

ALPHA-SYNUCLEIN SUBSTRATES AND METHODS FOR
MAKING AND USING THE SAME

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims priority from U.S. Provisional Patent Application No. 63/026,394, filed on May 18, 2020.

SEQUENCE LISTING

[0001] A Sequence Listing has been submitted electronically in ASCII format and is hereby incorporated by reference in its entirety. The ASCII copy, created on May 18, 2021, is named Amprion-SUBS-AS-PCT.txt and is 39,523 bytes in size.

BACKGROUND

[0002] Certain degenerative brain diseases, collectively termed “synucleinopathies,” involve pathological accumulation of misfolded alpha-synuclein (α S) protein in the brain of affected subjects. A misfolded α S protein is an α S protein having a different structural conformation than it has when involved in its typical, nonpathogenic normal function within a biological system. A misfolded α S protein may aggregate and may exist in or as an aggregate. A misfolded α S protein may localize in an α S protein aggregate. A misfolded α S protein may be a non-functional protein. A misfolded α S protein may be a pathogenic conformer of the α S protein.

[0003] Synucleinopathies include Parkinson’s disease (PD), dementia with Lewy bodies (DLB), and multiple system atrophy (MSA), as well as rare neuroaxonal dystrophies. Several lines of evidence indicate that the process of α S misfolding and aggregation may begin years or decades before the onset of clinical symptoms and substantial brain damage. Thus, detection of α S aggregates and/or misfolded α S protein to facilitate early diagnosis of synucleinopathies could prove crucial to permit intervention before irreversible neuropathological changes occur.

[0004] Unfortunately, soluble, misfolded α S protein is present in such low amounts in bodily fluids that it is very difficult to detect. Recently, however, significant advances have been made in the detection of misfolded α S aggregates (i.e., non-covalent associations of misfolded α S protein), particularly via Seed Amplification Assay (SAA) (formerly known as protein misfolding cyclic amplification (PMCA)). See, e.g., US20160077111A1, US20210063416A1, and U.S. Nonprovisional Patent Application No. 17/154,966, each of which is incorporated by reference herein in its entirety. Briefly, a biological sample (e.g., blood, skin, cerebrospinal fluid, or the like) is contacted with a pre-incubation mixture, the pre-incubation mixture comprising a monomeric α S substrate; a buffer composition; a salt; and an indicator, to form an incubation mixture. Multiple incubation cycles are conducted on the incubation mixture. Each incubation cycle comprises: (1) incubating the incubation mixture effective to cause misfolding and/or aggregation of the monomeric α S substrate in the presence of any soluble, misfolded α S protein present in the biological sample, and (2) physically disrupting the incubation mixture to “de-aggregate,” i.e., break up or disrupt, the misfolded α S aggregates to release smaller aggregates. The α S aggregates may then act as “seeds.” Detection of misfolded α S aggregate via indicator fluorescence indicates the presence of soluble, misfolded α S protein in the biological sample. An example depiction of the α S-SAA process with a biological sample containing soluble, misfolded α S protein is shown in **Figure 1**.

[0005] A significant limitation to α S-SAA technology is the difficulty of producing SAA-competent monomeric α S substrate at a sufficient scale for widespread testing. One approach to producing SAA-competent monomeric α S substrate has been to transform into an enterobacterial host cell (e.g., *Escherichia Coli* (“*E. Coli*”)) an expression vector comprising a nucleic acid sequence coding for human α S protein, such as the plasmid represented by SEQ ID NO: 4; culture the enterobacterial host cell under conditions effective to produce the

monomeric human α S protein; obtain the human α S protein from the enterobacterial host cell; and purify the human α S protein to yield recombinant monomeric α S substrate. However, the resultant monomeric α S substrate has, in some instances, contained “misincorporated” cysteine in place of tyrosine, especially at position 136. “Misincorporation” refers to the process by which certain amino acids, e.g., cysteine, may be unintentionally substituted for other amino acids, e.g., tyrosine, when expressing a nucleic acid for a protein of one organism, e.g., a human nucleic acid encoding human α S protein, in a host cell, e.g., a microorganism such as *E. Coli*. The presence of cysteine residues in monomeric α S substrate can lead to self-aggregation and misfolding by dimerization.

[0006] While aggregation of α S protein is an expected part of the pathology of misfolded protein disorders and is used to advantage in α S-SAA, it is crucial for monomeric α S substrate to self-aggregate as little as possible. Self-aggregation refers to aggregation that would occur even in the absence of soluble, misfolded α S protein in the biological sample. The propensity to self-aggregate is an intrinsic characteristic of α S protein that resides in its amino acid sequence.

[0007] Depending on the relative kinetics and extent, self-aggregation and misfolding of the monomeric α S substrate can prove to be a fatal factor in the use of α S-SAA to amplify and detect misfolded α S aggregate in a biological sample. When self-aggregation of the monomeric α S substrate occurs, the result may be a “false positive”—that is, misfolded α S aggregate is detected notwithstanding that no soluble, misfolded α S protein was present in the biological sample—or an inaccurately high quantitative assessment of soluble, misfolded α S protein in the biological sample.

[0008] Another limitation of existing monomeric α S substrates is the failure of the monomeric α S substrate to aggregate in the presence of soluble, misfolded α S protein in the

biological sample. This can occur, among other times, when the monomeric α S substrate lacks sufficient homology with the soluble, misfolded α S protein. When aggregation of the monomeric α S substrate and soluble, misfolded α S protein in the biological sample do not occur, the result may be a “false negative”—that is, misfolded α S aggregate is not detected notwithstanding that soluble, misfolded α S protein was present in the biological sample—or an inaccurately low quantitative assessment of soluble, misfolded α S protein in the biological sample.

[0009] Thus, self-aggregation and misfolding of the monomeric α S substrate on the one hand, and a failure to aggregate with soluble, misfolded α S protein in the biological sample on the other, can result in misdiagnosis or inaccurate prognosis of tested subjects. Accordingly, there is a need for monomeric α S substrate that, together with proper SAA conditions, reduces, slows, or prevents self-aggregation when used in α S-SAA, yet retains its activity in the presence of soluble, misfolded α S protein in a biological sample.

SUMMARY

[0010] An expression vector is provided for production of human α S protein or a conservative variant thereof (ultimately “monomeric α S substrate”) that exhibits a decreased tendency to self-aggregate in α S-SAA. The expression vector comprises a nucleic acid sequence coding for human α S protein or a conservative variant thereof, the nucleic acid sequence comprising codons that have been optimized to produce human α S protein or a conservative variant when expressed by an enterobacterial host cell such as *E. Coli*. In some aspects, the codons have been optimized to avoid amino acid misincorporation. In some aspects, the codons have been optimized to avoid cysteine misincorporation in the expressed protein. In further aspects, the codons have been optimized to avoid cysteine misincorporation in at least one of positions 39, 125, 133, and 136 of the expressed protein.

[0011] In one aspect, the expression vector may comprise a coding nucleic acid sequence represented by SEQ ID NO: 1 or having at least 90% identity with SEQ ID NO: 1. In one aspect, the expression vector may be a plasmid. For example, the expression vector may be a plasmid that enables the expression of a protein by means of the T7-lac operon system. In one aspect, the expression vector may be a plasmid comprising SEQ ID NO: 1, as represented by SEQ ID NO: 2 or having at least 90% identity with SEQ ID NO: 2. In some aspects, the monomeric α S substrate is represented by SEQ ID NO: 6.

[0012] Also provided is a method for making a purified monomeric human α S substrate or a conservative variant. The method comprises transforming into a host cell an expression vector comprising a nucleic acid sequence coding for monomeric human α S protein or a conservative variant, the nucleic acid sequence comprising codons that have been optimized to avoid amino acid misincorporation in the expressed protein; culturing the host cell under conditions effective to produce the monomeric human α S protein or a conservative variant; obtaining the monomeric human α S protein or a conservative variant from the host cell; and purifying the monomeric human α S protein or a conservative variant by subjecting the human α S protein or a conservative variant to one or more acid precipitation steps at a pH of about 3.5 or less, followed by chromatography, to yield the purified monomeric human α S substrate or a conservative variant. In some aspects, the host cell is an enterobacteria such as *E. Coli*, e.g., *E. Coli* BL21(DE3), BL21(DE3)-pLysS, and the like. In such aspects, the method further comprises eliminating bacterial lipids and/or adding lipopolysaccharide (“LPS”) to the monomeric human α S substrate or a conservative variant after purification.

[0013] Finally, a recombinant enterobacterial cell is provided, the recombinant enterobacterial cell comprising an expression vector for production of human α S protein or a conservative variant, the expression vector comprising a nucleic acid sequence coding for human α S protein or a conservative variant, the nucleic acid sequence comprising codons that

have been optimized to produce human α S protein or a conservative variant without amino acid misincorporation when expressed by the cell.

BRIEF DESCRIPTION OF THE FIGURES

[0014] The present invention may be more readily understood by reference to the following Figures, wherein:

[0015] **Figure 1** is an example depiction of the “fast” α S-SAA process using a biological sample that contains soluble, misfolded α S protein.

[0016] **Figures 2A** and **2B** show gel electrophoresis results of the monomeric α S substrate corresponding to SEQ ID NO: 6 expressed in *E. coli* strain BL21(DE3) transformed using the plasmid represented by SEQ ID NO: 2, wherein **Figure 2A** shows the results upon treatment with dithiothreitol (“DTT”), and **Figure 2B** shows the results without treatment.

[0017] **Figure 3** shows gel electrophoresis results of the substrate expressed in *E. coli* strain BL21(DE3) transformed using the plasmid represented by SEQ ID NO: 4 before and after treatment with DTT and filtrations with 30 or 50kDa filters.

[0018] **Figure 4A** and **Figure 4B** show α S-SAA “fast assay” aggregation curves (of three replicates individually) using the monomeric α S substrate corresponding to SEQ ID NO: 6 expressed in *E. coli* strain BL21(DE3) transformed using the plasmid represented by SEQ ID NO: 2, purified by acid precipitation to pH 4, in the presence of a confirmed PD sample (**Figure 4A**) and in a healthy control (HC) (**Figure 4B**).

[0019] **Figure 5A** and **Figure 5B** show α S-SAA “fast assay” aggregation curves (of three replicates individually) using the monomeric α S substrate corresponding to SEQ ID NO: 6 expressed in *E. coli* strain BL21(DE3) transformed using the plasmid represented by SEQ ID NO: 2, purified by acid precipitation to pH 3.5, in the presence of a confirmed PD sample (**Figure 5A**) and in an HC (**Figure 5B**).

[0020] **Figure 6A** and **Figure 6B** show α S-SAA “fast assay” aggregation curves (of three replicates individually) using the monomeric α S substrate corresponding to SEQ ID NO: 6 expressed in *E. coli* strain BL21(DE3) transformed using the plasmid represented by SEQ ID NO: 2, purified by acid precipitation to pH 3, in the presence of a confirmed PD sample (**Figure 6A**) and in an HC (**Figure 6B**).

[0021] **Figure 7A** and **Figure 7B** show α S-SAA “fast assay” aggregation curves (of three replicates individually) using the monomeric α S substrate corresponding to SEQ ID NO: 6 expressed in *E. coli* strain BL21(DE3) transformed using the plasmid represented by SEQ ID NO: 2, purified by acid precipitation to pH 2.5, in the presence of a confirmed PD sample (**Figure 7A**) and in an HC (**Figure 7B**).

[0022] **Figure 8A** and **Figure 8B** show α S-SAA “fast assay” aggregation curves (of three replicates individually) using the monomeric α S substrate corresponding to SEQ ID NO: 6 expressed in *E. coli* strain BL21(DE3) transformed using the plasmid represented by SEQ ID NO: 2, purified by acid precipitation to pH 2.0, in the presence of a confirmed PD sample (**Figure 8A**) and in an HC (**Figure 8B**).

[0023] **Figures 9A-9D** show α S-SAA “fast assay” aggregation curves (of three replicates individually) using the monomeric α S substrate corresponding to SEQ ID NO: 6 expressed in *E. coli* strain BL21(DE3) transformed using the plasmid represented by SEQ ID NO: 2, purified by acid precipitation to pH 2.5 and further supplemented with varying amounts of LPS, in the presence of confirmed PD samples (**Figures 9A** and **9C**) and in HCs (**Figure 9B** and **9D**).

[0024] **Figure 10A** and **Figure 10B** show α S-SAA “fast assay” aggregation curves (of two replicates individually) using the monomeric α S substrate corresponding to SEQ ID NO: 6 expressed in *E. coli* strain BL21(DE3) transformed using the plasmid represented by SEQ ID NO: 2, purified by acid precipitation to pH 2.5 and further supplemented with varying amounts of LPS, in the presence of confirmed MSA samples.

[0025] **Figure 11** shows a flowchart of an example method for purifying monomeric α S substrate that, together with proper SAA conditions, reduces, slows, or prevents altogether misfolding and self-aggregation when used in an α S-SAA assay, yet retains its activity in the presence of soluble, misfolded α S protein in a biological sample.

[0026] **Figures 12A-12D** show α S-SAA “fast assay” aggregation curves (of three replicates individually) using the monomeric α S substrate corresponding to SEQ ID NO: 6 expressed in *E. coli* strain BL21(DE3) transformed using the plasmid represented by SEQ ID NO: 2, purified by two acid precipitation steps – first at pH 3.5 and again at pH 2.0, as depicted in **Figure 11** – in the presence of confirmed PD samples (**Figures 12A-12C**) and in an HC (**Figure 12D**).

[0027] **Figures 13A-13D** show α S-SAA “fast assay” aggregation curves (of three replicates individually) using the monomeric α S substrate corresponding to SEQ ID NO: 6 expressed in *E. coli* strain BL21(DE3) transformed using the plasmid represented by SEQ ID NO: 2, purified by acid precipitation to pH about 3.1, and further purified by a second filtration of the dialyzed, filtered protein using a 50kDa filter in the presence of synthetic seeds (**Figure 13A**) and in an HC (**Figure 13B**) or using a 30kDa filter in the presence of synthetic seeds (**Figure 13C**) and in an HC (**Figure 13D**).

DETAILED DESCRIPTION

α S Expression Vector

[0028] An expression vector is provided for production of monomeric α S protein that, properly purified, exhibits a decreased tendency to self-aggregate in α S-SAA. The expression vector comprises a nucleic acid sequence coding for human α S protein or a conservative variant, the nucleic acid sequence comprising codons that have been optimized to produce human α S protein or a conservative variant when expressed by a suitable host cell. In one

aspect, the host cell is an enterobacterial host cell such as *E. Coli*. In another aspect the host cell is an S2 insect cell, a yeast cell, *Saccharomyces cerevisiae*, or *Pichia pastoris*. In some aspects, the codons have been optimized to avoid amino acid incorporation. In some aspects, the codons have been optimized to avoid cysteine misincorporation in the expressed protein. In further aspects, the codons have been optimized to avoid cysteine misincorporation in at least one of positions 39, 125, 133, and 136 of the expressed protein.

[0029] In one aspect, the expression vector may comprise a coding nucleic acid sequence represented by SEQ ID NO: 1 or having at least 90% identity with SEQ ID NO: 1. In one aspect, the expression vector may be a plasmid. For example, the expression vector may be a plasmid that enables the expression of a protein by means of the T7-lac operon system. In one aspect, the expression vector may be a plasmid comprising SEQ ID NO: 1, as represented by SEQ ID NO: 2 or having at least 90% identity with SEQ ID NO: 2.

[0030] Some amino acids can be coded for by more than one codon. A natural hierarchy exists for certain codons to be used in certain types of cells. Thus, a nucleic acid sequence configured for host cell expression of human α S protein may be different from a corresponding nucleic acid sequence that expresses human α S protein in human cells. For example, a nucleic acid sequence configured for expression of human α S protein in a microorganism such as *E. Coli*, may include, e.g., substitution of bacteria-typical codons for human-typical codons for selected amino acids. For example, the nucleic acid sequence configured for expression of human α S protein in a microorganism such as *E. Coli* may include a TAT codon expressing tyrosine rather than a TAC codon, which can lead to cysteine misincorporation.

[0031] SEQ ID NO: 1, upon expression in *E. Coli*, mitigates misincorporation of, among other amino acids, cysteine for tyrosine at one or more of positions 39, 125, 133, and 136 in SEQ ID NO: 6 compared to expression of human native nucleic acid (e.g., via an expression vector comprising a coding nucleic acid sequence comprising SEQ ID NO: 3, as represented

by SEQ ID NO: 4) in *E. Coli*. Cysteine misincorporation, as opposed to misincorporation of other amino acids, is detectable because of the formation of dimers (see **Figure 3**). Upon expression in the recombinant host cell, the α S protein (SEQ ID NO: 6) being produced has substantially decreased or no cysteine misincorporation at one or more of positions 39, 125, 133, and 136.

[0032] In various aspects, the expression vector may be operatively linked to regulatory sequences effective for expression of an optimized nucleic acid sequence (e.g., SEQ ID NO: 1) in the enterobacterial host cell. The term “operably linked” refers to the arrangement of various polynucleotide elements relative to each other such that the elements are functionally connected and can interact with each other. Such elements may include, without limitation, a promoter, an enhancer, a polyadenylation sequence, one or more introns and/or exons, and a coding sequence of a gene of interest to be expressed. The expression vector may be operatively linked to regulatory sequences effective for expression of the nucleic acid sequence represented by SEQ ID NO: 1 in *E. Coli*. Many suitable vectors are available. The vector components generally include, but are not limited to, one or more of the following: an origin of replication, one or more marker genes, an enhancer element, a promoter, and a transcription termination sequence.

[0033] The α S protein may be recombinantly produced without modification or may be produced as a fusion polypeptide with a heterologous polypeptide, e.g., a signal sequence or other polypeptide having a specific cleavage site at the N-terminus of the mature protein or polypeptide. In general, the signal sequence may be a component of the vector, or it may be a part of the coding sequence that is inserted into the vector. In one aspect, the heterologous signal sequence selected may be one that is recognized and processed (i.e., cleaved by a signal peptidase) by the host cell.

[0034] Expression vectors usually contain a selection gene, also termed a selectable marker. The selection gene encodes a protein necessary for the survival or growth of transformed host cells grown in a selective culture medium. Host cells not transformed with the vector containing the selection gene will not survive in the culture medium. Typical selection genes encode proteins that: (a) confer resistance to antibiotics or other toxins, e.g., ampicillin, neomycin, methotrexate, or tetracycline; (b) complement auxotrophic deficiencies; or (c) supply critical nutrients not available from complex media.

[0035] Expression vectors will contain a promoter that is recognized by the host organism and is operably linked to an orthogonal protein coding sequence. Promoters are untranslated sequences located upstream (5') to the start codon of a structural gene (generally within about 100 to 1000 bp) that control the transcription of the particular nucleic acid sequence to which they are operably linked. Such promoters typically fall into two classes: inducible and constitutive. Inducible promoters are promoters that initiate increased levels of transcription from DNA under their control in response to some change in culture conditions, e.g., the presence or absence of a nutrient or a change in temperature. Many promoters recognized by a variety of potential host cells are well known.

[0036] In one aspect, the expression vector may comprise a coding nucleic acid sequence represented by SEQ ID NO: 1 or a nucleic acid sequence having a sequence identity percentage to SEQ ID NO: 1 of at least about 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99. In some aspects, the nucleic acid sequence comprises a sequence having at least 95% identity with SEQ ID NO: 1. The nucleic acid sequence is not SEQ ID NO: 3.

[0037] The expression vector may be a plasmid. For example, the expression vector may be a plasmid represented by SEQ ID NO: 2. In some aspects, the plasmid nucleic acid sequence comprises a sequence having at least 95% identity with SEQ ID NO: 2. The plasmid nucleic acid sequence is not SEQ ID NO: 4.

[0038] To be clear, SEQ ID NO: 1 is the optimized DNA sequence that encodes for human α S with a C-terminal histag. SEQ ID NO: 1 is a part of SEQ ID NO: 2 (from 1459 to 1899bp of SEQ ID NO: 2). SEQ ID NO: 2 is the DNA sequence of the whole vector or plasmid including SEQ ID NO: 1. SEQ ID NO: 3 is the non-optimized DNA sequence that encodes for human α S with a C-terminal histag. SEQ ID NO: 3 is part of SEQ ID NO: 4 (from 1459 to 1899bp of SEQ ID NO: 4). SEQ ID NO: 4 is the DNA sequence of the whole vector or plasmid including SEQ ID NO: 3.

[0039] Due to the degenerate nature of the genetic code, a variety of different nucleotide sequences can be used to encode a given polypeptide. The present disclosure includes DNA compounds of any sequence that encode the amino acid sequences of the polypeptides described herein. In similar fashion, a polypeptide can typically tolerate one or more amino acid substitutions, deletions, and insertions in its amino acid sequence without loss or significant loss of a desired activity.

[0040] Nucleotide identity is determined by aligning the residues of the two polynucleotides to optimize the number of identical nucleotides along the lengths of their sequences. Gaps in either or both sequences are permitted in making the alignment in order to optimize the number of shared nucleotides, although the nucleotides in each sequence must nonetheless remain in their proper order. Preferably, two nucleotide sequences are compared using the Blastn program of the BLAST 2 search algorithm, as described by Tatusova, et al. (FEMS Microbiology Letters, 174, p. 247-50 (1999)), and available on the world wide web at the National Center for Biotechnology Information website, under BLAST in the Molecular Database section. Preferably, the default values for all BLAST 2 search parameters are used, including reward for match=1, penalty for mismatch=-2, open gap penalty=5, extension gap penalty=2, gap x dropoff=50, expect=10, wordsize=11, and optionally, filter on. In the

comparison of two nucleotide sequences using the BLAST search algorithm, nucleotide identity is referred to as “identities.”

Recombinant α S Protein-Producing Cells

[0041] A recombinant enterobacterial cell is also provided, wherein the cell comprises an expression vector for production of human α S protein or a conservative variant. The expression vector may comprise a nucleic acid sequence coding for human α S protein or a conservative variant, the nucleic acid sequence comprising codons that have been optimized to produce human α S protein or a conservative variant when expressed by the cell.

[0042] Enterobacteria, more formally known as *Enterobacteriaceae*, are a family of Gram-negative bacteria. Members of the *Enterobacteriaceae* family are bacilli (rod-shaped), are typically 1–5 μ m in length, and usually include flagella for movement. Examples of enterobacteria include *E. Coli*, Salmonella, Klebsiella, Shigella, Enterobacter, and Citrobacter.

[0043] In some aspects, the nucleic acid sequence coding for human α S protein or a conservative variant comprises codons that have been optimized to produce human α S protein or a conservative variant when expressed by an *E. Coli* host cell, e.g., *E. Coli* BL21(DE3), BL21(DE3)-pLysS, and the like. *E. coli* needs to express the gene for T7 RNA polymerase under control of a lacUV5 promoter, allowing expression of the T7 RNA polymerase to be induced with IPTG or by autoinduction. Once T7 RNA polymerase is expressed, it enables the transcription of SEQ ID NO: 1 in the plasmid. BL21(DE3) cells have the required phenotype.

[0044] In various aspects, the expression vector included in the enterobacterial cell may include any of the features for an expression vector described herein. For example, in some aspects, the expression vector comprises a nucleic acid sequence having at least 95% identity with SEQ ID NO: 1. In further aspects, the recombinant α S protein expressed in the recombinant host cell is characterized by the absence of or mitigation of amino acid

misincorporation, including cysteine misincorporation at one or more of positions 39, 125, 133, and 136.

[0045] The term “recombinant host cell” (or simply “host cell”) refers to a cell (the particular subject cell and the progeny of such a cell) into which a recombinant vector has been introduced. Because certain modifications may occur in succeeding generations due either to mutation or environmental influences, the progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term “host cell.” A recombinant host cell (e.g., a recombinant enterobacterial cell) may be an isolated cell, a cell line grown in culture, or may be a cell that resides in a living tissue or organism.

Methods for Preparing Monomeric α S Substrate

[0046] A method for preparing monomeric α S substrate for use in α S-SAA is also provided. The method comprises providing an enterobacterial host cell comprising a nucleic acid sequence coding for human α S protein or a conservative variant, the nucleic acid sequence comprising codons that have been optimized to produce human α S protein or a conservative variant; culturing the enterobacterial host cell under conditions effective to produce the human α S protein or a conservative variant; and obtaining the human α S protein or a conservative variant from the enterobacterial host cell. In some aspects, the method provides analogs or peptide fragments of human α S.

[0047] The phrases “monomeric α S protein” and “monomeric α S substrate” are used interchangeably and refer to one or more α S protein molecules or a conservative variant in their native, nonpathogenic configuration. In some aspects, the monomeric α S substrate comprises, consists essentially of, or consists of wildtype or recombinant human α S protein having 140 amino acids, having a molecular mass of 14,460 Da, and being represented by the sequence:

[0048] SEQ ID NO: 5:

MDVFMKGLSK AKEGVVAAAE KTKQGVAEAA GKTKEGVLYV GSKTKEGVVH
GVATVAEKTQ EQVTNVGGAV VTGVTAVAQK TVEGAGSIAA ATGFVKKDQL
GKNEEGAPQE GILEDMPVDP DNEAYEMPSE EGYQDYEPEA

[0049] In some aspects, the monomeric α S protein comprises, consists essentially of, or consists of a conservative variant of SEQ ID NO: 5. A conservative variant may be a peptide or amino acid sequence that deviates from SEQ ID NO: 5 only in the substitution or addition of one or several amino acids for amino acids having similar biochemical properties and having a minimal or beneficial impact on the activity of the resultant protein in the α S-SAA. A conservative variant must functionally perform substantially like the base component, i.e., SEQ ID NO: 5. For example, a conservative variant of SEQ ID NO: 5 will aggregate with misfolded α S protein and will form aggregates with substantially similar reaction kinetics as SEQ ID NO: 5 under similar reaction conditions.

[0050] Generally, a conservative variant (of SEQ ID NO: 5 or of any of the SEQ ID NOs disclosed herein) may have for example, one, two, three, four, five, six, seven (5%), and up to 14 (10%) substitutions, additions, or deletions in the amino acid sequence. In some aspects, the conservative variant of SEQ ID NO: 5 may include α S protein of other mammalian species, such as, for example, rodents and non-human primates. In some aspects, the conservative variant of SEQ ID NOs: 6-23 may include similarly tagged α S proteins of other mammalian species, such as, for example, rodents and non-human primates (that is, the variations is/are within the 140 amino acid sequence of the α S protein). In some aspects, the invention excludes SEQ ID NO: 5 as the monomeric α S substrate.

[0051] In some aspects, the monomeric α S substrate comprises a recombinant α S protein comprising six additional histidine amino acids (i.e., a polyHis purification tag) on the C-terminus of SEQ ID NO: 5, resulting in a molecular mass of 15,283 Da and being represented by the sequence:

[0052] SEQ ID NO: 6:

MDVFMKGLSK AKEGVVAAAE KTKQGVAEAA GKTKEGVLYV GSKTKEGVVH
GVATVAEKT EQVTNVGGAV VTGVTAVAQK TVEGAGSIAA ATGFVKKDQL
GKNEEGAPQE GILEDMPVDP DNEAYEMPSE EGYQDYEPEA HHHHHH

[0053] Thus, SEQ ID NO: 6 is distinguishable from SEQ ID NO: 5 by the six additional histidine amino acids on the C-terminus. The retention of the histidine tag may improve the ability of the human monomeric α S substrate to avoid self-aggregation. SEQ ID NO: 6 is further distinguishable from, e.g., a conservative variant of SEQ ID NO: 5 (and SEQ ID NO: 6) wherein one or more amino acids are added to the N-terminus. In some aspects, a conservative variant of SEQ ID NO: 5 wherein one or more amino acids are added to the N-terminus is excluded. However, some aspects include N-terminus additions. Thus:

[0054] SEQ ID NO: 7:

HHHHHH MDVFMKGLSK AKEGVVAAAE KTKQGVAEAA GKTKEGVLYV
GSKTKEGVVH GVATVAEKT EQVTNVGGAV VTGVTAVAQK TVEGAGSIAA
ATGFVKKDQL GKNEEGAPQE GILEDMPVDP DNEAYEMPSE EGYQDYEPEA

[0055] Additional purification tags are contemplated, including, e.g., FLAG, HA, Myc, and V5, thus generating the following SEQ ID NOs:

[0056] SEQ ID NO: 8:

MDVFMKGLSK AKEGVVAAAE KTKQGVAEAA GKTKEGVLYV GSKTKEGVVH
GVATVAEKT EQVTNVGGAV VTGVTAVAQK TVEGAGSIAA ATGFVKKDQL
GKNEEGAPQE GILEDMPVDP DNEAYEMPSE EGYQDYEPEA DYKDDDD

[0057] SEQ ID NO: 9:

DYKDDDD MDVFMKGLSK AKEGVVAAAE KTKQGVAEAA GKTKEGVLYV
GSKTKEGVVH GVATVAEKT EQVTNVGGAV VTGVTAVAQK TVEGAGSIAA
ATGFVKKDQL GKNEEGAPQE GILEDMPVDP DNEAYEMPSE EGYQDYEPEA

[0058] SEQ ID NO: 10:

MDVFMKGLSK AKEGVVAAAE KTKQGVAEAA GKTKEGVLYV GSKTKEGVVH
GVATVAEKT EQVTNVGGAV VTGVTAVAQK TVEGAGSIAA ATGFVKKDQL
GKNEEGAPQE GILEDMPVDP DNEAYEMPSE EGYQDYEPEA DYKDDDDK

[0059] SEQ ID NO: 11:

DYKDDDDK MDVFMKGLSK AKEGVVAAAE KTKQGVAEAA GKTKEGVLYV
GSKTKEGVVH GVATVAEKT EQVTNVGGAV VTGVTAVAQK TVEGAGSIAA
ATGFVKKDQL GKNEEGAPQE GILEDMPVDP DNEAYEMPSE EGYQDYEPEA

[0060] SEQ ID NO: 12:

MDVFMKGLSK AKEGVVAAAE KTKQGVAEAA GKTKEGVLYV GSKTKEGVVH
GVATVAEKT EQVTNVGGAV VTGVTAVAQK TVEGAGSIAA ATGFVKKDQL
GKNEEGAPQE GILEDMPVDP DNEAYEMPSE EGYQDYEPEA DYKDDDK

[0061] SEQ ID NO: 13:

DYKDDDK MDVFMKGLSK AKEGVVAAAE KTKQGVAEAA GKTKEGVLYV
GSKTKEGVVH GVATVAEKT EQVTNVGGAV VTGVTAVAQK TVEGAGSIAA
ATGFVKKDQL GKNEEGAPQE GILEDMPVDP DNEAYEMPSE EGYQDYEPEA

[0062] SEQ ID NO: 14:

MDVFMKGLSK AKEGVVAAAE KTKQGVAEAA GKTKEGVLYV GSKTKEGVVH
GVATVAEKT EQVTNVGGAV VTGVTAVAQK TVEGAGSIAA ATGFVKKDQL
GKNEEGAPQE GILEDMPVDP DNEAYEMPSE EGYQDYEPEA YPYDVDPDYA

[0063] SEQ ID NO: 15:

YPYDVDPDYA MDVFMKGLSK AKEGVVAAAE KTKQGVAEAA GKTKEGVLYV
GSKTKEGVVH GVATVAEKT EQVTNVGGAV VTGVTAVAQK TVEGAGSIAA
ATGFVKKDQL GKNEEGAPQE GILEDMPVDP DNEAYEMPSE EGYQDYEPEA

[0064] SEQ ID NO: 16:

MDVFMKGLSK AKEGVVAAAE KTKQGVAEAA GKTKEGVLYV GSKTKEGVVH
GVATVAEKT EQVTNVGGAV VTGVTAVAQK TVEGAGSIAA ATGFVKKDQL
GKNEEGAPQE GILEDMPVDP DNEAYEMPSE EGYQDYEPEA YAYDVPDYA

[0065] SEQ ID NO: 17:

YAYDVPDYA MDVFMKGLSK AKEGVVAAAE KTKQGVAEAA GKTKEGVLYV
GSKTKEGVVH GVATVAEKT EQVTNVGGAV VTGVTAVAQK TVEGAGSIAA
ATGFVKKDQL GKNEEGAPQE GILEDMPVDP DNEAYEMPSE EGYQDYEPEA

[0066] SEQ ID NO: 18:

MDVFMKGLSK AKEGVVAAAE KTKQGVAEAA GKTKEGVLYV GSKTKEGVVH
GVATVAEKT EQVTNVGGAV VTGVTAVAQK TVEGAGSIAA ATGFVKKDQL
GKNEEGAPQE GILEDMPVDP DNEAYEMPSE EGYQDYEPEA YDVPDYASL

[0067] SEQ ID NO: 19:

YDVPDYASL MDVFMKGLSK AKEGVVAAAE KTKQGVAEAA GKTKEGVLYV
GSKTKEGVVH GVATVAEKT EQVTNVGGAV VTGVTAVAQK TVEGAGSIAA
ATGFVKKDQL GKNEEGAPQE GILEDMPVDP DNEAYEMPSE EGYQDYEPEA

[0068] SEQ ID NO: 20:

MDVFMKGLSK AKEGVVAAAE KTKQGVAEAA GKTKEGVLYV GSKTKEGVVH
GVATVAEKT EQVTNVGGAV VTGVTAVAQK TVEGAGSIAA ATGFVKKDQL
GKNEEGAPQE GILEDMPVDP DNEAYEMPSE EGYQDYEPEA EQKLISEEDL

[0069] SEQ ID NO: 21:

EQKLISEEDL MDVFMKGLSK AKEGVVAAAE KTKQGVAEAA GKTKEGVLYV
GSKTKEGVVH GVATVAEKT EQVTNVGGAV VTGVTAVAQK TVEGAGSIAA
ATGFVKKDQL GKNEEGAPQE GILEDMPVDP DNEAYEMPSE EGYQDYEPEA

[0070] SEQ ID NO: 22:

MDVFMKGLSK AKEGVVAAAE KTKQGVAEAA GKTKEGVLYV GSKTKEGVVH
 GVATVAEKT EQVTNVGGAV VTGVTAVAQK TVEGAGSIAA ATGFVKKDQL
 GKNEEGAPQE GILEDMPVDP DNEAYEMPSE EGYQDYEPEA GKPIPNNLLGLDST

[0071] SEQ ID NO. 23:

GKPIPNNLLGLDST MDVFMKGLSK AKEGVVAAAE KTKQGVAEAA GKTKEGVLYV
 GSKTKEGVVH GVATVAEKT EQVTNVGGAV VTGVTAVAQK TVEGAGSIAA
 ATGFVKKDQL GKNEEGAPQE GILEDMPVDP DNEAYEMPSE EGYQDYEPEA

[0072] The preparation of monomeric α S substrate, polypeptide fragments thereof, mutants, truncations, derivatives, and splice variants that display substantially equivalent or altered α S activity relative to the wild-type protein are likewise contemplated. These variants may be deliberate, for example, such as modifications obtained through site-directed mutagenesis, or may be accidental, such as those obtained through mutations in hosts that are producers of the α -synuclein protein. Included within the scope of these terms are α S proteins specifically recited herein, and all substantially homologous analogs and allelic variants thereof.

[0073] Analogs may be made through conservative amino acid substitution. A “conservative amino acid substitution” may include, for example, one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains include amino acids with basic side chains (e.g., lysine, arginine, histidine), acidic side chains (e.g., aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (e.g., threonine, valine, isoleucine), and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine).

[0074] A “non-essential” amino acid residue is a residue that can be altered from the wild-type sequence of α S without abolishing or, more preferably, without substantially altering α S’s biological activity. Thus, a predicted non-essential amino acid residue in a α S protein is preferably replaced with another amino acid residue from the same side chain family. Alternatively, in another aspect, mutations can be introduced randomly along all or part of the α S coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for α S biological activity to identify mutants that retain activity. Following mutagenesis of the nucleotide sequence for α S, the encoded protein can be expressed recombinantly, and the activity of the protein can be determined.

[0075] The monomeric α S substrates produced herein exhibit a decreased tendency to self-aggregate. In some aspects, the monomeric α S substrate does not self-aggregate under α S-SAA conditions. In other aspects, the monomeric α S substrate self-aggregates at a much lower level and/or at a much slower rate than the level or rate of self-aggregation of monomeric α S substrate obtained by previous methods of preparing monomeric α S substrate. In such aspects, the rate and/or level of self-aggregation can be 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 60%, 70%, 80%, or 90% or less compared to the level of aggregation of monomeric α S substrate obtained using methods other than those described herein. In other aspects, the monomeric α S substrate self-aggregates at a slower rate or lower level than the rate or level of aggregation of monomeric α S substrate with soluble, misfolded α S protein. In such examples, α S-SAA detection is still successful because a synucleinopathy-positive sample may be distinguished from mere self-aggregation of monomeric α S substrate by the intensity of the detection (relative fluorescence units) or the time period that fluorescence increase begins. The decreased self-aggregation for monomeric α S substrate described herein can be a result of

differences in the method used to prepare the monomeric α S substrate, including differences in the resulting monomeric α S substrate composition.

[0076] The method of preparing human monomeric α S substrate or a conservative variant may comprise culturing a host cell comprising a nucleic acid sequence coding for human α S protein or a conservative variant, the nucleic acid sequence comprising codons that have been optimized to produce human α S protein or a conservative variant. Cells may be cultured in conventional nutrient media modified as appropriate for inducing promoters, selecting transformants, or amplifying the genes encoding the desired sequences. The culturing media may include buffers, nucleosides (such as adenosine and thymidine), antibiotics, trace elements, and glucose or an equivalent energy source. Any other necessary supplements may also be included at appropriate concentrations that would be known to those skilled in the art. The culture conditions, such as temperature, pH, and the like, are those previously used with the host cell selected for expression and will be apparent to the ordinarily skilled artisan. For a discussion of strategies for producing recombinant proteins using *E. Coli*, see Gopal G., Kumar A., Protein J., 32(6):419-25 (2013), the disclosure of which is incorporated herein by reference in its entirety.

[0077] The method of preparing monomeric α S substrate may include any of the expression vectors described herein. In some aspects, the enterobacteria is *E. Coli*. In further aspects, the nucleic acid sequence of the expression vector comprises a sequence having at least 95% identity with SEQ ID NO: 1. In yet further aspects, the expression vector is a plasmid comprising a sequence having at least 95% identity with SEQ ID NO: 2. In additional aspects, the codons have been optimized to avoid amino acid incorporation, including cysteine misincorporation, in the expressed human α S protein or a conservative variant.

[0078] Once cells have been allowed to grow in culture for an appropriate amount of time (e.g., 4-24 hours), the cells are lysed and monomeric α S protein is purified from the lysed cells.

A variety of methods may be used to lyse cells, such as use of a French press, sonication, freeze-thawing, chemical lysis, and microfluidizing. Purification may include a variety of different purification steps, such as centrifugation, column purification, dialysis, and acid precipitation to a pH of about 3.5 or less, optionally followed by addition of LPS and/or ultrafiltration (10kDa to 300kDa).

[0079] In some aspects, the method of preparing monomeric α S substrate from the host cell comprises lysing the cells using a microfluidizer. Microfluidizers break cells with high efficiency while maintaining intracellular content integrity by supplying constant, controlled shear rates, resulting in large cell membrane fragments that facilitate subsequent protein purification. Use of a microfluidizer for cell lysis can decrease the tendency of purified monomeric α S substrate to self-aggregate. An example of a suitable microfluidizer is the LM20 Microfluidizer® High Shear Fluid Homogenizer manufactured by Microfluidics.

[0080] The method of preparing monomeric α S substrate may comprise the step of separating the monomeric α S substrate from other components of the host cell such as lipids. In some aspects, obtaining the monomeric α S substrate from the host cell (e.g., *E. Coli*) comprises contacting the monomeric α S substrate with Lipid Removal Agent (“LRA”) to remove lipid contaminants (i.e., cell components). Thus, the monomeric α S substrate mixed with various other cell components is contacted with LRA after cell lysis, and the monomeric α S substrate is removed by centrifugation, which separates the protein from the lipids (the LRA and bound lipids go to the pellet fraction during centrifugation). Use of LRA for removal of the undesired lipid components may improve the ability of recombinant monomeric α S substrate in the resulting composition to avoid self-aggregation. LRA is a commercially available agent (available from Millipore Sigma) based on synthetic calcium silicate hydrate.

[0081] The method for preparing monomeric α S substrate or a monomeric α S substrate composition may comprise the step of initially separating the monomeric α S substrate from

lipids, such as LPS, or nucleic acids, such as DNA and RNA. In some aspects, obtaining the monomeric α S substrate from the host cell (e.g., *E. Coli*) comprises precipitation of non-synuclein components by addition of hydrochloric acid (HCl) to less than pH 3.50, including about pH 2.0 or lower. In some aspects, the method comprises adding LPS to the acid precipitated monomeric α S substrate.

[0082] In one aspect, obtaining the monomeric α S substrate from the host cell comprises precipitation of non-synuclein components by subjecting the human α S protein or a conservative variant to one or more acid precipitation steps at a pH of about 3.5 or less, e.g., first at pH 3.5 and again at pH 2, followed by chromatography, to yield the purified monomeric α S substrate or a conservative variant. An example of such an aspect is depicted by the flow chart shown in **Figure 11**.

[0083] Options to remove contaminants such as metal-binding proteins (e.g., Ferric Uptake Regulatory Protein; FUR) from the monomeric α S substrate composition include iron-IMAC, antibody depletion, genomic modification of endogenous *E. coli* FUR to include a purification tag (other than histag or other purification tag used by the α S protein), iron saturation to reduce binding to the nickel column, or elimination washes with Fe⁺⁺ during nickel-IMAC. In some aspects, purification of the monomeric α S substrate comprises excluding essentially all other proteins (e.g., metal-binding proteins). In further aspects, the metal-binding protein is a ferric uptake regulator (FUR).

α S Protein Compositions

[0084] The isolated human monomeric α S substrate or conservative variant composition can also include a suitable medium for suspending and/or storing the protein. For example, in some aspects, the monomeric α S substrate or conservative variant composition includes a buffer such as piperazine-N,N'-bis(2-ethanesulfonic acid) (PIPES) or phosphate buffered saline

(PBS). In further aspects, the composition consists essentially of one of SEQ ID NOs: 5-23 and PIPES. The monomeric α S substrate may be substantially free of other materials such as lipid contaminant from the lysed enterobacterial cell.

[0085] In some aspects, the human monomeric α S substrate or conservative variant composition is purified so that it is essentially free of contaminants. Most of the contaminants found in the purified monomeric α S substrate or conservative variant are proteins of host cell origin, with only a few potential contaminants of human origin. MS/MS identifies proteins by detecting fragments after enzymatic digestion and comparing the fragment footprint to the footprints of known proteins, which are available in public databases. Thus, some fragments may match microorganisms other than the host cell (e.g., *E. coli*). However, most prokaryotic contaminating proteins will be from the host cell.

[0086] The resulting relative abundance of all human-originated proteins is extremely low compared to monomeric α S substrate. The most represented contaminants of human origin are cytochrome B5 and keratin related peptides.

[0087] As for the contaminants with bacterial origins, most are from the enterobacterial host cell (e.g., *E. coli*), which is the expression host. By far, the most likely and abundant bacterial contaminant appears to be FUR from *E. coli*. This protein has affinity for iron and is responsible for controlling iron's intracellular concentration. When iron is bound to FUR, the iron becomes capable of binding DNA and can act as a regulator. Since iron and nickel are both divalent cations, FUR's ability to bind iron is potentially the reason why FUR co-purifies with monomeric α S substrate during IMAC purification. α S-SAA competent monomeric α S substrate or conservative variant may contain lower concentration of FUR protein than self-aggregating substrate. The potential mechanism by which FUR induces self-aggregation could include ionic interaction between monomeric α S substrate and residual iron or nickel carried

by FUR. The molecular weight of FUR is practically the same as α S, which may hide the contaminant in molecular weight based discriminating tools.

[0088] Some other notable contaminants from *E. coli* with relatively high abundance and matching scores are catabolite gene activator, HIT-like protein, FKBP-type peptidyl-prolyl cis-trans isomerase, and several ribosomal proteins. Catabolite gene activator is another DNA binding protein like FUR, but it does not require a metal cofactor. HIT proteins are proteins with histidine triad motifs that could be involved in binding zinc and have been shown to be able to bind nucleotides. Therefore, the triad motifs might also have affinity for nickel, which could make HIT proteins co-purify with monomeric α S substrate.

[0089] Accordingly, in some aspects, the human monomeric α S substrate or conservative variant composition is essentially free of other proteins. In some aspects, the other proteins comprise metal-binding proteins, while in further aspects the metal-binding proteins comprise FUR. Metal-binding proteins include proteins that bind (e.g., chelate) to metal ions such as sodium, potassium, magnesium, calcium, manganese, iron, cobalt, zinc, nickel, vanadium, molybdenum, and tungsten.

[0090] Potential contaminants of the human monomeric α S substrate or conservative variant composition, which are excluded from the composition in aspects of the invention, may include: Ferric uptake regulation protein OS=Escherichia coli; Catabolite gene activator OS=Escherichia coli; 30S ribosomal protein S12 OS=Escherichia coli; HIT-like protein hinT OS=Escherichia coli; FKBP-type peptidyl-prolyl cis-trans isomerase slyD OS=Escherichia coli; Bifunctional polymyxin resistance protein ArnA OS=Escherichia coli; 50S ribosomal protein L27 OS=Escherichia coli; Formyltetrahydrofolate deformylase OS=Escherichia coli; 30S ribosomal protein S15 OS=Escherichia coli; Glucosamine--fructose-6-phosphate aminotransferase [isomerizing] OS=Escherichia coli; Regulator of sigma D OS=Escherichia coli; Acyl-[acyl-carrier-protein]--UDP-N-acetylglucosamine O-acyltransferase

OS=Escherichia coli; UPF0047 protein yjbQ OS=Escherichia coli; Arabinose 5-phosphate isomerase GutQ OS=Escherichia coli; 50S ribosomal protein L28 OS=Escherichia coli; 30S ribosomal protein S16 OS=Parabacteroides distasonis; 30S ribosomal protein S20 OS=Escherichia coli; Arabinose 5-phosphate isomerase KdsD OS=Escherichia coli; 30S ribosomal protein S2 OS=Escherichia coli; 50S ribosomal protein L7/L12 OS=Kineococcus radiotolerans; Ribosomal RNA large subunit methyltransferase A OS=Escherichia coli; Nucleoside diphosphate kinase OS=Alcanivorax borkumensis; Uncharacterized HTH-type transcriptional regulator yeiE OS=Escherichia coli; 30S ribosomal protein S7 OS=Escherichia coli; NADH-quinone oxidoreductase subunit B OS=Geobacter sp.; NADH-quinone oxidoreductase subunit B OS=Jannaschia sp.; 30S ribosomal protein S16 OS=Bacteroides thetaiotaomicron; Elongation factor Tu OS=Actinobacillus pleuropneumoniae serotype 3; 50S ribosomal protein L4 OS=Acinetobacter sp.; 50S ribosomal protein L25 OS=Caldicellulosiruptor saccharolyticus; Spermidine N(1)-acetyltransferase OS=Escherichia coli; 50S ribosomal protein L4 OS=Neisseria meningitidis serogroup C; Aminoglycoside 3'-phosphotransferase OS=Escherichia coli; Putative uncharacterized protein yghX OS=Escherichia coli; N-hydroxyarylamine O-acetyltransferase OS=Escherichia coli; Outer membrane protein A OS=Escherichia coli; Ribosomal RNA small subunit methyltransferase H OS=Thermoanaerobacter tengcongensis; Ribosomal RNA small subunit methyltransferase H OS=Clostridium acetobutylicum; Phosphoribosylformylglycinamide cyclo-ligase OS=Geobacillus sp.; Uncharacterized protein yhbW OS=Escherichia coli; Putative acyl-[acyl-carrier-protein] desaturase desA1 OS=Mycobacterium tuberculosis; Riboflavin biosynthesis protein RibD OS=Escherichia coli; Ribosomal RNA large subunit methyltransferase G OS=Escherichia coli; Bifunctional protein putA OS=Escherichia coli; NADH-quinone oxidoreductase subunit G OS=Escherichia coli; D-amino acid dehydrogenase small subunit OS=Azotobacter vinelandii; Peptidase T OS=Erwinia carotovora subsp. atroseptica; Cobyrinic

acid A,C-diamide synthase OS=Rhodobacter capsulatus; ATP synthase subunit beta OS=Rickettsia akari; UPF0371 protein M6_Spy1067 OS=Streptococcus pyogenes serotype; NAD-reducing hydrogenase hoxS subunit alpha OS=Cupriavidus necator; Chaperone protein DnaK OS=Escherichia coli; Urease subunit alpha OS=Yersinia enterocolitica serotype; 1-deoxy-D-xylulose-5-phosphate synthase OS=Bordetella avium; and Translation initiation factor IF-2 OS=Proteus vulgaris.

Methods for Using Monomeric α S Substrate in SAA

[0091] Human monomeric α S substrate or conservative variant compositions described herein having a decreased tendency to self-aggregate are useful as substrate proteins in α S-SAA. Example α S-SAA methods include those disclosed in US20160077111A1 (“slow assay”), US20210063416A1 (“fast assay”), and U.S. Nonprovisional Patent Application No. 17/154,966.

[0092] Examples have been included to explain more clearly particular aspects of the invention.

EXAMPLES

Example 1: Synthesis of recombinant monomeric α S substrate SEQ ID NO: 6 using SEQ ID NO: 2

[0093] E. coli BL21(DE3) was transformed according to the manufacturer’s instructions (Lucigen® E. cloni® Express Chemically Competent Cells, MA019 Rev. 31OCT2016) with the expression vector comprising the plasmid represented by SEQ ID NO: 2. This plasmid comprises a codon optimized nucleic acid sequence represented by SEQ ID NO: 1, which encodes for the C-terminal HisTag α S protein represented by SEQ ID NO: 6 without amino acid misincorporation.

[0094] Bacterial pellets were grown in-house overnight using autoinduction media. Pellets were tested for inclusion bodies using B-Per reagent and SDS-PAGE. α S protein represented

by SEQ ID NO: 6 was highly expressed (~30% of all protein) and there were no inclusion bodies detected by B-Per. SEQ ID NO: 1 was verified by DNA sequencing.

Example 2: Purification of α S protein by microfluidization and acid precipitation

[0095] The bacterial pellets containing α S protein were prepared as described in Example 1, washed, frozen at -80 °C, and stored until use. For purification, the cells were thawed in a water bath set to 30 °C for 40-45 min in lysis buffer (50mM NaH₂PO₄ pH: 8.0, 0.3M NaCl, 0.2mM EDTA, 20mM Imidazole, 1mM PMSF, 0.1mM TCEP). The cells were resuspended in a final volume of lysis buffer equivalent to 4X the weight of the pellet (80 mL lysis buffer to 20 g of pellet). The resuspended cells were degassed using a standard vacuum pump. The resuspended cells were lysed by a microfluidizer (LM20 Microfluidizer®). The crude lysate was clarified by centrifugation to remove large cellular debris.

[0096] The clarified lysate was titrated by stepwise addition of 1M HCl during agitation. After reaching a target pH, the acidified lysate was incubated with agitation for 20-60 min. The acidified lysate was clarified by centrifugation, and the supernatant was neutralized using 1M NaOH to pH 8.00. The neutralized lysate was filtered by 0.22 μ m filters, and the material was loaded onto a column with nickel-Sepharose resin.

[0097] Chromatography was carried out using standard protocols. After loading the neutralized filtered lysate, the packed column was washed with a first wash buffer (50 mM NaH₂PO₄ pH: 7.4, 0.5M NaCl, 20 mM Imidazole, 0.1 mM TCEP) and a second wash buffer (50 mM NaH₂PO₄ pH: 7.4, 0.15M NaCl, 20 mM Imidazole, 0.1 mM TCEP). Protein was eluted with elution buffer (50 mM NaH₂PO₄ pH: 7.4, 0.15M NaCl, 250 mM Imidazole, 0.1 mM TCEP), and elution fractions were collected on ice. The elution fractions were evaluated by SDS-PAGE and Coomassie staining to determine the fractions with the lower amounts of contaminants and higher amounts of recombinant α S protein.

[0098] After pooling the elution fractions with high α S and low contaminants, the pool was dialyzed for 4-5 h in 1X PBS at a dilution factor of 1:200. A second dialysis was performed (1:400) overnight for a total 1:80,000 dilution factor.

[0099] The dialyzed material was filtered using a 50,000 Dalton MWCO Amicon centrifugal device. The 50kDa filtered material is the final α S monomer product, and protein concentration was determined by BCA, A280, or a combination of A280 and BCA for increased accuracy. The protein was aliquoted into single use aliquots containing around 6.5 mg each.

[00100] **Figures 2A and 2B** show electrophoresis results for C-terminal HisTag α S protein (SEQ ID NO: 6) prepared using the plasmid (SEQ ID NO: 2) comprising the nucleotide sequence (SEQ ID NO: 1). The C-terminal HisTag α S protein aliquot was separated into two fractions. One fraction of the C-terminal HisTag α S protein aliquot was contacted with DTT under disulfide bond reducing conditions. As shown in **Figures 2A and 2B**, the reduced and non-reduced fractions, respectively, of the C-terminal HisTag α S protein were essentially identical when run at protein amounts per lane of 1 μ g, 2 μ g, and 4 μ g. No band corresponding to dimer formation was visible. Accordingly, the plasmid (SEQ ID NO: 2) comprising the nucleotide sequence (SEQ ID NO: 1) produces the cysteine misincorporation-free C-terminal HisTag α S protein sequence (SEQ ID NO: 6) when expressed in *E. Coli*.

Comparative Example 1: Gel electrophoresis of recombinant human monomeric α S substrate prepared without using optimized codons

[00101] **Figure 3** shows electrophoresis results for C-terminal HisTag α S protein (intending SEQ ID NO: 6, but because of cysteine misincorporation, likely comprising a mix of SEQ ID NO: 6 and SEQ ID NO: 6-Y136C) prepared using the plasmid (SEQ ID NO: 4) comprising the nucleotide sequence (SEQ ID NO: 3) before and after treatment with DTT. Lane 1 (the first lane from the left) shows various molecular weight fractions of an incubation mixture. Lane 2

shows the recombinant monomeric α S protein reagent (SD*). Lane 3 shows the filtrate (F) through a 30 kDa cutoff filter. Lane 4 shows the retentate (R) caught by the 30 kDa cutoff filter. Lane 5 shows the filtrate through a 50 kDa cutoff filter. Lane 6 shows the retentate caught by the 50 kDa cutoff filter. The C-terminal HisTag α S protein has a nominal molecular weight of ~15 kDa, but it runs slightly higher than the 17kDa marker in SDS-PAGE. Accordingly, electrophoresis of the filtrate through the 30 kDa cutoff filter in Lane 3 showed a band at ~15 kDa, corresponding to the C-terminal HisTag α S protein. In addition to the intended 15 kDa recombinant monomeric folded α S protein, **Figure 3**, Lanes 4, 5, and 6 showed a band at ~36 kDa. After treatment of with DTT, the band at ~36 kDa disappeared, leaving only the expected band at ~15 kDa, indicating the presence and separation of dimers into monomers.

Example 3: α S-SAA of the acid precipitated monomeric α S substrate

[00102] α S-SAA using the acid precipitated α S monomer was conducted by the “fast assay” procedure, using the following parameters:

Parameter	Fast Assay (FA)
Buffer ([] (mM) and pH)	100mM PIPES pH 6.50
[NaCl] (mM)	500
ThT (μ M)	10
Substrate	Recombinant C-terminal histag α Syn (MW = 15,283mg/mmol) (SEQ ID NO: 6)
[substrate] (μ M)	19.6 μ M

[substrate] (mg/mL)	0.3
Sample (μ L)	40
Shaking type	Orbital
Shaking speed (rpm)	600-800
Shaking time (min)	1
Incubation time (min)	29
Beads material	Si ₃ N ₄
Beads size (mm)	2.38
Amount of beads	1 (ea)
Temperature ($^{\circ}$ C)	37
Reaction volume (μ L)	200

[00103] **Figure 4A** shows α S-SAA “fast assay” aggregation curves (of three replicates individually) in the presence of a confirmed PD sample, where the monomeric α S substrate was purified by acid precipitation to pH 4. The monomeric α S substrate aggregated as expected in the presence of the PD sample, reaching Fmax between 60-70 hours. However, the monomeric α S substrate showed high propensity for self-aggregation when analyzing a CSF sample from a HC (**Figure 4B**).

[00104] Likewise, **Figure 5A** shows α S-SAA “fast assay” aggregation curves (of three replicates individually) in the presence of a confirmed PD sample, where the monomeric α S substrate was purified by acid precipitation to pH 3.5. The monomeric α S substrate aggregated

as expected in the presence of the PD sample, reaching F_{max} between 60-70 hours, but showing greater variability than with the substrate produced using pH 4.0 acid precipitation. The monomeric αS substrate showed medium self-aggregation with an HC (**Figure 5B**).

[00105] **Figure 6A** shows αS -SAA “fast assay” aggregation curves (of three replicates individually) in the presence of a confirmed PD sample, where the monomeric αS substrate was purified by acid precipitation to pH 3. At pH 3, the PD aggregation is as expected for only two of the three wells. The third well showed a much lower fluorescence, which might be explained by self-aggregation. More surprisingly, the HC (**Figure 6B**) was “reproducibly positive,” indicating that the substrate purified in this way is more prone to self-aggregation (and further supporting the hypothesis that the third replicate from the PD sample could be self-aggregation).

[00106] **Figure 7A** shows αS -SAA “fast assay” aggregation curves (of three replicates individually) in the presence of a confirmed PD sample, where the monomeric αS substrate was purified by acid precipitation to pH 2.5. Strikingly, PD aggregation was heavily affected in terms of poor reproducibility, delayed aggregation of the positive wells (75-90 hours), and lack of amplification for one of the replicates, which could generate a false negative result. Conversely, the monomeric αS substrate did not self-aggregate (**Figure 7B**).

[00107] **Figure 8A** and **Figure 8B** show αS -SAA “fast assay” aggregation curves (of three replicates individually) using the monomeric αS substrate corresponding to SEQ ID NO: 6 expressed in *E. coli* strain BL21(DE3) transformed using the plasmid represented by SEQ ID NO: 2, purified by acid precipitation to pH 2.0, in the presence of a confirmed PD sample (**Figure 8A**) and in an HC (**Figure 8B**). The results were consistent with pH 2.5, with no self-aggregation in the HC, but low PD aggregation, as only two of the three replicates were positive.

Example 4: Addition of LPS to acid precipitated (pH 2.5) monomeric α S substrate prior to α S-SAA of PD sample

[00108] LPS was added to the acid precipitated (pH 2.5) monomeric α S substrate described in Example 3. **Figure 9A** shows α S-SAA “fast assay” aggregation curves (of three replicates individually) in the presence of a confirmed PD sample using 60,000 Endotoxin Units (EnU) of LPS. The LPS accelerated the aggregation, which started around 50 hours. Two of the replicates were clearly positive and showed reproducible aggregation, while the third one displayed a much lower maximum fluorescence (F_{max}). Despite the variability and low F_{max} of one replicate, this sample would be considered positive. 60,000 EnU of LPS induced very high levels of self-aggregation (**Figure 9B**). An LPS addition of 600 EnU substantially improved the reproducibility of the three replicates, which were all positive at the expected 50-75 hours (**Figure 9C**). Strikingly, 600 EnU of LPD did not induce self-aggregation, allowing clear identification of the HC sample as negative (**Figure 9D**).

Example 5: Addition of LPS to acid precipitated (pH 2.5) monomeric α S substrate prior to α S-SAA of MSA sample

[00109] LPS was added to the acid precipitated (pH 2.5) monomeric α S substrate. **Figure 10A** and **Figure 10B** show α S-SAA “fast assay” aggregation curves (of two replicates individually) in the presence of a confirmed MSA sample using 600 EnU and 120 EnU, respectively. The LPS induced aggregation with low fluorescence, consistent with MSA diagnosis.

Example 5: Purification of α S protein by microfluidization and double acid precipitation

[00110] **Figure 11** shows a flowchart of an example method for purifying monomeric α S substrate that, together with proper α S-SAA conditions, reduces, slows, or prevents altogether

misfolding and self-aggregation when used in an α S-SAA, yet retains its activity in the presence of soluble, misfolded α S protein in a biological sample.

[00111] Thus, the bacterial pellets containing α S protein were prepared as described in Example 1, washed, frozen at -80 °C, and stored until use. For purification, the cells were thawed in a water bath set to 30 °C for 40-45 min in lysis buffer (50mM NaH₂PO₄ pH: 8.0, 0.3M NaCl, 0.2mM EDTA, 20mM Imidazole, 1mM PMSF, 0.1mM TCEP). The cells were resuspended in a final volume of lysis buffer equivalent to 4X the weight of the pellet (80 mL lysis buffer to 20 g of pellet). The resuspended cells were degassed using a standard vacuum pump. The resuspended cells were lysed by a microfluidizer (LM20 Microfluidizer made by Microfluidics™). The crude lysate was clarified by centrifugation to remove large cellular debris.

[00112] The clarified lysate was titrated by stepwise addition of 1M HCl during agitation. After reaching a target pH of 3.5, the acidified lysate was incubated with agitation for 20-60 min. The acidified lysate was clarified by centrifugation. The clarified lysate was again titrated by stepwise addition of 1M HCl during agitation. After reaching a target pH of 2.0, the acidified lysate was incubated with agitation for 20-60 min. The double acidified lysate was clarified by centrifugation, filtered by 0.45 μ m filters, and the supernatant was neutralized using 1M NaOH to pH 8.00. The neutralized lysate was filtered by 0.22 μ m filters, and the material was loaded onto a column with nickel-Sepharose resin.

[00113] Chromatography was carried out using standard protocols. After loading the neutralized filtered lysate, the packed column was washed with a first wash buffer (50 mM NaH₂PO₄ pH: 7.4, 0.5M NaCl, 20 mM Imidazole, 0.1 mM TCEP) and a second wash buffer (50 mM NaH₂PO₄ pH: 7.4, 0.15M NaCl, 20 mM Imidazole, 0.1 mM TCEP). Protein was eluted with elution buffer (50 mM NaH₂PO₄ pH: 7.4, 0.15M NaCl, 250 mM Imidazole, 0.1 mM TCEP), and elution fractions were collected on ice. The elution fractions were evaluated

by SDS-PAGE and Coomassie staining to determine the fractions with the lower amounts of contaminants and higher amounts of recombinant α S protein.

[00114] After pooling the elution fractions with high α S and low contaminants, the pool was dialyzed for 4-5 h in 1X PBS at a dilution factor of 1:200. A second dialysis was performed (1:400) overnight for a total 1:80,000 dilution factor.

[00115] The dialyzed material was filtered using a 50,000 Dalton MWCO Amicon centrifugal device. The 50kDa filtered material is the final α S monomer product and protein concentration was determined by BCA, A280, or a combination of A280 and BCA for increased accuracy. The protein was aliquoted into single use aliquots containing around 6.5 mg each.

[00116] **Figures 12A-12D** show α S-SAA “fast assay” aggregation curves (of three replicates individually) of the resultant α S protein in the presence of three different confirmed PD samples (**Figures 12A-12C**) and in an HC (**Figure 12D**). The PD samples showed excellent reproducibility, and the HC showed no self-aggregation. This is surprising and counter-intuitive in view of the self-aggregation exhibited in the HC using a single acid precipitation step at pH 3.5 (**Figure 5B**) and the lack of desired aggregation in the PD sample using a single acid precipitation step at pH 2.0 (**Figure 8A**).

Example 6: Purification of α S protein by microfluidization, acid precipitation, and multiple filtration

[00117] The α S substrate was prepared as described in Example 2, with the acid precipitation pH target at about 3.1, except that one aliquot of the dialyzed, filtered α S substrate was filtered a second time at 50kDa and one aliquot of the dialyzed, filtered α S substrate was filtered a second time at 30kDa.

[00118] **Figures 13A-13D** show α S-SAA “fast assay” aggregation curves (of three replicates individually) using the monomeric α S substrate corresponding to SEQ ID NO: 6

expressed in *E. coli* strain BL21(DE3) transformed using the plasmid represented by SEQ ID NO: 2, purified by acid precipitation to pH about 3.1, and further purified by a second filtration of the dialyzed, filtered protein using a 50kDa filter in the presence of synthetic seeds (**Figure 13A**) and in an HC (**Figure 13B**) or using a 30kDa filter in the presence of synthetic seeds (**Figure 13C**) and in an HC (**Figure 13D**).

[00119] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs.

[00120] Where a range of values is provided, each intervening value, to the tenth of the unit of the lower limit unless the context clearly dictates otherwise, between the upper and lower limit of that range and any other stated or intervening value in that stated range, is encompassed within the invention. The upper and lower limits of these smaller ranges may independently be included in the smaller ranges and are also encompassed within the invention, subject to any specifically excluded limit in the stated range. Where the stated range includes one or both of the limits, ranges excluding either or both of those included limits are also included in the invention.

[00121] The term “about” in conjunction with a number is intended to include $\pm 10\%$ of the number. This is true whether “about” is modifying a stand-alone number or modifying a number at either or both ends of a range of numbers. In other words, “about 10” means from 9 to 11. Likewise, “about 10 to about 20” contemplates 9 to 22 and 11 to 18. In the absence of the term “about,” the exact number is intended. In other words, “10” means 10.

[00122] The singular forms “a”, “and”, and “the” include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to “a sample” also includes a plurality of such samples and reference to “a monomeric α S substrate” includes reference to one or more such molecule, and so forth.

[00123] The complete disclosure of all patents, patent applications, and publications, and electronically available material cited herein are incorporated by reference, whether or not the specific citation herein so states. The foregoing detailed description and examples have been given for clarity of understanding only. No unnecessary limitations are to be understood therefrom. The invention is not limited to the exact details shown and described, for variations obvious to one skilled in the art will be included within the invention defined by the claims.