the SCOP database (Murzin et al., (1995) 3 Mol Biol 247, 536-540) and has an all-beta structure with an OB fold (barrel, closed; n=5, S=10). It is family 50263. This family is distinct from SCOP family 50315 which has a structure of an OB fold (barrel, open; n*=5, S*=8). The SSB family of proteins play essential roles in DNA replication, recombination and repair both in prokaryotes and eukaryotes. Members of this family of proteins have been investigated from several organisms and there are structures of the protein in the presence and absence of DNA.

SEQUENCES

Suitably the invention may be applied to any *Plasmodium* single stranded DNA binding protein.

Most suitably the invention is applied to the *Plasmodium falciparum* single stranded DNA binding protein.

Most suitably the invention is applied to a polypeptide derived from, having or consisting of the sequence of SEQ ID NO:1

SEQ ID NO:1 - PfSSB aa78-285 (=207aa)

NEKSLNKIMLIGRVG**C**EPD IKILNGGDKVATFSLATNEF WRDRNTNELKSKTDWHRIVVY
DQNIVDLIDKYLRKGRRVY VQGSLHTRKWHTNDMNSQPK QITEIILSYNKGDLIFLDDKR
NFNQRNNSNNINSENQQHI NNEHINNNNINNGNDFMPLN SNDKIIEDKEFTDRLDDNNEE
NNFQSNSETFDKQEGIYDK MNVQEFEE

Regarding numbering: the wild type Plasmodium falciparum gene gives an extra 77 amino acids, not present in the exemplary sequence presented above. Thus the first amino acid in the exemplary sequences presented here is numbered 78 (N78). C93 (Cys 93 or cysteine 93) is in **bold**.

TRUNCATIONS

Suitably full length *Plasmodium* ssDNA binding protein is used.

It should be noted that the N-terminus of the protein comprises the DNA binding domain (DBD). Suitably, the N-terminus of the protein is not truncated. Suitably the protein contains the whole N-terminal region of the protein.

C-Terminal Truncations

The C terminal end of the *Plasmodium* ssDNA binding protein may be truncated. For example, up to 120 amino acids may be removed from the C terminal end of the polypeptide, with the resulting truncated protein expressing and purifying adequately. Thus the invention also includes sensor molecules as described herein with up to 120 amino acids truncated from the C terminal end; suitably 100 or fewer amino acids are truncated.

It is believed that the final ~87 residues are not likely to be involved in DNA binding. Therefore suitably 87 or fewer amino acids are truncated. If 87 residues are removed, then the protein may be less stable and may not be suitable when high stability is desired.

Thus, suitably 80 or fewer amino acids are truncated; suitably 60 or fewer amino acids are truncated.

It is likely that this deleterious effect on stability is present at least above ~50 C-terminal residues removed (truncated). Therefore, suitably 40 or fewer amino acids are truncated.

Up to 35 residues can be removed from the C-terminus without significantly changing the protein stability or fluorescence response. Thus, suitably 35 or fewer amino acids are truncated from the C terminal end of the sensor of the invention.

Most suitably 8 or fewer amino acids are truncated from the C terminal end of the sensor of the invention.

Several preferred C-terminal truncations are tested - results are in the examples section.

Exemplary truncations as discussed above:

SEQ ID NO:2 PfSSB •8: aa78-277

NEKSLNKIMLIGRVG**C**EPD IKILNGGDKVATFSLATNEF WRDRNTNELKSKTDWHRIVVY
DQNIVDLIDKYLRKGRRVY VQGSLHTRKWHTNDMNSQPK QITEIILSYNKGDLIFLDDKR
NFNQRNNSNNINSENQQHI NNEHINNNNINNGNDFMPLN SNDKIIEDKEFTDRLDDNNEE
NNFQSNSETFDKQEGIYDK

SEQ ID NO:3 PfSSB •35: aa78-250

NEKSLNKIMLIGRVGCEPD IKILNGGDKVATFSLATNEF WRDRNTNELKSKTDWHRIVVY
DQNIVDLIDKYLRKGRRVY VQGSLHTRKWHTNDMNSQPK QITEIILSYNKGDLIFLDDKR
NFNQRNNSNNINSENQQHI NNEHINNNNINNGNDFMPLN SNDKIIEDKEFTD

SEQ ID NO:4 PfSSB •87 (entire cooperativity domain deletion): aa78-198

NEKSLNKIMLIGRVG**C**EPD IKILNGGDKVATFSLATNEF WRDRNTNELKSKTDWHRIVVY
DQNIVDLIDKYLRKGRRVY VQGSLHTRKWHTNDMNSQPK QITEIILSYNKGDLIFLDDKR

N-Terminal Region

It should be noted that the apicoplast localisation sequence (ALS) is suitably not part of the sensor polypeptide of the invention. The ALS is 76 amino acids long at the N-terminus of the naturally occurring sequence. Suitably the polypeptide of the invention is expressed without the 76 amino acid ALS.

It should be noted that the first (77th) amino acid of *Plasmodium* protein without the ALS is methionine. However, as is commonly observed, this initiator methionine is cleaved from the mature protein. Thus, suitably the first amino acid of the biosensor of the invention is the 78th amino acid of the full wild-type sequence (i.e. the wild-type sequence minus the ALS and minus the first methionine at position 77).

For the avoidance of doubt, numbering of the amino acids residues is done with reference to the sequence of wild-type *Plasmodium falciparum* single stranded DNA binding protein having Uniprot Accession Number W7KL30 (see below).

Suitably the current version of sequence database(s) are relied upon. Alternatively, the release in force at the date of filing is relied upon. For the avoidance of doubt, UniProt Release 2015_07 is relied upon. In more detail, the UniProt consortium European Bioinformatics Institute (EBI), SIB Swiss Institute of Bioinformatics and Protein Information Resource (PIR)'s UniProt Knowledgebase (UniProtKB) Release 2015_07 (24-Jun-2015) is relied upon. UniProt (Universal Protein Resource) is a comprehensive catalogue of information on proteins.

For the avoidance of doubt, an exemplary wild-type *Plasmodium falciparum* single stranded DNA binding protein sequence (with the N-term ALS and Met 77 removed) forming the basis of preferred sensor molecules of the invention is provided as SEQ ID NO: 1.

Amino acid addresses given in the application correspond to the numbering of the *Plasmodium falciparum* single stranded DNA binding protein reference sequence of SEQ ID NO:1. Where truncated or extended forms of *Plasmodium* single stranded DNA binding protein(s) are used as polypeptides in molecules of the invention (e.g. where a 6his tag is added or where a section of the polypeptide is deleted) then the amino acid numbering should be treated as corresponding to the equivalent section of the full length *Plasmodium falciparum* single stranded DNA binding protein reference sequence and not as an 'absolute' or rigidly inflexible numeric address. By way of explanation, if the description mentions a substitution of C93, this means amino acid 93 of the *Plasmodium falciparum* single stranded DNA binding protein reference sequence of SEQ ID NO: 1. If the polypeptide used is truncated by deletion of the first 77 amino acids, the address given will still be C93 (rather than e.g. C16) – this will be easily understood by the skilled reader to refer to the amino acid of the corresponding context with reference to the *Plasmodium falciparum* single stranded DNA binding protein sequence of SEQ ID NO:1, as is conventional in the art.

Clearly there are elements of the *Plasmodium* single stranded DNA binding protein wild type sequence which we teach are important to mutate, such as by substitution, to achieve certain advantages. However, there are also numerous residues which may or may not be mutated depending on operator choice. Clearly there are also numerous residues which should not be mutated in case such mutation would interfere with the function of the polypeptide. Typically it would be expected that if the skilled operator had a concern whether or not a particular residue could be mutated or not, they could

make the mutation and then test the resulting polypeptide to ensure that the desired property was retained in the mutated version. However, in order to provide further guidance on this point, the following comments are made:

The exemplary *Plasmodium falciparum* single stranded DNA binding protein has been studied in detail and each residue has been classified as set out in Table A below. Table A lists all residues, based on surface exposure. The PDB file used for the analysis was the 3ULP that was modified by removing the ssDNA. The structure does not include the ALS sequence or the unstructured C-terminal cooperativity domain.

Table A

# Residue	Amino acid	Accessibility % (ASA)	Buried/Intermediate/Exposed
77	M	n/a	n/a
78	N	64.2	Exposed
79	E	39.3	Intermediate
80	K	48.8	Exposed
81	S	9.6	Buried
82	L	4.1	Buried
83	N	6.3	Buried
84	K	7.0	Buried
85	I	0.1	Buried
86	M	0.0	Buried
87	L	0.1	Buried
88	I	0.0	Buried
89	G	3.3	Buried
90	R	35.2	Intermediate
91	V	2.6	Buried
92	G	35.3	Intermediate
93	C	33.0	Intermediate
94	E	55.8	Exposed
95	P	14.5	Intermediate
96	D	49.8	Exposed
97	I	26.6	Intermediate
98	K	60.9	Exposed
99	I	68.6	Exposed
100	L	40.6	Exposed
101	N	105.8	Exposed
102	G	104.0	Exposed
103	G	56.3	Exposed
104	D	45.8	Exposed
105	K	30.1	Intermediate
106	V	12.1	Intermediate
107	A	0.0	Buried
108	T	17.5	Intermediate
109	F	2.5	Buried

110	S	29.1	Intermediate
111	L	0.0	Buried
112	A	10.0	Buried
113	T	0.0	Buried
114	N	17.0	Intermediate
115	E	25.3	Intermediate
116	F	38.2	Intermediate
117	W	44.5	Exposed
118	R	49.1	Exposed
119	D	25.0	Intermediate
120	R	64.2	Exposed
121	N	38.4	Intermediate
122	T	47.9	Exposed
123	N	59.5	Exposed
124	E	44.7	Exposed
125	L	70.1	Exposed
126	K	39.1	Intermediate
127	S	59.1	Exposed
128	K	33.9	Intermediate
129	T	37.0	Intermediate
130	D	3.0	Buried
131	W	54.9	Exposed
132	H	4.4	Buried
133	R	40.6	Exposed
134	I	0.0	Buried
135	V	1.6	Buried
136	V	0.5	Buried
137	Y	46.8	Exposed
138	D	11.8	Intermediate
139	Q	29.6	Intermediate
140	N	78.8	Exposed
141	I	9.1	Buried
142	V	0.0	Buried
143	D	45.0	Exposed
144	L	41.3	Exposed
145	I	0.6	Buried
146	D	47.7	Exposed
147	K	55.1	Exposed
148	Y	57.0	Exposed
149	L	0.3	Buried
150	R	48.7	Exposed
151	K	68.8	Exposed
152	G	47.3	Exposed
153	R	31.9	Intermediate
154	R	4.1	Buried
155	V	0.0	Buried
156	Y	0.0	Buried
157	V	0.2	Buried
158	Q	8.9	Buried
159	G	2.3	Buried
160	S	15.4	Intermediate
161	L	1.8	Buried
162	H	14.0	Intermediate

163	T	18.9	Intermediate
164	R	47.6	Exposed
165	K	72.1	Exposed
166	W	42.4	Exposed
167	H	69.4	Exposed
168	T	61.1	Exposed
169	N	66.6	Exposed
170	D	96.0	Exposed
171	M	n/a	n/a*
172	N	89.7	Exposed
173	S	90.1	Exposed
174	Q	55.8	Exposed
175	P	52.1	Exposed
176	K	54.5	Exposed
177	Q	61.3	Exposed
178	I	27.3	Intermediate
179	T	18.9	Intermediate
180	E	3.7	Buried
181	I	1.6	Buried
182	I	11.0	Intermediate
183	L	0.6	Buried
184	S	42.6	Exposed
185	Y	37.1	Intermediate
186	N	37.3	Intermediate
187	K	51.4	Exposed
188	G	18.0	Intermediate
189	D	24.3	Intermediate
190	L	3.7	Buried
191	I	1.0	Buried
192	F	19.2	Intermediate
193	L	0.4	Buried
194	D	47.2	Exposed

*Not visible in the crystal structure.

The classification is as follows, making use of information on surface exposure:

Buried: Residues in the core of the structure where mutations are likely to affect the function of the sensor.

Exposed: Residues on the surface of the protein. Mutations of these residues are likely to retain the sensor function. Possibly any mutation may be allowed

Intermediate: Change in sensor function when mutating these residues is possible, but may be difficult to predict and conservative mutations are likely to be more successful e.g. routine testing of resulting mutants is particularly preferred when mutating these residues. Most of these are partially buried.

In more detail, the relative accessible surface area (ASA) of a residue is its ASA calculated using DSSP (Kabsch, W. and C. Sander (1983). "Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical

features." Biopolymers 22(12): 2577-637.) divided by its nominal maximum area as defined by (Chothia, C. (1976). "The nature of the accessible and buried surfaces in proteins." J Mol Biol 105(1): 1-12). Here, a residue is defined as exposed if its relative ASA is at least 40% of its nominal maximum area. A residue is defined as buried if its relative ASA is less than 10% of its nominal maximum area.

Buried Residues

A few conservative changes may be possible in the buried region. Suitably residues in the polypeptide part of a sensor molecule of the invention have at least 90% sequence identity to SEQ ID NO:1. Suitably any differences are conservative substitutions. Suitably residues in the polypeptide part of a sensor molecule of the invention have 100% similarity to SEQ ID NO:1. More suitably residues shown as 'buried' in table A are not mutated. Thus suitably residues in the polypeptide part of a sensor molecule of the invention which correspond to residues shown as 'buried' in table A are not mutated relative to SEQ ID NO:1. In other words, suitably residues in the polypeptide part of a sensor molecule of the invention shown as 'buried' in table A comprise the same residue as at the corresponding position in SEQ ID NO:1.

Thus in some embodiments, the polypeptide component of the sensor molecule of the invention suitably comprises amino acid sequence having 100% sequence identity to those residues shown as 'buried' in table A.

Intermediate Residues

Suitably residues shown as 'intermediate' in table A may be mutated. Thus suitably residues in the polypeptide part of a sensor molecule of the invention which correspond to residues shown as 'intermediate' in table A may be mutated relative to SEQ ID NO:1. In other words, suitably residues in the polypeptide part of a sensor molecule of the invention shown as 'intermediate' in table A may comprise a different residue (or no residue) from the corresponding position in SEQ ID NO:1.

Suitably a biosensor of the invention has at least 60% sequence identity to SEQ ID NO: 1. In some embodiments, the polypeptide component of the sensor of the invention suitably comprises amino acid sequence having at least 60% sequence identity to those residues shown as 'intermediate' in table A. In some embodiments, the polypeptide component of the sensor of the invention suitably comprises amino acid sequence having at least 68% sequence identity to those residues shown as 'intermediate' in table A, suitably least 70% sequence identity, suitably least 74%

sequence identity, suitably least 78% sequence identity, suitably least 82% sequence identity, suitably least 86% sequence identity, suitably least 90% sequence identity, suitably least 94% sequence identity, suitably least 98% sequence identity to those residues shown as 'intermediate' in table A.

5

However, it should be noted that these 'intermediate' residues are in fact partially buried. Thus, suitably these 'intermediate' residues are only mutated by substitution with a conservative residue relative to SEQ ID NO: 1. In other words, suitably the non-identical residues noted above comprise only conservative substitutions relative to the
10 corresponding residue in SEQ ID NO: 1.

Unless otherwise apparent from the text, mentions of 'sequence homology' suitably refer to sequence identity. However, occasionally it is more appropriate to refer to sequence similarity. Sequence similarity takes account of sequence identity and also
15 takes account of conservative substitutions (i.e. non-identical residues but where the residue is similar or conserved compared to the original residue). Assessing sequence similarity is well known in the art. Examples are provided in the examples section. In case any further guidance is needed, suitably the following parameters are used in the algorithm for calculating sequence similarity: BLOSUM62 matrix, gap penalty 10.0,
20 gapextend penalty 0.5; more suitably BLOSUM62 matrix, gapopen 10.0, gapextend 0.5, endopen 10.0, endextend 0.5, pairwise alignment. Suitably the polypeptide component of the sensor of the invention has at least 50% sequence similarity to SEQ ID NO: 1, suitably at least 60% sequence similarity, suitably at least 63% sequence similarity, suitably at least 65% sequence similarity, suitably at least 70% sequence similarity,
25 suitably at least 75% sequence similarity, suitably at least 80% sequence similarity, suitably at least 85% sequence similarity, suitably at least 90% sequence similarity, suitably at least 95% sequence similarity, suitably at least 98% sequence similarity, suitably at least 99% sequence similarity, most suitably 100% similarity to SEQ ID NO: 1.

30

Conservative substitutions may be made for example according to the table below. Amino acids in the same block in the second column and preferably in the same line in the third column may be substituted for each other:

ALIPHATIC	Non-polar	G A P
		I L V
	Polar - uncharged	C S T M
		N Q

	Polar - charged	D E
		K R
AROMATIC		H F W Y

More suitably residues shown as 'intermediate' in table A are not mutated. Thus suitably residues in the polypeptide part of a sensor of the invention which correspond to residues shown as 'intermediate' in table A are not mutated relative to SEQ ID NO:1. In other words, suitably residues in the polypeptide part of a sensor of the invention shown as 'intermediate' in table A comprise the same residue as at the corresponding position in SEQ ID NO:1.

In all cases, sequence similarity or sequence homology (identity) scores are calculated for the section or segment of the polypeptide/polynucleotide which corresponds to or aligns with the reference sequence such as SEQ ID NO: 1. Therefore, if the target sequence is shorter than the reference sequence such as SEQ ID NO: 1 (e.g. a truncation) then the score is calculated only for the length of the shorter or truncated sequence, i.e. for the 'overlapping region' between the two sequences.

Exposed Residues

Suitably residues shown as 'exposed' in table A (i.e. column II of table A) may be mutated. Thus suitably residues in the polypeptide part of a sensor of the invention which correspond to residues shown as 'exposed' in table A may be mutated relative to SEQ ID NO:1. In other words, suitably residues in the polypeptide part of a sensor of the invention shown as 'exposed' in table A may comprise a different residue (or no residue) from the corresponding position in SEQ ID NO:1.

In some embodiments, the polypeptide component of the sensor of the invention suitably comprises amino acid sequence having at least 21% sequence identity to those residues shown as 'exposed' in table A, suitably at least 30% sequence identity, suitably at least 40% sequence identity, suitably at least 50% sequence identity, suitably at least 60% sequence identity, suitably at least 63% sequence identity, suitably at least 65% sequence identity, suitably at least 70% sequence identity, suitably at least 75% sequence identity, suitably at least 80% sequence identity, suitably at least 85% sequence identity, suitably at least 90% sequence identity, suitably at least 95% sequence identity, suitably at least 98% sequence identity, suitably at least 99% sequence identity, most suitably 100% identity to those residues shown as 'exposed' in table A.

The invention has been exemplified using Plasmodium falciparum SSB. Other SSBs, in particular other Plasmodium SSBs, may be equally used. Exemplary sequences are presented below. Sequences are given for a variety of Plasmodium species with sequence alignment after. This illustrates the close sequence homology to *P. falciparum* protein, the exemplary sensor protein, in the DNA-binding region:

MNEKSLNKIMLIGRVGCEPD IKILNGGDKVATFSLATNEF WRDRNTNELKSKTDWHRIVV
YDQNIVDLIDKYLRKGRRVY VQGSLHTRKWHTNDMNSQPK QITEIILSY NKGDLIFLDD
KRNFNQRNNSNNINSENQQH INNEHINNNNINNGNDFMPL NSNDKIIEDKEFTDRLDDNN
EENNFQSNSETFDKQEGIYD KMNVQEFEE

Bold: N-terminal methionine (amino acid residue number 77), spliced (lost)
Italics: C-terminal co-cooperativity domain
underlined: C-terminal Acidic tip
double underlined: Cysteine residue 93, labelled with IDCC
Standard text: DNA binding region

>tr|W7KL30|W7KL30_PLAFO Uncharacterized protein OS=Plasmodium falciparum (isolate NF54) GN=PFNF54_00914 PE=3 SV=1
MGKRMLCIVFPLLIYFNYVLHRTYGYIIGDVKVHQLNNIINKRIIKSRKISMFKINLQND
FNYENKRFYNNMNRNV**M**NEKSLNKIMLIGRVGCEPDIKILNGGDKVATFSLATNEFWRDR
NTNELKSKTDWHRIVVYDQNIVDLIDKYLRKGRRVYVQGSLHTRKWHTNDMNSQPKQITE
IILSYNKGDLIFLDDKRNFNQRNNSNNINSENQQHINNEHINNNNINNGNDFMPLNSNDK
IIEDKEFTDRLDDNNEENNFQSNSETFDKQEGIYDKMNVQEFEE

>tr|A5KB16|A5KB16_PLAVS Single-strand binding protein, putative OS=Plasmodium vivax (strain Salvador I) GN=PVX_098000 PE=3 SV=1
MKALLLTSVMYVLCLPAIWGYVGDPKMQKLTVGGVWSRRKTSTRKVNLLKTKLQSDFNYE
GKRYYNNRPMTGEKSLNKISLIGRVGCEPDIKILNGGDKVATFSLATNEFWRDRNTNELK
SKTDWHRIVVYDQNIVDLVDKFLRKGRRVYIQGSLHTRRWFGNDLTNQPRQITEVVLSYN
KGDLIFLDDKRNFISRNASNVQSSESPPSTSEANVANSLGGNSDEGITPPNSHSDGAAEQ
DDFAHGLEGAAAEGFDSLEDGLGKEEGGIYDKMNTQEFDE

>tr|A0A060RP35|A0A060RP35_PLARE Single-stranded DNA-binding protein OS=Plasmodium reichenowi GN=SSB PE=3 SV=1
MGKRMLCIVFPLLIYFNYVLQRTYGYIIGDVKVHQLNNIINKRIIKNRKISMFKINLQND
FNYENKRFYNNMNRNV**M**NEKSLNKIMLIGRVGCEPDIKILNGGDKVATFSLATNEFWRDR
NTNELKSKTDWHRIVVYDQNVVDLIDKYLRKGRRVFVQGSLHTRKWHSNDMNSQPKQITE
IILSYNKGDLIFLDDKRNFNQRNNSNNINSENQQHINNEHINNNNNNNNNNINNGNDFMP
LNSNDKIIEDKEFTDRLDDNNEENNFQSNSDTFDKQEGIYDKMNVQEFEE

>tr|A0A0D9QP36|A0A0D9QP36_PLAFR Uncharacterized protein OS=Plasmodium fragile GN=AK88_01545 PE=3 SV=1
MKALLLSSVLYVLCLPAIWGYVGDPQLQKFTLGGISGQRRNSTRKVNFFKTKLQSDYNYE
GKRYYNNRPMTGEKSLNKIMLIGRVGCEPDIKILNGGDKVATFSLATNEFWRDRNTNELK
SKTDWHRVVVYDQNIVDLVDKFLRKGRRVYIQGSLHTRRWFGNDLNNQPRQITEVVLSYN
KGDLIFLDDKRNFISRNASNAQSSESSTSTSEANVSSSLGGNSDEGITPSNGHSDGDVEQ
DDFAHGLDGAAAEGFDSLEDGLGKEEGIYDKMNTQEFEE

>tr|Q4Z5Q5|Q4Z5Q5_PLABA Single-strand binding protein, putative (Fragment) OS=Plasmodium berghei (strain Anka) GN=PB000544.00.0 PE=3 SV=1
KIMKLILLVSSIYLTSILKIWGYIGNLKIHTFNDVSNKRITSRKVNNLKINMQNDFNYEN
NNKRFYNNRPVMNEKSLNKITLIGRVGCEPDIKILNGGDKVATFSLATNEFWRDRNSNEL

```
KSKTDWHRIVVYDQNIVDLIDKYLRKGRRVYIQGSLHTRKWFGNDLNNQPRQITEIVLSY
NKGDLIFLDDRRNFTSRNISNMPNSENKNEINNMNQNIGMDSDMISSMDNVEDDQSISSF
DHNPDDNSNNFESTGNPIDKDEGGIYDKMNTQEFEE

>tr|Q7PDM6|Q7PDM6_PLAYO SINGLE-STRAND BINDING PROTEIN OS=Plasmodium
yoelii yoelii GN=PY05238 PE=3 SV=1
MKLIFLVITIYLNSILKIWGYIGNLKIHTFNGGSNKRITSRKVHNLKINMQNDFNYENNN
KRFYNNRPVMSEKSLNKITLIGRVGCEPDIKILNGGDKVATFSLATNEFWRDRNSNELKS
KTDWHRIVVYDQNIVDLIDKYLRKGRRVYIQGSLHTRKWFGNDLNNQPRQITEIVLSYNK
GDLIFLDDRRNFTSRNISNMPNSENKNEINNMNQNGGMDNDIISSIDNIEDDQSISSFDH
NLDDNSNNFESTDNPIDKDEGGIYDKMNTQEFEE

>tr|Q4XMV9|Q4XMV9_PLACH Single-strand binding protein, putative
(Fragment) OS=Plasmodium chabaudi GN=PC000797.04.0 PE=3 SV=1
QIMKLIFLLSIIYLTNILKVWGYIGNLKIHTFSGVSNKRTTSRKVNNLKINMQNDFNYEN
NNKRFYNNRPPMSEKSLNKITLIGRVGCEPDIKILNGGDKVATFSLATNEFWRDRNTNEL
KSKTDWHRVVVYDQNIVDLIDKYLRKGRRVYIQGSLHTRKWFGNELNNQPRQITEIVLSY
NKGDLIFLDDRRNFSPRNALNMSNGENKNEINMNQHGGMDSDMISSMDNMEDDQSISSFD
HSMDDNSNTFDSTDNPIDKDEGGIYDKMNTQEFEE

>tr|W7A536|W7A536_9APIC Single-strand DNA-binding protein
OS=Plasmodium inui San Antonio 1 GN=C922_01230 PE=3 SV=1
MKALLLTSVMYVLCLPAIWGYVGDPKMQKLTVGGVCSQRRNSTRKANLFKTKLQSDFNYE
GKRYYNNRPMTGEKSLNRIMLIGRVGCEPDIKILNGGDKVATFSLATNEFWRDRNTNELK
SKTDWHRVVVYDQNIVDLVDKFLRKGRRVYIQGSLHTRRWFGNDLTNQPRQITEVVLSYN
KGDLIFLDDKRNFISRNVSNVQSSESTSSNSEANVASSLGGNSDEGVTPSNSHSDGGAEQ
DDFAHGLDGAPAEGFDSLEDGLEKEEGIYDKMNSQEFDE

>tr|W7AIE7|W7AIE7_PLAVN Single-strand DNA-binding protein
OS=Plasmodium vinckei petteri GN=YYG_01232 PE=3 SV=1
MKLIFLLSIIYLTNILKIWGYIGNLKIHTFSGVSNKRTTSRKVNNLKINMQNDFNYENNN
KRFYNNRPSMSEKSLNKITLIGRVGCEPDIKILNGGDKVATFSLATNEFWRDRNTNELKS
KTDWHRVVVYDQNIVDLIDKYLRKGRRVYIQGSLHTRKWFGNELNNQPRQITEIVLSYNK
GDLIFLDDRRNFIQRNAPNMSNGENKNEINNMNQHVGMDTDMISSMDNMEDDQSISSFDH
GMDDNSNSFDSTDNPIDKDEGGIYDKMNTQEFEE

>tr|K6UKM8|K6UKM8_9APIC Single-strand binding protein (Fragment)
OS=Plasmodium cynomolgi strain B GN=PCYB_103280 PE=3 SV=1
MKVLLLTSVMYVLCLPAIWGYVGDPKMHKLTVGGVWSERRSSTRKVNPFKTKLQSDFNYE
GKRYYNNRPMTGEKSLNKIMLIGRVGCEPDIKILNGGDKVATFSLATNEFWRDRNTNELK
SKTDWHRVVVYDQNIVDLVDKFLRKGRRVYIQGSLHTRRWFGNDLTNQPRQITEVVLSYN
KGDLIFLDDKRNFISRNASNVQSSESPSSTSEPSVANSLGGNSDEGITPPNSHTDGGVEQ
DDFAHGLEGAAAEGFDSLEDGLGKEEGIYDKMNTQEFDE

>tr|B3L6I1|B3L6I1_PLAKH Single-strand binding protein, putative
OS=Plasmodium knowlesi (strain H) GN=PKH_102410 PE=3 SV=1
MKALLLSTVMYVLCLPAIWGYVGEPKMHKLTMGGISNQRRTSTRKMNLLKTKMQSDFNYE
GSKRYYNNRPMTGEKSLNKISLIGRVGCEPDIKILNGGDKVATFSLATNEFWRDRNTNEL
KSKTDWHRIVVYDQNIVDLVDKFLRKGRRVYIQGSLHTRRWFGNDLTNQPRQITEVVLSY
NKGDLIFLDDKRNFISRNTSNVQSTESASSSNEANIANSLGGNADEAISASNSDSHGGVE
QDEFAHGLEGAAAEGFDSLEDGLGKEDGIYDKMNTQEFDE
```

The skilled person can choose the residues needed to be mutated and/or labelled by reference to the PfSSB of SEQ ID NO:1. In case further guidance is needed, exemplary Plasmodium sequences are aligned below:

CLUSTAL O(1.2.1) multiple sequence alignment

```
tr|B3L6I1|B3L6I1_PLAKH          -----MKALLLSTVMYVLCLPAIWGY-VGEPKMHKLTMGGISNQRRTSTRKMNLLKTKMQ
tr|A0A0D9QP36|A0A0D9QP36_PLAFR  -----MKALLLSSVLYVLCLPAIWGY-VGDPQLQKFTLGGISGQRRNSTRKVNFFKTKLQ
tr|W7A536|W7A536_9APIC          -----MKALLLTSVMYVLCLPAIWGY-VGDPKMQKLTVGGVCSQRRNSTRKANLFKTKLQ
tr|A5KB16|A5KB16_PLAVS          -----MKALLLTSVMYVLCLPAIWGY-VGDPKMQKLTVGGVWSRRKTSTRKVNLLKTKLQ
tr|K6UKM8|K6UKM8_9APIC          -----MKVLLLTSVMYVLCLPAIWGY-VGDPKMHKLTVGGVWSERRSSTRKVNPFKTKLQ
tr|W7KL30|W7KL30_PLAFO          MGKRMLCIVFPLLIYFNYVLHRTYGYIIGDVKVH--QLNNIINKRIIKSRKISMFKINLQ
tr|A0A060RP35|A0A060RP35_PLARE  MGKRMLCIVFPLLIYFNYVLQRTYGYIIGDVKVH--QLNNIINKRIIKNRKISMFKINLQ
tr|Q4XMV9|Q4XMV9_PLACH          ---QIMKLIFLLSIIYLTNILKVWGY-IGNLKIH--TFSGVSNKRT-TSRKVNNLKINMQ
tr|W7AIE7|W7AIE7_PLAVN          -----MKLIFLLSIIYLTNILKIWGY-IGNLKIH--TFSGVSNKRT-TSRKVNNLKINMQ
tr|Q4Z5Q5|Q4Z5Q5_PLABA          ---KIMKLILLVSSIYLTSILKIWGY-IGNLKIH--TFNDVSNKRI-TSRKVNNLKINMQ
tr|Q7PDM6|Q7PDM6_PLAYO          -----MKLIFLVITIYLNSILKIWGY-IGNLKIH--TFNGGSNKRI-TSRKVHNLKINMQ
                                     :  ::     :   :   :** :*: :::   .     .*  ..**   :* ::*

tr|B3L6I1|B3L6I1_PLAKH          SDFNYEGSK-RYYNNRPMTGEKSLNKISLIGRVGCEPDIKILNGGDKVATFSLATNEFWR
tr|A0A0D9QP36|A0A0D9QP36_PLAFR  SDYNYEG-K-RYYNNRPMTGEKSLNKIMLIGRVGCEPDIKILNGGDKVATFSLATNEFWR
tr|W7A536|W7A536_9APIC          SDFNYEG-K-RYYNNRPMTGEKSLNRIMLIGRVGCEPDIKILNGGDKVATFSLATNEFWR
tr|A5KB16|A5KB16_PLAVS          SDFNYEG-K-RYYNNRPMTGEKSLNKISLIGRVGCEPDIKILNGGDKVATFSLATNEFWR
tr|K6UKM8|K6UKM8_9APIC          SDFNYEG-K-RYYNNRPMTGEKSLNKIMLIGRVGCEPDIKILNGGDKVATFSLATNEFWR
tr|W7KL30|W7KL30_PLAFO          NDFNYENKRFYNNMNRNVMNEKSLNKIMLIGRVGCEPDIKILNGGDKVATFSLATNEFWR
tr|A0A060RP35|A0A060RP35_PLARE  NDFNYENKRFYNNMNRNVMNEKSLNKIMLIGRVGCEPDIKILNGGDKVATFSLATNEFWR
tr|Q4XMV9|Q4XMV9_PLACH          NDFNYENNNKRFYNNRPPMSEKSLNKITLIGRVGCEPDIKILNGGDKVATFSLATNEFWR
tr|W7AIE7|W7AIE7_PLAVN          NDFNYENNNKRFYNNRPSMSEKSLNKITLIGRVGCEPDIKILNGGDKVATFSLATNEFWR
tr|Q4Z5Q5|Q4Z5Q5_PLABA          NDFNYENNNKRFYNNRPVMNEKSLNKITLIGRVGCEPDIKILNGGDKVATFSLATNEFWR
tr|Q7PDM6|Q7PDM6_PLAYO          NDFNYENNNKRFYNNRPVMSEKSLNKITLIGRVGCEPDIKILNGGDKVATFSLATNEFWR
                                .*:***  .     **    *****:* ********************************

tr|B3L6I1|B3L6I1_PLAKH          DRNTNELKSKTDWHRIVVYDQNIVDLVDKFLRKGRRVYIQGSLHTRRWFGNDLTNQPRQI
tr|A0A0D9QP36|A0A0D9QP36_PLAFR  DRNTNELKSKTDWHRVVVYDQNIVDLVDKFLRKGRRVYIQGSLHTRRWFGNDLNNQPRQI
tr|W7A536|W7A536_9APIC          DRNTNELKSKTDWHRVVVYDQNIVDLVDKFLRKGRRVYIQGSLHTRRWFGNDLTNQPRQI
tr|A5KB16|A5KB16_PLAVS          DRNTNELKSKTDWHRIVVYDQNIVDLVDKFLRKGRRVYIQGSLHTRRWFGNDLTNQPRQI
tr|K6UKM8|K6UKM8_9APIC          DRNTNELKSKTDWHRVVVYDQNIVDLVDKFLRKGRRVYIQGSLHTRRWFGNDLTNQPRQI
tr|W7KL30|W7KL30_PLAFO          DRNTNELKSKTDWHRIVVYDQNIVDLIDKYLRKGRRVYVQGSLHTRKWHTNDMNSQPKQI
tr|A0A060RP35|A0A060RP35_PLARE  DRNTNELKSKTDWHRIVVYDQNVVDLIDKYLRKGRRVFVQGSLHTRKWHSNDMNSQPKQI
tr|Q4XMV9|Q4XMV9_PLACH          DRNTNELKSKTDWHRVVVYDQNIVDLIDKYLRKGRRVYIQGSLHTRKWFGNELNNQPRQI
tr|W7AIE7|W7AIE7_PLAVN          DRNTNELKSKTDWHRVVVYDQNIVDLIDKYLRKGRRVYIQGSLHTRKWFGNELNNQPRQI
tr|Q4Z5Q5|Q4Z5Q5_PLABA          DRNSNELKSKTDWHRIVVYDQNIVDLIDKYLRKGRRVYIQGSLHTRKWFGNDLNNQPRQI
tr|Q7PDM6|Q7PDM6_PLAYO          DRNSNELKSKTDWHRIVVYDQNIVDLIDKYLRKGRRVYIQGSLHTRKWFGNDLNNQPRQI
                                ***:************:******:***:**:*******::*******:*. *:::..**:**

tr|B3L6I1|B3L6I1_PLAKH          TEVVLSYNKGDLIFLDDKRNFISRNTSNVQSTESASSSNEANIANSL-------GGNADE
tr|A0A0D9QP36|A0A0D9QP36_PLAFR  TEVVLSYNKGDLIFLDDKRNFISRNASNAQSSESSTSTSEANVSSSL-------GGNSDE
tr|W7A536|W7A536_9APIC          TEVVLSYNKGDLIFLDDKRNFISRNVSNVQSSESTSSNSEANVASSL-------GGNSDE
tr|A5KB16|A5KB16_PLAVS          TEVVLSYNKGDLIFLDDKRNFISRNASNVQSSESPPSTSEANVANSL-------GGNSDE
tr|K6UKM8|K6UKM8_9APIC          TEVVLSYNKGDLIFLDDKRNFISRNASNVQSSESPSSTSEPSVANSL-------GGNSDE
tr|W7KL30|W7KL30_PLAFO          TEIILSYNKGDLIFLDDKRNFNQRNNSNNINSENQQHINNEHINN------NNINNGNDF
tr|A0A060RP35|A0A060RP35_PLARE  TEIILSYNKGDLIFLDDKRNFNQRNNSNNINSENQQHINNEHINNNNNNNNNNINNGNDF
tr|Q4XMV9|Q4XMV9_PLACH          TEIVLSYNKGDLIFLDDRRNFSPRNALNMSNGENKNEI-NMNQH-----------GGMDS
tr|W7AIE7|W7AIE7_PLAVN          TEIVLSYNKGDLIFLDDRRNFIQRNAPNMSNGENKNEINNMNQH-----------VGMDT
tr|Q4Z5Q5|Q4Z5Q5_PLABA          TEIVLSYNKGDLIFLDDRRNFTSRNISNMPNSENKNEINNMNQN-----------IGMDS
tr|Q7PDM6|Q7PDM6_PLAYO          TEIVLSYNKGDLIFLDDRRNFTSRNISNMPNSENKNEINNMNQN-----------GGMDN
                                **::**************:***  **  *  . *.     :                 *

tr|B3L6I1|B3L6I1_PLAKH          AISASNSDSHGGVEQDEFAHGLEG-AAAEGFDSLEDGLGKED-GIYDKMNTQEFDE
tr|A0A0D9QP36|A0A0D9QP36_PLAFR  GITPSNGHSDGDVEQDDFAHGLDG-AAAEGFDSLEDGLGKEE-GIYDKMNTQEFEE
tr|W7A536|W7A536_9APIC          GVTPSNSHSDGGAEQDDFAHGLDG-APAEGFDSLEDGLEKEE-GIYDKMNSQEFDE
tr|A5KB16|A5KB16_PLAVS          GITPPNSHSDGAAEQDDFAHGLEG-AAAEGFDSLEDGLGKEEGGIYDKMNTQEFDE
tr|K6UKM8|K6UKM8_9APIC          GITPPNSHTDGGVEQDDFAHGLEG-AAAEGFDSLEDGLGKEE-GIYDKMNTQEFDE
tr|W7KL30|W7KL30_PLAFO          MPLNSNDKIIEDK---EFTDRLDDNNEENNFQSNSETFDK-QEGIYDKMNVQEFEE
tr|A0A060RP35|A0A060RP35_PLARE  MPLNSNDKIIEDK---EFTDRLDDNNEENNFQSNSDTFDK-QEGIYDKMNVQEFEE
tr|Q4XMV9|Q4XMV9_PLACH          DMISSMDNMEDDQSISSFDHSMD--DNSNTFDSTDNPIDKDEGGIYDKMNTQEFEE
tr|W7AIE7|W7AIE7_PLAVN          DMISSMDNMEDDQSISSFDHGMD--DNSNSFDSTDNPIDKDEGGIYDKMNTQEFEE
tr|Q4Z5Q5|Q4Z5Q5_PLABA          DMISSMDNVEDDQSISSFDHNPD--DNSNNFESTGNPIDKDEGGIYDKMNTQEFEE
tr|Q7PDM6|Q7PDM6_PLAYO          DIISSIDNIEDDQSISSFDHNLD--DNSNNFESTDNPIDKDEGGIYDKMNTQEFEE
                                       .       .* .  :    : *:*  : : * : ******* ***:*
```

DIMERS

The homo-tetramers into which the sensor molecules associate are formed by binding of 2 dimer pairs to one another. It is possible to introduce mutation(s) to reduce the interaction of the pairs of dimers. For example, a Y156 mutation such as Y156R may be made to reduce the association of dimers into tetramers, thereby promoting dimer formation.

SEQ ID NO:5 PfSSB Y156R:

NEKSLNKIMLIGRVG**C**EPD IKILNGGDKVATFSLATNEF WRDRNTNELKSKTDWHRIVVY
DQNIVDLIDKYLRKGRRV**R** VQGSLHTRKWHTNDMNSQPK QITEIILSYNKGDLIFLDDKR
NFNQRNNSNNINSENQQHI NNEHINNNNINNGNDFMPLN SNDKIIEDKEFTDRLDDNNEE
NNFQSNSETFDKQEGIYDK MNVQEFEE

Thus, in one embodiment the invention relates to ssDNA sensors comprised of dimers of two individual sensor polypeptides of the invention.

The SSB structure indicates weaker interactions between pairs of subunits: the inventors teach that disruption of this dimer-dimer interface leads to viable dimeric proteins. In *Ec*SSB changing tyrosine 156 into arginine in the equivalent location, *Ec*SSB-Y78R, has been shown to lead to stable dimer of *Ec*SSB (Landwehr, M., Curth, U., and Urbanke, C. (2002) A dimeric mutant of the homotetrameric single-stranded DNA binding protein from Escherichia coli, Biol Chem 383, 1325-1333). Kozlov, A. G.; Weiland, E.; Mittal, A.; Waldman, V.; Antony, E.; Fazio, N.; Pappu, R. V.; Lohman, T. M., Intrinsically Disordered C-Terminal Tails of E. coli Single-Stranded DNA Binding Protein Regulate Cooperative Binding to Single-Stranded DNA. J. Mol. Biol. 2015, 427, (4), 763-774.

This mutation has been made in PfSSB and the dimer has been labelled with IDCC on the C93 location. This molecule is stable and has a 19-fold signal change when binding to ssDNA and can be used as a dimeric form PfSSB-based ssDNA biosensor. The location of Y156 residues on each subunit is shown in Figure 21.

There are several possible residues on the dimer-dimer interaction surface of the PfSSB tetramer that may be forming salt bridges, listed in table SB below.

Table SB - salt bridges

PfSSB mutation	Effect on PfSSB
K84D	Disrupt the possible salt bridge with D189
Y156R	Polar to positive charged residue
K187D	Disrupt the possible salt bridge with D189
D189K	Disrupt the possible salt bridge with K7 or K187

Suitably the sensor of the invention has one or more mutations from Table SB; suitably the sensor of the invention has two or more mutations from Table SB;
suitably the sensor of the invention has three or more mutations from Table SB;

suitably the sensor of the invention has all four mutations from Table SB.

Aspartate, D189, has two lysine residues in close proximity, which are K84 and K187. Either one of the lysines is close enough for possible salt bridge formation. The pair K84 and D189 is the most likely as they do fall closer to each other on the crystal structure. Suitable mutation to disrupt the dimer-dimer interaction would be to oppose the charges individually and this would break the salt bridge and lead destabilisation of the dimer-dimer interface (K84D, K187D and D189K).

In another embodiment, the invention relates to a tetramer, each tetramer comprising 4 individual polypeptides of the invention.

TAGS

It is often useful to tag proteins of the invention, for example to facilitate their purification after recombinant production.

Suitably tags may be placed at the extreme C-terminus or the extreme N-terminus of the sensor molecule.

Most suitably tags are placed at the C-terminus of the protein. Most suitably the N-terminus of the protein is not tagged.

Suitably one tag per protein molecule is used. Multiple tags per protein molecule may be used if desired, including multiple copies of the same tag or two or more different tags, for example it may be desirable to use a 6 His-tag for purification and in addition to use a Myc-tag for detection.

Tags may be removed from the sensor protein, for example by proteolytic cleavage, or may be retained on the sensor protein during use.

A hexahistidine tag (6his) may be added to the polypeptide part of the sensor molecule of the invention to simplify purification; most suitably a C-terminal hexahistidine tag is used. 6 His is a particularly useful tag for purification on nickel substrates. However, any suitable tag known in the art may be used.

Alternatively, the sensor molecule of the invention may be tagless. Tagless purification (if needed) is well known in the art.

MANUFACTURE AND PREPARATION

The polypeptide components of the sensor molecules of the invention may be produced by standard recombinant techniques, such as creating a nucleic acid encoding the amino acid sequence of the polypeptide, and then expressing the polypeptide in a host such as *E.coli*. Alternatively an in vitro translation may be used. Alternatively the polypeptide itself may be chemically synthesised.

The polypeptide(s) may be purified by any suitable method known in the art, such as 6His tagging the protein then purification using Ni-NTA beads.

To introduce the desired amino acid substitutions into the polypeptide of the invention, any suitable technique may be used such as site directed mutagenesis. For example, mutant PCR primers or oligonucleotides containing the desired nucleotide sequence may be annealed to a template and ligated, extended or amplified to produce a mutated nucleotide sequence encoding the desired substitution. Alternatively the desired nucleotide sequence may be synthesised chemically.

Exemplary techniques are presented in the Examples below.

pH

The sensors of the invention have the advantage of being usable under a wide range of pH conditions.

Suitably the pH of the assay is in the range 6.0 to 9.0. More suitably the pH of the assay is in the range about 6.5 to about 8.5. Most suitably the pH of the assay is in the range 6.5 to 8.5.

SOLUTION CONDITIONS/SALT CONDITIONS

Suitably salt conditions for assays or use of the sensors of the invention are from 20 mM to >200 mM. Preferred salts are those mentioned in the examples section.

ASSAY METHODS

The labelled SSBs of the invention can be used in assays for general biochemical use for detecting single stranded DNA in a sample. For example, in detecting strand separation of double stranded DNA which is catalysed by helicases and allows further processing such as repair and replication. Alternatively the labelled SSBs of the invention can be used in assays to measure the removal of single stranded DNA, or in assays to measure the decrease in double stranded DNA. These assays can be qualitative or quantitative. The invention is particularly useful for following the kinetics of reactions, because of the rapid reaction time of the SSBs. The assays can be used in screening methods for inhibitors of DNA processing enzymes. Such assays comprise assaying single stranded DNA levels *in vitro* using a SSB according to the present invention in the presence and absence of the inhibitors and assaying for an alteration in the single stranded DNA levels.

The single stranded DNA may be a single strand of DNA or may be a single stranded region of a DNA duplex.

The sample may be from any source, including serum, urine, saliva, sweat, tissue culture, cell extracts, cell lines, food, beverages, pharmaceuticals and environmental (for example water). If concentrations of single stranded DNA in the sample are high, samples may be diluted as necessary to achieve accurate quantification of single stranded DNA levels.

The kinetics of these methods may depend on salt concentration or the presence of a particular anion. A skilled person would be able to select the appropriate conditions to carry out the methods of the invention. The sensors of the invention are advantageously able to perform in a wide range of salt conditions, including low salt conditions, which is an advantage over prior art sensors.

These methods will typically be *in vitro* assays.

Thus the invention provides a method for detecting single stranded DNA in a sample comprising the steps of:

(i) mixing the sample with the modified SSB comprising at least one detectable label; and

(ii) detecting a change in the mixture arising from the interaction between the single stranded DNA and the SSB.

The change detected in (ii) can be related to the concentration of single stranded DNA in the sample.

By employing this method, using a modified SSB, it is possible to follow the kinetics of biological systems due to the extremely rapid reaction time of the method.

The invention also provides a modified SSB of the invention, for use in an assay of single stranded DNA.

A further aspect of the invention provides a polypeptide sequence of SEQ ID NO: 1 in which one or more wild type amino acid residues are changed to a cysteine residue.

A further aspect of the invention provides a polypeptide sequence of SEQ ID NO: 1 in which the wild type cysteine residue has a label attached.

The invention further provides PfSSB polypeptide sequence as disclosed herein with a label attached, suitably at C93 or the amino acid residue corresponding to same.

The invention further provides the nucleotide sequences which encode the polypeptide sequences disclosed herein.

FURTHER DEFINITIONS

The term "comprising" encompasses "having/including" as well as "consisting" *e.g.* a composition "comprising" X may consist exclusively of X or may include something additional *e.g.* X + Y.

The word "substantially" does not exclude "completely" *e.g.* a composition which is "substantially free" from Y may be completely free from Y. Where necessary, the word "substantially" may be omitted from the definition of the invention.

The term "about" in relation to a numerical value x means, for example, $x \pm 10\%$.

FURTHER ASPECTS

In another aspect, the invention relates to a process for making a SSB as described above which method comprises modification of a SSB to include a detectable label attached to an amino acid of the protein.

In another aspect, the invention relates to a fluorescent reagentless biosensor for single-stranded DNA using *Plasmodium* SSB.

In one aspect, the invention relates to a fluorescent biosensor for single-stranded DNA using *Plasmodium* such as *Plasmodium falciparum* single stranded DNA binding protein as the polypeptide, or protein scaffold.

INDUSTRIAL APPLICATION

The sensors of the invention are structurally different from the SSB proteins as they occur in nature, due to specific mutational differences compared to the naturally

occurring sequences and due also to the presence of dye molecules attached to the polypeptides which are not present in nature. Therefore the invention clearly involves the 'hand of man' and is eligible for patent protection, presenting molecules and methods which do not occur in the natural world.

The utility (industrial application) of the invention is demonstrated by its use in assays and experimental systems as set out in the following examples.

Further particular and preferred aspects are set out in the accompanying independent and dependent claims. Features of the dependent claims may be combined with features of the independent claims as appropriate, and in combinations other than those explicitly set out in the claims.

Where an apparatus feature is described as being operable to provide a function, it will be appreciated that this includes an apparatus feature which provides that function or which is adapted or configured to provide that function.

BRIEF DESCRIPTION OF THE DRAWINGS

Embodiments of the present invention will now be described further, with reference to the accompanying drawings, in which:

Figure 1 shows fluorescent spectra and titrations of DCC-PfSSB with ssDNA.

All data were at high salt: see Methods for details. The measurements were corrected for dilution. (A) Excitation and emission spectra were measure with 250 nM DCC-PfSSBC93 in the presence and absence of 595 nM of dT70. (B) Titrations of 250 nM DCC-PfSSBC93 with ssDNA. The polydT concentration is given in terms of 70-base binding sites. The average length of polydT is 500 bases. The breakpoint for dT70 titration is 256 nM. For dT35 titration this is 533 nM and for polydT it is 266 nM. (C) and (D) Titrations of a solution containing 250 nm DCC-PfSSB and 250 nM unlabeled PfSSB with dT35 or dT70.

Figure 2 shows effect of ionic strength on PfDCC-SSBC93 binding to ssDNA. A) dT35 titration. B) dT70 titration. C) polydT titration at high and low salt conditions.

The titrations were performed at same conditions as in Figure 1B. The solutions contained 250 nM PfDCC-SSBC93 in Graph A and C and 200 nM in Graph B. A) The breakpoint of fluorescence increase is 530 nM of dT35 at high salt and 544 nM at low salt concentration. B) The breakpoint for fluorescence increase for dT70 is 187 nM of dT70 at high salt and 193 nM dT70 at low salt. C) The breakpoint for fluorescence

increase for polydT was at 266 nM at high salt and 269 nM at low salt concentration of polydT.

Figure 3. DCC-PfSSBC93 association kinetics to various lengths of ssDNA with ssDNA in excess.

A) DCC-PfSSBC93 Association kinetics at high ionic strength to ssDNA. (●) dT70, (○) dT35 and (triangle) polydT. The data was fitted to double exponential, where the first phase rate constant followed concentration dependence. The second phase was significantly smaller in amplitude for all ssDNA lengths and less than 20% of the entire signal at all ssDNA concentrations. Only second slow phase fitted for dT70 followed concentration dependence. B) Association kinetics at low ionic strength conditions (B) fast phase and (C) slow phase. At low ionic strength the traces fitted well double exponentials and the observed rate constants for both phases followed concentrations dependence. (○)dT70, (◇) dT35 and (triangle) polydT. The second order rate constants determined from the linear fits are given in Table 3. The measurements were done on the stopped-flow apparatus and the final concentration of the DCC-PfSSBC93 was 5 nM and the ssDNA was varied from 25 to 400 nM.

Figure 4 shows DCC-PfSSBC93 association kinetics to various lengths of ssDNA when the DCC-PfSSBC93 in excess at high ionic strength.

(A) dT35, (B) dT70 and (C) polydT. The measurements were done at high ionic strength buffer with final concentration of DCC-PfSSBC93 varying from 50 nM to 750 nM with dT35 and dT70 and 50 nM to 1000 nM with polydT. The graphs show the traces for 100 nM and 500 nM DCC-PfSSBC93. The ssDNA concentration was kept at 10 % of the concentration of the DCC-PfSSBC93

Figure 5 shows association kinetics with excess DCC-PfSSB over ssDNA at high ionic strength.

Dependence of the observed rate constant of fast and slow phase on the concentration of the DCC-PfSSBC. The measurements as in Figure 4 were done at high ionic strength buffer with the ssDNA concentration was 10% of that of DCC-PfSSB and the data were fit to two exponentials. (A) Fast phases with linear fits linear fits. (B) Slow phases with linear fits. The rate the second order rate constants from these fits are in Table 4.

Figure 6 shows dissociation kinetics at high ionic strength.

The measurement were done by premixing 20 nM DCC-PfSSBC93, 25 nM dT70/50 nM dT35/25 nM polydT to form the PfSSB·ssDNA complex before mixing with varying concentrations of wtSSB. The final concentrations in the reaction were 10 nM DCC-PfSSBC93, 12.5 nM dT70/ 25 nM dT35/ 12.5 nM polydT and 0.25 - 4 μM wtSSB. The measurements were repeated at high (200 mM NaCl) and low (20 mM NaCl) ionic strength buffers.

Figure 7 shows dissociation kinetics at low ionic strength.

The measurements were done same as Figure 5 except the buffer had 20 mM NaCl. Left (A) to (C) graphs have not been normalised, whereas the (A) to (C) graphs on the right have been normalised and show the total fluorescence change.

Figure 8 shows dependence of dissociation kinetics on the wtSSB concentration.

The measurements (A)-(C) have been measured at high ionic strength (200 mM NaCl) and the measurements (D)-(F) have been done at low ionic strength (20 mM NaCl). (A) dT35, high salt. (B) dT70, high salt. (C) polydT, high salt. (D) dT35, low salt. (E) dT70, low salt. The values for gradients for all linear fits in graphs (A) to (E) are given in Table 5. (F) polydT low salt. The average koff of four concentrations of wtSSB measured is 1.3 s-1 for the fast phase and 0.12 s-1 for the slow phase.

Figure 9 shows plasmid unwinding by PcrA helicase.

(A) Fluorescent traces of PcrA unwinding 3094 bp plasmid with DCC-EcSSBG26C biosensor and DCC-PfSSBC93. (B) Plasmid unwinding of different plasmid lengths with DCC-PfSSBC93. (C) Plasmid unwinding durations plotted against plasmid lengths. The linear fir gives a gradient of 0.0575 (SEM 4.7 x 10-3). The reciprocal of the gradient gives the unwinding rate, which is 17.4 bp s-1 (+/- 1.4).

Figure 10 shows AddAB helicase assay with linear dsDNA.

(A) Fluorescence time courses of AddAxBx unwinding different dsDNA lengths using DCC-PfSSB as a reporter for production of ssDNA. (B) Linear fit of unwinding durations, plotted against the dsDNA lengths. The fit has a slope of 0.0114 s bp-1 (SEM 0.14 x 10-3). This gives unwinding rate of 97.7 bp s-1. (C) Fluorescence change plotted against the dsDNA lengths. The data points were fitted linear regression and this had a gradient of 8.4 x 10-2 (SEM 3 x 10-4) fluorescence units per base pair.

Figure 11 shows Plasmodium falciparum ssDNA bp; DNA is shown in brown (see the strand wound around the protein structure of PfSSB). The labelling position is highlighted only on one of the four subunits in the structure of PfSSB (C93 in magenta - see amino acid on central lower region of PfSSB).

Figure 12 shows *Escherichia coli* ssDNA bp; DNA is shown in blue (see the strand wound around the protein structure of EcSSB). The labelling position is highlighted only on one of the four subunits in the structure of EcSSB (G26 in red - see amino acid on far right lower loop of EcSSB).

Figure 13 shows a graph.

Figure 14 shows a graph.

Figure 15 shows a graph.

Figure 16 shows a graph.

Figure 17 shows a graph.

Figure 18 shows a graph.

Figure 19 shows a diagram.

Figure 20 shows a graph.

Figure 21 shows a diagram.

Figure 22 shows a diagram.

EXAMPLES

Example 1: Location of main labelling sites on PfSSB (invention); comparison to EcSSB (comparative only - not part of invention)

The position of C93 in PfSSB (Figure 11) and G26C in EcSSB (Figure 12) are shown on crystal structures. Note that the numbering system differs for each species, but is that used in the published crystal structures (Antony, E.; Weiland, E. A.; Korolev, S.; Lohman, T. M., Plasmodium falciparum SSB Tetramer Wraps Single-Stranded DNA with Similar Topology but Opposite Polarity to E. coli SSB. *J. Mol. Biol.* **2012**, 420, (4–5), 269-283; Raghunathan, S.; Kozlov, A. G.; Lohman, T. M.; Waksman, G., Structure of the DNA binding domain of *E. coli* SSB bound to ssDNA. *Nature Struct. Biol.* **2000**, 7, (8), 648-652).

The four subunits are coloured (shaded) differently and DNA is shown in brown (PfSSB) and blue (EcSSB). The labelling position is highlighted only on one of the four subunits in each structure for illustration purposes - suitably each subunit is labelled. PfSSB (C93 in magenta); EcSSB (G26 in red).

The crystal structure publications define names for the beta sheets and connecting loops, which are the same for each structure, and so this terminology is standard in the art and can be used herein:

C93 of PfSSB is located on a short loop of residues 91 – 96 (L1-2), joining two beta sheets (β1 and β2).

G26C of EcSSB is located on the end of a hairpin loop, residues 21 -29 (L1-1'), joining two interacting beta sheets (β1 and β1').

Example 2: Comparative Data

In this example we demonstrate technical differences between *Plasmodium* proteins according to the invention and prior art proteins such as from *E. coli*. The *Plasmodium*

proteins according to the invention show significant and demonstrable advantages which are not possessed by the prior art proteins.

For example, major differences are seen at low salt (where multiple binding modes with the prior art *E. coli* protein (EcSSB) are favoured). In addition the higher signal with *Plasmodium* SSB (PfSSB) proteins of the invention is shown.

The prior art *E. coli* data are taken with permission from:
Dillingham, M. S.; Tibbles, K. L.; Hunter, J. L.; Bell, J. C.; Kowalczykowski, S. C.; Webb, M. R., Fluorescent single-stranded DNA binding protein as a probe for sensitive, real time assays of helicase activity. Biophys. J. 2008, 95, 3330-3339; and
Kunzelmann, S.; Morris, C.; Chavda, A. P.; Eccleston, J. F.; Webb, M. R., Mechanism of interaction between single-stranded DNA binding protein and DNA. Biochemistry 2010, 49, 843-852.

Plasmodium data relate to the invention.

Example 2.1: Fluorescence response size.

Bigger fluorescence change is achieved with DCC-PfSSB compared to prior art *E.coli* proteins. Comparative data are shown as fluorescence spectra, using dT70 as DNA. In each case the "background" fluorescence in the absence of DNA was normalised to "1". Excitation and emission spectra are shown. Fig 13 shows prior art DCC-EcSSB, Fig 14 shows an example of the invention DCC-PfSSB.

Example 2.2: Better linear response to ssDNA.

Non-linear response is demonstrated with prior art DCC-EcSSB at low salt (20 mM NaCl), titrating dT70 as DNA into ~250 nM biosensor. Fig 15 shows prior art DCC-EcSSB, Fig 16 shows an example of the invention DCC-PfSSB. The improved response using the invention is clearly apparent.

Example 2.3: Multiple binding modes with prior art DCC-EcSSB.

Slow fluorescence decrease is seen with binding transient of 10-fold excess prior art DCC-EcSSB to dT70 at high concentrations. This is due to slow transition to 35-base-mode binding giving lower fluorescence with prior art DCC-EcSSB.

Advantageously, such a decrease not seen with these conditions using an example of the invention DCC-PfSSB.

Fig 17 shows prior art DCC-EcSSB, Fig 18 shows an example of the invention DCC-PfSSB.

200 mM NaCl was used.

Insets show short time scales

Example 2.4: Helicase assay

The assay measures double-stranded DNA (dsDNA) unwinding by the helicase AddAB. The diagram in Figure 19 illustrates the basis of the assay. The reaction, in the presence of prior art DCC-EcSSB or an example of the invention DCC-PfSSB (at identical concentrations) and at low ionic strength (20 mM NaCl), was initiated by addition of ATP. ATP-coupled unwinding leads to an increase of ssDNA to which DCC-SSB binds. This was observed as an increase in fluorescence. Once the entire dsDNA is unwound, a steady state of fluorescence should be observed.

Apart from the lower signal and sensitivity with prior art DCC-EcSSB, the fluorescence slowly decreases after an initial increase. This is due to slow transition to 35-base-mode binding, giving lower fluorescence, with prior art DCC-EcSSB. As well as the advantageous higher signal and sensitivity provided by the invention such as DCC-PfSSB, the invention also shows sustained signal i.e. advantageously does not show decreasing fluorescence as with the prior art protein. Results are shown in Fig 20.

Thus it is clearly demonstrated that proteins of the invention possess and display a range of advantageous characteristics and properties which are not present in the prior art proteins. The comparative data illustrates the advances and new functions offered by the proteins of the invention.

Example 3 - Production, Labelling and Use of SSB Sensors

Methods. PfSSB, wild-type and single cysteine mutants, were expressed from pET22b in C41(DE3) *E. coli* cells using the T7 promotor system (Lucigen, WI, U.S.A). The cells were grown at 37 °C in 0.5 L x 4 LB cultures to OD_{595} of 0.6 and the expression induced using 1 mM IPTG. The cells were harvested after 5 h by centrifugation and resuspended in 80 ml 50 mM Tris·HCl (pH 7.5), 200 mM NaCl, 1 mM EDTA, 1 mM dithiothreitol (DTT) and 10% sucrose, and stored -80 °C. PfSSB

was purified as described for *E. coli* SSB[2] with minor modification and omitting the polymine P precipitation step. The purification was done at 0-4 °C.

The cells were thawed and 1 protease inhibitor tablet (Roche) per 50 ml of cells was added. Cells were lysed using probe sonication, 4 x 20 s. Lysate was spun for 20 min at 38000 g. Solid ammonium sulphate was gradually add to 150 g/L and stirred for 30 min. The solution was spun at 14 000 g for 20 min and the pellet was resuspended with 50 mM Tris·HCl (pH 8.3), 200 mM NaCl, 1 mM EDTA and 20% glycerol (v/v) and stirred for 30 min to improve the resuspension of PfSSB. The solution was spun for 20 min at 38 000 g to remove any impurities and passed through 0.2 μm syringe filter membrane before loading on to a 5 ml heparin column (GE healthcare, Little Chalfont, U.K.). The column was pre-equilibrated with buffer A consisting of 50 mM Tris·HCl (pH 8.3), 1 mM EDTA, 1 mM DTT and 20% glycerol (v/v). The sample was loaded by mixing with the buffer A at ratio of 16/84 for sample/buffer A immediately before loading to column to prevent *Pf*SSB precipitation known to take place at low salt conditions. The column was washed with 10 times the column volume and a NaCl gradient from 0 to 1 M NaCl over 150 ml was run to elute the *Pf*SSB. *Pf*SSB was dialysed in a SnakeSkin Pleated Dialysis Tubing (Life technologies, Thermoscientific, Glasgow, U.K.)) to storage buffer of 50 mM Tris·HCl (pH 8.3), 1 mM EDTA, 500 mM NaCl, 1 mM DTT and 50% glycerol (v/v) overnight and concentrated using Viva Spin concentrator with a molecular weight cut-off of 10 000 Da. The concentration of *Pf*SSB was determined using absorbance at 280 nm with molar extinction coefficient of 95 800 M^{-1} cm^{-1} for the tetramer. *Pf*SSB was stored at -80 °C.

AddAB was a generous gift from Mark Dillingham (Bristol University, U.K.). PcrA and RepD proteins were expressed using standard purification protocols (Refs!). pCERoriD plasmids for PcrA helicase assay were prepared as prescribed before[4]. The linear dsDNA was prepared by digestion of pCER*oriD* plasmids with EcoRI restriction endonuclease (Roche) and gel purified from 1 % agarose gels using QIAGEN gel extraction kit (QIAGEN Ltd, Manchester, U.K.). dT_{35}, dT_{70} and polydT and other materials unless otherwise stated were obtained from Sigma-Aldrich (Gillingham, U.K.). IDCC, N-[2-(iodoacetamido)ethyl]-7-diethylaminocoumarin-3-carboxamide) was a gift from Dr J.E.T. Corrie (NIMR, London, U.K.)

Fluorophore labelling of PfSSB. Typically, 10 mg of *Pf*SSB wild-type and its cysteine mutants were labelled using iodoacetamide-thiol chemistry and all labelling steps were performed at room temperature. First, the PfSSB was incubated with 10-fold

excess of DTT over PfSSB monomers for 20 min. DTT was removed by elution through PD-10 column with the labelling buffer (20 mM Tris·HCl pH 7.5, 1.0 mM EDTA, 500 mM NaCl, 20% (v/v) glycerol). IDCC was added in 2.5-fold excess over ~ 25 μM PfSSB monomers and incubated 2 hours on end-to-end rotor. Excess IDCC was removed from the reaction using 2-mercaptoethanol sulphate by adding it in 10-fold excess over monomers and incubating for 20 min. The incubation mixture was passed through 0.2 μm syringe filter membrane (Pall Life Sciences, Portsmouth, U.K.) before loading on to a 10 ml P4 gel filtration column (Biorad) pre-equilibrated with the labelling buffer. Labelled-PfSSB was isolated by elution with labelling buffer and if required concentrated using Viva spin concentration with a molecular weight cut-off of 10 000 Da. The concentration of IDCC-labelled PfSSB was determined from the 430 nm absorbance using extinction coefficient for the IDCC, which 44 800 M^{-1} cm^{-1} at that wavelength. A mass spectroscopic analysis of DCC-PfSSBC93 gave a mass of 24 917.3 +/- 5 Da. This is close to the theoretical mass (24 913.8 Da) assuming the N-terminal methionine has been cleaved. The labelled-PfSSB was stored at -80 °C.

Absorbance and Fluorescence Measurements. Absorbance measurements were done on a Jasco spectrophotometer (Jasco Analytical Instruments Inc., Easton, MD, U.S.A.) using 3 mm pathlength quartz cells at room temperature. Fluorescence measurements were taken on Cary Eclipse spectrofluorometer (Varian, Palo Alto, CA, U.S.A.) with a xenon lamp using 3 mm pathlength quartz cells at room temperature. The fluorescence spectra were measured at the excitation and emission wavelengths given in Table 1, determined from spectra. All fluorescent property and kinetic measurements were performed at 20 °C. When the measurements were at high ionic strength buffer, the buffer had 25 mM Tris·HCl (pH 7.5), 200 mM NaCl and 10 μM BSA. At low ionic strength buffer NaCl concentration was lowered to 20 mM.

Fluorescence titrations. The titrations were done by excitation at 433 nm and emission at 473 nm for all ssDNA lengths.

dT_{70} titration curves for a solution containing equimolar labeled and unlabeled PfSSB were fitted using the following equation:

$$\Delta\text{Fluorescence} = (F_{max} - F_{min})(-b+\text{sqrt}(b^2 +4(K\text{-}1)PL))/(2L(K\text{-}1))). \quad (1)$$

Where $b = KX + P + L - XP$, X = [unlabelled PfSSB], L = [DCC-PfSSB], P = [total ssDNA], K = ratio of K_d values for the two SSB species, F_{min} = fluorescence in the absence of DNA, F_{max} = fluorescence of DCC-PfSSB·ssDNA.

Stopped-flow fluorescence. All assays with diethylammoniumcoumarin fluorescence were performed on a HiTech Scientific SF61MX stopped-flow (TgK Scientific Ltd., Bradford-on Avon, U.K.) with a mercury-xenon lamp. Excitation was at 436 nm and fluorescent emission was measured with a 455 nm cut-off-filter (Schott glass) and using 4 mm slits. KineticStudio software (TgK Scientific Ltd., Bradford-on Avon, U.K.) was used to record and quantify the data. Further analysis of the results was done using Grafit 7 (Erithacus Software Ltd.)[12].The concentrations given are final concentrations for the components. The fitting was done to minimum of three acquisitions and the measurements were repeated three times unless otherwise stated. The traces shown are examples of single acquisition.

Plasmid unwinding assay using PcrA helicase. Plasmid unwinding was measured on a stopped flow apparatus at 30 °C in buffer containing 50 mM Tris·HCl pH 7.5, 100 mM KCl, 1 mM EDTA, 10mM $MgCl_2$ and 10% (v/v) ethanediol. A typical reaction contained 0.5 nM pCER*oriD* plasmid, 190 nM PcrA, 2 nM RepD, 200 nM DCC-SSB or DCC-PfSSB and 1 mM ATP. Firstly, the plasmid was incubated with RepD for 30 s to allow RepD nicking. To this PcrA was added and incubated further 30 s. After this, the SSB was added to this syringes, A and B. Finally, the ATP was added to syringe B and the reaction was mixed.

Linear dsDNA unwinding assay with AddA^{D1172A} B^{D961A} helicase. The AddAB helicase assay was performed on the stopped-flow apparatus at 37°C in buffer containing 25 mM Tris·HCl (pH 7.5), 20 mM NaCl and 2 mM $MgCl_2$. The final reaction contained 0.5 nM *Eco*RI-linearized pCER*oriD* DNA, 1.25 nM AddA^{D1172A}B^{D961A} helicase, 175 nM DCC-PfSSBC93 and 250 nM ATP. dsDNA was first incubated 2 min with the DCC-PfSSB in syringe A. ATP was added to syringe B. After helicase and dsDNA incubation, the DCC-PfSSB was added to both syringes A and B, and incubated for 2 min before rapid mixing.

Abbreviations used:

IDCC, N-[2-(iodoacetamido)ethyl]-7-diethylaminocoumarin-3-carboxamide)

6-IATR, 6-iodoacetamidotetramethylrhodamine

5-IATR, 5-iodoacetamidotetramethylrhodamine

Production

Native *Plasmodium falciparum* SSB is transcribed and translated from nuclear DNA with a 76 amino acid apicoplast localization sequence (ALS) that is cleaved off once delivered to apicoplast. Here, the PfSSB was expressed without the localization sequence, but amino acids are numbered according the entire translated amino acid sequence. PfSSB has a single cysteine C93 that is located on the surface of PfSSB and was expected to be accessible to labelling using iodoacetamide/maleimide chemistry. Other labelling sites were also selected and tested, based on inventors' insights after previous work on *E.coli* SSB. Cysteines were introduced onto surface positions of the (C93A)PfSSB (Table 1). Cysteines were introduced in to positions that were equivalent to labeling sites used in the *E.* coli SSB[2, 13, 14]. G102C G103C W166C in the former are equivalent to G26C, G27C and W88C, respectively, in the latter, based on sequence homology and structure comparison. As the binding site of *Pf*SSB has been measured as ~65-70 bases, the measurements with dT_{70} were used as the simplest model for ssDNA binding, but along with shorter and longer lengths.

Fluorescent spectra and ssDNA equilibrium binding. The highest fluorescence increase was observed with labeling the wild type cysteine-93 of PfSSB with IDCC. When binding to dT70, the signal change was ~20-fold (Figure 1A). Fluorescence changes for other labeling sites are listed in Table 1 and for other lengths of DNA in Table 2.

Table 1. Signal changes of various PfSSB-fluorophore combinations.

Excitation and emission spectra were measure with 250 nM DCC-PfSSB in the presence and absence of 595 nM ssDNA at high salt. The polydT concentration was calculated as concentration of 70-base binding sites. Average length of polydT is 500 bases. The signal change is given as a ratio of fluorescence with nucleic acid divided by fluorescence without nucleic acid at wavelength of maximum emission.

PfSSB variant	**Labeling reagent**	**Excitation wavelength (nm)**	**Emission wavelength (nm)**	**Fluorescence ratio (± dT_{35})**	**Fluorescence ratio (± dT_{70})**	**Fluorescence ratio (± polydT)**
C93	IDCC	432	473	19.3	18.5	18.3
	5-IATR	558	577		9.1	

G103C	IDCC	433	473	2.7	3.5	
G102C	IDCC	435	467	-	-	1.6
W166C	5-IATR	556	577		12.8	
	6-IATR	557	577		3.5	

Table 2. Effect of ionic strength on the signal change of DCC-PfSSB.

The measurements were at same conditions as in Table 2, except the low ionic strength buffer had 20 mM NaCl. Excitation was at 432-437 nm, emission at 472-475 nm.

Buffer conditions	dT_{20}	dT_{35}	dT_{55}	dT_{70}	polydT
High salt	10.3	19.3	20.8	18.5	18.3
Low salt	-	16.9	-	15.0	13.5

PfSSB, fully labeled at C93 with IDCC (named hereafter DCC-PfSSB), overall gave the best characteristics and these are now described in detail.

The titration of DCC-PfSSB with dT_{70} gave linear increase in fluorescence, followed by a sharp break to a constant level (Figure 1B). This indicates that the binding of DCC-PfSSB is tight and stoichiometric with one DCC-PfSSB binding to one dT_{70}. The fluorescence increase with dT_{35} titration is also linear but shows two phases with different amplitudes, before the breakpoint. This suggests that one dT_{35} first binds to each DCC-PfSSB and the second phase is the binding of a second dT_{35} with a final ratio of DCC-PfSSB:dT_{35} being 1:2. The polydT titration also showed a linear fluorescence change.

The effect of IDDC labelling of C93 on the PfSSB affinity to ssDNA was tested by repeating the ssDNA titrations with the presence of equimolar unlabeled PfSSB and DCC-PfSSB. The traces with dT35 and dT70 show a low increase in fluorescence at the beginning of titration where the amount of available ssDNA is low compared to end of titration (Figure 1C and D). This indicates that the DNA binds tighter, and so preferentially, to unlabeled PfSSB. The weaker binding of DCC-PfSSB is most likely due to the bulk of the fluorescent modification on the surface. Fitting the titration curve with unlabeled and labeled PfSSB using Equation 1 gives unlabeled protein binding as 25-fold tighter.

Effect of salt concentration on SSB binding mode to ssDNA. This was measured by titrations at low salt conditions (20 mM NaCl), compared to the high salt concentration (200 mM NaCl). The fluorescence change upon binding to ssDNA at low salt was ~20% less, depending on the length of ssDNA (Table 2 and Figure 2). The stoichiometric binding was not affected by the change in salt concentration as the fluorescent breakpoint of linear fluorescence increase and constant fluorescence state in titration assays remain unaffected. The titrations with all ssDNA lengths also confirmed that maximum fluorescence observed was lower at low salt.

Association kinetics with excess DNA. Labeling PfSSB with an environmentally sensitive fluorophore enabled the determination of the binding kinetics of ssDNA. Second order binding constants were determined at high salt by rapidly mixing various lengths of ssDNA with DCC-PfSSB on stopped-flow apparatus. These included dT35, dT55, dT70 and polydT. The experiments were done at pseudo-first order conditions where the ssDNA was at high excess over PfSSB. The fluorescence time courses were biphasic with a small slow phase. They were therefore fit to two exponentials. For dT70 the second-order binding rate constant was 309 $\mu M^{-1}s^{-1}$, possibly diffusion controlled. The second, slower phase was also observed but this had amplitude less than 20% of the first phase and the observed rate constant for the phase did not follow concentration dependence, suggesting it possibly represents a slow rearrangement. The dissociation rate constant determined from the intercept of y-axis was 7.6 s^{-1}. The binding rate constant for different lengths of ssDNA are summarised in Table 3.

Table 3. Effect of ionic strength on association kinetics.

	Fast phase	Slow phase	
dT_{35}	116	-	High salt
dT_{35}	35.4	7.6	Low salt
dT_{70}	309	4.8	High salt
dT_{70}	174	0.6	Low salt
dT_{55}	256	5.5	High salt
polydT	137	-	High salt
polydT	44.3	3.9	Low salt

When the same assay was repeated at low salt conditions (20 mM NaCl), the binding rate constant for dT70 was 174 $\mu M^{-1} s^{-1}$.

Association kinetics with excess SSB. The association kinetics of DCC-PfSSB binding to ssDNA was also measured with high excess of DCC-PfSSB over ssDNA (Figure 4). At the high salt conditions the DCC-PfSSB binding to dT_{70} showed biphasic increases in fluorescence that fit well to a double exponential. The observed rate constant for the first phase, representing the primary binding process, increased hyperbolically with DCC-PfSSB concentration (Figure 5B). The linear dependence at low concentrations (Figure 5A) gives a second order rate constant of 354 $\mu M^{-1}s^{-1}$. This value is similar to that found by measurements with excess dT70. The slow phase was almost independent of concentration (Figure 5C), similar to that observed with excess DNA and may be a rearrangement of *Pf*SSB on dT_{70} rather than a direct binding.

The measurements were also done with dT_{35} (Figure 4) and these traces qualitatively similar to those with dT70. The rapid phase gave second order rate constant for dT_{35} binding is 266 $\mu M^{-1}s^{-1}$ (Figure 5A). However, there was an increase in the slow phase rate constant, albeit at much slower value than the fast. The data did not give a good linear fit.

Measurements with polydT showed biphasic traces with a rapid increase in fluorescence and followed by a slow small decreased. The traces were fitted with double exponentials and both phases showed concentration dependence with increasing DCC-PfSSB concentration. The second order rate constant for the first fast phase 197 $\mu M^{-1}s^{-1}$ (Figure 5A). The slow phase was almost independent of concentration, although as evidenced from the time courses (Figure 4), the amplitude increased at high SSB concentration.

The measurements were repeated at low ionic strength with dT_{70} and polyd and the rate constants are summarized in Table 4.

Table 4. Association kinetics parameters with excess DCC-PfSSB concentration over ssDNA.

Values are from the fits in Figure 5

		Fast phase		Slow phase	
ssDNA	Buffer	k_{on}	y-intercept	k_{on}	y-intercept

		($\mu M^{-1} s^{-1}$)	(s^{-1})	($\mu M^{-1} s^{-1}$)	(s^{-1})
dT35	High salt	266	4.02	n/a	n/a
dT70	High salt	350	6.60	4	0.68
polydT	High salt	200	1.91	5	0.21
dT70	Low salt	150	1.88	n/a	n/a
polydT	Low salt	100	1.62	8	0.11

The binding with excess DCC-PfSSB over ssDNA showed that the rate constant of binding were lower at low ionic strength buffer than at high ionic strength buffer.

PfSSB dissociation kinetics from ssDNA. DCC-PfSSB dissociation kinetics from ssDNA were measured done by rapidly mixing the DCC-PfSSB·ssDNA complex with a large excess of the unlabelled wtSSB to trap reased ssDNA. Unlabeled SSB has a tighter affinity to ssDNA than its labelled counterpart as was shown by titration assays (Figure 1C and D). If the PfSSB dissociation takes place as a single step, much slower than binding to the unlabelled trap, the dissociation kinetics should not be affected by the concentration of the wtSSB. As the dT70 binding and dissociation is seen as the simplest model, the dissociation experiments with dT70 are discussed first. The series of measurements at high ionic strength (200 mM NaCl) buffer with dT70 and varying concentrations of wtSSB showed time courses with a fluorescent decrease (Figure 6). These traces were fitted using double exponentials and both phases were dependent on the concentration of wtSSB. This indicated that the dissociation takes place in multiple steps. Such steps may include interactions of the wtSSB with the DCC-PfSSB and different binding modes before ssDNA has been fully trapped by unlabeled wtSSB.

Repeating the dissociation measurement with dT_{35}, dT_{70} and polydT at low ionic strength conditions also showed traces with decreasing in fluorescence (Figure 7). These were fitted to double exponentials and results are listed in Table 5.

Table 5. The second order rate constants determined from DCC-PfSSBC93, dependence on the wtSSB concentration.

Values are from the fits in Figure 8

		Fast phase		Slow phase	
ssDNA	Buffer	k_{obs}	y-intercept	k_{obs}	y-intercept

	conditions	($\mu M^{-1} s^{-1}$)	(s^{-1})	($\mu M^{-1} s^{-1}$)	(s^{-1})
dT_{35}	High salt	0.759	0.26	0.181	0.047
dT_{35}	Low salt	0.165	0.058	0.022	0.018
dT_{70}	High salt	0.271	0.405	0.158	0.115
dT_{70}	Low salt	0.156	0.13	n/a	n/a
polydT	High salt	1.763	3.12	0.097	0.088
polydT	Low salt	n/a	n/a	n/a	n/a

Traces with dT_{35} showed two phases that were both dependent on the wtSSB concentration. With dT_{70} only the fast phase rate constant showed dependence on the wtSSB concentration. Measurements with polydT fitted to double exponentials but the rate constants were not dependent on the wtSSB concentration and the average of rate constant observed for all measured wtSSB concentration was 1.3 s^{-1} for the fast phase and 0.12 s^{-1} for the slow phase. As the association and dissociation measurement enabled determination relatively accurate k_{on} and k_{off} values the equation $K_d = k_{off}/k_{on}$ can be used to determine the dissociation constant (K_d). This gives a value of 29.3 nM for DCC-PfSSB for polydT at low salt conditions.

Plasmid unwinding by PcrA. The activity and possible use of DCC-PfSSB was tested in two double-stranded DNA (dsdDNA) unwinding assays, mediated by helicases. In the first assay, PcrA helicase unwound circular plasmids, which was first nicked at its double stranded origin of replication by an initiator protein, RepD[4]. On nicking and removing the supercoiling, RepD covalently binds to the plasmid DNA and increases the processivity of PcrA helicase. In the assay, the unwinding complex of PcrA, RepD and plasmid was pre-formed, then mixed rapidly in stopped-flow apparatus with ATP in presence of a large excess of SSB. As the PcrA moves along DNA ssDNA is produced, to which the SSB binds. Previously this has been done using the fluorescent SSB from *E. coli*[4]. DCC-PfSBB was tested in the same assay (Figure 9). The fluorescence increase duration and amplitude were dependent on the length of the plasmid. The unwinding rate determined from fitting the unwinding duration against the plasmid lengths gives the unwinding rate of 17 bp s^{-1}. With *E. coli* SSB this was 27 bp s^{-1}. The total fluorescence change was 5.5-fold larger with DCC-PfSSB than with *E.coli* DCC-SSB (Figure 9A).

AddAB helicase assay. The DCC-PfSSB was also tested in another helicase assay. This assayed the $AddA^xB^x$ unwinding five different lengths of linear dsDNA. AddAB binds to ends of linear dsDNA having a linear overhang of a few bases. AddAB was pre-mixed with linear dsDNA before rapidly mixing the complex with ATP in presence of DCC-PfSSB. With all dsDNA lengths a linear increase in fluorescence was observed initially (Figure 10a). This phase was followed by phase with almost constant fluorescence. The duration of increasing fluorescence, as well as the amplitude, was linearly dependent on the length of the dsDNA (Figure 10B and C) and represents the AddAB unwinding dsDNA that leads to increasing amounts of ssDNA, which the DCC-PfSSB binds until all the dsDNA was unwound. Plotting the unwinding duration against linear dsDNA lengths the unwinding rate for AddAB was determined from reciprocal of the gradient of the fit, and this was 90 bp s^{-1}.

Overview of Example 3

We demonstrate that a fluorophore-PFSSB adduct has significantly improved properties over the previously reported DCC-SSB from *E. coli*[2, 14], in terms of its ability to measure ssDNA. This improvement might particularly be in the relative significance of multiple binding modes for the from *E. coli* protein, where the stoichiometry of the SSB:DNA complex is dependent both on solution conditions and ionic strength, but importantly, very dependent on the ratio of SSB to ssDNA in the medium and the absolute concerntrations[5, 6]. This could be viewed in terms of the following equation:

$$\mathrm{DNA70} \underset{1}{\overset{\mathrm{SSB}}{\rightleftharpoons}} \mathrm{SSB.DNA70} \underset{2}{\overset{\mathrm{SSB}}{\rightleftharpoons}} \mathrm{SSB_2.DNA70} \qquad (2)$$

DNA70 represents a 70-base stretch of ssDNA, such as dT70. At high salt, the equilibrium constant for step 2 (K_2) is low and so formation of SSB_2.DNA70, which represents the 35-base mode, is disfavoured thermodynamically. At low salt, K_2 is larger and, as [SSB] increases, SSB_2.DNA70 becomes more favoured.

This manifested itself with the *E. coli* DCC-SSB in several ways. Firstly, the titrations of DCC-SSB with ssDNA at low ionic strength were not linear (as in Figure 3 of Dillingham et al, 2008[2]). This was interpreted as follows. At low concentrations of ssDNA, the ratio of SSB:ssDNA is high and mainly 35-base binding was seen, which has a much lower fluorescence than the 65-70-base binding observed later in the

titration, when the ratio is low. Secondly, the binding measurements with excess SSB at low salt show a significant second phase decrease in fluorescence in the time courses (see Figure 2 of Kunzelmann et al, 2010[14]). This was interpreted as rapid binding in the 65-70-base mode (rapid step 1 of Equation 2) followed by subsequent rearrangement to some 35-base mode with its decreased fluorescence (slow step 2). A similar phenomenon was observed in unwinding assays with AddAB, where biphasic time courses, a fluorescence increase followed by slow decrease, occurred under some conditions[3].

With DCC-PfSSB, there is one immediate advantage in that the fluorescence enhancement on DNA binding at 20-fold (Figure 1A) is ~3-fold higher than DCC-SSB. This translates to a higher sensitivity, as seen with the plasmid unwinding assay (Figure 9A). Several pieces of data suggest that there is a lower tendency to 35-base binding, in other words K_2 is lower with DCC-PfSSB. The titrations at low ionic strength are all fairly linear, particularly at <50% saturation (Figure 1). That is the saturation region where most assays would be used. At least with dT70, there is no fluorescence decrease in a slow second phase of the binding experiments with excess protein (Figure 4), nor in the linear dsDNA unwinding assay (Figure 10A). However, the binding kinetics at high SSB concentration with polydT still show the second phase that may be due to 35-base binding. It seems likely that K_2 is significantly reduced with the Plasmodium protein some occurs only under extreme conditions. As for all fluorescence assays, a calibration is useful under the conditions being used. This would greatly increase the range of assay conditions amenable to simple treatment and without the multiple binding complications.

Example 4

An updated version of the table above is presented with further diethylaminocoumarin-labeled (IDCC) constructs according to the present invention:

PfSSB variant	Excitation wavelength (nm)	Emission wavelength (nm)	Fluorescence ratio (± dT_{35})	Fluorescence ratio (± dT_{70})	Fluorescence ratio (± polydT)
Wild-type	432	473	19.3	18.5	18.3
Y156R	432	472	10.3	12.3	11.5
Δ8	433	472	-	-	19.0
Δ35	433	473	24.3	21.1	25.4
Δ87	433	473	-	-	0.2
C93A G103C	433	473	2.7	3.5	-

C93A G102C	435	467	-	-	1.6
C93AY156RW166C	432	474	-	2.7	-

PfSSBC93AG102C:

MNEKSLNKIMLIGRVG**A**EPD IKILN**C**GDKVATFSLATNEF WRDRNTNELKSKTDWHRIVV
YDQNIVDLIDKYLRKGRRVY VQGSLHTRKWHTNDMNSQPK QITEIILSYNKGDLIFLDDK
RNFNQRNNSNNINSENQQHI NNEHINNNNNINNGNDFMPLN SNDKIIEDKEFTDRLDDNNE
ENNFQSNSETFDKQEGIYDK MNVQEFEE

PfSSBC93AG103C:

MNEKSLNKIMLIGRVG**A**EPD IKILNG**C**DKVATFSLATNEF WRDRNTNELKSKTDWHRIVV
YDQNIVDLIDKYLRKGRRVY VQGSLHTRKWHTNDMNSQPK QITEIILSYNKGDLIFLDDK
RNFNQRNNSNNINSENQQHI NNEHINNNNNINNGNDFMPLN SNDKIIEDKEFTDRLDDNNE
ENNFQSNSETFDKQEGIYDK MNVQEFEE

In one embodiment, W166C mutation may be made in the *Plasmodium* ssDNA binding protein. Suitably the W166C mutation is made in conjunction with a C93 mutation such as C93 A. In this embodiment, approximate 12-15 times signal may be achieved compared to the unbound state.

PfSSBC93AY156RW166C:

MNEKSLNKIMLIGRVG**A**EPD IKILNGGDKVATFSLATNEF WRDRNTNELKSKTDWHRIVV
YDQNIVDLIDKYLRKGRRV**R** VQGSLHTRK**C**HTNDMNSQPK QITEIILSYNKGDLIFLDDK
RNFNQRNNSNNINSENQQHI NNEHINNNNNINNGNDFMPLN SNDKIIEDKEFTDRLDDNNE
ENNFQSNSETFDKQEGIYDK MNVQEFEE

Most suitably polypeptides comprising dye attached at one or more of W166C, G102C and G103C are included for comparative purposes since the best sensors of the invention have label located on the (L1-1') loop of residues joining the two beta sheets (β1) and (β1').

References

1. Kunzelmann, S.; Solscheid, C.; Webb, M. R., Fluorescent biosensors: design and application to motility proteins. In *Fluorescent Methods Applied to Molecular Motors, Experientia Supplementum* Toseland, C. P.; Fili, N., Eds. Springer: Basel, 2014; Vol. 105, pp 25-47.
2. Dillingham, M. S.; Tibbles, K. L.; Hunter, J. L.; Bell, J. C.; Kowalczykowski, S. C.; Webb, M. R., Fluorescent single-stranded DNA binding protein as a probe for sensitive, real time assays of helicase activity. *Biophys. J.* **2008,** 95, 3330-3339.
3. Yeeles, J. T. P.; Gwynn, E. J.; Webb, M. R.; Dillingham, M. S., The AddAB helicase-nuclease catalyses rapid and processive DNA unwinding using a single Superfamily 1A motor domain. *Nucleic Acids Res.* **2011,** 39, 2271-2285.
4. Slatter, A. F.; Thomas, C. D.; Webb, M. R., PcrA helicase tightly couples ATP hydrolysis to unwinding double-stranded DNA, modulated by the replication initiator protein, RepD. *Biochemistry* **2009,** 48, 6326-6334.
5. Bujalowski, W.; Lohman, T. M., *Escherichia coli* single-strand binding protein forms multiple, distinct complexes with single-stranded DNA. *Biochemistry* **1986,** 25, (24), 7799-7802.
6. Lohman, T. M.; Ferrari, M. E., *Escherichia Coli* single-stranded DNA-binding protein: multiple DNA-binding modes and cooperativities. *Annu. Rev. Biochem.* **1994,** 63, (1), 527-570.

7. Antony, E.; Kozlov, A. G.; Nguyen, B.; Lohman, T. M., Plasmodium falciparum SSB Tetramer Binds Single-Stranded DNA Only in a Fully Wrapped Mode. *J. Mol. Biol.* **2012,** 420, (4–5), 284-295.

8. Kozlov, A. G.; Weiland, E.; Mittal, A.; Waldman, V.; Antony, E.; Fazio, N.; Pappu, R. V.; Lohman, T. M., Intrinsically Disordered C-Terminal Tails of E. coli Single-Stranded DNA Binding Protein Regulate Cooperative Binding to Single-Stranded DNA. *J. Mol. Biol.* **2015,** 427, (4), 763-774.

9. Antony, E.; Weiland, E. A.; Korolev, S.; Lohman, T. M., Plasmodium falciparum SSB Tetramer Wraps Single-Stranded DNA with Similar Topology but Opposite Polarity to E. coli SSB. *J. Mol. Biol.* **2012,** 420, (4–5), 269-283.

10. Raghunathan, S.; Kozlov, A. G.; Lohman, T. M.; Waksman, G., Structure of the DNA binding domain of *E. coli* SSB bound to ssDNA. *Nature Struct. Biol.* **2000,** 7, (8), 648-652.

11. Raghunathan, S.; Ricard, C. S.; Lohman, T. M.; Waksman, G., Crystal structure of the homo-tetrameric DNA binding domain of *Escherichia coli* single-stranded DNA-binding protein determined by multiwavelength x-ray diffraction on the selenomethionyl protein at 2.9-A resolution. *Proc. Natl. Acad. Sci. U.S.A.* **1997,** 94, (13), 6652-6657.

12. Leatherbarrow, R. J., *GraFit Version 7*. Erithacus Software Ltd: Horley, U.K., 2009.

13. Fili, N.; Mashanov, G.; Toseland, C. P.; Batters, C.; Wallace, M. I.; Yeeles, J. T. P.; Dillingham, M. S.; Webb, M. R.; Molloy, J. E., Visualizing DNA unwinding by helicases at the single molecule level. *Nucleic Acids Res.* **2010,** 38, 4448-4457.

14. Kunzelmann, S.; Morris, C.; Chavda, A. P.; Eccleston, J. F.; Webb, M. R., Mechanism of interaction between single-stranded DNA binding protein and DNA. *Biochemistry* **2010,** 49, 843-852.

Although illustrative embodiments of the invention have been disclosed in detail herein, with reference to the accompanying drawings, it is understood that the invention is not limited to the precise embodiment and that various changes and modifications can be effected therein by one skilled in the art without departing from the scope of the invention as defined by the appended claims and their equivalents.

CLAIMS

1. A protein of the single stranded DNA binding domain (SSB) family, wherein said protein is a modified *Plasmodium* protein, said protein comprising at least one detectable label attached to an amino acid of said protein, wherein said amino acid is located on the (L1-1') loop of residues joining the two beta sheets (β1) and (β1'), wherein the characteristics of the detectable label alter on binding single stranded DNA.

2. A protein according to claim 1 wherein the detectable label is attached to a region of the protein surface.

3. A protein according to claim 1 or claim 2 wherein the detectable label is attached via a cysteine residue in the protein.

4. A protein according to claim 3 wherein the cysteine residue is a naturally occurring cysteine residue.

5. A protein according to claim 4 wherein the cysteine residue is, or corresponds to, C93 (Cys 93) of SEQ ID NO: 1.

6. A protein according to claim 3, wherein said cysteine residue is a cysteine residue engineered into the protein at a position which is, or corresponds to, one selected from the group consisting of: G92, E94 and K151 of SEQ ID NO: 1.

7. A protein according to any of claims 1 to 6, wherein said modified *Plasmodium* protein is a modified *Plasmodium falciparum* protein.

8. A protein according to any of claims 1 to 7, said protein comprising amino acid sequence corresponding to at least amino acids 78 to 198 of SEQ ID NO:1,
said amino acid sequence having at least 50% sequence identity to amino acids 78 to 198 of SEQ ID NO:1.

9. A protein according to claim 8, said protein comprising amino acid sequence corresponding to at least amino acids 78 to 250 of SEQ ID NO:1,
said amino acid sequence having at least 50% sequence identity to amino acids 78 to 250 of SEQ ID NO:1.

10. A protein according to claim 8, said protein comprising amino acid sequence corresponding to at least amino acids 78 to 277 of SEQ ID NO:1,
said amino acid sequence having at least 50% sequence identity to amino acids 78 to 277 of SEQ ID NO:1.

11. A protein according to claim 8, said protein comprising amino acid sequence corresponding to at least amino acids 78 to 285 of SEQ ID NO:1,
said amino acid sequence having at least 50% sequence identity to amino acids 78 to 285 of SEQ ID NO:1.

12. A protein according to claim 8, the polypeptide part of said protein consisting of the amino acid sequence of SEQ ID NO:1.

13. A protein according to any preceding claim wherein the detectable label is a fluorescent label.

14. The protein of claim 13 wherein the detectable label is a coumarin.

15. A protein according to claim 14 wherein the label is selected from the group consisting of *N*-[2-(1-maleimidyl)ethyl]-7-diethylaminocoumarin-3-carboxamide and *N*-[2-(iodoacetamido)ethyl]-7-diethylaminocoumarin-3-carboxamide (IDCC).

16. A protein according to claim 15 wherein the label is IDCC.

17. A protein according to any preceding claim which further comprises a mutation compared to SEQ ID NO: 1 at a position selected from the group consisting of K84, Y156, K187 and D189, preferably Y156.

18. A protein according to claim 17 wherein said mutation is selected from the group consisting of K84D, Y156R, K187D and D189K, preferably Y156R.

19. A method for detecting single stranded DNA in a sample comprising the steps of:
(i) mixing the sample with the protein of any one of claims 1-18 and
(ii) detecting a change in the mixture arising from the interaction between the single stranded DNA and the protein.

20. A method for monitoring changes in ssDNA concentration in a sample comprising contacting said sample with a protein according to any of claims 1 to 18 and determining changes in the characteristics of the detectable label, wherein changes in the characteristics of the detectable label indicate changes in the concentration of ssDNA in said sample.

21. A method according to claim 20 wherein the characteristics of the detectable label are monitored by measurement of changes in fluorescence of a fluorophore comprised by said protein.

22. A method of screening for inhibitors of DNA processing enzymes which comprises assaying single stranded DNA levels *in vitro* using a protein according to any one of claims 1-18 in the presence and absence of the inhibitors and assaying for an alteration in the single stranded DNA levels.

23. Use of a protein according to any of claims 1 to 18 in the determination of ssDNA concentration in a sample.

24. A nucleic acid having a nucleotide sequence encoding the polypeptide portion of the protein according to any of claims 1 to 18.

25. A ssDNA binding protein or method substantially as disclosed herein.

FIGURE 1

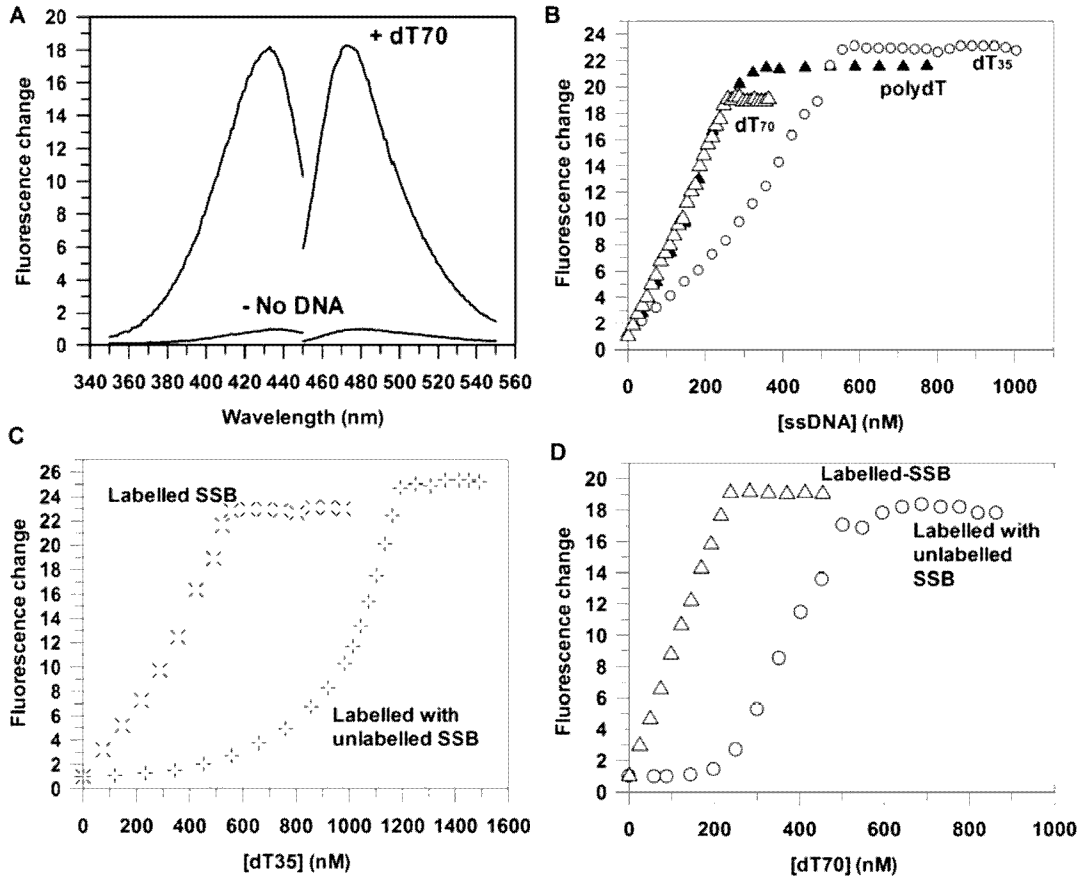

A
+ dT70
- No DNA
Fluorescence change
Wavelength (nm)
B
dT35
polydT
dT70
Fluorescence change
[ssDNA] (nM)
C
Labelled SSB
Labelled with unlabelled SSB
Fluorescence change
[dT35] (nM)
D
Labelled-SSB
Labelled with unlabelled SSB
Fluorescence change
[dT70] (nM)

FIGURE 2

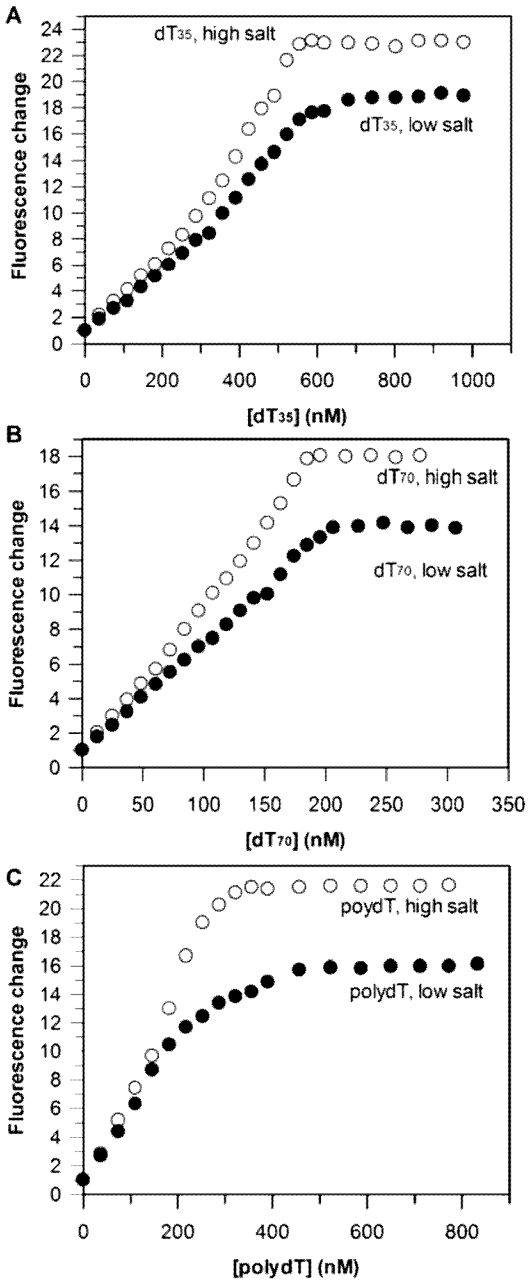
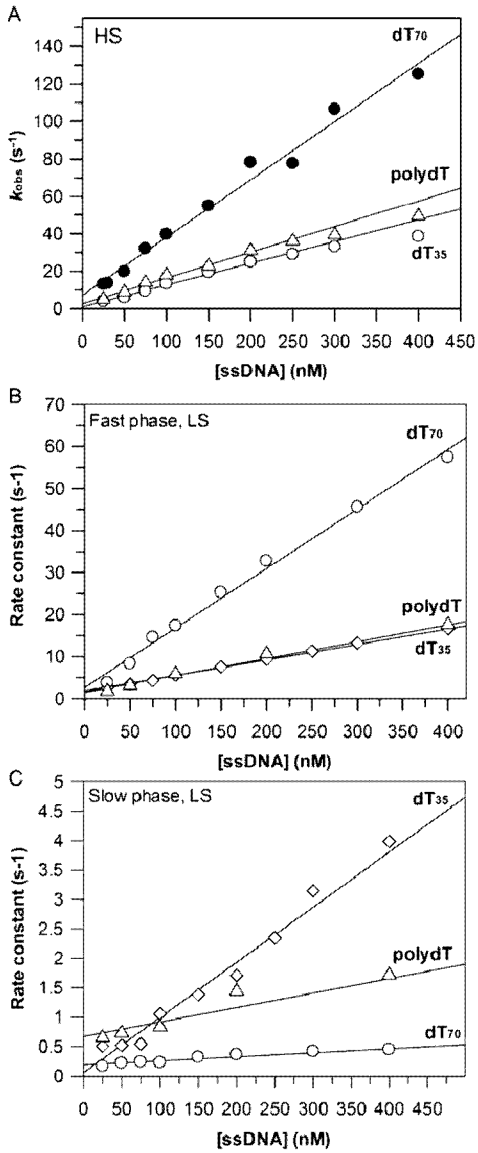
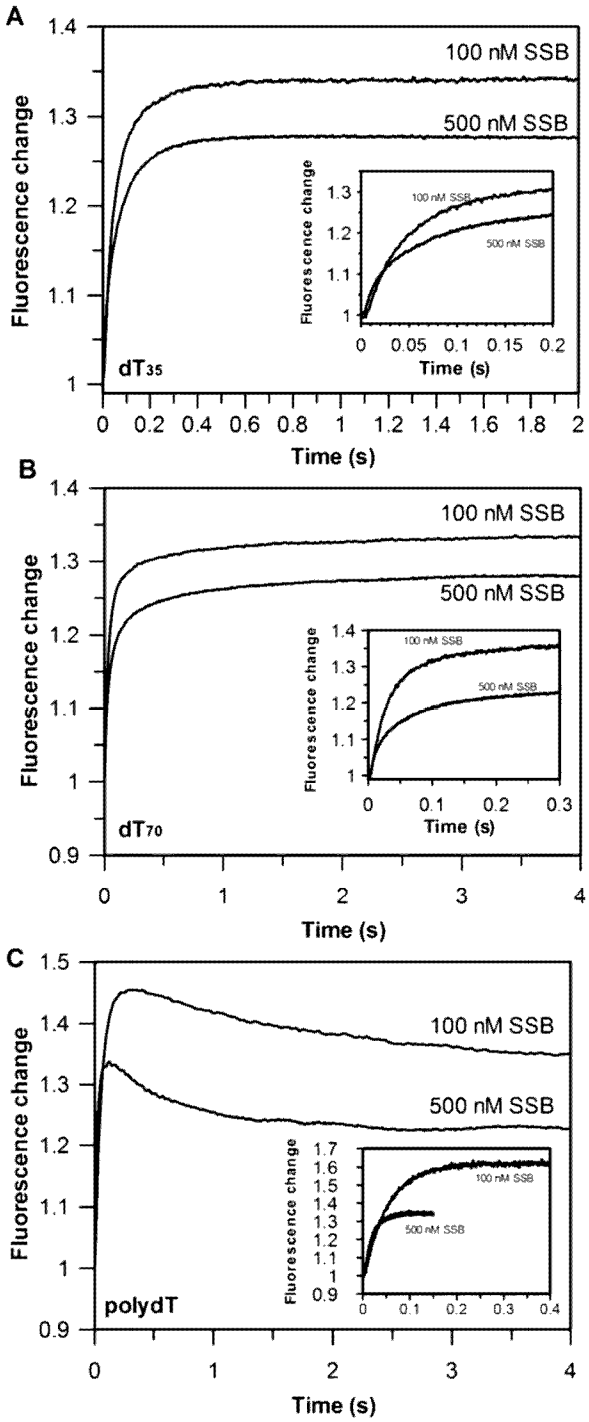


FIGURE 3



10

FIGURE 4


A
100 nM SSB
500 nM SSB
Fluorescence change
dT35
Time (s)
B
100 nM SSB
500 nM SSB
Fluorescence change
dT70
Time (s)
C
100 nM SSB
500 nM SSB
Fluorescence change
polydT
Time (s)

.

FIGURE 5

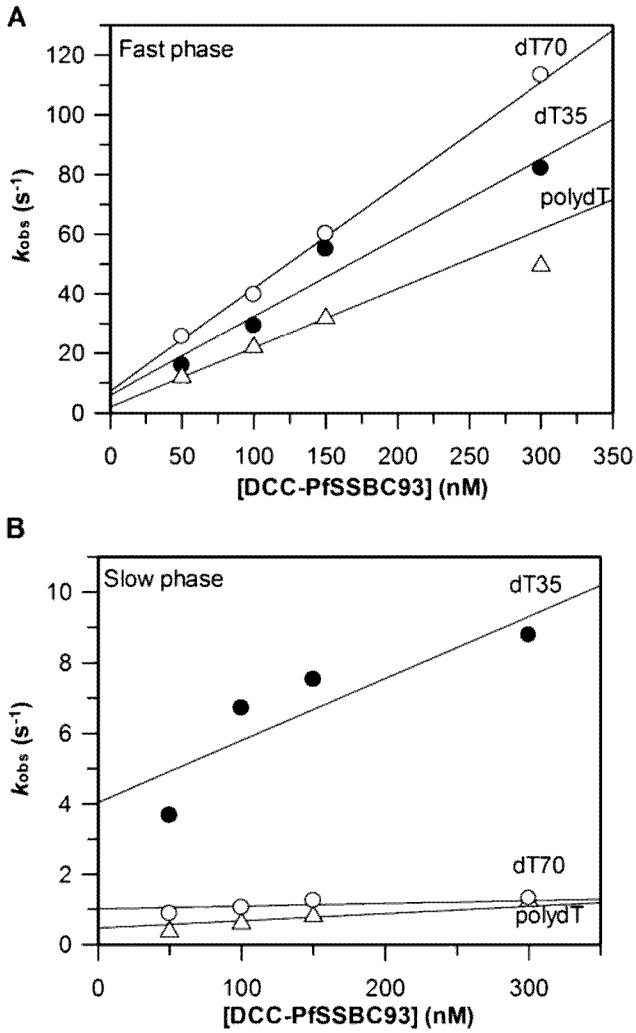

A
Fast phase
dT70
dT35
polydT
k_{obs} (s^{-1})
[DCC-PfSSBC93] (nM)
B
Slow phase
dT35
dT70
polydT
k_{obs} (s^{-1})
[DCC-PfSSBC93] (nM)

FIGURE 6

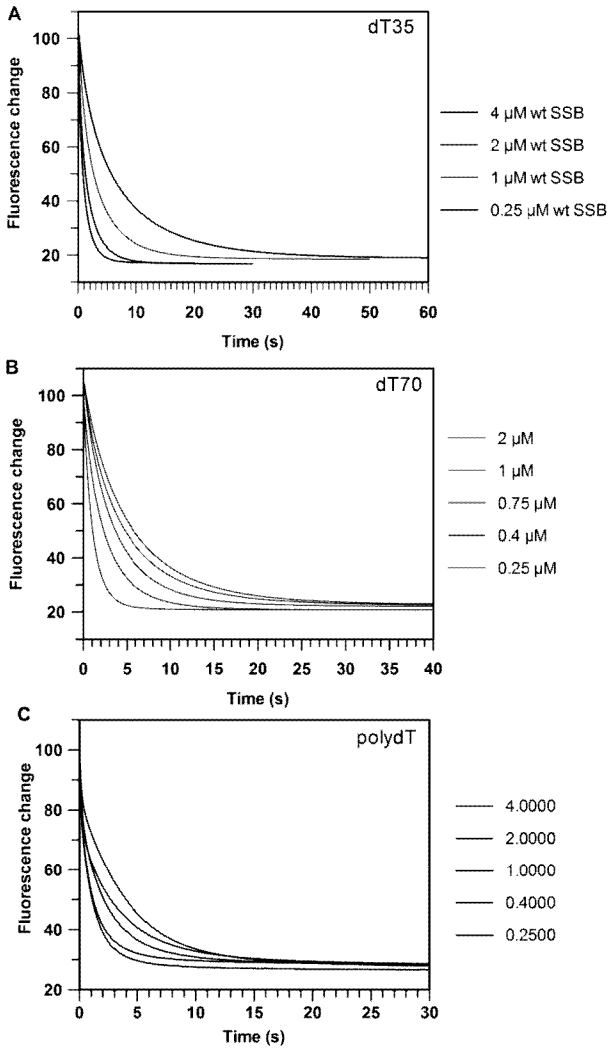
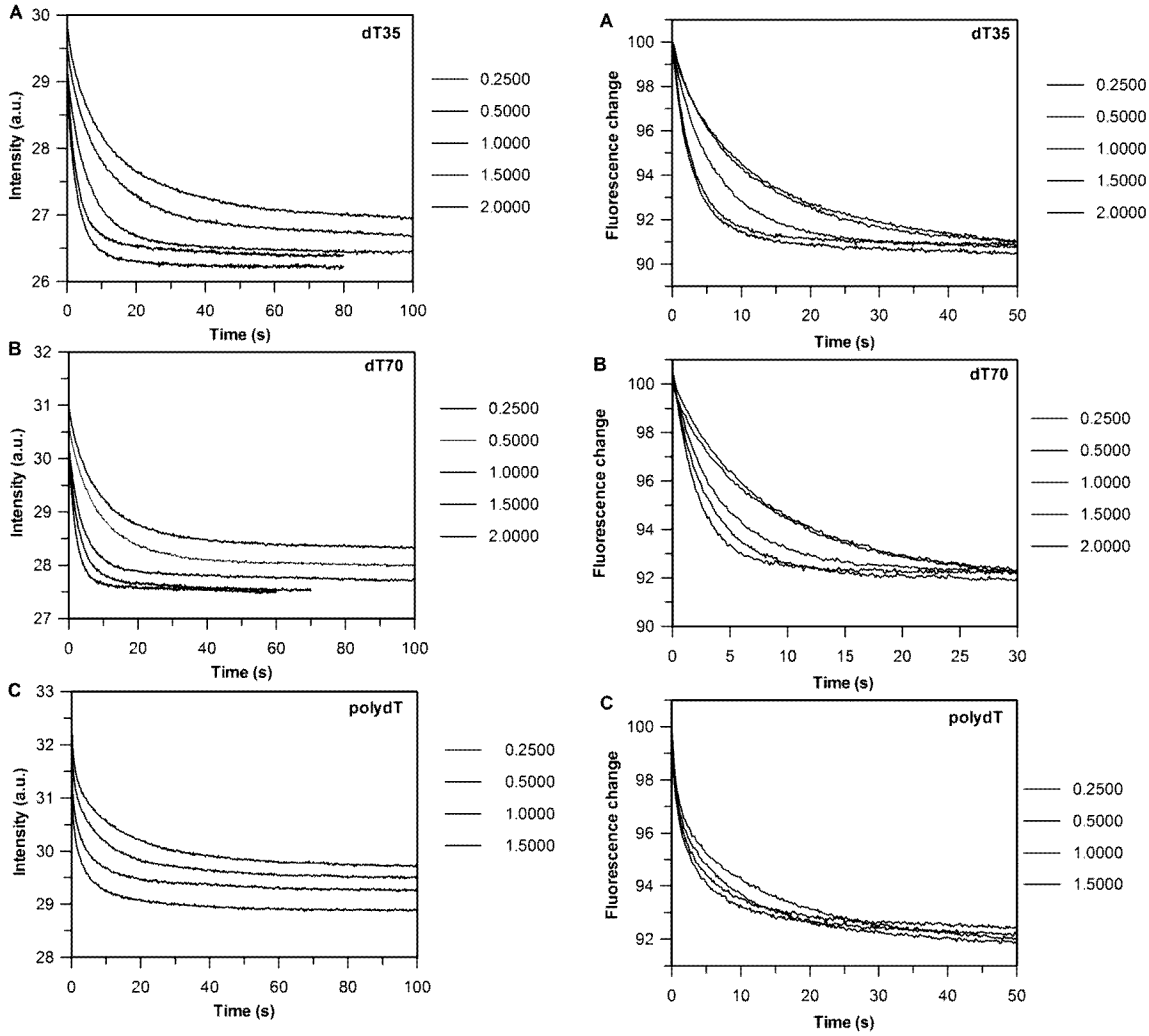

A
dT35
Fluorescence change
Time (s)
4 μM wt SSB
2 μM wt SSB
1 μM wt SSB
0.25 μM wt SSB
B
dT70
Fluorescence change
Time (s)
2 μM
1 μM
0.75 μM
0.4 μM
0.25 μM
C
polydT
Fluorescence change
Time (s)
4.0000
2.0000
1.0000
0.4000
0.2500

FIGURE 7


A
dT35
Intensity (a.u.)
Time (s)
0.2500
0.5000
1.0000
1.5000
2.0000
B
dT70
Intensity (a.u.)
Time (s)
0.2500
0.5000
1.0000
1.5000
2.0000
C
polydT
Intensity (a.u.)
Time (s)
0.2500
0.5000
1.0000
1.5000
A
dT35
Fluorescence change
Time (s)
0.2500
0.5000
1.0000
1.5000
2.0000
B
dT70
Fluorescence change
Time (s)
0.2500
0.5000
1.0000
1.5000
2.0000
C
polydT
Fluorescence change
Time (s)
0.2500
0.5000
1.0000
1.5000

5

FIGURE 8

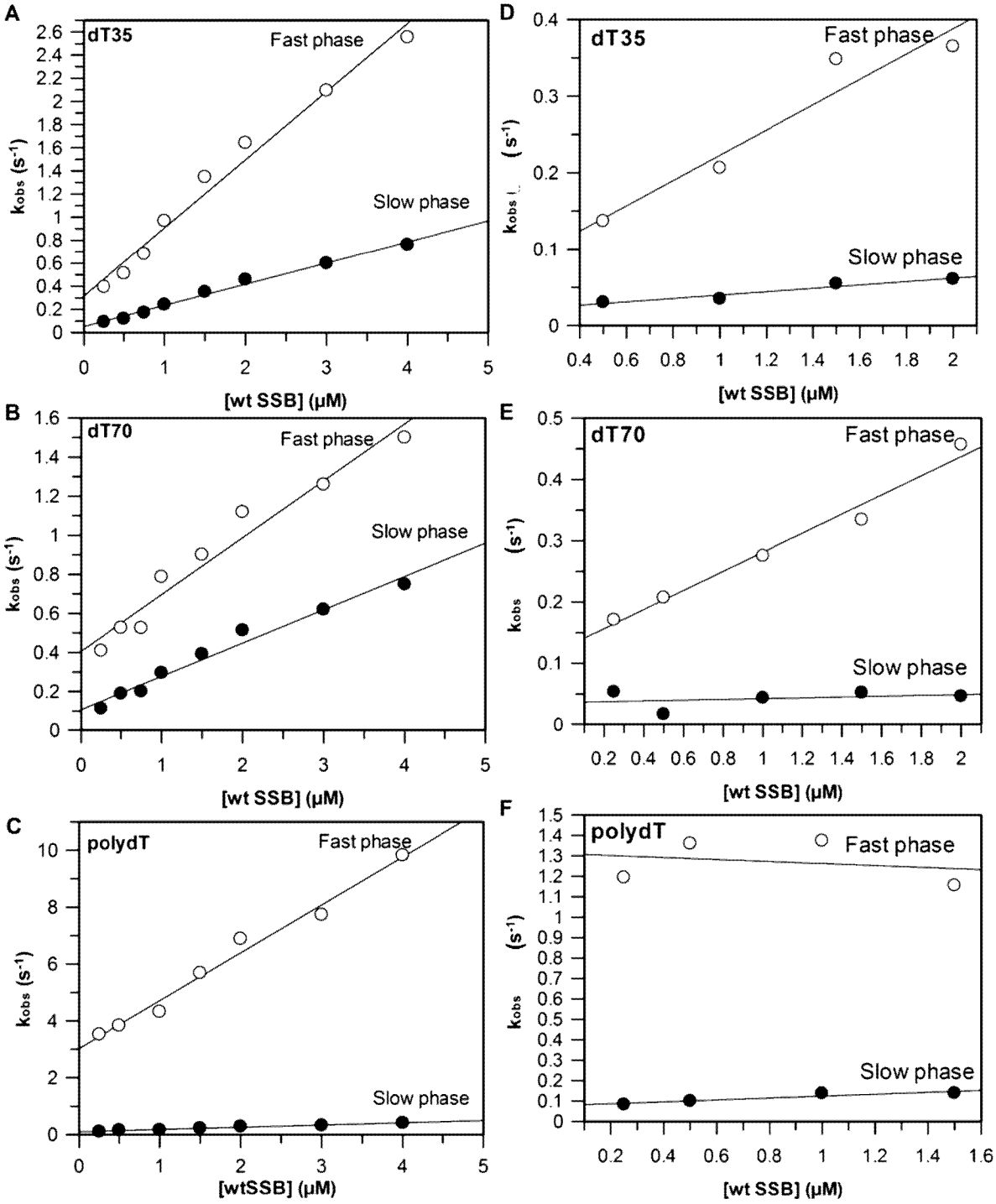
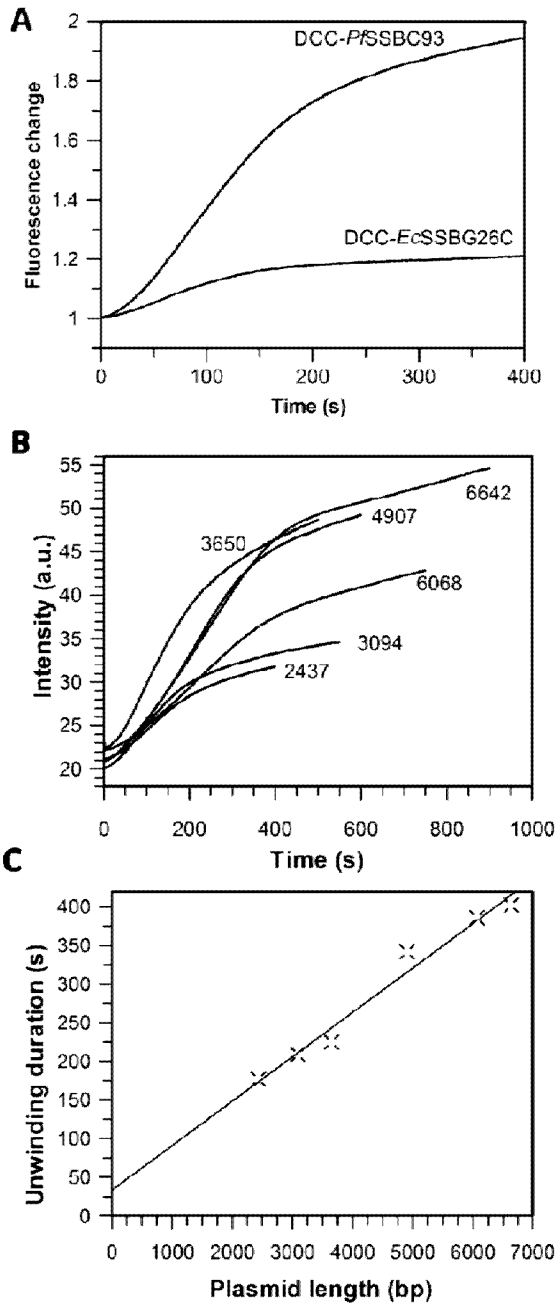
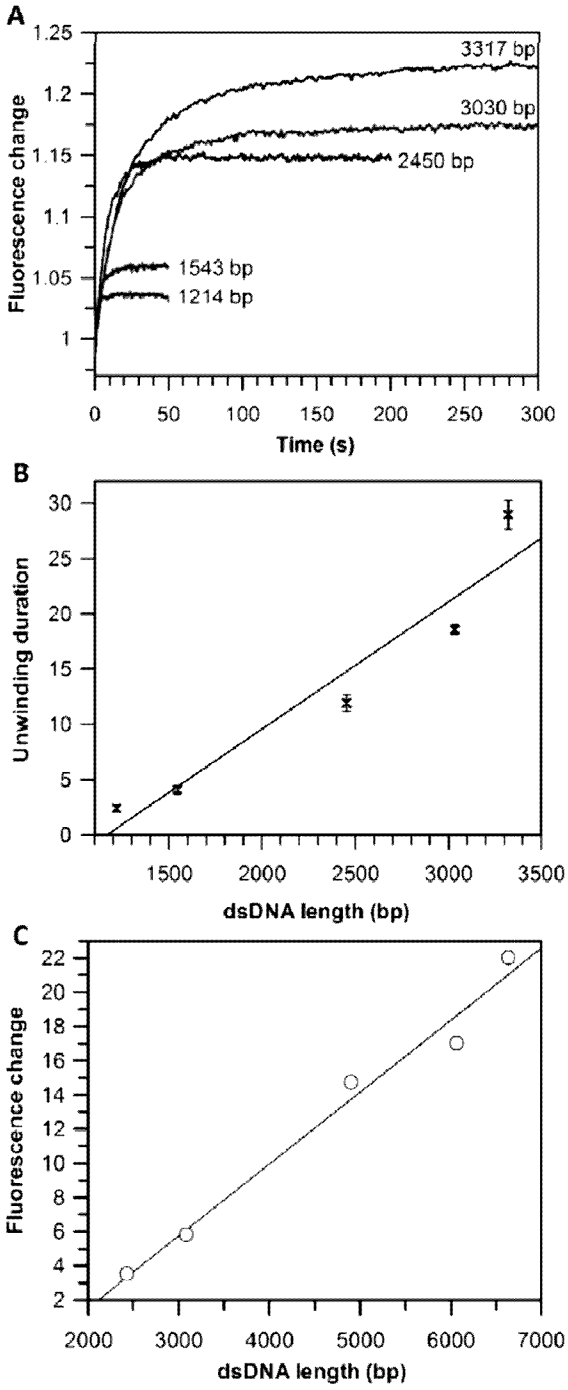

A
dT35
Fast phase
Slow phase
kobs (s-1)
[wt SSB] (μM)
D
dT35
Fast phase
Slow phase
[wt SSB] (μM)
B
dT70
Fast phase
Slow phase
kobs (s-1)
[wt SSB] (μM)
E
dT70
Fast phase
Slow phase
kobs (s-1)
[wt SSB] (μM)
C
polydT
Fast phase
Slow phase
kobs (s-1)
[wtSSB] (μM)
F
polydT
Fast phase
Slow phase
kobs (s-1)
[wt SSB] (μM)

FIGURE 9


A
DCC-PfSSBC93
DCC-EcSSBG26C
Fluorescence change
Time (s)
B
6642
4907
3650
6068
3094
2437
Intensity (a.u.)
Time (s)
C
Unwinding duration (s)
Plasmid length (bp)

5

FIGURE 10


A
Fluorescence change
Time (s)
3317 bp
3030 bp
2450 bp
1543 bp
1214 bp
B
Unwinding duration
dsDNA length (bp)
C
Fluorescence change
dsDNA length (bp)

5

FIGURE 11

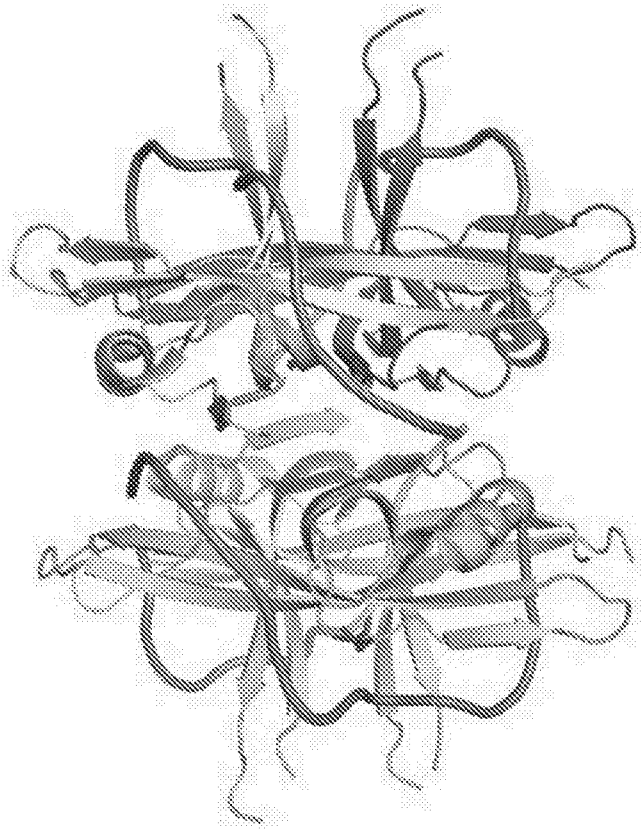


FIGURE 12

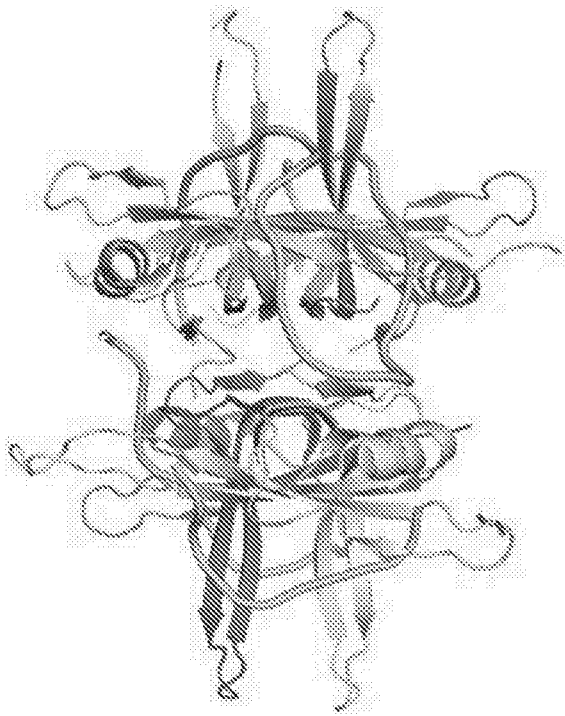


FIGURE 13

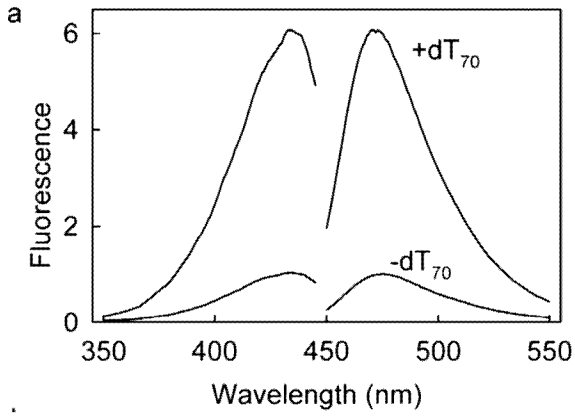


FIGURE 14

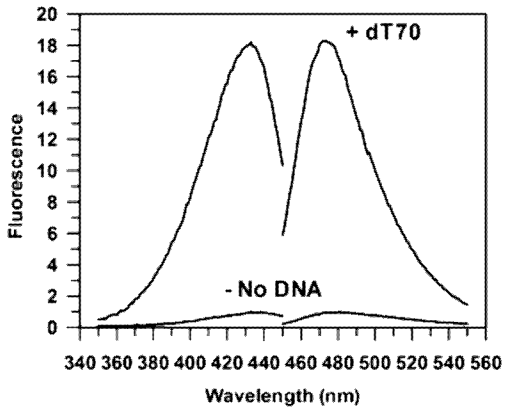


FIGURE 15

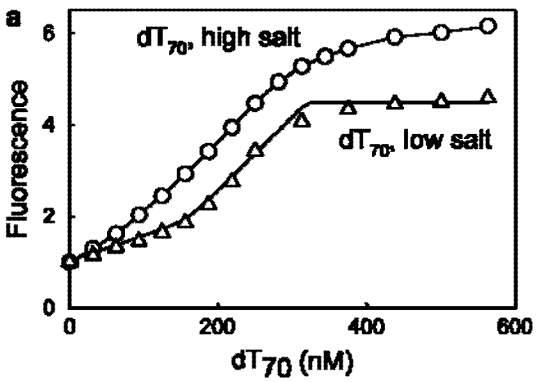


FIGURE 16

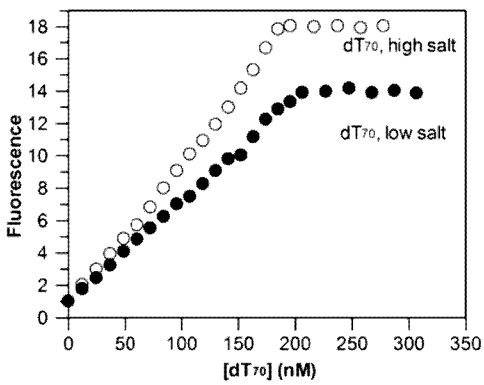


FIGURE 17

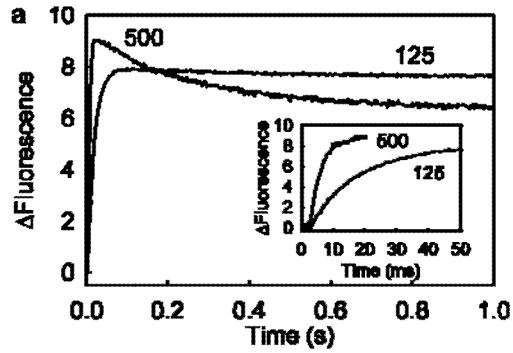


FIGURE 18

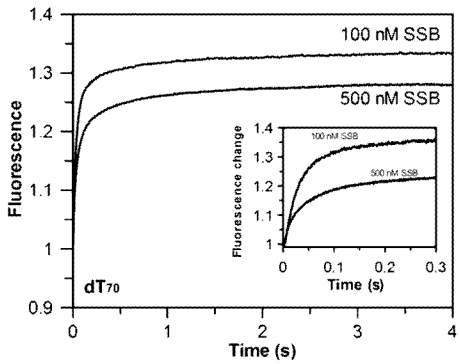


FIGURE 19

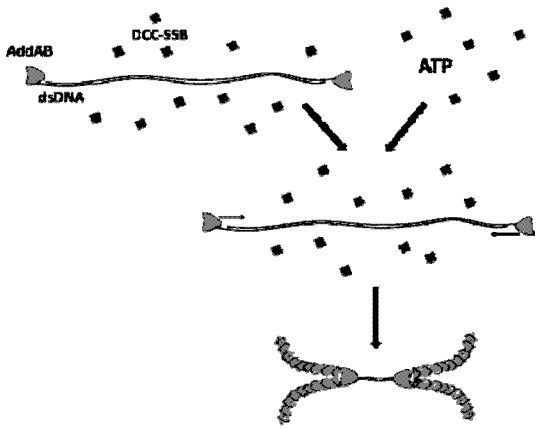


FIGURE 20

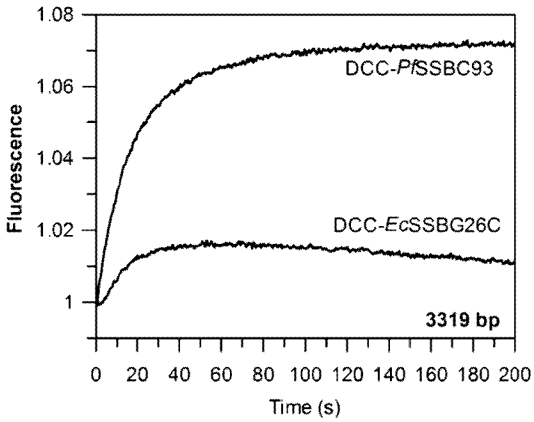


FIGURE 21

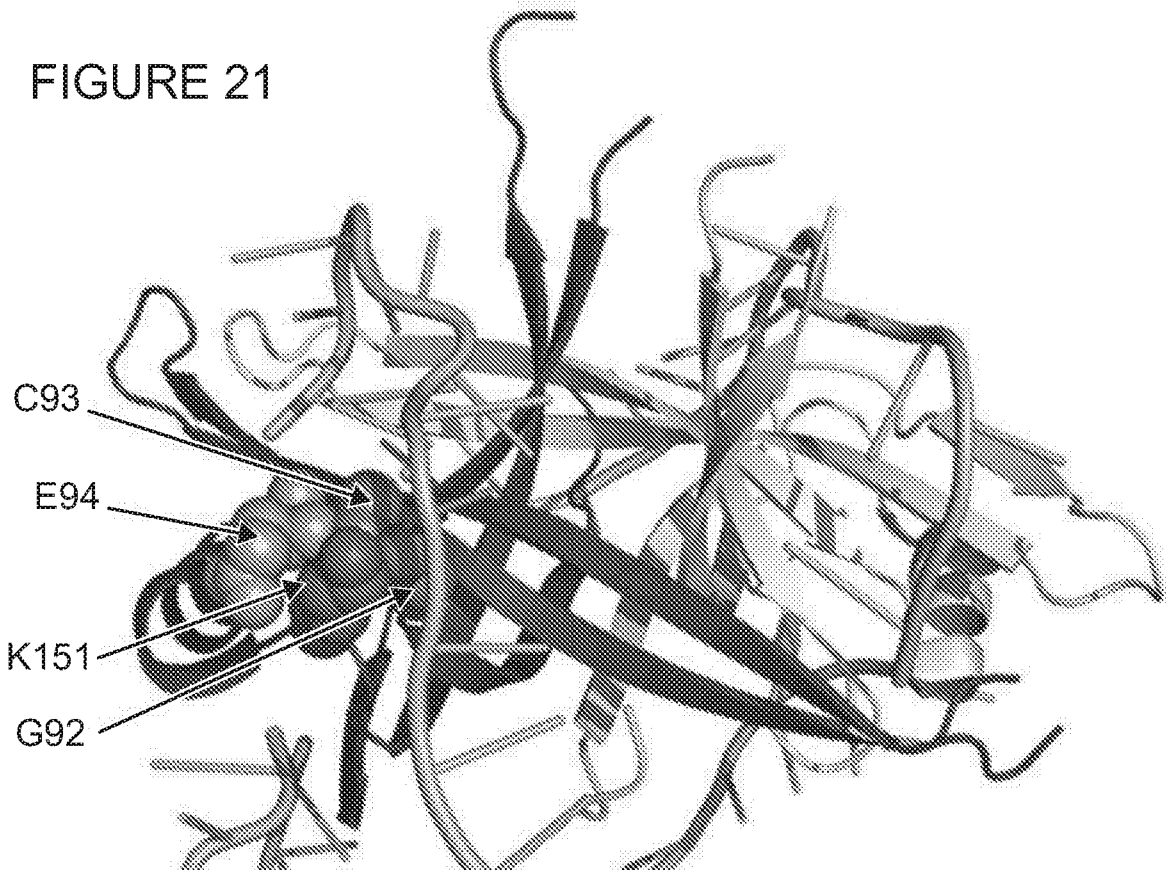
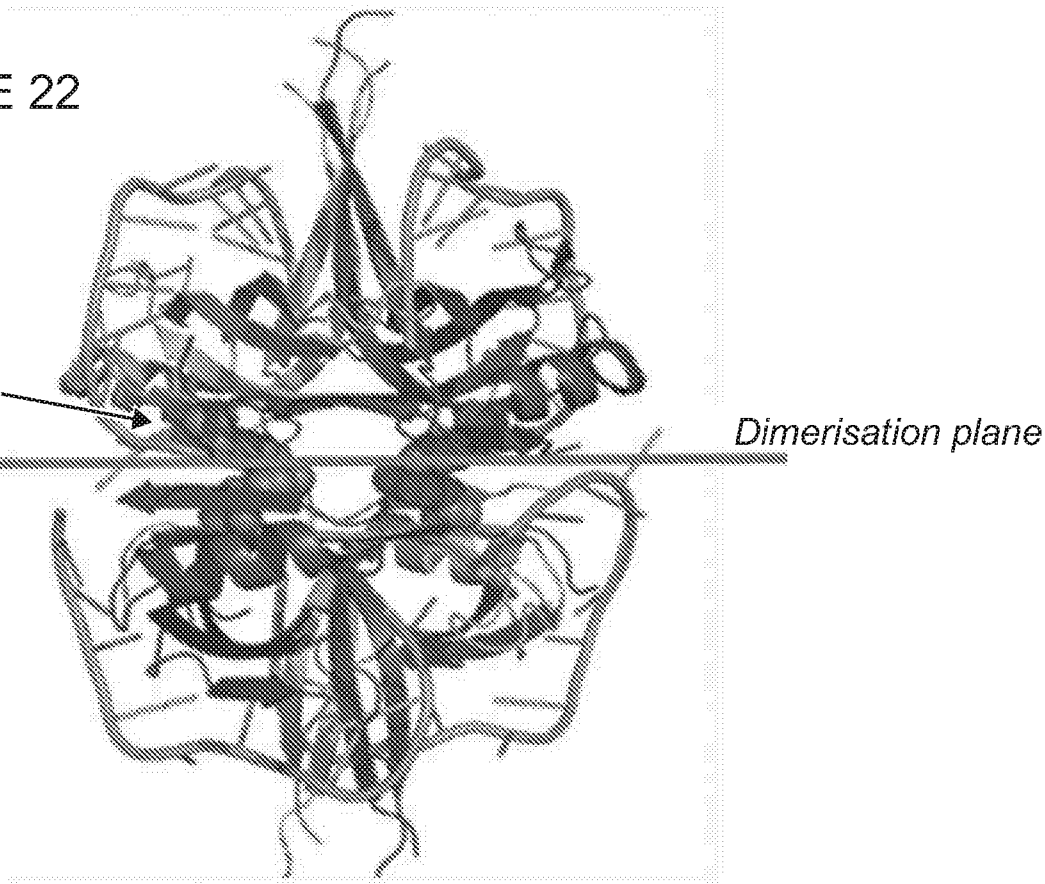


FIGURE 22



PfSSB tetramer

INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2016/052124

A. CLASSIFICATION OF SUBJECT MATTER
INV. C07K14/00 G01N33/53 G01N33/533
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C07K G01N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, BIOSIS, Sequence Search, EMBASE, INSPEC, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2008/152379 A1 (MEDICAL RES COUNCIL [GB]; DILLINGHAM MARK [GB]; KOWALCZYKOWSKI STEPHEN) 18 December 2008 (2008-12-18) cited in the application	25
Y	the whole document	1-25

	-/--	

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
18 November 2016	06/12/2016

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Ury, Alain

INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2016/052124

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	EDWIN ANTONY ET AL: "Plasmodium falciparum SSB Tetramer Wraps Single-Stranded DNA with Similar Topology but Opposite Polarity to E. coli SSB", JOURNAL OF MOLECULAR BIOLOGY, vol. 420, no. 4-5, 1 July 2012 (2012-07-01), pages 269-283, XP055320510, United Kingdom ISSN: 0022-2836, DOI: 10.1016/j.jmb.2012.04.021 cited in the application see whole document and in particular page 271 and Fig.3	1-25
A	ANTONY EDWIN ET AL: "Plasmodium falciparum SSB Tetramer Binds Single-Stranded DNA Only in a Fully Wrapped Mode", JOURNAL OF MOLECULAR BIOLOGY, vol. 420, no. 4-5, July 2012 (2012-07), pages 284-295, XP002764378, ISSN: 0022-2836 cited in the application the whole document	1-25
A	KOZLOV ALEXANDER G ET AL: "Intrinsically Disordered C-Terminal Tails ofE.coliSingle-Stranded DNA Binding Protein Regulate Cooperative Binding to Single-Stranded DNA", JOURNAL OF MOLECULAR BIOLOGY, vol. 427, no. 4, 3 January 2015 (2015-01-03), pages 763-774, XP029136785, ISSN: 0022-2836, DOI: 10.1016/J.JMB.2014.12.020 cited in the application the whole document	1-25
A	US 5 747 247 A (KOWALCZYKOWSKI STEPHEN C [US] ET AL) 5 May 1998 (1998-05-05) the whole document	1-25
A	KUNZELMANN SIMONE ET AL.: "Fluorescent biosensors: design and application to motor proteins", EXS, BIRKHAEUSER VERLAG, BASEL, CH, vol. 105, 1 January 2014 (2014-01-01), pages 25-47, XP009190657, ISSN: 1023-294X, DOI: 10.1007/978-3-0348-0856-9_2 cited in the application page 31	1-25

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/GB2016/052124

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2008152379 A1	18-12-2008	DK 2170937 T3	10-06-2014
		EP 2170937 A1	07-04-2010
		ES 2473279 T3	04-07-2014
		US 2011177501 A1	21-07-2011
		WO 2008152379 A1	18-12-2008
US 5747247 A	05-05-1998	NONE	