

## TARGETING OF CHONDROITIN SULFATE GLYCANS

## FIELD OF THE INVENTION

The present invention relates to functional binding fragments comprising the minimal binding fragments of VAR2CSA, to antibodies against such binding fragments of VAR2CSA, nucleic acids encoding such fragments of VAR2CSA as well as methods for their production. The invention further relates to conjugates and fusion proteins of VAR2CSA polypeptides including the minimal binding fragments and their use, in particular in the treatment of conditions associated with expression of chondroitin sulfate A (CSA), such as an inappropriate expression of chondroitin sulfate A (CSA).

## 10 BACKGROUND OF THE INVENTION

Proteoglycans are proteins conjugated to one or more glycosaminoglycan (GAG) chains. These proteins are distributed inside cells, on the cell membrane and in the extracellular matrix serving a variety of functions: cartilage matrix formation; the structural organization of tissues; organizations of basement membranes; regulating the role of secretory vesicles; binding of cytokines, chemokines, growth factors, and morphogens; protease receptors and protease inhibitors; co-receptors, tyrosine-kinase-type growth factor receptors; as endocytic receptors; facilitate cell attachment, cell-cell interactions, and cell motility as well as cell migration.

The malaria parasite *Plasmodium falciparum* utilizes host cell proteoglycans in almost all stages of its complex life cycle. The sporozoite infects hepatocytes in the liver through surface-expressed circumsporozoite protein interacting with highly sulfated heparan sulfate proteoglycans (HSPG). Merozoite infection of the erythrocytes is mediated by EBA-175 binding to sialic acid on glycophorin A. In addition, a number of *Plasmodium falciparum* Erythrocyte Membrane Protein 1 (PfEMP1) proteins, mediating host endothelial adhesion, have been described as glycan-binding. One of these is VAR2CSA, which is a unique member of the PfEMP1 protein family. VAR2CSA binds with high affinity to an unusual, low-sulfated form of chondroitin sulfate A (CSA), attached to proteoglycans, so called Chondroitin Sulfate Proteoglycan (CSPG), in the intervillous spaces of the placenta. VAR2CSA is a large multi-domain protein (350 kDa) expressed on the surface of *P. falciparum*-infected erythrocytes (IEs), and the VAR2CSA-CSA interaction is responsible for placenta specific sequestration in

placental malaria (PM). Importantly, recombinant full-length VAR2CSA ecto-domain from FCR3 and 3D7 type parasites has shown affinity for CSA in the low nano-molar range.

CSA belongs to the family of glycosaminoglycans (GAGs), which are linear polymers of alternating amino sugars and hexuronic acid residues, attached to proteoglycans. There are several types of GAGs including; chondroitin sulfate (CS), dermatan sulfate (DS or CSB), heparan sulfate (HS) and heparin. While the polysaccharide backbone of these GAGs is simple, considerable diversity arises in modifications such as sulfation and uronate epimerization. This is the basis for the wide variety in domain structure and biological activities of different GAGs.

CS interacts with many important factors such as growth hormones, cytokines, chemokines, and adhesion molecules and is thought to be involved in structural stabilization, cytokinesis, cell proliferation, differentiation, cell migration, tissue morphogenesis, organogenesis, infection, and wound repair. CS chains are composed of alternating units of N-acetyl-D-galactosamine (GalNAc) and glucuronic acid residues. Glucuronic acid can be sulfated at its C2 position and GalNAc can be sulfated at C4 and/or C6, giving rise to various disaccharide units. Varying modifications of the sugar backbone allows structural and functional heterogeneity of the CS chains. Placenta adhering *P. falciparum* parasites specifically associate with low sulfated CSA with sulfation only at C4 of GalNAc.

Early studies pinpointed CSA as being responsible for IE sequestration in the placenta. The specific receptor was however not known. Upon further research it was found that the human placenta contained three distinct types of chondroitin sulfate proteoglycans (CSPG), but that the IE adhered specifically to low sulfated CSPG in the intervillous spaces. What is special for this type of CSPG is that only 2-8% of the disaccharide units are C4 sulfated. In an accompanying study, aimed to identify the specific structural requirements for the CSA, it was found that parasite adhesion to CSPG is inhibited by CSA containing between 30-50% C4 sulfation, with the remaining 50-70% disaccharide units being unsulfated. The minimal inhibition of binding requirements for CSA was shown to be a dodecasaccharide (six disaccharides) containing a minimum of 2-3 or 4-5 C4 sulfated disaccharide units.

Chondroitin sulfate proteoglycan 4 (CSPG4), also known as High Molecular Weight-Melanoma Associated Antigen (HMW-MAA) or melanoma-associated chondroitin sulfate proteoglycan (MSCP), is a cell surface proteoglycan which has been shown to be expressed by melanoma cells.

CSPG4/MSCP/HMV-MAA is a large proteoglycan characterized by having CS chains on the protein backbone. The sulfation of these CS chains seems to be primarily on C4 of GalNAc (CSA), although the degree of sulfation is not known.

#### OBJECT OF THE INVENTION

- 5 It is an object of embodiments of the invention to provide minimal functional binding fragments of VAR2CSA suitable for the targeting and/or detection of chondroitin sulfate glycans. Other objects of embodiments of the invention is to provide methods for treating conditions associated with expression, such as inappropriate expression of the chondroitin sulfate glycans, such as CSA, wherein VAR2CSA polypeptides or fragments thereof, either  
10 alone or as part of conjugates or fusion proteins are used to target and/or detect tissue or cells having an expression, such as inappropriate expression of the chondroitin sulfate glycans.

#### SUMMARY OF THE INVENTION

- 15 It has been found by the present inventors that VAR2CSA retains its ability to bind with high affinity and specificity to certain chondroitin sulfate proteoglycans with minimal structural elements of the polypeptide sequence. More importantly, the present inventors have found that VAR2CSA polypeptides bind with high and specific affinity to cancer cells and tissues, which binding by the present inventors is suggested to be through a specific interaction with  
20 chondroitin sulfate proteoglycans expressed on the surface of the cancer cells or in the surrounding extracellular matrix. Accordingly, the present inventors suggest to use this specific and high affinity binding for the targeting of cancer cells or other tissues or cells with high or otherwise expression, such as inappropriate expression of this particular type of chondroitin sulfate proteoglycans.

- 25 So, in a first aspect the present invention relates to an isolated protein fragment of VAR2CSA, which fragment consists of a sequential amino acid sequence of  
a) ID1, and  
b) DBL2Xb, and optionally  
c) ID2a.

- 30 In some embodiments the isolated protein fragment of VAR2CSA according to the present invention comprises ID2a.

In a second aspect the present invention relates to an antibody that specifically binds a protein fragment of VAR2CSA, which fragment consists of a sequential amino acid sequence

of a) ID1, and b) DBL2Xb, and optionally c) ID2a. In some embodiments, the antibody according to the present invention does not bind full length VAR2CSA polypeptides.

In a third aspect the present invention relates to nucleic acid molecules encoding a protein fragment of VAR2CSA, which fragment consist of a sequential amino acid sequence of a) ID1, and b) DBL2Xb, and optionally c) ID2a. The invention further relates to a nucleic acid probe capable of hybridizing to such nucleic acid sequence under stringent conditions.

In a further aspect the present invention relates to a vector comprising an isolated nucleic acid molecule according to the invention.

In a further aspect the present invention relates to a host cell comprising a vector comprising an isolated nucleic acid molecule according to the invention.

In a further aspect the present invention relates to a method for producing the protein fragment according to the invention, the method comprising cultivating a cell as defined herein in an appropriate growth medium under conditions allowing expression of the polynucleotide construct and recovering the resulting protein fragment from the culture medium.

In a further aspect the present invention relates to a conjugate or fusion protein comprising a VAR2CSA polypeptide, and a therapeutic or diagnostic effector moiety, such as a cytotoxic moiety, fluorescent label, and/or a radiolabel.

It is to be understood that for a conjugate, fusion or chimeric protein comprising a VAR2CSA polypeptide, any VAR2CSA polypeptide as defined herein may be used. Accordingly, this aspect is not limited to the use of minimal binding fragments. This applies whenever the term VAR2CSA polypeptide is used and are accordingly equally relevant when used for medical treatment, targeting or diagnosing.

In a further aspect the present invention relates to a composition comprising the protein fragment as defined herein, the antibody according to the invention, or a conjugate according to the invention.

In a further aspect the present invention relates to a protein fragment as defined herein, the antibody according to the invention, a VAR2CSA polypeptide, or a conjugate according to the invention for use as a medicament or diagnostic agent.

In a further aspect the present invention relates to a protein fragment as defined herein, the antibody according to the invention, a VAR2CSA polypeptide, or a conjugate according to the invention for use in diagnosis.

5 In a further aspect the present invention relates to a pharmaceutical composition comprising the protein fragment as defined herein, a VAR2CSA polypeptide, or a conjugate according to the invention.

10 In a further aspect the present invention relates to a method for detecting a protein fragment as defined herein, or a conjugate according to the invention in a biological sample, said method comprising: a) obtaining a biological sample; b) contacting said biological sample with an antibody according to the invention; and c) detecting complexes of said antibody and said protein fragment or conjugate, if any; as an indication of the presence of said protein fragment or conjugate in said sample.

15 Accordingly, methods are provided to measure the level of protein fragments of VAR2CSA in a biological sample. This may be used and applied as part of a treatment, for monitoring the progress of a treatment, or alternatively as part of a production method producing VAR2CSA polypeptides according to the present invention.

In a further aspect the present invention relates to the use of a protein fragment as defined herein, a VAR2CSA polypeptide, or a conjugate according to the invention; for the preparation of a medicament.

20 In a further aspect the present invention relates to a protein fragment as defined herein, a VAR2CSA polypeptide, or a conjugate according to the invention for the treatment of any indications associated with a condition involving expression, such as inappropriate expression of CSA, such as in cancer, arthritis, arthrosis, multiple sclerosis,,healing after neural damage, cartilage repair, wound healing, and in psoriasis.

25 In a further aspect the present invention relates to a method for the treatment of any indication associated with expression, such as inappropriate expression of CSA, such as in cancer, arthritis, arthrosis, multiple sclerosis, pathological conditions caused by neural damage, conditions of the cartilage and scar tissue, such as in rheumatism, cartilage repair or wound healing, or in psoriasis; the method comprising administering a therapeutically or  
30 prophylactically effective amount of a protein fragment as defined herein, a VAR2CSA polypeptide, or a conjugate according to the invention to a subject in need thereof.

In a further aspect the present invention relates to a method for preventing the occurrence of an indication or condition associated with expression, such as inappropriate expression of CSA, such as in cancer, multiple sclerosis, arthritis, arthrosis, pathological conditions caused by neural damage, conditions of the cartilage and scar tissue, such as in rheumatism, cartilage repair or wound healing, or in psoriasis; the method comprising administering a therapeutically or prophylactically effective amount of a protein fragment as defined herein, a VAR2CSA polypeptide, or a conjugate according to the invention to a subject in need thereof.

In a further aspect the present invention relates to the use of a protein fragment as defined herein, a VAR2CSA polypeptide, or a conjugate according to the invention, as a biomarker, such as tool to detect expression, such as inappropriate expression of CSA in body fluids such as blood, plasma, urine, saliva, feces, cerebrospinal fluid, lymph, gastric fluid, pleural fluid, cartilage fluid, sperm, and/or tissue for the diagnosis and/or prognosis of an indication or condition associated with expression, such as inappropriate expression of CSA, such as a malignant disease, arthritis, arthrosis, pathological conditions caused by neural damage, conditions of the cartilage and scar tissue, such as in rheumatism or wound healing, or a cancer disease, such as brain tumors, liver tumors and tumors in the reproductive tract.

It is to be understood that as used herein the term biomarker is intended to refer to the use of VAR2CSA polypeptides, conjugates and fusion proteins according to the present invention when introduced into an organism to detect CSA expression as a mean for diagnosis and/or prognosis of an indication or condition associated with expression of CSA, such as inappropriate expression of CSA.

In a further aspect the present invention relates to the use of a protein fragment as defined herein, a VAR2CSA polypeptide, or a conjugate according to the invention for the immunization of a subject, such as in a vaccine.

In a further aspect the present invention relates to the use of a protein fragment as defined herein, a VAR2CSA polypeptide, or a conjugate according to the invention, as a targeting moiety for the isolation of a cell expressing CD44 and/or CSPG4, and/or any other proteoglycan, such as a proteoglycan listed in Table 1.

Table 1. Potential molecules targeted by a VAR2CSA polypeptide

| Protein ID 1    | Protein ID 2 | Gene name |
|-----------------|--------------|-----------|
| NG2             | CSPG4        | cspg4     |
| Neuroglycan and | CSPG5        | ngc       |

|                                                                        |                                               |               |
|------------------------------------------------------------------------|-----------------------------------------------|---------------|
| Neuroglycan-C                                                          |                                               |               |
| Neuropilin-1 CS                                                        | NRP-1-CS                                      | <b>NRP1</b>   |
| APLP2 and APP (and when CSA is added the proteins are called Appicans) | amyloid precursor- like protein 2             | <b>APLP2</b>  |
| Snorc                                                                  |                                               | Snorc         |
| <b>Tomoregulin-2</b>                                                   | <b>TENB2</b>                                  | <b>TMEFF2</b> |
| Thrombomodulin ,                                                       | CD141                                         | THBD          |
| Betaglycan                                                             | Transforming growth factor beta receptor III  | <b>TGFBR3</b> |
| Syndecan 1                                                             | CD138                                         | <b>SDC1</b>   |
| Syndecan 2                                                             | CD362                                         | <b>SDC2</b>   |
| Syndecan 3                                                             |                                               | <b>SDC3</b>   |
| Syndecan 4                                                             | Amphiglycan                                   | <b>SDC4</b>   |
| CSPG8                                                                  | CSPG8                                         | Cd44          |
| Glypican1-6 (kun 1 og 5)                                               |                                               | GPC1-6        |
| Brevican                                                               | CSPG7                                         | bcan          |
| lubricin                                                               | Proteoglycan 4                                | <i>PRG4</i>   |
| Dentin matrix protein 1                                                |                                               | DMP1          |
| Neurocan                                                               | CSPG3                                         | ncan          |
| Versican                                                               | CSPG2                                         | vcan          |
| Aggrecan                                                               | CSPG1                                         | acan          |
| Bamecan                                                                | CSPG6                                         | smc3          |
| SRPX2                                                                  | Sushi repeat-containing protein               | <b>SRPX2</b>  |
| Serglycin                                                              | Hematopoietic proteoglycan core protein       | <b>SRGN</b>   |
| Decorin                                                                | Small leucine-rich proteoglycan (SLRP) family | dcn           |
| Biglycan                                                               | Small leucine-rich proteoglycan (SLRP) family | bgn           |
| Lumican                                                                | Small leucine-rich proteoglycan (SLRP) family | lum           |
| Fibromodulin                                                           | Small leucine-rich proteoglycan (SLRP) family | fmod          |
| Keratocan                                                              | Small leucine-rich proteoglycan (SLRP) family | kera          |
| Mimecan                                                                | osteoglycin                                   | ogn           |
| Testican 1-3                                                           | BM-40/SPARC/ osteonectin                      | <i>SPOCK1</i> |

|            |                                                                      |               |
|------------|----------------------------------------------------------------------|---------------|
|            | family of extracellular calcium-binding proteins.                    |               |
| phosphacan | Receptor-type tyrosine-protein phosphatase zeta                      | <i>PTPRZ1</i> |
| Leprecan   | Leucine Proline-Enriched Proteoglycan 1                              | LEPRE1        |
| Perlecan   | basement membrane-specific heparan sulfate proteoglycan core protein | <b>HSPG2</b>  |

In a further aspect the present invention relates to a method for the isolation of cells expressing CD44, and/or CSPG4 and/or any other proteoglycan, such as a proteoglycan listed in Table 1, such as cancer stem cells in a biological sample, said method comprising:

- 5                   a) obtaining a biological sample comprising cells expressing CD44, and/or CSPG4 and/or any other proteoglycan, such as a proteoglycan listed in Table 1;
- b) contacting said biological sample with a protein fragment as defined herein, a VAR2CSA polypeptide, or a conjugate according to the invention, optionally coupled to a solid support; and
- 10               c) purifying or isolating the complexes of said cells expressing CD44, and/or CSPG4 and/or any other proteoglycan, such as a proteoglycan listed in Table 1 and said protein fragment or conjugate.

In a further aspect the present invention relates to a diagnostic method for detecting elevated CSA levels in a body fluid, such as blood, plasma, urine, spinal fluid, pleura effusions, joint fluid, bone marrow, gastric fluid, faeces, semen, sperm, prostate fluid, saliva, eye fluid, lung aspirate, and lymph, in response to malignancy or other conditions associated with inappropriate CSA expression, the method comprising the steps of contacting said body fluid with a protein fragment as defined herein, a VAR2CSA polypeptide, or a conjugate according to the invention and detecting the complexes formed with CSA in said body fluid.

20

In a further aspect the present invention relates to a method for the purification of CD44, and/or CSPG4, and/or any other proteoglycan, such as a proteoglycan listed in Table 1 in a biological sample, said method comprising:

- 25                   a) obtaining a biological sample comprising CD44, and/or CSPG4 and/or any other proteoglycan, such as a proteoglycan listed in Table 1;

- b) contacting said biological sample with a protein fragment as defined herein, a VAR2CSA polypeptide, or a conjugate according to the invention, optionally coupled to a solid support; and
- c) purifying or isolating the complexes of said CD44, and/or CSPG4, and/or any other proteoglycan, such as a proteoglycan listed in Table 1 and said protein fragment or conjugate.

In a further aspect the present invention relates to a protein fragment as defined herein, a VAR2CSA polypeptide, or a conjugate according to the invention, or pharmaceutical composition according to the invention in combination with any other suitable anticancer agent.

#### DETAILED DISCLOSURE OF THE INVENTION

This invention is based on the fact that a part of a malaria protein, the so-called VAR2CSA, can bind to a cancer specific antigen and extra-cellular CSPG with very high specificity and very high binding strength.

VAR2CSA mediates parasite adhesion exclusively to low-sulfated chondroitin sulfate A (CSA) attached to proteoglycans (CSPG) in the placenta of pregnant women. Recombinant protein has been shown to bind with unprecedented high affinity and specificity to CSA. This may be due to an interaction with CSA that is not only dependent on the charged sulfates but also on the CS backbone. The inventors of the present invention envision that CS present in the placenta is very similar to the CS presented on various cancer cells including cancer stem cells. This is substantiated by the fact that VAR2CSA expressing malaria parasites adhere specifically to CSA on C32 melanoma cells and to human brain cancer cells.

Accordingly, the current invention relies on the high affinity and specificity between VAR2CSA recombinant proteins and low-sulfated CSA. By tagging this protein the invention can be used in a wide range of applications including the tracking of metastases in vivo and to diagnose metastatic disease. By coupling VAR2CSA to an apoptotic or cytotoxic reagent the invention can be used to specifically target and eliminate cancer cells and cancer stem cells. By simple therapy using VAR2CSA recombinant protein it will be possible to neutralize the activity of CSA thereby inhibiting tumorigenesis and/or metastasis of CSA-expressing cancer cells. CSA can be present on a number of protein backbones, e.g. CSPG4, CD44, biglycan, decorin, versican, aggrecan (the major CSPG in cartilage), perlecan, syndecan, and others listed in Table 1.

The present invention is envisioned to be particularly relevant to malignant melanoma cancer including cutaneous, ocular and conjunctival melanoma having CSPG4 with CSA chains on the surface of the melanoma cells. This GAG chain is believed to be involved in mitoses and metastases. However, CSPG4 is not only specific to melanoma. Micro- and tissue array analyses, performed by the inventors on data from large panels of human tissue and cell lines, suggest that CSPG4 and other types of CSA-containing proteoglycans may be present on a wide range of cancer types originating from all three cellular germ layers. These cancer types include carcinomas (Breast carcinoma, Pancreatic carcinoma, Ovarian carcinoma, Endometrial carcinoma, Hepatocellular carcinoma, Lung carcinoma, Colon carcinoma, Prostate carcinoma, Cervix carcinoma, Testis carcinoma, Basal cell skin carcinoma, Clear cell renal cell carcinoma, Kراتينized head and neck squamous cell carcinoma, Skin squamous cell carcinoma, Vulvar kراتينized squamous cell carcinoma and Vulvar basal cell carcinoma), sarcomas (Breast liposarcoma, Fibrosarcoma, Dedifferentiated chondro- and liposarcoma, Leiomyosarcoma, Liposarcoma, Myxoid liposarcoma, Uterine corpus leiomyosarcoma, Osteosarcoma, Ewing sarcoma and Rhabdomyosarcoma), hematopoietic cancers (Chronic lymphatic leukaemia (CLL), Acute lymphatic leukaemia (ALL), Acute myeloid leukaemia (AML), B-cell, T-cell and large granular lymphoma), tumours of neuroepithelial tissue, such as Astrocytomas (Pleomorphic Xanthoastrocytoma, Fibrillary Astrocytomas, Anaplastic astrocytoma, Glioblastoma Multiforme), Oligodendroglioma, Ependymoma, Choroid plexus tumor, Oligoastrocytoma, gliosarcoma, Ganglioglioma, Retinoblastoma, Neurocytoma, Neuroblastomas (Esthesioneuroblastoma and Ganglioneuroblastoma), Medulloblastoma, Atypical Teratoid Rhabdoid tumors and all types of neuroendocrine cancer.

Chondroitin sulfate proteoglycans (CSPG) also constitute an important component of the extracellular matrix of the central nerve system (CNS) including the eye and of joint cartilage. Extra-cellular CSPG is critically involved in the pathogenesis of arthritis and the lack of regeneration after neural damage. Loss of extra-cellular CSPG is critical for the development of arthritis and arthrosis, and high local concentrations of extra-cellular CSPG prevents neural out growth after neural damage.

VAR2CSA recombinant proteins may not only be used in the treatment of indications associated with malignant growth, such as in cancers. Therapies to either increase or decrease CSPG presence in the extracellular environment may be used to treat arthritis, athrosis and to enhance neural recovery after neurite damage, including multiple sclerosis. For these strategies the inventors of the present invention envision that VAR2CSA may be used either as a direct inhibitor or as a molecule targeting and maintaining drugs changing CSPG metabolism to the affected tissues.

The inventors of the present invention have identified a malaria protein that binds CSA in the intervillous spaces of the placenta with an affinity below 10 nM. Smaller recombinant parts of VAR2CSA have been produced at high yields that bind CSA with characteristics similar to that of the full-length and native VAR2CSA protein. The recombinant VAR2CSA protein does not  
5 bind other CS such as chondroitin sulfate C (C6S) or highly sulfated GAGs such as heparan sulfate (HS). Recombinant proteins can be produced to bind with high affinity to CSA in various expression systems, here among S2 cells, T.Ni cells, CHO cells and E. coli strains including BL21 and Shuffle cells (tm Lifetechnologies).

The inventors of the present invention have also identified a single small (75 kDa) antigen  
10 that binds CSA with very high affinity (nM) and high specificity. Table 3 (See example 2) lists the CSA affinity of all the analyzed VAR2CSA proteins using biosensor technology.

The inventors of the present invention have shown that this VAR2CSA recombinant protein binds strongly at low concentrations to a wide range of cancer cell lines including cutaneous Melanoma (C32, MeWo), Lung carcinoma (A549), Breast carcinoma (HCC1395),  
15 Osteosarcoma (U2OS, MNNG/HOS), Rhabdomyosarcoma (RH30) and cutaneous T-cell lymphoma (Table 4 and 5). As a control molecule another VAR2CSA protein was used, which is identical to the minimal binding VAR2CSA construct except for a 151 amino acids truncation in the C-terminal part of the molecule. This truncation removes the CSA binding. Recombinant VAR2CSA binds all CSPG4 expressing cell lines and cancer cell lines expressing  
20 other CSPG molecules having CSA chains (e.g. T-cell lymphoma). Recombinant VAR2CSA protein fails to interact with human red blood cells and peripheral blood mononuclear cells (PBMC) (Table 4).

The inventors of the present invention have shown herein that malaria parasites adhere to C32 melanoma cells, probably through a specific interaction between CSPG4 and VAR2CSA.  
25 Thus, it is envisioned that recombinant VAR2CSA and conjugates thereof may be used as a therapeutic compound targeting CSA on various cancer cells.

The advantages of targeting CSA on cancer cells with VAR2CSA over other current therapies in development are numerous:

1) The interaction between VAR2CSA and CSA is of unprecedented high affinity and highly  
30 specific.

2) VAR2CSA is an evolutionary refined malaria protein and it is thus unlikely that therapy will break tolerance and cause autoimmune reactions in the patient.

3) VAR2CSA is a stable protein that is well characterized and can be highly expressed in organisms compatible with large-scale protein production.

4) VAR2CSA is a polymorphic protein with a number of serovariants. Repeated therapy could be offered by different serovariants to avoid issues with neutralizing antibodies.

- 5) VAR2CSA is naturally exposed extracellularly on the *P. falciparum*-infected the red blood cell and is thus by nature a stable protein in human serum and has been shown to be highly protease resistant.

The present invention is centred on the interaction between VAR2CSA and CSA. This interaction is a high affinity interaction and the main use is to target CSA expressing cancer cells.

CSA may also be involved in other diseases and pathological conditions like for example arthritis, arthrosis, multiple sclerosis and healing after neural damage, cartilage repair, wound healing, and in psoriasis. Accordingly, VAR2CSA polypeptides or conjugates may be used in the treatment of any such disease or condition.

- In addition the interaction between VAR2CSA and CSA could be used as a biotechnological tool, for example for protein purification and cell sorting assays.

Accordingly, the inventors of the present invention envision several uses of this invention:

- 1) Traceable recombinant VAR2CSA polypeptides or conjugates may be used to track tumors and metastases in cancer patients.
- 2) Recombinant VAR2CSA polypeptides or conjugates may be used to directly target and neutralize CSA activity in cancer cells.
- 3) Recombinant VAR2CSA polypeptides or conjugates, such as VAR2CSA polypeptides coupled to a cytotoxic molecule may be used to target cancer cells with minimal adverse toxicity to CSA-negative tissue.
- 4) Tagged recombinant VAR2CSA polypeptides or conjugates may be used as a research or clinically developmental tool studying CSA on cancer cells.
- 5) A tagged recombinant VAR2CSA polypeptides or conjugates may be used in assays to sort CSA-positive cells in biotechnology and life sciences. This could be done by coupling

recombinant VAR2CSA to resins so that it can be used to purify CSPG4-expressing cells, such as cancer stem cells, providing a novel and efficient biotechnological tool.

6) VAR2CSA polypeptides or conjugates may be used for in vitro depletion of CSPG4-expressing cells, such as cancer cells, as part of autologous transplantations.

5

7) VAR2CSA polypeptides or conjugates could be used in an anti-CSPG4 vaccine. By immunizing animals with CSPG4-VAR2CSA complexes or conjugates, VAR2CSA might act as a carrier and enhancer for an immune response towards CSPG4 with the aim of breaking tolerance to CSPG4.

10

8) VAR2CSA polypeptides or conjugates could be used in monitoring increased CSA levels in body fluids (i.e. urine, spinal fluid, pleural effusions, joints, bone marrow, and lymph) in response to malignancy. This is based on the fact that VAR2CSA polypeptides have specificity for low sulfated CSA and could detect tumor progression as a function of an increased proportion of un-sulfated CS (C0S).

15

9) VAR2CSA polypeptides or conjugates could be used in treatment of arthritis and arthrosis. The VAR2CSA polypeptides could block or target drugs that block protease mediated degradation of aggrecan during arthritis and arthrosis. VAR2CSA polypeptides could also be used to target anti-inflammatory drugs to the affected tissues and to deliver inhibitors such as ADAMTS4 and -5 inhibitors. VAR2CSA polypeptides could be used to target drugs that stimulate the production of aggrecan by chondrocytes. Repeated i.v. injections of aggrecan coupled to VAR2CSA polypeptides could be used to induce tolerance to aggrecan.

20

10) VAR2CSA polypeptides or conjugates could by binding to extracellular CSPG in neural tissue inactivate the CSPGs effect on neurite outgrowth for instance by blocking the signaling through the tyrosine phosphatase-sigma receptor. VAR2CSA peptides could target drugs degrading CSPG or inhibiting CSPG production in affected neural tissue. For examples the following drugs could be considered to be coupled to VAR2CSA: chondroitinase ABC, which cut the sugar chains of the protein core of CSPG molecules. Xylocides, which reduce CSPG production, or drugs that inhibit enzymes important foe CSPG production such as chondroitin synthase or chondroitin polymerizing factor. Examples for such drugs include: 4-flouro-glucosamine, p-nitrophenyl-beta-D-xylooxide, 4-methyl-umbelliferyl-beta-D-xylopyranoside.

25

30

11) VAR2CSA polypeptides or conjugates could also be used to target and maintain cytokines such as IL1-alfa, which stimulate production of ADAMTS4, which subsequently cleave CSPG.

12) CSPG4 expression on cancer cells can influence drug resistance. Tumors in many patients usually initially respond to therapy but chemoresistance develops over time and cancer progresses. CSPG4 expression is associated with multidrug resistance and is mediated by its association with integrin-induced activation of PI3K pathways. Recombinant VAR2CSA polypeptide targeting CSPG4 on cancer cells can reduce or hinder chemoresistance and could thus be used in combination therapies with for example PLX4032, a BRAFV600E inhibitor.

### Definitions

The term "VAR2CSA polypeptide" as used herein refers to the extracellular part of a specific Erythrocyte Membrane Protein 1 (PfEMP1) protein expressed by *Plasmodium falciparum* interacting with chondroitin sulfate proteoglycans (CSPG) and characterized by having a sequence of SEQ ID NO:55 or SEQ ID NO:56 or fragments or variants thereof with the ability to bind chondroitin sulfate A (CSA) that could be presented on a proteoglycans (CSPG).

In some embodiments, the VAR2CSA polypeptide according to the present invention at least comprises the protein fragment of VAR2CSA, which fragment consist of a sequential amino acid sequence of a) ID1, and b) DBL2Xb.

In some embodiments, the VAR2CSA polypeptide according to the present invention at least comprises the protein fragment of VAR2CSA, which fragment consist of a sequential amino acid sequence of a) ID1, and b) DBL2Xb, and c) ID2a.

Included within the definition of a VAR2CSA polypeptide is polypeptides described in Salanti A. et al Mol. Micro 2003 Jul;49(1):179-91, in Khunrae P. et al, J Mol Biol. 2010 Apr 2;397(3):826-34, in Srivastava A. et al, Proc Natl Acad Sci U S A. 2010 Mar 16;107(11):4884-9, in Dahlbäck M. et al, J Biol Chem. 2011 May 6;286(18):15908-17, or in Srivastava A. et al, PLoS One. 2011;6(5):e20270.

The term "ID1" as used herein refers to a domain of VAR2CSA characterized by having an amino acid sequence with at least 70 % sequence identity to an amino acid sequence identified by 1-152 of SEQ ID NO:1.

The term "DBL2Xb" as used herein refers to a domain of VAR2CSA characterized by having an amino acid sequence with at least 70 % sequence identity with to amino acid sequence identified by 153-577 of SEQ ID NO:1.

The term "ID2a" as used herein refers to a domain of VAR2CSA characterized by having an amino acid sequence of at least 20, at least 21, at least 22, at least 23, at least 24, at least 25, at least 26, at least 27, at least 28, at least 29, at least 30, at least 31, at least 32, at least 33, at least 34, at least 35, at least 36, at least 37, at least 38, at least 39, at least 40, at least 41, at least 42, at least 43, at least 44, at least 45, at least 46, at least 47, at least 48, at least 49, at least 50, at least 51, at least 52, at least 53, at least 54, at least 55, at least 56, at least 57, at least 58, at least 59, at least 60, at least 61, or at least 62, such as the 63 consecutive amino acids from the N-terminal of amino acids 578-640 of SEQ ID NO:1 and with at least 70 % sequence identity to such a sequence of consecutive amino acids.

- 10 In some embodiments an amino acid sequence identity referred to herein of at least 70 % of any one sequence identified by SEQ ID NO:1-75 or a fragment thereof, refers to a sequence with at least 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99 % sequence identity to this sequence.

- 15 The terms "variant" or "variants", as used herein, refers to a VAR2CSA polypeptide having an amino acid sequence of SEQ ID NO:55 or SEQ ID NO:56 or a fragments a VAR2CSA polypeptide comprising an amino acid sequence of SEQ ID NO:1-54, which fragments or variants retain the ability to bind chondroitin sulfate A (CSA) on proteoglycans (CSPG), wherein one or more amino acids have been substituted by another amino acid and/or wherein one or more amino acids have been deleted and/or wherein one or more amino acids have been inserted in the polypeptide and/or wherein one or more amino acids have been added to the polypeptide. Such addition can take place either at the N-terminal end or at the C-terminal end or both. The "variant" or "variants" within this definition still have functional activity in terms of being able to bind chondroitin sulfate A (CSA). In some embodiment a variant has at least 70 %, such as at least 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99 % sequence identity with the sequence of SEQ ID NO:1-75, such as the sequence of SEQ ID NO:1, 3-5, 10, 11, 29, 34, 36-38, 41, 43-45, 48, 53-56, 60-70, 72-75.

- 30 The phrases "functional variant", "functional fragment", and "functional derivatives" as used herein refers to variants, fragments, truncated versions, as well as derivatives of SEQ ID NO:55 or SEQ ID NO:56, such as any one of SEQ ID NO:1, 3-5, 10, 11, 29, 34, 36-38, 41, 43-45, 48, 53-56, 60-70, 72-75, which polypeptides comprises essential binding sequence parts of SEQ ID NO:55 or SEQ ID NO:56 and at least possesses the ability to bind chondroitin sulfate A (CSA). It is to be understood that a VAR2CSA functional variant or functional fragment may have two or three features selected from being a both a variant, and/or a fragment and/or a derivative.

A functional variant or fragment of a VAR2CSA polypeptide encompass those that exhibit at least about 25%, such as at least about 50%, such as at least about 75%, such as at least about 90% of the binding affinity of wild-type VAR2CSA polypeptide that has been produced in the same cell type, when tested in the assays as described herein.

- 5     The term "immunologic fragment" as used herein refers to fragment of an amino acid sequence that possesses essentially the same functional activities and the same spatial orientation to be recognized by an antibody. Accordingly a specific antibody will bind both the polypeptide and immunologic fragments thereof.

- 10    The term "another amino acid" as used herein means one amino acid that is different from that amino acid naturally present at that position. This includes but is not limited to amino acids that can be encoded by a polynucleotide. In some embodiments the different amino acid is in natural L-form and can be encoded by a polynucleotide.

- 15    The term "derivative" as used herein, is intended to designate a VAR2CSA polypeptide exhibiting substantially the same or improved biological activity relative to wild-type VAR2CSA identified by SEQ ID NO:55 or SEQ ID NO:56, or a fragment thereof, in which one or more of the amino acids of the parent peptide have been chemically modified, e.g. by alkylation, PEGylation, acylation, ester formation or amide formation or the like.

- 20    The term "construct" is intended to indicate a polynucleotide segment which may be based on a complete or partial naturally occurring nucleotide sequence encoding the polypeptide of interest. The construct may optionally contain other polynucleotide segments. In a similar way, the term "amino acids which can be encoded by polynucleotide constructs" covers amino acids which can be encoded by the polynucleotide constructs defined above, i.e. amino acids such as Ala, Val, Leu, Ile, Met, Phe, Trp, Pro, Gly, Ser, Thr, Cys, Tyr, Asn, Glu, Lys, Arg, His, Asp and Gln.

- 25    The term "vector", as used herein, means any nucleic acid entity capable of the amplification in a host cell. Thus, the vector may be an autonomously replicating vector, i.e. a vector, which exists as an extra-chromosomal entity, the replication of which is independent of chromosomal replication, e.g. a plasmid. Alternatively, the vector may be one which, when introduced into a host cell, is integrated into the host cell genome and replicated together  
30    with the chromosome(s) into which it has been integrated. The choice of vector will often depend on the host cell into which it is to be introduced. Vectors include, but are not limited to plasmid vectors, phage vectors, viruses or cosmid vectors. Vectors usually contain a replication origin and at least one selectable gene, i.e., a gene which encodes a product which is readily detectable or the presence of which is essential for cell growth.

As used herein the term "appropriate growth medium" means a medium containing nutrients and other components required for the growth of cells and the expression of the nucleic acid sequence encoding the VAR2CSA polypeptide of the invention.

In the present context, the term "treatment" is meant to include prevention, curing and alleviating the symptoms of a disease, disorder or condition involving expression, such as inappropriate expression of CSA, such as in cancer. Prophylactic and therapeutic administration of VAR2CSA polypeptide, conjugate or derivative according to the invention is thus included in the term "treatment".

The term "subject" as used herein means any animal, in particular mammals, such as humans, and may, where appropriate, be used interchangeably with the term "patient".

The term "sequence identity" as known in the art, refers to a relationship between the sequences of two or more polypeptide molecules or two or more nucleic acid molecules, as determined by comparing the sequences. In the art, "identity" also means the degree of sequence relatedness between nucleic acid molecules or between polypeptides, as the case may be, as determined by the number of matches between strings of two or more nucleotide residues or two or more amino acid residues. "Identity" measures the percent of identical matches between the smaller of two or more sequences with gap alignments (if any) addressed by a particular mathematical model or computer program (i.e., "algorithms").

The term "similarity" is a related concept, but in contrast to "identity", refers to a sequence relationship that includes both identical matches and conservative substitution matches. If two polypeptide sequences have, for example, (fraction (10/20)) identical amino acids, and the remainder are all non-conservative substitutions, then the percent identity and similarity would both be 50%. If, in the same example, there are 5 more positions where there are conservative substitutions, then the percent identity remains 50%, but the percent similarity would be 75% ((fraction (15/20))). Therefore, in cases where there are conservative substitutions, the degree of similarity between two polypeptides will be higher than the percent identity between those two polypeptides.

Conservative modifications to the amino acid sequence of SEQ ID NO: 1-56, 60-70, and 72-75 (and the corresponding modifications to the encoding nucleotides) will produce VAR2CSA polypeptides having functional and chemical characteristics similar to those of naturally occurring VAR2CSA polypeptides. In contrast, substantial modifications in the functional and/or chemical characteristics of a VAR2CSA polypeptide may be accomplished by selecting substitutions in the amino acid sequence of SEQ ID NO: 1-56, 60-70, and 72-75 that differ significantly in their effect on maintaining (a) the structure of the molecular backbone in the

area of the substitution, for example, as a sheet or helical conformation, (b) the charge or hydrophobicity of the molecule at the target site, or (c) the bulk of the side chain.

For example, a "conservative amino acid substitution" may involve a substitution of a native amino acid residue with a nonnative residue such that there is little or no effect on the

5 polarity or charge of the amino acid residue at that position. Furthermore, any native residue in the polypeptide may also be substituted with alanine, as has been previously described for "alanine scanning mutagenesis" (see, for example, MacLennan et al., 1998, *Acta Physiol. Scand. Suppl.* 643:55-67; Sasaki et al., 1998, *Adv. Biophys.* 35:1-24, which discuss alanine scanning mutagenesis).

10 Desired amino acid substitutions (whether conservative or non-conservative) can be determined by those skilled in the art at the time such substitutions are desired. For example, amino acid substitutions can be used to identify important residues of a VAR2CSA polypeptide, or to increase or decrease the affinity of a VAR2CSA polypeptide described herein.

15 Naturally occurring residues may be divided into classes based on common side chain properties:

- 1) hydrophobic: norleucine, Met, Ala, Val, Leu, Ile;
- 2) neutral hydrophilic: Cys, Ser, Thr, Asn, Gln;
- 3) acidic: Asp, Glu;
- 20 4) basic: His, Lys, Arg;
- 5) residues that influence chain orientation: Gly, Pro; and
- 6) aromatic: Trp, Tyr, Phe.

For example, non-conservative substitutions may involve the exchange of a member of one of these classes for a member from another class. Such substituted residues may be  
25 introduced into regions of the *Plasmodium falciparum* VAR2CSA polypeptide that are homologous with non-*Plasmodium falciparum* VAR2CSA polypeptides, or into the non-homologous regions of the molecule.

In making such changes, the hydropathic index of amino acids may be considered. Each amino acid has been assigned a hydropathic index on the basis of their hydrophobicity and  
30 charge characteristics, these are: isoleucine (+4.5); valine (+4.2); leucine (+3.8); phenylalanine (+2.8); cysteine/cystine (+2.5); methionine (+1.9); alanine (+1.8); glycine (-

0.4); threonine (-0.7); serine (-0.8); tryptophan (-0.9); tyrosine (-1.3); proline (-1.6); histidine (-3.2); glutamate (-3.5); glutamine (-3.5); aspartate (-3.5); asparagine (-3.5); lysine (-3.9); and arginine (-4.5).

The importance of the hydropathic amino acid index in conferring interactive biological function on a protein is understood in the art. Kyte et al., J. Mol. Biol., 157:105-131 (1982). It is known that certain amino acids may be substituted for other amino acids having a similar hydropathic index or score and still retain a similar biological activity. In making changes based upon the hydropathic index, the substitution of amino acids whose hydropathic indexes are within  $\pm 2$  is preferred, those that are within  $\pm 1$  are particularly preferred, and those within  $\pm 0.5$  are even more particularly preferred.

It is also understood in the art that the substitution of like amino acids can be made effectively on the basis of hydrophilicity, particularly where the biologically functionally equivalent protein or peptide thereby created is intended for use in immunological embodiments, as in the present case. The greatest local average hydrophilicity of a protein, as governed by the hydrophilicity of its adjacent amino acids, correlates with its immunogenicity and antigenicity, i.e., with a biological property of the protein.

The following hydrophilicity values have been assigned to amino acid residues: arginine (+3.0); lysine (+3.0); aspartate (+3.0 $\pm$ 1); glutamate (+3.0 $\pm$ 1); serine (+0.3); asparagine (+0.2); glutamine (+0.2); glycine (0); threonine (-0.4); proline (-0.5 $\pm$ 1); alanine (-0.5); histidine (-0.5); cysteine (-1.0); methionine (-1.3); valine (-1.5); leucine (-1.8); isoleucine (-1.8); tyrosine (-2.3); phenylalanine (-2.5); tryptophan (-3.4). In making changes based upon similar hydrophilicity values, the substitution of amino acids whose hydrophilicity values are within  $\pm 2$  is preferred, those that are within  $\pm 1$  are particularly preferred, and those within  $\pm 0.5$  are even more particularly preferred. One may also identify epitopes from primary amino acid sequences on the basis of hydrophilicity. These regions are also referred to as "epitopic core regions."

A skilled artisan will be able to determine suitable variants of the polypeptide as set forth in SEQ ID NO:1-75 using well known techniques. For identifying suitable areas of the molecule that may be changed without destroying activity, one skilled in the art may target areas not believed to be important for activity. For example, when similar polypeptides with similar activities from the same species or from other species are known, one skilled in the art may compare the amino acid sequence of a VAR2CSA polypeptide to such similar polypeptides. With such a comparison, one can identify residues and portions of the molecules that are conserved among similar polypeptides. It will be appreciated that changes in areas of a VAR2CSA polypeptide that are not conserved relative to such similar polypeptides would be

less likely to adversely affect the biological activity and/or structure of the VAR2CSA polypeptide. One skilled in the art would also know that, even in relatively conserved regions, one may substitute chemically similar amino acids for the naturally occurring residues while retaining activity (conservative amino acid residue substitutions). Therefore, even areas that  
5 may be important for biological activity or for structure may be subject to conservative amino acid substitutions without destroying the biological activity or without adversely affecting the polypeptide structure.

Additionally, one skilled in the art can review structure-function studies identifying residues in similar polypeptides that are important for activity or structure. In view of such a  
10 comparison, one can predict the importance of amino acid residues in a VAR2CSA polypeptide that correspond to amino acid residues that are important for activity or structure in similar polypeptides. One skilled in the art may opt for chemically similar amino acid substitutions for such predicted important amino acid residues of VAR2CSA polypeptides and other polypeptides of the invention.

One skilled in the art can also analyze the three-dimensional structure and amino acid  
15 sequence in relation to that structure in similar polypeptides. In view of that information, one skilled in the art may predict the alignment of amino acid residues of a VAR2CSA polypeptide with respect to its three dimensional structure. One skilled in the art may choose not to make radical changes to amino acid residues predicted to be on the surface of the protein, since  
20 such residues may be involved in important interactions with other molecules. Moreover, one skilled in the art may generate test variants containing a single amino acid substitution at each desired amino acid residue. The variants can then be screened using activity assays as described herein. Such variants could be used to gather information about suitable variants. For example, if one discovered that a change to a particular amino acid residue resulted in  
25 destroyed, undesirably reduced, or unsuitable activity, variants with such a change would be avoided. In other words, based on information gathered from such routine experiments, one skilled in the art can readily determine the amino acids where further substitutions should be avoided either alone or in combination with other mutations.

A number of scientific publications have been devoted to the prediction of secondary  
30 structure. See Moulton J., Curr. Op. in Biotech., 7(4):422-427 (1996), Chou et al., Biochemistry, 13(2):222-245 (1974); Chou et al., Biochemistry, 113(2):211-222 (1974); Chou et al., Adv. Enzymol. Relat. Areas Mol. Biol, 47:45-148 (1978); Chou et al., Ann. Rev. Biochem., 47:251-276 and Chou et al., Biophys. J., 26:367-384 (1979). Moreover, computer programs are currently available to assist with predicting secondary structure. One method of  
35 predicting secondary structure is based upon homology modeling. For example, two polypeptides or proteins, which have a sequence identity of greater than 30%, or similarity

greater than 40% often have similar structural topologies. The recent growth of the protein structural data base (PDB) has provided enhanced predictability of secondary structure, including the potential number of folds within a polypeptide's or protein's structure. See Holm et al., Nucl. Acid. Res., 27(1):244-247 (1999). It has been suggested (Brenner et al., Curr. Op. Struct. Biol., 7(3):369-376 (1997)) that there are a limited number of folds in a given polypeptide or protein and that once a critical number of structures have been resolved, structural prediction will gain dramatically in accuracy.

Additional methods of predicting secondary structure include "threading" (Jones, D., Curr. Opin. Struct. Biol., 7(3):377-87 (1997); Sippl et al., Structure, 4(1):15-9 (1996)), "profile analysis" (Bowie et al., Science, 253:164-170 (1991); Gribskov et al., Meth. Enzymol., 183:146-159 (1990); Gribskov et al., Proc. Nat. Acad. Sci., 84(13):4355-4358 (1987)), and "evolutionary linkage" (See Home, supra, and Brenner, supra).

Identity and similarity of related polypeptides can be readily calculated by known methods. Such methods include, but are not limited to, those described in Computational Molecular Biology, Lesk, A. M., ed., Oxford University Press, New York, 1988; Biocomputing: Informatics and Genome Projects, Smith, D. W., ed., Academic Press, New York, 1993; Computer Analysis of Sequence Data, Part 1, Griffin, A. M., and Griffin, H. G., eds., Humana Press, New Jersey, 1994; Sequence Analysis in Molecular Biology, von Heinje, G., Academic Press, 1987; Sequence Analysis Primer, Gribskov, M. and Devereux, J., eds., M. Stockton Press, New York, 1991; and Carillo et al., SIAM J. Applied Math., 48:1073 (1988).

Preferred methods to determine identity and/or similarity are designed to give the largest match between the sequences tested. Methods to determine identity and similarity are described in publicly available computer programs. Preferred computer program methods to determine identity and similarity between two sequences include, but are not limited to, the GCG program package, including GAP (Devereux et al., Nucl. Acid. Res., 12:387 (1984); Genetics Computer Group, University of Wisconsin, Madison, Wis.), BLASTP, BLASTN, and FASTA (Altschul et al., J. Mol. Biol., 215:403-410 (1990)). The BLASTX program is publicly available from the National Center for Biotechnology Information (NCBI) and other sources (BLAST Manual, Altschul et al. NCB/NLM/NIH Bethesda, Md. 20894; Altschul et al., supra). The well known Smith Waterman algorithm may also be used to determine identity.

Certain alignment schemes for aligning two amino acid sequences may result in the matching of only a short region of the two sequences, and this small aligned region may have very high sequence identity even though there is no significant relationship between the two full length sequences. Accordingly, in a preferred embodiment, the selected alignment method

(GAP program) will result in an alignment that spans at least 50 contiguous amino acids of the target polypeptide.

For example, using the computer algorithm GAP (Genetics Computer Group, University of Wisconsin, Madison, Wis.), two polypeptides for which the percent sequence identity is to be determined are aligned for optimal matching of their respective amino acids (the "matched span", as determined by the algorithm). A gap opening penalty (which is calculated as 3 times the average diagonal; the "average diagonal" is the average of the diagonal of the comparison matrix being used; the "diagonal" is the score or number assigned to each perfect amino acid match by the particular comparison matrix) and a gap extension penalty (which is usually  $\{ \text{fraction} (1/10) \}$  times the gap opening penalty), as well as a comparison matrix such as PAM 250 or BLOSUM 62 are used in conjunction with the algorithm. A standard comparison matrix (see Dayhoff et al., Atlas of Protein Sequence and Structure, vol. 5, supp.3 (1978) for the PAM 250 comparison matrix; Henikoff et al., Proc. Natl. Acad. Sci. USA, 89:10915-10919 (1992) for the BLOSUM 62 comparison matrix) is also used by the algorithm.

Preferred parameters for a polypeptide sequence comparison include the following:

Algorithm: Needleman et al., J. Mol. Biol, 48:443-453 (1970); Comparison matrix: BLOSUM 62 from Henikoff et al., Proc. Natl. Acad. Sci. USA, 89:10915-10919 (1992); Gap Penalty: 12, Gap Length Penalty: 4, Threshold of Similarity: 0.

The GAP program is useful with the above parameters. The aforementioned parameters are the default parameters for polypeptide comparisons (along with no penalty for end gaps) using the GAP algorithm.

Preferred parameters for nucleic acid molecule sequence comparisons include the following:

Algorithm: Needleman et al., J. Mol Biol., 48:443-453 (1970); Comparison matrix: matches=+10, mismatch=0, Gap Penalty: 50, Gap Length Penalty: 3.

The GAP program is also useful with the above parameters. The aforementioned parameters are the default parameters for nucleic acid molecule comparisons.

Other exemplary algorithms, gap opening penalties, gap extension penalties, comparison matrices, thresholds of similarity, etc. may be used, including those set forth in the Program Manual, Wisconsin Package, Version 9, September, 1997. The particular choices to be made will be apparent to those of skill in the art and will depend on the specific comparison to be made, such as DNA to DNA, protein to protein, protein to DNA; and additionally, whether the

comparison is between given pairs of sequences (in which case GAP or BestFit are generally preferred) or between one sequence and a large database of sequences (in which case FASTA or BLASTA are preferred).

The inventors of the present invention has now addressed and found the answers to the following key questions related to the molecular mechanism behind placental adhesion in PM:

- 1) is the described differential CSA adhesion related to the VAR2CSA sequence
- 2) what are the exact minimal structural requirements for VAR2CSA binding to CSA
- 3) what type of chemical interaction exists between VAR2CSA and CSA and finally
- 4) can this information be used to design an optimal vaccine antigen?

By expressing identical FCR3 and 3d7 VAR2CSA truncations, the present inventors showed that VAR2CSA bind CSA with similar affinity and specificity, regardless of parasite strain origin. These two sequences has a sequence identity of 79,6 %. The present inventors further demonstrate that the high CSA binding-affinity is retained in several shorter fragments, and that DBL2X, including small regions from the flanking interdomains, form a compact core that contains the high affinity CSA binding site. In silico the present inventors defined putative GAG binding sites in VAR2CSA and by deletion and substitution the present inventors showed that mutations in these sites have no effect on CSPG binding. Using the theory of polyelectrolyte-protein interactions the present inventors have shown that the VAR2CSA-CSA interaction may not, solely, be dependent on ionic interactions. Finally, the present inventors have shown that several short VAR2CSA fragments are capable of inducing the production of adhesion-blocking antibodies and that the anti-adhesive antibodies target the proposed CSA binding region. These data provide the first detailed insight into the biochemical nature of the interaction between a PfEMP1 molecule and its ligand.

### **Preparation of VAR2CSA polypeptides and other polypeptides of the invention**

The invention also relates to a method of preparing VAR2CSA polypeptides and other polypeptides of the invention as mentioned above. The VAR2CSA polypeptides and other polypeptides of the invention described herein may be produced by means of recombinant nucleic acid techniques. In general, a cloned wild-type VAR2CSA nucleic acid sequence is modified to encode the desired protein. This modified sequence is then inserted into an expression vector, which is in turn transformed or transfected into host cells. Higher eukaryotic cells, in particular cultured mammalian cells, may be used as host cells. Prokaryotic cells such as *Lactococcus lactis* or *E. coli* can also be used to express the polypeptides as long as these prokaryotes are able to produce disulfide bonds or the protein is or may be refolded correctly. In addition, Yeast strains can also be used to express the protein, here among *Saccharomyces cerevisiae* and *P. Pichia*.

The amino acid sequence alterations may be accomplished by a variety of techniques. Modification of the nucleic acid sequence may be by site-specific mutagenesis. Techniques for site-specific mutagenesis are well known in the art and are described in, for example, Zoller and Smith (DNA 3:479-488, 1984) or "Splicing by extension overlap", Horton et al., Gene 77, 1989, pp. 61-68. Thus, using the nucleotide and amino acid sequences of VAR2CSA, one may introduce the alteration(s) of choice. Likewise, procedures for preparing a DNA construct using polymerase chain reaction using specific primers are well known to persons skilled in the art (cf. PCR Protocols, 1990, Academic Press, San Diego, California, USA).

The polypeptides of the present invention can also comprise non-naturally occurring amino acid residues. Non-naturally occurring amino acids include, without limitation, beta-alanine, desaminohistidine, trans-3-methylproline, 2,4-methanoproline, cis-4-hydroxyproline, trans-4-hydroxyproline, N-methylglycine, allo-threonine, methylthreonine, hydroxyethylcysteine, hydroxyethylhomocysteine, nitroglutamine, homoglutamine, pipecolic acid, thiazolidine carboxylic acid, dehydroproline, 3- and 4-methylproline, 3,3-dimethylproline, tert-leucine, nor-valine, 2-azaphenylalanine, 3-azaphenylalanine, 4-azaphenylalanine, and 4-fluorophenylalanine. Several methods are known in the art for incorporating non-naturally occurring amino acid residues into polypeptides. For example, an in vitro system can be employed wherein nonsense mutations are suppressed using chemically aminoacylated suppressor tRNAs. Methods for synthesizing amino acids and aminoacylating tRNA are known in the art. Transcription and translation of plasmids containing nonsense mutations is carried out in a cell-free system comprising an E. coli S30 extract and commercially available enzymes and other reagents. Polypeptides are purified by chromatography. See, for example, Robertson et al., J. Am. Chem. Soc. 113:2722, 1991; Ellman et al., Methods Enzymol. 202:301, 1991; Chung et al., Science 259:806-9, 1993; and Chung et al., Proc. Natl. Acad. Sci. USA 90:10145-9, 1993). In a second method, translation is carried out in *Xenopus* oocytes by microinjection of mutated mRNA and chemically aminoacylated suppressor tRNAs (Turcatti et al., J. Biol. Chem. 271:19991-8, 1996). Within a third method, E. coli cells are cultured in the absence of a natural amino acid that is to be replaced (e.g., phenylalanine) and in the presence of the desired non-naturally occurring amino acid(s) (e.g., 2-azaphenylalanine, 3-azaphenylalanine, 4-azaphenylalanine, or 4-fluorophenylalanine). The non-naturally occurring amino acid is incorporated into the polypeptide in place of its natural counterpart. See, Koide et al., Biochem. 33:7470-6, 1994. Naturally occurring amino acid residues can be converted to non-naturally occurring species by in vitro chemical modification. Chemical modification can be combined with site-directed mutagenesis to further expand the range of substitutions (Wynn and Richards, Protein Sci. 2:395-403, 1993).

The nucleic acid construct encoding the VAR2CSA polypeptides and other polypeptides of the invention of the invention may suitably be of genomic or cDNA origin, for instance obtained by preparing a genomic or cDNA library and screening for DNA sequences coding for all or part of the polypeptide by hybridization using synthetic oligonucleotide probes in accordance with standard techniques (cf. Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 2nd. Ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1989).

The nucleic acid construct encoding a VAR2CSA polypeptide may also be prepared synthetically by established standard methods, e.g. the phosphoramidite method described by Beaucage and Caruthers, *Tetrahedron Letters* 22 (1981), 1859 - 1869, or the method described by Matthes et al., *EMBO Journal* 3 (1984), 801 - 805. According to the phosphoramidite method, oligonucleotides are synthesised, e.g. in an automatic DNA synthesiser, purified, annealed, ligated and cloned in suitable vectors. The DNA sequences encoding the *Plasmodium falciparum* VAR2CSA polypeptides and other polypeptides of the invention may also be prepared by polymerase chain reaction using specific primers, for instance as described in US 4,683,202, Saiki et al., *Science* 239 (1988), 487 - 491, or Sambrook et al., *supra*.

Furthermore, the nucleic acid construct may be of mixed synthetic and genomic, mixed synthetic and cDNA or mixed genomic and cDNA origin prepared by ligating fragments of synthetic, genomic or cDNA origin (as appropriate), the fragments corresponding to various parts of the entire nucleic acid construct, in accordance with standard techniques.

The nucleic acid construct is preferably a DNA construct. DNA sequences for use in producing VAR2CSA polypeptides and other polypeptides according to the present invention will typically encode a pre-pro polypeptide at the amino-terminus of VAR2CSA to obtain proper posttranslational processing and secretion from the host cell.

The DNA sequences encoding the *Plasmodium falciparum* VAR2CSA polypeptides and other polypeptides according to the present invention are usually inserted into a recombinant vector which may be any vector, which may conveniently be subjected to recombinant DNA procedures, and the choice of vector will often depend on the host cell into which it is to be introduced. Thus, the vector may be an autonomously replicating vector, i.e. a vector, which exists as an extrachromosomal entity, the replication of which is independent of chromosomal replication, e.g. a plasmid. Alternatively, the vector may be one which, when introduced into a host cell, is integrated into the host cell genome and replicated together with the chromosome(s) into which it has been integrated.

The vector is preferably an expression vector in which the DNA sequence encoding the Plasmodium falciparum VAR2CSA polypeptides and other polypeptides according to the present invention is operably linked to additional segments required for transcription of the DNA. In general, the expression vector is derived from plasmid or viral DNA, or may contain  
5 elements of both. The term, "operably linked" indicates that the segments are arranged so that they function in concert for their intended purposes, e.g. transcription initiates in a promoter and proceeds through the DNA sequence coding for the polypeptide.

Expression vectors for use in expressing VAR2CSA polypeptides and other polypeptides according to the present invention will comprise a promoter capable of directing the  
10 transcription of a cloned gene or cDNA. The promoter may be any DNA sequence, which shows transcriptional activity in the host cell of choice and may be derived from genes encoding proteins either homologous or heterologous to the host cell.

Examples of suitable promoters for directing the transcription of the DNA encoding the Plasmodium falciparum VAR2CSA polypeptide in mammalian cells are the SV40 promoter  
15 (Subramani et al., Mol. Cell Biol. 1 (1981), 854 -864), the MT-1 (metallothionein gene) promoter (Palmiter et al., Science 222 (1983), 809 - 814), the CMV promoter (Boshart et al., Cell 41:521-530, 1985) or the adenovirus 2 major late promoter (Kaufman and Sharp, Mol. Cell. Biol, 2:1304-1319, 1982).

An example of a suitable promoter for use in insect cells is the polyhedrin promoter (US  
20 4,745,051; Vasuvedan et al., FEBS Lett. 311, (1992) 7 - 11), the P10 promoter (J.M. Vlak et al., J. Gen. Virology 69, 1988, pp. 765-776), the Autographa californica polyhedrosis virus basic protein promoter (EP 397 485), the baculovirus immediate early gene 1 promoter (US 5,155,037; US 5,162,222), or the baculovirus 39K delayed-early gene promoter (US 5,155,037; US 5,162,222).

Examples of suitable promoters for use in yeast host cells include promoters from yeast glycolytic genes (Hitzeman et al., J. Biol. Chem. 255 (1980), 12073 - 12080; Alber and Kawasaki, J. Mol. Appl. Gen. 1 (1982), 419 - 434) or alcohol dehydrogenase genes (Young et al., in Genetic Engineering of Microorganisms for Chemicals (Hollaender et al, eds.), Plenum Press, New York, 1982), or the TPI1 (US 4,599,311) or ADH2-4c (Russell et al., Nature 304  
25 (1983), 652 - 654) promoters.  
30

Examples of suitable promoters for use in filamentous fungus host cells are, for instance, the ADH3 promoter (McKnight et al., The EMBO J. 4 (1985), 2093 - 2099) or the tpiA promoter. Examples of other useful promoters are those derived from the gene encoding A. oryzae TAKA amylase, Rhizomucor miehei aspartic proteinase, A. niger neutral alpha-amylase, A.

niger acid stable alpha-amylase, *A. niger* or *A. awamori* glucoamylase (gluA), *Rhizomucor miehei* lipase, *A. oryzae* alkaline protease, *A. oryzae* triose phosphate isomerase or *A. nidulans* acetamidase. Preferred are the TAKA-amylase and gluA promoters. Suitable promoters are mentioned in, e.g. EP 238 023 and EP 383 779.

- 5 The DNA sequences encoding the *Plasmodium falciparum* VAR2CSA polypeptides and other polypeptides according to the present invention may also, if necessary, be operably connected to a suitable terminator, such as the human growth hormone terminator (Palmiter et al., Science 222, 1983, pp. 809-814) or the TPI1 (Alber and Kawasaki, J. Mol. Appl. Gen. 1, 1982, pp. 419-434) or ADH3 (McKnight et al., The EMBO J. 4, 1985, pp. 2093-2099)
- 10 terminators. Expression vectors may also contain a set of RNA splice sites located downstream from the promoter and upstream from the insertion site for the VAR2CSA sequence itself. Preferred RNA splice sites may be obtained from adenovirus and/or immunoglobulin genes. Also contained in the expression vectors is a polyadenylation signal located downstream of the insertion site. Particularly preferred polyadenylation signals
- 15 include the early or late polyadenylation signal from SV40 (Kaufman and Sharp, *ibid.*), the polyadenylation signal from the adenovirus 5 Elb region, the human growth hormone gene terminator (DeNoto et al. Nucl. Acids Res. 9:3719-3730, 1981) or the polyadenylation signal from *Plasmodium falciparum*, human or bovine genes. The expression vectors may also include a noncoding viral leader sequence, such as the adenovirus 2 tripartite leader, located
- 20 between the promoter and the RNA splice sites; and enhancer sequences, such as the SV40 enhancer.

To direct the *Plasmodium falciparum* VAR2CSA polypeptides and other polypeptides of the present invention into the secretory pathway of the host cells, a secretory signal sequence (also known as a leader sequence, prepro sequence or pre sequence) may be provided in the

25 recombinant vector. The secretory signal sequence is joined to the DNA sequences encoding the *Plasmodium falciparum* VAR2CSA polypeptides and other polypeptides according to the present invention in the correct reading frame. Secretory signal sequences are commonly positioned 5' to the DNA sequence encoding the peptide. The secretory signal sequence may be that, normally associated with the protein or may be from a gene encoding another

30 secreted protein.

For secretion from yeast cells, the secretory signal sequence may encode any signal peptide, which ensures efficient direction of the expressed *Plasmodium falciparum* VAR2CSA polypeptides and other polypeptides according to the present invention into the secretory pathway of the cell. The signal peptide may be naturally occurring signal peptide, or a

35 functional part thereof, or it may be a synthetic peptide. Suitable signal peptides have been found to be the alpha-factor signal peptide (cf. US 4,870,008), the signal peptide of mouse

salivary amylase (cf. O. Hagenbuchle et al., Nature 289, 1981, pp. 643-646), a modified carboxypeptidase signal peptide (cf. L.A. Valls et al., Cell 48, 1987, pp. 887-897), the yeast BAR1 signal peptide (cf. WO 87/02670), or the yeast aspartic protease 3 (YAP3) signal peptide (cf. M. Egel-Mitani et al., Yeast 6, 1990, pp. 127-137).

- 5 For efficient secretion in yeast, a sequence encoding a leader peptide may also be inserted downstream of the signal sequence and upstream of the DNA sequence encoding the Plasmodium falciparum VAR2CSA polypeptides and other polypeptides according to the present invention. The function of the leader peptide is to allow the expressed peptide to be directed from the endoplasmic reticulum to the Golgi apparatus and further to a secretory vesicle for secretion into the culture medium (i.e. exportation of the Plasmodium falciparum VAR2CSA polypeptides and other polypeptides according to the present invention across the cell wall or at least through the cellular membrane into the periplasmic space of the yeast cell). The leader peptide may be the yeast alpha-factor leader (the use of which is described in e.g. US 4,546,082, US 4,870,008, EP 16 201, EP 123 294, EP 123 544 and EP 163 529).
- 10
- 15 Alternatively, the leader peptide may be a synthetic leader peptide, which is to say a leader peptide not found in nature. Synthetic leader peptides may, for instance, be constructed as described in WO 89/02463 or WO 92/11378.

- For use in filamentous fungi, the signal peptide may conveniently be derived from a gene encoding an Aspergillus sp. amylase or glucoamylase, a gene encoding a Rhizomucor miehei lipase or protease or a Humicola lanuginosa lipase. The signal peptide is preferably derived from a gene encoding A. oryzae TAKA amylase, A. niger neutral alpha-amylase, A. niger acid-stable amylase, or A. niger glucoamylase. Suitable signal peptides are disclosed in, e.g. EP 238 023 and EP 215 594.
- 20

- For use in insect cells, the signal peptide may conveniently be derived from an insect gene (cf. WO 90/05783), such as the lepidopteran Manduca sexta adipokinetic hormone precursor signal peptide (cf. US 5,023,328).
- 25

- The procedures used to ligate the DNA sequences coding for the Plasmodium falciparum VAR2CSA polypeptides and other polypeptides according to the present invention, the promoter and optionally the terminator and/or secretory signal sequence, respectively, and to insert them into suitable vectors containing the information necessary for replication, are well known to persons skilled in the art (cf., for instance, Sambrook et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor, New York, 1989).
- 30

Methods of transfecting mammalian cells and expressing DNA sequences introduced in the cells are described in e.g. Kaufman and Sharp, J. Mol. Biol. 159 (1982), 601 - 621; Southern

and Berg, J. Mol. Appl. Genet. 1 (1982), 327 - 341; Loyter et al., Proc. Natl. Acad. Sci. USA 79 (1982), 422 - 426; Wigler et al., Cell 14 (1978), 725; Corsaro and Pearson, Somatic Cell Genetics 7 (1981), 603, Graham and van der Eb, Virology 52 (1973), 456; and Neumann et al., EMBO J. 1 (1982), 841 - 845.

- 5 Cloned DNA sequences are introduced into cultured mammalian cells by, for example, calcium phosphate-mediated transfection (Wigler et al., Cell 14:725-732, 1978; Corsaro and Pearson, Somatic Cell Genetics 7:603-616, 1981; Graham and Van der Eb, Virology 52d:456-467, 1973) or electroporation (Neumann et al., EMBO J. 1:841-845, 1982). To identify and select cells that express the exogenous DNA, a gene that confers a selectable phenotype (a  
10 selectable marker) is generally introduced into cells along with the gene or cDNA of interest. Preferred selectable markers include genes that confer resistance to drugs such as neomycin, hygromycin, and methotrexate. The selectable marker may be an amplifiable selectable marker. A preferred amplifiable selectable marker is a dihydrofolate reductase (DHFR) sequence. Selectable markers are reviewed by Thilly (Mammalian Cell Technology,  
15 Butterworth Publishers, Stoneham, MA, incorporated herein by reference). The person skilled in the art will easily be able to choose suitable selectable markers.

- Selectable markers may be introduced into the cell on a separate plasmid at the same time as the gene of interest, or they may be introduced on the same plasmid. If on the same plasmid, the selectable marker and the gene of interest may be under the control of different  
20 promoters or the same promoter, the latter arrangement producing a dicistronic message. Constructs of this type are known in the art (for example, Levinson and Simonsen, U.S. 4,713,339). It may also be advantageous to add additional DNA, known as "carrier DNA," to the mixture that is introduced into the cells.

- After the cells have taken up the DNA, they are grown in an appropriate growth medium,  
25 typically 1-2 days, to begin expressing the gene of interest. As used herein the term "appropriate growth medium" means a medium containing nutrients and other components required for the growth of cells and the expression of the Plasmodium falciparum VAR2CSA polypeptide of interest. Media generally include a carbon source, a nitrogen source, essential amino acids, essential sugars, vitamins, salts, phospholipids, protein and growth factors.  
30 Drug selection is then applied to select for the growth of cells that are expressing the selectable marker in a stable fashion. For cells that have been transfected with an amplifiable selectable marker the drug concentration may be increased to select for an increased copy number of the cloned sequences, thereby increasing expression levels. Clones of stably transfected cells are then screened for expression of the Plasmodium falciparum VAR2CSA  
35 polypeptide of interest.

The host cell into which the DNA sequences encoding the *Plasmodium falciparum* VAR2CSA polypeptides and other polypeptides according to the present invention is introduced may be any cell, which is capable of producing the posttranslational modified polypeptides and includes yeast, fungi and higher eucaryotic cells.

5 Examples of mammalian cell lines for use in the present invention are the COS-1 (ATCC CRL 1650), baby hamster kidney (BHK) and 293 (ATCC CRL 1573; Graham et al., J. Gen. Virol. 36:59-72, 1977) cell lines. A preferred BHK cell line is the tk- ts13 BHK cell line (Waechter and Baserga, Proc. Natl. Acad. Sci. USA 79:1106-1110, 1982, incorporated herein by  
10 reference), hereinafter referred to as BHK 570 cells. The BHK 570 cell line has been deposited with the American Type Culture Collection, 12301 Parklawn Dr., Rockville, Md. 20852, under ATCC accession number CRL 10314. A tk- ts13 BHK cell line is also available from the ATCC under accession number CRL 1632. In addition, a number of other cell lines may be used within the present invention, including Rat Hep I (Rat hepatoma; ATCC CRL 1600), Rat Hep II (Rat hepatoma; ATCC CRL 1548), TCMK (ATCC CCL 139), Human lung  
15 (ATCC HB 8065), NCTC 1469 (ATCC CCL 9.1), CHO (ATCC CCL 61) and DUKX cells (Urlaub and Chasin, Proc. Natl. Acad. Sci. USA 77:4216-4220, 1980).

Examples of suitable yeasts cells include cells of *Saccharomyces* spp. or *Schizosaccharomyces* spp., in particular strains of *Saccharomyces cerevisiae* or *Saccharomyces kluyveri*. Methods for transforming yeast cells with heterologous DNA and producing  
20 heterologous poly-peptides there from are described, e.g. in US 4,599,311, US 4,931,373, US 4,870,008, 5,037,743, and US 4,845,075, all of which are hereby incorporated by reference. Transformed cells are selected by a phenotype determined by a selectable marker, commonly drug resistance or the ability to grow in the absence of a particular nutrient, e.g. leucine. A preferred vector for use in yeast is the POT1 vector disclosed in US 4,931,373. The  
25 DNA sequences encoding the *Plasmodium falciparum* VAR2CSA polypeptides and other polypeptides according to the present invention may be preceded by a signal sequence and optionally a leader sequence, e.g. as described above. Further examples of suitable yeast cells are strains of *Kluyveromyces*, such as *K. lactis*, *Hansenula*, e.g. *H. polymorpha*, or *Pichia*, e.g. *P. pastoris* (cf. Gleeson et al., J. Gen. Microbiol. 132, 1986, pp. 3459-3465; US  
30 4,882,279).

Examples of other fungal cells are cells of filamentous fungi, e.g. *Aspergillus* spp., *Neurospora* spp., *Fusarium* spp. or *Trichoderma* spp., in particular strains of *A. oryzae*, *A. nidulans* or *A. niger*. The use of *Aspergillus* spp. for the expression of proteins is described in, e.g., EP 272 277, EP 238 023, EP 184 438 The transformation of *F. oxysporum* may, for  
35 instance, be carried out as described by Malardier et al., 1989, Gene 78: 147-156. The

transformation of *Trichoderma* spp. may be performed for instance as described in EP 244 234.

When a filamentous fungus is used as the host cell, it may be transformed with the DNA construct of the invention, conveniently by integrating the DNA construct in the host chromosome to obtain a recombinant host cell. This integration is generally considered to be an advantage as the DNA sequence is more likely to be stably maintained in the cell. Integration of the DNA constructs into the host chromosome may be performed according to conventional methods, e.g. by homologous or heterologous recombination.

Transformation of insect cells and production of heterologous polypeptides therein may be performed as described in US 4,745,051; US 4,879,236; US 5,155,037; 5,162,222; EP 397,485) all of which are incorporated herein by reference. The insect cell line used as the host may suitably be a Lepidoptera cell line, such as *Spodoptera frugiperda* cells or *Trichoplusia ni* cells (cf. US 5,077,214). Culture conditions may suitably be as described in, for instance, WO 89/01029 or WO 89/01028, or any of the aforementioned references.

The transformed or transfected host cell described above is then cultured in a suitable nutrient medium under conditions permitting expression of the *Plasmodium falciparum* VAR2CSA polypeptide after which all or part of the resulting peptide may be recovered from the culture. The medium used to culture the cells may be any conventional medium suitable for growing the host cells, such as minimal or complex media containing appropriate supplements. Suitable media are available from commercial suppliers or may be prepared according to published recipes (e.g. in catalogues of the American Type Culture Collection). The *Plasmodium falciparum* VAR2CSA polypeptide produced by the cells may then be recovered from the culture medium by conventional procedures including separating the host cells from the medium by centrifugation or filtration, precipitating the proteinaceous components of the supernatant or filtrate by means of a salt, e.g. ammonium sulfate, purification by a variety of chromatographic procedures, e.g. ion exchange chromatography, gel filtration chromatography, affinity chromatography, or the like, dependent on the type of polypeptide in question.

Transgenic animal technology may be employed to produce the VAR2CSA polypeptides and other polypeptides of the invention. It is preferred to produce the proteins within the mammary glands of a host female mammal. Expression in the mammary gland and subsequent secretion of the protein of interest into the milk overcomes many difficulties encountered in isolating proteins from other sources. Milk is readily collected, available in large quantities, and biochemically well characterized. Furthermore, the major milk proteins are present in milk at high concentrations (typically from about 1 to 15 g/l).

From a commercial point of view, it is clearly preferable to use as the host a species that has a large milk yield. While smaller animals such as mice and rats can be used (and are preferred at the proof of principle stage), it is preferred to use livestock mammals including, but not limited to, pigs, goats, sheep and cattle. Sheep are particularly preferred due to such factors as the previous history of transgenesis in this species, milk yield, cost and the ready availability of equipment for collecting sheep milk (see, for example, WO 88/00239 for a comparison of factors influencing the choice of host species). It is generally desirable to select a breed of host animal that has been bred for dairy use, such as East Friesland sheep, or to introduce dairy stock by breeding of the transgenic line at a later date. In any event, animals of known, good health status should be used.

To obtain expression in the mammary gland, a transcription promoter from a milk protein gene is used. Milk protein genes include those genes encoding caseins (see U.S. 5,304,489), beta lactoglobulin, a lactalbumin, and whey acidic protein. The beta lactoglobulin (BLG) promoter is preferred. In the case of the ovine beta lactoglobulin gene, a region of at least the proximal 406 bp of 5' flanking sequence of the gene will generally be used, although larger portions of the 5' flanking sequence, up to about 5 kbp, are preferred, such as a ~4.25 kbp DNA segment encompassing the 5' flanking promoter and non coding portion of the beta lactoglobulin gene (see Whitelaw et al., Biochem. J. 286: 31-39 (1992)). Similar fragments of promoter DNA from other species are also suitable.

Other regions of the beta lactoglobulin gene may also be incorporated in constructs, as may genomic regions of the gene to be expressed. It is generally accepted in the art that constructs lacking introns, for example, express poorly in comparison with those that contain such DNA sequences (see Brinster et al., Proc. Natl. Acad. Sci. USA 85: 836-840 (1988); Palmiter et al., Proc. Natl. Acad. Sci. USA 88: 478-482 (1991); Whitelaw et al., Transgenic Res. 1: 3-13 (1991); WO 89/01343; and WO 91/02318, each of which is incorporated herein by reference). In this regard, it is generally preferred, where possible, to use genomic sequences containing all or some of the native introns of a gene encoding the protein or polypeptide of interest, thus the further inclusion of at least some introns from, e.g., the beta lactoglobulin gene, is preferred. One such region is a DNA segment that provides for intron splicing and RNA polyadenylation from the 3' non coding region of the ovine beta lactoglobulin gene. When substituted for the natural 3' non coding sequences of a gene, this ovine beta lactoglobulin segment can both enhance and stabilize expression levels of the protein or polypeptide of interest. Within other embodiments, the region surrounding the initiation ATG of the VAR2CSA sequence is replaced with corresponding sequences from a milk specific protein gene. Such replacement provides a putative tissue specific initiation environment to enhance expression. It is convenient to replace the entire VAR2CSA pre pro

and 5' non coding sequences with those of, for example, the BLG gene, although smaller regions may be replaced.

For expression of VAR2CSA polypeptides and other polypeptides according to the present invention in transgenic animals, a DNA segment encoding VAR2CSA is operably linked to additional DNA segments required for its expression to produce expression units. Such additional segments include the above mentioned promoter, as well as sequences that provide for termination of transcription and polyadenylation of mRNA. The expression units will further include a DNA segment encoding a secretory signal sequence operably linked to the segment encoding modified VAR2CSA. The secretory signal sequence may be a native secretory signal sequence or may be that of another protein, such as a milk protein (see, for example, von Heijne, Nucl. Acids Res. 14: 4683 4690 (1986); and Meade et al., U.S. 4,873,316, which are incorporated herein by reference).

Construction of expression units for use in transgenic animals is conveniently carried out by inserting a VAR2CSA sequence into a plasmid or phage vector containing the additional DNA segments, although the expression unit may be constructed by essentially any sequence of ligations. It is particularly convenient to provide a vector containing a DNA segment encoding a milk protein and to replace the coding sequence for the milk protein with that of a VAR2CSA variant; thereby creating a gene fusion that includes the expression control sequences of the milk protein gene. In any event, cloning of the expression units in plasmids or other vectors facilitates the amplification of the VAR2CSA sequence. Amplification is conveniently carried out in bacterial (e.g. E. coli) host cells, thus the vectors will typically include an origin of replication and a selectable marker functional in bacterial host cells. The expression unit is then introduced into fertilized eggs (including early stage embryos) of the chosen host species. Introduction of heterologous DNA can be accomplished by one of several routes, including microinjection (e.g. U.S. Patent No. 4,873,191), retroviral infection (Jaenisch, Science 240: 1468 1474 (1988)) or site directed integration using embryonic stem (ES) cells (reviewed by Bradley et al., Bio/Technology 10: 534 539 (1992)). The eggs are then implanted into the oviducts or uteri of pseudopregnant females and allowed to develop to term. Offspring carrying the introduced DNA in their germ line can pass the DNA on to their progeny in the normal, Mendelian fashion, allowing the development of transgenic herds. General procedures for producing transgenic animals are known in the art (see, for example, Hogan et al., Manipulating the Mouse Embryo: A Laboratory Manual, Cold Spring Harbor Laboratory, 1986; Simons et al., Bio/Technology 6: 179 183 (1988); Wall et al., Biol. Reprod. 32: 645 651 (1985); Buhler et al., Bio/Technology 8: 140 143 (1990); Ebert et al., Bio/Technology 9: 835 838 (1991); Krimpenfort et al., Bio/Technology 9: 844 847 (1991); Wall et al., J. Cell. Biochem. 49: 113 120 (1992); U.S. 4,873,191; U.S. 4,873,316; WO 88/00239, WO 90/05188, WO 92/11757; and GB 87/00458). Techniques for introducing

foreign DNA sequences into mammals and their germ cells were originally developed in the mouse (see, e.g., Gordon et al., Proc. Natl. Acad. Sci. USA 77: 7380 7384 (1980); Gordon and Ruddle, Science 214: 1244 1246 (1981); Palmiter and Brinster, Cell 41: 343 345 (1985); Brinster et al., Proc. Natl. Acad. Sci. USA 82: 4438 4442 (1985); and Hogan et al. (ibid.)).

5 These techniques were subsequently adapted for use with larger animals, including livestock species (see, e.g., WO 88/00239, WO 90/05188, and WO 92/11757; and Simons et al., Bio/Technology 6: 179 183 (1988)). To summarize, in the most efficient route used to date in the generation of transgenic mice or livestock, several hundred linear molecules of the DNA of interest are injected into one of the pro nuclei of a fertilized egg according to  
10 established techniques. Injection of DNA into the cytoplasm of a zygote can also be employed.

Production in transgenic plants may also be employed. Expression may be generalised or directed to a particular organ, such as a tuber (see, Hiatt, Nature 344:469 479 (1990); Edelbaum et al., J. Interferon Res. 12:449 453 (1992); Sijmons et al., Bio/Technology 8:217  
15 221 (1990); and EP 0 255 378).

#### *VAR2CSA purification*

The VAR2CSA polypeptides and other polypeptides of the invention may be recovered from cell culture medium or milk. The VAR2CSA polypeptides and other polypeptides of the present invention may be purified by a variety of procedures known in the art including, but  
20 not limited to, chromatography (e.g., ion exchange, affinity, hydrophobic, chromatofocusing, and size exclusion), electrophoretic procedures (e.g., preparative isoelectric focusing (IEF), differential solubility (e.g., ammonium sulfate precipitation), or extraction (see, e.g., Protein Purification, J.-C. Janson and Lars Ryden, editors, VCH Publishers, New York, 1989).

Preferably, they may be purified by affinity chromatography on an anti-VAR2CSA antibody  
25 column. Additional purification may be achieved by conventional chemical purification means, such as high performance liquid chromatography. Other methods of purification, including barium citrate precipitation, are known in the art, and may be applied to the purification of the novel VAR2CSA polypeptides and other polypeptides described herein (see, for example, Scopes, R., Protein Purification, Springer-Verlag, N.Y., 1982).

30 For therapeutic purposes it is preferred that the VAR2CSA polypeptides and other polypeptides of the invention are substantially pure. Thus, in a preferred embodiment of the invention the and other polypeptides of the invention are purified to at least about 90 to 95% homogeneity, preferably to at least about 98% homogeneity. Purity may be assessed by e.g. gel electrophoresis and amino-terminal amino acid sequencing.

The term "isolated polypeptide" refers to a polypeptide of the present invention that (1) has been separated from at least about 50 percent of polynucleotides, lipids, carbohydrates or other materials (i.e., contaminants) with which it is naturally associated. Preferably, the isolated polypeptide is substantially free from any other contaminating polypeptides or other  
5 contaminants that are found in its natural environment, which would interfere with its therapeutic, diagnostic, prophylactic or research use.

The term "microorganism" as used herein refers to bacteria, fungi, archaea, protists; microscopic plants and animals (such as green algae or plankton), the planarian and amoeba. Included within this definition are pathogenic microorganisms.

## 10 **Administration and pharmaceutical compositions**

### *Combination treatments*

The VAR2CSA polypeptide, derivative, or conjugate as defined in the present specification may be administered simultaneously or sequentially with one or more other cancer agent, and/or be used in a combination treatment with other known therapies. The factors may be  
15 supplied in single-dosage form wherein the single-dosage form contains both compounds, or in the form of a kit-of-parts comprising a preparation of a VAR2CSA polypeptide as a first unit dosage form and a preparation of the one or more other compound as a second unit dosage form. Whenever a first or second or third, etc., unit dose is mentioned throughout this specification this does not indicate the preferred order of administration, but is merely done  
20 for convenience purposes.

Suitable other cancer agents or therapies that may be used in combination with a VAR2CSA polypeptide includes antibodies already on the market or in development, including Vemurafenib (Hoffmann-La Roche), human monoclonal antibodies against MCSP, Therapeutical (Micromet Inc) anti-MCSP using BiTE antibody platform technology, and  
25 Adoptive transfer of cytotoxic T cells with specificity for MCSP.

By "simultaneous" dosing of a preparation of a VAR2CSA polypeptide and a preparation of one or more other compound is meant administration of the compounds in single-dosage form, or administration of a first agent followed by administration of a second agent with a time separation of no more than 15 minutes, preferably 10, more preferred 5, more preferred  
30 2 minutes. Either factor may be administered first.

By "sequential" dosing is meant administration of a first agent followed by administration of a second agent with a time separation of more than 15 minutes. Either of the two unit dosage form may be administered first. Preferably, both products are injected through the same intravenous access.

- 5 Another object of the present invention is to provide a pharmaceutical formulation comprising a VAR2CSA polypeptide which is present in a serum/plasma concentration from 0 mg/ml to 1 mg/ml, and wherein the formulation has a pH from 2.0 to 10.0. The formulation may further comprise a buffer system, preservative(s), tonicity agent(s), chelating agent(s), stabilizers and surfactants. In some embodiments of the invention the pharmaceutical formulation is an aqueous formulation, i.e. formulation comprising water. Such formulation is typically a solution or a suspension. In a further embodiment of the invention the pharmaceutical formulation is an aqueous solution. The term "aqueous formulation" is defined as a formulation comprising at least 50 % w/w water. Likewise, the term "aqueous solution" is defined as a solution comprising at least 50 %w/w water, and the term "aqueous suspension" is defined as a suspension comprising at least 50 % w/w water.

In other embodiments the pharmaceutical formulation is a freeze-dried formulation, whereto the physician or the patient adds solvents and/or diluents prior to use.

In other embodiments the pharmaceutical formulation is a dried formulation (e.g. freeze-dried or spray-dried) ready for use without any prior dissolution.

- 20 In a further aspect the invention relates to a pharmaceutical formulation comprising an aqueous solution of a VAR2CSA polypeptide, and a buffer, wherein the VAR2CSA polypeptide is present in a serum/plasma concentration from 0-1 mg/ml or above, and wherein the formulation has a pH from about 2.0 to about 10.0.

- In a other embodiments of the invention the pH of the formulation is selected from the list consisting of 2.0, 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 2.7, 2.8, 2.9, 3.0, 3.1, 3.2, 3.3, 3.4, 3.5, 3.6, 3.7, 3.8, 3.9, 4.0, 4.1, 4.2, 4.3, 4.4, 4.5, 4.6, 4.7, 4.8, 4.9, 5.0, 5.1, 5.2, 5.3, 5.4, 5.5, 5.6, 5.7, 5.8, 5.9, 6.0, 6.1, 6.2, 6.3, 6.4, 6.5, 6.6, 6.7, 6.8, 6.9, 7.0, 7.1, 7.2, 7.3, 7.4, 7.5, 7.6, 7.7, 7.8, 7.9, 8.0, 8.1, 8.2, 8.3, 8.4, 8.5, 8.6, 8.7, 8.8, 8.9, 9.0, 9.1, 9.2, 9.3, 9.4, 9.5, 9.6, 9.7, 9.8, 9.9, and 10.0.

- 30 In a further embodiment of the invention the buffer is selected from the group consisting of sodium acetate, sodium carbonate, citrate, glycylglycine, histidine, glycine, lysine, arginine, sodium dihydrogen phosphate, disodium hydrogen phosphate, sodium phosphate, and tris(hydroxymethyl)-aminomethan, bicine, tricine, malic acid, succinate, maleic acid, fumaric

acid, tartaric acid, aspartic acid or mixtures thereof. Each one of these specific buffers constitutes an alternative embodiment of the invention.

In a further embodiment of the invention the formulation further comprises a pharmaceutically acceptable preservative. In a further embodiment of the invention the preservative is selected from the group consisting of phenol, o-cresol, m-cresol, p-cresol, methyl p-hydroxybenzoate, propyl p-hydroxybenzoate, 2-phenoxyethanol, butyl p-hydroxybenzoate, 2-phenylethanol, benzyl alcohol, chlorobutanol, and thiomerosal, bronopol, benzoic acid, imidurea, chlorohexidine, sodium dehydroacetate, chlorocresol, ethyl p-hydroxybenzoate, benzethonium chloride, chlorphenesine (3p-chlorphenoxypropane-1,2-diol) or mixtures thereof. In a further embodiment of the invention the preservative is present in a concentration from 0.1 mg/ml to 20 mg/ml. In a further embodiment of the invention the preservative is present in a concentration from 0.1 mg/ml to 5 mg/ml. In a further embodiment of the invention the preservative is present in a concentration from 5 mg/ml to 10 mg/ml. In a further embodiment of the invention the preservative is present in a concentration from 10 mg/ml to 20 mg/ml. Each one of these specific preservatives constitutes an alternative embodiment of the invention. The use of a preservative in pharmaceutical compositions is well-known to the skilled person. For convenience reference is made to Remington: *The Science and Practice of Pharmacy*, 19<sup>th</sup> edition, 1995.

In a further embodiment of the invention the formulation further comprises an isotonic agent. In a further embodiment of the invention the isotonic agent is selected from the group consisting of a salt (e.g. sodium chloride), a sugar or sugar alcohol, an amino acid (e.g. L-glycine, L-histidine, arginine, lysine, isoleucine, aspartic acid, tryptophan, threonine), an alditol (e.g. glycerol (glycerine), 1,2-propanediol (propyleneglycol), 1,3-propanediol, 1,3-butanediol) polyethyleneglycol (e.g. PEG400), or mixtures thereof. Any sugar such as mono-, di-, or polysaccharides, or water-soluble glucans, including for example fructose, glucose, mannose, sorbose, xylose, maltose, lactose, sucrose, trehalose, dextran, pullulan, dextrin, cyclodextrin, soluble starch, hydroxyethyl starch and carboxymethylcellulose-Na may be used. In some embodiments the sugar additive is sucrose. Sugar alcohol is defined as a C4-C8 hydrocarbon having at least one --OH group and includes, for example, mannitol, sorbitol, inositol, galactitol, dulcitol, xylitol, and arabitol. In some embodiments the sugar alcohol additive is mannitol. The sugars or sugar alcohols mentioned above may be used individually or in combination. There is no fixed limit to the amount used, as long as the sugar or sugar alcohol is soluble in the liquid preparation and does not adversely effect the stabilizing effects achieved using the methods of the invention. In some embodiments, the sugar or sugar alcohol concentration is between about 1 mg/ml and about 150 mg/ml. In a further embodiment of the invention the isotonic agent is present in a concentration from 1 mg/ml to 50 mg/ml. In a further embodiment of the invention the isotonic agent is present in a

concentration from 1 mg/ml to 7 mg/ml. In a further embodiment of the invention the isotonic agent is present in a concentration from 8 mg/ml to 24 mg/ml. In a further embodiment of the invention the isotonic agent is present in a concentration from 25 mg/ml to 50 mg/ml. Each one of these specific isotonic agents constitutes an alternative embodiment of the invention. The use of an isotonic agent in pharmaceutical compositions is well known to the skilled person. For convenience reference is made to Remington: *The Science and Practice of Pharmacy*, 19<sup>th</sup> edition, 1995.

In a further embodiment of the invention the formulation further comprises a chelating agent. In a further embodiment of the invention the chelating agent is selected from salts of ethylenediaminetetraacetic acid (EDTA), citric acid, and aspartic acid, and mixtures thereof. In a further embodiment of the invention the chelating agent is present in a concentration from 0.1 mg/ml to 5 mg/ml. In a further embodiment of the invention the chelating agent is present in a concentration from 0.1 mg/ml to 2 mg/ml. In a further embodiment of the invention the chelating agent is present in a concentration from 2 mg/ml to 5 mg/ml. Each one of these specific chelating agents constitutes an alternative embodiment of the invention. The use of a chelating agent in pharmaceutical compositions is well-known to the skilled person. For convenience reference is made to Remington: *The Science and Practice of Pharmacy*, 19<sup>th</sup> edition, 1995.

In a further embodiment of the invention the formulation further comprises a stabilizer. The use of a stabilizer in pharmaceutical compositions is well-known to the skilled person. For convenience reference is made to Remington: *The Science and Practice of Pharmacy*, 19<sup>th</sup> edition, 1995.

More particularly, compositions of the invention are stabilized liquid pharmaceutical compositions whose therapeutically active components include a polypeptide that possibly exhibits aggregate formation during storage in liquid pharmaceutical formulations. By "aggregate formation" is intended a physical interaction between the polypeptide molecules that results in formation of oligomers, which may remain soluble, or large visible aggregates that precipitate from the solution. By "during storage" is intended a liquid pharmaceutical composition or formulation once prepared, is not immediately administered to a subject. Rather, following preparation, it is packaged for storage, either in a liquid form, in a frozen state, or in a dried form for later reconstitution into a liquid form or other form suitable for administration to a subject. By "dried form" is intended the liquid pharmaceutical composition or formulation is dried either by freeze drying (i.e., lyophilization; see, for example, Williams and Polli (1984) *J. Parenteral Sci. Technol.* 38:48-59), spray drying (see Masters (1991) in *Spray-Drying Handbook* (5th ed; Longman Scientific and Technical, Essex, U.K.), pp. 491-676; Broadhead et al. (1992) *Drug Devel. Ind. Pharm.* 18:1169-1206; and Mumenthaler et

al. (1994) Pharm. Res. 11:12-20), or air drying (Carpenter and Crowe (1988) Cryobiology 25:459-470; and Roser (1991) Biopharm. 4:47-53). Aggregate formation by a polypeptide during storage of a liquid pharmaceutical composition can adversely affect biological activity of that polypeptide, resulting in loss of therapeutic efficacy of the pharmaceutical composition. Furthermore, aggregate formation may cause other problems such as blockage of tubing, membranes, or pumps when the polypeptide-containing pharmaceutical composition is administered using an infusion system.

The pharmaceutical compositions of the invention may further comprise an amount of an amino acid base sufficient to decrease aggregate formation by the polypeptide during storage of the composition. By "amino acid base" is intended an amino acid or a combination of amino acids, where any given amino acid is present either in its free base form or in its salt form. Where a combination of amino acids is used, all of the amino acids may be present in their free base forms, all may be present in their salt forms, or some may be present in their free base forms while others are present in their salt forms. In some embodiments, amino acids to use in preparing the compositions of the invention are those carrying a charged side chain, such as arginine, lysine, aspartic acid, and glutamic acid. Any stereoisomer (i.e., L, D, or DL isomer) of a particular amino acid (e.g. glycine, methionine, histidine, imidazole, arginine, lysine, isoleucine, aspartic acid, tryptophan, threonine and mixtures thereof) or combinations of these stereoisomers, may be present in the pharmaceutical compositions of the invention so long as the particular amino acid is present either in its free base form or its salt form. In some embodiments the L-stereoisomer is used. Compositions of the invention may also be formulated with analogues of these amino acids. By "amino acid analogue" is intended a derivative of the naturally occurring amino acid that brings about the desired effect of decreasing aggregate formation by the polypeptide during storage of the liquid pharmaceutical compositions of the invention. Suitable arginine analogues include, for example, aminoguanidine, ornithine and N-monoethyl L-arginine, suitable methionine analogues include ethionine and buthionine and suitable cysteine analogues include S-methyl-L cysteine. As with the other amino acids, the amino acid analogues are incorporated into the compositions in either their free base form or their salt form. In a further embodiment of the invention the amino acids or amino acid analogues are used in a concentration, which is sufficient to prevent or delay aggregation of the protein.

In a further embodiment of the invention methionine (or other sulphuric amino acids or amino acid analogous) may be added to inhibit oxidation of methionine residues to methionine sulfoxide when the polypeptide acting as the therapeutic agent is a polypeptide comprising at least one methionine residue susceptible to such oxidation. By "inhibit" is intended minimal accumulation of methionine oxidized species over time. Inhibiting methionine oxidation results in greater retention of the polypeptide in its proper molecular

form. Any stereoisomer of methionine (L, D, or DL isomer) or combinations thereof can be used. The amount to be added should be an amount sufficient to inhibit oxidation of the methionine residues such that the amount of methionine sulfoxide is acceptable to regulatory agencies. Typically, this means that the composition contains no more than about 10% to about 30% methionine sulfoxide. Generally, this can be achieved by adding methionine such that the ratio of methionine added to methionine residues ranges from about 1:1 to about 1000:1, such as 10:1 to about 100:1.

In a further embodiment of the invention the formulation further comprises a stabilizer selected from the group of high molecular weight polymers or low molecular compounds. In a further embodiment of the invention the stabilizer is selected from polyethylene glycol (e.g. PEG 3350), polyvinyl alcohol (PVA), polyvinylpyrrolidone, carboxy-/hydroxycellulose or derivatives thereof (e.g. HPC, HPC-SL, HPC-L and HPMC), cyclodextrins, sulphur-containing substances as monothioglycerol, thioglycolic acid and 2-methylthioethanol, and different salts (e.g. sodium chloride). Each one of these specific stabilizers constitutes an alternative embodiment of the invention.

The pharmaceutical compositions may also comprise additional stabilizing agents, which further enhance stability of a therapeutically active polypeptide therein. Stabilizing agents of particular interest to the present invention include, but are not limited to, methionine and EDTA, which protect the polypeptide against methionine oxidation, and a nonionic surfactant, which protects the polypeptide against aggregation associated with freeze-thawing or mechanical shearing.

In a further embodiment of the invention the formulation further comprises a surfactant. In a further embodiment of the invention the surfactant is selected from a detergent, ethoxylated castor oil, polyglycolized glycerides, acetylated monoglycerides, sorbitan fatty acid esters, polyoxypropylene-polyoxyethylene block polymers (eg. poloxamers such as Pluronic® F68, poloxamer 188 and 407, Triton X-100 ), polyoxyethylene sorbitan fatty acid esters, polyoxyethylene and polyethylene derivatives such as alkylated and alkoxyated derivatives (tweens, e.g. Tween-20, Tween-40, Tween-80 and Brij-35), monoglycerides or ethoxylated derivatives thereof, diglycerides or polyoxyethylene derivatives thereof, alcohols, glycerol, lectins and phospholipids (eg. phosphatidyl serine, phosphatidyl choline, phosphatidyl ethanolamine, phosphatidyl inositol, diphosphatidyl glycerol and sphingomyelin), derivatives of phospholipids (eg. dipalmitoyl phosphatidic acid) and lysophospholipids (eg. palmitoyl lysophosphatidyl-L-serine and 1-acyl-sn-glycero-3-phosphate esters of ethanolamine, choline, serine or threonine) and alkyl, alkoxy (alkyl ester), alkoxy (alkyl ether)- derivatives of lysophosphatidyl and phosphatidylcholines, e.g. lauroyl and myristoyl derivatives of lysophosphatidylcholine, dipalmitoylphosphatidylcholine, and modifications of the polar head

group, that is cholines, ethanolamines, phosphatidic acid, serines, threonines, glycerol, inositol, and the positively charged DODAC, DOTMA, DCP, BISHOP, lysophosphatidylserine and lysophosphatidylthreonine, and glycerophospholipids (eg. cephalins), glyceroglycolipids (eg. galactopyransoide), sphingoglycolipids (eg. ceramides, gangliosides),

5 dodecylphosphocholine, hen egg lysolecithin, fusidic acid derivatives- (e.g. sodium tauro-dihydrofusidate etc.), long-chain fatty acids and salts thereof C6-C12 (eg. oleic acid and caprylic acid), acylcarnitines and derivatives, N<sup>α</sup>-acylated derivatives of lysine, arginine or histidine, or side-chain acylated derivatives of lysine or arginine, N<sup>α</sup>-acylated derivatives of dipeptides comprising any combination of lysine, arginine or histidine and a neutral or acidic  
10 amino acid, N<sup>α</sup>-acylated derivative of a tripeptide comprising any combination of a neutral amino acid and two charged amino acids, DSS (docusate sodium, CAS registry no [577-11-7]), docusate calcium, CAS registry no [128-49-4]), docusate potassium, CAS registry no [7491-09-0]), SDS (sodium dodecyl sulfate or sodium lauryl sulfate), sodium caprylate, cholic acid or derivatives thereof, bile acids and salts thereof and glycine or taurine  
15 conjugates, ursodeoxycholic acid, sodium cholate, sodium deoxycholate, sodium taurocholate, sodium glycocholate, N-Hexadecyl-N,N-dimethyl-3-ammonio-1-propanesulfonate, anionic (alkyl-aryl-sulphonates) monovalent surfactants, zwitterionic surfactants (e.g. N-alkyl-N,N-dimethylammonio-1-propanesulfonates, 3-cholamido-1-propyldimethylammonio-1-propanesulfonate, cationic surfactants (quaternary ammonium  
20 bases) (e.g. cetyl-trimethylammonium bromide, cetylpyridinium chloride), non-ionic surfactants (eg. Dodecyl β-D-glucopyranoside), poloxamines (eg. Tetronic's), which are tetrafunctional block copolymers derived from sequential addition of propylene oxide and ethylene oxide to ethylenediamine, or the surfactant may be selected from the group of imidazoline derivatives, or mixtures thereof. Each one of these specific surfactants constitutes  
25 an alternative embodiment of the invention.

The use of a surfactant in pharmaceutical compositions is well known to the skilled person. For convenience reference is made to Remington: *The Science and Practice of Pharmacy*, 19<sup>th</sup> edition, 1995.

It is possible that other ingredients may be present in the peptide pharmaceutical formulation  
30 of the present invention. Such additional ingredients may include wetting agents, emulsifiers, antioxidants, bulking agents, tonicity modifiers, chelating agents, metal ions, oleaginous vehicles, proteins (e.g., human serum albumin, gelatine or proteins) and a zwitterion (e.g., an amino acid such as betaine, taurine, arginine, glycine, lysine and histidine). Such additional ingredients, of course, should not adversely affect the overall stability of the  
35 pharmaceutical formulation of the present invention.

Pharmaceutical compositions containing a VAR2CSA polypeptide according to the present

invention may be administered to a patient in need of such treatment at several sites, for example, at topical sites, for example, skin and mucosal sites, at sites which bypass absorption, for example, administration in an artery, in a vein, in the heart, and at sites which involve absorption, for example, administration in the skin, under the skin, in a muscle or in the abdomen.

Topical administration may be a particular advantage in the treatment of conditions associated with local inflammation, such as in the treatment of inflammation associated with burn or other conditions associated with the skin. Accordingly, in some embodiments administration is by topical administration.

- 10 In some particular embodiments, eye droplets may be used in conditions associated with the eye, such as keratitis, such as diffuse lamellar keratitis (DLK).

Administration of pharmaceutical compositions according to the invention may be through several routes of administration, for example, lingual, sublingual, buccal, in the mouth, oral, in the stomach and intestine, nasal, pulmonary, for example, through the bronchioles and alveoli or a combination thereof, epidermal, dermal, transdermal, vaginal, rectal, ocular, for examples through the conjunctiva, uretal, and parenteral to patients in need of such a treatment.

Compositions of the current invention may be administered in several dosage forms, for example, as solutions, suspensions, emulsions, microemulsions, multiple emulsion, foams, 20 salves, pastes, plasters, ointments, tablets, coated tablets, rinses, capsules, for example, hard gelatine capsules and soft gelatine capsules, suppositories, rectal capsules, drops, gels, sprays, powder, aerosols, inhalants, eye drops, ophthalmic ointments, ophthalmic rinses, vaginal pessaries, vaginal rings, vaginal ointments, injection solution, in situ transforming solutions, for example in situ gelling, in situ setting, in situ precipitating, in situ 25 crystallization, infusion solution, and implants.

Compositions of the invention may further be compounded in, or attached to, for example through covalent, hydrophobic and electrostatic interactions, a drug carrier, drug delivery system and advanced drug delivery system in order to further enhance stability of the VAR2CSA polypeptide, increase bioavailability, increase solubility, decrease adverse effects, 30 achieve chronotherapy well known to those skilled in the art, and increase patient compliance or any combination thereof. Examples of carriers, drug delivery systems and advanced drug delivery systems include, but are not limited to, polymers, for example cellulose and derivatives, polysaccharides, for example dextran and derivatives, starch and derivatives, poly(vinyl alcohol), acrylate and methacrylate polymers, polylactic and polyglycolic acid and

block co-polymers thereof, polyethylene glycols, carrier proteins, for example albumin, gels, for example, thermogelling systems, for example block co-polymeric systems well known to those skilled in the art, micelles, liposomes, microspheres, nanoparticulates, virus like particles, bacteria like particles, liquid crystals and dispersions thereof, L2 phase and  
5 dispersions thereof, well known to those skilled in the art of phase behaviour in lipid-water systems, polymeric micelles, multiple emulsions, self-emulsifying, self-microemulsifying, cyclodextrins and derivatives thereof, and dendrimers.

Compositions of the current invention are useful in the formulation of solids, semisolids, powder and solutions for pulmonary administration of the VAR2CSA polypeptide, using, for  
10 example a metered dose inhaler, dry powder inhaler and a nebulizer, all being devices well known to those skilled in the art.

Compositions of the current invention are specifically useful in the formulation of controlled, sustained, protracting, retarded, and slow release drug delivery systems. More specifically, but not limited to, compositions are useful in formulation of parenteral controlled release and  
15 sustained release systems (both systems leading to a many-fold reduction in number of administrations), well known to those skilled in the art. Even more preferably, are controlled release and sustained release systems administered subcutaneous. Without limiting the scope of the invention, examples of useful controlled release system and compositions are hydrogels, oleaginous gels, liquid crystals, polymeric micelles, microspheres, nanoparticles,

20 Methods to produce controlled release systems useful for compositions of the current invention include, but are not limited to, crystallization, condensation, co-crystallization, precipitation, co-precipitation, emulsification, dispersion, high pressure homogenisation, encapsulation, spray drying, microencapsulating, coacervation, phase separation, solvent evaporation to produce microspheres, extrusion and supercritical fluid processes. General  
25 reference is made to Handbook of Pharmaceutical Controlled Release (Wise, D.L., ed. Marcel Dekker, New York, 2000) and Drug and the Pharmaceutical Sciences vol. 99: Protein Formulation and Delivery (MacNally, E.J., ed. Marcel Dekker, New York, 2000).

Parenteral administration may be performed by subcutaneous, intramuscular, intraperitoneal or intravenous injection by means of a syringe, optionally a pen-like syringe. Alternatively,  
30 parenteral administration can be performed by means of an infusion pump. A further option is a composition, which may be a solution or suspension for the administration of the VAR2CSA polypeptide in the form of a nasal or pulmonal spray. As a still further option, the pharmaceutical compositions containing the VAR2CSA polypeptide of the invention can also be adapted to transdermal administration, *e.g.* by needle-free injection or from a patch,  
35 optionally an iontophoretic patch, or transmucosal, *e.g.* buccal, administration.

The term "stabilized formulation" refers to a formulation with increased physical stability, increased chemical stability or increased physical and chemical stability.

The term "physical stability" of the protein formulation as used herein refers to the tendency of the protein to form biologically inactive and/or insoluble aggregates of the protein as a result of exposure of the protein to thermo-mechanical stresses and/or interaction with interfaces and surfaces that are destabilizing, such as hydrophobic surfaces and interfaces. Physical stability of the aqueous protein formulations is evaluated by means of visual inspection and/or turbidity measurements after exposing the formulation filled in suitable containers (e.g. cartridges or vials) to mechanical/physical stress (e.g. agitation) at different temperatures for various time periods. Visual inspection of the formulations is performed in a sharp focused light with a dark background. The turbidity of the formulation is characterized by a visual score ranking the degree of turbidity for instance on a scale from 0 to 3 (a formulation showing no turbidity corresponds to a visual score 0, and a formulation showing visual turbidity in daylight corresponds to visual score 3). A formulation is classified physical unstable with respect to protein aggregation, when it shows visual turbidity in daylight. Alternatively, the turbidity of the formulation can be evaluated by simple turbidity measurements well known to the skilled person. Physical stability of the aqueous protein formulations can also be evaluated by using a spectroscopic agent or probe of the conformational status of the protein. The probe is preferably a small molecule that preferentially binds to a non-native conformer of the protein. One example of a small molecular spectroscopic probe of protein structure is Thioflavin T. Thioflavin T is a fluorescent dye that has been widely used for the detection of amyloid fibrils. In the presence of fibrils, and perhaps other protein configurations as well, Thioflavin T gives rise to a new excitation maximum at about 450 nm and enhanced emission at about 482 nm when bound to a fibril protein form. Unbound Thioflavin T is essentially non-fluorescent at the wavelengths.

Other small molecules can be used as probes of the changes in protein structure from native to non-native states. For instance the "hydrophobic patch" probes that bind preferentially to exposed hydrophobic patches of a protein. The hydrophobic patches are generally buried within the tertiary structure of a protein in its native state, but become exposed as a protein begins to unfold or denature. Examples of these small molecular, spectroscopic probes are aromatic, hydrophobic dyes, such as anthracene, acridine, phenanthroline or the like. Other spectroscopic probes are metal-amino acid complexes, such as cobalt metal complexes of hydrophobic amino acids, such as phenylalanine, leucine, isoleucine, methionine, and valine, or the like.

The term "chemical stability" of the protein formulation as used herein refers to chemical covalent changes in the protein structure leading to formation of chemical degradation

products with potential less biological potency and/or potential increased immunogenic properties compared to the native protein structure. Various chemical degradation products can be formed depending on the type and nature of the native protein and the environment to which the protein is exposed. Elimination of chemical degradation can most probably not be completely avoided and increasing amounts of chemical degradation products is often seen during storage and use of the protein formulation as well-known by the person skilled in the art. Most proteins are prone to deamidation, a process in which the side chain amide group in glutaminy or asparaginy residues is hydrolysed to form a free carboxylic acid. Other degradation pathways involves formation of high molecular weight transformation products where two or more protein molecules are covalently bound to each other through transamidation and/or disulfide interactions leading to formation of covalently bound dimer, oligomer and polymer degradation products (*Stability of Protein Pharmaceuticals*, Ahern. T.J. & Manning M.C., Plenum Press, New York 1992). Oxidation (of for instance methionine residues) can be mentioned as another variant of chemical degradation. The chemical stability of the protein formulation can be evaluated by measuring the amount of the chemical degradation products at various time-points after exposure to different environmental conditions (the formation of degradation products can often be accelerated by for instance increasing temperature). The amount of each individual degradation product is often determined by separation of the degradation products depending on molecule size and/or charge using various chromatography techniques (e.g. SEC-HPLC and/or RP-HPLC).

Hence, as outlined above, a "stabilized formulation" refers to a formulation with increased physical stability, increased chemical stability or increased physical and chemical stability. In general, a formulation must be stable during use and storage (in compliance with recommended use and storage conditions) until the expiration date is reached.

In some embodiments of the invention the pharmaceutical formulation comprising the VAR2CSA polypeptide is stable for more than 6 weeks of usage and for more than 3 years of storage. In other embodiments of the invention the pharmaceutical formulation comprising the VAR2CSA polypeptide is stable for more than 4 weeks of usage and for more than 3 years of storage. In a further embodiment of the invention the pharmaceutical formulation comprising the VAR2CSA polypeptide is stable for more than 4 weeks of usage and for more than two years of storage. In an even further embodiment of the invention the pharmaceutical formulation comprising the VAR2CSA polypeptide is stable for more than 2 weeks of usage and for more than two years of storage.

#### **Indications for use of VAR2CSA polypeptide and conjugates thereof**

The VAR2CSA polypeptides or conjugates thereof may be used in a wide range of indications associated with expression, such as inappropriate expression of CSA, such as in various cancers, such as metastatic cancers including melanomas, such as C32 melanoma, sarcomas, lung carcinomas, oligodendrocytomas, human brain tumours including gliomas, leukaemia, such as lymphoblastic leukemia and acute myeloid leukemia, and carcinoma, such as squamous cell carcinomas and breast carcinomas, renal cell carcinomas, chondrosarcomas, and pancreatic cell carcinomas. The VAR2CSA polypeptides or conjugates thereof may also be used for cancer stem cells and accordingly target the cells before development into a cancer. Other conditions associated with expression, such as inappropriate expression of CSA are conditions of the cartilage and/or the development of scar tissue.

The VAR2CSA polypeptides or conjugates thereof may be used in identifying, tracking and targeting distant micro-metastasis in vivo. Virtually all primary tumours, including cancers of the hematopoietic system, have the potential of developing into metastatic disease, which is highly associated with poor therapeutic outcome of the patients.

The VAR2CSA polypeptides or conjugates thereof may be used to target compounds that prevent degradation of or repair extracellular CSPG such as growth hormones, anti-inflammatory compounds or protein inhibitors, to cartilage tissue, joints, and neural tissue.

The VAR2CSA polypeptides or conjugates thereof may be used to target compounds that enhance degradation or prevent production of extracellular CSPG such as chondroitinase ABC, which cut the sugar chains of the protein core of CSPG molecules. Xylocides, which reduce CSPG production, or drugs that inhibit enzymes important for CSPG production such as chondroitin synthase or chondroitin polymerizing factor (such as 4-fluoro-glucosamine, p-nitrophenyl-beta-D-xyloside, 4-methyl-umbelliferyl-beta-D-xylopyranoside), to damaged neural tissue.

VAR2CSA conjugated to a nucleic acid, here among small interfering RNA (siRNA), antisense peptide nucleic acids (PNA), small hairpin RNA (shRNA) and locked nucleic acids (LNA<sup>TM</sup>), can be used to remove RNA encoding CSA presenting molecules.

### **Conjugates of VAR2CSA polypeptide**

### **Therapeutic or diagnostic effector moiety, such as cytotoxic and detecting moieties**

In some aspects of the present invention, there are provided VAR2CSA polypeptides, fusion protein or conjugate as defined in the present disclosure, further comprising a therapeutic

effector moiety, such as an inflammatory agent, a steroid hormone, a cytotoxic or detecting agent or moiety, such as an organic molecule, radionuclide, or cytotoxic enzyme.

In some aspects of the present invention, the VAR2CSA polypeptide or VAR2CSA fusion protein according to the present invention comprises a sequence as defined by one or more sequences selected from SEQ ID NO 57-59, and 71 or a functional variant or fragment thereof.

In some embodiments the VAR2CSA polypeptide or VAR2CSA fusion protein according to the present invention may comprise a protease inhibitor, such as basic pancreatic trypsin inhibitor (BPTI) in the terminal, such as the N-terminal of the protein sequence, such as a sequence defined by SEQ ID NO:57.

In some embodiments the VAR2CSA polypeptide or VAR2CSA fusion protein according to the present invention may comprise a toxin protein sequence, such as a sequence as defined by one or more sequences selected from SEQ ID NO 58, 59 and 71, such as a toxin protein sequence have an optimized to be less immunogenic, such as a sequence defined by SEQ ID NO:59. In some embodiments the signal sequence KDEL of SEQ ID NO 58 or 59 is present in a VAR2CSA fusion protein according to the present invention and in some embodiments the signal sequence KDEL of SEQ ID NO 58 or 59 is absent in a VAR2CSA fusion protein according to the present invention. Accordingly, the signal sequence KDEL may be optional for the constructs according to the present invention.

Non-limiting examples of cytotoxic moieties which may be fused or conjugated to VAR2CSA polypeptides according to the invention, are chemotherapeutics selected from calicheamycin, cisplatin, adriamycin, auristatin, doxorubicin, maytansinoid, taxol, ecteinascidin, geldanamycin, methotrexate and their derivatives, and combinations thereof and the like suitable for cancer therapy. Examples of cytotoxic proteins fused to VAR2CSA polypeptides are *Pseudomonas* exotoxin A, diphtheria toxin, ricin toxin, pokeweed antiviral protein, saporin, gelonin and variants hereof.

Conjugates of albumin with doxorubicin for use in cancer have been described (Kratz et al, J Med Chem 45: 5523-33, 2002) and with metotrexate in rheumatoid arthritis (Wunder et al, J Immunol 170:4793-4801, 2003). Compounds that increase reactive oxygen species, i.e. Piperlongumine have also been described (Raj et al, Nature 475: 231-234, 2011). Also, therapeutic enzymes, agents that induce apoptosis, and the like in order to provide for targeted cytotoxicity, i.e. killing of tumor cells, may be used.

The VAR2CSA polypeptides described herein may mediate killing of cells by inducing complement dependent cytotoxicity (CDC) mediated lysis, antibody dependent cellular cytotoxicity (ADCC) mediated lysis, apoptosis, homotypic adhesion, and/or phagocytosis, such as by inducing CDC mediated lysis and/or ADCC mediated lysis. The VAR2CSA

5 polypeptides described herein may interact with components of the immune system, preferably through ADCC or CDC. However, VAR2CSA polypeptides of the invention may also exert an effect simply by binding to tumor antigens on the cell surface, thus, e.g. blocking proliferation of the cells.

According to the invention, the term "therapeutic effector moiety" means any molecule, which may exert a therapeutic effect. According to the invention, a therapeutic effector molecule is preferably selectively guided to a cell, which expresses CSA and includes anticancer agents, radioisotopes, toxins, cytostatic or cytolytic drugs, etc. Anticancer agents comprise, for example, Anthracyclins (doxorubicin, daunorubicin, epirubicin, idarubicin, valrubicin, , mitoxantrone), Platinum and non-platinum based alkylating agents (cisplatin, carboplatin, oxaliplatin, mechlorethamine, cyclophosphamide, chlorambucil, ifosfamide, busulfan, carmustine, dacarbazine, lomustine, procarbazine), Vinca alkaloids (vincristine, vinblastine, vinorelbine, vindesine), Taxanes (taxol and decetaxel), Topoisomerase I inhibitors (camptothecin, irinotecan, topotecan), Topoisomerase II inhibitors (amsacrine, etoposide, etoposide phosphate, teniposide and other alkaloid-derivates naturally occurring in the root of American Mayapple (*Podophyllum peltatum*)), Non-anthracyclin cytotoxic antibiotics (dactinomycin, bleomycin, plicamycin and mitomycins), Anti-steroids (such as aminoglutethimide), Nucleoside analogues (cytarabidine, fluorouracil and mercaptopurine), Antimetabolites (methotrexate and thioguanine), dichlorodiphenyltrichloroethane analogues (like mitotane), and reactive oxygen species (ROS)-inducing compounds (including but not limited to piperlongumine, and beta-phenylethyl isothiocyanate). Other anticancer agents are described, for example, in Goodman and Gilman, "The Pharmacological Basis of Therapeutics", 8th Edition, 1990, McGraw-Hill, Inc., in particular Chapter 52 (Antineoplastic Agents (Paul Calabresi and Bruce A. Chabner). Toxins may be proteins such as pokeweed antiviral protein, cholera toxin, pertussis toxin, ricin, gelonin, abrin, diphtheria exotoxin or Pseudomonas exotoxin. Toxin residues may also be high energy-emitting radionuclides such as cobalt-60. A VAR2CSA polypeptide may be used together with cell-penetrating peptides (CPP) to facilitate transport of the VAR2CSA polypeptide and any thereto-linked molecule across cell plasma membranes. Cell-penetrating peptides have found numerous applications in medicine as drug delivery agents in the treatment of different diseases including cancer and virus inhibitors. Examples on CPP include but are not limited to: trans-activating transcriptional activator (Tat) from human immunodeficiency virus; pep-1 (Chariot<sup>TM</sup>); R8, azo-R8; SMOc. (Okuyama M et al. Nat Methods. 2007 Feb;4(2):153-9M; Soane L and Fiskum

GJ Neurochem. 2005 Oct;95(1):230-43; Loudet A et al. Org Biomol Chem. 2008 Dec 21;6(24):4516-22).

#### Radionuclides

- 5 A VAR2CSA polypeptides, a fusion protein or conjugate according to the aspects described herein coupled to a polyaminopolycarboxylate chelator may be used to provide a radiolabeled polypeptide consisting of a radiochelate of the VAR2CSA polypeptide, fusion protein or conjugate coupled to the chelator and a radionuclide suitable for medical imaging, the radionuclide being selected from the group consisting of  $^{61}\text{Cu}$ ,  $^{64}\text{Cu}$ ,  $^{66}\text{Ga}$ ,  $^{67}\text{Ga}$ ,  $^{68}\text{Ga}$ ,  $^{110}\text{In}$ ,  
10  $^{111}\text{In}$ ,  $^{44}\text{Sc}$ ,  $^{89}\text{Zr}$  and  $^{86}\text{Y}$ , or with a radionuclide suitable for therapy, the radionuclide being selected from the group consisting of  $^{225}\text{Ac}$ ,  $^{212}\text{Bi}$ ,  $^{213}\text{Bi}$ ,  $^{67}\text{Cu}$ ,  $^{166}\text{Ho}$ ,  $^{177}\text{Lu}$ ,  $^{212}\text{Pb}$ ,  $^{149}\text{Pm}$ ,  $^{153}\text{Sm}$ ,  $^{227}\text{Th}$  and  $^{90}\text{Y}$ , wherein the radionuclide is complexed with the VAR2CSA polypeptide, such as via a chelator.

- Accordingly VAR2CSA polypeptides, a fusion protein or conjugate according to the aspects  
15 described herein may be used for radioimaging of cancer cells, including solid tumors or metastases, such as in melanoma patients.

- In embodiments thereof, the polypeptide may also be radiolabeled with non-metal radioisotopes using so called indirect labelling. Thus, for labelling with for example  $^{18}\text{F}$ ,  $^{76}\text{Br}$ ,  
20 different iodine isotopes and  $^{211}\text{At}$ , intermediate "linker molecules" are used for labelling. Such a linker molecule should contain two functional moieties, one providing rapid and efficient radiolabeling, and another enabling rapid and efficient coupling to the proteins, e.g. to amine groups, or preferably to the thiol group of a unique cysteine. For example a maleimide group reacts with thiol groups to form a stable thioether bond. The "linker  
25 molecule" may first be reacted with the radiolabel and subsequently with the thiol or the selenothiol group of the protein.

- Other alternative detecting moieties includes fluorophores or fluorochromes such as any one selected from Hydroxycoumarin, Aminocoumarin, Methoxycoumarin, Cascade Blue, Pacific Blue, Pacific Orange, Lucifer yellow, NBD, R-Phycoerythrin (PE), PE-Cy5 conjugates, PE-Cy7  
30 conjugates, Red 613, PerCP, TruRed, FluorX, Fluorescein, BODIPY-FL, TRITC, X-Rhodamine, Lissamine Rhodamine B, Texas Red, Allophycocyanin (APC), and APC-Cy7 conjugates.

Such conjugates with detecting moieties include fluorophores or fluorochromes may be used for imaging of cancer cells or tumors.

#### Steroid hormones or anti-inflammatory agents

In some embodiments according to the invention the VAR2CSA polypeptides are conjugated with an anti-inflammatory agent, including steroid hormones.

Cartilage and scar tissue is known to contain CSPG in high amounts. Accordingly, it may be attractive to direct anti-inflammatory agents such as non-steroid anti-inflammatory compounds, disease modifying anti-rheumatic drugs (such as methotrexate, azathioprine, sulfasalazine, ciclosporine, penicillamine, leflunomide, or gold), biological anti-rheumatic drugs (such as Tumor Necrosis Factor inhibitors, interleukin-1-receptor antagonists, CD20-antibody, Insulin Growth Factor 1) and steroid hormones or alternative compounds to such tissues.

- 10 In some embodiments according to the invention the VAR2CSA polypeptides are conjugated with an anti-inflammatory agent, such as non-steroid anti-inflammatory compounds, disease modifying anti-rheumatic drugs (such as methotrexate, azathioprine, sulfasalazine, ciclosporine, penicillamine, leflunomide, or gold), biological anti-rheumatic drugs (such as Tumor Necrosis Factor inhibitors, interleukin-1-receptor antagonists, CD20-antibody, Insulin Growth Factor 1) and steroid hormones or alternative compounds to such tissues.
- 15

#### **Conjugates with CSPG4**

In some embodiments according to the invention the VAR2CSA polypeptides are conjugated with CSPG4.

- It is assumed that conjugates of VAR2CSA polypeptides with CSPG4 may be used as an immunization agent. For the purpose of this use, it is assumed that the VAR2CSA polypeptide may function as a chaperone that could facilitate a display of CSPG4 to T-cells in a conformation that would provide antibodies. Accordingly, it is assumed that VAR2CSA polypeptides conjugated with CSPG4 may be used in a vaccine.
- 20

- As used herein the term "CSPG4" refers to the 2322 amino acid full length human Chondroitin sulfate proteoglycan 4 identified by Uniprot as Q6UVK1 (CSPG4\_HUMAN) as well as variants, functional fragments, and orthologs thereof. CSPG4 may also be referred to as melanoma-associated chondroitin sulfate proteoglycan (MCSP), High Molecular Weight-Melanoma Associated Antigen (HMW-MAA) or neuron-glial antigen 2 (NG2).
- 25

#### **Targeting of CD44 or other proteoglycans**

- For the purpose of use of conjugates of VAR2CSA polypeptides in the treatment of cancer indications, it is assumed that the conjugates according to the present invention may be used
- 30

to target not only CSPG4 expressing tumor cells, but also CD44 expressing cells, such as cancer stem cells, and cells expressing proteoglycans exemplified but not limited to those of Table 1. This targeting is mediated through the binding to CSA on the CD44 antigen.

Accordingly, the conjugates according to the present invention may be used to target CSPG4 negative but CD44 positive cells. This may be used as an alternative to or simultaneously with the targeting of CSPG4 expressing tumor cells.

**Use in isolation of cancer stem cells through binding to CD44, and/or CSPG4, and/or other proteoglycans, such as those in Table 1.**

The specific and high affinity binding of the VAR2CSA polypeptides according to the present invention, such as in the form of conjugates of VAR2CSA polypeptides, may be used to isolate stem cells, such as cancer stem cells expressing CD44 and/or CSPG4.

**Use in isolation or detection of circulating tumor cells (CTC) through binding to CSA-containing proteoglycans**

The specific and high affinity binding of the VAR2CSA polypeptides according to the present invention, such as in the form of conjugates of VAR2CSA polypeptides, may be used to isolate or detect CTCs of epithelial and non-epithelial origin, which express one or more CSA-containing proteoglycans, such as those described in Table 1.

**Anti-idiotypic antibodies**

As an alternative or supplement to the use of VAR2CSA polypeptides, it is also possible to use anti-idiotypic antibodies or even mimotopes that mimic VAR2CSA. The technologies for preparing anti-idiotypic antibodies that mimic an antigen epitope are known in the art and entail provision of a first monoclonal antibody binding VAR2CSA polypeptides followed by subsequent production of a second antibody that binds the idiotype of said first antibody. Mimotopes can be isolated from libraries of random peptides that are screened in phage display against antibodies that bind VAR2CSA polypeptides specifically.

Anti-idiotypic antibodies may also be prepared by immunization with inhibitory host or patient derived antibodies against VAR2CSA in order to obtain and screen for polyclonal and/or monoclonal antibodies, such as human antibodies against and inhibiting the host derived antibodies. Although VAR2CSA generally is an evolutionary refined malaria protein unlikely to cause autoimmune reactions in the patient, such an immune reaction cannot be excluded after a period of treatment. An anti-idiotypic antibody used in combination with or as an alternative to VAR2CSA polypeptides may then be used.

*Specific embodiments of the invention*

As described herein the present invention relates to an isolated protein fragment of VAR2CSA, which fragment consist of a sequential amino acid sequence of

- a) ID1, and
- b) DBL2Xb, and optionally
- c) ID2a.

In some embodiments the isolated protein fragment of VAR2CSA according to the present invention comprises ID2a.

- 10 In some embodiments the isolated protein fragment of VAR2CSA according to the present invention do not comprise ID2a.

In some embodiments the isolated protein fragment of VAR2CSA according to the present invention further comprises an amino acid sequence in the N- or C-terminal, or within the sequence of the protein fragment of VAR2CSA of not more than 100 amino acids, such as not  
15 more than 90 amino acids, such as not more than 80 amino acids, such as not more than 70 amino acids, such as not more than 60 amino acids, such as not more than 50 amino acids, such as not more than 40 amino acids, such as not more than 30 amino acids, such as not more than 20 amino acids, such as not more than 18 amino acids, such as not more than 16 amino acids, such as not more than 14 amino acids, such as not more than 12 amino acids,  
20 such as not more than 10 amino acids, such as not more than 8 amino acids, such as not more than 6 amino acids, such as not more than 4 amino acids, such as not more than 2 amino acids derived from any part of a VAR2CSA polypeptide as defined herein, which is not part of ID1, DBL2Xb, or ID2a.

25 In some embodiments the isolated protein fragment of VAR2CSA according to the present invention further comprises an amino acid sequence in the N- or C-terminal, or within the sequence of the protein fragment of VAR2CSA of not more than 100 amino acids, such as not more than 90 amino acids, such as not more than 80 amino acids, such as not more than 70 amino acids, such as not more than 60 amino acids, such as not more than 50 amino acids,  
30 such as not more than 40 amino acids, such as not more than 30 amino acids, such as not more than 20 amino acids, such as not more than 18 amino acids, such as not more than 16 amino acids, such as not more than 14 amino acids, such as not more than 12 amino acids, such as not more than 10 amino acids, such as not more than 8 amino acids, such as not more than 6 amino acids, such as not more than 4 amino acids, such as not more than 2  
35 amino acids, which amino acid sequence is not derived from any part of a VAR2CSA polypeptide as defined herein.

In some embodiments the protein fragment according to the present invention binds chondroitin sulfate A (CSA) on proteoglycans (CSPG) with an affinity as measured by a  $K_D$  lower than 100 nM, such as lower than 80 nM, such as lower than 70 nM, such as lower than 60 nM, such as lower than 50 nM, such as lower than 40 nM, such as lower than 30 nM, such as lower than 26 nM, such as lower than 24 nM, such as lower than 22 nM, such as lower than 20 nM, such as lower than 18 nM, such as lower than 16 nM, such as lower than 14 nM, such as lower than 12 nM, such as lower than 10 nM, such as lower than 9 nM, such as lower than 8 nM, such as lower than 7 nM, such as lower than 6 nM, or lower than 4 nM.

- 10 In some embodiments the protein fragment according to the present invention comprises an amino acid sequence having at least 70 % sequence identity with any one amino acid sequence of 1-577 of SEQ ID NO:1, 1-592 of SEQ ID NO:3, 1-579 of SEQ ID NO:4, 1-576 of SEQ ID NO:5, 1-586 of SEQ ID NO:10, 1-579 of SEQ ID NO:11, 1-565 of SEQ ID NO:29, 1-584 of SEQ ID NO:34, 1-569 of SEQ ID NO:36, 1-575 of SEQ ID NO:37, 1-592 of SEQ ID NO:38, 1-603 of SEQ ID NO:41, 1-588 of SEQ ID NO:43, 1-565 of SEQ ID NO:44, 1-589 of SEQ ID NO:45, 1-573 of SEQ ID NO:48, 1-583 of SEQ ID NO:53, or 1-569 of SEQ ID NO:54.

- 20 In some embodiments the protein fragment according to the present invention comprises an amino acid sequence having at least 70 % sequence identity with an amino acid sequence of 578-640 of SEQ ID NO:1, 593-656 of SEQ ID NO:3, 580-643 of SEQ ID NO:4, 577-640 of SEQ ID NO:5, 587-650 of SEQ ID NO:10, 580-643 of SEQ ID NO:11, 566-628 of SEQ ID NO:29, 585-647 of SEQ ID NO:34, 570-632 of SEQ ID NO:36, 576-639 of SEQ ID NO:37, 593-655 of SEQ ID NO:38, 604-667 of SEQ ID NO:41, 589-652 of SEQ ID NO:43, 566-628 of SEQ ID NO:44, 590-653 of SEQ ID NO:45, 574-637 of SEQ ID NO:48, 584-646 of SEQ ID NO:53, or 570-632 of SEQ ID NO:54.

- 25 In some embodiments the protein fragment according to the present invention comprises an amino acid sequence having at least 70 % sequence identity with an amino acid sequence of SEQ ID NO:2, 6, 8, 9, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 30, 31, 32, 33, 35, 39, 40, 42, 46, 47, 49, 50, 51, or 52.

- 30 In some embodiments the protein fragment according to the present invention consists of an amino acid sequence having at least 70 % sequence identity with any one amino acid sequence of 1-577 of SEQ ID NO:1, 1-592 of SEQ ID NO:3, 1-579 of SEQ ID NO:4, 1-576 of SEQ ID NO:5, 1-586 of SEQ ID NO:10, 1-579 of SEQ ID NO:11, 1-565 of SEQ ID NO:29, 1-584 of SEQ ID NO:34, 1-569 of SEQ ID NO:36, 1-575 of SEQ ID NO:37, 1-592 of SEQ ID NO:38, 1-603 of SEQ ID NO:41, 1-588 of SEQ ID NO:43, 1-565 of SEQ ID NO:44, 1-589 of SEQ ID NO:45, 1-573 of SEQ ID NO:48, 1-583 of SEQ ID NO:53, or 1-569 of SEQ ID NO:54.

In some embodiments the protein fragment according to the present invention consists of an amino acid sequence selected from the list consisting of SEQ ID NO:1, 3-5, 10, 11, 29, 34, 36-38, 41, 43-45, 48, 53, and 54.

5 In some embodiments the protein fragment according to the present invention fragment consists of an amino acid sequence having a length of less than 700 amino acids, such as less than 690 amino acids, such as less than 680 amino acids, such as less than 670 amino acids, such as less than 660 amino acids, such as less than 650 amino acids, such as less than 640 amino acids, such as less than 630 amino acids, such as less than 620 amino acids, such as less than 610 amino acids, such as less than 600 amino acids, such as less than 590  
10 amino acids, such as less than 580 amino acids, such as less than 570 amino acids.

In some embodiments the protein fragment according to the present invention is substantially pure.

In some embodiments the protein fragment according to the present invention has a molecular mass of less than about 100 kDa under non-reducing conditions on an SDS-PAGE.

15 In some embodiments the protein fragment according to the present invention is a recombinant protein.

In some embodiments the protein fragment according to the present invention is non-glycosylated.

20 The invention further relates to a protein fragment as defined herein, a VAR2CSA polypeptide, or a conjugate according to the invention for the treatment of any indications associated with a condition involving expression, such as inappropriate expression of CSA, such as in cancer, arthritis, multiple sclerosis and healing after neural damage, cartilage repair, wound healing, and in psoriasis.

25 In some embodiments, a VAR2CSA polypeptide, conjugate or fusion protein is or comprises a protein fragment of VAR2CSA according to the present invention.

Accordingly, a VAR2CSA polypeptide, conjugate or fusion protein according to the present invention may comprise an amino acid sequence with at least 70 % sequence identity to an amino acid sequence identified by any sequence of SEQ ID NO: 1-75.

30 In some embodiments a VAR2CSA polypeptide according to the present invention consist of an amino acid sequence selected from SEQ ID NO: 60-70, 72-75.

In some embodiments, the cancer is selected from Cutaneous, Ocular or Conjunctival melanoma. Carcinomas (Triple negative- and metaplastic breast carcinoma, Pancreatic carcinoma, Ovarian carcinoma, Endometrial carcinoma, Hepatocellular carcinoma, Lung carcinoma, Colon carcinoma, Prostate carcinoma, Cervix carcinoma, Testis carcinoma, Basal cell skin carcinoma, Clear cell renal cell carcinoma, Kراتينized head and neck squamous cell carcinoma, Skin squamous cell carcinoma, Vulvar kراتينized squamous cell carcinoma and Vulvar basal cell carcinoma), sarcomas (Breast liposarcoma, Fibrosarcoma, Dedifferentiated chondro- and liposarcoma, Leiomyosarcoma, Liposarcoma, Myxoid liposarcoma, Uterine corpus leiomyosarcoma, Osteosarcoma, Ewing sarcoma and Rhabdomyosarcoma), hematopoietic cancers (Chronic lymphatic leukaemia (CLL), Acute lymphatic leukaemia (ALL), Acute myeloid leukaemia (AML), B-cell, T-cell and large granular lymphoma), tumours of neuroepithelial tissue, such as Astrocytomas (Pleomorphic Xanthoastrocytoma, Fibrillary Astrocytomas, Anaplastic astrocytoma, Glioblastoma Multiforme), Oligodendroglioma, Ependymoma, Choroid plexus tumor, Oligoastrocytoma, gliosarcoma, Ganglioglioma, Retinoblastoma, Neurocytoma, Neuroblastomas (Esthesioneuroblastoma and Ganglioneuroblastoma), Medulloblastoma and Atypical Teratoid Rhabdoid tumors, and any other CSA-expressing cancer subtype.

In some embodiments, the cancer is selected from all CSA-expressing malignancies including carcinomas (including but not limited to Breast carcinoma, Pancreatic carcinoma, Ovarian carcinoma, Endometrial carcinoma, Hepatocellular carcinoma, Lung carcinoma, Colon carcinoma, Prostate carcinoma, Cervix carcinoma, Testis carcinoma, Basal cell skin carcinoma, Clear cell renal cell carcinoma, Head and neck squamous cell carcinoma, Skin squamous cell carcinoma, Vulvar kراتينized squamous cell carcinoma and Vulvar basal cell carcinoma), sarcomas (including but not limited to Fibrosarcoma, Dedifferentiated chondro- and liposarcoma, Leiomyosarcoma, Liposarcoma, Myxoid liposarcoma, Uterine corpus leiomyosarcoma, Osteosarcoma, Ewing sarcoma and Rhabdomyosarcoma, Synovial sarcoma, Solitary Fibrous tumor), hematopoietic cancers (including but not limited to Chronic lymphatic leukaemia (CLL), Acute lymphatic leukaemia (ALL), Acute myeloid leukaemia (AML), B-cell, T-cell and large granular lymphoma), tumours of neuroepithelial tissue, such but not limited to Astrocytomas (Pleomorphic Xanthoastrocytoma, Fibrillary Astrocytomas, Anaplastic astrocytoma, Glioblastoma Multiforme), Oligodendroglioma, Ependymoma, Choroid plexus tumor, Oligoastrocytoma, gliosarcoma, Ganglioglioma, Retinoblastoma, Neurocytoma, Neuroblastomas (Esthesioneuroblastoma and Ganglioneuroblastoma), Medulloblastoma, Atypical Teratoid Rhabdoid tumors and all types of neuroendocrine cancer.

Sequences, including sequences of VAR2CSA polypeptides:

>fcr3 745 amino acids | 640 aa; underlined sequence corresponds to the ID1 domain of FCR3, Sequence in bold corresponds to DBL2Xb domain of FCR3. Remaining sequence is ID2a (SEQ ID NO:1)

NYIKGDPYFAEYATKLSFILNPSDANNPSGETANHND EACNCNESGISSVGQAQTSGPSSNKT CITHSSIK  
 5 TNKKKECKDVKLG VRENDKDLKICVIEDTSLSGVDNCCCQDLLGILQENCSDNKRGSSSNDSCDNKNO  
DECQKKLEKVFAS**LTNGYKCDKCKSGTSRSK****KKWIWKSSGNEEGLQEEYANTIGLPPRTQSLY**  
**LGNLPKLENVCEDVKDINFD****TKKFLAGCLIVSFHEGKNLKKRYPQNKNSGNKENLCKALEYSF**  
**ADYGDLIKGT****SIWDNEYTKDLELNLQNNFGKLF****GKYIKKNNTAEQDTSYSSLDELRESWWNT**  
 10 **NKKYIWTAMKHGAEMNITTCNADGSVTGSGSSCDDIPTIDLIPQYLRFLQEWVENFCEQRQA**  
**KVKD****VITNCKSCKESGNKCKTECKTKCKDECEKYKKFIEACGTAGGGIGTAGSPWSKRWDQIY**  
**KRYSKHIEDAKRNRKAGTKNCGTSSTTNA****AAS****TDENKCVQSDIDSFFKHLIDIGLTPSSYLSN**  
**VLDDNICGADKAPWTTYTTTTEKCNKERDKSKSQSSDTLVVVNVPSPLGNTPYRYKYACQC**  
 KIPTNEETCDDRKEYMNQWSCGSARTMKRGYKNDNYELCKYNGVDVKPTTVRSNSSKLD

>gi|254952610|gb|ACT97135.1| VAR2CSA [Plasmodium falciparum] | 341 aa (SEQ ID NO:2)  
 15 KCDKCKSGTSRSRKIWTWRKSSGNKEGLQEEYANTIGLSPRTQLLYLGNLRKLENVCEDVTDINFDTKK  
 FLAGCLIAAFHEGKNLKKRYLEKKKGDNNSKCKDLKYSFADYGDLIKGTSIWDNDFTKDLELNLQQIFGK  
 LFRKYIKKNISTEQDTSYSSLDELRESWWNTNKKYIWLAMKHGAGMNSTTCSGSGDSSSGENQTNSC  
 DDIPTIDLIPQYLRFLQEWVEHFCEQRQAKVKDVTNCNSCKESGGTCNSDCEKKCKNKCDAYKTFIEDC  
 KGVGGTGTAGSSWVKRWYQIYMRYSKYIEDAKRNRKAGTKSCGTSSTTNVSVSTDENKCVQS-

>M24 745 amino acids | 656 aa (SEQ ID NO:3)  
 20 **DYIKGDPYFAEYATKLSFILNSSDANNPSGETANHND****EV****CNPNESEISSVGQAQTSDPSSNKT**  
**CNTHSSIKANKKKVCKHV****KLGINNNDKVL****RVCVIEDTSLSGVENCCFKDLLGILQENCSDNKS**  
**GSSSNGSCNNKNQEACEKNLEKVLASLTNCYKCDKCKSGTSTVNKNWIWKSSGNKEGLQKE**  
**YANTIGLPPRTHSLYLGNLPKLENVCEDVKDINFD****TKKFLAGCLIAAFHEGKNLKKRYPQNK**  
 25 **DDNSKLCKALEYSFADYGDLIKGT****SIWDNEYTKDLELNLQQIFGKLFRKYIKKNISTEQD****TL**  
**SSLDELRESWWNTNKKYIWLAMKHGAGMNITTCG****DGSVTGSGSSCDDIPTIDLIPQYLRFL**  
**QEWVEHFCKQRQEKVKDVINSCNSCKNTSSKTKLGDTCNSDCEKKCKIECEKYKKFIEECRTA**  
**VGGTAGSSWSKRWDQIYKMYSKHIEDAKRNRKAGTKNCGITTGTISGESSGANS****GVTTTENK**  
**CVQSDIDSFFKHLIDIGLTPSSYLSIVLDDNICGDDKAPWTTYTTTTEKCNKERDKSK**  
 30 **SQQSNTSVVVNVPSPLGNTPHGYKYACQCKIPTNEETCDDRKEYMNQWISDTSKNPKGSGSTNNDY**  
 ELYTYNGVKETKLPKLNPKLD

>KMWII 745 amino acids | 643 aa (SEQ ID NO:4)  
 35 **DYIKDDPYSKEYTTKLSFILNSSDANTSSGETANHND****EACNCNESEISSVGQAQTSGPSSNKT**  
**ITHSFIKANKKKVCKDVKLG VRENDKVL****RVCVIEDTSLSGVDNCCCQDLLGILQENCSDNKR**  
**GSSSNGSCNNKNQDECQKKLEKVFVSLTNGYKCDKCKSGTSTVNKKWIWKSSGNEKGLQKEY**  
**ANTIGLPPRTQSLYLGNLPKLG****NVCEDVTDINFDTKKFLAGCLIAAFHEGKNL****KISHEKKKGD**  
**NGKKLCKALEYSFADYGDLIKGT****SIWDNEYTKDLELNLQKAFGKLFGKYIKKNIASDENTS****SYSS**  
**LDELRESWWNTNKKYIWTAMKHGAEMNSTMCNADGSVTGSGSSCDDIPTTDFIPQYLRFLQ**  
**EWVEHFCKQRQEKVNAVIENCNSCKNTSGERKIGGTCNGDCKTECKNKCEAYKNFIEDCKGG**  
 40 **DGTAGSSWVKRW****DQIYKRYSKHIEDAKRNRKAGTKSCGPSSITNASVSTDENKCVQSDIDSF**  
**FKHLIDIGLTPSSYLSIVLDENNCGEDNAPWTTYTTTTEKCNKD****KKKSKSQSCNTAVVVNV**  
**PSPLGNTPHEYKYACQCKIPTTEETCDDRKEYMNQWISDTSKKQKGSGSTNNDYELYTYTG****VKETKLP**  
 KKLNSPKLD

>1248 745 amino acids | 640 aa (SEQ ID NO:5)  
 45 **SYVKNDPYSKEYVT****KLSFILNPSDANNPSGETANHND****EACNPNESEIASVGQAQTSDRLS****QKA**  
**CITHSFIGANKKIVCKDVKLG VREKDKDLKICVIEDDSL****RGVENCCFKDLLGILQENCSDNKS**  
**GSSSNGSCNNKNQDECQKKLDEALASLHNGYKCDKCKSGTSRSK****KIWTWRKFPNGEGLQKE**  
**YANTIGLPPRTQSLYLGNLRKLENVCKG****VTDINFDTKKFLAGCLIAAFHEGKNL****KISNKKKND**  
**DNGKKLCKDLKYSFADYGDLIKGT****SIWDNEYTKDLELNLQKIFGKLFRKYIKKNIASDENT****LYS**  
 50 **SLDELRESWWNTNKKYIWLAMKHGTTCS****SGSGDNGDGSVTGSGSSCDDMSTIDLIPQYLRFL**  
**QEWVEHFCKQRQEKVKDV****IENTCKSCKNTSGERIIGGTCGSDCKTKCKGECDAYKNFIEECKRG**  
**DGTAGSPWSKRWDQIYMRYSKYIEDAKRNRKAGTKNCGTSSTTNA****AENKCVQSDIDSFFKHL**  
**IDIGLTPSSYLSIVLDENICGDDKAPWTTYTTTTEKCNKETDKSKSQSCNTAVVVNVPSPL**  
**GNTPHGYKYACECKIPTTEETCDDRKEYMNQWISDTSKKPKGGRSTNNDYELYTYNGVKETKLP**  
 55 **KKSS**  
 SSKLD

>gi|254952618|gb|ACT97139.1| VAR2CSA [Plasmodium falciparum] | 358 aa (SEQ ID NO:6)  
 KCEKCKSEQSKNNNIWIWRKFPNGEGLQKEYANTIGLPPRTHSLYLGNLPKLENVCKDVKDINFDTKE  
 KFLAGCLIAAFHEGKNLKTTPQKNADNNSKLCKDLKYSFADYGD LIKGTSIWDNDFTKDLELNLQKIFG  
 KLFRKYIKKNIASDENTLYSSLDELRESWWNTNKKYIWLAMKHGAEMNSTMCNGDGSVTGSSDSGSTT  
 5 CSGDNNGSISCDDIPTIDLIPQYLRFLQEWVEHFCKQRQEKVKPVIENCKSCKNTSGERIIGGTGSDCEK  
 KCKGECDAYKKFIEECKGGGGGTGTAGSPWSKRWDQIYKRYSKYIEDAKRNRKAGTKSCGPSSTTNA  
 ASTTESKCVQS

>gi|254952592|gb|ACT97126.1| VAR2CSA [Plasmodium falciparum] | 333 aa (SEQ ID NO:7)  
 KCDKCKSEQSKNNKNWIWKQFPNGEGLQKEYANTIGLPPRTHSLYLGNLPKLENVCKGVTDINFDTK  
 10 EKFLAGCLIAAFHEGKNLKTSHKKKGDNKKLCKDLKYSFADYGD LIKGTSIWDNDFTKDLELNLQKIF  
 GKLFYIKKNISAEQDTSYSSDELRESWWNTNKKYIWLAMKHGTTCSGSGDNNGDGSVTGSGSSCD  
 DMPTTDFIPQYLRFLQEWVEHFCKQRQEKVNAVITNCKSCKESGGTCNSDCEKKCKDECEKYKKFIEECR  
 TAADGTAGSSWSKRWDQIYKMYSKHIEDAKRNRKAGTKNCGTSSTTNAENKCVQS

>gi|90193467|gb|ABD92329.1| erythrocyte membrane protein 1 [Plasmodium falciparum] |  
 15 269 aa (SEQ ID NO:8)  
 DYIKDDPYSKEYTTKLSFILNSSDANTSSGETANHNDACNCNESEIASVEQASISDRSSQKAYITHSSIK  
 TNKKKVCKYVKLGINNNDKVLRVCIEDTSLSGVENCCFKDLLGILQENCSDNKRGSFNDSCNNNNEE  
 ACQKKLEKVLASLTNGYKCEKCKSGTSRSKKKWIWKSSGKEGGLQKEYANTIGLPPRTQSLYLGNLPKL  
 ENVCKGVTDINFDTKEKFLAGCLIAAFHEGKNLKP SHQKNDDNNSKLCKDLKYSFADY

>gi|254952616|gb|ACT97138.1| VAR2CSA [Plasmodium falciparum] | 333 aa (SEQ ID NO:9)  
 KCDKCKSGTSRSKKKWTWRKSSGNKEGLQKEYANTIGLPPRTHSLYLGNLRKLENVCEDVTDINFDTKE  
 20 KFLAGCLIAAFHEGKNLKTTPQKNDDNNSKLCKALKYSFADYGD LIKGTSIWDNDFTKDLELNLQKIFG  
 KLFRKYIKKNISTEQHTSYSSLDELRESWWNTNKKYIWLAMKHGAEMNGTTCSCSGDSSDDIPTIDLIPQ  
 YLRFLQEWVEHFCKQRQAKVNAVINSNCNCKNTSGERKLGGTGCGSECKTECKNKCDAYKEFIDGTGSGG  
 25 GTGTAGSSWVKRWDQIYKRYSKYIEDAKRNRKAGSKNCGTSSTTNAESKCVQS

>hb31 745 amino acids | 650 aa (SEQ ID NO:10)  
**SYVKNNPYSAEYVTKLSFILNSSDANTSSETPSKYDEVVCNCNESEISSVGQAQTSGPSSNKTCTC**  
**ITHSSIKTNKKKVCKDVKLGINNNDKVLRVCIEDTSLSGVDNCCCQDLLGILQENCSDKNQS**  
**GSSSNGSCNNKNQDECQKKLEKVFASLTNGYKCDKCKSGTSRSKKKWIWRKSSGNEEGLQKE**  
 30 **YANTIGLPPRTQSLYLGNLRKLENVCKGVTDINFDTKEKFLAGCLIAAFHEGKNLKTTPQNK**  
**KLCKDLKYSFADYGD LIKGTSIWDNEYTKDLELNLQKAFGKLFRKYIKKNISTEQHTLYSSLDE**  
**LRESWWNTNKKYIWLAMKHGAGMNSTTCCGDGSGVTGSGSSCDDIPTIDLIPQYLRFLQEWV**  
**EHFCKQRQEKVNAVIENTNSCKEKGDTNCECKTECEKKCKIECEKYKTFIEECVTAVGGTSGS**  
**PWSKRWDQIYKRYSKYIEDAKRNRKAGTKNCGITTGTISGESSGANSVTTTENKCVQSDID**  
 35 **SFFKHLIDIGLTPSSYLSIVLDDNICGADNAPWTTYTTYTTYTTTKNCDIKKKTPKSQPINTSV**  
**VVNVPSPLGNTPHGYKYACQCKIPTTEESCDRKEYMNQWIIDTSKKQKSGSGSTNNDYELYTYNGVK**  
 ETKLPKSSSSKLD

>hb32 745 amino acids | 643 aa (SEQ ID NO:11)  
**SYVKDDPYSAEYVTKLSFILNSSDANTSSETPSKYDEVVCNCNESEISSVGQAQTSGPSSNKTCTC**  
 40 **ITHSSIKTNKKKVCKDVKLGINNNDKVLRVCIEDTSLSGVDNCCCQDLLGILQENCSDKNQS**  
**GSSSNGSCNNKNQDECQKKLEKVFASLTNGYKCDKCKSGTSRSKKKWIWRKSSGNEEGLQKE**  
**YANTIGLPPRTQSLYLGNLPKLENVCKGVTDIIYDTKEKFLSGCLIAAFHEGKNLKTSHKKND**  
**DNGKKLCKALEYSFADYGD LIKGTSIWDNDFTKDLELNLQKIFGKLFRKYIKKNNTAEQDTSY**  
**SSLDELRESWWNTNKKYIWTAMKHGAGMNSTTCSGDGSGVTGSGSSCDDMPTIDLIPQYLRFL**  
 45 **QEWVEHFCKQRQEKVKDVITNCSCKEKGDTNCECKTECKTKCKGECEKYKNFIEECNGTAD**  
**GGTSGSSWSKRWDQIYKRYSKYIEDAKRNRKAGTKNCGTSSTTNAASTTENKCVQSDIDSF**  
**FKHLIDIGLTPSSYLSNVLDDNICGEDKAPWTTYTTYTTTKNCDIQKKT PKPQSCDTLVVVNVP**  
**SPLGNTPHGYKYVCECKIPTTEETCDDRKEYMNQWIIDTSKKQKSGSGSTNNDYELYTYNGVQIKQAAG**  
 TLKNSKLD

>gi|90193475|gb|ABD92333.1| erythrocyte membrane protein 1 [Plasmodium falciparum] |  
 50 269 aa (SEQ ID NO:12)  
 NYIKGDPYSAEYATKLSFILNSSDTENASEKIQKNNDDEVVCNCNESEIASVEQAPISDRSSQKACITHSSIK  
 ANKKKVCKHVKLGVRENDKDLKICVIEDTSLSGVDNCCCQDLLGILQENCSDNKS GSSSNGSCNNNNEE  
 ICQKKLEKVLASLTNGYKCDKCKSGTSTVNKNWIWKYSGKEGGLQEYANTIGLPPRTQSLYLGNLPKL  
 55 ENVCEDVKDINFDTKEKFLAGCLIAAFHEGKNLKT SNKKKNDDNNSKLCKALKYSFADY

>gi|254952600|gb|ACT97130.1| VAR2CSA [Plasmodium falciparum] | 344 aa (SEQ ID NO:13)

KCDKCKSGTSTVNKKWIWKKYSGTEGGLQEEYANTIALPPRTQSLYLGNLPKLENVCKDVTDFINFDTK  
5 FLAGCLIAAFHEGKNLKTYYLEKKKGDNKKNDNNSKLCKALKYSFADYGDLIKGTSIWDNDFTKDLEL  
NLQQIFGKLFKRYIKKNIASDENTLYSSLDELRESWWNTNKKYIWLAMKHGAGMNSTMCNADGSVTGS  
GSSCDDIPTIDLIPQYLRFLQEWVEHFCKQRQAKVKDVITNCNSCKECCGGTCNGECKTECEKKCKGECD  
AYKKFIEECKGKADEGTSGSSWSKRWDQIYKRYSKYIEDAKRNRKAGTKNCGPSSTTSTAESKCVQS

>gi|254952598|gb|ACT97129.1| VAR2CSA [Plasmodium falciparum] | 334 aa (SEQ ID NO:14)

10 KCDKCKSEQSKKNNNIWIWKKSSGTEGGLQKEYANTIALPPRTQSLYLGNLRKLENVCEDVKDINFDTK  
KFLAGCLIAAFHEGKNLKKRYLEKKNGDNNSKLCKALKYSFADYGDLIKGTSIWDNEYTKDLELNLQKIFG  
KLFRKYIKKNTAEQHTSYSSLDELRESWWNTNKKYIWLAMKHGTTCSGSGDNGSISCDIPTIDLIPQ  
YLRFLQEWVEHFCEQRQGVNAVIENCNSCKNTSSKTKLGGTCNGECKTECKGECDAYKEFIECKKGTA  
AEGTSGSSWVKRWYQIYMRYSKYIEDAKRNRKAGTKNCGTSSTTSTAESKCVQS

15 >gi|254952596|gb|ACT97128.1| VAR2CSA [Plasmodium falciparum] | 332 aa (SEQ ID NO:15)

KCDKCKSEQSKKNNNIWIWKKSSGTEGGLQKEYANTIALPPRTQSLYLGNLRKLENVCEDVKDINFDTK  
KFLAGCLIAAFHEGKNLKKRYLEKKNGDNNSKLCKALKYSFADYGDLIKGTSIWDNEYTKDLELNLQKIFG  
20 KLFRKYIKKNTAEQDTSYSSLDELRESWWNTNKKYIWTAMKHGTTCSGSGDNGSISCDIPTIDLIPQ  
YLRFLQEWVEHFCEQRQEKVKDVKNCSCKECCGGTCNGECKTECKNKCKDECDAYKKFIEECKGAAE  
GTSGSSWSKRWDQIYKRYSKYIEDAKRNRKAGTKNCGTSSTTSTAENKCVQS

>gi|90193465|gb|ABD92328.1| erythrocyte membrane protein 1 [Plasmodium falciparum] | 267 aa (SEQ ID NO:16)

25 NYIKDDPYSAEYTTKLSFILNSSDTENASEKIQKNNDEVCPNPNESGIACVELAQTSGSSSNKTCNTHSFIK  
ANKKKVCKDVKLGINKKDKDLKICVIEDDSLGRVDNCCCQDLLGILQENCSDKNQSGSSSNGSCNNKN  
QEACQKKLENVFASLTNGYKCEKCKSEQSKKNNKNWIWKKYSVKEEGLQKEYANTIALPPRTQSLYLGNL  
PKLGNVCKGVTDINFDTKKFLAGCLIAAFHEGKNLKTYYLQNKKKLCKALKYSFADY

>gi|90193477|gb|ABD92334.1| erythrocyte membrane protein 1 [Plasmodium falciparum] | 263 aa (SEQ ID NO:17)

30 DYIKGDPYFAEYATKLSFILNSSDANTSSGETANHNDEACNPNESEIASVEQASISDRSSQKACNTHSSIK  
ANKKKCKECHKVGLGVRENDKDLKICVIEDTSLSGVDNCCCQDLLGILQENCSDNKRGSSSNGSCDKNSEE  
ICQKKLDEALASLHNGYKNQKCKSEQSKKNNKNWIWKKSSGNEKGLQKEYANTIGLPPRTQSLYLGNLP  
KLENVCEDVTDFINFDTKKFLAGCLIAAFHEGKNLKTTPQNKNDDNGKKLCKD

35 >gi|254952594|gb|ACT97127.1| VAR2CSA [Plasmodium falciparum] | 338 aa (SEQ ID NO:18)

KCDKCKSEQSKKNNNIWIWKKSSGNKKGLQKEYANTIGLPPRTQSLYLGNLPKLENVCKDVTDFINFDTK  
KFLAGCLIAAFHEGKNLKNISNEKKNDNGKKLCKDLKYSFADYGDLIKGTSIWDNEYTKDLELNLQNNFG  
40 KLFRKYIKKNTAEQHTLYSSLDELRESWWNTNKKYIWLAMKHGTTCSGSGDNGDGSVTGSGSSCDD  
MSTIDLIPQYLRFLQEWVEHFCKQRQEKVNAVIENCNSCKNTSSKTKLGGTCNGECKTECEKKCKDECEK  
YKEFIEECKRGDGTAGSPWVKRWQIYMRYSKYIEDAKRNRKAGTKSCGTSAAENKCVQS

>gi|254952602|gb|ACT97131.1| VAR2CSA [Plasmodium falciparum] | 341 aa (SEQ ID NO:19)

45 KCDKCKSEQSKKNNNIWIWKKSSGDEKGLQKEYANTIALPPRTQSLYLGNLPKLENVCKDVTDFINFDTK  
KFLAGCLIAAFHEGKNLKTSHQKNADNGKKNDNGKKLCKALKYSFADYGDLIKGTSIWDNEYTKDLE  
LNLQQIFGKLFKRYIKRNTAEQHTLYSSLDELRESWWNTNKKYIWLAMKHGTTCSGSGDNGDGSVTG  
SGSSCDDMSTIDLIPQYLRFLQEWVEHFCKQRQEKVKDVITNCNSCKECCGGTCGSDCKTKCEAYKKFIEE  
CNGTADGGTSGSSWSKRWDQIYKRYSKYIEDAKRNRKAGTKNCGPSSGANSVTTTENKCVQS

>gi|254952660|gb|ACT97160.1| VAR2CSA [Plasmodium falciparum] | 352 aa (SEQ ID NO:20)

50 KCEKCESEQSKNNKYWIWKKSSGNGEGLQEEYANTIALPPRTHSLCLVCLHEKEGKKTQELKNIRTNSE  
LLKERIIAAFHEGKNLKTSPQKNNDNGKKLCKDLKYSFADYGDLIKGTSIWDNEYTKDLELNLQKIFGKLF  
RKYIKKNTAEQHTLYSSLDELRESWWNTNKKYIWLAMKHGAGMNSTMCNADGSVTGSSDSGSTTCC  
GDNGSISCDMDPTIDLIPQYLRFLQEWVEHFCEQRQEKVNAVITNCKSCKECCGGTCNSDCEKKCKAYKE  
55 FIECKGGGTGEGTSGSSWSKRWDQIYKRHSKHIEDAKRNRKAGTKNCGITTGTISGESSGANSVTTTE  
NKCVCQS

>gi|254952652|gb|ACT97156.1| VAR2CSA [Plasmodium falciparum] | 344 aa (SEQ ID NO:21)

KCDKCKSGTSRSRKIWTWRKFRGNGLQKEYANTIGLSPRTQLLYLVCLHEKGKKTQELKNISTNSELL  
KEWIIAAFHEGKNLKTTPQKKNDNDNGKKLCKALKYSFADYGD LIKGT SIWDNDFTKDLELNLQKIFGKLF  
5 RKYIKKNIASDENTSYS SLD ELRESWWNTNKKYIWTAMKHGAGMNGTTC CGD GSVTGSSDSGSTTCCG  
DGSVTGSGSSCDDIPTIDLIPQYLRFLQEWVEHFCEQRQEKVKDVTNCKSCKESEKKCKNKCDAYKEFI  
DGTGSGGGGTGTAGSSWSKRWDQIYMRYSKYIEDAKRNRKAGTKNCGTSSGANSVTTTENKCVQS

>gi|254952622|gb|ACT97141.1| VAR2CSA [Plasmodium falciparum] | 350 aa (SEQ ID NO:22)

KCEKCKSEQSKNNKIWTWRKFPNGEGLQKEYANTIGLSPRTQLLYLVCLHEKGKKTQHK TISTNSELL  
KEWIIAAFHEGKNLKKRYLEKKKGDNNSKLCKDLKYSFADYGD LIKGT SIWDNDFTKDLELNLQKIFGKLF  
10 RKYIKKNIASDENTSYS SLD ELRESWWNTNKKYIWTAMKHGAGMNSTMCNGD GSVTGSSDSGSTTCS  
GDNGSIS CDDIPTIDLIPQYLRFLQEWVEHFCEQRQEKVKDVIKNCNSCKECCGTCNGECKTECKNKCK  
DECEKYKNFIEVCTGGDGTAGSPWSKRWYQIYMRYSKYIEDAKRNRKAGTKSCGTSSGANSVTTTESK  
15 CVQS

>gi|254952626|gb|ACT97143.1| VAR2CSA [Plasmodium falciparum] | 359 aa (SEQ ID NO:23)

KCEKCKSEQSKNNKNWIWRKFPNGEGLQKEYANTIGLPPRTHSLYLVCLHEKGKKTQELKNIRTNSEL  
LKEWIIAAFHEGKNLKKRYHQNNNSGNKKKCKALEYSFADYGD LIKGT SIWDNEYTKDLELNLQKIFGKLF  
20 LFRKYIKKNISTEQDTLYSSLD ELRESWWNTNKKYIWLAMKHGAGMNSTTCCGD GSVTGSSDSGSTTCS  
GDNGSIS CDDMPTIDLIPQYLRFLQEWVEHFCEQRQEKVKDVIENCKSCKNTSGERIIGGTCNGECKTEC  
EKKCKAACEAYKTFIEECEGKAAEGTSGSSWSKRWYQIYMRYSKYIEDAKRNRKAGTKNCGKSSGANSV  
VTTENKCVQS

>gi|90193469|gb|ABD92330.1| erythrocyte membrane protein 1 [Plasmodium falciparum] | 270 aa (SEQ ID NO:24)

NYIKDDPYSKEYVTKLSFIPNSSDANNPSGETANHNDEV CNPNESEISSVEHAQTSVLLSQKAYITHSSIK  
ANKKKVCKYVKLG VRENDKDLKICVIEDDSL RGVENCCFKDFLRILQENCSDNKRESSSNGSCNNNNEE  
ACEKNLDEALASLTNCKYKQCKSGTSTVNNKNWIWKSSGKEGGLQKEYANTIGLPPRTQSLCLVVCL  
DEKEGKTQELKNIRTNSELLKEWIIAAFHEGKNLKKRYHQNKNDNNSKLCKALKYSFADY

>gi|254952644|gb|ACT97152.1| VAR2CSA [Plasmodium falciparum] | 334 aa (SEQ ID NO:25)

KCDKCKSEQSKNNKYWIWKYSVKEGGLQKEYANTIALPPRTQSLCLVVCLDEKEGKTQELKNIRTNSE  
LLKERIIAAFHEGKNLKT YHEKKKGDDGKKLCKDLKYSFADYGD LIKGT SIWDNDFTKDLELNLQKIFGKLF  
30 FRKYIKKNNTAEQHTSYSSLD ELRESWWNTNKKYIWTAMKHGAEMNGTTCSCSGDSSNDIPTIDLIPQY  
35 LRFLQEWVEHFCEQRQAKVNAVIKNCCKECCGTCNGECKTECKTKCKGECEKYKEFIEKCEGQAAEG  
TSGSSWSKRWYQIYMRYSKYIEDAKRNRKAGTKNCGTSSGANSVTTTENKCVQS

>gi|254952642|gb|ACT97151.1| VAR2CSA [Plasmodium falciparum] | 351 aa (SEQ ID NO:26)

KCDKCKSEQSKNNKNWIWKYSGTEGGLQKEYANTIALPPRTQSLYLVCLHEKEEKTQELKNISTNSEL  
LKEWIIAAFHEGKNL KISPQNKNDNGKNLCKDLKYSFADYGD LIKGT SIWDNDFTKDLELNLQKIFGKLF  
40 KYIKKNNTAEQDTLYSSLD ELRESWWNTNKKYIWTAMKHGAGMNGTTC CGD GSVTGSSDSGSTTCCG  
DGSVTGSGSSCDDIPTIDLIPQYLRFLQEWVEHFCEQRQAKVKDVIKNCNSCKECCGTCNGECKTECEK  
KCKGECEAYKKFIEKCN GGGGEGTSGSSWSKRWDQIYMRYSKYIEDAKRNRKAGTKNCGTSSTTNAE  
NKCVQS

>gi|254952658|gb|ACT97159.1| VAR2CSA [Plasmodium falciparum] | 353 aa (SEQ ID NO:27)

KCDKCKSGTSTVNKKWIWKKFPKGEGGLQEEYANTIALPPRTQSLCLVVCLDEKEGKTQHK TISTNSELL  
KEWIIAAFHEGKNL KISNKKKNDENNSKLCKDLKYSFADYGD LIKGT SIWDNDFTKDLELNLQKIFGKLF  
50 KYIKKNNTAEQDTSYS SLD ELRESWWNTNKKYIWLAMKHGTTCCSGSGDNGD GSVTGSSDSGSTTCC  
GDGSVTGSGSSCDDIPTIDLIPQYLRFLQEWVEHFCKQRQAKVKDVIENCKSCKNTSSKTKLGDT CNSD  
CKTKCKVACEKYKEFIEKCVSAAGGTSGSSWVKRWDQIYMRYSKYIEDAKRNRKAGTKNCGPSSTTSTA  
ESKCVQS

>gi|254952640|gb|ACT97150.1| VAR2CSA [Plasmodium falciparum] | 327 aa (SEQ ID NO:28)

KCDKCKSGTSTVNKKWIWKYSGKEGGLQKEYANTIGLPPRTQSLCLVCLHEKEGKTQELKNISTNSELL

KEWIIAAFHEGKNLKISNKKKNDNDNGKKLCKDLKYSFADYGD LIKGT SIWDNDFTKDLELNLQKIFGKLF  
 RKYIKKNNTAEQDTLYSSDELRESWWNTNKKYIWTAMKHGAGMNSTTCSCSGDSSNDIPTIDLIPQYL  
 RFLQEWVEHFCKQRQEKVNAVITNCKSCKESGGTCNSDCEKKCKIECEKYKNFIEKCVTAAGGTSGSSW  
 SKRWDQIYKMYSKYIEDAKRNRKAGTKNCGPSSTTNAASTDENKCVQS

5 >dd2full 745 amino acids | 628 aa (SEQ ID NO:29)  
**NYIKGDPYFAEYATKLSFILNSSDTENASETPSKYYDEACNCNESEIASVGQAQTS GPSSNKT  
 CITHSSIKTNKKKECKDVKLGINNNNDKVL RVCVIEDTSLSGVDNCCCQDLLGILQENCSDNKR  
 GSSNGSCDKNSEEICQKKLEKVFASLTNGYKCDKCKSGTSRSKKKWIWKKSSGNEEGLQKEY  
 ANTIGLPPRTQSLCLVCLHEKEGKTQHKTISTNSELLKEWIIAAFHEGKNLKTSHQNNNSGNK  
 10 KLCKALEYSFADYGD LIKGT SIWDNEYTKDLELNLQKIFGKLF RKYIKKNNTAEQHTSYSS  
 LDELRESWWNTNKKYIWTAMKHGAGMNGTTCSCSGDSSNDMPTIDLIPQYLRFLQEWVEHFCKQ  
 RQEKVNAVIENTCNSCKESGGTCNSDCKTECKNKCEAYKEFIEDCKGGGTGTAGSPWSKRWDQ  
 IYKRYSKHIEDAKRNRKAGTKNCGTSSTTNAASTDENKCVQSDVDSFFKHLIDIGLTPSSYL  
 SNVLDDNICGADKAPWTTYTTTTTKNC DIQKKTPKSQSCDTLVVVNVPSPLGNTPHEYKYAC  
 15 ECKIPTTEETCDDRKEYMNQWSCGSAQTVRGRSGKDDYELTYNGVKETKPLGLTKNSKLD**

>gi|254952636|gb|ACT97148.1| VAR2CSA [Plasmodium falciparum] | 350 aa (SEQ ID  
 NO:30)  
 KCEKCKSEQSKKNNKNWIWRKFRGTEGGLQEEYANTIGLPPRTQSLCLVCLDEKGGKTQELKNIRTNSE  
 LLKEWIIAAFHEGKNLKPSHQNKNSGNKENLCKALKYSFADYGD LIKGT SIWDNDFTKDLELNLQKIFGK  
 20 LFRKYIKKNNTAEQHTSYSSDELRESWWNTNKKYIWTAMKHGAEMNGTTCNADGSVTGSSDSGTTCS  
 GDNGSISCDIPTIDLIPQYLRFLQEWVEHFCKQRQEKVNAVINSCKNTSSKTKLGDTCSNDCCKTC  
 KIECEKYKTFIEKCVTAAGGTSGSPWSKRWDQIYKRYSKYIEDAKRNRKAGTKNCGPSSTTSTAESKCVQ  
 S

>gi|254952638|gb|ACT97149.1| VAR2CSA [Plasmodium falciparum] | 330 aa (SEQ ID  
 NO:31)  
 KCDKCKSEQSKKNNKNWIWRKYSGNGEGLQKEYANTIGLPPRTHSLYLVLCLHEKEGKTQELKNIRTNSEL  
 LKEWIIAAFHEGKNLKTTYLENKNNDENKKLCKALKYSFADYGD LIKGT SIWDNDFTKDLELNLQKIFGK  
 LFRKYIKKNIASDENTLYSSDELRESWWNTNKKYIWTAMKHGAEMNGTTCSSGSGDNGSISCDIPTID  
 LIPQYLRFLQEWVGHFCKQRQEKVNAVITNCNSCKESGGTCNSDCEKKCKIECEKYKFFIEECRTAAGGT  
 30 SGSPWSKRWDQIYKMYSKYIEDAKRNRKAGTKNCGPSSTTSTAESKCVQS

>gi|254952628|gb|ACT97144.1| VAR2CSA [Plasmodium falciparum] | 334 aa (SEQ ID  
 NO:32)  
 KCDKCKSEQSKKNNKNWIWRKYSGNGEGLQKEYANTIGLPPRTHSLYLVLCLHEKEGKTQHKTISTNSELL  
 KEWIIAAFHEGKNLKKRYPQNNNSGNKKLCKDLKYSFADYGD LIKGT SIWDNEYTKDLELNLQKAFGK  
 LFRKYIKKNIASDENTLYSSDELRESWWNTNKKYIWLAMKHGAEMNGTTCNADGSVTGSGSSCDDMST  
 35 IDLIPQYLRFLQEWVEHFCEQRQAKVKDVINSCKSCKESGDTCSNDCCKCKNKCDAYKTFIEEFCTADG  
 GTAGSPWSKRWDQIYKRYSKYIEDAKRNRKAGTKNCGTSSGANSVTTTENKCVQS

>gi|254952630|gb|ACT97145.1| VAR2CSA [Plasmodium falciparum] | 350 aa (SEQ ID  
 NO:33)  
 KCDKCKSGTSTVNKNWIWKKYSGKEEGLQKEYANTIALPPRTHSLYLVLCLHEKGKKTQELKNIRTNSEL  
 KEWIIAAFHEGKNLKTSPQNNNSGNKKLCKALKYSFADYGD LIKGT SIWDNDFTKDLELNLQKIFGKLF  
 RKYIKKNNTAEQHTSYSSDELRESWWNTNKKYIWLAMKHGAEMNGTTCGDSVTGSSDSGTTCS  
 GDNGSISCDDMPTTDFIPQYLRFLQEWVEHFCKQRQEKVKHVMESCKSCKEKGDTCSNGECKTECEKK  
 40 KKNKCEAYKTFIEKCVSADGGTSGSSWSKRWDQIYMRYSKYIEDAKRNRKAGTKNCGTSSTTNAASTAE  
 NKCVQS

>P13 745 amino acids | 647 aa (SEQ ID NO:34)  
**DYIKDDPYSAEYATKLSFILNPSDANTSSGETANHND EVCNCNESEIASVELAPISDSSSNKT  
 CITHSFIGANKKKECKDVKLG VREKDKDLKICVIEDDSL RGVENCCCQDLLGILQENCSDNKS  
 50 GSSNGSCDKNSEDECQKKLENVFASLKNYKCDKCKSGTSTVNKKWIWRKYSGNGEGLQKEYA  
 NTIGLPPRTHSLYLVLCLHEKEGKTQHKTISTNSELLKEWIIAAFHEGKNLKTSHQNNNSGNK  
 KLCKALKYSFADYGD LIKGT SIWDNDFTKDLELNLQKIFGKLF RKYIKKNIASDENTSYSS  
 LDELRESWWNTNKKYIWLAMKHGAEMNSTMCNGDGSVTGSSDSGTTCSGDNGSISCDIPTID  
 LIPQYLRFLQEWVEHFCKQRQEKVKDVITNCKSCKESGDTCSNDCCKCKNKCEAYKFFIEER  
 RTAAQGTAESSWVKRWDQIYMRYSKYIEDAKRNRKAGTKSCGPSSTTNAASTAENKCVQS  
 55 DIDSFFKHLIDIGLTPSSYLSIVLDDNICGADNAPWTTYTTTTTKNC DIKKKTPKPQSCDTL**

5 KCDKCKSGTSTVNKKWIWRKSSGNKEGLQKEYANTIGLPPRTQSLYLGNLPKLENVCEDVKDINFDTKEK  
FLAGCLIVSFHEGKNLKTSHEKKNDNGKKLKCALEYSFADYGD LIKGT SIWDNEYTKDLELNLQKIFGKL  
FRKYIKNNTAEQDTSYSSLDLRESWWNTNKKYIWTAMKHGAGMNITTCGGDSSGENQTNSCDDIP  
TIDLIPQYLRFLQEWVEHFCKQRQEKVNAVVTNCKSCKESGGTCNGECKTKCKNKCEVYKTFIDNVGDG  
TAGSPWVKRWDQIYKRYSKHIEDAKRNRKAGTKNCGITTTGTISGESSGATSGVTTTENKCVQS

NYIKDDPYSKEYVTKLSFIPNSSDANTSSEKIQKNNDDEVCPNPESGISSVEQAQTS GPSSNKT  
CITHSSIAKANKKKECKDVKLG VRENDKDLKICVIEDTSLSGVDNCCCQDLLGILQENCSDNKR  
SSSNDSCDNKNQDECQKKLDEALES LHNGYKNQKCKSGTSTVNKKWIWKSSGNKEGLQKE  
YANTIGLP PRTQSLYLG NLPKLE NVSKGVTDIIYDTKEKFLAGCLIVSFHEGKNLKTSHEKND  
DNGKKLCKALEYSFADYGD LIKGT SIWDNEYTKDLELNLQKAFGKLF RKYIKKNISAEQDTSYS  
SLDELRESWWNTNKKYIWIAMKHGAGMNGTTCCGDGSSGENQTNSCDDIPTIDLIPQYLRFL  
QEWVEHFCEQRQAKVKDVITNCKSCKNTSGERKIGGTCNGECKTKCKNKCEAYKT FIEHCKGG  
DGTAGSSWVKRWDQIYKRYSKHIEDAKRNRKAGTKSCGTSTAENKCVQSDIDSFFKHLIDIG  
LTPSSYLSIVLDENNCGEDKAPWTTYTTTNCNDIQKDKSKSQSSDTLVVVNVPSPLGNTPHG  
YKYACQCKIPTTEETCDDRKEYMNQWSCGSARTMKRGYKNDNYELCKYNGVDVKPTTVRSSTKLD

25 DYIKGDPYSAEYVTKLSFIPNSSDANNPSEKIQKNNDDEVNCNNESEISSVGQASISDPSSNKT  
NTHSSIKANKKKVKCKDVKLGVRENDKVLKICVIEHTSLRGVDNCCFKDLLGILQEPRIDKNQS  
GSSSNGSCDKNSEEACEKNLEKVLASLTNGYKCDKCKSGTSRSKKKWIWKYSGKEGGLQEE  
YANTIGLPPRTQSLCLVVCLDEKEGKTQELKNISTNSELLKEWIIAAFPEGKNLKPSPEKKKG  
NGKKLCKDLKYSFADYGDLIKGTSIWDNEYTKDLELNLQKIFGKLFKRYIKKNIASDENTLYSS  
LDELRESWWNTNKKYIWLAMKHGAGMNSTMCNADGSVTGSGSSCDDMPTIDLIPQYLRFLQ  
EWVEHFCKQRQEKVKPVIENCNSCKNTSSERKIGGTCNSDCKTECKNKCEVYKKFIEDCKGGD  
GTAGSSWSKRWDQIYKRYSKYIEDAKRNRKAGTKNCGPSSTTNAENKCVQSDIDSFFKHLI  
30 DGLTTPSSYLSTVLDDNICGEDNAPWTTYTTYTTTKNCDKDKKKSKSQSCDTLVVVNVPSPL  
GINTPHEYKYACECRTPNKQESCDDRKEYMNQWISDNTKNPKGSGSGKDYYELYTYNGVDVKPTTVRS  
SSTKLD

35 DYIKGDPYFAEYATKLSFILNSSDANTSSGETANHND EACNCNESEISSVEHASISDPSSNKT  
NTHSSIKANKKKVKHVKLGVRENDKDLRVCVIEHTSLSGVENCCFKDFLRILQENCSDNKS  
SSSNGSCDKNNEEACEKNLEKVFASLTNCYKCEKCKSEQSKNNKKWTWRKSSGNKGGQEE  
YANTIGLPPRTQSLCLVCLDEKEGKKTQELKNIRTNSELLKEWIIAAFHEGKNLKPSHEKND  
DNGKKNDNNSKLCKDLKYSFADYGDLIKGTSIWDNEYTKDLELNLQKIFGKLFRKYIKKNIA  
40 SIDENTLYSSLDELRESWWNTNKKYIWLAMKHGAEMNGTTCNADGSVTGSGSSCDDIPTIDLI  
PQYLRFLQEWVEHFCKQRQAKVKDVIENCKSCKESGNKCKTECKNKCEAYKKFIENCKGGDG  
TAGSSVVKRWDQIYMRSKYIEDAKRNRKAGTKNCGPSSITNVSASTDENKCVQSDIDSFFK  
HLIDIGLTTPSSYLSIVLDDNICGDDKAPWTTYTTYTTYTTYTTYTTYTTYTTTKNCDKERDKS  
SQSCNTAVVNVPSPLGNTPHEYKYACECRTPSNKLCDRKEYMNQWSSGSAQTVRDRSGKDY  
ELYTYNGVKETKLPKKNLSSKLD

50 KCDKCKSEQSKKNKYWIWKSSVKEEGLQKEYANTIALPPRTHSLCLVVLDEKGKKTQELKNISTNSE  
LLKERIIAAFHEGKNLKTYYLEKKNADNNSKCLKALKYSFADYGDLIKGTSIWDNEYTKDLELNLQQIFGKL  
FRKYIKKNNTAEQHTLYSSLDELRESWWNTNKKYIWLAMKHGAGMNGTCCGDSVTGSSDSGTTCS  
GDNGSISCDDMPTTDFIPQYLRFLQEWVEHFCKQRQEKVKDVIENCNSCKNNLGKTEINEKCKTECKNK  
CEAYKNFIEKFCTADGGTSGSPWSKRWDOIYKRYSKYIEDAKRNRKAGTKNCGTSSTTSTAENKCVOS

55 KCEKCKSGTSTVNKYWIWRKSSGNKEGLQKEYANTIALPPRTHSLCLVCLDEKEGKTQELKNISTNSEL  
LKERIIAAFHEGENLKTSHEKKKGDDGKKNADNNSKLCKALKYSFADYGDLIKGTSIWDNEYTKDLELNL

QKIFGKLF RKYIKKNIASDENTSYS SLDLRESWWNTNKKYIWLAMKHGAGMNGTTCSCSGDSSDDMP  
TTDFIPQYLRFLQEWVEHFCKQRQENVNAVIENCNSCKECCGGTCNSDCEKKCKTECKNKCEAYKNFIEKF  
CTADGGTSGYSWSKRWDQIYKRYSKYIEDAKRNRKAGTKSCGTSSTSTAESKCVQS

>ghana2 745 amino acids | 667 aa (SEQ ID NO:41)

5 **SYVKNNPYSKEYVTKLSFILNPSDANNPSETPSKYIYDEV C NCNESGIACV GQAQTSGPSSNKT**  
**CITHSFIGANNKKKVCKDVKLG VREKDKDLKICVIEDT YLSGVDNCCFKDFLGMLQENCSDNKS**  
**GSSSNGSCNNKNQDECEKNLDEALASLTNGYKCEKCKSGTSTVNKYWIWRKSSGNKEGLQKE**  
**YANTIALPPRTHSLCLVCLDEKEGKTQHKTISTNSELLKEWIIAAFHEGKNLKTSHEKKKGDD**  
**GKKNADNNSKLCKALKYSFADYGD L IKGTSIWDNDFTKDLELNLQKIFGKLF RKYIKKNIASD**  
10 **ENTSYS SLDLRESWWNTNKKYIWLAMKHGAGMNSTTCCGDGSGVTGSSDSGSTTCCGDGSGV**  
**TGSGSSCDDMPTTDFIPQYLRFLQEWVEHFCKQRQENVNAVIENCNSCKECCGGTCNSDCEKK**  
**CKTECKGECDAYKEFIEKCN GGAAGTSGSSWSKRWDQIYKRYSKYIEDAKRNRKAGTKNCG**  
**TSSTTSTAESKCVQSDIDSFFKHLIDIGLTPSSYLSIVLDENICGADNAPWTTYTTTTTTTTT**  
**TTEKCNKETDKSKLQQCNTSVVNVPSPLGNTPHGYKYVCECRTPNKQETCDDRKEYMNQWISD**  
15 **NTKNPKGSRSTNNDYELYTYNGVQIKPTTVRSNSTKLD**

>gi|254952634|gb|ACT97147.1| VAR2CSA [Plasmodium falciparum] | 348 aa (SEQ ID NO:42)

KCDKCKSEQSKNNKNWIWKKSSGNEKGLQKEYANTIGLPRTQSLCLVCLDEKEGKTQELKNIRTNS  
ELLKEWIIAAFHEGKNLKTSHEKKKGDNNSKLCKDLKYSFADYGD L IKGTSIWDNEYTKDLELNLQNNFG  
20 **KLFRKYIKKNIASDENTSYS SLDLRESWWNTNKKYIWLAMKHGAGMNSTTCSGSGSTTCSGSGSTT**  
**CSSGSGDSCDDMPTID L IPQYLRFLQEWVEHFCKQRQEKVNAVIKNCNSCKESGGTCNGECKTECKNKCE**  
**EAYKTFIEEFCTADGGTSGSPWSKRWDQIYKMYSKHIEDAKRNRKAGTKNCGPSSTTNVSVSTDENKCV**  
**QS**

>ghana1 745 amino acids | 652 aa (SEQ ID NO:43)

25 **DYIKDDPYFAEYVTKLSFILNSSDANNPSGETANHND EVCNP NESGIASVEQAQTSDPSSNKT**  
**CNTHSSIKANKKKKVCKHVKLGVRENDKDLKICVIEHTSLSGVENCCCQDFLRILQENCSDNKS**  
**GSSSNGSCNNKNQEACEKNLEKVLASLTNCYKCDKCKSEQSKNNKNWIWKKSSGNEKGLQ**  
**KEYANTIGLPRTQSLCLVCLDEKEGKTQELKNIRTNSELLKEWIIAAFHEGKNLKKRYPQNK**  
**NDDNNSKLCKDLKYSFADYGD L IKGTSIWDNEYTKDLELNLQNNFGKLF RKYIKKNIESTE QDT**  
30 **LYSSLDLRESWWNTNKKYIWLAMKHGAGMNSTTCSGSGSTTCSGSGSTTCSGSGDSCD**  
**DMPTTDFIPQYLRFLQEWVEHFCKQRQEKVNAVIKNCNSCKESGGTCNGECKTECKNKCEAY**  
**KTFIEEFCTADGGTSGSPWSKRWDQIYKMYSKHIEDAKRNRKAGTKNCGPSSTTNVSVSTDE**  
**NKCVQSDIDSFFKHLIDIGLTPSSYLSIVLDDNIGCEDKAPWTTYTTTTTTTCKCNKETDKSKS**  
**QSCNTAVVNVPSPLGNTPHGYK YACECKIPTTEETCDDRKEYMNQWIIDTSKKQKSGSGGKDDYE**  
35 **LYTYNGVDVKPTTVRSNSTKLD**

>V1S1 745 amino acids | 628 aa (SEQ ID NO:44)

**DYIKDDPYSAQYTTKLSFILNPSDANTSSEKI QKNND EACNCNESGISSVGQAQTSGPSSNKT**  
**CITHSSIKANKKKKVCKDVKLG INNNDKVL RVCVIEDTSLSGVDNCCCQDLLGILQENCSDNKR**  
**GSSSNGSCNNNNEEACEKNLDEAPASLHNGYKNQCKCKSGTSRSKKKWIWKKSSGNEKGLQE**  
40 **EYANTIGLPRTQSLCLVCLHEKEGKTQHKTISTNSELLKEWIIAAFHEGKNLKTSHEKKNDN**  
**GKKLCKALEYSFADYGD L IKGTSIWDNEYTKDLELNLQKAFGKLF RKYIKKNTAEQDTSYS SLDL**  
**RESWWNTNKKYIWIAMKHGAGMNGTTCSCSGDSSNDMPTID L IPQYLRFLQEWVEHFC**  
**EQRQAKVKDVITNCKSCKESGNCKCKTECKTKCKDECEKYKTFIEDCN GGGGTAGSSWVKRW**  
**DQIYKRYSKHIEDAKRNRKAGTKNCGPSSITNAAASTDENKCVQSDIDSFFKHLIDIGLTPSS**  
45 **YLSNVLDENSCGDDKAPWTTYTTTTTTKNC DIQKDKSKSQPINTSVVNVPSPLGNTPYRYKY**  
**ACECKIPTTEESCDDRKEYMNQWSCGSARTMKRGYKNDNYELCKYNGVDVKPTTVRSNSSKLD**

>raj116\_var25 745 amino acids | 653 aa (SEQ ID NO:45)

**DYIKGDPYFAEYATKLSFILNPSDTENASETPSKYIYDEACNP NESEIASVEQAQTSGPSSNKTC**  
**ITHSSIKTNKKKECKDVKLG VRENDKDLKICVIEDTSLSGVDNCCFKDLLGILQENCSDNKRGS**  
50 **SSNDSCNNNNEEACEKNLDEALASLTNGYKCDKCKSGTSTVNKKWTRKSSGNEEGLQKEYA**  
**NTIGLPRTQSLCLVCLHEKEGKTQHK TISTNSELLKEWIIAAFHEGKNLKTSHEKKNDN GKK**  
**LCKALEYSFADYGD L IKGTSIWDNEYTKDLELNLQKAFGKLF RKYIKKNTAEQDTSYS SLDL**  
**RESWWNTNKKYIWTAMKHGAEMNGTTCSSGSGDNGDSSITGSSDSGSTTCSGDNGSISCD**  
**IPTTDFIPQYLRFLQEWVEHFCEQRQAKVKDVINSCNSC NESGGTCNGECKTKCKDECEKYK**  
55 **FIEDCN GGGGTAGSSWVKRW DQIYKRYSKHIEDAKRNRKAGTKNCGPSSITNAAASTDENKC**  
**VQSDVDSFFKHLIDIGLTPSSYLSIVLDENSCGDDKAPWTTYTTTTTTEKCNKERDKSKSQSS**

**DTLVVVNVPSPLGNTPEHYKYACECKIPTNEETCDDRDKYMNQWISDTSKKQKSGSGSKDYELYTY**  
**NGVQIKQAAGRSSTKLD**

>gi|31323048|gb|AAP37940.1| var2csa [Plasmodium falciparum] | 490 aa (SEQ ID NO:46)  
 KCDKCKSEQSKKNNNKWIWKKYSGNGEGLQKEYANTIGLPPRTQSLCLVCLHEKEGKTQHKTIISTNSEL  
 5 LKEWIIAAFHEGKNLKKRYPQKNDDNNSKLCKALEYSFADYGDLIKGTSIWDNEYTKDLELNLQKAFGK  
 LFRKYIKKNNTAEQDTSYSSSDELRESWWNTNKKYIWTAMKHGAEMNGTTCSSGSGDNGDSSCDDIPT  
 IDLIPQYLRFLQEWVEHFCKQRQAKVKDVINSCNSCKNTSGERKIGGTCNSDCEKKCKVACDAYKTFIEE  
 CRTAVGGTAGSSWVKRWDQIYKRYSKHIEDAKRNRKAGTKNCGPSSTTNAAENKCVQSDIDSFFKHLID  
 10 IGLTTPSSYLSNVLDENSCGADKAPWTTYTTYTTYTTYTTTEKCNKERDKSKSQSNTSVVVNVPSPL  
 GNTPEHYKYACECKIPTTEETCDDRKEYMNQWIIDNTKNPKGSGSTDNDYELYTYNGVQIKQAAGRSST  
 KLD

>gi|254952620|gb|ACT97140.1| VAR2CSA [Plasmodium falciparum] | 335 aa (SEQ ID  
 NO:47)  
 KCEKCKSGTSTVNNKWIWRKSSGKEGGLQKEYANTIGLPPRTQSLYLGNLPKLENVCKGVTDIIYDTKEK  
 15 FLSGCLIAAFHEGKNLKTTYLEKKNDNGKKLCKALEYSFADYGDLIKGTSIWDNEYTKDLELNLQKIFGK  
 LFRKYIKKNNTAEQDTSYSSSDELRESWWNTNKKYIWIAMKHGAGMNGTTCSSGSGDSSNDIPTTDFIP  
 QYLRFLQEWVENFCEQRQAKVKPVIENCNSCKESGGTCNGECKTKCKVACDAYKKFIDGTGSGGGSRP  
 GIAGSSWSKRWDQIYKRYSKHIEDAKRNRKAGTKNCGPSSTITNVSVSTDENKCVQS

>T2C6 745 amino acids | 637 aa (SEQ ID NO:48)  
 20 **NYIKDDPYSKEYVTKLSFIPNSSDANTSSEKIQKNNDEVCPNPESGISSVEQAQTS DPSSNKT**  
**CITHSSIKANKKKECKDVKLGVRENDKDLKICVIEHTSLSGVDNCCFKDFLRMLQEPRIDKNQ**  
**RGSSNGSCDKNSEEACEKNLDEALASLTNGYKCDKCKSEQSKKNNNKWIWKKFPGKEGGLQ**  
**EEYANTIGLPPRTQYLCLVCLDEKEGKTQELKNIRTNSELLKEWIIAAFHEGKNLKTTPQKK**  
**NDDNGKKLCKDLKYSFADYGDLIKGTSIWDNEYTKNVELNLQNNFGKLFRKYIKKNNTAEQD**  
 25 **TSYSSSDELRESWWNTNKKYIWLAMKHGAEMNSTTCCGDGSVTSGSGSSCDDIPTIDLIPQYL**  
**RFLQEWVEHFCKQRQAKVKDVITNCNSCKESGNCKCKTECKNKCKDECEKYKKFIEACGTAVG**  
**GTGTAGSPWSKRWDQIYKRYSKHIEDAKRNRKAGTKNCGPSSTTNAAENKCVQSDIDSFFKH**  
**LIDIGLTPSSYLSIVLDDNICGADKAPWTTYTTYTTTENCIDIQKTPKSQSCDTLVVVNVPSPL**  
**GNTPHGYKYACQCRTPNKQESCDDRKEYMNQWIIDNTKNPKGSGSGKDYELCKYNGVKETKPLGTL**  
 30 **KNSKLD**

>gi|254952632|gb|ACT97146.1| VAR2CSA [Plasmodium falciparum] | 330 aa (SEQ ID  
 NO:49)  
 KCDKCKSEQSKKNNNKWIWRKFPKGEGGLQKEYANTIGLPPRTQSLCLVCLHEKEGKTQHKTIISTNSEL  
 35 KEWIIAAFHEGKNLKTTYLEKKNAENKKKLCALKYSFADYGDLIKGTSIWDNEYTKDLELNLQKIFGKLF  
 RKYIKKNNTAEQDTSYSSSDELRESWWNTNKKYIWTAMKHGAGMNGTMCNADGSVTGSGSSCDDMPT  
 TDFIPQYLRFLQEWVEHFCKQRQAKVKDVIENCKSCKESGNCKCKTECKNKCDAYKTFIEECGTAVGGTAG  
 SSWVKRWDQIYKRYSKHIEDAKRNRKAGTKNCGTSSTTNAAASTAENKCVQS

>gi|90193487|gb|ABD92339.1| erythrocyte membrane protein 1 [Plasmodium falciparum] |  
 269 aa (SEQ ID NO:50)  
 40 NYIKDDPYSKEYVTKLSFILNSSDAENASETPSKYYDEACNCNESGISSVEQASISDRSSQKACNTHSFIG  
 ANKKKVCKHVKLGVRENDKDLKICVIEDDSLGRVENCCFKDFLRMLQEPRIDKNQRGSSSNDSCNNNE  
 EACEKNLDEALASLHNGYKNQKCKSEQSKKNNNKWIWKKSSGKEGGLQKEYANTIGLPPRTQSLCLVCL  
 HEKEGKTQHKTIISTNSELLKEWIIADFHEGKNLKTTYLEKKKGDNKKLCKALKYSFADY

>gi|254952646|gb|ACT97153.1| VAR2CSA [Plasmodium falciparum] | 347 aa (SEQ ID  
 NO:51)  
 45 KCDKCKSEQSKKNNNKWIWKKSSGKEGGLQKEYANTIALPPRTQSLCLVCLHEKEGKTQHKTIISTNSE  
 LLKEWIIADFHEGKNLKTTYLEKQNADNGKKNADNNSKLCKDLKYSFADYGDLIKGTSIWDNEYTKDLEL  
 NLQQIFGKLFRKYIKKNIASDENTLYSSSDELRESWWNTNKKYIWTAMKHGAEMNGTTCSSGSGDSSSG  
 50 ENQTNSCDDIPTIDLIPQYLRFLQEWVEHFCEQRQAKVKDVITNCKSCKESGGTCNSDCKTKCKGECEKY  
 KKFIECKGGGTEGTSGSSWVKRWYQIYMRYSKYIEDAKRNRKAGTKSCGTSSGANSGVTTTESKCVQ  
 S

>gi|90193485|gb|ABD92338.1| erythrocyte membrane protein 1 [Plasmodium falciparum] |  
 269 aa (SEQ ID NO:52)  
 55 DYIKDDPYSKEYTTKLSFILNSSDANTSSEKIQKNNDEVCPNNESEISSVEQAQTSRPSSNKTCTHSSIK  
 ANKKKVCKDVKLGVRENDKVLRVCVIEHTSLSGVENCCQDLLGILQENCSDNKRGSNGSCDKNSEE

ACEKNLDEALASLTNCYKNQKCKSEQSKKNNNKWIWKKSSGNEKGLQKEYANTIGLPPRTQSLCLVCLH  
EKEGKTQELKNISTNSELLKEWIIAAFHEGKNLKTTPQNKNDDNGKKLFDLKYSFADY

>MTS1 745 amino acids | 646 aa (SEQ ID NO:53)

**DYIKDDPYSKEYTTKLSFILNSSDANTSSEKIQKNNDEVCPNNESEISSVEQAQTSRPSSNKT  
ITHSSIKANKKKVCKDVKLGVRENDKVLRCVIEHTSLSGVENCCCQDLLGILQENCSDNKR  
SSNGSCDNSEEACEKNLDEALASLTNCYKNQKCKSEQSKKNNNKWIWKKSSGKEGGLQKE  
YANTIGLPPRTQSLYLGNLPKLENVCKGVTDINFDTKEKFLAGCLIAAFHEGKNLKTTYLEKKN  
DDNGKKLCKALEYSFADYGDLIKGTSIWDNEYTKDLELNLQKAFGKLFRKYIKKNNTAEQDTS  
YSSLDELRESWWNTNKKYIWTAMKHGAGMNGTTCSSGSGDSSNDIPTTDFIPQYLRFLQEW  
VENFCEQRQAKVKDVIENCNSCKNTSGERKIGDTCNSDCEKKCKDECEKYKKFIEDCKGGDGT  
AGSSWVKRWDQIYKRYSKHIEDAKRNRKAGTKNCGITTGTISGESSGATSGVTTTENKCVQS  
DIDSFFKHLIDIGLTTTPSSYLSNVLDDNICGEDNAPWTTYTTTTEKCNKETDKSKSQSNTAV  
VVNVPSPLGNTPHGYKYACECKIPTTEETCDDRKEYMNQWSCGSAQTVRDRSGKDDYELCKYNGVQI  
KQAAGTLKNSKLD**

>Q8I639 (Q8I639\_PLAF7) Plasmodium falciparum (isolate 3D7), 632 aa extracellular part  
(SEQ ID NO:54)

**NYIKGDPYFAEYATKLSFILNSSDANNPSEKIQKNNDEVCPNNESEIASVEQEQISDPSSNKT  
ITHSSIKANKKKVCKHVKLGVRENDKDLRCVIEHTSLSGVENCCCQDFLRILQENCSDNKS  
SSNGSCNNKNQEACEKNLEKVLASLTNCYKCDKCKSEQSKKNNKNWIWKKSSGKEGGLQK  
EYANTIGLPPRTQSLCLVCLDEKGGKTQELKNIRTNSELLKEWIIAAFHEGKNLKPSHEKKND  
DNGKKLCKALEYSFADYGDLIKGTSIWDNEYTKDLELNLQKIFGKLFRKYIKKNNTAEQDTSY  
SLDELRESWWNTNKKYIWLAMKHGAGMNSTTCCGDGSVTSGSGSSCDDIPTIDILIPQYLRFLQ  
EWVEHFCKQRQEKVKPVIENCCKESGGTCNGECKTECKNKCEVYKKFIEDCKGGDGTAGSS  
WVKRWDQIYKRYSKYIEDAKRNRKAGTKNCGPSSTTNAENKCVQSDIDSFFKHLIDIGLTT  
PSSYLSIVLDDNICGADKAPWTTYTTTTEKCNKETDKSKLQQCNTAVVVNVPSPLGNTPHG  
YKYACQCKIPTNEETCDDRKEYMNQWSCGSARTMKRGYKNDNYELCKYNGVDVKPTTVRSNSSKLD**

>Q8I639 (Q8I639\_PLAF7) Plasmodium falciparum (isolate 3D7), complete 2730 aa  
extracellular part (SEQ ID NO:55)

**MDKSSSIANKIEAYLGAKSDDSKIDQSLKADPSEVQYYGSGGDGYLRKNICKITVNHSDSGTNDPCDRIP  
PPYGDNDQWKCAIILSKVSEKPENVFVPPRRQRMCIINLEKLNVDKIRDKHAFLADVLLTARNERIVQ  
NHPDTNSSNVCNALERSFADIADIIRGTDLWKGNTSNLEQNLMFAKIRENDKVLQDKYPKDQNYRKL  
REDWWNANRQKVWEVITCGARSNDLLIKRGWRTSGKSNGDNKLELCRCGHHYEEKVPTKLDYVPQFLR  
WLTEWIEDFYREKQNLIDDMERHREECTSEDHKSKEGTSYCSTCKDKCKKYCECVKKWKSEWENQKNK  
YTELYQQNKNETSQKNTSRYDDYVKDFFKKLEANYSSLENYIKGDPYFAEYATKLSFILNSSDANNPSEKI  
QKNNDEVCPNNESEIASVEQEQISDPSSNKT  
CITHSSIKANKKKVCKHVKLGVRENDKDLRCVIEHTSL  
SGVENCCCQDFLRILQENCSDNKS  
GSSSSNGSCNNKNQEACEKNLEKVLASLTNCYKCDKCKSEQSKKN  
NKNWIWKKSSGKEGGLQKEYANTIGLPPRTQSLCLVCLDEKGGKTQELKNIRTNSELLKEWIIAAFHEG  
KNLKPSHEKKND  
DNGKKLCKALEYSFADYGDLIKGTSIWDNEYTKDLELNLQKIFGKLFRKYIKKNNTAEQ  
DTSYSSLDELRESWWNTNKKYIWLAMKHGAGMNSTTCCGDGSVTSGSGSSCDDIPTIDILIPQYLRFLQ  
EWVEHFCKQRQEKVKPVIENCCKESGGTCNGECKTECKNKCEVYKKFIEDCKGGDGTAGSSWVKRW  
DQIYKRYSKYIEDAKRNRKAGTKNCGPSSTTNAENKCVQSDIDSFFKHLIDIGLTTTPSSYLSIVLDDNIC  
GADKAPWTTYTTTTEKCNKETDKSKLQQCNTAVVVNVPSPLGNTPHGYKYACQCKIPTNEETCDDRKE  
YMNQWSCGSARTMKRGYKNDNYELCKYNGVDVKPTTVRSNSSKLD  
DKDVTFNLFQWNKEIQYQIEQ  
YMTNTKISCNNEKNVLSRVSEAAQPKFSDNERDRNSITHEDKNCKEKCCKCYSLWIEKINDQWDKQKD  
NYNKFQRKQIYDANKGSQNKVVLSNLFSCWEEYIQKYFNGDWSKIKNIGSDTFEFLIKKCGND  
SGDGETIFSEKLNNAEKKCKENESTNNKMSSETSCDCSEPIYIRGCQPKIYDGKIFPGKGGEKQWICKDTII  
HGDTNGACIPPRQNLVCGELWDKRYGGRSNIKNDTKESLKQKIKNAIQKETELLYEYHDKGTAIISRNP  
MKGQKEKEEKNND  
SNGLPKGFCHAVQRSFIDYKNMILGTSVNIYIYIGKLQEDIKKIEKGTTKQNGKTV  
GSGAENVNAWWKGIEGEMWDAVRCAITKINKQKKNGTFSIDECGIFPPTGNDEDQSVSWFKEWSEQF  
CIERLQYEKNIRDACTNNGQGDQKIQGDCKRKCEEYKKYISEKKQEWQDKQTKYENKYVGKSASDLLKEN  
YPECISANFDFIFNDNIEYKTYYPYGDYSSICSCEQVKYIYENNAEKKNNKSLCHEKGNDRTWSKKYIKKL  
ENGRTLEGVYVPPRRQQLCLYELFPPIIKNKNNDITNAKKELLETLQIVAEREAYYLWKQYHAHNDTTYLAHK  
KACCAIRGSFYDLEDIIKGNLDLVHDEYTKYIDSKLNEIFDSSNKNNDIETKRARTDWWENEIAVPNITGAN**

KSDPKTIRQLVWDAMQSGVRKAIDEEKEKKPNENFPPCMGVQHIGIAKPQFIRWLEEWTFNEFCEKYTKY  
FEDMKSNCNLKRGADDCDDNSNIECKKACANYTNWLNPKRIEWNGMSNYNKIYRKSKESEDGKDYS  
MIMEPTVIDYLNKRCNGEINGNYICCSCKNIGENSTSGTVNKKLQKKETQCEDNKGPLDLMNKVLNKM  
PKYSEHKMKCTEVYLEHVEEQLEIDNAIKDYKLYPLDRCFDDKSKMKVCDLIGDAIGCKHKTKLDELDE  
5 WNDVDMRDPYNKYKGVLIPIPRRRQLCFSRIVRGPANLRNLKEFKEEILKGAQSEGKFLGNYYNEDKDKEK  
ALEAMKNSFYDYEYIIGSDMLTNIQFKDIKRKLDRLLEKETNNTKVDWETNKKSIWNAMLCGYKKS  
GNKIIDPSWCTIPTTETPPQFLRWIKWGTNVCIQKEEHKEYVSKSCSNVTNLGAQESSEKNTSEIKKY  
QEWSRKRSIQWEAISEGYKKYKGMDEFKNTFKNIKEPDANEPNANEYLKKHCSKPCPGFNDMQEITKYT  
NIGNEAFKQIKEQVDIPAELEDVIYRLKHHEYDKGNDYICNKYKNINVMKKNNDDTWTDLVKNSSDINK  
10 GVLLPPRRKNLFLKIDESDICKYKRDPKLFKDFIYSSAISEVERLKKVYGEAKTKVHAMKYSFADIGSIIKG  
DDMMENNSSDKIGKILGDGVGQNEKRKKWWDMMNKYHIWESMLCGYKHAYGNISENDRKMLDIPNND  
DEHQFLRWFQEWTFNCTKRNELYENMVTACNSAKCNTSNGSVDKKECTEACKNYSNFILIKKKEYQSL  
NSQYDMNYKETKAEEKESPEYFKDKCNGECSCLSEYFKDETRWKNPYETLDDTEVKNNCMCKPPPPASN  
NTSDILQKTIPFGIALALGSIAFLFMKKKPKTPVDLLRVLDIPKGDYGIPTPKSSNRYIPYASDRYKGGKTYIY  
15 MEGDTSGDDDKYIWDL

>FCR3 (SEQ ID NO:56) complete 2734 aa extracellular part (577 aa highlighted corr. ID1-DBL2b)

MDSTSTIANKIEEYLGAKSDDSKIDELLKADPSEVEYYRSGGDG DYLNKNICKITVNHSDSGKYDPCEKKL  
20 PPYDDNDQWKCQONSSDGSGKPENICVPPRRERLCTYNLENLKF DKIRDNN AFLADVLLTARNEGEKIVQ  
NHPDTNSSNVCNALERSFADLADIIRGTDQWKG TNSNLEKNLQMF AKIRENDKVLQDKYPKDQKYTKL  
REAWWNANRQKVWEVITCGARSNDLLIKRGWRTSGKSDRKKNFELCRKCGHYEKEVPTKLDYVPQFLR  
WLTEWIEDFYREKQNLIDDMERHREECTREDHKSKEGTSYCSTCKDKCKKYCECVKKWKTEWENQENK  
YKDLYEQNKNTSQKNTSRYDDYVKDFFEKLEANYSSLE**NYIKGDPYFAEYATKLSFILNP****SDANNP**  
25 **SGETANHND EACNCNESG****ISSVGQAQTS GPSSNKT CITHSSIKTNKKKECKDVKLGVRENDKD**  
**LKICVIEDTSLSGVDNCCCQDLLGILQENCSDNKR GSSSNDSCDNKNQDECQKKLEKVFASLT**  
**NGYKCDKCKSGTSRSKKKWIWKSSGNEEGLQEYANTIGLPPRTQSLYLG NLPKLENVCEDV**  
**KDINFDTKEKFLAGCLIVSFHEGKNLKKRYPQNKNSGNKENLCKALEYSFADYGD LIKGTSIW**  
**DNEYTKDLELNLQNNFGKLF GKYIKKNNTAEQDTSYSSLDELRESWWNTNKKYIWTAMKHG**  
30 **AEMNITTCNADGSVTGSGSSCDDIPTIDLIPQYLRFLQEWVENFCEQRQAKVKDVITNCKSCK**  
**ESGNKCKTECKTKCKDECEKYKKFIEACGTAGGGIGTAGSPWSKRWDQIYKRYSKHIEDAKR**  
**NRKAGTKNCGTSSTTNAASTDENKCVQSDIDSFFKHLIDIGLTPSSYLSNVLDN****ICGADK**  
**APWTTYTTYTTTEKCNKERDKSKSQSSDTLVVVNVPSPLGNTPYRYKYACQCKIPTNEETCDDRK**  
EYMNQWSCGSARTMKRGYKNDNYELCKYNGVDVKPTTVRSNSSKLDGNDVTFNLF EQWNKEIQYQIE  
35 QYMTNANISCIDEKEVLDSVSDEGTPKVRGGYEDGRNNNTDQGTNCKEKKCYKLWIEKINDQWGKQK  
DNYNKFRSKQIYDANKGSQNKKVVSLSNFFSCWEEYIQKYFNGDWSKIKNIGSDTFEFLIKKCGNNSA  
HGEEIFNEKLKNAEKKCKENESTDTNINKSETSCDLNATNYIRGCQSKTYDGKIFPGKGGEKQWICKDTII  
HGD TNGACIPPR TQNL CVGELWDKSYGGRSNIKNDTKELLKEIKNAIHKETELLYEYHDTGTAIISKNDK  
KGQKGKNDPNGLPKGFCHAVQRSFIDYKNMILGTSVNIYEHIGKLQEDIKKIIEKGTPQQKDKIGGVGSS  
40 TENVNAWWKGIEREMWDAVRCAITKINKNNNSIFNGDECGVSPPTGNDEDQSVSWFKEWGEQFCIER  
LRYEQNIREACTINGKNEKKCINSKSGQGDKIQGACKRKCEKYKKYISEKKQEWQDKQTKYENKYVGKS  
ASDLLKENYPECISANFDFIFNDNIEYKTYYPYGDYSSICSCEQVKYKYNNNAEKKNNKSLCYEKDNDMTW  
SKKYIKKLENGRSLEGVYVPPRRQQLCLYELFPIIKNEEGMEKAKEELLETLQIVAEREAYYLWKQYNPTG  
KGIDDANKKACCAIRGSFYDLEDIIKGNLDLVHDEYTKYIDSKLNEIFGSSDTNDIDTKRARDWENETIT  
45 NGTDRKTIRQLVWDAMQSGVRYAVEEKNENFPLCMGVEHIGIAKPQFIRWLEEWTFNEFCEKYTKYFEDM  
KSKCDPPKRADTCG DNSNIECKKACANYTNWLNPKRIEWNGMSNYNKIYRKSKESEGGKDYSMIMA  
PTVIDYLNKRCHGEINGNYICCSCKNIGAYNTTSGTVNKKLQKKETECEEEKGPLDLMNEVLNKMMDKKYS  
AHKMKCTEVYLEHVEEQLEIDNAIKDYKLYPLDRCFDDQTKMKVCDLIADAIGCKDKTKLDELDEWND  
MDLRGTYNKHKGVLIPPRRRQLCFSRIVRGPANLRSLNEFKEEILKGAQSEGKFLGNYYKEHKDEKALEA  
50 MKNSFYDYEDIIGKTDMLTNI EFKDIKILDRLLEKETNNTKKAEDWWKTNKKSIWNAMLCGYKKSGNKI  
IDPSWCTIPTTETPPQFLRWIKWGTNVCIQKQEHEKEYVSKSCSNVTNLGAQASESNNCTSEIKKYQEWS  
RKRSIRWETISKRYKKYKRM DILKDVKEPDANTYLREHCSKCPG FNDMEEMNNNEDNEKEAFKQIKEQ  
VKIPAELEDVIYRIKHHEYDKGNDYICNKYKNIHDRMKKNNGNFVTDNFVKKSWEISNGVLIPPRRKNLFL  
YIDPSKICEYKDPKLFKDFIYWSAFTEVERLKKAYGGARAKVVHAMKYSFTDIGSIIKGDDMMMEKNSSD  
55 KIGKILGDTDGQNEKRKKWWDMMNKYHIWESMLCGYREAEGDTETNENCRFPDIESVPQFLRWFQEWSE  
NFCDRRQKLYDKLNSECISAECTNGSVDNSKCTHACVNYKNYILT KTEYEIQTNKYDNEFKNKNNSNDKD  
APDYLKECNDNKCECLNKHIDDKNKTWKNPYETLEDTFKSKCDCPKPLPSPIKPDDLPPQADEPFDPTIL

QTTIPFGIALALGSIAFLFMKVIIYIYVCCICMYVCMYVCMYVCMYVCMHVCMLCVYVIYVFKICIYI  
EKEKRKK

>BPTI, protease inhibitor (SEQ ID NO:57)

5 RPDFCLEPPYTGPCKARIIRYFYNAGLCQTFVYGGCRAKRNNFKSAEDCMRTCGGA

>PE38, *Pseudomonas* exotoxin A (SEQ ID NO:58), (underlining of KDEL represent a signal sequence, which may be optional for the constructs according to the present invention)

10 RHRQPRGWEQLEQCGYPVQRLVALYLAARLSWNQVDQVIRNALASPGSGDLGEAIREQPEQARLALT  
AAAESERFVRQGTGNDEAGAANGPADSGDALLERNYPTGAFLGDGGDISFSTRGTQNWTVRLLQAH  
RQLEERGYVFGYHGTFLAAQSIIVFGGVRARSQDLDAIWRGFYIAGDPALAYGYAQDQEPDARGRIRNG  
ALLRVYVPRSSLPGFYRTSLTLAAPEAAGEVERLIGHPLRLDAITGPEEEGGRLETILGWPLAERTVVIPS  
AIPTDPRNVGGDLDPSSIPDKEQAISALPDYASQPGKPPRKKDEL

15 >PE38LR, variant of PE38 (SEQ ID NO:59) (underlining of KDEL represent a signal sequence, which may be optional for the constructs according to the present invention)

RHRQPRGWEQLYPTGAFLGDGGDISFSTRGTQNWTVRLLQAHQRQLEERGYVFGYHGTFLAAQSIIV  
FGGVRARSQDLDAIWRGFYIAGDPALAYGYAQDQEPDARGRIRNGALLRVYVPRSSLPGFYRTSLTLAAP  
EAAGEVERLIGHPLRLDAITGPEEEGGRLETILGWPLAERTVVIPSAIPTDPRNVGGDLDPSSIPDKEQAI  
20 SALPDYASQPGKPPRKKDEL

Sequences of VAR2CSA polypeptides fused with truncated fragments of *Pseudomonas* exotoxin A (PE38)

25 Fused VAR2CSA-PE38 proteins may have modifications such as a protease inhibitor (BPTI) in the N-terminal and/or an optimized PE38 sequence that is less immunogenic (PE38LR)

>BPTI-ID1-ID2aFCR3-PE38LR, (SEQ ID NO:60) underlined sequence corresponds to the ID1 domain of FCR3, sequence in bold corresponds to DBL2Xb domain of FCR3, underlined and bold sequence is ID2a

30 RPDFCLEPPYTGPCKARIIRYFYNAGLCQTFVYGGCRAKRNNFKSAEDCMRTCGGANYIKGDPYFAEYA  
TKLSFILNPSDANNPSGETANHNDACNCNESGISSVGQAQTSGPSSNKTCTHSSIKTNKKKECKDVKL  
GVRENDKDLKICVIEDTSLSGVDNCCCQDLLGILQENCSDNKRGSNDSCDNKNQDECQKKLEKVFAS  
LTNGYKCDKCKSGTSRSKKKWIWKSSGNEEGLQEYANTIGLPPRTQSLYLGNLPKLENVCE  
35 DVKDINFDTKEKFLAGCLIVSFHEGKNLKKRYPQNKNSGNKENLCKALEYSFADYGDLIKGTISI  
WDNEYTKDLELNLQNNFGKLFQYIKKNNTAEQDTSYSSLDLRESWWNTNKKYIWTAMKH  
GAEMNITTCNADGSVTGSGSSCDDIPTIDLIPQYLRFLQEWVENFCEQRQAKVKDVITNCKSC  
KESGNKCKTECKTKCKDECEKYKKFIEACGTAGGGIGTAGSPWSKRWDQIYKRYSKHIEDAK  
RNRKAGTKNCGTSSTTNAAASTDENKCVQSDIDSFFKHLIDIGLTPSSYLSNVLDDNICGAD  
40 KAPWTTYTTYTTTEKCNKERDKSKSQSSDTLVVVNVPSPLGNTPYRYKYACOCKIPTNEETCD  
DRKEYMNQWSCGSARTMKRGYKNDNYELCKYNGVDVKPTTVRSNSSKLD RHRQPRGWEQLYP  
TGAFLGDGGDISFSTRGTQNWTVRLLQAHQRQLEERGYVFGYHGTFLAAQSIIVFGGVRARSQDLDAI  
WRGFYIAGDPALAYGYAQDQEPDARGRIRNGALLRVYVPRSSLPGFYRTSLTLAAPEAAGEVERLIGHPL  
LRLDAITGPEEEGGRLETILGWPLAERTVVIPSAIPTDPRNVGGDLDPSSIPDKEQAISALPDYASQPGKPP  
RKDEL

45 >BPTI-ID1-ID2aFCR3-PE38 (SEQ ID NO:61)

RPDFCLEPPYTGPCKARIIRYFYNAGLCQTFVYGGCRAKRNNFKSAEDCMRTCGGANYIKGDPYFAEYA  
TKLSFILNPSDANNPSGETANHNDACNCNESGISSVGQAQTSGPSSNKTCTHSSIKTNKKKECKDVKL  
GVRENDKDLKICVIEDTSLSGVDNCCCQDLLGILQENCSDNKRGSNDSCDNKNQDECQKKLEKVFAS  
50 LTNGYKCDKCKSGTSRSKKKWIWKSSGNEEGLQEYANTIGLPPRTQSLYLGNLPKLENVCE  
DVKDINFDTKEKFLAGCLIVSFHEGKNLKKRYPQNKNSGNKENLCKALEYSFADYGDLIKGTISI  
WDNEYTKDLELNLQNNFGKLFQYIKKNNTAEQDTSYSSLDLRESWWNTNKKYIWTAMKHGAEMNITTCNADGSVTGSGS  
CDDIPTIDLIPQYLRFLQEWVENFCEQRQAKVKDVITNCKSCKESGNKCKTECKTKCKDECEKYKKFIEA  
CGTAGGGIGTAGSPWSKRWDQIYKRYSKHIEDAKRNRKAGTKNCGTSSTTNAAASTDENKCVQSDIDS  
55 FFKHLIDIGLTPSSYLSNVLDDNICGADKAPWTTYTTYTTTEKCNKERDKSKSQSSDTLVVVNVPSPLGN  
TPYRYKYACOCKIPTNEETCD DRKEYMNQWSCGSARTMKRGYKNDNYELCKYNGVDVKPTTVRSNSSKL  
DPEGGSALAALTAHQACHLPLETFTRHRQPRGWEQLEQCGYPVQRLVALYLAARLSWNQVDQVIRNALAS  
PGSGGDLGEAIREQPEQARLALTAAAESERFVRQGTGNDEAGAANGPADSGDALLERNYPTGAFLGD

GGDISFSTRGTQNWTVRLLQAHQRQLEERGYVFGYHGTFLAAQSIVFGGVRARSQDLDAIWRGFYIA  
GDPALAYGYAQDQEPDARGRIRNGALLRVYVPRSSLPGFYRTSLTAAPEAAAGEVERLIGHPLPLRLDAITG  
PEEEGGRLETILGWPLAERTVVIPSAIPTDPRNVGGDLDPSSIPDKEQAISALPDYASQPGKPPRKDEL

5 >ID1-ID2aFCR3-PE38 (SEQ ID NO:62)  
NYIKGDPYFAEYATKLSFILNPSDANNPSGETANHNDACNCNESGISSVGQAQTS GPSSNKCITHSSIK  
TNKKKECKDVKLGVRENDKDLKICVIEDTSLSGVDNCCCQDLLGILQENCSDNKRGSNDSCDNKNQ  
DECQKKLEKVFASLTNGYKCDKCKSGTSRSKKKWIWKSSGNEEGLQEEYANTIGLPPRTQSLYLGNLKP  
10 LENVCEDVKDINFDTKEKFLAGCLIVSFHEGKNLKKRYPQNKNSGNKENLCKALEYSFADYGDLIKGTISI  
DNEYTKDLELNLQNNFGKLFYIKKNNTAEQDTSYSSLDELRESWWNTNKKYIWTAMKHGAEMNITTC  
NADGSVTGSGSSCDDIPTIDLIPQYLRFLQEWVENFCEQRQAKVKDVITNCKSCKESGNKCKTECKTKCK  
DECEKYKKFIEACGTAGGGIGTAGSPWSKRWDQIYKRYSKHIEDAKRNRKAGTKNCGTSSTTNAASTD  
ENKCVQSDIDSFFKHLIDIGLTPSSYLSNVLDDNICGADKAPWTTYTTYTTTEKCNKERDKSQSQSSDTL  
VVVNVPSPLGNTPYRYKYACQCKIPTNEETCDDRKEYMNQWSCGSARTMKRGYKNDNYELCKYNGVDV  
15 KPTTVRSNSSKLDPEGGLAALTAHQACHLPLETFRHRQPRGWEQLEQCGYPVQRLVALYLAARLSWN  
QVDQVIRNALASPGSGGDLGEAIREQPEARLALTAAAESERFVRQGTGNDEAGAANGPADSGDALLE  
RNYPTGAEFLDGGDISFSTRGTQNWTVRLLQAHQRQLEERGYVFGYHGTFLAAQSIVFGGVRARSQ  
DLDAIWRGFYIAGDPALAYGYAQDQEPDARGRIRNGALLRVYVPRSSLPGFYRTSLTAAPEAAAGEVERLI  
GHPLPLRLDAITGPEEEGGRLETILGWPLAERTVVIPSAIPTDPRNVGGDLDPSSIPDKEQAISALPDYASQ  
20 PGKPPRKDEL

>ID1-ID2aFCR3-PE38LR (SEQ ID NO:63)  
NYIKGDPYFAEYATKLSFILNPSDANNPSGETANHNDACNCNESGISSVGQAQTS GPSSNKCITHSSIK  
TNKKKECKDVKLGVRENDKDLKICVIEDTSLSGVDNCCCQDLLGILQENCSDNKRGSNDSCDNKNQ  
25 DECQKKLEKVFASLTNGYKCDKCKSGTSRSKKKWIWKSSGNEEGLQEEYANTIGLPPRTQSLYLGNLKP  
LENVCEDVKDINFDTKEKFLAGCLIVSFHEGKNLKKRYPQNKNSGNKENLCKALEYSFADYGDLIKGTISI  
DNEYTKDLELNLQNNFGKLFYIKKNNTAEQDTSYSSLDELRESWWNTNKKYIWTAMKHGAEMNITTC  
NADGSVTGSGSSCDDIPTIDLIPQYLRFLQEWVENFCEQRQAKVKDVITNCKSCKESGNKCKTECKTKCK  
DECEKYKKFIEACGTAGGGIGTAGSPWSKRWDQIYKRYSKHIEDAKRNRKAGTKNCGTSSTTNAASTD  
30 ENKCVQSDIDSFFKHLIDIGLTPSSYLSNVLDDNICGADKAPWTTYTTYTTTEKCNKERDKSQSQSSDTL  
VVVNVPSPLGNTPYRYKYACQCKIPTNEETCDDRKEYMNQWSCGSARTMKRGYKNDNYELCKYNGVDV  
KPTTVRSNSSKLDHRHRQPRGWEQLYPTGAEFLDGGDISFSTRGTQNWTVRLLQAHQRQLEERGYVFG  
YHGTFLAAQSIVFGGVRARSQDLDAIWRGFYIAGDPALAYGYAQDQEPDARGRIRNGALLRVYVPRSSL  
PGFYRTSLTAAPEAAAGEVERLIGHPLPLRLDAITGPEEEGGRLETILGWPLAERTVVIPSAIPTDPRNVGGD  
35 LDPSSIPDKEQAISALPDYASQPGKPPRKDEL

>BPTI-DBL1-ID2aFCR3-PE38LR (SEQ ID NO:64)  
RPDFCLEPPYTGPCKARIIRYFYNAGLCQTFVYGGCRAKRNNFKSAEDCMRTCGGANHSDSGKYDPCE  
KKLPPYDDNDQWKCCQNSSDGSCKPENICVPPRRERLCTYNLENLKFDKIRDNNAFLADVLLTARNEGE  
40 KIVQNHPTDNSSNVCNALERSFADLADIIRGTDQWKGTSNLEKNLQMFAKIRENDKVLQDKYPKDQK  
YTKLREAWWNANRQKVWEVITCGARSNDLLIKRGWRTSGKSDRKKNFELCRKCGHYEKEVPTKLDYVP  
QFLRWLTWIEDFYREKQNLIDDMERHREECTREDHKSKEGTSYCSTCKDKCKKYCECVKKWKTEWEN  
QENKYKDLYEQNKNTSQKNTSRYDDYVKDFFEKLEANYSSLENYIKGDPYFAEYATKLSFILNPSDANNP  
SGETANHNDACNCNESGISSVGQAQTS GPSSNKCITHSSIKTNKKKECKDVKLGVRENDKDLKICVIE  
45 DTSLSGVDNCCCQDLLGILQENCSDNKRGSNDSCDNKNQDECQKKLEKVFASLTNGYKCDKCKSGT  
SRSKKKWIWKSSGNEEGLQEEYANTIGLPPRTQSLYLGNLPLENVCEDVKDINFDTKEKFLAGCLIVSF  
HEGKNLKKRYPQNKNSGNKENLCKALEYSFADYGDLIKGTISIWDNEYTKDLELNLQNNFGKLFYIKKN  
NTAEQDTSYSSLDELRESWWNTNKKYIWTAMKHGAEMNITTCNADGSVTGSGSSCDDIPTIDLIPQYLR  
FLQEWVENFCEQRQAKVKDVITNCKSCKESGNKCKTECKTKCKDECEKYKKFIEACGTAGGGIGTAGSP  
50 WSKRWDQIYKRYSKHIEDAKRNRKAGTKNCGTSSTTNAASTDENKCVQSDIDSFFKHLIDIGLTPSSY  
LSNVLDDNICGADKAPWTTYTTYTTTEKCNKERDKSQSQSSDTLVVVNVPSPLGNTPYRYKYACQCKIPT  
NEETCDDRKEYMNQWSCGSARTMKRGYKNDNYELCKYNGVDVKPTTVRSNSSKLDHRHRQPRGWEQLY  
PTGAEFLDGGDISFSTRGTQNWTVRLLQAHQRQLEERGYVFGYHGTFLAAQSIVFGGVRARSQDLDA  
AIWRGFYIAGDPALAYGYAQDQEPDARGRIRNGALLRVYVPRSSLPGFYRTSLTAAPEAAAGEVERLIGH  
55 PLPLRLDAITGPEEEGGRLETILGWPLAERTVVIPSAIPTDPRNVGGDLDPSSIPDKEQAISALPDYASQPGK  
PPRKDEL

>BPTI-DBL1-ID2aFCR3-PE38 (SEQ ID NO:65)

RPDFCLEPPYTGPCKARIIRYFYNAKAGLCQTFVYGGCRAKRNNFKSAEDCMRTCGGANHSDSGKYDPCE  
KKLPPYDDNDQWKCQQNSSDGSQKPENICVPPRRERLCTYNLENLKF DKIRDNN AFLADVLLTARNEGE  
KIVQNHPTDNSSNVCNALERSFADLADIIRGTDQWKG TNSNLEKNLQMF AKIRENDKVLQDKY PKDQK  
YTKLREAWWNANRQKVWEVITCGARSNDLLIKRGWRTSGKSDRKKNFELCRKCGHYEKEVPTKLDYVP  
5 QFLRWLTEWIEDFYREKQNLIDDMERHREECTREDHKSKEGTSYCSTCKDKCKKYCECVKKWKTEWEN  
QENKYKDLYEQNKNTSQKNTSRYDDYVKDFFEKLEANYSSLENYIKGDPYFAEYATKLSFILNPSDANNP  
SGETANHND EACNCNESGISSVGQAQTS GPSSNKT CITHSSI KTNKKKECKDVKLG VRENDKDLKICVIE  
DTSLSGVDNCCCQDLLGILQENCSDNKR GSSSNDSCDNKNQDECQKKLEKVFASLTNGYKCDKCKSGT  
SRSKKKWIWKSSGNEEGLQEEYANTIGLPPRTQSLYLGNLPKLENVCEDVKDINFDTKEKFLAGCLIVSF  
10 HEGKNLKKRYPQNKNSGNKENLCKALEYSFADYGD LKGT SIWDNEYTKDLELNLQNNFGKLF GKYIKKN  
NTAEQDTSYSSLD ELRESWWNTNKKYIWTAMKHGAEMNITTCNADGSVTGSGSSCDDIPTIDLIPQYLR  
FLQEWVENFCEQRQAKVKDVITNCKSCKESGNKCKTECKTKCKDECEKYKKFIEACGTAGGGIGTAGSP  
WSKRWDQIYKRYSKHIEDAKRNRKAGTKNCGTSSTTNA AASTDENKCVQSDIDSFFKHLIDIGLTPSSY  
LSNVLDDNICGADKAPWTTYTTYTTTEKCNKERDKSKSQSSDTLVVVNVPSPLGNTPYRYKYACQCKIPT  
15 NEETCDDRKEYMNQWSCGSARTMKRGYKNDNYELCKYNGVDVKPTTVRSNSSKLDPEGGS LAALTAHQ  
ACHLPLETFTRHRQPRGWEQLEQCGYPVQRLVALYLAARLSWNQVDQVIRNALASPGSGGDLGEAIREQ  
PEQARLALT LAAAESERFVRQGTGNDEAGAA nGPADSGDALLERNYPTGAEFLGDGGDISFSTRGTQNW  
TVERLLQAHRQLEERGYV FVG YHGT FLEAAQSIVFGGVRARSQDLDAIWRGFYIAGDPALAYGYAQDQEP  
DARGRIRNGALLRVYVPRSSLPGFYRTSLT LAAPEAAGEVERLIGHPLPLRLDAITGPEEEGGRLETILGWPL  
20 AERTVVI PSAIPTDPRNVGGDLDPSSIPDKEQAISALPDYASQPGKPPRKDEL

>DBL1-ID2aFCR3-PE38LR (SEQ ID NO:66)

NHSDSGKYDPCEKKLPPYDDNDQWKCQQNSSDGSQKPENICVPPRRERLCTYNLENLKF DKIRDNN AFL  
ADVLLTARNEGEKIVQNHPTDNSSNVCNALERSFADLADIIRGTDQWKG TNSNLEKNLQMF AKIREND  
25 KVLQDKY PKDQKYTKLREAWWNANRQKVWEVITCGARSNDLLIKRGWRTSGKSDRKKNFELCRKCGHY  
EKEVPTKLDYVPQFLRWLTEWIEDFYREKQNLIDDMERHREECTREDHKSKEGTSYCSTCKDKCKKYCEC  
VKKWKTEWENQENKYKDLYEQNKNTSQKNTSRYDDYVKDFFEKLEANYSSLENYIKGDPYFAEYATKLS  
SFILNPSDANNPSGETANHND EACNCNESGISSVGQAQTS GPSSNKT CITHSSI KTNKKKECKDVKLGVR  
ENDKDLKICVIEDTSLSGVDNCCCQDLLGILQENCSDNKR GSSSNDSCDNKNQDECQKKLEKVFASLTN  
30 GYKCDKCKSGTSRSKKKWIWKSSGNEEGLQEEYANTIGLPPRTQSLYLGNLPKLENVCEDVKDINFDTK  
EKFLAGCLIVSFHEGKNLKKRYPQNKNSGNKENLCKALEYSFADYGD LKGT SIWDNEYTKDLELNLQNN  
FGKLF GKYIKKNNTAEQDTSYSSLD ELRESWWNTNKKYIWTAMKHGAEMNITTCNADGSVTGSGSSCD  
DIPTIDLIPQYLRFLQEWVENFCEQRQAKVKDVITNCKSCKESGNKCKTECKTKCKDECEKYKKFIEACGT  
AGGGIGTAGSPWSKRWDQIYKRYSKHIEDAKRNRKAGTKNCGTSSTTNA AASTDENKCVQSDIDSFFK  
35 HLIDIGLTPSSYLSNVLDDNICGADKAPWTTYTTYTTTEKCNKERDKSKSQSSDTLVVVNVPSPLGNTPY  
RYKYACQCKIPTNEETCDDRKEYMNQWSCGSARTMKRGYKNDNYELCKYNGVDVKPTTVRSNSSKLD  
HRQPRGWEQLYPTGAEFLGDGGDISFSTRGTQNW TVERLLQAHRQLEERGYV FVG YHGT FLEAAQSIVF  
GGVRARSQDLDAIWRGFYIAGDPALAYGYAQDQEPDARGRIRNGALLRVYVPRSSLPGFYRTSLT LAAPE  
AAGEVERLIGHPLPLRLDAITGPEEEGGRLETILGWPLAERTVVI PSAIPTDPRNVGGDLDPSSIPDKEQAIS  
40 ALPDYASQPGKPPRKDEL

>DBL1-ID2aFCR3-PE38 (SEQ ID NO:67)

NHSDSGKYDPCEKKLPPYDDNDQWKCQQNSSDGSQKPENICVPPRRERLCTYNLENLKF DKIRDNN AFL  
ADVLLTARNEGEKIVQNHPTDNSSNVCNALERSFADLADIIRGTDQWKG TNSNLEKNLQMF AKIREND  
45 KVLQDKY PKDQKYTKLREAWWNANRQKVWEVITCGARSNDLLIKRGWRTSGKSDRKKNFELCRKCGHY  
EKEVPTKLDYVPQFLRWLTEWIEDFYREKQNLIDDMERHREECTREDHKSKEGTSYCSTCKDKCKKYCEC  
VKKWKTEWENQENKYKDLYEQNKNTSQKNTSRYDDYVKDFFEKLEANYSSLENYIKGDPYFAEYATKLS  
SFILNPSDANNPSGETANHND EACNCNESGISSVGQAQTS GPSSNKT CITHSSI KTNKKKECKDVKLGVR  
ENDKDLKICVIEDTSLSGVDNCCCQDLLGILQENCSDNKR GSSSNDSCDNKNQDECQKKLEKVFASLTN  
50 GYKCDKCKSGTSRSKKKWIWKSSGNEEGLQEEYANTIGLPPRTQSLYLGNLPKLENVCEDVKDINFDTK  
EKFLAGCLIVSFHEGKNLKKRYPQNKNSGNKENLCKALEYSFADYGD LKGT SIWDNEYTKDLELNLQNN  
FGKLF GKYIKKNNTAEQDTSYSSLD ELRESWWNTNKKYIWTAMKHGAEMNITTCNADGSVTGSGSSCD  
DIPTIDLIPQYLRFLQEWVENFCEQRQAKVKDVITNCKSCKESGNKCKTECKTKCKDECEKYKKFIEACGT  
AGGGIGTAGSPWSKRWDQIYKRYSKHIEDAKRNRKAGTKNCGTSSTTNA AASTDENKCVQSDIDSFFK  
55 HLIDIGLTPSSYLSNVLDDNICGADKAPWTTYTTYTTTEKCNKERDKSKSQSSDTLVVVNVPSPLGNTPY  
RYKYACQCKIPTNEETCDDRKEYMNQWSCGSARTMKRGYKNDNYELCKYNGVDVKPTTVRSNSSKLD  
EGGSLAALTAHQACHLPLETFTRHRQPRGWEQLEQCGYPVQRLVALYLAARLSWNQVDQVIRNALASPG  
SGGDLGEAIREQPEQARLALT LAAAESERFVRQGTGNDEAGAA nGPADSGDALLERNYPTGAEFLGDGG

DISFSTRGTQNWTVRLLQAHQRLEERGYVFGYHGTFLAAQSIIVFGGVRARSQDLDAIWRGFYIAGDP  
ALAYGYAQDQEPDARGRIRNGALLRVYVPRSSLPGFYRTSLTLAAPEAAGEVERLIGHPLPLRLDAITGPEE  
EGGRLETILGWPLAERTVVIPSAIPTDPRNVGGDLDPSSIPDKEQAISALPDYASQPGKPPRKDEL

5 >ID1-ID2a3D7-PE38 (SEQ ID NO:68)  
LSFILNSSDANNPSEKIQKNNDDEVCCNCNESGSIASVEQEIQISDPSSNKTCTHSSIKANKKKVCKHVKLGV  
RENDKDLRVCVIEHTSLSGVENCCCQDFLRILQENCSDNKS GSSSSNGSCNNKNQEACEKNLEKVLASLT  
NCYKCDKCKSEQSKNNKNWIWKKSSGKEGGLQKEYANTIGLPPRTQSLCLVVCLDEKGKKTQELKNIR  
10 TNSSELLKEWIIAAFHEGKNLKP SHEKKNDNGKKLCKALEYSFADYGD LKGT SIWDNEYTKDLELNLQKI  
FGKLF RKYIKKNNTAEQDTSYSSLD ELRESWWNTNKKYIWLAMKHGAGMNSTTCCGDG SVTSGSGSSCD  
DIPTIDLIPQYLRFLQEWVEHFCKQRQEKVKPVIENCKSCKESGGTCNGECKTECKNKCEVYKKFIEDCKG  
GDGTAGSSWVKRWDQIYKRYSKYIEDAKRNRKAGTKNCGPSSTTNA AENKCVQSDIDSFFKHLIDIGLT  
15 TPSSYLSIVLDDNICGADKAPWTTYTTYTTTEKCNKETDKSKLQQCNTAVVVNVPSPLGNTPHGYKYACQ  
CKIPTNEETCDDRKEYMNQWSCGSARTMKRGYKNDNYELCKYNGVDVKPTTVRSNSSLDRHRRQPRG  
LTAHQACHLPLETFRHRQPRGWEQLEQCGYPVQRLVALYLAARLSWNQVDQVIRNALASPGSGGDLGE  
AIREQPEQARLALTAAAESERFVRQGTGNDEAGAANGPADSGDALLERNYPTGAEFLDGGDISFSTRG  
TQNWTVRLLQAHQRLEERGYVFGYHGTFLAAQSIIVFGGVRARSQDLDAIWRGFYIAGDPALAYGYA  
QDQEPDARGRIRNGALLRVYVPRSSLPGFYRTSLTLAAPEAAGEVERLIGHPLPLRLDAITGPEEEGGRLET  
20 ILGWPLAERTVVIPSAIPTDPRNVGGDLDPSSIPDKEQAISALPDYASQPGKPPRKDEL

>ID1-ID2a3D7-PE38LR (SEQ ID NO:69)  
LSFILNSSDANNPSEKIQKNNDDEVCCNCNESGSIASVEQEIQISDPSSNKTCTHSSIKANKKKVCKHVKLGV  
RENDKDLRVCVIEHTSLSGVENCCCQDFLRILQENCSDNKS GSSSSNGSCNNKNQEACEKNLEKVLASLT  
NCYKCDKCKSEQSKNNKNWIWKKSSGKEGGLQKEYANTIGLPPRTQSLCLVVCLDEKGKKTQELKNIR  
25 TNSSELLKEWIIAAFHEGKNLKP SHEKKNDNGKKLCKALEYSFADYGD LKGT SIWDNEYTKDLELNLQKI  
FGKLF RKYIKKNNTAEQDTSYSSLD ELRESWWNTNKKYIWLAMKHGAGMNSTTCCGDG SVTSGSGSSCD  
DIPTIDLIPQYLRFLQEWVEHFCKQRQEKVKPVIENCKSCKESGGTCNGECKTECKNKCEVYKKFIEDCKG  
GDGTAGSSWVKRWDQIYKRYSKYIEDAKRNRKAGTKNCGPSSTTNA AENKCVQSDIDSFFKHLIDIGLT  
TPSSYLSIVLDDNICGADKAPWTTYTTYTTTEKCNKETDKSKLQQCNTAVVVNVPSPLGNTPHGYKYACQ  
30 CKIPTNEETCDDRKEYMNQWSCGSARTMKRGYKNDNYELCKYNGVDVKPTTVRSNSSLDRHRRQPRG  
WEQLYPTGAEFLDGGDISFSTRGTQNWTVRLLQAHQRLEERGYVFGYHGTFLAAQSIIVFGGVRAR  
SQDLDAIWRGFYIAGDPALAYGYAQDQEPDARGRIRNGALLRVYVPRSSLPGFYRTSLTLAAPEAAGEVER  
LIGHPLPLRLDAITGPEEEGGRLETILGWPLAERTVVIPSAIPTDPRNVGGDLDPSSIPDKEQAISALPDYAS  
QPGKPPRKDEL

35 >ID1-DBL2bFCR3-PE38LR (SEQ ID NO:70)  
NYIKGDPYFAEYATKLSFILNPSDANNPSGETANHND EACNCNESGISSVGQAQTSGPSSNKTCTHSSIK  
TNKKKECKDVKLGVRENDKDLKICVIEDTSLSGVDNCCCQDLLGILQENCSDNKR GSSSSNDSCDNKNQ  
DECQKKLEKVFASLTNGYKCDKCKSGTSRSKKKWIWKKSSGNEEGLQEYANTIGLPPRTQSLYLGNLPK  
40 LENVCEDVKDINFDTKEKFLAGCLIVSFHEGKNLKKRYPQNKNSGNKENLCKALEYSFADYGD LKGT SIW  
DNEYTKDLELNLQNNFGKLF GKYIKKNNTAEQDTSYSSLD ELRESWWNTNKKYIWTAMKHGAEMNITTC  
NADGSVTGSGSSCDDIPTIDLIPQYLRFLQEWVENFCEQRQAKVDVITNCKSCKESGNKCKTECKTKCK  
DECEKYKFFIEACGTAGGGIGTAGSPWSKRWDQIYKRYSKHIEDAKRNRKAGTKNCGTSSTTNA AASTD  
ENKCVQSDIDSFFKHLIDIGLTTPSSYLSNVLDDNICGADKAPWTTYTTYTTTEKCNKERDKSKSQSSDTL  
45 VVVNVPSPLGNTPYRYKRRHRQPRGWEQLYPTGAEFLDGGDISFSTRGTQNWTVRLLQAHQRLEERG  
YVFGYHGTFLAAQSIIVFGGVRARSQDLDAIWRGFYIAGDPALAYGYAQDQEPDARGRIRNGALLRVYV  
PRSSLPGFYRTSLTLAAPEAAGEVERLIGHPLPLRLDAITGPEEEGGRLETILGWPLAERTVVIPSAIPTDPR  
NVGGDLDPSSIPDKEQAISALPDYASQPGKPPRKDEL

50 >DT388, sequence of diphtheria toxin (SEQ ID NO:71)  
MGADDVVDSSKSFVMENFSSYHGT KPGYVDSIQKGIQKPKSGTQGN YDDDWKGFYSTDNKYDAAGYS  
VDNENPLSGKAGGVVKVTYPGLTKVLALKVDNAETIKKELGLSLTEPLMEQVGTEEFIKRFGDGASRVVLS  
LPFAEGSSSVEYINNWEQAKALSVELEINFETRGRGQDAMYEYMAQACAGNRVRRSVGSSLSCLNDW  
DVIRDKT KTKIESLKEHGPIKNKMS EPNKTVSEEKAKQYLEEFHQTALEHPELSELKTVTGTNPVFAGAN  
55 YAAWAVNVAQVIDSETADNLEKTTAALSILPGIGSVMGIADGAVHHNTEEIVAQSIALSSLMVAQAIPLVG  
ELVDIGFAAYNFVESIINLFQVVHNSYNRPAYSPGHKTQPMHEF

Sequences of VAR2CSA polypeptides fused with truncated fragments of diphtheria toxin

>DT388-DBL1-ID2a 3D7 (SEQ ID NO:72)

MGADDVVDSSKSFVMENFSSYHGTPGYVDSIQKGIQKPKSGTQGNYYYYDDWKGFYSTDNKNYDAAGYS  
 5 VDNENPLSGKAGGVVKVTPGLTKVLALKVDNAETIKKELGLSLTEPLMEQVGTEEFIKRFGDGASRVVLS  
 LPFAEGSSSVVEYINNWEQAKALSVELEINFETRGRGQDAMYEYMAQACAGNRVRRSVGSSSLSCINLDW  
 DVIRDKTKTKIESLKEHGPIKNKMSESPNKTVSEEEKAKQYLEEFHQTALEHPELSELKTVTGTNPVFAGAN  
 YAAWAVNVAQVIDSETADNLEKTTAALSILPGIGSVMGIADGAVHHNTEEIVAQSIALSSLMVAQAIPLVG  
 10 ELVDIGFAAYNFVESIINLFQVVHNSYNRPAYSPGHKTQPMHEFHSDSGTNDPCDRIPPPYGDNDQWKCA  
 IILSKVSEKPEENVFVPPRRQRMCIINLEKLNVDKIRDKHAFLADVLLTARNEGERIVQNHPTDNSSNVCNA  
 LERSFADIADIIRGTDLWKGTSNLEQNLKQMFAKIRENDKVLQDKYPKDQNYRKLREDWWNANRQKV  
 WEVITCGARSNDLLIKRGWRTSGKSNNGDNKLELCRKCGHYEEKVPTKLDYVPQFLRWLTEWIEDFYREK  
 QNLIDDMERHREECTSEDHKSKEGTSYCSTCKDKCKKYCECVKKWKSEWENQKNKYTELYQQNKNETS  
 15 QKNTSRYDDYVKDFFKKLEANYSSLENYIKGDPYFAEYATKLSFILNSSDANNPSEKIQKNNDVENCNCNES  
 GIASVEQEQISDPSSNKTCTHSSIKANKKKVCKHVKLGVRENDKDLRVCVIEHTSLSGVENCCCQDFLRI  
 LQENCSDNKGSSSSNGSCNNKNQEACEKNLEKVLASLTNCYKCDKCKSEQSKNNKNKIWKSSSGKE  
 GGLQKEYANTIGLPPRTQSLCLVCLDEKGKKTQELKNIRTNSELLKEWIIAAFHEGKNLKPSEKKNDDN  
 GKKLCKALEYSFADYGDLIKGTSIWDNEYTKDLELNLQKIFGKLFRKYIKNNNTAEQDTSYSSLDELRESW  
 WNTNKKYIWLAMKHGAGMNSTTCCGDSVTGSGSSCDDIPTIDLIPQYLRFLQEWVEHFCKQRQEKVK  
 20 PVIENCKSCKESGGTCNGECKTECKNKCEVYKKFIEDCKGGDGTAGSSWVKRWDQIYKRYSKYIEDAKR  
 NRKAGTKNCGPSSTTNAAENKCVQSDIDSFFKHLIDIGLTPSSYLSIVLDDNICGADKAPWTTYTTYTTT  
 ECKNKETDKSKLQQCNTAVVVNVPSPLGNTPHGYKYACQCKIPTNEETCDDRKEYMNQWSCGSARTMK  
 RGYKNDNYELCKYNGVDVKPTTVRSNSSKLDSSGR

>DT388-DBL1-ID2a FCR3 (SEQ ID NO:73)

MGADDVVDSSKSFVMENFSSYHGTPGYVDSIQKGIQKPKSGTQGNYYYYDDWKGFYSTDNKNYDAAGYS  
 VDNENPLSGKAGGVVKVTPGLTKVLALKVDNAETIKKELGLSLTEPLMEQVGTEEFIKRFGDGASRVVLS  
 LPFAEGSSSVVEYINNWEQAKALSVELEINFETRGRGQDAMYEYMAQACAGNRVRRSVGSSSLSCINLDW  
 30 DVIRDKTKTKIESLKEHGPIKNKMSESPNKTVSEEEKAKQYLEEFHQTALEHPELSELKTVTGTNPVFAGAN  
 YAAWAVNVAQVIDSETADNLEKTTAALSILPGIGSVMGIADGAVHHNTEEIVAQSIALSSLMVAQAIPLVG  
 ELVDIGFAAYNFVESIINLFQVVHNSYNRPAYSPGHKTQPMHEFHSDSGKYDPCEKKLPPYDDNDQWKC  
 QQNSSDGSKGKPEVICVPPRRERLCTYNLENLKFDKIRDNNAFLADVLLTARNEGEKIVQNHPTDNSSNVC  
 NALERSFADLADIIRGTDQWKGTSNLEKLNKQMFAKIRENDKVLQDKYPKDQYTKLREAWWNANRQ  
 KVWEVITCGARSNDLLIKRGWRTSGKSDRKNFELCRKCGHYEKEVPTKLDYVPQFLRWLTEWIEDFYRE  
 35 KQNLIDDMERHREECTREDHKSKEGTSYCSTCKDKCKKYCECVKKWKTEWENQENKYKDLYEQNKNT  
 SQKNTSRYDDYVKDFFEKLEANYSSLENYIKGDPYFAEYATKLSFILNPSDANNPSETANHNDEACNCN  
 ESGISSVGQAQTS GPSSNKTCTHSSIKTNKKKECKDVKLGVRENDKDLKICVIEDTSLSGVDNCCCQDL  
 LGILQENCSDNKRGSNDSCDNKNQDECQKKLEKVFASLTNGYKCDKCKSGTSRSKKKWIWKSSSG  
 NEEGLQEEYANTIGLPPRTQSLYLGNLPKLENVCEDVKDINFDTKEKFLAGCLIVSFHEGKNLKKRYPQNK  
 40 NSGNKENLCKALEYSFADYGDLIKGTSIWDNEYTKDLELNLQNNFGKLFKGYIKNNNTAEQDTSYSSLDE  
 LRESWWNTNKKYIWTAMKHGAEMNITTCNADGSVTGSGSSCDDIPTIDLIPQYLRFLQEWVENFCEQRQ  
 AKVKDVITNCKSKESGNKCKTECKTKCKDECEKYKKFIEACGTAGGGIGTAGSPWSKRWDQIYKRYSK  
 HIEDAKRNRKAGTKNCGTSSTTNAASTDENKCVQSDIDSFFKHLIDIGLTPSSYLSNVLDDNICGADK  
 APWTTYTTYTTTECKNKERDKSKSQSSDTLVVVNVPSPLGNTPYRYKYACQCKIPTNEETCDDRKEYMNQ  
 45 WSCGSARTMKRGYKNDNYELCKYNGVDVKPTTVRSNSSKLDSSGR

>DT388-ID1-ID2a 3D7 (SEQ ID NO:74)

MGADDVVDSSKSFVMENFSSYHGTPGYVDSIQKGIQKPKSGTQGNYYYYDDWKGFYSTDNKNYDAAGYS  
 VDNENPLSGKAGGVVKVTPGLTKVLALKVDNAETIKKELGLSLTEPLMEQVGTEEFIKRFGDGASRVVLS  
 50 LPFAEGSSSVVEYINNWEQAKALSVELEINFETRGRGQDAMYEYMAQACAGNRVRRSVGSSSLSCINLDW  
 DVIRDKTKTKIESLKEHGPIKNKMSESPNKTVSEEEKAKQYLEEFHQTALEHPELSELKTVTGTNPVFAGAN  
 YAAWAVNVAQVIDSETADNLEKTTAALSILPGIGSVMGIADGAVHHNTEEIVAQSIALSSLMVAQAIPLVG  
 ELVDIGFAAYNFVESIINLFQVVHNSYNRPAYSPGHKTQPMHEFLSFILNSSDANNPSEKIQKNNDVENCNC  
 55 NESGIASVEQEQISDPSSNKTCTHSSIKANKKKVCKHVKLGVRENDKDLRVCVIEHTSLSGVENCCCQD  
 FLRILQENCSDNKGSSSSNGSCNNKNQEACEKNLEKVLASLTNCYKCDKCKSEQSKNNKNKIWKSS  
 GKEGGLQKEYANTIGLPPRTQSLCLVCLDEKGKKTQELKNIRTNSELLKEWIIAAFHEGKNLKPSEKKN  
 DDNGKKLCKALEYSFADYGDLIKGTSIWDNEYTKDLELNLQKIFGKLFRKYIKNNNTAEQDTSYSSLDEL  
 RESWWNTNKKYIWLAMKHGAGMNSTTCCGDSVTGSGSSCDDIPTIDLIPQYLRFLQEWVEHFCKQRQE

KVKPVIENCKSCKESGGTCNGECKTECKNKCEVYKKFIEDCKGGDGTAGSSWVKRWDQIYKRYSKYIED  
AKRNRKAGTKNCGPSSTTNAAENKCVQSDIDSFFKHLIDIGLTPSSYLSIVLDDNICGADKAPWTTYTTY  
TTTEKCNKETDKSKLQQCNTAVVVNVPSPLGNTPHGYKYACQCKIPTNEETCDDRKEYMNQWSCGSART  
MKRGYKNDNYELCKYNGVDVKPTTVRSNSSKLDGR

5

>DT388-ID1-ID2a FCR3 (SEQ ID NO:75)

10

MGADDVVDSSKSFVMENFSSYHGTPGYVDSIQKGIQKPKSGTQGNYYYYWKGFYSTDNKYDAAGYS  
VDNENPLSGKAGGVVKVTPGLTKVLALKVDNAETIKKELGLSLTEPLMEQVGTEEFIKRFGDGASRVVLS  
LPFAEGSSSVEYINNWEQAKALSVELEINFETRGRGQDAMYEYMAQACAGNRVRRSVGSSLSCINLDW  
DVIRDKTKTKIESLKEHGIKNKMSSEPNKTVSEEKAKQYLEEFHQTALEHPELSEKTVGTNPVFAGAN  
YAAWAVNVAQVIDSETADNLEKTTAALSILPGIGSVMGIADGAVHHNTEEIVAQSIALSSLMVAQAIPLVG  
ELVDIGFAAYNFVESIINLFQVVHNSYNRPAYSPGHKTQPMHEFNYIKGDPYFAEYATKLSFILNPSDANNP  
SGETANHND EACNCNESGISSVGQAQTS GPSSNKT CITHSSIKNKKKECKDVKLG VRENDKDLKICVIE  
DTSLSGVDNCCCQDLLGILQENCSDNKRGSNDSCDNKNQDECQKKLEKVFASLTNGYKCDKCKSGT  
15 SRSKKKWIWKKSSGNEEGLQEEYANTIGLPRTQSLYLGNLPKLENVCEVDKINFDTKKFLAGCLIVSF  
HEGKNLKKRYPQNKNSGNKENLCKALEYSFADYGDLIKGT SIWDNEYTKDLELNLQNNFGKLF GKYIKKN  
NTAEQDTSYSSLDLRESWWNTNKKYIWTAMKHGAEMNITTCNADGSVTGSGSSCDDIPTIDLIPQYLR  
FLQEWVENFCEQRQAKVKD VITNCKSCKESGNKCKTECKTKCKDECEKYKKFIEACGTAGGGIGTAGSP  
WSKRWDQIYKRYSKHIEDAKRNRKAGTKNCGTSSTTNAAASTDENKCVQSDIDSFFKHLIDIGLTPSSY  
20 LSNVLDNICGADKAPWTTYTTYTTTEKCNKERDKSKSQSSDTLVVVNVPSPLGNTPYRYKYACQCKIPT  
NEETCDDRKEYMNQWSCGSARTMKRGYKNDNYELCKYNGVDVKPTTVRSNSSKLDGR

## EXAMPLES

### EXAMPLE 1 Production of truncated recombinant VAR2CSA proteins

All protein truncations were produced according to previously defined domain borders

25

(Dahlbäck M, Jorgensen LM, Nielsen MA, Clausen TM, Ditlev SB, et al. J Biol Chem 286:

15908-15917). For the purpose of simplification we have divided the CIDR<sub>PAM</sub> domain into  
two domains ID2a and ID2b, where ID2a is the N-terminal part of CIDR<sub>PAM</sub> not containing the  
CIDR-like sequence and ID2b corresponds to the CIDR-like sequence. We also used a new  
DBL2X border incorporating 93 amino acids of ID2a. For simplification we call this border  
30 DBL2Xb, while the old border will be referred to as DBL2Xa. Primers used in cloning are listed  
in Table 2. Fragments were expressed in baculovirus-infected insect cells as soluble proteins  
as described in Method 1. Most proteins were produced based on the FCR3 genotype. Some  
FCR3 fragments did not express and these were instead made based on the 3D7 genotype.

35

The proteins were used interchangeably in the analysis since we show that recombinant  
VAR2CSA from both genotypes bind equally to CSA. All proteins showed a shift in gel mobility  
when comparing reduced and non-reduced samples by SDS-PAGE (Method 2). This is  
consistent with the formation of intra-molecular disulfide bridges. Some proteins formed  
high-molecular weight complexes detected by non-reduced SDS-PAGE. This is probably due  
to the formation of inter-molecular disulfide bridges between unpaired cysteines. This was  
40 confirmed by reducing the complexes to monomeric protein using DTT.

Table 2: Cloning Primers

| FCR3 Primers                        |                                    |                                        |
|-------------------------------------|------------------------------------|----------------------------------------|
| Protein                             | Forward Primer                     | Reverse Primer                         |
| ID1-ID2b                            | AACTACATCAAGGGCGAC (SEQ ID NO:76)  | CTTGTTGATATTGGTGTCGGT (SEQ ID NO:77)   |
| DBL1X-ID2a                          | CACAGCGATAGCGGCAAG (SEQ ID NO:78)  | GTCCAGCTTGCTGGAGTT (SEQ ID NO:79)      |
| ID1-ID2a                            | AACTACATCAAGGGCGAC (SEQ ID NO:80)  | GTCCAGCTTGCTGGAGTT (SEQ ID NO:81)      |
| ID1-DBL2Xa                          | AACTACATCAAGGGCGAC (SEQ ID NO:82)  | AGCGGCGTTGGTGGTGA (SEQ ID NO:83)       |
| ID1-DBL2Xb                          | AACTACATCAAGGGCGAC (SEQ ID NO:84)  | GTACTTGTACCGGTAGGG (SEQ ID NO:85)      |
| DBL1X-DBL2Xb                        | CACAGCGATAGCGGCAAG (SEQ ID NO:86)  | GTACTTGTACCGGTAGGG (SEQ ID NO:87)      |
| 3d7 Primers                         |                                    |                                        |
| Protein                             | Forward Primer                     | Reverse Primer                         |
| DBL2X-DBL4ε                         | CTGACCAACTGCTACAAG (SEQ ID NO:88)  | GGTCCAGAGGGTACAGCTT (SEQ ID NO:89)     |
| ID1-DBL3ε                           | CTGTCCTTCATCCTGAAC (SEQ ID NO:90)  | TTCAGCGTTGTTGTA                        |
| ID1-DBL4ε                           | CTGTCCTTCATCCTGAAC (SEQ ID NO:92)  | GTCCAGAGGGTACAGCTT (SEQ ID NO:93)      |
| DBL1X-ID2b                          | CACTCTGACTCTGGCACC (SEQ ID NO:94)  | AGAGGACTTCATCTTGTGTTGGT (SEQ ID NO:95) |
| ID1-ID2b                            | CTGTCCTTCATCCTGAAC (SEQ ID NO:96)  | AGAGGACTTCATCTTGTGTTGGT (SEQ ID NO:97) |
| DBL1X-ID2a                          | CACTCTGACTCTGGCACC (SEQ ID NO:98)  | GTCCAGCTTAGAGGAGTT (SEQ ID NO:99)      |
| ID1-ID2a                            | CTGTCCTTCATCCTGAAC (SEQ ID NO:100) | GTCCAGCTTAGAGGAGTT (SEQ ID NO:101)     |
| DBL1X-DBL2Xa                        | CACTCTGACTCTGGCACC (SEQ ID NO:102) | GGCGGCGTTGGTGGTAGA (SEQ ID NO:103)     |
| ID1-DBL2Xa                          | CTGTCCTTCATCCTGAAC (SEQ ID NO:104) | GGCGGCGTTGGTGGTAGA (SEQ ID NO:105)     |
| DBL1X-DBL2Xb                        | CACTCTGACTCTGGCACC (SEQ ID NO:106) | GTACTTGTATCCGTGGGG (SEQ ID NO:107)     |
| ID1-DBL2Xb                          | CTGTCCTTCATCCTGAAC (SEQ ID NO:108) | GTACTTGTATCCGTGGGG (SEQ ID NO:109)     |
| Mutating Putative CSA Binding Sites |                                    |                                        |
| Protein                             | PCR1                               |                                        |
|                                     | Fragment 1                         |                                        |
|                                     | Forward                            | Reverse                                |

|                                                                                                           |                                                                                                                                                    |                                                                                                                                                                      |
|-----------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| DBL1X-ID2a (DSM Deletion)<br>Alanine sub.<br>K(626,629,630), R(631)<br>Alanine sub.<br>K(459,460,461,464) | CACAGCGATAGCGGCAAG (SEQ ID NO:110)<br>CACAGCGATAGCGGCAAG (SEQ ID NO:112)<br>CACAGCGATAGCGGCAAG (SEQ ID NO:114)                                     | GGTGTCTGAAGTTGATGTCGGGCAGATTGCCCAGGTA (SEQ ID NO:111)<br>AGCTGCGGCCAGATTAGCGCCCTCGTGAAGGACAC (SEQ ID NO:113)<br>AGCGCATTGAGCTGCGGCGTTGGTCTTGATGGAGCT (SEQ ID NO:115) |
| <b>Fragment 2</b>                                                                                         |                                                                                                                                                    |                                                                                                                                                                      |
| <b>Protein</b>                                                                                            | <b>Forward</b>                                                                                                                                     | <b>Reverse</b>                                                                                                                                                       |
| DBL1X-ID2a (DSM Deletion)<br>Alanine sub.<br>K(626,629,630), R(631)<br>Alanine sub.<br>K(459,460,461,464) | CACAGCGATAGCGGCAAG (SEQ ID NO:116)<br>GCTAATCTGGCCGCAGCTTACCCCCAGAATAAGAAC (SEQ ID NO:118)<br>GCCGCAGCTGAATGCGCTGACGTGAAGCTGGGCGTG (SEQ ID NO:120) | GTCCAGCTTGCTGGAGTT (SEQ ID NO:117)<br>GTCCAGCTTGCTGGAGTT (SEQ ID NO:119)<br>GTCCAGCTTGCTGGAGTT (SEQ ID NO:121)                                                       |
| <b>PCR2</b>                                                                                               |                                                                                                                                                    |                                                                                                                                                                      |
| <b>Final Construct</b>                                                                                    |                                                                                                                                                    |                                                                                                                                                                      |
| <b>Protein</b>                                                                                            | <b>Forward</b>                                                                                                                                     | <b>Reverse</b>                                                                                                                                                       |
| DBL1X-ID2a (DSM Deletion)<br>Alanine sub.<br>K(626,629,630), R(631)<br>Alanine sub.<br>K(459,460,461,464) | CACAGCGATAGCGGCAAG (SEQ ID NO:122)<br>CACAGCGATAGCGGCAAG (SEQ ID NO:124)<br>CACAGCGATAGCGGCAAG (SEQ ID NO:126)                                     | GTCCAGCTTGCTGGAGTT (SEQ ID NO:123)<br>GTCCAGCTTGCTGGAGTT (SEQ ID NO:125)<br>GTCCAGCTTGCTGGAGTT (SEQ ID NO:127)                                                       |

### EXAMPLE 2 VAR2CSA from FCR3 and 3D7 Binds CSA with similar Affinity and Specificity

FCR3 infected erythrocytes (IE) adhere much stronger to CSA *in vitro* compared to 3D7 or NF54 (IE). If the differences are related to the sequence differences of the expressed VAR2CSA the information could be used to define residues involved in the adhesion process. To test this, we produced a series of overlapping 3D7 VAR2CSA fragments, identical to the ones we have previously tested for FCR3.

The proteins were first screened for specific CSPG binding in a solid phase binding assay (ELISA) (described in Method 3). Proteins binding specifically to CSPG were then further purified by size exclusion chromatography, to obtain pure monomeric protein, and subjected to kinetic analysis on a Quartz Crystal Microbalance (Attana A100) (Method 2 and 4, respectively). The 3D7 VAR2CSA fragments showed binding characteristics very similar to their FCR3 counterparts in the solid state binding assay. The same is true in the kinetic analysis (Table 3). The sensorgrams show association and dissociation data collected at different protein concentrations. This allows determination of the association rate constant ( $k_{on}$ ), disassociation rate constant ( $k_{off}$ ), and the equilibrium constant ( $K_D$ ). Together with the peak response levels these parameters give estimation for the CSPG binding affinity. There is no apparent difference between 3D7 and FCR3 fragments. Some fragments show lower affinity, but this characteristic is maintained in the fragments counterpart. This indicates that the 3D7 and FCR3 VAR2CSA proteins fold and function in the same way.

**Table 3: CSA binding affinity of produced VAR2CSA proteins.** Affinity is given as a  $K_D$  (nM) value determined in kinetics experiments using a quartz crystal microbalance biosensor (Attana A100). N/A: proteins for which no  $K_D$  value could be determined, due to a lack of binding to CSA.

| VAR2CSA Fragment | FCR3   |         | 3D7    |
|------------------|--------|---------|--------|
|                  | Baculo | E. coli | Baculo |
| FV2              | 5.2*   |         | 8.2    |
| ID1-DBL4ε        | 8.6*   |         | 9.4    |
| ID1-DBL3ε        | 0.3*   |         | 8.5    |
| DBL2X-DBL4ε      | 2.4*   |         | 1.2    |
| DBL1-ID2b        | 1.5*   |         |        |
| DBL1-ID2a        | 8.0    | 3.5     | 29.5   |
| ID1-ID2a         | 7.6    | 18.3    | 5.7    |
| DBL1X-DBL2Xb     |        | 14.6    |        |
| DBL1X-DBL2Xa     | N/A    |         |        |
| ID1-DBL2Xb       |        |         | 21.8   |
| ID1-DBL2Xa       |        |         | N/A    |

\* Proteins published in (Dahlbäck et al, JBC, 2011)

**EXAMPLE 3 - The core-CSA binding site lies within the DBL2X domain**

It has been suggested that the minimal CSA binding region in VAR2CSA lies within DBL2X-ID2b, with the need of flanking domains for full affinity binding (Dahlbäck M, Jorgensen LM, Nielsen MA, Clausen TM, Ditlev SB, et al. J Biol Chem 286: 15908-15917). Here we have analyzed shorter fragments of VAR2CSA to further map the regions required for CSA binding.

The truncated proteins were first screened for binding to a CSA proteoglycan (CSPG) in ELISA and then further purified to obtain monomers for examination on the Quartz Crystal Microbalance (Methods 3, 2 and 4, respectively). The minimal binding region is ID1-DBL2Xb (Table 3). This region showed a binding affinity of 21.8 nM, which is comparable to that of full-length VAR2CSA.

Placental IEs are highly selective for low-sulfated placental CSPG. They do not adhere to any other glycosaminoglycans (GAG), such as heparan sulfate (HS). The same is true for the full-length recombinant VAR2CSA protein. The solid state binding assay showed that the VAR2CSA fragments, containing the minimal CSA binding region, bound specifically to CSA. To confirm this the minimal binding fragments were further tested for binding to a heparan sulfate proteoglycan (HSPG) on the Quartz Crystal Microbalance (Method 4). None of the fragments bound HSPG.

**EXAMPLE 4 - Antibodies induced against novel minimum binding regions induce a potent parasite anti-adhesive immune response**

A VAR2CSA based vaccine against PM must be able to induce a strong protective immune response. In this, the most important aspect is the formation of anti-VAR2CSA IgG antibodies capable of inhibiting placental sequestration. We have examined the molecular mechanism underlying the VAR2CSA-CSA interaction for the purpose of designing optimal vaccine antigens. To test whether our produced VAR2CSA recombinant fragments showed the capacity to induce an adhesion blocking immune response, they were used in rat immunizations (Method 6).

VAR2CSA fragment-specific serum was tested for ability to inhibit IE adhesion to CSPG (Method 11). Antibodies raised against all CSA-binding fragments were very potent inhibitors of binding. In fact binding was inhibited nearly 100% in all cases. DBL1X-DBL2Xa and ID1-

DBL2Xa were not good inhibitors, consistent with the lack of CSA binding of these fragments (Table 3). The data implies that the CSA-binding proteins are properly folded and support the localization of the above-defined minimal binding region.

**EXAMPLE 5 - Epitopes responsible for the induction of anti-adhesive antibodies lies within the minimal binding region**

To examine if the inhibitory anti-FV2 response is directed towards the minimal binding region, we affinity purified FV2 antibodies on four of the previous described VAR2CSA fragments (Method 7). The fragment-specific antibodies were then tested for the capacity to inhibit VAR2CSA expressing parasite binding to CSPG (Method 11). Antibodies purified on immobilized ID1-DBL4 $\epsilon$ , DBL1X-ID2a and ID1-ID2a, fully inhibited parasite adhesion. Furthermore, the depleted FV2 samples lost a significant portion of their inhibitory capacity. This indicates that epitopes inducing anti-adhesive antibodies are present within these fragments. Antibodies purified on DBL1X-DBL2Xa show a reduced inhibitory capacity, consistent with the lack of CSA binding of this fragment (Table 3). The data suggest that epitopes responsible for induction of inhibitory antibodies are located within the minimal binding region (here illustrated by ID1-ID2a).

**EXAMPLE 6 - Mutating putative GAG binding sites in the minimal binding region has no effect on CSPG binding**

Characterizing the nature of the interaction between VAR2CSA and CSA is important for the design of multivalent PM vaccines. In this, a major part is identification of the specific CSA-binding site and characterization of the underlying chemical interactions. Sequence analysis of the minimal CSA-binding region revealed two conserved putative GAG binding sites. One is located in the ID1 region and has the classic Cardin-Weintraub XBBBXXBX motif (Cardin, A. D., and Weintraub, H. J. (1989) *Arteriosclerosis* **9**, 21-32) (458-NKKKECKD-465). Another, in DBL2X, has the same motif in reverse (625-GKNLKKRY-632). It has also been hypothesized that a dimorphic sequence motif (DSM), found in the N-terminal part of DBL2X, is involved in binding CSA (Sander, A. F., Salanti, A., Lavstsen, T., Nielsen, M. A., Magistrado, P., Lusingu, J., Ndam, N. T., and Arnot, D. E. (2009) *PLoS One* **4**, e6667). To test whether these putative sites had a function in CSA binding, we substituted basic amino acids in the classic GAG binding sites with alanines and made a ten amino acid (590-KLENVCEDVK-603) deletion in the middle of a surface exposed loop within the DSM region. All mutations were performed in the DBL1X-ID2a fragment.

Substituting basic amino acids in the putative ID1 and DBL2X GAG binding sites, with alanines had no effect on CSPG binding. No decrease in CSPG binding was seen compared to

the wild-type protein in ELISA (Method 3). The construct with four alanine substitutions, Alanine Sub. K(459,460,461,464), shows considerable HSPG binding, which could be caused by a change in protein structure in response to mutation. The two mutants, Alanine Sub. K(626, 629, 630), R(631) and Alanine Sub. K(459,460,461,464), show CSPG binding kinetics similar to the positive control (Method 4). This is evident by similar  $K_D$  values and peak responses.

The deletion of the DSM region did not reduce binding to CSPG (Methods 3 and 4). The DSM knock out mutant shows considerable binding to HSPG in ELISA. This is likely caused by an erroneous cloning where 100 amino acids of DBL1X were lost. Importantly CSPG binding was not affected.

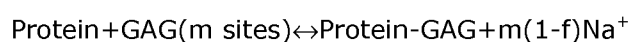
## EXAMPLE 7

### VAR2CSA binding to CSPG does not depend on ionic interactions

Mutation of the classic Cardin-Weintraub GAG binding motifs had no effect on CSPG binding. This indicates that the VAR2CSA-CSA binding mechanism differs from the general mode of sulfate binding in classic GAG binding models. There are examples of GAG binding proteins showing little dependence on ionic interactions with the sulfated GAG structure. To test if this was the case, we examined ionic dependence according to the polyelectrolyte theory (Record, M. T., Jr., Lohman, M. L., and De Haseth, P. (1976) *J Mol Biol* **107**, 145-158).

Glycosaminoglycans, like DNA, are highly charged polymers often referred to as polyelectrolytes. The negatively charged groups incur a high degree of repulsive energy within each polymer. Monovalent cations, such as  $\text{Na}^+$ , interact with the negatively charged groups to minimize the repulsive energy. Binding of basic amino acids to the sulfate groups displaces the bound cations and leads to the release of free energy. The favorable release of bound  $\text{Na}^+$  ions is referred to as the polyelectrolyte effect.

The theory states that the binding of a protein to a GAG can be described by:



Where  $m$  is the number of  $\text{Na}^+$  ions released upon binding of a single protein and  $f$  is the fraction of anions not shielded by  $\text{Na}^+$  ions. According to the theory the observed  $K_D$  value is related to ionic and non-ionic contributions by:

$$\text{Log}K_{D,\text{observed}} = \text{Log}K_{D,\text{nonionic}} + m(1-f)\text{Log}[\text{Na}^+]$$

Where  $K_{D,\text{nonionic}}$  is the disassociation constant in the absence of ionic interactions. A plot of  $\text{Log}K_{D,\text{observed}}$  vs  $\text{Log} [\text{Na}^+]$  is linear with a slope of  $m(1-f)$ . Thus, if the fraction of unshielded anions ( $f$ ) is known, the number of ionic interactions involved in the binding can be determined. For heparin ( $1-f$ ) is 0.8 (Olson, S. T., Halvorson, H. R., and Bjork, I. (1991) *J Biol Chem* **266**, 6342-6352). The value is not known for CSA, but ( $1-f$ ) cannot exceed 1. We can therefore estimate the maximal number of ionic interactions involved. Furthermore, when  $[\text{Na}^+] = 1 \text{ M}$ ,  $\text{Log} [\text{Na}^+] = 0$ , which means that at this  $\text{Na}^+$  concentration  $\text{Log}K_{D,\text{observed}} = \text{Log}K_{D,\text{nonionic}}$ .

We tested the binding of FV2, DBL1X-ID2a and ID1-ID2a to CSPG in a solid state binding assay at different concentrations of NaCl (150 mM, 200 mM, 250 mM, 300 mM), by performing titrations of binding from 400 nM-1.65 nM protein in a 1:2 dilution series (Method 5). The observed  $K_D$  values were determined as the protein concentration giving half-maximum ( $B_{\text{max}}$ ) response. This was done using non-linear regression (least squares fit with Hill slope) in Graphpad Prism. Higher salt concentrations were not included in the analysis as binding was almost completely inhibited. This is probably due to a change in protein structure. This notion is supported by the fact that  $\text{Log}K_{D,\text{observed}}$  vs  $\text{Log} [\text{Na}^+]$  was linear only between 150 mM and 300 mM, suggesting that other factors play a role at higher concentrations of NaCl.

$\text{Log}K_{D,\text{observed}}$  vs  $\text{Log} [\text{Na}^+]$  shows a linear relationship. The slope  $m(1-f)$  ranges between 2.7 for ID1-ID2a and 3.4 for full-length (FV2). We do not know the value for  $f$ , but the maximal number of ionic interactions involved in the binding must be between 2 and 3. It is interesting that the value for the full-length protein is higher than for the short fragments, indicating that this protein makes an extra ionic interaction with CSPG. The  $K_D$  values at 150 mM NaCl serves as our reference point, as this is the physiological NaCl concentration. By extrapolating the linear relationship and finding the y-intercept we find that  $K_{D,\text{nonionic}} = 5.9 \mu\text{M}$  for FV2,  $K_{D,\text{nonionic}} = 3.4 \mu\text{M}$  for DBL1X-ID2a, and  $K_{D,\text{nonionic}} = 0.7 \mu\text{M}$  for ID1-ID2a. Comparing the logarithmic values of these and the reference point (150 mM NaCl), we estimate that between 25-35% of the VAR2CSA binding can be accounted for by ionic interactions. This suggests that the high CSA affinity for VAR2CSA cannot be explained by ionic interactions with the sulfated GAG structure alone. The high affinity may be achieved through a complex binding site making a multivalent interaction with the CSA carbohydrate backbone.

## EXAMPLE 8

**VAR2CSA minimal CSA binding region binds specifically to a wide panel of cancer cells**

Many different cancer cells have been associated with high expression of the proteoglycan CSPG4. This molecule was initially described as a marker for melanoma but it has recently been found in many cancer forms, including cancer stem cells. The CS chain(s) attached to CSPG4 is known to be primarily CSA. One of the smallest VAR2CSA fragments (ID1-ID2a) was analyzed for binding to a large panel of various cancer cell lines by flow-cytometry (Method 12a and 12b). The non-CSA binding protein ID1-DBL2Xa was used as a negative control. The VAR2CSA recombinant protein (ID1-ID2a) binds strongly at 75 nM to all cancer cell lines transcribing CSPG4 (microarray data) including cutaneous Melanoma (C32, MeWo), Lung carcinoma (A549), Breast carcinoma (HCC1395), Osteosarcoma (U2OS, MNNG/HOS), Rhabdomyosarcoma (RH30) (Table 4 and 5). This protein also binds strongly to cutaneous T-cell lymphoma, which does not express CSPG4 (Table 4). The negative control protein ID1-DBL2Xa did not bind to any of the cell lines tested (Table 4). In addition, ID1-ID2a did not interact with human red blood cells, which were used as control cells. Wild type and GAG-deficient Chinese hamster ovary (CHO) cells were also analyzed for ID1-ID2a interaction. The strong interaction seen for ID1-ID2a with wild-type CHO cells was completely abolished when analyzing the CHO-745 cell line, in which the GAG-synthesis is disrupted. The CSA specificity of the interaction was also verified by inhibiting VAR2CSA binding to cells by pre-mixing VAR2CSA with CSA, CSC or HS. CSC and HS did not have any effect on the binding, whereas CSA efficiently abrogated binding of VAR2CSA to the cancer cells. Following these results, a larger panel of cancer cells were screened by flow cytometry (Table 6 and 7) using the DBL1-ID2a or ID1-ID2a fragment of VAR2CSA. The main purpose of this screening is to identify cell lines suitable for xenograft modeling *in vivo*.

**Table 4. Staining of cancer cell lines and negative control cells using the minimal binding domain of VAR2CSA (ID1-ID2a).**

Cells were incubated with medium alone (blank) or recombinant proteins (ID1-DBL2 or ID1-ID2a) at 75 nM for 30 minutes, followed by incubation with anti-V5-FITC (Invitrogen) at 1:800, cells were washed thrice between each incubation. Shown are the mean FITC fluorescence values recorded from a minimum of 5000 cells using a FC500 flowcytometer (Becton Dickinson).

| Cell type | Blank | ID1-DBL2Xa | ID1-ID2a |
|-----------|-------|------------|----------|
| C32       | 5,77  | 6,94       | 63,81    |
| MyLa 2059 | 5,61  | 5,61       | 145,35   |
| MyLa 1850 | 5,87  | 5,6        | 137,86   |

|              |      |      |       |
|--------------|------|------|-------|
| Cho WT       | 3,09 | 4,35 | 34,79 |
| Cho 745      | 4,24 | 4,29 | 4,38  |
| PBMC         | 1,34 | 1,36 | 1,67  |
| Erythrocytes | 1,11 | 1,17 | 1,07  |

**Table 5. Staining of cancer cell lines using recombinant VAR2CSA**

Cells were incubated with medium alone (blank) or recombinant proteins (DBL1-ID2a or ID1-ID2a) at 75 nM for 30 minutes, followed by incubation with anti-V5-FITC (Invitrogen) at 1:800, cells were washed thrice between each incubation. Shown are the medium score of FITC fluorescence intensity recorded from a minimum of 4 high power field images using a HAL100 Zeiss microscope. NS: No staining; +: weak; ++: medium; +++: strong; ++++: Very strong.

| Cell type  | Blank | DBL1-ID2a |
|------------|-------|-----------|
| U2OS       | NS    | +++       |
| MG63       | NS    | ++++      |
| MDA-MB-231 | NS    | +++       |
| TC32       | NS    | +         |
| TC71       | NS    | ++        |
| MNNG       | NS    | +++       |
| CHLA9      | NS    | ++        |
| CHLA10     | NS    | ++        |
| RH30       | NS    | +++       |
| RH18       | NS    | ++        |
| PC3        | NS    | +++       |

**Table 6 Screening of diverse human cancer cell lines for binding of recombinant VAR2CSA (using DBL1-ID2a or ID1-ID2a).**

Binding was measured by flow cytometry as described in METHOD 12.

NS: No staining; +: weak; ++: medium; +++: strong; ++++: Very strong.

| Cell line | Control | 75 nM<br>VAR2CSA | 150 nM<br>VAR2CSA | Comments                                                 |
|-----------|---------|------------------|-------------------|----------------------------------------------------------|
| MeWo      | NS      | +++              | ++++              | Melanoma (Fibroblast morphology, derived from lymphnode) |
| A549      | NS      | +++              | +++               | Lung Adenocarcinoma (K-RasG12S)                          |

|            |    |     |      |                                                                                                                            |
|------------|----|-----|------|----------------------------------------------------------------------------------------------------------------------------|
| HCC1395    | NS | +++ | ++++ | Invasive ductal breast carcinoma TNM stage 1 grade 3; no lymphnode metastasis; Her2-neg, ER-neg, PR-neg (Triple-negative)  |
| RH30       | NS | +++ | ++++ | Rhabdomyosarcoma (TPp53 negativ; PAX7-FOXO1A fusion positive; highly genomic instable (>50 chromosome rearrangements))     |
| MNNG       | NS | +++ | +++  | Osteosarcoma from 13 year old female caucasian (TPR-Met positive)                                                          |
| U2OS       | NS | +++ | +++  | Osterosarcoma from 15 year old female caucasian (IGF-R1 and IGFR-II positive; TPp53 wt, pRb wt, p16-neg; highly aneuploid) |
| H1792      | NS | ++  | ++   | Lung Adenocarcinoma (K-RasG12S: TPp53het))                                                                                 |
| MDA-MD-435 | NS | ++  | +++  | Breast carcinoma of melanocytic origin (ER-neg, Her2-pos, PR-pos)                                                          |
| MG63       | NS | +++ | ++++ | Osteosarcoma                                                                                                               |
| TC32       | NS | ++  | ++   | Ewing's sarcoma                                                                                                            |
| CHLA9      | NS | ++  | ++   | Ewing's sarcoma                                                                                                            |
| CHLA10     | NS | ++  | ++   | Ewing's sarcoma                                                                                                            |
| TC71       | NS | ++  | ++   | Ewing's sarcoma                                                                                                            |
| HOS        | NS | +++ | ++++ | Osteosarcoma                                                                                                               |
| PC3        | NS | ++  | ++   | Prostate carcinoma                                                                                                         |
| SKNMC      | NS | ++  | +++  | Ewing's sarcoma                                                                                                            |
| MCF-7      | NS | +   | ++   | Breast carcinoma                                                                                                           |

**Table 7 Screening of more human cell cancer cell lines for binding of recombinant VAR2CSA (using DBL1-ID2a or ID1-ID2a)**

- 5 Binding was measured by flow cytometry as described in METHOD 12.  
Values shown are mean fluorescence intensity using protein concentration of 200 nM.

| Cell type | Negative control | ID1-ID2a | Comments          |
|-----------|------------------|----------|-------------------|
| GP202     | 21.63            | 111.37   | Gastric Carcinoma |
| NCI-N87   | 7.18             | 207.72   | Gastric Carcinoma |

| MKN45     | 4.22  | 55.4      | Gastric Carcinoma                                |
|-----------|-------|-----------|--------------------------------------------------|
| MKN28     | 6.9   | 103.84    | Gastric Carcinoma                                |
| AGS       | 7.25  | 18.21     | Gastric Carcinoma                                |
| KatoIII   | 7.33  | 18.76     | Gastric Carcinoma                                |
| SNU-1     | 4.33  | 155.79    | Gastric Carcinoma                                |
| SNU-638   | 8.47  | 8.49      | Gastric Carcinoma                                |
| IPA220    | 7.72  | 13.67     | Gastric Carcinoma                                |
| MDA-231   | 3.39  | 63.43     | Triple negative Breast                           |
| T47D      | 3.63  | 48.13     | Luminal Breast                                   |
| LNCap     | 6.58  | 24.86     | Prostate                                         |
| PC3       | 5.2   | 29.82     | Prostate                                         |
| Ovc316    | 1.89  | 7.24      | Ovarian cancer stem cells                        |
| Cell type | Blank | DBL1-ID2a |                                                  |
| NALM-6    | 6,19  | 8,22      | Acute lymphatic leukaemia (ALL)                  |
| 697       | 3,23  | 30,36     | ALL                                              |
| AMO-1     | 2,68  | 35,22     | Myelomatosis                                     |
| KMM-1     | 2,82  | 16,1      | Myelomatosis                                     |
| MOLP-8    | 2,44  | 19,24     | Myelomatosis                                     |
| KMS-12-PE | 3,02  | 7,14      | Myelomatosis                                     |
| KMS-12-BM | 2,2   | 3,25      | Myelomatosis                                     |
| U2932     | 4,24  | 16,83     | Diffuse Large B-cell lymphoma (DLBCL)            |
| SU-DHL8   | ND    | 3,75      | DLBCL                                            |
| SU-DHL5   | 2,19  | 10,28     | DLBCL                                            |
| Oci_Ly19  | 3,38  | 18,96     | DLBCL                                            |
| HBL1      | 6,53  | 39,53     | DLBCL                                            |
| Farage    | 2,8   | 3,28      | DLBCL                                            |
| RIVA      | 2,26  | 3,32      | DLBCL                                            |
| WSU-FSCCL | 4,89  | 22,32     | Low-grade follicular small cleaved cell lymphoma |
| U-698-M   | 2,24  | 2,85      | Lymphoblastic lymphoma del(6)(q15q22)            |

**EXAMPLE 9*****Recombinant VAR2CSA binds to cancer cells with high affinity***

The binding affinity of the recombinant VAR2CSA fragment DBL1-ID2a to the cancer cell

- 5 lines, C32 melanoma and two Cho cell lines (described in example 8) was analysed using a Quartz Crystal Microbalance biosensor (Attana Cell200). A 2-fold dilution series (25-400 nM)

of the protein was analysed for binding to the cell surface, with regeneration of the binding surface in between each new protein injection. The binding affinity was estimated to lie in the nano-molar range (Table 8), which is similar to the binding affinity to pure receptor (Table 3).

**5 Table 8. Estimated binding affinity ( $K_D$ ) of recombinant DBL1-ID2a (E. coli) to cancer cells expressing CSA (C32 and Cho WT) and lack of binding to a CSA-negative cell line (Cho 745)**

N/A:  $K_D$  could not be determined due to lack of binding to the cells

| Cell type          | $K_D$ (nM) |
|--------------------|------------|
| C32 melanoma cells | 13         |
| Cho WT             | 1.4        |
| Cho 745            | N/A        |

10

**EXAMPLE 10**

***Recombinant VAR2CSA protein binds to cancer tissue with high specificity***

- 15 The binding of recombinant VAR2CSA to primary cancer tissue obtained from human patients is investigated using immunohistochemistry (IHC). The method was developed using human placenta tissue as positive control and Tonsil and liver tissue as negative control. The staining protocol was optimized on the Ventana Discovery XT platform with no epitope retrieval. Paraffin embedded tissue spotted on glass slides was incubated with 0.1-500 nM V5-
- 20 VAR2CSA (ID1-ID2a) or V5-Control protein (DBL4) for 1 h in room temperature, washed for 8 minutes, incubated with 1:700 mouse anti-V5 antibody for 30 minutes, washed for 8 minutes. Bound anti-V5 was subsequently detected using UltraMap anti-mouse HRP. V5-VAR2CSA stains human placenta in 0.5 nM concentrations with no staining in Tonsil or normal liver. The staining can be completely blocked by adding 200  $\mu\text{g}/\mu\text{l}$  CSA to the reaction buffer.
- 25 V5-control protein does not stain human placenta tissue at any concentrations tested. A multi-organ tissue micro-array (TMA) representing 24 normal organs showed low or absent staining when stained with 1 nM V5-VAR2CSA, while cancer specimens of breast, colon, Rectum, Prostate, kidney, liver, bladder, pancreas, squamous cell, Lung, Gall bladder, Stomach, Testis, Ovary, Uterus, Adrenal gland, Thyroid and Thymus, hematopoietic system,
- 30 and the connective tissue (sarcomas) stained positive with intensities equal or higher than human placenta positive control tissue (Table 9).

**Table 9. Detection of CSA on primary human tumor specimens using recombinant VAR2CSA. Table shows number of positive/ total number of cases stained as described in Example 10 for main cancer groups. Positive staining is defined as intensity equal or higher than that observed in placenta tissue.**

| Cancer group                      | Positive ratio |
|-----------------------------------|----------------|
| Bladder carcinoma                 | 44/56          |
| Prostate carcinoma                | 71/76          |
| Breast carcinoma                  | 64/75          |
| Melanoma                          | 5/6            |
| Sarcoma                           | 23/25          |
| Esophagus Squamous cell carcinoma | 2/3            |
| Stomach Adenocarcinoma            | 3/3            |
| Colon carcinoma                   | 2/3            |
| Rectal Adenocarcinoma             | 3/3            |
| Liver carcinoma                   | 3/3            |
| Renal carcinoma                   | 3/3            |
| Lung carcinoma                    | 2/3            |
| Cervix carcinoma                  | 3/3            |
| Ovarian carcinoma                 | 2/3            |
| Diffuse B-cell lymphoma           | 1/3            |
| Astrocytoma                       | 3/3            |
| Pancreatic carcinoma              | 3/3            |

#### EXAMPLE 11

##### 10 ***Inhibition of transformation-parameters in vitro by recombinant VAR2CSA proteins***

The inhibitory effect of un-coupled VAR2CSA on tumor cell morphology *in vitro* is investigated by three different assays:

- i) The **soft agar colony formation assay** addresses whether VAR2CSA can inhibit the ability of cancer cells to proliferate in a three dimensional matrix.
- 15 ii) The **migration assay** addresses whether VAR2CSA can inhibit the ability of cancer cells to migrate vertically towards a chemo-attractant in a boyden chamber.
- iii) The **invasion assay** addresses whether VAR2CSA can inhibit the ability of cancer cells to invade through an artificial basement membrane.

**Soft agar colony formation assay:** Cells are treated with 25-100 nM VAR2CSA for 24 hours before seeded in soft agar matrix, and left for 10-12 days at 37°C. Images are captured by phase contrast microscope and quantified by ImageJ software. Recombinant VAR2CSA inhibits soft agar colony formation of MG63 osteosarcoma and RH30

5 Rhabdomyosarcoma cells in concentrations between 75 and 150 nM.

**Basement Membrane Extract (BME)-coated cell invasion assay:** To model the invasive process, we utilize the CultreCoat® 24 Well BME-Coated Cell Invasion platform (Cedarlane) according to the manufacturer's protocols, with the following modifications. Cells are serum starved one day before assays in the presence or absence of 25-100 nM VAR2CSA. On the

10 second day, cells maintained under the above conditions are plated in the top chambers ( $1 \times 10^5$  cells/well) of plates, while lower chambers contained either serum depleted media as a negative control, or media supplemented with 10% FBS. Cells are then incubated for another 18 hours. Cells invading through the BME are collected using dissociation buffer

15 containing Calcein AM, which converts into a highly fluorescent compound in living cells. Emitted fluorescence are measured using a fluorescent plate reader, analyzed by the FLUOStar software, fitted on a standard curve, and converted into corresponding number of cells.

20 **Migration assay.** The Migration assay is essentially the same procedure as the Basement Membrane Extract (BME)-coated cell invasion assay, but without BME.

Migration and Invasion capacity of MG63 osteosarcoma, RH30 Rhabdomyosarcoma, and MDA-MB-231 triple-negative breast cancer are inhibited by 75-150 nM recombinant

25 VAR2CSA.

## EXAMPLE 12

### ***Analyzing intracellular signalling events controlling cancer cell transformation-parameters regulated by CSA-containing proteoglycans.***

30 CSPG4 facilitates proliferation, migration and invasion via a Ras, Rac1 and PI3 kinase-dependent mechanism. Based on the results obtained in EXAMPLE 10, we will investigate intracellular signalling events leading to potential VAR2CSA-mediated inhibition of proliferation, migration and invasion. This is done with state-of-the-art biochemical and molecular biology methods including, but not limited to, Rac1 activation assays,

35 immunoblotting of pathway components and in-cell measurements of reactive oxygen species (ROS) generation. This line of experiments will clarify the signalling pathways affected by VAR2CSA binding to CSA-containing proteoglycans.

**Rac1 activity assay:** Rac1 activity assays are performed on appropriate human cancer cell lines left untreated or treated with recombinant VAR2CSA, according to the manufacturer's protocols (Thermo Scientific).

**Reactive oxygen species (ROS) assays:** Crude ROS levels are measured by CM-

- 5 H2DCFDA (Invitrogen) according to the manufacturer's guidelines. Superoxide levels will be measured using dihydroethidium (DHE). In the presence of the superoxide anion  $O_2^-$ , dihydroethidium is rapidly oxidized to oxyethidium, which binds DNA and emits light in the 570-580 nm ranges when excited at 488 nm. For cell culture, after appropriate treatments, cells are washed in Hank's Balanced Salt Solution (HBSS), incubated for 30-60 minutes in  
10 HBSS containing 10  $\mu$ M DHE, washed in HBSS and directly analyzed for oxyethidium fluorescence with an epi-fluorescence HAL100 microscope (Zeiss). For tumor sections, snap-frozen tumors are cut in 20  $\mu$ m sections using a cryostat, washed and DHE-treated as described for cell lines, mounted on cover slides and analyzed as for cell lines. Oxyethidium emission are analyzed and quantified using ImageJ software. For all tumor specimens,  
15 hematoxylin and eosin (H&E) staining are performed side-by-side to verify tissue integrity and pathology, using standard methods. Preliminary data indicates that recombinant VAR2CSA inhibits ROS-generation in MG63 and U2OS cells.

- Immunodetection.** For immunoblotting, proteins separated by SDS-PAGE and transferred to a nitrocellulose membrane are detected with relevant primary and appropriate secondary  
20 antibodies, ECL Western blotting reagents (Thermo Scientific), and film (Kodak and Covance (HA). For microscopy, cells are fixed in 4% formaldehyde, incubated with appropriate primary antibodies, incubated with appropriate secondary FITC-conjugated antibodies and analyzed by microscopy as described in EXAMPLE 9. Human cancer cell lines (MDA-MB-231, MG63, U2OS, TC32, TC71 and RH30) were serum starved for 24 h with recombinant VAR2CSA (ID1-  
25 ID2a) or Control protein (DBL4), and lysates prepared at 0, 1, 2, 3, 4, 5, 6 and 12 h after serum was added back to the cells. Using this approach, 100 nM VAR2CSA efficiently inhibited proto-oncogene tyrosine-protein kinase Src phosphorylation on T416, Focal Adhesion Kinase (FAK) phosphorylation at T397, Extracellular-Signal-regulated Kinase (ERK) 1- and 2-phosphorylation at Thr202/Tyr204 for human ERK1 and Thr185/Tyr187 for human  
30 ERK2. This suggests that recombinant VAR2CSA inhibits canonical ERK signaling in cancer cells.

### EXAMPLE 13

- Unbiased analysis of intracellular signalling events modified by recombinant***  
35 ***VAR2CSA.***

The broad impact of VAR2CSA on intracellular signalling events can be analysed using expression microarray technology. MG63 osteosarcoma cells were serum starved for 24 h with no treatment, VAR2CSA or Control (DBL4) and RNA was harvested after 1h serum addition. The total RNA was quality tested (RIN<8), used as a template for Affymetrix®

probe construction and hybridized to the Affymetrix U133Aplus2.0® chip system. This readout provides a snapshot of activated or inactivated signalling pathways after 1 h of serum was added back. Preliminary data confirmed an inhibitory effect on ERK signalling.

#### 5 **EXAMPLE 14**

##### ***Inhibition of cancer cell growth in vivo by recombinant VAR2CSA proteins***

Based on the results from the *in vitro* analysis appropriate cell lines will be selected for *in vivo* subcutaneous and metastatic xenograft models in immuno-compromised mice. The *in vivo* study addresses five main questions:

- 10 i) can *i.v.* or *i.p.* administrated recombinant VAR2CSA trace and bind human cancer cells *in vivo*?
- ii) can *i.v.* or *i.p.* administration of recombinant VAR2CSA inhibit tumor formation *in vivo*?
- 15 iii) can *i.v.* or *i.p.* administration of recombinant VAR2CSA inhibit growth of established tumors *in vivo*?
- iv) can *i.v.* or *i.p.* administration of recombinant VAR2CSA inhibit metastatic spread of human cancer cells *in vivo*?
- 20 v) does *i.v.* or *i.p.* administration of recombinant VAR2CSA change CSA-containing proteoglycan-governed signaling events in human cancer cells *in vivo* (post mortem pathology and biochemistry)?

**In vivo models:** Selected human cancer cell lines representing cancer types showing a strong binding to VAR2CSA are inoculated subcutaneously into Rag2m or SCID immuno-compromised mice at approximately  $5 \times 10^6$  cells/animal. When the tumor is established, the mice receive the first injection of vehicle (Saline) and recombinant VAR2CSA (1 mg/Kg).  
25 Treatment is repeated once a week throughout an experimental period of approximately 30 days. Animal weights and tumor volumes are measured every second or third day and at termination, tumors are removed and divided into two halves, with one half snap-frozen in liquid nitrogen and the other half fixed in paraffin. Snap-frozen tumors are processed for (DHE) superoxide detection as described in EXAMPLE 11 (along with corresponding  
30 hematoxylin and eosin [H&E] staining of the same tumor specimens).

#### **EXAMPLE 15**

##### ***Tracking micro-metastasis in vivo by tracer-coupled recombinant VAR2CSA peptides.***

Recombinant VAR2CSA will be coupled to different applicable tracer-molecules in  
35 collaboration with external partners or outsourced on a contract-based agreement. The traceable recombinant VAR2CSA molecules are analyzed for their ability to track and report

micro-metastasis in both xenograft and transgenic mouse models. In vivo models are established as described in EXAMPLE 12. For testing of tracer-coupled VAR2CSA *in vivo*, mice with metastatic cancer are analyzed by *in vivo* imaging for the ability of VAR2CSA to track and bind micro-metastasis.

5

#### EXAMPLE 16

##### **Internalization of recombinant VAR2CSA proteins**

Recombinant VAR2CSA is internalized by cancer cells. This was shown by first conjugating VAR2CSA fragment (DBL1-ID2a) with a fluorophore and then analysing VAR2CSA uptake

10 both by live imaging and on fixed cells. Cancer cell lines (C32 melanoma and MDA-MB-231) were seeded and grown overnight to 60-80% confluency. Cells were incubated with fluorophore-conjugated VAR2CSA for 10-15 min at 4°C to allow for surface binding of VAR2CSA. Cells were then washed to remove unbound VAR2CSA, and subsequently incubated at 37°C to initiate internalization for 10 min, 1 h, 2 h, 4 h, and up to 22 h.

15 Fluorophore-conjugated transferrin was used for following classical clathrin-dependent uptake of transferrin ending up in the lysosomes. In addition, for some experiments fluorophore-conjugated dextran was used for detecting lysosomes. The live imaging analysis showed that VAR2CSA starts to reach lysosomes after around 4 h, and after 22 h all VAR2CSA can be localized to the lysosomal compartments. However, colocalization of VAR2CSA and transferrin

20 was scarce, and VAR2CSA was taken up much slower than transferrin. The fact that recombinant VAR2CSA is taken up by cancer cells, allows us to fuse or conjugate VAR2CSA to cytotoxic compounds that become active inside the cancer cell. Table 10 summarizes the result from indicated cancer cell lines tested for internalization of recombinant VAR2CSA.

25 **Table 10. Cells were incubated with medium alone (blank) or recombinant proteins (DBL1-ID2a or ID1-ID2a) at 75 nM for 1h, followed by incubation with anti-V5-FITC (Invitrogen) at 1:800, cells were washed thrice between each incubation. Shown are the medium score of FITC fluorescence intensity at either the plasma membrane or intracellular structures recorded from a minimum of 4 high power field images using a HAL100 Zeiss microscope.**

30

**Scoring system is:**

**+: weak; ++: medium; +++: strong; ++++: Very strong.**

| Cell line | Plasma membrane localization (after 1h) | Intracellular localization (after 1h) |
|-----------|-----------------------------------------|---------------------------------------|
| U2OS      | +                                       | ++++                                  |
| RH30      | +                                       | +++                                   |
| MG63      | +                                       | ++++                                  |
| MeWo      | +                                       | ++++                                  |
| HOS       | +                                       | +++                                   |

|            |      |      |
|------------|------|------|
| MDA-MB-231 | +    | ++++ |
| SKNMC      | ++++ | (+)  |
| RH18       | +    | +++  |
| TC71       | +    | ++   |
| TC3        | +    | ++   |

**EXAMPLE 17*****Fused VAR2CSA-toxin protein kills cancer cells***

- 5 DBL1-ID2a and ID1-ID2a VAR2CSA gene fragments have been fused to *Pseudomonas* exotoxin A and diphtheria toxin as various constructs (SEQ ID NO:60-70, 72). These fused VAR2CSA-toxin proteins are expressed in *E. coli*. The protein construct called BPTI-ID1-ID2aFCR3-PE38LR (SEQ ID NO:60), which is based on ID1-ID2a from VAR2CSA and PE38 has been successfully produced and analysed for binding to cancer cells (Table 11) as well as
- 10 cytotoxic activity as described in Method 13.
- Preliminary data show that this fused VAR2CSA-toxin protein binds to CSA-expressing cancer cells and is able to induce cell death (IC<sub>50</sub> for the U2OS cell line is below 1 nM).

**Table 11. Binding of VAR2CSA-PE38 to cancer cells analysed by flow cytometry**

- 15 Binding of DBL1-ID2a (naked protein) and ID1-ID2a-PE38 at 200 nM to Myla2059 cells (T cell lymphoma) was detected with anti-PENTA HIS antibody and anti-mouse-FITC antibody and analysed by flow cytometry. Binding is given as mean fluorescence intensity (MFI). a) cells were treated with chondroitinase ABC to remove CS chains from cell surface, b) protein was mixed with soluble CSA (400 ug/m) prior to addition to cells, c) control equals cells stained
- 20 with the first and second layer of antibodies only.

|                                       | DBL1-ID2a | ID1-ID2a-PE38 | Control <sup>c</sup> |
|---------------------------------------|-----------|---------------|----------------------|
| Binding to cells                      | 24.7      | 12.4          | 2.3                  |
| Binding to treated cells <sup>a</sup> | 4.4       | 2.5           | 2.5                  |
| Inhibition of binding <sup>b</sup>    | 3.2       | 2.1           | -                    |

**EXAMPLE 18*****Analyzing the anti-tumor effect of cytotoxic compounds coupled to recombinant VAR2CSA***

- 25 Based on the results in EXAMPLE 14 recombinant VAR2CSA will be sought coupled to relevant cytotoxic compounds and tested *in vivo* for performance. Coupling of relevant compounds to VAR2CSA will be performed in collaboration with external partners or outsourced on a

contract-based agreement. In particular we analyze whether these VAR2CSA:compound-fusions can:

- i) be delivered specifically to the tumor environment *in vivo*.
- ii) be up-concentrated and retained specifically in the tumor environment *in vivo*.
- iii) specifically kill tumor cells with minimal damage to normal tissues *in vivo*.

In vivo models are established as described in EXAMPLE 12. The mice are treated with cytotoxic VAR2CSA conjugates, and the effect is assayed as described for un-conjugated protein in EXAMPLE 12.

#### EXAMPLE 19

##### ***Purification of CSA-expressing stem cells from heterogeneous cell populations.***

Pluripotent stem cells have been reported to express high levels of CSPG4. Stem cells also express other CSA-containing proteoglycans, such as CD44, which VAR2CSA can bind to. Accordingly, recombinant VAR2CSA will be conjugated to an appropriate resin (beads), mixed with a heterogeneous but stem cell or cancer stem cell-containing cell population and sought purified by conventional centrifugation protocols. Purified cells will be analyzed for expression of diverse stem cell markers including CD44, CD31, CD4, OCT4, SOX2, Nestin and Nanog, by immunoblotting (as in EXAMPLE 11), microscopy and FACS (as in EXAMPLE 9). A common trait of cancer stem cells is high expression of Aldehyde dehydrogenase 1A (ALDH1 High). This can be conveniently measured using the AldeFluor® Kit (Stem Cell Technologies). Recombinant VAR2CSA binding to MDA-MB-231 detects a subpopulation of ALDH1 High cells, suggesting that VAR2CSA can bind human cancer stem cells.

#### EXAMPLE 20

##### ***Identification and targeting of CD44-expressing cancer stem cells.***

CD44 is currently the most popular marker for cancer stem cells and it is a CSA-containing proteoglycan that can bind recombinant VAR2CSA. By using the same approaches as in EXAMPLE 12-15, it will be investigated whether un-modified and modified recombinant VAR2CSA peptides can locate, bind, purify and potentially kill the highly resistant CD44-positive cancer stem cells.

#### EXAMPLE 21

##### ***Detection of circulating tumor cells***

We will examine whether recombinant VAR2CSA can be used as a prognostic marker for cancer recurrences. Cancer cells spread through the blood system after detachment from the primary tumor. A subsequent risk of the occurrence of circulating tumor cells (CTCs) is extravasation and metastasis. Current assays used for detecting CTCs have a poor sensitivity and cannot be directly correlated with risk of metastases. Using VAR2CSA-coupled magnetic

beads and flow cytometry, we will investigate the prognostic value of detecting CS expressing cancer cells in the blood flow. This method could be used as a fast and painless examination of patients.

## 5 EXAMPLE 22

### ***Identification of potential CSPG molecules that were targeted by VAR2CSA***

Recombinant VAR2CSA protein (DBL1-ID2a) with a V5-tag was screened for binding to a panel of transfected HEK293 cells expressing >3000 human membrane receptors. A set of 25 receptors have been identified as potential targets of VAR2CSA (Table 12). The interaction  
10 between VAR2CSA and these receptors will be further verified by analysis of the binding specificity through inhibition with CSA and HS, both in the HEK293 system and in ELISA.

**Table 12. Receptors that were experimentally identified as potential targets of VAR2CSA.**

| Gene ID | Name                                                                | UniProt/SwissProt                   |
|---------|---------------------------------------------------------------------|-------------------------------------|
| BCAN    | Brevican                                                            | <a href="#">PGCB HUMAN, Q96GW7</a>  |
| BDKRB2  | Bradykinin receptor B2                                              | <a href="#">BKRB2 HUMAN, P30411</a> |
| CA9     | Carbonic anhydrase IX                                               | <a href="#">CAH9 HUMAN, Q16790</a>  |
| CCR10   | chemokine (C-C motif) receptor 10                                   | <a href="#">CCR10 HUMAN, P46092</a> |
| CD44    | CD44 molecule (Indian blood group)                                  | <a href="#">CD44 HUMAN, P16070</a>  |
| CDH8    | Cadherin 8, type 2                                                  | <a href="#">CADH8 HUMAN, P55286</a> |
| CFB     | Complement factor B                                                 | <a href="#">CFAB HUMAN, P00751</a>  |
| GABBR2  | gamma-aminobutyric acid (GABA) B receptor, 2                        | <a href="#">GABR2 HUMAN, O75899</a> |
| GPC3    | Glypican 3                                                          | <a href="#">GPC3 HUMAN, P51654</a>  |
| GPC5    | Glypican 5                                                          | <a href="#">GPC5 HUMAN, P78333</a>  |
| GPR65   | G-protein coupled receptor 65                                       | <a href="#">PSYR HUMAN, Q8IYL9</a>  |
| GPRC5B  | G protein-coupled receptor, family C, group 5, member B             | <a href="#">GPC5B HUMAN, Q9NZH0</a> |
| KCNA2   | potassium voltage-gated channel, shaker-related subfamily, member 2 | <a href="#">KCNA2 HUMAN, P16389</a> |
| PKD2    | polycystic kidney disease 2 (autosomal dominant)                    | <a href="#">PKD2 HUMAN, Q13563</a>  |
| PODXL2  | podocalyxin-like 2                                                  | <a href="#">PDXL2 HUMAN, Q9NZ53</a> |

|              |                                                                        |                                            |
|--------------|------------------------------------------------------------------------|--------------------------------------------|
| PTPRG        | protein tyrosine phosphatase, receptor type, G                         | <u><a href="#">PTPRG HUMAN, P23470</a></u> |
| S100A9       | S100 calcium binding protein A9                                        | <u><a href="#">S10A9 HUMAN, P06702</a></u> |
| SDC1         | Syndecan 1                                                             | <u><a href="#">SDC1 HUMAN, P18827</a></u>  |
| SDC4         | Syndecan 4                                                             | <u><a href="#">SDC4 HUMAN, P31431</a></u>  |
| STX2         | Syntaxin 2                                                             | <u><a href="#">STX2 HUMAN, P32856</a></u>  |
| STXBP5       | syntaxin binding protein 5 (tomosyn)                                   | <u><a href="#">STXB5 HUMAN, Q5T5C0</a></u> |
| TGFBR3       | transforming growth factor, beta receptor III                          | <u><a href="#">TGBR3 HUMAN, Q03167</a></u> |
| TMEFF1       | transmembrane protein with EGF-like and two follistatin-like domains 1 | <u><a href="#">TEFF1 HUMAN, Q8IYR6</a></u> |
| TMEFF2/TENB2 | transmembrane protein with EGF-like and two follistatin-like domains 2 | <u><a href="#">TEFF2 HUMAN, Q9UIK5</a></u> |
| TMEM154      | Transmembrane protein 154                                              | (None)                                     |

| Gene ID | Name                               | UniProt/SwissProt                          |
|---------|------------------------------------|--------------------------------------------|
| BCAN    | Brevican                           | <u><a href="#">PGCB HUMAN, Q96GW7</a></u>  |
| BDKRB2  | Bradykinin receptor B2             | <u><a href="#">BKRB2 HUMAN, P30411</a></u> |
| CA9     | Carbonic anhydrase IX              | <u><a href="#">CAH9 HUMAN, Q16790</a></u>  |
| CCR10   | chemokine (C-C motif) receptor 10  | <u><a href="#">CCR10 HUMAN, P46092</a></u> |
| CD44    | CD44 molecule (Indian blood group) | <u><a href="#">CD44 HUMAN, P16070</a></u>  |
| CDH8    | Cadherin 8, type 2                 | <u><a href="#">CADH8 HUMAN,</a></u>        |

|               |                                                                        |                                         |
|---------------|------------------------------------------------------------------------|-----------------------------------------|
|               |                                                                        | <a href="#">P55286</a>                  |
| CFB           | Complement factor B                                                    | <a href="#">CFAB_HUMAN,<br/>P00751</a>  |
| GABBR2        | gamma-aminobutyric acid (GABA) B receptor, 2                           | <a href="#">GABR2_HUMAN,<br/>Q75899</a> |
| GPC3          | Glypican 3                                                             | <a href="#">GPC3_HUMAN,<br/>P51654</a>  |
| GPC5          | Glypican 5                                                             | <a href="#">GPC5_HUMAN,<br/>P78333</a>  |
| GPR65         | G-protein coupled receptor 65                                          | <a href="#">PSYR_HUMAN, Q8IYL9</a>      |
| GPRC5B        | G protein-coupled receptor, family C, group 5, member B                | <a href="#">GPC5B_HUMAN,<br/>Q9NZH0</a> |
| KCNA2         | potassium voltage-gated channel, shaker-related subfamily, member 2    | <a href="#">KCNA2_HUMAN,<br/>P16389</a> |
| PKD2          | polycystic kidney disease 2 (autosomal dominant)                       | <a href="#">PKD2_HUMAN,<br/>Q13563</a>  |
| PODXL2        | podocalyxin-like 2                                                     | <a href="#">PDXL2_HUMAN,<br/>Q9NZ53</a> |
| PTPRG         | protein tyrosine phosphatase, receptor type, G                         | <a href="#">PTPRG_HUMAN,<br/>P23470</a> |
| S100A9        | S100 calcium binding protein A9                                        | <a href="#">S10A9_HUMAN,<br/>P06702</a> |
| SDC1          | Syndecan 1                                                             | <a href="#">SDC1_HUMAN,<br/>P18827</a>  |
| SDC4          | Syndecan 4                                                             | <a href="#">SDC4_HUMAN,<br/>P31431</a>  |
| STX2          | Syntaxin 2                                                             | <a href="#">STX2_HUMAN,<br/>P32856</a>  |
| STXBP5        | syntaxin binding protein 5 (tomosyn)                                   | <a href="#">STXB5_HUMAN,<br/>Q5T5C0</a> |
| TGFBR3        | transforming growth factor, beta receptor III                          | <a href="#">TGBR3_HUMAN,<br/>Q03167</a> |
| TMEFF1        | transmembrane protein with EGF-like and two follistatin-like domains 1 | <a href="#">TEFF1_HUMAN,<br/>Q8IYR6</a> |
| TMEFF2/TE NB2 | transmembrane protein with EGF-like and two follistatin-like domains 2 | <a href="#">TEFF2_HUMAN,<br/>Q9UIK5</a> |
| TMEM154       | Transmembrane protein 154                                              | (None)                                  |

|        |                                                    |                                      |
|--------|----------------------------------------------------|--------------------------------------|
| THBD   | Thrombomodulin                                     | <u>TRBM_HUMAN,</u><br><u>P07204</u>  |
| CSPG5  | chondroitin sulfate proteoglycan 5 (neuroglycan C) | <u>CSPG5_HUMAN,</u><br><u>O95196</u> |
| STXBP5 | syntaxin binding protein 5 (tomosyn)               | <u>STXB5_HUMAN,</u><br><u>Q5T5C0</u> |

## DISCUSSION

Malaria is one of the most common infectious diseases and one of the largest global health problems. Pregnant women are especially vulnerable to infection, despite previously acquired immunity. In this study we have addressed key questions related to the molecular mechanism behind the VAR2CSA-CSA interaction in PM.

Previous work has suggested that the minimal CSA binding region in VAR2CSA is DBL2X-ID2b, with the need for DBL1X or DBL3X for full affinity binding (Dahlback, M., Jorgensen, L. M., Nielsen, M. A., Clausen, T. M., Ditlev, S. B., Resende, M., Pinto, V. V., Arnot, D. E., Theander, T. G., and Salanti, A. *J Biol Chem* 286, 15908-15917). In continuation of this work we made further truncations of VAR2CSA, focusing on the DBL2X region. We show that the core CSA-binding site lies within the DBL2X domain including small parts of the flanking interdomain regions. The binding does not depend on the ID2b region, or on the DBL1X or DBL3X flanking domains, as previously suggested. This is evident by the specific CSPG binding of ID1-ID2a and ID1-DBL2Xb (Table 3). The minimal binding region is ID1-DBL2Xb, which bound CSPG with characteristics comparable to that of full-length VAR2CSA.

It is interesting that these new data maps the core-CSA binding site onto a single domain. Binding of DBL2X (and any other single DBL domain) to CSA has previously been shown to be non-specific and of weak affinity (Resende, M., Ditlev, S. B., Nielsen, M. A., Bodevin, S., Bruun, S., Pinto, V. V., Clausen, H., Turner, L., Theander, T. G., Salanti, A., and Dahlback, M. (2009) *Int J Parasitol* 39, 1195-1204). It is clear that the ID1 and parts of the ID2a interdomains are essential for CSA binding. DBL1X-DBL2Xa and ID1-DBL2Xa did not bind CSPG. The two C-terminal DBL2X borders (DBL2Xa and DBL2Xb) differ by 93 amino acids. Since deletion of these amino acids eliminates binding they must be important for CSA binding.

The ID1-DBL2Xb minimal binding region is much smaller than full-length VAR2CSA, having a molecular weight of only 62 kDa. It is unlikely that further substantial truncations of

VAR2CSA will be functional in binding CSA. Our data redefines DBL2X as a larger functional domain, incorporating parts of the flanking ID1 and ID2a interdomains.

A VAR2CSA based vaccine against PM must be able to induce a strong protective immune response. In this, the most important aspect is the generation of IgG antibodies capable of inhibiting placental-specific parasite adhesion. To test the immunogenic characteristics of our produced fragments, we used them in the immunization of rats. Sera raised against all fragments containing the CSA binding site inhibited parasite adhesion to CSA. Importantly, sera raised against ID1-ID2a resulted in almost complete inhibition. This suggests that the minimal CSA binding fragments retain the capacity for inducing a strong anti-adhesive immune response. This conclusion was further supported by the fact that antibodies purified from anti-FV2 serum on ID1-ID2a retained most of the adhesion blocking activity, and that the anti-ID1-ID2a antibody depleted anti-FV2 sample lost most of its activity. This indicates that epitopes required for the induction of adhesion blocking antibodies are located within this region.

In this study we have tested anti-VAR2CSA sera in homologous inhibition of FCR3 parasites binding to CSA. It is important that a vaccine is capable of inhibiting placental adhesion regardless of parasite strain origin. A major concern in vaccine development is therefore the high interclonal diversity among parasite variants. While recombinant full-length VAR2CSA is very immunogenic the antibodies produced are not cross-inhibitory (Avril, M., Hathaway, M. J., Srivastava, A., Dechavanne, S., Hommel, M., Beeson, J. G., Smith, J. D., and Gamain, B. *PLoS One* 6, e16622). A recent study shows that DNA-vaccination with ID1-DBL2X from FCR3, induces antibodies that are cross-inhibitory, inhibiting CSA adhesion of other laboratory strains as well as parasites isolated in the field (Bordbar, B., Tuikue-Ndam, N., Bigey, P., Doritchamou, J., Scherman, D., and Deloron, P. *Vaccine*). This supports the use of this small fragment in a PM vaccine.

Cardin and Weintraub predicted that a GAG binding site would take one of two forms (Cardin, A. D., and Weintraub, H. J. (1989) *Arteriosclerosis* 9, 21-32). These are X-B-B-X-B-X and X-B-B-B-X-X-B-X, where X is any hydrophobic residue and B is any basic residue, with a preference for arginine. Both of these describe a binding site for a sulfated disaccharide. While many interactions may occur, the ionic interaction between negatively charged sulfates and basic amino acids are thought to be most important. We mutated two such sites within the minimal binding region; 625-GKNLKKRY-632 in DBL2X and 458-NKKKECKD-465 in ID1. We also deleted a large region within a dimorphic sequence motif (DSM) located in the N-terminal part of DBL2X, as this has been suggested to have a function in binding. Deletion of the DSM region had no effect on CSA binding. Neither did any substitutions in the putative

GAG binding sites. This is a clear indication that these sites have little or no function in CSA binding.

It has been shown that the minimal binding requirement for the human CSA receptor is a dodecasaccharide with 2-4 C4 sulfated GalNAc monosaccharides (Alkhalil, A., Achur, R. N., Valiyaveetil, M., Ockenhouse, C. F., and Gowda, D. C. (2000) *J Biol Chem* 275, 40357-40364). It is remarkable that the VAR2CSA expressing parasites, *in vivo*, are very specific for CSA carrying only 2-8% C4 sulfated disaccharide units. To examine if the VAR2CSA-CSA complex formation is dependent on ionic interactions, we tested binding at different salt concentrations. Binding of ID1-ID2a, DBL1X-ID2a and FV2 in 150 mM-300 mM NaCl show a linear relationship when plotting  $\text{Log}(K_{D,\text{observed}})$  vs.  $\text{Log}[\text{Na}^+]$ . We find that binding depends on a maximum of 2-3 ionic interactions. It is interesting that the value for the full-length protein is higher than for the shorter fragments, indicating that this protein makes an additional ionic interaction with CSA. We have in this study screened for fragments containing the CSA specific high-affinity binding region. It is possible that more interactions occur in downstream regions of the protein, but that the core site lies within DBL2X. Extrapolating and finding the Y intercept ( $[\text{Na}^+] = 1 \text{ M}$ ,  $\text{Log}[\text{Na}^+] = 0$ ) tells us that  $K_{D,\text{nonionic}} = 5.9 \mu\text{M}$  for FV2,  $K_{D,\text{nonionic}} = 3.4 \mu\text{M}$  for DBL1X-ID2a, and  $K_{D,\text{nonionic}} = 0.7 \mu\text{M}$  for ID1-ID2a. This indicates that only 25-30% of the VAR2CSA-CSA binding can be accounted for by ionic interactions. This is in contrast to other GAG binding proteins, which have shown up to 80-90% dependency on ionic interactions in similar assays (Faller, B., Mely, Y., Gerard, D., and Bieth, J. G. (1992) *Biochemistry* 31, 8285-8290; Hileman, R. E., Fromm, J. R., Weiler, J. M., and Linhardt, R. J. (1998) *Bioessays* 20, 156-167).

Our data suggest that the VAR2CSA-CSA interaction does not conform to conventional GAG-protein interactions. We hypothesize that the high CSA affinity is achieved through a multivalent interaction, which may include multiple binding sites making nonionic interactions with the CSA carbohydrate backbone. Some of the interaction is ionic and some degree of sulfation is needed for VAR2CSA binding. It is therefore likely that there is an interaction between basic amino acids and sulfates, but that this is not the determining factor in the affinity.

In this study we have defined a small single-domain VAR2CSA fragment that can be produced in eukaryotic cells as a functional CSA-binding protein, and has the capacity to induce highly adhesion-blocking antibodies. This fragment has the potential to be a powerful candidate for a vaccine against PM.

The data identifies a small recombinant part of VAR2CSA that binds specifically to CSA thereby mediating placental binding of infected erythrocytes. We show that this VAR2CSA

fragment also binds specifically to cancer cells, through an interaction with CSA presented on CSPG4 or other protein backbones that were identified in this study. In addition, we find that binding of VAR2CSA polypeptides, based on this small fragment, to cancer cells inhibits migration and invasion of the cells. These VAR2CSA polypeptides also inhibit canonical ERK signaling, and we find that VAR2CSA polypeptides that are fused to a toxin efficiently kill the cancer cells.

## METHODS

### METHOD 1 - Cloning and protein expression in insect cells

VAR2CSA sequence fragments were amplified from codon optimized FCR3 (GenBank accession no. GU249598) or 3D7 (GenBank accession no. JQ247428) VAR2CSA genes using specific primers (Table 2). Simple fragments were amplified in a one-step PCR. Amino acid substitution constructs were made in a two-step PCR. First PCR amplified two fragments from the codon optimized FCR3 template, containing overlapping complimentary ends. Second PCR amplified the total construct, using the two overlapping fragments as template with primers specific for the outer borders. All fragments were sequenced for verification. Fragments were cloned into the baculovirus vector pAcGP67-A (BD Biosciences), modified to contain a V5 and His tag at the C-terminal. The proteins were expressed in baculovirus-infected insect cells as soluble protein secreted into the cell culture supernatant. Briefly, linearized Bakpak6 Baculovirus DNA (BD Biosciences) was co-transfected with the pAcGP67-A plasmids, into Sf9 insect cells for generation of recombinant virus particles. 10 ml of the second amplification was used to infect High-Five cells in 400 ml serum-free medium (10486, GIBCO) at a density of  $1 \times 10^6$  cells/ml. The secreted recombinant protein was harvested from the supernatant 3 days after initial infection. The supernatant was filtered (0.2  $\mu$ m), dialyzed and concentrated before protein purification.

### METHOD 2 - Protein Purification and SDS-PAGE

The filtered supernatant containing the secreted recombinant protein was dialyzed using an ÄKTA cross-flow (GE Healthcare). The dialysis was performed in 10 mM  $\text{NaH}_2\text{PO}_4$  (pH 7.4, Sigma-Aldrich) and 500 mM NaCl. The resulting solution was filtered (0.2  $\mu$ m) and imidazole was added to a final concentration of 15 mM. The protein was then purified on a 1-ml HisSelect column (H8286, Sigma-Aldrich). Bound protein was eluted with 10 mM  $\text{NaH}_2\text{PO}_4$  (pH 7.4), 500 mM NaCl, and 500 mM imidazole. Proteins needed for Quartz Crystal Microbalance measurements and SAXS were further purified to obtain monomers by size exclusion chromatography using a HiLoad 16/60 Superdex 200 column (GE Healthcare) in 20 mM Tris

(pH 8) and 200 mM NaCl. The purity and structural integrity of the protein was verified by SDS-PAGE.

### METHOD 3 - ELISA

Falcon microtiter plates (351172, BD Biosciences) were incubated at a concentration of 3µg/ml for CSPG (bovine) (D8428, Sigma) or HSPG (H4777, Sigma) and 100µg/ml for CSA (C9819, Sigma), CSC (400675, Seikagaku), and CSB (C3788, Sigma) overnight at 4°C. The plates were then blocked with TSM binding buffer (20mM Tris, 150mM NaCl, 2mM CaCl<sub>2</sub>, 0.05% Tween-20, 1% BSA, PH7.4 at 25°C) for 2 hours at 37°C on a shaker. A 2-fold dilution series (1.56mM-100mM) of protein was prepared in TSM binding buffer and added to the plates, which was incubated 1hr at 37°C on a shaker. All measurements were performed in triplicates. The plates were washed three times in TSM washing buffer (20mM Tris, 150mM NaCl, 2mM CaCl<sub>2</sub>, 0.05% Tween-20, PH7.4 at 25°C). The plates were then incubated with 1:3000 anti-V5-HRP antibody (R96125, Invitrogen) in TSM binding buffer 1hr at 37°C on a shaker. The plates were washed three times in TSM washing buffer. Finally the plates were developed with o-phenylenediamine substrate (DAKO) for 15min. The reaction was quenched with 2.5M H<sub>2</sub>SO<sub>4</sub>. Absorbance was measured at 490nm.

### METHOD 4 - Quartz Crystal Microbalance (Attana A100)

Experiments were performed on an Attana A100 (Attana AB), using gold plated 10MHz, AT-cut quartz crystal, polystyrene chips (3611-3103 Attana AB). All buffers and reagents were filtered to 0.2µm. The ligand was CSPG (Bovine) (D8428, Sigma) or HSPG (H4777, Sigma), coated at a concentration of 100µg/ml. Coating was done in steady state by adding ligand solution and incubation 30 minutes at room temperature. This was followed by blocking the plate with PBS containing 0.1% Ig-free BSA (BSA-50, Rockland), 30 minutes at room temperature. The Attana A100 was washed with 1% SDS prior to every experiment, using the manufacturers predefined daily wash program. Following the wash, the running buffer was switched to PBS at a flow rate of 25µL/min, at 25°C, and the machine was allowed to stabilize at a maximum change in frequency of 0.5Hz/min. Once stabilized PBS was injected multiple times to show that the injection process minimally affected the baseline. Prior to sample injection PBS was injected as a blank. Analyte was injected in a 1:3 dilution series (0.25µg/ml-60µg/ml) starting with the lowest concentration. Association time was set to 84 seconds and disassociation time to 5 minutes. Due to high affinity of binding it was not possible to regenerate binding surface following injections. The data collected was processed in the Attaster Evaluation software (Attana AB). Curves were fitted in a simple 1:1 model.  $k_{on}$  and  $k_{off}$  were determined by curve fitting and  $K_D$  was calculated based  $K_D = k_{off} / k_{on}$ .

### METHOD 5 - Salt Titration assay

The ionic dependency of VAR2CSA-CSA binding was tested in an ELISA based binding assay. CSPG was coated at 3 µg/ml. A 1:2 dilution series (400-1.56 nM) of protein was added in several different NaCl concentrations (150 mM, 200 mM, 250 mM, and 300 mM). All experiments were performed in triplicates. The  $K_D$  values were calculated for each titration series in Graphpad Prism using non-linear regression (Least squares fit with hill slope).

#### **METHOD 6 - Animal immunizations and serum extraction**

All animal immunizations complied with national and European regulations. Wistar rats were injected subcutaneously with 30 µg recombinant protein in Freund's complete adjuvant (F5881, Sigma-Aldrich). The immunization was boosted three times at 3-week intervals with 15 µg protein in Freund's incomplete adjuvant (F5506, Sigma-Aldrich). Blood samples were taken one week after each boost, and serum was extracted by centrifugation.

#### **METHOD 7 - IgG Affinity purification**

Pools of sera from rats immunized with full-length FCR3 VAR2CSA (FV2) were affinity purified on 1 ml NHS-activated HP column (HiTrap NHS-activated HP, 17-0716-01, GE Healthcare), containing immobilized multidomain FCR3 proteins (DBL1X-DBL2Xa, DBL1X-ID2a, ID1-ID2a, or ID1-DBL4e) and full-length FV2. Purification was done according to the manufacturer's protocol. In short, coupling of ligand to column was done by adding 1 ml 1:1 solution of coupling buffer (0.2 M NaHCO<sub>3</sub>, 0.5 M NaCl, pH 8.3) and ligand (concentration 0.5-10 mg/ml) to the column. The column was sealed and incubated for 30 min at room temperature, followed by incubation at 4°C overnight. The column was washed with 6 ml Buffer A (0.5 M ethanolamine, 0.5 M NaCl, pH 8.3), 6 ml Buffer B (0.1 M acetate, 0.5 M NaCl, pH 4) and finally 6 ml Buffer A. After an incubation period of 30 min at room temperature, the washing was repeated in reverse order (Buffer B, A, B). 8-10 ml PBS was injected to adjust the pH before purifying the serum. The sample was passed through the column 3-5 times. The column was washed with 10 ml PBS before antibodies were eluted with 10 ml elution buffer (0.1 M citric acid, pH 2.7).

#### **METHOD 8 - *P. falciparum* parasite cultures**

*P. falciparum* FCR3 type parasites were maintained in culture using 5% hematocrit (Human Blood-group 0 Rh+) in parasite medium RPMI-1640 (BE12115F, Lonza) supplemented with 25 mM NaHCO<sub>3</sub>, 0.125 µg/ml gentamycin sulfate (BE02012E, Lonza), 0.125 µg/ml AlbuMAX II (11021029, Invitrogen) and 2% normal human serum. IEs were repeatedly panned on BeWo cells (CCL98, ATCC) to maintain the CSA adhering phenotype. Furthermore, isolates

were tested to be mycoplasma negative and were regularly genotyped by PCR using nested GLURP (Glutamate-rich protein) and MSP-2 (Merozoite surface protein 2) primers.

#### **METHOD 9 - Purification of late stage trophozoites**

Parasite cultures were enriched for late trophozoite and schizont stage in a strong magnetic field using a MACS CS-column (130-041-305, Miltenyi Biotec) and a Vario-MACS magnet (Miltenyi Biotec). In brief, the parasite culture suspension was applied to the column. The column was then washed with 2% fetal calf serum (F6178, Sigma-Aldrich) in PBS. Late-stage infected erythrocytes were elute from the column after separation from the magnet, spun down and resuspended in 2% fetal calf serum in PBS and diluted to a concentration of  $2 \times 10^6$  IEs/ml.

#### **METHOD 10 - Flow Cytometry (FCM)**

Antibody binding to native VAR2CSA on the purified late-stage trophozoites infected erythrocytes, was measured by flow cytometry (FCM). 100  $\mu$ l purified late-stage parasites at a concentration of  $2 \times 10^5$  IEs/ml in PBS with 2% FCS were labeled with serum (depleted for non-specific binding by pre-incubation with non-infected erythrocytes) in a final concentration of 1:10. The cells were washed three times in PBS with 2% FCS. The cells were then further labeled with ethidium bromide (15585011, Invitrogen) in a final concentration of 2  $\mu$ g/ml and a 1:100 dilution of FITC labeled secondary anti-rat-IgG antibody (62-9511, Invitrogen). As negative controls, late-stage parasites were also incubated with serum from rats immunized with an antigen other than VAR2CSA and with secondary antibodies alone. Data from 5000 ethidium bromide positive IEs were collected using a FC500 flow cytometer (Beckmann Coulter). Finally the median fluorescence intensity was determined using the WinList 5.0 software (Verify Software House).

#### **METHOD 11 - Inhibition of parasites binding CSPG**

Serum antibodies were analyzed for their ability to inhibit IE binding to CSPG. This was done in a 96-well plate format using a robot-standardized washing method. Wells were coated with 2  $\mu$ g/ml CSPG (D8428, Sigma-Aldrich). A total of  $2 \times 10^5$  tritium labeled (Hypoxanthine Monohydrochloride, PerkinElmer, NET177005MC) late stage IEs in 100  $\mu$ L were added in triplicates to the wells. The labeled IEs were then incubated with serum for 90 min at 37°C. Unbound IEs were washed away by a pipetting robot (Beckman Coulter). The proportion of adhering IEs was determined by liquid scintillation counting on a Topcount NXT (Perkin Elmer).

**METHOD 12a - Cancer cell binding assays**

Flow-cytometry (FCM) was used to test the reactivity of the VAR2CSA minimal binding polypeptide to CSPG expressed on the surface of various cell lines. Cells were cultured in RPMI supplemented with 10 % foetal calf serum (CHO cells, C32), Hams F12 (BeWo), kept in 5% carbon dioxide at 37°C or purified from a human blood sample in CPD buffer (red blood cells). Aliquots of cells ( $1 \times 10^5$ ) were sequentially exposed to the VAR2CSA minimal binding polypeptide (150, 75 or 37 nM) and  $\alpha$ -V5-FITC (1:800)(Invitrogen) diluted in FACS2 (PBS+2% FCS) for 30 minutes at +4°C in dark with smooth agitation. As negative controls a truncated version of the minimal binding polypeptide and FACS2 buffer were used. Intact cells were gated based on the forward and side scatter signal. Data were acquired using a FC500 flow-cytometer (Beckman Coulter) from a minimum of 5000 cells. All samples relating to a particular cell line were processed and analyzed in a single assay.

**METHOD 12b - Cancer cell binding assays**

As an alternative to the flow-cytometry assay above, cells were incubated with VAR2CSA minimal binding polypeptide and  $\alpha$ -V5-FITC (1:500)(Invitrogen) diluted in HBSS. VAR2CSA polypeptide was used at the same concentrations as written above. Following  $\alpha$ -V5-FITC staining cells were washed 3 times in HBSS, collected in Enzyme-free cell detachment buffer (Invitrogen) and analyzed on a FACS Calibur (BD Biosciences) for FL-1 signal intensity.

Abbreviations CIDR, cysteine-rich inter-domain region; CSA, chondroitin sulfate A; CSPG, chondroitin sulfate proteoglycan; DBL, Duffy binding-like domain; FCM, flow-cytometry; FV2, full-length ecto-domain of the VAR2CSA protein without N-terminal segment; HSPG, heparan sulfate proteoglycan; ID, inter-domain; IE, *P. falciparum*-infected erythrocyte; NTS, N-terminal segment; PM, placental malaria; PfEMP1, *Plasmodium falciparum* erythrocyte membrane protein 1; PM, placental malaria.

**METHOD 13 – Cytotoxicity test *in vitro* of fused VAR2CSA-toxin proteins**

Cancer cell lines were seeded in a 96-well plate, with 500,0 cells/well one day before the experiment. On the day of experiment a 10-fold dilution series (ranging from 10 µg/ml to 0.01 ng/ml) of fused VAR2CSA-toxin, and control protein (VAR2CSA without toxin) was added to separate wells. Similar dilution series, which also contained 400 µg/ml of CSA, was made for both proteins and added to separate wells. The cells with proteins were incubated for 72 hours at 37 °C. Cell death was analyzed by a MTT cell proliferation assay, where readout is absorbance at 570 nm.

**METHOD 14 – Staining of paraffin-embedded human tissue samples**

The binding of recombinant VAR2CSA to primary cancer tissue obtained from human patients is investigated using immunohistochemistry (IHC). Paraffin embedded tissue spotted on glass slides subjected no antigen retrieval was incubated with 0.1-500 nM V5-VAR2CSA variants or V5-Control protein (DBL4) for 1 h in room temperature, washed for 8 minutes, incubated with 1:700 mouse anti-V5 antibody for 30 minutes, washed for 8 minutes. Bound anti-V5 was subsequently detected using UltraMap anti-mouse HRP using the Ventana Discovery XT platform.

5

## CLAIMS

1. An isolated protein fragment of VAR2CSA, which fragment consist of a sequential amino acid sequence of

a) ID1, and

b) DBL2Xb, and optionally

c) ID2a.

2. The isolated protein fragment according to claim 1, which fragment binds chondroitin sulfate A (CSA) on proteoglycans (CSPG) with an affinity as measured by a  $K_D$  lower than 100 nM, such as lower than 80 nM, such as lower than 70 nM, such as lower than 60 nM, such as lower than 50 nM, such as lower than 40 nM, such as lower than 30 nM, such as lower than 26 nM, such as lower than 24 nM, such as lower than 22 nM, such as lower than 20 nM, such as lower than 18 nM, such as lower than 16 nM, such as lower than 14 nM, such as lower than 12 nM, such as lower than 10 nM, such as lower than 9 nM, such as lower than 8 nM, such as lower than 7 nM, such as lower than 6 nM, or lower than 4 nM.

3. The isolated protein fragment according to any one of claims 1-2, which fragment comprises an amino acid sequence having at least 70 % sequence identity with any one amino acid sequence of 1-577 of SEQ ID NO:1, 1-592 of SEQ ID NO:3, 1-579 of SEQ ID NO:4, 1-576 of SEQ ID NO:5, 1-586 of SEQ ID NO:10, 1-579 of SEQ ID NO:11, 1-565 of SEQ ID NO:29, 1-584 of SEQ ID NO:34, 1-569 of SEQ ID NO:36, 1-575 of SEQ ID NO:37, 1-592 of SEQ ID NO:38, 1-603 of SEQ ID NO:41, 1-588 of SEQ ID NO:43, 1-565 of SEQ ID NO:44, 1-589 of SEQ ID NO:45, 1-573 of SEQ ID NO:48, 1-583 of SEQ ID NO:53, or 1-569 of SEQ ID NO:54.

4. The isolated protein fragment according to any one of claims 1-3, which fragment comprises an amino acid sequence having at least 70 % sequence identity with an amino acid sequence of 578-640 of SEQ ID NO:1, 593-656 of SEQ ID NO:3, 580-643 of SEQ ID NO:4, 577-640 of SEQ ID NO:5, 587-650 of SEQ ID NO:10, 580-643 of SEQ ID NO:11, 566-628 of SEQ ID NO:29, 585-647 of SEQ ID NO:34, 570-632 of SEQ ID NO:36, 576-639 of SEQ ID NO:37, 593-655 of SEQ ID NO:38, 604-667 of SEQ ID NO:41, 589-652 of SEQ ID NO:43, 566-628 of SEQ ID NO:44, 590-653 of SEQ ID NO:45, 574-637 of SEQ ID NO:48, 584-646 of SEQ ID NO:53, or 570-632 of SEQ ID NO:54.

5. The isolated protein fragment according to any one of claims 1-4, which fragment comprises an amino acid sequence having at least 70 % sequence identity with an amino acid sequence of SEQ ID NO:2, 6, 8, 9, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 30, 31, 32, 33, 35, 39, 40, 42, 46, 47, 49, 50, 51, or 52.

6. The isolated protein fragment according to any one of claims 1-4, which fragment consists of an amino acid sequence having at least 70 % sequence identity with any one amino acid sequence of 1-577 of SEQ ID NO:1, 1-592 of SEQ ID NO:3, 1-579 of SEQ ID NO:4, 1-576 of SEQ ID NO:5, 1-586 of SEQ ID NO:10, 1-579 of SEQ ID NO:11, 1-565 of SEQ ID NO:29, 1-584 of SEQ ID NO:34, 1-569 of SEQ ID NO:36, 1-575 of SEQ ID NO:37, 1-592 of SEQ ID NO:38, 1-603 of SEQ ID NO:41, 1-588 of SEQ ID NO:43, 1-565 of SEQ ID NO:44, 1-589 of SEQ ID NO:45, 1-573 of SEQ ID NO:48, 1-583 of SEQ ID NO:53, or 1-569 of SEQ ID NO:54.
7. The isolated protein fragment according to claim 1, which fragment consists of an amino acid sequence selected from the list consisting of SEQ ID NO:1, 3-5, 10, 11, 29, 34, 36-38, 41, 43-45, 48, 53, and 54.
8. The isolated protein fragment according to any one of claims 1-5, which fragment consists of an amino acid sequence having a length of less than 700 amino acids, such as less than 690 amino acids, such as less than 680 amino acids, such as less than 670 amino acids, such as less than 660 amino acids, such as less than 650 amino acids, such as less than 640 amino acids, such as less than 630 amino acids, such as less than 620 amino acids, such as less than 610 amino acids, such as less than 600 amino acids, such as less than 590 amino acids, such as less than 580 amino acids, such as less than 570 amino acids.
9. The isolated protein fragment according to any one of claims 1-8, which is substantially pure.
10. The isolated protein fragment according to any one of claims 1-9, wherein said protein fragment has a molecular mass of less than about 100 kDa under non-reducing conditions on an SDS-PAGE.
11. The isolated protein fragment according to any one of claims 1-10, wherein said protein fragment is a recombinant protein.
12. The isolated protein fragment according to any one of claims 1-11, wherein said protein fragment is non-glycosylated.
13. An antibody that specifically binds a protein fragment according to any one of claims 1-11.
14. An isolated nucleic acid molecule encoding a protein fragment of any one of the claims 1-11.

15. A vector comprising an isolated nucleic acid molecule according to claim 14.
16. A host cell comprising a vector according to claim 15.
17. A method for producing the protein fragment as defined in any one of claims 1-11, the method comprising cultivating a cell as defined in claim 16 in an appropriate growth medium  
5 under conditions allowing expression of the polynucleotide construct and recovering the resulting protein fragment from the culture medium.
18. A conjugate or fusion protein comprising a VAR2CSA polypeptide, and a therapeutic or diagnostic effector moiety, such as a cytotoxic moiety, fluorescent label, and/or a radiolabel.
19. The conjugate according to claim 18, wherein the VAR2CSA polypeptide is or  
10 comprises a protein fragment as defined in any one of claims 1-11.
20. The conjugate according to any one of claims 18 or 19, wherein the cytotoxic moiety is selected from calicheamycin, auristatin, doxorubicin, maytansinoid, taxol, ecteinascidin, geldanamycin, methotrexate and their derivatives, cytotoxic proteins such as *Pseudomonas* exotoxin A, diphtheria toxin, ricin toxin, pokeweed antiviral protein, saporin, gelonin and  
15 functional variants, fragments, and combinations thereof.
21. The conjugate according to any one of claims 18-20, wherein the cytotoxic moiety is a synthetic chemotoxin not derived from a naturally occurring compound.
22. The conjugate according to any one of claims 18-21, in which said label is selected from the group consisting of fluorescent dyes and metals, chromophoric dyes,  
20 chemiluminescent compounds and bioluminescent proteins, enzymes, radionuclides and particles.
23. The conjugate according to any one of claims 18-20, wherein the therapeutic or diagnostic effector moiety is an anti-inflammatory agent.
24. The conjugate according to any one of claims 18-20, wherein the therapeutic or  
25 diagnostic effector moiety is CSPG4, CD44, or other proteoglycans exemplified but not limited to those in Table 1.
25. A composition comprising the protein fragment as defined in any one of claims 1-11, the antibody according to claim 13, or a conjugate according to any one of claim 18-24.

26. The protein fragment as defined in any one of claims 1-11, the antibody according to claim 13, a VAR2CSA polypeptide, or a conjugate according to any one of claim 18-24 for use as a medicament or diagnostic agent.

27. The protein fragment as defined in any one of claims 1-11, the antibody according to claim 13, a VAR2CSA polypeptide, or a conjugate according to any one of claim 18-24 for use in diagnosis.

28. A pharmaceutical composition comprising the protein fragment as defined in any one of claims 1-11, a VAR2CSA polypeptide, or a conjugate according to any one of claim 18-24.

29. A method for detecting a protein fragment as defined in any one of claims 1-11, or a conjugate according to any one of claim 18-24 in a biological sample, said method comprising:

- a) obtaining a biological sample;
- b) contacting said biological sample with an antibody according to claim 13; and
- c) detecting complexes of said antibody and said protein fragment or conjugate, if any;

as an indication of the presence of said protein fragment or conjugate in said sample.

30. Use of a protein fragment as defined in any one of claims 1-11, a VAR2CSA polypeptide, or a conjugate according to any one of claim 18-24; for the preparation of a medicament.

31. A protein fragment as defined in any one of claims 1-11, a VAR2CSA polypeptide, or a conjugate according to any one of claim 18-24 for the treatment of any indications associated with a condition involving expression, such as inappropriate expression of CSA, such as in cancer, arthritis, arthrosis, multiple sclerosis and healing after neural damage, cartilage repair, wound healing, and in psoriasis.

32. A protein fragment according to claim 31, wherein said cancer is selected from all CSA-expressing malignancies including carcinomas (including but not limited to Breast carcinoma, Pancreatic carcinoma, Ovarian carcinoma, Endometrial carcinoma, Hepatocellular carcinoma, Lung carcinoma, Colon carcinoma, Prostate carcinoma, Cervix carcinoma, Testis carcinoma, Basal cell skin carcinoma, Clear cell renal cell carcinoma, Head and neck squamous cell carcinoma, Skin squamous cell carcinoma, Vulvar kreatorized squamous cell

carcinoma and Vulvar basal cell carcinoma), sarcomas (including but not limited to Fibrosarcoma, Dedifferentiated chondro- and liposarcoma, Leiomyosarcoma, Liposarcoma, Myxoid liposarcoma, Uterine corpus leiomyosarcoma, Osteosarcoma, Ewing sarcoma and Rhabdomyosarcoma, Synovial sarcoma, Solitary Fibrous tumor), hematopoietic cancers (including but not limited to Chronic lymphatic leukaemia (CLL), Acute lymphatic leukaemia (ALL), Acute myeloid leukaemia (AML), B-cell, T-cell and large granular lymphoma), tumours of neuroepithelial tissue, such but not limited to Astrocytomas (Pleomorphic Xanthoastrocytoma, Fibrillary Astrocytomas, Anaplastic astrocytoma, Glioblastoma Multiforme), Oligodendroglioma, Ependymoma, Choroid plexus tumor, Oligoastrocytoma, gliosarcoma, Ganglioglioma, Retinoblastoma, Neurocytoma, Neuroblastomas (Esthesioneuroblastoma and Ganglioneuroblastoma), Medulloblastoma, Atypical Teratoid Rhabdoid tumors and all types of neuroendocrine cancer, and any other CSA-expressing cancer.

33. A method for the treatment of any indication associated with expression, such as inappropriate expression of CSA, such as in cancer, arthritis, arthrosis, multiple sclerosis, pathological conditions caused by neural damage, conditions of the cartilage and scar tissue, such as in rheumatism, cartilage repair or wound healing, or in psoriasis; the method comprising administering a therapeutically or prophylactically effective amount of a protein fragment as defined in any one of claims 1-11, a VAR2CSA polypeptide, or a conjugate according to any one of claim 18-24 to a subject in need thereof.

34. A method for preventing the occurrence of an indication or condition associated with expression, such as inappropriate expression of CSA, such as in cancer, multiple sclerosis, arthritis, arthrosis, pathological conditions caused by neural damage, conditions of the cartilage and scar tissue, such as in rheumatism, cartilage repair or wound healing, or in psoriasis; the method comprising administering a therapeutically or prophylactically effective amount of a protein fragment as defined in any one of claims 1-11, a VAR2CSA polypeptide, or a conjugate according to any one of claim 18-24 to a subject in need thereof.

35. A nucleic acid probe capable of hybridizing to a nucleic acid molecule according to any one of claims 14-15 under stringent conditions.

36. Use of a protein fragment as defined in any one of claims 1-11, a VAR2CSA polypeptide, or a conjugate according to any one of claim 18-24, as a biomarker in body fluids, such as blood, plasma, urine, saliva, feces, cerebrospinal fluid, lymph, gastric fluid, pleural fluid, cartilage fluid, sperm, and/or tissue for the diagnosis and/or prognosis of an indication or condition associated with expression, such as inappropriate expression of CSA, such as a malignant disease, arthritis, arthrosis, pathological conditions caused by neural

damage, conditions of the cartilage and scar tissue, such as in rheumatism or wound healing, or a cancer disease, such as brain tumors, liver tumors and tumors in the reproductive tract.

37. Use of a protein fragment as defined in any one of claims 1-11 a VAR2CSA polypeptide, or a conjugate according to claim 24 for the immunization of a subject, such as in a vaccine.

38. Use of a protein fragment as defined in any one of claims 1-11 a VAR2CSA polypeptide, or a conjugate according to claim 24 as a targeting moiety for the isolation of a cell expressing CD44, and/or CSPG4, and/or any other proteoglycan, such as a proteoglycan listed in Table 1.

39. A method for the isolation and/or detection of cells expressing CD44, and/or CSPG4, and/or any other proteoglycan, such as a proteoglycan listed in Table 1, such as cancer stem cells or circulating cancer cells (CTC) in a biological sample, said method comprising:

- a) obtaining a biological sample comprising cells expressing CD44 and/or CSPG4, and/or any other proteoglycan, such as proteoglycans listed in Table 1;
- b) contacting said biological sample with a protein fragment as defined in any one of claims 1-11, a VAR2CSA polypeptide, or a conjugate according to any one of claim 18-24, optionally coupled to a solid support; and
- c) purifying, isolating and/or detecting the complexes of said cells expressing CD44, and/or CSPG4, and/or any other proteoglycan, such as a proteoglycan listed in Table 1 and said protein fragment or conjugate.

40. A diagnostic method for detecting elevated CSA levels in a body fluid, such as blood, plasma, urine, spinal fluid, pleura effusions, joints fluid, bone marrow, semen, sperm, prostate fluid, faeces, gastric fluid, saliva, eye fluid, lung aspirate, and lymph, in response to malignancy, the method comprising the steps of contacting said body fluid with a protein fragment as defined in any one of claims 1-11, a VAR2CSA polypeptide, or a conjugate according to any one of claim 18-24 and detecting the complexes formed with CSA in said body fluid.

41. A method for the purification of CD44, and/or CSPG4, and/or any other proteoglycan, such as a proteoglycan listed in Table 1 in a biological sample, said method comprising:

- a) obtaining a biological sample comprising CD44, and/or CSPG4, and/or any other proteoglycan, such as a proteoglycan listed in Table 1;

- b) contacting said biological sample with a protein fragment as defined in any one of claims 1-11, a VAR2CSA polypeptide, or a conjugate according to any one of claim 18-24, optionally coupled to a solid support; and
- c) purifying or isolating the complexes of said CD44, and/or CSPG4, and/or any other proteoglycan, such as a proteoglycan listed in Table 1 and said protein fragment or conjugate.

42. A protein fragment as defined in any one of claims 1-11, a VAR2CSA polypeptide, or a conjugate according to any one of claim 18-24, or pharmaceutical composition according to claim 28 in combination with any other suitable anticancer agent.