

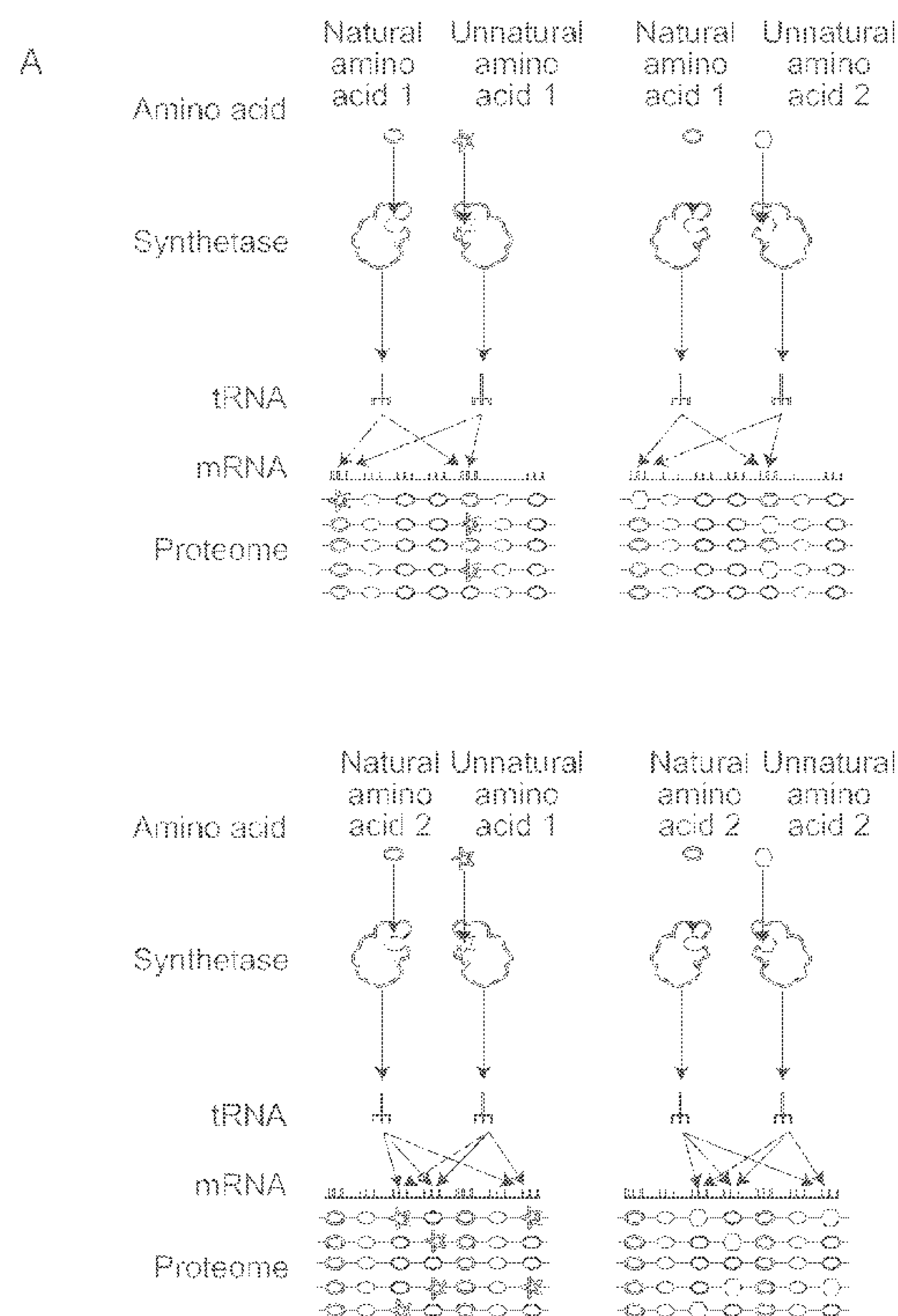


(86) **Date de dépôt PCT/PCT Filing Date:** 2015/03/10
(87) **Date publication PCT/PCT Publication Date:** 2015/09/17
(85) **Entrée phase nationale/National Entry:** 2016/08/29
(86) **N° demande PCT/PCT Application No.:** GB 2015/050694
(87) **N° publication PCT/PCT Publication No.:** 2015/136265
(30) **Priorité/Priority:** 2014/03/14 (GB1404569.4)

(51) **Cl.Int./Int.Cl.** **C12N 9/36** (2006.01),
C07K 1/13 (2006.01), **C07K 14/435** (2006.01)
(71) **Demandeur/Applicant:**
MEDICAL RESEARCH COUNCIL, GB
(72) **Inventeur/Inventor:**
ELLIOTT, THOMAS, GB
(74) **Agent:** OSLER, HOSKIN & HARCOURT LLP

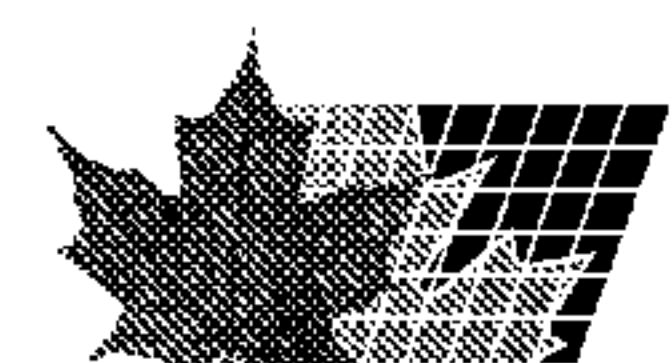
(54) **Titre : ACIDES AMINES DE CYCLOPROPENE ET PROCEDES**
(54) **Title: CYCLOPROPENE AMINO ACIDS AND METHODS**

FIG. 1



(57) **Abrégé/Abstract:**

The invention relates to a polypeptide comprising an amino acid having a cyclopropene group wherein said cyclopropene group is joined to the amino acid via a carbamate group. Suitably the cyclopropene group is a 1,3-disubstituted cyclopropene such as a 1,3-



(57) Abrégé(suite)/Abstract(continued):

di methylcyclopropene. Suitably the cyclopropene group is present as a residue of a lysine amino acid. The invention also relates to methods of making the polypeptides. The invention also relates to an amino acid comprising cyclopropene wherein said cyclopropene group is joined to the amino acid moiety via a carbamate group.

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

(19) World Intellectual Property
Organization
International Bureau

(10) International Publication Number

WO 2015/136265 A1(43) International Publication Date
17 September 2015 (17.09.2015)**WIPO | PCT**

(51) International Patent Classification:

C12N 9/36 (2006.01) *C07K 1/13* (2006.01)
C07K 14/435 (2006.01)

(21) International Application Number:

PCT/GB2015/050694

(22) International Filing Date:

10 March 2015 (10.03.2015)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

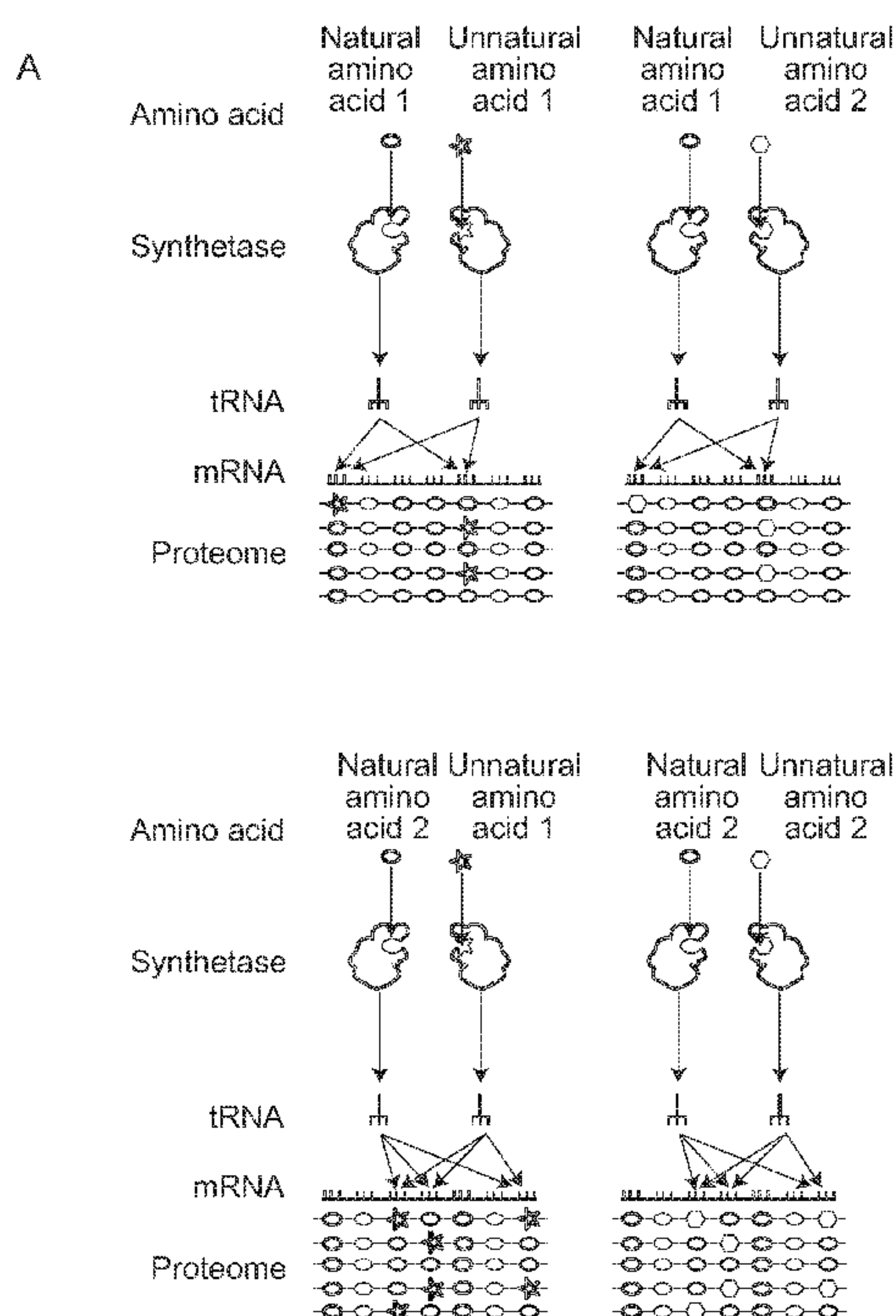
1404569.4 14 March 2014 (14.03.2014) GB

(71) Applicant: **MEDICAL RESEARCH COUNCIL**
[—/GB]; 2nd Floor, David Philips Building, Polaris
House, North Star Avenue, Swindon SN2 1FL (GB).(72) Inventor: **ELLIOTT, Thomas**; MRC Laboratory of Mo-
lecular Biology, Francis Crick Avenue, Cambridge Bio-
medical Campus, Cambridge Cambridgeshire CB2 0QH
(GB).(74) Agent: **SCRIPT IP LIMITED**; 18 Bridge Street, Frome,
Somerset BA11 1BB (GB).(81) Designated States (unless otherwise indicated, for every
kind of national protection available): AE, AG, AL, AM,
AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY,
BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM,
DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT,
HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR,
KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG,
MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM,
PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC,
SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN,
TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.(84) Designated States (unless otherwise indicated, for every
kind of regional protection available): ARIPO (BW, GH,
GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ,
TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU,
TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE,
DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU,
LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK,

[Continued on next page]

(54) Title: CYCLOPROPENE AMINO ACIDS AND METHODS

FIG. 1



(57) Abstract: The invention relates to a polypeptide comprising an amino acid having a cyclopropene group wherein said cyclopropene group is joined to the amino acid via a carbamate group. Suitably the cyclopropene group is a 1,3-disubstituted cyclopropene such as a 1,3-di methylcyclopropene. Suitably the cyclopropene group is present as a residue of a lysine amino acid. The invention also relates to methods of making the polypeptides. The invention also relates to an amino acid comprising cyclopropene wherein said cyclopropene group is joined to the amino acid moiety via a carbamate group.

WO 2015/136265 A1



SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

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

CYCLOPROPENE AMINO ACIDS AND METHODS

FIELD OF THE INVENTION

The invention relates to site-specific incorporation of bio-orthogonal groups via the
5 (expanded) genetic code. In particular the invention relates to incorporation of
carbamate-bonded cyclopropenes into polypeptides via genetically incorporated amino
acids such as lysines. Such cyclopropene groups are useful for addition of further
chemical groups such as tetrazines.

10 BACKGROUND TO THE INVENTION

The site-specific incorporation of bio-orthogonal groups via genetic code expansion
provides a powerful general strategy for site specifically labelling proteins with any
probe. However, the slow reactivity of the bio-orthogonal functional groups that can be
genetically encoded, and/or their need for photoactivation, has limited this strategy's
15 utility.

The rapid, site-specific labeling of proteins with diverse probes remains an outstanding
challenge for chemical biologists; enzyme mediated labeling approaches may be rapid,
but use protein or peptide fusions that introduce perturbations into the protein under
20 study and may limit the sites that can be labeled, while many 'bio-orthogonal' reactions
for which a component can be genetically encoded are too slow to effect the
quantitative and site specific labeling of proteins on a time-scale that is useful to study
many biological processes.

25 There is a pressing need for general methods to site-specifically label proteins, in
diverse contexts, with user-defined probes.

Inverse electron demand Diels-Alder reactions involving tetrazines have emerged as an
important class of rapid bio-orthogonal reactions. The rates reported for some of these
30 reactions are very fast.

Yu et al 2012 (Angew. Chem. Int. Ed. Volume 51, pages 10600-10604) disclose
Genetically Encoded Cyclopropene Directs Rapid, Photoclick Chemistry Mediated
Protein Labelling in Mammalian Cells. The authors report the synthesis of a stable
35 cyclopropene amino acid, the characterisation of its reactivity in a photo induced
cycloaddition reaction with two tetrazoles, its site-specific incorporation into proteins
both in *E.coli* and in mammalian cells, and its use in directing bioorthogonal labelling of

proteins both *in vitro* and *in vivo*. In order to incorporate their cyclopropene containing amino acid into proteins, the authors had to evolve an orthogonal tRNA/tRNA synthetase pair that selectively charges their cyclopropene lysine amino acid in response to a TAG amber codon. This required a synthetase library to be
5 constructed, five positions within that synthetase to be randomised, together with at least five rounds of positive and negative selection screening. It is a drawback of this work that it relies on the specific mutant synthetase produced. In joining their tetrazole compounds to the cyclopropene moiety in their modified amino acids, Yu *et al* use photo activation. Photo activation is carried out at either 302 nano metres or 365
10 nano metres. The requirement for photo activation in joining tetrazoles to the amino acid of Yu *et al* is a drawback in the art. This is a laborious extra step in the conjugation chemistry. UV is also damaging to cells and so is disadvantageous in the *in vivo*/cellular setting.

15 Kamber *et al* disclose Isomeric Cyclopropenes Exhibiting Unique Bioorthogonal Reactivities (2013 JACS Volume 135, pages 13680-13683). The authors discuss two reactions that can be used to tag biomolecules in complex environments: the inverse electron demand Diels-Alder reaction of tetrazines with 1,3-disubstituted cyclopropenes, and the 1,3-dipolar cycloaddition of nitrile imines with 3,3-disubstituted
20 cyclopropenes. The authors discuss various chemical reaction schemes used to generate stable cyclo adducts. None of the molecules discussed by Kamber *et al* are amino acids. There is no reason to imagine that the compounds as described could be incorporated into amino acids. Even if any such incorporation was attempted, there is absolutely no suggestion or guidance which might allow such compounds to be
25 incorporated into polypeptides. No schemes for synthesis of amino acids comprising any of the chemical groups described are presented by Kamber *et al*. There are no biochemical tools for incorporation into proteins mentioned anywhere in this document. Kamber *et al* are solely concerned with examining the substitution pattern on the cyclopropene, one such pattern allowing reactions with tetrazines and one such
30 pattern not being permissive of reactions with tetrazines.

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

SUMMARY OF THE INVENTION

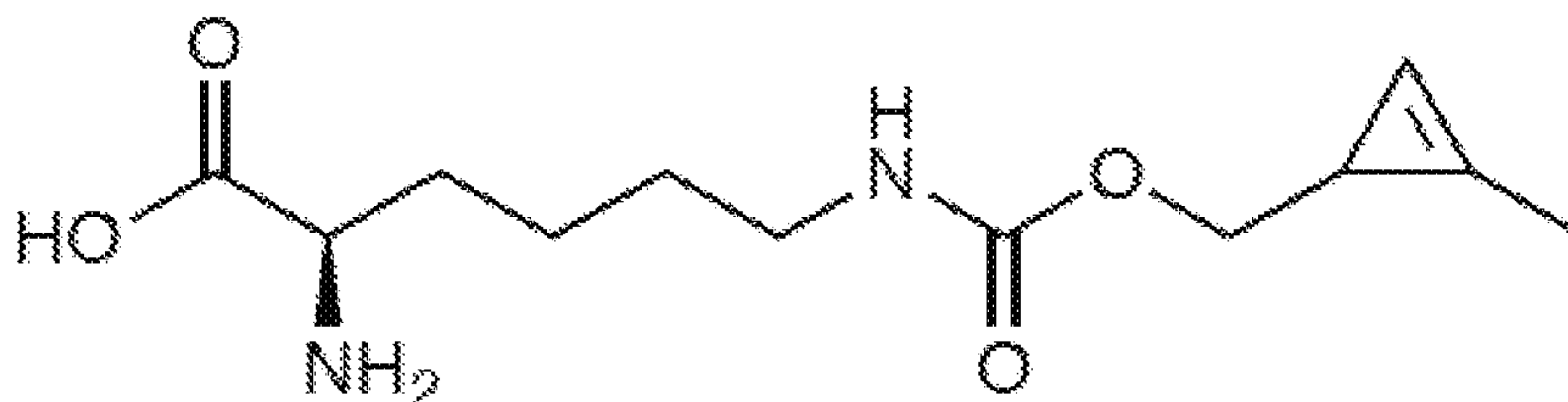
In one aspect the invention provides a polypeptide comprising an amino acid having a cyclopropene group wherein said cyclopropene group is joined to the amino acid via a carbamate group.

Suitably said cyclopropene group is a 1,3-disubstituted cyclopropene. Suitably said cyclopropene is a 1,3-dimethylcyclopropene. Suitably said cyclopropene group is present as a residue of a lysine amino acid. Suitably said polypeptide further comprises a tetrazine compound linked to said cyclopropene group.

In another aspect, the invention relates to an amino acid comprising cyclopropene wherein said cyclopropene group is joined to the amino acid moiety via a carbamate group.

Suitably said cyclopropene is a 1,3-disubstituted cyclopropene. Suitably said cyclopropene is a 1,3-dimethylcyclopropene. Suitably said amino acid is a lysine amino acid. Suitably said amino acid comprises *N*^ε-[((2-methylcycloprop-2-en-1-yl)methoxy)carbonyl]-l-lysine.

Suitably said amino acid comprises, or more suitably consists of:



In another aspect, the invention relates to a method of producing a polypeptide comprising a cyclopropene group wherein said cyclopropene group is joined to the amino acid moiety via a carbamate group, said method comprising genetically incorporating an amino acid comprising a cyclopropene group joined to the amino acid moiety via a carbamate group, into a polypeptide.

Suitably producing the polypeptide comprises

- providing a nucleic acid encoding the polypeptide which nucleic acid comprises an orthogonal codon encoding the amino acid having a cyclopropene group;

(ii) translating said nucleic acid in the presence of an orthogonal tRNA synthetase/tRNA pair capable of recognising said orthogonal codon and incorporating said amino acid having a cyclopropene group into the polypeptide chain.

5 Suitably said orthogonal codon comprises an amber codon (TAG), said tRNA comprises *MbtRNA*_{CUA} and said tRNA synthetase comprises *MbPylRS*.

Suitably said orthogonal codon comprises an amber codon (TAG), said tRNA comprises *MmtRNA*_{CUA} and said tRNA synthetase comprises *MmPylRS*.

10

In another aspect, the invention relates to a method as described above wherein said amino acid comprising a cyclopropene group is an amino acid as described above.

15 In another aspect, the invention relates to a method of producing a polypeptide comprising a tetrazine group, said method comprising providing a polypeptide comprising a cyclopropene group as described above, contacting said polypeptide with a tetrazine compound, and incubating to allow joining of the tetrazine to the cyclopropene group by an inverse electron demand Diels-Alder cycloaddition reaction.

20 Suitably said reaction is allowed to proceed for 10 minutes or less, preferably for 1 minute or less, preferably for 30 seconds or less. Reactions *in vivo*, or in eukaryotic culture conditions such as tissue culture medium or other suitable media for eukaryotic cells, may need to be conducted for longer than 30 seconds to achieve maximal labelling. The skilled operator can determine optimum reaction times by trial and error
25 based on the guidance provided herein.

In another aspect, the invention relates to a polypeptide as described above wherein said polypeptide comprises two or more amino acids each having a cyclopropene group, wherein each said cyclopropene group is joined to each said amino acid via a carbamate
30 group. Provision of two or more cyclopropene groups on the polypeptide advantageously allows joining of two or more conjugated groups (functional groups) to the polypeptide. This is especially helpful when the conjugated groups (functional groups) comprise drug molecules such as cytotoxic molecules such as in an antibody-drug-conjugate.

35

Suitably said polypeptide comprises four amino acids each having a cyclopropene group.

Suitably the antibody drug conjugate (ADC) comprising a polypeptide as described above comprises four amino acids each having a cyclopropene group. This is especially advantageous for the joining of four cytotoxic molecules to the ADC of interest.

5

In another aspect, the invention relates to an antibody drug conjugate (ADC) comprising a polypeptide as described above. Suitably the polypeptide is an antibody polypeptide such as whole antibody (e.g. a monoclonal antibody (mAb)) or is an antibody fragment (e.g. a single-chain variable fragment [scFv]), suitably an antibody
10 fragment comprising CDR amino acid sequence.

Suitably the antibody polypeptide (or fragment) may advantageously be humanised by manufacture of chimaeric antibody polypeptide(s); suitably the antibody polypeptide (or fragment) may advantageously be CDR-grafted; suitably the antibody polypeptide (or fragment) may advantageously be fully humanised to the extent that the technology
15 permits.

Suitably the antibody polypeptide (or fragment) may be fused to another polypeptide of interest such as such as a ligand for the transferrin receptor, for example transferrin or a part thereof, to assist in transport and/or targeting of the ADC.

20 In another aspect, the invention relates to a polypeptide as described above wherein said tetrazine group is further joined to a fluorophore.

Suitably said fluorophore comprises fluorescein, tetramethyl rhodamine (TAMRA) or boron-dipyrromethene (BODIPY).

25

Suitably said fluorophore may comprise one or more Alexa fluorophore(s). Suitably said fluorophore may comprise one or more Cyanine based fluorophore(s).

DETAILED DESCRIPTION

30 Genetic code expansion methods allow the quantitative, site-specific, and genetically directed incorporation of unnatural amino acids with diverse chemical structures and bearing diverse functional groups. This is most commonly achieved by inserting the unnatural amino acid in response to an amber stop codon introduced into a gene of interest.^{12, 13} Genetic code expansion is achieved via the introduction of an orthogonal
35 aminoacyl-tRNA synthetase/tRNA_{CUA} pair into cells. The pyrrolysyl-tRNA

synthetase/tRNA_{CUA} pair is amongst the most useful pairs for genetic code expansion,¹³ because it 1) can specifically recognize a range of useful unnatural amino acids, 2) can be evolved to recognize an extended range of chemical structures, and 3) can be used as an orthogonal pair for genetic code expansion in *E. coli*,¹⁴ yeast,¹⁵ mammalian cells,¹⁶⁻¹⁸
5 *C. elegans*¹⁹ and *D. melanogaster*.²⁰

We demonstrate production of newly synthesized proteins with cyclopropene groups that can be labelled with tetrazine probes introduced via a chemoselective inverse electron demand Diels-Alder reaction.

10

In another aspect, the invention relates to a homogenous recombinant polypeptide as described above. Suitably said polypeptide is made by a method as described above.

Also disclosed is a polypeptide produced according to the method(s) described herein.
15 As well as being the product of those new methods, such a polypeptide has the technical feature of comprising cyclopropene suitably carbamate-linked cyclopropene.

Mutating has its normal meaning in the art and may refer to the substitution or truncation or deletion of the residue, motif or domain referred to. Mutation may be
20 effected at the polypeptide level e.g. by synthesis of a polypeptide having the mutated sequence, or may be effected at the nucleotide level e.g. by making a nucleic acid encoding the mutated sequence, which nucleic acid may be subsequently translated to produce the mutated polypeptide. Where no amino acid is specified as the replacement amino acid for a given mutation site, suitably a randomisation of said site is used. As a
25 default mutation, alanine (A) may be used. Suitably the mutations used at particular site(s) are as set out herein.

A fragment is suitably at least 10 amino acids in length, suitably at least 25 amino acids, suitably at least 50 amino acids, suitably at least 100 amino acids, suitably at least 200
30 amino acids, suitably at least 250 amino acids, suitably at least 300 amino acids, suitably at least 313 amino acids, or suitably the majority of the polypeptide of interest.

The methods of the invention may be practiced *in vivo* or *in vitro*.

In one embodiment, suitably the methods of the invention are not applied to the human or animal body. Suitably the methods of the invention are *in vitro* methods. Suitably the methods do not require the presence of the human or animal body. Suitably the methods are not methods of diagnosis or of surgery or of therapy of the
5 human or animal body.

The term 'comprises' (comprise, comprising) should be understood to have its normal meaning in the art, i.e. that the stated feature or group of features is included, but that the term does not exclude any other stated feature or group of features from also being
10 present.

ADVANTAGES

Cyclopropene is a less carbon rich group than known protein labelling groups.
15

Cyclopropene amino acid of the current invention leads to more rapid protein labelling than prior art techniques.

Using the cyclopropene amino acid of the present invention leads to a more efficient
20 incorporation than prior art labelled amino acids.

It has been known to incorporate amino acids bearing norbornene groups into proteins. The present invention offers specific advantages over prior art methods involving norbornene groups. For example, although the conjugation chemistry for cyclopropene
25 amino acids of the invention is similar to that of norbornene containing amino acids, conjugation to cyclopropene amino acids can be faster.

Incorporation of cyclopropene amino acids according to the invention can be more efficient than incorporation of prior art unnatural amino acids. The incorporation of
30 cyclopropene amino acids according to the invention can lead to a higher level of incorporation than prior art unnatural amino acids.

It is an advantage of the invention that the cyclopropene amino acids taught can be incorporated using wild type tRNA synthetases. Prior art unnatural amino acids have
35 tended to require mutant tRNA synthetases for their incorporation, such as, for example, amino acids incorporating BCN groups.

Rapid conjugation reactions for unnatural amino acids incorporated into polypeptides have been mentioned in the prior art. For example, TCO/BCN amino acids offer rapid reaction times, which can be faster than norbornene reaction times. However, it is an advantage of the cyclopropene amino acids that very rapid reaction times are provided.

5

Certain known unnatural amino acids are able to use the wild type tRNA synthetases. For example, amino acids comprising norbornene groups can be incorporated using wild type tRNA synthetase. However, by using cyclopropene containing amino acids of the invention a higher level of incorporation is achieved. In other words, the amount of material produced which comprises the unnatural amino acid is greater when using cyclopropene containing amino acids of the invention than when using prior art unnatural amino acids such as those comprising norbornene.

10

It is an advantage of the invention that the cyclopropene amino acids form excellent substrates for the tRNA synthetases noted herein, most suitably the wild type tRNA synthetases noted herein.

15

It is an advantage of the invention that the cyclopropene containing amino acids support excellent linker chemistry, for example rapid and specific reaction with tetrazine containing compounds.

20

It is an advantage of the invention that the cyclopropene containing amino acids are smaller in size than known unnatural amino acids previously used to label proteins. For example, a known unnatural amino acid comprising norbornene can be incorporated into polypeptides, but cyclopropene containing amino acids of the invention are advantageously of smaller size than the norbornene containing amino acids of the prior art.

25

It is an advantage of the invention that the cyclopropene amino acids are less likely to perturb protein structure when incorporated into polypeptides. At least part of this advantageous effect may be attributed to the small size of the cyclopropene molecular group.

30

A key advantage of incorporation of a cyclopropene group is that it permits a range of extremely useful further compounds such as labels to be easily and specifically attached to the cyclopropene group.

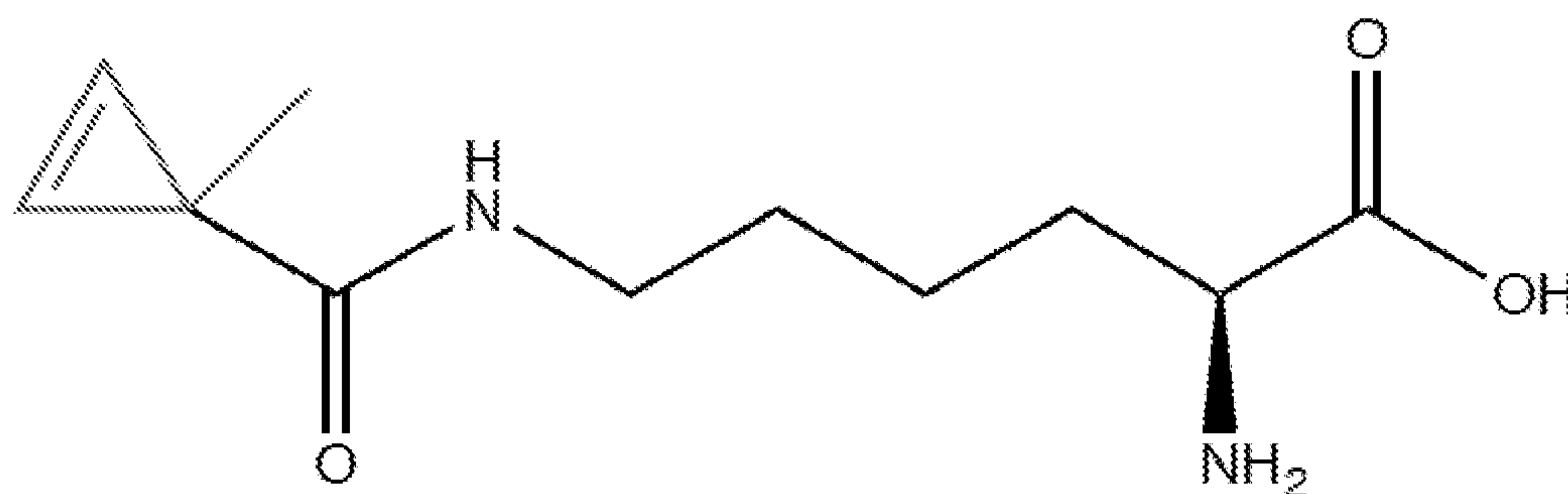
35

In another aspect, the invention relates to a polypeptide as described above wherein said cyclopropene group is joined to a tetrazine group.

CYCLOPROPENE – CARBAMATE LINKAGE

5

An unnatural amino acid comprising an amide bonded cyclopropene has been described in the prior art (Yu *et al* 2012). This amino acid is 3,3 disubstituted. This amino acid is as follows:

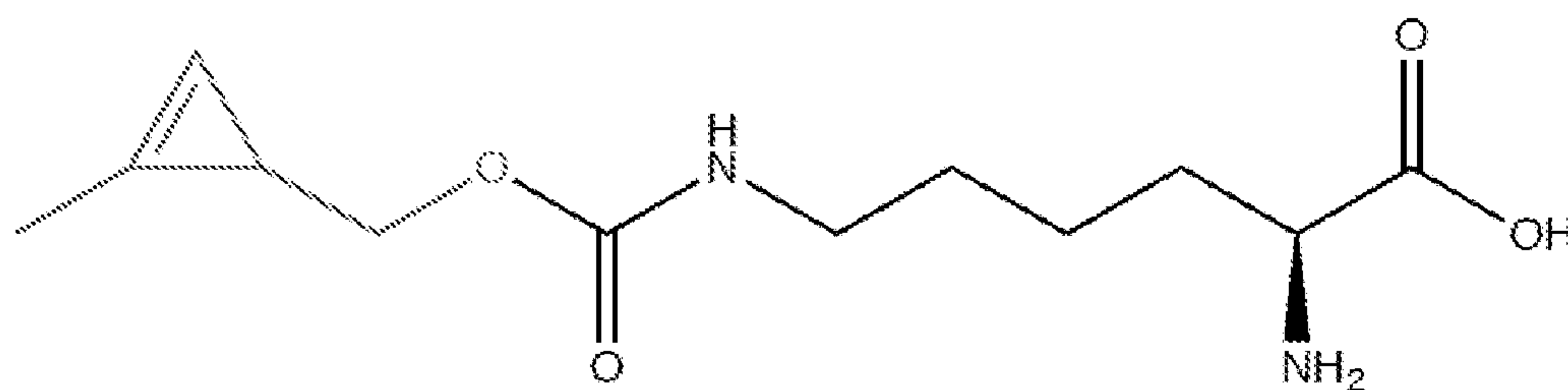


10 In order to incorporate this amino acid into polypeptides, it is essential to use a mutant tRNA synthetase.

In contrast, the amino acid comprising cyclopropene of the present invention contains a carbamate group (rather than an amide group). The cyclopropene containing amino acid of the present invention is therefore chemically distinct from the amide bonded cyclopropene amino acid in the art.

20

An exemplary amino acid of the invention is 1,3 disubstituted. An exemplary amino acid of the invention is as follows:



It is an advantage of the carbamate – cyclopropene amino acid of the invention that it is incorporated well by the wild type tRNA synthetase. This has the advantage of requiring less biological manipulation in order to obtain good incorporation. This also provides the advantage of enhanced or increased incorporation. In other words, the

cyclopropene – carbamate amino acid of the present invention is incorporated to higher levels and/or more efficiently than known unnatural amino acids.

5 Use of the cyclopropene amino acid of the invention may provide a superior rate of reaction with tetrazine compounds.

The carbamate chemistry of the invention provides the advantage of more degrees of freedom in the chemical structure of the incorporated amino acid. In particular, the carbamate cyclopropene of the invention has more degrees of freedom compared to the
10 amide cyclopropene known in the art. Similarly, the carbamate cyclopropene of the invention is more accessible when present in the polypeptide chain.

By comparison with the amide bonded cyclopropene known in the art, the carbamate cyclopropene of the present invention is a slightly “longer” amino acid. This provides
15 the advantage of a greater “reach” for the groups of the amino acid protruding away from the amino acid backbone. Again, this can render those groups more accessible for further labelling or conjugation reactions.

The chemical structure of the carbamate cyclopropene of the invention advantageously
20 provides more conformational degrees of freedom. In other words, the carbamate cyclopropene group of the invention can adopt more conformations within a protein structure than prior art amide bonded cyclopropene amino acids.

In more detail, this may arise from the nature of the bonding between cyclopropene
25 group and amino acid group. In the prior art amide arrangement, the important bond is SP² hybridised. In the invention, the important bond is SP³ hybridised, which is a more flexible bonding arrangement.

Moreover, the cyclopropene carbamate arrangement of the invention comprises a
30 methylene group between the carbamate and the cyclopropene group. Firstly, this provides a longer molecule. The prior art amide bonded version is a less advantageous shorter molecule. More specifically, the methylene carbon in the amino acid of the present invention corresponds to a double bonded oxygen group (=O) instead of the advantageous methylene carbon of the present invention. The double bonded version
35 in the prior art amide amino acid cannot rotate as freely as the methylene carbon bonded group in the amino acid of the invention.

The fact that the amino acid of the present invention is smaller than prior art norbornene containing amino acids and yet still preserves the advantageous carbamate chemistry is a benefit of the invention. This benefit provides, among other things, better incorporation of the amino acid into the polypeptide chain.

5

In addition, the joining to tetrazine compounds (tetrazine conjugation) is advantageously facilitated by the carbamate cyclopropene arrangement in the amino acid of the present invention.

10

Suitably said tetrazine group is further joined to a fluorophore.

Suitably said tetrazine group is further joined to a polyethylene glycol (PEG) group.

Suitably said fluorophore comprises fluorescein, tetramethyl rhodamine (TAMRA) or boron-dipyrromethene (BODIPY).

15

INCORPORATION

Suitably the cyclopropene amino acid of the invention is incorporated into a polypeptide using the wild type tRNA synthetase.

20

Suitably the amino acid having a cyclopropene group is incorporated at a position corresponding to a lysine residue in the wild type polypeptide. This has the advantage of maintaining the closest possible structural relationship of the cyclopropene containing polypeptide to the wild type polypeptide from which it is derived.

25

Suitably the polypeptide comprises a single cyclopropene group. This has the advantage of maintaining specificity for any further chemical modifications which might be directed at the cyclopropene group. For example when there is only a single cyclopropene group in the polypeptide of interest then possible issues of partial modification (e.g. where only a subset of cyclopropene groups in the polypeptide are subsequently modified), or issues of reaction microenvironments varying between alternate cyclopropene groups in the same polypeptides (which could lead to unequal reactivity between different cyclopropene group(s) at different locations in the polypeptide) are advantageously avoided.

30

Suitably the polypeptide comprises two cyclopropene groups; suitably the polypeptide comprises three cyclopropene groups; suitably the polypeptide comprises four

cyclopropene groups; suitably the polypeptide comprises five cyclopropene groups; suitably the polypeptide comprises ten cyclopropene groups or even more.

5 In principle multiple cyclopropene containing amino acids could be incorporated by the same or by different orthogonal codons/orthogonal tRNA pairs. Suitably multiple cyclopropene containing amino acids are incorporated by insertion of multiple amber codons (together with a suitable orthogonal tRNA synthetase as described herein).

10 Suitably the amino acid comprising cyclopropene is a lysine amino acid. In one embodiment, the tRNA may be from one species such as *Methanosarcina barkeri*, and the tRNA synthetase may be from another species such as *Methanosarcina mazei*. In another embodiment, tRNA may be from a first species such as *Methanosarcina mazei* and the tRNA synthetase may from a second species such as *Methanosarcina barkeri*. When an orthogonal pair comprises tRNA and tRNA synthetase from different species,
15 it is always with the proviso that the orthogonal pair work effectively together ie. that the tRNA synthetase will effectively amino acylate the tRNA of the amino acid of interest. Equally, mutant tRNAs or mutant tRNA synthetases may be used provided they have the correct amino acylation activity. Although it is an advantage of the invention that the cyclopropene containing amino acids of the invention are effectively
20 charged onto tRNAs using the wild type PylRS synthetase, it is equally possible to use mutant PylRS synthetases provided they are effective in charging the tRNA with the cyclopropene containing amino acid of the invention. Most suitably, orthogonal pairs comprise the tRNA and a tRNA synthetase from the same species.

25 Of course it is possible to evolve the wild type synthetase (or another variant of a suitable synthetase) to make a synthetase for incorporation of the cyclopropene amino acid of the invention which may have increased efficiency. In principle, a Pyl derived tRNA synthetase might be of use. Chimeric tRNA synthetases may be produced provided that the charging/acetylation part of the tRNA synthetase molecule is based
30 on or derived from Pyl tRNA synthetase. In other words, the anti-codon part of the tRNA molecule may be varied according to operator choice, for example to direct tRNA in recognising an alternate codon such as a sense codon, a quadruplet codon, an amber codon or another "stop" codon. However, the functional acylation/charging part of the tRNA molecule should be conserved in order to preserve the cyclopropene charging
35 activity.

Either of the *Methanosarcina barkeri* and *Methanosarcina mazei* species pyrrolysine tRNA synthetases are suitable.

Both the *Methanosarcina barkeri* and *Methanosarcina mazei* tRNAs are suitable. In any case these tRNAs differ by only one nucleotide. This one nucleotide difference has no impact on their activity in connection with cyclopropene containing amino acids. Therefore, either tRNA is equally applicable in the present invention.

The tRNA used may be varied such as mutated. In all cases, any such variants or mutants of the Pyl tRNA should always retain the capacity to interact productively with the tRNA synthetase used to charge the tRNA with the cyclopropene containing amino acid.

Genetic Incorporation and Polypeptide Production

In the method according to the invention, said genetic incorporation preferably uses an orthogonal or expanded genetic code, in which one or more specific orthogonal codons have been allocated to encode the specific amino acid residue with the cyclopropene group so that it can be genetically incorporated by using an orthogonal tRNA synthetase/tRNA pair. The orthogonal tRNA synthetase/tRNA pair can in principle be any such pair capable of charging the tRNA with the amino acid comprising the cyclopropene group and capable of incorporating that amino acid comprising the cyclopropene group into the polypeptide chain in response to the orthogonal codon. The orthogonal codon may be the orthogonal codon amber, ochre, opal or a quadruplet codon. The codon simply has to correspond to the orthogonal tRNA which will be used to carry the amino acid comprising the cyclopropene group. Preferably the orthogonal codon is amber.

It should be noted that many of the specific examples shown herein have used the amber codon and the corresponding tRNA/tRNA synthetase. As noted above, these may be varied. Alternatively, in order to use other codons without going to the trouble of using or selecting alternative tRNA/tRNA synthetase pairs capable of working with the amino acid comprising the cyclopropene group, the anticodon region of the tRNA may simply be swapped for the desired anticodon region for the codon of choice. The anticodon region is not involved in the charging or incorporation functions of the tRNA nor recognition by the tRNA synthetase so such swaps are entirely within the ambit of the skilled operator. Thus in some embodiments the anticodon region of the tRNA used in the invention such as *MbtRNA*_{CUA} or *MmtRNA*_{CUA} may be exchanged i.e. a chimeric tRNA_{CUA} may be used such that the anticodon region is swapped to recognise

an alternate codon so that the cyclopropene containing amino acid may be incorporated in response to a different orthogonal codon as discussed herein including ochre, opal or a quadruplet codon, and the nucleic acid encoding the polypeptide into which the cyclopropene amino acid is to be incorporated is correspondingly mutated to introduce
 5 the cognate codon at the point of incorporation of the cyclopropene amino acid. Most suitably the orthogonal codon is amber.

Thus alternative orthogonal tRNA synthetase/tRNA pairs may be used if desired.

10 Preferably the orthogonal synthetase/tRNA pair are *Methanosarcina barkeri* MS pyrrolysine tRNA synthetase (*MbPylRS*) and its cognate amber suppressor tRNA (*MbtRNA_{CUA}*).

The *Methanosarcina barkeri* PylT gene encodes the *MbtRNA_{CUA}* tRNA.

The *Methanosarcina barkeri* PylS gene encodes the *MbPylRS* tRNA synthetase protein.

15 When particular amino acid residues are referred to using numeric addresses, the numbering is taken using *MbPylRS* (*Methanosarcina barkeri* pyrrolysyl-tRNA synthetase) amino acid sequence as the reference sequence (i.e. as encoded by the publicly available wild type *Methanosarcina barkeri* PylS gene Accession number Q46E77):

20	MDKKPLDVLI	SATGLWMSRT	GTLHKIKHYE	VSRSKIYIEM	ACGDHLVVNN
	SRSCRTARAF	RHHKYRKTCK	RCRVSEDEDIN	NFLTRSTEGK	TSVKVKVSA
	PKVKKAMPKS	VSRAPKPLEN	PVSAKASTDT	SRSVPSPAKS	TPNSPVPTSA
	PAPSLTRS QL	DRVEALLSPE	DKISLNI AKP	FRELESELVT	RRKNDFQRLY
	TNDREDYL GK	LERDITKFFV	DRDFLEIKSP	ILIPA EYVER	MGINNDTELS
25	KQIFRVDKNL	CLRPMLAPTL	YNYLRKLDRI	LPDPIKIFEV	GPCYRKESDG
	KEHLEEF TMV	NFCQMGS GCT	RENLESLIKE	FLDYLEIDFE	IVGDSCMVYG
	DTLDIMHGDL	ELSSAVVGPV	PLDREWGIDK	PWIGAGFGLE	RLLKVMHGFK
	NIKRASRSES	YNGISTNL			

30 If required, the person skilled in the art may adapt *MbPylRS* tRNA synthetase protein by mutating it so as to optimise for the cyclopropene amino acid to be used. The need for mutation (if any) depends on the cyclopropene amino acid used. An example where the *MbPylRS* tRNA synthetase may need to be mutated is when the cyclopropene amino acid is not processed by the *MbPylRS* tRNA synthetase protein.

35

Such mutation (if desired) may be carried out by introducing mutations into the *MbPylRS* tRNA synthetase, for example at one or more of the following positions in the *MbPylRS* tRNA synthetase: M241, A267, Y271, L274 and C313.

5 tRNA Synthetases

The tRNA synthetase of the invention may be varied. Although specific tRNA synthetase sequences may have been used in the examples, the invention is not intended to be confined only to those examples.

10

In principle any tRNA synthetase which provides the same tRNA charging (aminoacylation) function can be employed in the invention.

15

For example the tRNA synthetase may be from any suitable species such as from archaea, for example from *Methanosarcina barkeri* MS; *Methanosarcina barkeri* str. Fusaro; *Methanosarcina mazei* Go1; *Methanosarcina acetivorans* C2A; *Methanosarcina thermophila*; or *Methanococcoides burtonii*. Alternatively the tRNA synthetase may be from bacteria, for example from *Desulfitobacterium hafniense* DCB-2; *Desulfitobacterium hafniense* Y51; *Desulfitobacterium hafniense* PCP1; *Desulfotomaculum acetoxidans* DSM 771.

20

Exemplary sequences from these organisms are the publically available sequences. The following examples are provided as exemplary sequences for pyrrolysine tRNA synthetases:

25

>*M.barkeri*MS/1-419/
Methanosarcina barkeri MS
 VERSION Q6WRH6.1 GI:74501411

30

MDKKPLDVLISATGLWMSRTGTLHKIKHHEVSRSKIYIEMACGDHLVVNNSRSCRTA
 RAFRHHKYRKTCKRCRVSDDEDINNFLTRSTESKNSVKVRVVSAPKVKKAMPKSVSRAP
 KPLENSVSAKASTNTSRSVSPSPAKSTPNSSVPASAPAPSLTRSQ LDRVEALLSPEDKISL
 NMAKPFRELEPELVTRRKND FQRLYTNDREDYLGKLERDITKFFVDRG FLEIKSPILIP
 AEYVERMGINNDTELSKQIFRVDKNLCLRPMLAPTLYNYLRK LDRILPGPIKIFEVGPC
 35 YRKESDGKEHLEEF TMVNFCQMGSGCTRENLEALIKEFLDY LEIDFEIVGDSCMVYGD
 TLDIMHGDLELSSAVVGPVSLDREW GIDKPWIGAGFGLERLLKVMHGFKNIKRASRS
 ESYNGISTNL

>*M.barkeri*F/1-419/

Methanosarcina barkeri str. Fusaro

VERSION YP_304395.1 GI:73668380

5

MDKKPLDVLISATGLWMSRTGTLHKIKHYEVSRSKIYIEMACGDHLVVNNSRSCRTA
RAFRHHKYRKTCKRCRVSEDEDINNFLTRSTEGKTSVKVKVVSAPKVKKAMPKSVSRAP
KPLENPVSAKASTDTSRSVSPAKSTPNPVPPTSAPAPSLTRSQDRVEALLSPEDKISL
NIAKPFRELESELVTRRKNDFFQRLYTNDREDYLGKLERDITKFFVDRDFLEIKSPILIPA
10 EYVERMGINNDTELSKQIFRVDKNLCLRPMLAPTLYNYLRKLDRLPDPIKIFEVGPCY
RKESDGKEHLEEFMTMVNFCQMSGGCTRENLESLIKEFLDYLEIDFEIVGDSCMVYGDT
LDIMHGDLELSSAVVGPVPLDREWGDIDKPWIGAGFGLERLLKVMHGFKNIKRASRSE
SYNGISTNL

15

>*M.mazei*/1-454

Methanosarcina mazei Go1

VERSION NP_633469.1 GI:21227547

20

MDKKPLNTLISATGLWMSRTGTIHKIKHHEVSRSKIYIEMACGDHLVVNNSRSSRTAR
ALRHHKYRKTCKRCRVSEDEDLNKFLTKANEDQTSVKVKVVSAPTRTKKAMPKSVARA
PKPLENTEAAQAQPSGSKFSPAIPVSTQESVSPASVSTSISSISTGATASALVKGNTNPI
TSMSAPVQASAPALTKSQTDRLLEVLLNPKDEISLNSGKPFRELESELLSRRKKDLQQIY
AEERENYLGKLEREITRFFVDRGFLEIKSPILIPLEYIERMGIDNDTELSKQIFRVDKNF
CLRPMLAPNLYNYLRKLDRLPDPIKIFEIGPCYRKESDGKEHLEEFMTMLNFCQMSG
25 CTRENLESIITDFLNHLGIDFKIVGDSCMVYGDTLDVMHGDLELSSAVVGPIPLDREW
GIDKPWIGAGFGLERLLKVKHDFKNIKRAARSESYYNGISTNL

30

>*M.acetivorans*/1-443

Methanosarcina acetivorans C2A

VERSION NP_615128.2 GI:161484944

35

MDKKPLDTLISATGLWMSRTGMIHKIKHHEVSRSKIYIEMACGERLVVNNSRSSRTAR
ALRHHKYRKTCHRCRVSEDEDINNFLTSTSEKTTVKVKVVSAPRVRKAMPKSVARAP
KPLEATAQVPLSGSKPAPATPVSAQAQAPAPSTGSASATSASAQRMANSAAAPVPT
SAPALTKGQLDRLEGLLSPKDEISLDSEKPFRELESELLSRRKKDLKRIYAEERENYLG
KLEREITKFFVDRGFLEIKSPILIPAEYVERMGINSDELSKQVFRIDKNFCLRPMLAPN
LYNYLRKLDRLPDPIKIFEIGPCYRKESDGKEHLEEFMTMLNFCQMSGGCTRENLEAH

TEFLNHLGIDFEIIGDSCMVYGNTLDVMHDDLELSSAVVGPVPLDREWGIDKPWIGA
GFGLERLLKVMHGFKNIKRAARSESYNGISTNL

>*M.thermophila*/1-478

5 *Methanosarcina thermophila*, VERSION DQ017250.1 GI:67773308

MDKKPLNTLISATGLWMSRTGKLHKIRHHEVSKRKIYIEMECGERLVVNNNSRSCRAA
RALRHHKYRKICKHCRVSEDLNKFLTRTNEDKSNKVTTVVSAPKIRKVMPSVART
PKPLENTAPVQTLPSESQPAPTTPISASTTAPASTSTTAPAPASTTAPAPASTTAPASAST
10 TISTSAMPASTSAQGTTFKNYISGGFPRPIPVQASAPALTKSQIDRLQGLLSPKDEISLDS
GTPFRKLESELLSRRRKDLKQIYAEEREHYLGKLEREITKFFVDRGFLEIKSPILIPMEYI
ERMGIDNDKELSKQIFRVDNNFCLRPMLAPNLYNLRKLNRLALPDPIKIFEIGPCYRK
ESDGKEHLEEFMNLNFCQMSGCTRENLEAIKDFLDYLGIDFEIVGDSCMVYGDITLD
VMHGDLELSSAVVGPVPMDRDWGINKPWIGAGFGLERLLKVMHNFKNIKRASRSES
15 YNGISTNL

>*M.burtonii*/1-416

Methanococcoides burtonii DSM 6242, VERSION YP_566710.1 GI:91774018

20 MEKQLLDVLVELNGVWLSRSGLLHGIRNFEITTKHHIETDCGARFTVRNSRSSRSAR
SLRHNKYRKPCKRCPADEQIDRFVKKTFKEKRQTVSVFSSPKKHVPKPKVAVIKSFS
ISTPSPKEASVSNSIPTPSISVVKDEVKVPEVKYTPSQIERLKTLMSPDDKIPIQDELPEF
KVLEKELIQRRRDDLLKKMYEEDREDRLGKLERDITEFFVDRGFLEIKSPIMIPFEYIER
MGIDKDDHLNKQIFRVDESMCLRPMLAPCLYNLYLRKLDKVLDPPIRIFEIGPCYRKES
25 DGSSHLEEFMNVNFCQMSGCTRENMEALIDEFLEHLGIEYEIEADNCMVYGDITDI
MHGDLELSSAVVGPIPLDREWGVNKPWMGAGFGLERLLKVRHNYTNIRRASRSELYY
NGINTNL

>*D.hafniense*_DCB-2/1-279

30 *Desulfitobacterium hafniense* DCB-2
VERSION YP_002461289.1 GI:219670854

MSSFWTKVQYQRLKELNASGEQLEMGFSDALSRDRAFQGIEHQLMSQGKRHLEQLR
TVKHRPALLELEEGALAKALHQQGFVQVVTPTIITKSALAKMTIGEDHPLFSQVFWLDG
35 KKCLRPMLAPNLYTLWRELERLWDKPIRIFEIGTCYRKESQGAQHLNEFTMLNLTEL
GTPLEERHQRLDMARWVLEAAGIREFELVTESSVYGDITVDVMKGDLELASGAMG
PHFLDEKWEIVDPWVGLGFGLERLLMIREGTQHVQSMARSLSYLDGVRLNIN

>*D.hafniense*_Y51/1-312

Desulfitobacterium hafniense Y51

VERSION YP_521192.1 GI:89897705

5

MDRIDHTDSKFVQAGETPVLPATFMFLTRRDPPLSSFWTKVQYQRLKELNASGEQLE
MGFSDALSRDRAFQGIEHQLMSQGKRHLEQLRTVKHRPALLEEEGLAKALHQQGF
VQVVTPTIITKSALAKMTIGEDHPLFSQVFWLDGKKCLRPMLAPNLYTLWRELERLW
DKPIRIFEIGTCYRKESQGAQHLNEFTMLNLTELGTPLERHQRLEDMARWVLEAAGI
10 REFELVTESSVVGDTVDVMKGDLELASGAMGPHFLDEKWEIVDPWVGLGFGLERLL
MIREGTQHVQSMARSLSYLDGVRLNIN

>*D.hafniense*PCP1/1-288

Desulfitobacterium hafniense

15 VERSION AY692340.1 GI:53771772

MFLTRRDPPLSSFWTKVQYQRLKELNASGEQLEMFGSDALSRDRAFQGIEHQLMSQG
KRHLEQLRTVKHRPALLEEEKLAKALHQQGFVQVVTPTIITKSALAKMTIGEDHPLF
SQVFWLDGKKCLRPMLAPNLYTLWRELERLWDPKPIRIFEIGTCYRKESQGAQHLNEF
20 TMLNLTELGTPLERHQRLEDMARWVLEAAGIREFELVTESSVVGDTVDVMKGDLE
LASGAMGPHFLDEKWEIFDPWVGLGFGLERLLMIREGTQHVQSMARSLSYLDGVRL
NIN

>*D.acetoxidans*/1-277

25 *Desulfotomaculum acetoxidans* DSM 771

VERSION YP__003189614.1 GI:258513392

MSFLWTVSQQKRLSELNASEEEKNMSFSSTSDREAAYKRVEMRLINESKQRLNKLRLH
ETRPAICALENRLAAALRGAGFVQVATPVILSKKLLGKMTITDEHALFSQVFWIEENKC
30 LRPMPLAPNLYYILKDLLRLWEKPVRIIFEIGSCFRKESQGSNHLNEFTMLNLVEWGLPE
EQRQKRISLAKLVMDETGIDEYHLEHAESVVGDTVDVMHRDIELGSGALGPHFLD
GRWGVVGPWVGIGFGLERLLMVEQGGQNVRSMGKSLTYLDGVRLNI

When the particular tRNA charging (aminoacylation) function has been provided by
35 mutating the tRNA synthetase, then it may not be appropriate to simply use another
wild-type tRNA sequence, for example one selected from the above. In this scenario, it
will be important to preserve the same tRNA charging (aminoacylation) function. This

is accomplished by transferring the mutation(s) in the exemplary tRNA synthetase into an alternate tRNA synthetase backbone, such as one selected from the above.

5 In this way it should be possible to transfer selected mutations to corresponding tRNA synthetase sequences such as corresponding pylS sequences from other organisms beyond exemplary *M.barkeri* and/or *M.mazei* sequences.

Target tRNA synthetase proteins/backbones, may be selected by alignment to known tRNA synthetases such as exemplary *M.barkeri* and/or *M.mazei* sequences.
10

This subject is now illustrated by reference to the pylS (pyrrolysine tRNA synthetase) sequences but the principles apply equally to the particular tRNA synthetase of interest.

15 For example, an alignment of all PylS sequences may be prepared. These can have a low overall % sequence identity. Thus it is important to study the sequence such as by aligning the sequence to known tRNA synthetases (rather than simply to use a low sequence identity score) to ensure that the sequence being used is indeed a tRNA synthetase.

20 Thus suitably when sequence identity is being considered, suitably it is considered across the sequences of the examples of tRNA synthetases as above. Suitably the % identity may be as defined from an alignment of the above sequences.

25 It may be useful to focus on the catalytic region. The aim of this is to provide a tRNA catalytic region from which a high % identity can be defined to capture/identify backbone scaffolds suitable for accepting mutations transplanted in order to produce the same tRNA charging (aminoacylation) function, for example new or unnatural amino acid recognition.

30 Thus suitably when sequence identity is being considered, suitably it is considered across the catalytic region. Suitably the % identity may be as defined from the catalytic region.

35 'Transferring' or 'transplanting' mutations onto an alternate tRNA synthetase backbone can be accomplished by site directed mutagenesis of a nucleotide sequence encoding the tRNA synthetase backbone. This technique is well known in the art. Essentially the backbone pylS sequence is selected (for example using the active site alignment

discussed above) and the selected mutations are transferred to (i.e. made in) the corresponding/homologous positions.

When particular amino acid residues are referred to using numeric addresses, unless
 5 otherwise apparent, the numbering is taken using MbPylRS (*Methanosarcina barkeri* pyrrolysyl-tRNA synthetase) amino acid sequence as the reference sequence (i.e. as encoded by the publicly available wild type *Methanosarcina barkeri* PylS gene Accession number Q46E77):

```

10  MDKKPLDVLI  SATGLWMSRT  GTLHKIKHYE  VSRSKIYIEM  ACGDHLVVNN
    SRSCRTARAF  RHHKYRKTCK  RCRVSDDEDIN  NFLTRSTEGK  TSVKVKVWSA
    PKVKKAMPKS  VSRAPKPLEN  PVSAAKASTDT  SRSVPSPAKS  TPNSPVPTSA
    PAPSLTRS QL  DRVEALLSPE  DKISLNI AKP  FRELESELVT  RRKNDFQRLY
    TNDREDYL GK  LERDITKFFV  DRDFLEIKSP  ILIPA EYVER  MGINNDTELS
15  KQIFRV DKNL  CLR PMLAPTL  YNYLRKLDRI  LPDPIKIFEV  GPCYRKESDG
    KEHLEEFTMV  NFCQM GSGCT  RENLESLIKE  FLDYLEIDFE  IVGDSCMVYG
    DTLDIMHGDL  ELSSAVVGPV  PLDREW GIDK  PWIGAGFGLE  RLLKVMHGFK
    NIKRASRSES YYNGISTNL
  
```

20 This is to be used as is well understood in the art to locate the residue of interest. This is not always a strict counting exercise -- attention must be paid to the context or alignment. For example, if the protein of interest is of a slightly different length, then location of the correct residue in that sequence corresponding to (for example) L266 may require the sequences to be aligned and the equivalent or corresponding residue
 25 picked, rather than simply taking the 266th residue of the sequence of interest. This is well within the ambit of the skilled reader.

Notation for mutations used herein is the standard in the art. For example L266M means that the amino acid corresponding to L at position 266 of the wild type sequence
 30 is replaced with M.

The transplantation of mutations between alternate tRNA backbones is now illustrated with reference to exemplary *M.barkeri* and *M.mazei* sequences, but the same principles apply equally to transplantation onto or from other backbones.

35

For example Mb AcKRS is an engineered synthetase for the incorporation of AcK
 Parental protein/backbone: *M. barkeri* PylS

Mutations: L266V, L270I, Y271F, L274A, C317F

Mb PCKRS: engineered synthetase for the incorporation of PCK

Parental protein/backbone: *M. barkeri* PylS

5 Mutations: M241F, A267S, Y271C, L274M

Synthetases with the same substrate specificities can be obtained by transplanting these mutations into *M. mazei* PylS. Thus the following synthetases may be generated by transplantation of the mutations from the Mb backbone onto the Mm tRNA backbone:

10 Mm AcKRS introducing mutations L301V, L305I, Y306F, L309A, C348F into *M. mazei* PylS,

and

Mm PCKRS introducing mutations M276F, A302S, Y306C, L309M into *M. mazei* PylS.

15 Full length sequences of these exemplary transplanted mutation synthetases are given below.

>Mb_PylS/1-419

MDKKPLDVLISATGLWMSRTGTLHKIKHHEVSRSKIYIEMACGDHLVVNNSRSCRTA
20 RAFRHHKYRKTCKRCRVSDDEDINNFLTRSTESKNSVKVRVVSAPKVKKAMPKSVSRAP
KPLENSVSAKASTNTSRVSPSPAKSTPNSSVPASAPAPSLTRSQDRVEALLSPEDKISL
NMAKPFRELEPELVTRRKNDFFQRLYTNDREDYLGKLERDITKFFVDRGFLEIKSPILIP
AEYVERMGINNDTELSKQIFRVDKNLCLRPMLAPTLYNYLRKLDRLPGPIKIFEVGPC
YRKESDGKEHLEEFMTMVNFCQMSGCTRENLEALIKEFLDYLEIDFEIVGDSCMVYGD
25 TLDIMHGDLELSSAVVGPVSLDREWIDKPGWAGFGLERLLKVMHGFKNIKRASRS
ESYNGISTNL

>Mb_AcKRS/1-419

MDKKPLDVLISATGLWMSRTGTLHKIKHHEVSRSKIYIEMACGDHLVVNNSRSCRTA
30 RAFRHHKYRKTCKRCRVSGEDINNFLTRSTESKNSVKVRVVSAPKVKKAMPKSVSRAP
KPLENSVSAKASTNTSRVSPSPAKSTPNSSVPASAPAPSLTRSQDRVEALLSPEDKISL
NMAKPFRELEPELVTRRKNDFFQRLYTNDREDYLGKLERDITKFFVDRGFLEIKSPILIP
AEYVERMGINNDTELSKQIFRVDKNLCLRPMVAPTIFNYARKLDRLPGPIKIFEVGPC
YRKESDGKEHLEEFMTMVNFFQMSGCTRENLEALIKEFLDYLEIDFEIVGDSCMVYGD
35 TLDIMHGDLELSSAVVGPVSLDREWIDKPGWAGFGLERLLKVMHGFKNIKRASRS
ESYNGISTNL

>Mb_PCKRS/1-419

MDKKPLDVLISATGLWMSRTGTLHKIKHHEVSRSKIYIEMACGDHLVVNNSRSCRTA
 RAFRHHKYRKTCKRCRVSEDEDINNFLTRSTESKNSVKVRVVSAPKVKKAMPKSVSRAP
 KPLENSVSAKASTNTSRVSPSPAKSTPNSSVPASAPAPSLTRSQDRVEALLSPEDKISL
 5 NMAKPFRELEPELVTRRKND FQRLYTNDREDYLGKLERDITKFFVDRGFLEIKSPILIP
 AEYVERFGINNDTELSKQIFRVDKNLCLRPMLSPTLCNYMRKLDRLPGPIKIFEVGPC
 YRKESDGKEHLEEFMTMVNFCQMGSGCTRENLEALIKEFLDYLEIDFEIVGDSCMVYGD
 TLDIMHGDLELSSAVVGPVSLDREW GIDKPWIGAGFGLERLLKVMHGFKNIKRASRS
 ESYNGISTNL

10

>Mm_PylS/1-454

MDKKPLNTLISATGLWMSRTGTIHKIKHHEVSRSKIYIEMACGDHLVVNNSRSSRTAR
 ALRHHKYRKTCKRCRVSEDEDLNKFLTKANEDQTSVKVKVVSAPTRTKKAMPKSVARA
 PKPLENTEAAQAQPSGSKFSPAIPVSTQESVSPASVSTSISSISTGATASALVKGNTNPI
 15 TSMSAPVQASAPALTKSQTDRLEVLLNPKDEISLNSGKPFRELESELLSRRKKDLQQIY
 AEERENYLGKLEREITRFFVDRGFLEIKSPILIPLEYIERMGIDNDTELSKQIFRVDKNF
 CLRPMLAPNLYNLRKLDRALPDPIKIFEIGPCYRKESDGKEHLEEFMTMLNFCQMGSG
 CTRENLESIITDFLNHLGIDFKIVGDSCMVYGDTLDVMHGDLELSSAVVGPIPLDREW
 GIDKPWIGAGFGLERLLKVKHDFKNIKRAARSESYNGISTNL

20

>Mm_AcKRS/1-454

MDKKPLNTLISATGLWMSRTGTIHKIKHHEVSRSKIYIEMACGDHLVVNNSRSSRTAR
 ALRHHKYRKTCKRCRVSEDEDLNKFLTKANEDQTSVKVKVVSAPTRTKKAMPKSVARA
 PKPLENTEAAQAQPSGSKFSPAIPVSTQESVSPASVSTSISSISTGATASALVKGNTNPI
 25 TSMSAPVQASAPALTKSQTDRLEVLLNPKDEISLNSGKPFRELESELLSRRKKDLQQIY
 AEERENYLGKLEREITRFFVDRGFLEIKSPILIPLEYIERMGIDNDTELSKQIFRVDKNF
 CLRPVAPNIFNYARKLDRALPDPIKIFEIGPCYRKESDGKEHLEEFMTMLNFFQMGSG
 CTRENLESIITDFLNHLGIDFKIVGDSCMVYGDTLDVMHGDLELSSAVVGPIPLDREW
 GIDKPWIGAGFGLERLLKVKHDFKNIKRAARSESYNGISTNL

30

>Mm_PCKRS/1-454

MDKKPLNTLISATGLWMSRTGTIHKIKHHEVSRSKIYIEMACGDHLVVNNSRSSRTAR
 ALRHHKYRKTCKRCRVSEDEDLNKFLTKANEDQTSVKVKVVSAPTRTKKAMPKSVARA
 PKPLENTEAAQAQPSGSKFSPAIPVSTQESVSPASVSTSISSISTGATASALVKGNTNPI
 35 TSMSAPVQASAPALTKSQTDRLEVLLNPKDEISLNSGKPFRELESELLSRRKKDLQQIY
 AEERENYLGKLEREITRFFVDRGFLEIKSPILIPLEYIERFGIDNDTELSKQIFRVDKNFC
 LRPMLSPNLCNYMRKLDRALPDPIKIFEIGPCYRKESDGKEHLEEFMTMLNFCQMGSGC

TRENLESIITDFLNHLGIDFKIVGDSCMVYGDTLDVMHGDLELSSAVVGPIPLDREWGI
DKPWIGAGFGLERLLKVKHDFKNIKRAARSESYNGISTNL

The same principle applies equally to other mutations and/or to other backbones.

5

Transplanted polypeptides produced in this manner should advantageously be tested to ensure that the desired function/substrate specificities have been preserved.

10

Polynucleotides encoding the polypeptide of interest for the method described above can be incorporated into a recombinant replicable vector. The vector may be used to replicate the nucleic acid in a compatible host cell. Thus in a further embodiment, the invention provides a method of making polynucleotides of the invention by introducing a polynucleotide of the invention into a replicable vector, introducing the vector into a compatible host cell, and growing the host cell under conditions which bring about replication of the vector. The vector may be recovered from the host cell. Suitable host cells include bacteria such as *E. coli*.

15

Preferably, a polynucleotide of the invention in a vector is operably linked to a control sequence that is capable of providing for the expression of the coding sequence by the host cell, i.e. the vector is an expression vector. The term "operably linked" means that the components described are in a relationship permitting them to function in their intended manner. A regulatory sequence "operably linked" to a coding sequence is ligated in such a way that expression of the coding sequence is achieved under condition compatible with the control sequences.

20

Vectors of the invention may be transformed or transfected into a suitable host cell as described to provide for expression of a protein of the invention. This process may comprise culturing a host cell transformed with an expression vector as described above under conditions to provide for expression by the vector of a coding sequence encoding the protein, and optionally recovering the expressed protein.

25

The vectors may be for example, plasmid or virus vectors provided with an origin of replication, optionally a promoter for the expression of the said polynucleotide and optionally a regulator of the promoter. The vectors may contain one or more selectable marker genes, for example an ampicillin resistance gene in the case of a bacterial plasmid. Vectors may be used, for example, to transfect or transform a host cell.

30

Control sequences operably linked to sequences encoding the protein of the invention include promoters/enhancers and other expression regulation signals. These control sequences may be selected to be compatible with the host cell for which the expression vector is designed to be used in. The term promoter is well-known in the art and

35

encompasses nucleic acid regions ranging in size and complexity from minimal promoters to promoters including upstream elements and enhancers.

Another aspect of the invention is a method, such as an *in vitro* method, of incorporating the cyclopropene containing amino acid(s) genetically and site-specifically into the protein of choice, suitably in a eukaryotic cell. One advantage of incorporating genetically by said method is that it obviates the need to deliver the proteins comprising the cyclopropene amino acid into a cell once formed, since in this embodiment they may be synthesised directly in the target cell. The method comprises the following steps:

- i) introducing, or replacing a specific codon with, an orthogonal codon such as an amber codon at the desired site in the nucleotide sequence encoding the protein
- ii) introducing an expression system of orthogonal tRNA synthetase/tRNA pair in the cell, such as a pyrrolysyl-tRNA synthetase/tRNA pair
- iii) growing the cells in a medium with the cyclopropene containing amino acid according to the invention.

Step (i) entails or replacing a specific codon with an orthogonal codon such as an amber codon at the desired site in the genetic sequence of the protein. This can be achieved by simply introducing a construct, such as a plasmid, with the nucleotide sequence encoding the protein, wherein the site where the cyclopropene containing amino acid is desired to be introduced/replaced is altered to comprise an orthogonal codon such as an amber codon. This is well within the person skilled in the art's ability and examples of such are given here below.

Step (ii) requires an orthogonal expression system to specifically incorporate the cyclopropene containing amino acid at the desired location (e.g. the amber codon). Thus a specific orthogonal tRNA synthetase such as an orthogonal pyrrolysyl-tRNA synthetase and a specific corresponding orthogonal tRNA pair which are together capable of charging said tRNA with the cyclopropene containing amino acid are required. Examples of these are provided herein.

Protein Expression and Purification

Host cells comprising polynucleotides of the invention may be used to express proteins of the invention. Host cells may be cultured under suitable conditions which allow expression of the proteins of the invention. Expression of the proteins of the invention may be constitutive such that they are continually produced, or inducible, requiring a stimulus to initiate expression. In the case of inducible expression, protein production

can be initiated when required by, for example, addition of an inducer substance to the culture medium, for example dexamethasone or IPTG.

Proteins of the invention can be extracted from host cells by a variety of techniques
5 known in the art, including enzymatic, chemical and/or osmotic lysis and physical disruption.

Proteins of the invention can be purified by standard techniques known in the art such as preparative chromatography, affinity purification or any other suitable technique.
10

FURTHER ADVANTAGES

Yu *et al* join tetrazoles to cyclopropene amino acids in polypeptides. Yu *et al* require the use of ultraviolet irradiation in order to photoactivate their conjugation groups.
15 Their best reaction rates were achieved with 302 nano metres UV irradiation. However, this type of UV irradiation has high ionisation potential. This means that the molecules and/or cells upon which the radiation is directed are likely to be damaged by this UV energy. By contrast, the conjugations of the present invention do not require any UV step for photoactivation. Even when Yu *et al* use a less damaging source of UV
20 irradiation (eg. 365 nano metre UV irradiation), the observed reaction rates are considerably slower than those provided by the present invention. Thus, even if the UV irradiation is adjusted in Yu *et al* in an attempt to try to avoid or reduce some of the drawbacks associated with UV treatment, the same laborious irradiation step must still be carried out and slower reaction rates are achieved. It is an advantage of the present
25 invention that UV irradiation can be omitted, and that excellent reaction rates are obtained even without photoactivation.

It is an advantage of the cyclopropene amino acids of the present invention that they are easy to manufacture. For example, the number steps in the synthetic pathway is
30 advantageously few.

It should be noted that the prior art cyclopropene amino acid of Yu *et al* contains an amide group. This amide bond is a potential substrate for peptidases. Peptidase action on the amide bond of the prior art cyclopropene amino acid would cleave the
35 cyclopropene part of the molecule off the polypeptide. This is clearly a disadvantage. By contrast, it is an advantage of the carbamate linked cyclopropene groups of the present invention that carbamate bonded cyclopropene is not a target for peptidases.

Prior art based techniques rely on tetrazole chemistry for conjugation. In contrast, the present invention teaches the use of advantageous tetrazine chemistry.

5 It is an advantage of the carbamate bonded cyclopropene amino acids of the present invention that they enable the use of the wild type PylRS synthetase. Making use of the wild type synthetase is advantageous as it involves less labour by alleviating the need to prepare mutant synthetases. In addition, the mutant synthetases do not always amino acylate in tRNA to the same level as wild type tRNA synthetases. In other words, the
10 mutations required to be made to a synthetase in order to handle prior art cyclopropene amide bonded amino acids can cause a loss of efficiency of amino acylation. In contrast, it is demonstrated herein that amino acylation using the wild type synthetase with the amino acid of the present invention is a very efficient process, which is a further advantage over prior art techniques.

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

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

25 BRIEF DESCRIPTION OF THE DRAWINGS

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

Figure 1 SORT-M enables proteome tagging and labelling at diverse codons, with
30 diverse chemistries, and in genetically targeted cells and tissues. **(a)** Proteome tagging via SORT (stochastic orthogonal recoding of translation) uses an orthogonal aminoacyl-tRNA synthetase/tRNA pair. The pyrrolysyl-tRNA synthetase/tRNA pair is used in this study. This synthetase (and its previously evolved active-site variants) recognizes a range of unnatural amino acids (yellow star, and yellow hexagon), does not
35 aminoacylate endogenous tRNAs, but efficiently aminoacylates its cognate tRNA - without regard to anticodon identity; PyltRNA is not a substrate for endogenous aminoacyl-tRNA synthetases. Orthogonal pyrrolysyl-tRNA synthetase/tRNA_{xxx} pairs

(XXX indicates choice of anticodon, yellow) in which the anticodon has been altered compete for the decoding of sense codons (dark blue and pink) via a pathway that is orthogonal to that used by natural synthetases and tRNAs (dark blue and pink) to direct natural amino acids. SORT allows the incorporation of diverse chemical groups into the proteome, in response to diverse codons. Since there is no competition at the active site of the orthogonal synthetase, starvation and minimal media are not required. In addition the expression pattern of the orthogonal proteome tagging system can be genetically directed allowing tissue specific proteome labelling. Selective pressure incorporation approaches are shown in **Supplementary Fig. 1** for comparison to SORT. **(b)** The combination of encoding amino acids (**1-3**) across the proteome *via* SORT and chemoselective modification of **3** with tetrazine probes (**4a-g**, **5**, **6** and **7**) allows detection of labelled proteins *via* SORT-M (stochastic orthogonal recoding of translation and chemoselective modification). Amino acid structures: *N*^ε-((tert-butoxy)carbonyl)-L-lysine **1**, *N*^ε-(1-propynlyoxy)carbonyl)-L-lysine **2** and *N*^ε-(((2-methylcycloprop-2-en-1-yl)methoxy)carbonyl)-L-lysine.

Figure 2 (Supplementary Figure 2) shows Quantitative site-specific incorporation of 3 into proteins expressed in *E. coli* and its rapid and quantitative labelling with tetrazine probes

- A.** The PylRS/tRNA_{CUA} pair directs efficient, site-specific incorporation of **3** into sfGFP bearing an amber stop codon at position 150. Incorporation of **3** is more efficient than **1** a well-established excellent substrate for the PylRS/tRNA_{CUA} pair.
- B.** Specific and quantitative labelling of 2 nmol sfGFP bearing **3** with 10 equivalents of tetrazine fluorophore **4a**. ESI-MS analysis of sfGFP-**3** purified from *E. coli* grown with 1 mM **3** bearing the PylRS/tRNA_{CUA} pair and SfGFP150TAG confirms the incorporation of **3**. sfGFP150-**3**: Expected mass: 27951.5 Da, Found mass: 27950 ± 1.0 Da, minor peak 27820 corresponding to loss of N-terminal methionine. Labelling sfGFP150-**3** with **4a** is quantitative, as judged by ESI-MS of the labelling reaction. Expected mass: 28758.4 Da, Found mass: 28758 ± 1.0 Da, minor peak 28627 corresponds to loss of N-terminal methionine.
- C.** Determining the rate constant for labelling of sfGFP-**3** (10.6 μM, sfGFP incorporating **3** at position 150), with 10 equivalents of **4a**. 2 nmol of purified sfGFP-**3**, (10.6 μM in 20 mM Tris-HCl, 100 mM NaCl, 2 mM EDTA, pH 7.4) were incubated with 20 nmol of tetrazine-dye conjugate **4a** (10 μl of a 2 mM solution in DMSO). At different time points 8 μL aliquots were taken from the solution and quenched with a 700-fold excess of BCN and plunged into liquid nitrogen. Samples were mixed with NuPAGE LDS sample buffer supplemented with 5 % β-mercaptoethanol, heated for 10 min to

90°C and analyzed by 4-12% SDS page. The amounts of labelled proteins were quantified by scanning the fluorescent bands with a Typhoon Trio phosphorimager (GE Life Sciences). Bands were quantified with the ImageQuant™ TL software (GE Life Sciences) using rubber band background subtraction. The rate constant was
 5 determined by fitting the data to a single-exponential equation. The calculated observed rate k' was divided by the concentration of **4a** to obtain rate constant k for the reaction. Measurements were done in triplicate. All data processing was performed using Kaleidagraph software (Synergy Software, Reading, UK). For comparison the rate of labelling sfGFP bearing N ϵ -5-norbornene-2-yloxycarbonyl-L-lysine (NorK), a known
 10 substrate for PylRS, was determined in a similar way using 11.25 μ M sfGFP bearing NorK at position 150 (SfGFP-NorK) and 20 equivalents of **4a**.

Figure 3 shows Supplementary Table 1 - Primers

15 **Figure 4 (Supplementary Figure 3) shows SORT-M enables codon specific proteome tagging and labelling in *E. coli***

A. Proteome labelling with **3** via the indicated PylRS/tRNA_{xxx} pair. Cells contained two plasmids, one encoding MbPylRS, the other encoding T4 lysozyme and the indicated tRNA_{xxx}. Cells were grown in the presence of 0.1 mM **3** from OD₆₀₀=0.2 and T4
 20 lysozyme expression, induced by the addition of 0.2 mM arabinose after 1h. After a further 3 h cells were harvested. Tagged proteins in the lysate were detected via an inverse electron demand Diels-Alder reaction between incorporated **3** and tetrazine fluorophore **4a** (20 mM, 1h, RT). The amino acids in parentheses are the natural amino acids encoded by the endogenous tRNA bearing the corresponding anti-codon.

25 **B.** Lane profile analysis for each codon.

Figure 5 (Supplementary Figure 4) shows Specific amino acid replacement in SORT demonstrated by ESI-MS

T4 lysozyme isolated after SORT with UUU(Lys) in the presence of 1mM **3**. Expected
 30 mass WT T4 lysozyme: 19512.2 Da, Found mass: 19510 $\pm$ 2.0 Da. Expected mass WT T4 lysozyme Lys→**3** single mutation: 19622.3 Da, Found mass: 19620 $\pm$ 2.0 Da.

Figure 6 (Supplementary Figure 5) shows Incorporation of **3 (0.1 mM) via SORT-M is not toxic to cells**

35 Chemically competent DH10B cells were transformed with two plasmids: pBKwtPylRS necessary for expression of PylRS, and pBAD_wtT4L_MbPylT_{xxx} plasmids that is required for expression of PyltRNA_{xxx} and expresses lysozyme under arabinose control.

The cells were recovered in 1 ml SOB medium for one hour at 37°C prior to aliquoting to 10 ml LB-KT (LB media with 50 µg ml⁻¹ kanamycin, and 25 µg ml⁻¹ tetracycline) and incubated overnight (37°C, 250 rpm, 12 h). The overnight culture (OD₆₀₀≈3) was diluted to a OD₆₀₀~0.3 in 10 mL LB-KT_{1/2} (LB media with 25 µg ml⁻¹ kanamycin, and 12.5 µg ml⁻¹ tetracycline) supplemented with **3** at different concentrations, 0, 0.1, 0.5 mM. 200 µL aliquots of these cultures were transferred into a 96-well plate and OD₆₀₀ measured using a Microplate reader, Infinite 200 Pro (TECAN). OD₆₀₀ was measured for each sample every 10 min with linear 1 mm shaking between the measurements.

Figure 7 (Supplementary Figure 6) shows Measurement of time-dependent variation in incorporation of **3 in proteome via SORT-M at different concentrations of **3** in response to AAA codon**

Chemically competent DH10B cells were transformed with two plasmids: pBKwtPylRS necessary for expression of PylRS, and pBAD_wtT4L_MbPylT_{UUU} plasmid that is required for expression of PyltRNA_{UUU}. pBAD_wtT4L_MbPylT_{UUU} plasmid also contains the gene for expression of T4 lysozyme that is downstream of arabinose-inducible promoter. After transformation, cells were recovered in 1 ml SOB medium for one hour at 37°C prior to inoculation in 10 ml LB-KT (LB media with 50 µg ml⁻¹ kanamycin, and 25 µg ml⁻¹ tetracycline). The culture was incubated overnight (37°C, 250 rpm, 12 h) and subsequently diluted to an OD₆₀₀~0.3 in 30 mL LB-KT_{1/2} (LB media with 25 µg ml⁻¹ kanamycin, and 12.5 µg ml⁻¹ tetracycline) supplemented with **3** at different concentrations, 0, 0.1, 0.5 mM. The cultures was incubated (37°C, 250 rpm) for 1 h, when OD₆₀₀ reached approximately 0.6. 2 ml culture aliquot was collected in a separate tube for each of three cultures. This is the pre-induction culture (lane labelled as1 in the gel image). Subsequently arabinose was added at a final concentration of 0.2% (v/v) to induce expression of T4 lysozyme and culture aliquots of 2 mL were collected every hour (lanes labelled as 2, 3 and 4 corresponding to 1, 2 and 3h culture collection after induction). For each of the collected cultures, bacterial cells were pelleted by centrifugation at 4 °C, washed with ice cold PBS (3 x 1 mL) and subsequently the pellets were frozen and stored at -20 °C. The pellets were then thawed in 200 µL of ice cold PBS and lysed by sonication (9 x 10 s ON / 20 s OFF, 70% power). The lysates were clarified by centrifugation at 15,000 RPM, 4 °C for 30 minutes. The supernatants were transferred to fresh 1.5 mL tubes. 50 µL of supernatant was transferred to a new tube for the labeling reactions, and the rest was frozen in liquid nitrogen and stored at -80C. To the 50 µL of supernatant, 0.5 µL of 2 mM **4a** was added and the lysates were incubated at 25°C for 1 hour. After 1h, 17 µL of 4X LDS sample buffer supplemented (6mM BCN and 5% BME) was added and mixed by vortexing gently. Samples were incubated for 10min before boiling at 90 °C for 10 min.

Samples were analysed by 4-12% SDS-PAGE and fluorescent images were acquired using Typhoon Trio phosphoimager (GE Life Sciences)

Figure 8 shows Site-specific incorporation of **3** into proteins at diverse codons and specific proteome labelling using SORT-M in human cells. **(a)** Western blot analysis demonstrates the efficient amino acid dependant expression of an mCherry-EGFP fusion protein separated by an amber stop codon bearing a C-terminal HA-tag (mCh-TAG-EGFP-HA) in HEK293T cells. Anti-FLAG detected tagged PylRS **(b)** Specific labelling of mCh-TAG-EGFP-HA (immunoprecipitated from 10^6 cells) with **4a** (20 μ M in 50 μ L PBS, 1h, RT) confirms the incorporation of **3** into protein in HEK293 cells. **(c)** SORT-M labelling of **3** that is statistically incorporated into newly synthesised proteins across the whole proteome of mammalian cells directed by six different PylRS/PyltRNA_{xxx} mutants using 0.5 mM **3**. Labeling with **4g** (20 μ M in PBS, 1h, RT, as above). The amino acids in parentheses are the natural amino acids encoded by the endogenous tRNA bearing the corresponding anti-codon.

Figure 9 (Supplementary Figure 11) shows

A. Full blots from **Figure 8**.

B. Full blots from **Figure 10**.

20

Figure 10 shows Site-specific incorporation of amino acid **3** into protein produced in *Drosophila melanogaster*. **(a)** Incorporation of **3** demonstrated by a dual luciferase reporter. Dual luciferase assay on ovary extract from 10 female flies expressing Triple-Rep-L in the presence or absence of 10 mM **1** or 10mM **3**. The data show a representative example from 1 of 3 biological replicates. The error bars represent the standard deviation of 3 technical replicates from a single biological replicate. **(b)** Site-specific incorporation of **3** (or **1**) into GFP_TAG_mCherry-HA in flies expressing PylRS/PyltRNA_{CUA}. The full-length protein resulting from unnatural amino acid incorporation is detected by anti-HA western blot. **(c)** Specific labelling of encoded **3** with tetrazine probes. Flies were fed with no amino acid, amino acid **1** (500 flies) or amino acid **3** (100 flies). 5 times more flies were fed with **1** in order to generate comparable amount of reporter protein. The full-length protein containing the unnatural amino acid was immunoprecipitated from lysed ovaries with anti-GFP beads. The beads were labelled (**4g**, 4 μ M, 200 μ L PBS, RT, 2h) washed. Full length protein was detected by anti-HA blot and the same gel imaged on a fluorescence scanner shows specific fluorescent labelling of the protein incorporating **3** but not **1**, confirming the identity of the incorporated amino acid.

30

35

Figure 11 (example 6) Specific protein labeling at genetically encoded unnatural amino acids **1** and **2**. (a) Genetically encoded **1**, but not **2**, in calmodulin is specifically labeled with probe **3**. Coomassie and fluorescence images demonstrate the specificity of labeling and ESI MS before labelling (black, expected mass: 17875, found mass: 17874) and after labelling (red, expected mass: 18553, found mass: 18552) demonstrate the reaction is quantitative. (b) Genetically encoded **2**, but not **1**, in calmodulin is specifically labeled with probe **4**. Coomassie and fluorescence images demonstrate the specificity of labeling and ESI MS before labeling (black, expected mass: 17930, found mass: 17930) and after labelling (green, expected mass: 18484, found mass: 18485) demonstrate the reaction is quantitative. Raw (before deconvolution) ESI-MS spectra are not shown.

Figure 12 (example 6) Incorporating **1** and **2** at positions 1 and 40 of Calmodulin and the kinetics of specific labelling. (a) Expression was performed in *E. coli* bearing ribo-Q1, *O-gst-cam1TAG-40AGTA*, the PylRS/tRNA^{UACU} pair and the *M/PrpRS/tRNA^{CUA}* pair. Amino acids **1** and **2** were used at 4 and 1 mM, respectively. (b) Labelling time course for reaction of CaM_{1,240} with **3** and **4**. Each reaction was followed for 2h by in gel fluorescence and mobility shift.

Figure 13 (example 6) Concerted, quantitative one-pot, dual labeling of Calmodulin in 30 minutes. (a) Dye dependent labeling of CaM_{1,240}; sequential labeling with purification after first labeling in lane 4, sequential labeling without purification in lane 5, one-pot dual labeling in lane 6. (b) ESI-MS of one-pot protein labeling, before labeling (black, expected mass: 18000 found mass: 18000), after labeling (gold, expected mass: 19233 found mass: 19234). Raw (before deconvolution) ESI-MS spectra are not shown.

Figure 14 shows Scheme A

Figure 15 (supplementary figure 15) shows Amino acid and DNA sequence of *Drosophila* GFP-amber-mCherry-HA.

GFP (amino acid residues 1-238), Amber codon at position 248, mCherry (amino acid residues 255-489), HA tag (amino acid residues 491-499), Myc tag (amino acid residues 500-509), His tag (amino acid residues 510-515) and SV40 NLS (amino acid residues 523-528).

Figure 16 shows structure of exemplary amino acid *N*^ε-[((2-methylcycloprop-2-en-1-yl)methoxy)carbonyl]-l-lysine.

EXAMPLES - DESCRIPTION OF THE EMBODIMENTS

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

Chemical syntheses - general methods

All chemicals and solvents were purchased from Sigma-Aldrich, Alfa Aesar or Fisher Scientific and used without further purification unless otherwise stated. Qualitative

analysis by thin layer chromatography (TLC) was performed on aluminium sheets coated with silica (Merck TLC 60F-254). The spots were visualized under short wavelength ultra-violet lamp (254nm) or stained with basic, aqueous potassium permanganate, ethanolic ninhydrin or vanillin. Flash column chromatography was performed with specified solvent systems on silica gel 60 (mesh 230-400).

LC-MS analysis was performed on Agilent 1200 machine. The solvents used consisted of 0.2 % formic acid in water (buffer A) and 0.2 % formic acid in acetonitrile (buffer B). LC was performed using Phenomenex Jupiter C18 column (150 x 2 mm, 5µm) and monitored using variable wavelengths. Retention times (R_t) are recorded to a nearest 0.1 min and m/z ratio to nearest 0.01 mass units. The following programme was used for small molecule LC gradient: 0-1 min (A:B 10:90-10:90, 0.3 mL/min), 1-8 min (A:B 10:90-90:10, 0.3 mL/min), 8-10 min (A:B 90:10-90:10, 0.3 mL/min), 10-12 (A:B 90:10-10:90, 0.3 mL/min).

Mass spectrometry analysis following LC was carried out in ESI mode on a 6130 Quadrupole spectrometer and recorded in both positive and negative ion modes. NMR analysis was carried out on a Bruker 400MHz instrument. All reported chemical shifts (δ) relative to TMS were referenced to the residual protons in deuterated solvents used: d_1 – chloroform (^1H δ = 7.26 ppm, ^{13}C δ = 77.16 ppm), d_6 – dimethylsulfoxide (^1H δ = 2.49 ppm, ^{13}C δ = 39.52 ppm), D_2O (^1H δ = 4.70). APT or two-dimensional experiments (COSY, HSQC) were always performed to provide additional information used for analysis where needed. Coupling constants are given in Hz and described as: singlet – s, doublet – d, triplet – t, quartet – q, broad singlet – br, multiplet – m, doublet of doublets – dd, etc. and combinations thereof.

25

Protein expression, purification and labelling of site-specifically incorporated **3 in *E. coli***

Expression and purification of sfGFP-3 from E. coli Electrocompetent *E. coli* DH10B cells were co-transformed with pBK-MbPylRS and psfGFP150TAG PylT^{14, 26}. Transformed cells were recovered in S.O.B. (1 mL, supplemented with 0.2% glucose) for 1 h at 37 °C and used to inoculate LB containing 50 µg/mL kanamycin and 25 µg/mL tetracycline (LB-KT). The cells were incubated with shaking overnight at 37 °C, 250 r.p.m. 1 mL of overnight culture was used to inoculate 100 mL of LB-KT½, the day culture was then incubated (37 °C, 250 r.p.m). At $\text{O.D.}_{600} \sim 0.3$, the culture was divided equally and supplemented with either **3** (1 mM) or H_2O (500 µL) and incubated further (37 °C, 250 r.p.m). At $\text{O.D.}_{600} \sim 0.6$ protein expression was

35

induced by the addition of arabinose (0.2%), after 4 h, the cells were harvested by centrifugation (4000 r.p.m, 20 min) and the pellet frozen until further use.

The frozen bacterial pellet was thawed on ice and resuspended in 2.5 mL lysis buffer (Bugbuster®, Novagen®, 50 µg/mL DNase 1, Roche inhibitor cocktail and 20 mM imidazole). Cells were incubated (4 °C, 30 minutes) then clarified by centrifugation (16000 g, 4 °C, 30 minutes). The clarified lysates were transferred to fresh tubes and 100 µL Ni-NTA slurry added. The mixtures was incubated with agitation (4 °C, 1 h) and then collected by centrifugation (1000 g, 4 °C, 5 min). The beads were resuspended three times in 500 µL wash buffer (10 mM Tris-HCL, 40 mM imidazole, 200 mM NaCl, pH 8) and collected by centrifugation (1000 g, 4 °C, 5 min). Finally, the beads were resuspended in 100 µL elution buffer (10 mM Tris-HCL, 300 mM imidazole, 200 mM NaCl, pH 8), pelleted by centrifugation (1000 g, 4 °C, 5 min) and the supernatant collected into fresh tubes. The elution was repeated three times with 100 µL of elution buffer. The purified proteins were analysed by 4-12% SDS-PAGE and LC-MS.

Protein Mass Spectrometry

Using an Agilent 1200 LC-MS system, ESI-MS was additionally carried out with a 6130 Quadrupole spectrometer. The solvent system consisted of 0.1 % formic acid in H₂O as buffer A, and 0.1 % formic acid in acetonitrile (MeCN) as buffer B. Protein UV absorbance was monitored at 214 and 280 nm. Protein MS acquisition was carried out in positive ion mode and total protein masses were calculated by deconvolution within the MS Chemstation software (Agilent Technologies).

In vitro labeling of purified sfGFP150-3

To Purified sfGFP150-1 or sfGFP150-3 protein (~30 µM, in elution buffer) was added **4a** (10 molar equivalents, from a 2 mM stock solution in DMSO). The reactants were mixed by aspirating several times and the mixture then incubated at room temperature for 2 hours, a sample was analysed by ESI-MS. Following incubation the proteins were separated by 4-12% SDS-PAGE and analysed by using Typhoon Trio phosphoimager (GE Life Sciences).

Time course of sfGFP150-3 and sfGFP150-NorK labelling and rate constant determination

2 nmol sfGFP-3 (10.6 µM) was labeled at room temperature by the addition of 20 nmol of tetrazine-dye conjugate **4a** (10µl of a 2 mM solution in DMSO) the samples were mixed by

aspirating several times. At different time points, 8 μ L aliquots were taken from the solution and quenched with a 700-fold excess of bicyclo[6.1.0]non-4-yn-9-ylmethanol (BCN) and plunged into liquid nitrogen. Samples were mixed with NuPAGE LDS sample buffer supplemented with 5 % β -mercaptoethanol, heated for 10 min to 90°C and analyzed by 4-12% SDS page. The

5 amounts of labelled proteins were quantified by scanning the fluorescent bands with a Typhoon Trio phosphoimager (GE Life Sciences). Bands were quantified with the ImageQuantTM TL software (GE Life Sciences) using rubber band background subtraction. The rate constant was determined by fitting the data to a single-exponential equation. The calculated observed rate k' was divided by the concentration of **4a** to obtain rate constant k for the reaction. Measurements

10 were done in triplicate. All data processing was performed using Kaleidagraph software (Synergy Software, Reading, UK). For comparison the rate of labelling sfGFP bearing Ne-5-norbornene-2-yloxycarbonyl-L-lysine (NorK), a known substrate for PylRS, was determined in a similar way using 11.25mM sfGFP bearing NorK at position 150 (SfGFP-NorK) and 20 equivalents of **4a**.

15 *Plasmid construction for pBAD_wtT4L_MbPylT_{XXX}*

pBAD_T4L83TAG_MbPylT_{CUA} was digested with *NcoI* and *KpnI* restriction enzymes. The same restriction enzymes were also used to digest the wild-type T4 lysozyme from (D67) pBAD_wtT4L. The insert and backbone were ligated in 3:1 ratio using T4 DNA ligase (RT, 2 hours), transformed into chemically competent DH10B cells and grown on Tetracycline agar

20 plates (37°C, 18 hours). Single colonies were picked and the correct sequence was confirmed by DNA sequencing (GATC GmbH.), this step created pBAD_wtT4L_MbPylT_{CUA}. All final constructs were confirmed by DNA sequencing.

Proteomic incorporation of 3 via SORT in E. coli expressing T4 lysozyme

25 Electrocompetent *E. coli* DH10B cells (50 μ L) were either doubly transformed with pBAD_wtT4L_MbPylT_{XXX} plasmid (2 μ L, necessary for expression of PyltRNA_{XXX} and expresses T4 lysozyme under arabinose control) and pBKwtPylS plasmid (2 μ L necessary for expression of PylRS) or singly transformed with pBAD_wtT4L_MbPylT_{XXX} alone. Transformed cells were recovered in 1 mL S.O.B. (supplemented with 0.2% glucose) for 1 h at

30 37 °C. 100 μ L of the recovery was used to inoculate 5 mL LB-KT (50 μ g/mL kanamycin and 25 μ g/mL tetracycline) or LB-T (25 μ g/mL tetracycline). Cultures were incubated overnight (37 °C, 250 r.p.m.). 1 mL of each overnight culture was used to inoculate 15 mL ½ strength antibiotic containing media LB-T or LB-KT. Cultures were incubated at 37 °C until O.D.₆₀₀ ~0.3 was reached, at this time each culture was divided into 5 mL aliquots and supplemented

35 with either **3** (0.1 mM final conc.) or H₂O (50 μ L). Cultures were then incubated (37 °C, 250

r.p.m.). At O.D.₆₀₀ 0.6. T4 lysozyme expression was initiated by the addition of arabinose (0.2% final conc.) and cultures incubated for a further 4 hours. Cells were harvested by centrifugation (4000 rpm, 4 °C, 20 minutes) and then resuspended three times in 1 mL of ice cold PBS and collected by centrifugation (4000 rpm, 4 °C, 20 minutes). The final bacterial pellets were
 5 immediately frozen for storage.

E.coli: Chemoselective labelling proteomes tagged with 3 with tetrazine-dye conjugates

Frozen bacterial pellets were resuspended in 500 µL PBS and lysed using a bath
 10 sonicator (energy output 7.0, 90 s total sonication time. 10 s blasts and 20 s breaks, Misonix Sonicator 3000). The lysate was cleared by centrifugation (4 °C, 14000 r.p.m., 30 min) and the supernatant aspirated to a fresh tube. To 50 µL of cleared cell lysate was added **4a** (2 mM, stock in DMSO, final concentration – 20 µM). The reactions were mixed by aspirating several times and the samples then incubated in the dark
 15 (room temperature, 1 h). After this time 17 µL of 4X LDS sample buffer supplemented (6 mM BCN and 5% BME) was added and mixed by vortexing gently. Samples were incubated for 10 min before boiling at 90 °C for 10 min. Samples were analysed by 4-12% SDS-PAGE and fluorescent images were acquired using Typhoon Trio phosphoimager (GE Life Sciences).
 20 The same protocol for fluorescent labelling of the *E. Coli* proteins was applied for all tetrazine-dye conjugates.

Site-specific incorporation of 3 in HEK293 cells and chemoselective labelling with tetrazine probes

25 *Site specific incorporation of 3 in HEK cells*

HEK293 Cells (ATCC CRL-1573) were plated on 24 well plates and grown to near confluence. The cells were transfected using Lipofectamine 2000 (Invitrogen) with the pMmPylS-mCherry-TAG-EGFP-HA construct and the p4CMVE-U6-PylT construct.¹⁸ After 16hrs growth with or without 1mM **3** or with 1mM **1** the cells were lysed on ice
 30 using RIPA buffer (Sigma). The lysates were spun down and the supernatant was added to 4X LDS sample buffer (Life technologies). The samples were run out by SDS-PAGE, transferred to a nitrocellulose membrane and blotted using primary rat anti-HA(clone 3F10, Roche, No. 1 867 423) and mouse anti-FLAG (clone G191, Abnova, cat. MAB8183), the secondary antibodies were anti-rat (Invitrogen, A11077) and anti-
 35 mouse (Cell Signaling Technologies, No. 7076S).

Labelling site-specifically incorporated 3 from HEK 293 cells

Adherent HEK293T cells (ATCC CRL-11268; 4×10^6 per immunoprecipitation) were transfected with 7.5 μ g p4CMVE-U6-PylT and 7.5 μ g pPylRS-mCherry-TAG-EFGP-HA¹⁸ using TransIT-293 transfection reagent according to the manufacturer's protocol and cultured for 48 hours in DMEM/10%FBS, supplemented with 0.5 mM **1** or 2 mM **3** where indicated. Cells were washed twice with PBS and lysed on ice for 30 minutes in 1mL Lysis Buffer (150 mM NaCl, 1% Triton X-100, 50 mM Tris HCl (pH 8.0). After clarifying the lysate by centrifugation (10 min at 16000g), HA-tagged proteins were captured using 50 μ L μ MACS HA-tag MicroBeads (Miltenyl Biotec) per transfection, washed with 0.5 mL RIPA (150 mM NaCl, 1% Igepal CA-630, 0.5% sodium deoxycholate, 0.1% SDS, 50 mM Tris HCl (pH 8.0) and 0.5 mL PBS (pH 7.4). The suspension of MicroBeads was incubated with 50 μ L PBS (pH 7.4), 20 μ M **4a** for 1 hour and subsequently washed with 0.5 mL RIPA to remove excess dye. HA-tagged proteins were eluted from beads using SDS sample buffer and separated on a 4-12% Bis-Tris PAGE gel (Invitrogen), imaged using a Typhoon imager (GE Healthcare) and subsequently stained with DirectBlue or transferred for western blotting with Anti-HA-tag pAb-HRP-Direct (MBL).

Expression and purification of SfGFP from mammalian cells

HEK293T were transfected in a 10cm tissue culture dish with 15 μ g DNA using PEI and incubated for 72 hours with **3** (0.5 μ M). Cells were washed twice with PBS and lysed in 1mL RIPA buffer. Cleared lysate was added to 50 μ L GFP-Trap® M (ChromoTek) and incubated for 4h. Beads were washed with 1mL RIPA, 1mL PBS, 1mL PBS+500mM NaCl, 1mL ddH₂O and eluted in 1%Acetic Acid/ddH₂O. Purified protein was labeled with 2 μ M **4a** for 4h and loaded on a 4-12% Bis-Tris PAGE gel. Fluorescence of **4a**-labeled sfGFP was detected on a Typhoon imager and gel was stained subsequently with DirectBlue.

Fly plasmids, transgenic flies and culture

For all fly experiments no randomisation or blinding was used within this study.

Plasmid construction for transgenic fly line generation

The PyltRNA_{CUA} anticodon was mutated using the QuikChange mutagenesis kit and pSG108 (pJet 1.2-U6-PylT, gift from S. Greiss) as a template. This contains the PylT gene without its 3' terminal CCA fused to the *Drosophila* U6-b promoter. Primers FMT19 and FMT20 were used to generate PyltRNA_{TGC} to decode alanine codons (creating pFT18); primers FMT23 and FMT24 were used to generate PyltRNA_{GCT} to decode serine codons (creating pFT20); primers FMT27 and FMT28 were used to generate PyltRNA_{CAG} to decode leucine codons (creating pFT22) and primers FMT29

and FMT30 were used to generate PyltRNA_{CAT} to decode methionine codons (creating pFT23). The mutated tRNA expression cassettes were subcloned from pFT18, pFT20, pFT22 and pFT23 into pUC18 using EcoRI and HindIII then multimerised using AsiSI, BamHI and BglII to create 2, then 4 copies of the tRNA. The 4 copy versions of the
 5 tRNA cassette were subcloned into pSG118 using AsiSI and MluI to create pFT58 (Ala), pFT60 (Ser), pFT62 (Leu) and pFT63 (Met). pSG118 contains the *M.mazei* PylRS gene.²⁰

Fly lines and culture conditions

10 Transgenic lines were created by P element insertion using a *Drosophila* embryo injection service (BestGene Inc.). Lines were generated using the following plasmids: pFT58 (Ala), pFT60 (Ser), pFT62 (Leu) and pFT63 (Met). nos-Gal4-VP16 (Bloomington 4937) and MS1096-Gal4 (Bloomington 8860) were used as Gal4 drivers. All flies were grown at 25°C on standard Iberian medium. Flies were fed unnatural
 15 amino acids by mixing dried yeast with the appropriate concentration of amino acid (usually 10mM) diluted in dH₂O to make a paste. Ovaries were prepared from females that were grown on Iberian fly food supplemented with a yeast paste with or without the amino acid for a minimum of 48 hours. For proteome labelling experiments transgenic male flies of constructs FT58, FT60, FT62 and FT63 were crossed with nos-
 20 vp16-GAL4 virgins to generate FT58/nos-vp16-GAL4, FT60/nos-vp16-GAL4, FT62/nos-vp16-GAL4 and FT63/nos-vp16-GAL4 respectively.

Site specific incorporation of 3 in D. melanogaster

Luciferase assays

25 Ovaries from 10 females of Triple Rep-L flies recombined with nos-Gal4-VP16 fed **3**, **1** or no amino acid were dissected in 100µl 1x Passive lysis buffer and processed for luciferase assays as previously described ²⁰.

Immunoprecipitation and labelling of site specifically incorporated 3

30 Ovaries from 100 (for control and **3**) or 500 (for **1**) females were dissected in PBS then lysed in 300 or 1500 µl RIPA buffer containing 1x complete protease inhibitor cocktail (Roche). A sample was taken into 4 x LDS buffer as a total lysate control then the remainder was used for immunoprecipitation with GFP-TRAP agarose beads (Chromotek) following the manufacturer's instructions. The total volume of the IP was
 35 3ml. After overnight incubation, the beads were washed 2 x with RIPA buffer then 2 x with PBS. For tetrazine labeling, the beads were resuspended in 200µl PBS + 4µM **4g**

and incubated for 2 hours on a roller at RT. The beads were washed 3 times with 500 μ L of wash buffer then resuspended in 4x LDS sample buffer.

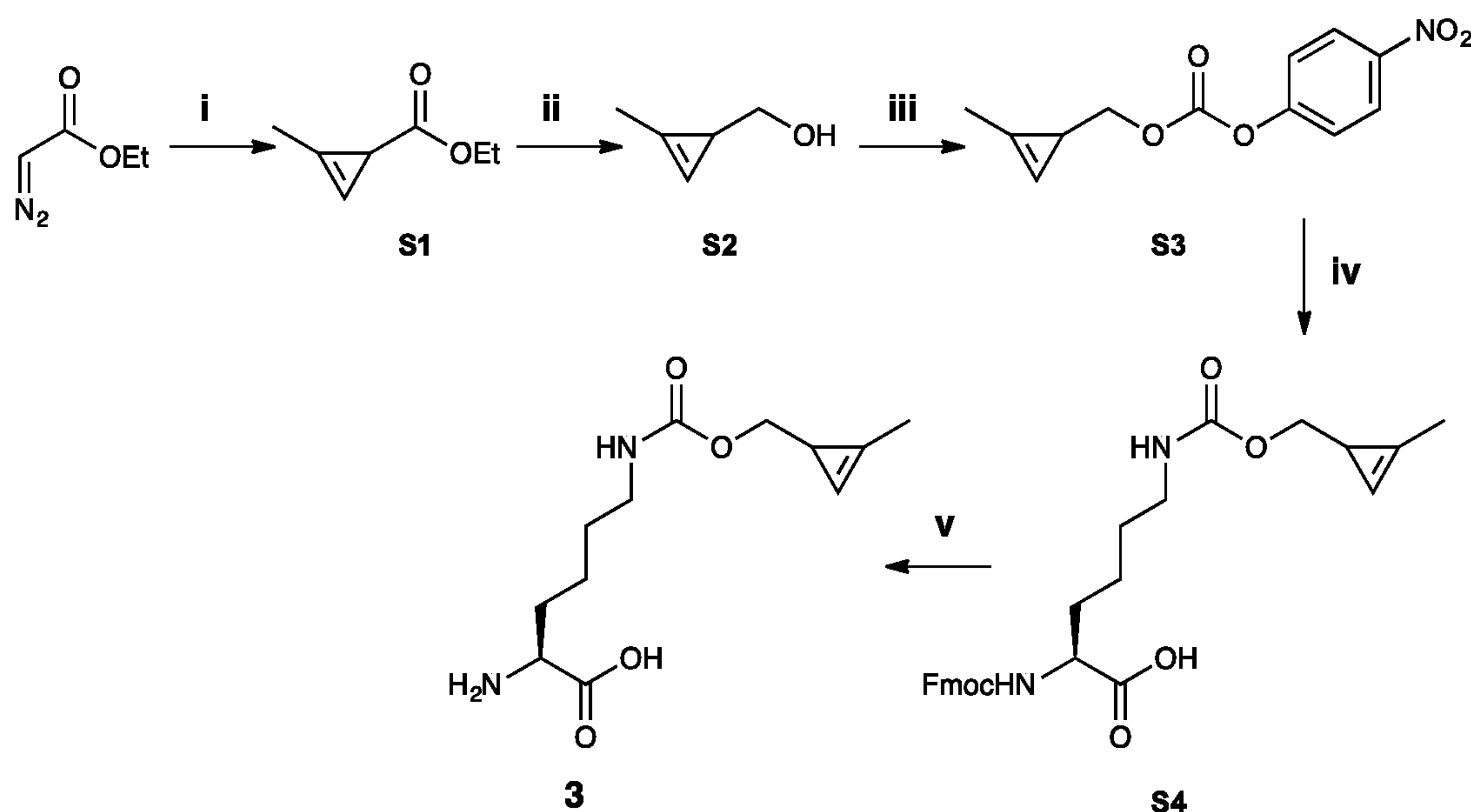
Example 1 - Synthesis of *N*^ε-[((2-methylcycloprop-2-en-1-yl)methoxy)carbonyl]-L-lysine 3

A class of reaction useful in protein labelling is the very rapid and specific inverse electron demand Diels-Alder reaction between strained alkenes (or alkynes) and tetrazines.²¹⁻²⁵

While we, and others, have previously encoded unnatural amino acids bearing strained alkenes, alkynes and tetrazines via genetic code expansion and demonstrated their use for site-specific protein labelling via inverse electron demand Diels-Alder reactions,²⁶⁻³⁰ all the molecules used to date are rather large. We have previously shown that a variety of carbamate derivatives of lysine are good substrates for PylRS,³¹ and it has been demonstrated that 1,3 disubstituted cyclopropenes, unlike 3,3 disubstituted cyclopropenes,^{32,24} react efficiently with tetrazines.²² We therefore designed and synthesized a carbamate derivative of lysine, bearing a 1,3 disubstituted cyclopropene (*N*^ε-[((2-methylcycloprop-2-en-1-yl)methoxy)carbonyl]-L-lysine **3**, **Fig. 1b**), for incorporation into proteins and labelling with tetrazines.

20

Synthesis of methylcycloprop-2-en-1-yl}methoxy)carbonyl]-L-lysine (3)



Scheme 1. Synthesis of *N* ϵ -[(2-methylcycloprop-2-en-1-yl)methoxy]carbonyl]-L-lysine **3**. *Reagents and conditions:* i. $\text{Rh}_2(\text{OAc})_4$, propyne, CH_2Cl_2 , 4 °C to RT, 75% yield; ii. DIBAL-H, CH_2Cl_2 , 0 °C to RT; iii. 4-nitrophenyl chloroformate, Hünig's base, CH_2Cl_2 , RT, 73% yield; iv. Fmoc-Lys-OH, Hünig's base, THF/DMF, 4 °C to RT, 82% yield; v. NaOH, THF/ H_2O , RT, 68% yield.

i. Ethyl 2-methylcycloprop-2-ene-1-carboxylate **S1**

A 100 mL 2-neck round bottom flask was charged with CH_2Cl_2 (2 mL) and rhodium acetate (442 mg, 1 mmol, 0.05 eq), and fitted with a dry ice condenser. Propyne (approx. 10 mL) was condensed into the rhodium acetate suspension and the flask lowered into a water bath (20 °C), a steady reflux of propyne was obtained. Ethyl diazoacetate (2.1 mL, 20 mmol, 1eq) was added to the stirred propyne solution drop-wise over 1 h using a syringe pump. The reaction was stirred at room temperature for a further 10 minutes whereby TLC analysis showed the reaction to be complete by after this time. The cyclopropene product was then purified by silica gel flash column chromatography eluting with pentane and diethyl ether (90:10). This gave the desired product **S1** as a colourless volatile liquid (1.9 g, 75% yield). ^1H NMR analysis δ_{H} (400 MHz, CDCl_3) 6.35 (1H, t, J 1.4), 4.18-4.09 (2H, m), 2.16 (3H, d, J 1.3), 2.12 (1H, d, J 1.6), 1.26 (3H, t, J 7.1); LRMS m/z (ES^+) 127.2 $[\text{M}+\text{H}]^+$.

These values are in good agreement with literature.{Liao, 2004 #1}

ii. and iii. (2-Methylcycloprop-2-en-1-yl)methyl (4-nitrophenyl) carbonate **S3**

DIBAL-H (22.5 mL of a 1M solution in CH_2Cl_2 , 22.5 mmol, 1.5 eq) was added drop-wise to a stirred solution of cyclopropene ester **S1** (1.9 g, 15 mmol, 1 eq) in CH_2Cl_2 (15 mL) at -10 °C. The reaction was stirred at -10 °C for 20 minutes before quenching with the

cautious addition of H₂O (1 mL), then NaOH (1 mL of a 1 M solution in H₂O) and H₂O (2.3 mL). The mixture was stirred for a further 2h at room temperature before it was dried (Na₂SO₄) and filtered. Hunig's base (3.9 mL, 22.5 mmol, 1.5eq) was added to the filtrate (containing crude cyclopropene alcohol **S2**) followed by the addition of 4-nitrophenyl chloroformate (3.3 g, 16.5 mmol, 1.1 eq). After stirring at room temperature for 18 hours a significant colourless precipitate formed, and TLC analysis showed complete consumption of the crude cyclopropene alcohol **S2**. The reaction was diluted with CH₂Cl₂ and then dry loaded onto silica gel, whereby the activated carbonate **S3** was purified by silica gel column chromatography eluting with ethyl acetate and hexane (20:80). This gave the desired cyclopropene carbonate **S3** as a colourless oil (2.7 g, 73% yield over 2 steps). ¹H NMR analysis δ_H (400 MHz, CDCl₃) 8.28 (2H, d, *J* 9.2), 7.39 (2H, d, *J* 9.2), 6.62 (1H, s), 4.21 (1H, dd, *J* 10.9, 5.3), 4.14 (1H, dd, *J* 10.9, 5.3), 2.18 (3H, d, *J* 1.3), 1.78 (1H, td, *J* 5.3, 1.3).

iv. *N*^α-(Fmoc)-*N*^ε-(((2-methylcycloprop-2-en-1-yl)methoxy)carbonyl)-L-lysine **S4**

Fmoc-Lys-OH·HCl (6.7 g, 16.5 mmol, 1.5 eq) was dissolved in THF (30 mL) and DMF (10 mL), to this solution was added Hünig's base (9.0 mL, 55.0 mmol, 5 eq) followed by cyclopropene carbonate **S3** (2.7 g, 11.0 mmol, 1 eq) an immediate yellow coloration was observed upon addition of the carbonate. The reaction was stirred at room temperature for 6 hours and was adjudged complete by the consumption of starting material after this time as shown by TLC analysis. The crude reaction mixture was dry loaded onto silica gel and the major product purified by silica gel column chromatography eluting with ethyl acetate, hexane and acetic acid (50:49:1 then 99:0:1). This gave the desired product **S4** as a colourless gum (4.3 g, 82% yield). ¹H NMR analysis δ_H (400 MHz, CDCl₃) 7.77 (2H, t, *J* 7.6), 7.65-7.55 (2H, m), 7.39 (2H, t, *J* 7.6), 7.31 (2H, t, *J* 7.3), 6.54 (1H, s), 5.68-5.57 (1H, m), 4.84 (1H, br-s), 4.44-4.32 (2H, m), 4.22 (1H, t, *J* 7.0), 3.98-3.87 (1H, m), 3.17-3.09 (2H, m), 2.15-2.06 (6H, m), 1.99-1.86 (1H, m), 1.84-1.70 (1H, m), 1.68-1.59 (1H, m), 1.58-1.34 (2H, m); LRMS *m/z* (ES⁺) 479.3 [M+H]⁺, 501.3 [M+Na]⁺, *m/z* (ES⁻) 477.2 [M-H]⁻.

v. *N*^ε-[(((2-methylcycloprop-2-en-1-yl)methoxy)carbonyl)]-L-lysine **3**

N^α-(Fmoc)-*N*^ε-(((2-methylcycloprop-2-en-1-yl)methoxy)carbonyl)-L-lysine **S4** (3.5 g, 7.0 mmol, 1 eq) was dissolved in THF and H₂O (3:1 40 mL), to this solution was added sodium hydroxide (0.9 g, 22.6 mmol, 3.1 eq). The reaction was stirred at room temperature for 8 hours after which time the reaction was adjudged complete by LC-MS analysis. The reaction mixture was diluted with H₂O (100 mL) and the pH adjusted to ~5 by the addition of HCl (1M). The aqueous solution was washed with Et₂O (5× 100 mL), then concentrated to dryness yielding a colourless solid. The solid was purified by

preparative HPLC, the product fractions were combined and the solvent removed by freeze-drying. This gave *N*^ε-[(2-methylcycloprop-2-en-1-yl)methoxy]carbonyl-L-lysine **3** as a colourless solid. δ_{H} (400 MHz, D₂O) 6.45 (1H, s), 3.90-3.61 (2H, m), 3.09 (1H, t, *J* 6.4), 2.98-2.86 (2H, m), 1.92 (3H, s), 1.52-1.37 (2H, m), 1.37-1.22 (2H, m), 1.21-1.08 (2H, m), 0.83 (1H, d, *J* 5.2). LRMS *m/z* (ES⁺) 257.2 [M+H]⁺, *m/z* (ES⁻) 255.2 [M-H]⁻. δ_{C} (100 MHz, D₂O) 101.1 (CH), 72.3 (CH₂), 55.9 (CH), 40.2 (CH₂), 34.3 (CH₂), 28.9 (CH₂), 20.3 (CH₂), 16.6 (CH₃), 10.8 (CH) HRMS (ES⁺) Found: (M+Na)⁺ 279.1302. C₁₂H₂₀O₄N₂Na required M⁺, 279.1315.

10 Example 2 - Encoding the site-specific incorporation of **3** in *E. coli*

We demonstrated that **3** is efficiently and site-specifically incorporated into recombinant proteins in response to the amber codon using the PylRS/tRNA_{CUA} pair and an SfGFP gene bearing an amber codon at position 150 (**Supplementary Fig. 2a**). The yield of protein is 8 mg per litre of culture, which is greater than that obtained for a well-established efficient substrate for PylRS *N*^ε-[(tert-butoxy)carbonyl]-L-lysine **1** (4 mg per litre of culture) ³³ Electrospray ionisation mass spectrometry of SfGFP bearing **3** at position 150 (SfGFP-**3**) confirms the incorporation of the unnatural amino acid (**Supplementary Fig. 2b**). SfGFP-**3** was specifically labelled with the fluorescent tetrazine probe **4a**, while SfGFP-**1** was left unlabelled (**Supplementary Fig. 2b**). 2 nmol of SfGFP-**3** was quantitatively labelled with 10 equivalents of **4a** in 30 minutes, as judged by both fluorescence imaging and mass spectrometry (**Supplementary Fig. 2b**). The second order rate constant for labelling SfGFP-**3** with **4a** was $27 \pm 1.8 \text{ M}^{-1}\text{s}^{-1}$ (**Supplementary Fig. 2c**).²⁶

25

Since PylRS does not recognize the anticodon of its cognate tRNA³⁴ it is possible to alter the anticodon of this tRNA to decode distinct codons. We created a new tRNA in which the anticodon of PyltRNA_{CUA} was converted from CUA to UUU (**Supplementary Table 1**), to decode a set of lysine codons. We added 0.1 mM **3** to cells containing PylRS, PyltRNA_{UUU}, and the gene for T4 lysozyme. Following expression of T4

30

lysozyme we detected proteins in the lysate bearing **3** with the tetrazine probe **4a** (20 microM 1h, **Supplementary Fig. 3**). Control experiments show that the observed labelling requires the presence of the synthetase and tRNA, and electrospray ionization mass spectrometry demonstrates the incorporation of **3** in place of lysine in T4
 5 lysozyme (**Supplementary Fig. 4**). The addition of **3** (0.1 or 0.5 mM) has little or no effect on cell growth (**Supplementary Fig. 5**) suggesting that the amino acid is not toxic at the concentration used, and there is substantial labelling within 1h of amino acid addition (**Supplementary Fig. 6**).

10 **Example 3 - Genetic encoding of 3 in human cells**

Full-length mCherry-**3**-GFP-HA was expressed in HEK293 cells carrying the PylRS/tRNA_{CUA} pair and mCherry-TAG-EGFP-HA (a fusion between the mCherry gene and the EGFP gene with a C-terminal HA tag, separated by the amber stop codon (TAG)).¹⁸ Full-length protein was detected only in the presence of the **3** (**Fig. 8a**. Full
 15 gels in **Supplementary Fig. 11**). mCherry-**3**-EGFP-HA was selectively labelled with **4a**, while mCherry-**1**-EGFP-HA was not labelled (**Fig. 8b**)¹⁸ demonstrating the site-specific incorporation of **3** with the PylRS/tRNA_{CUA} pair in human cells.

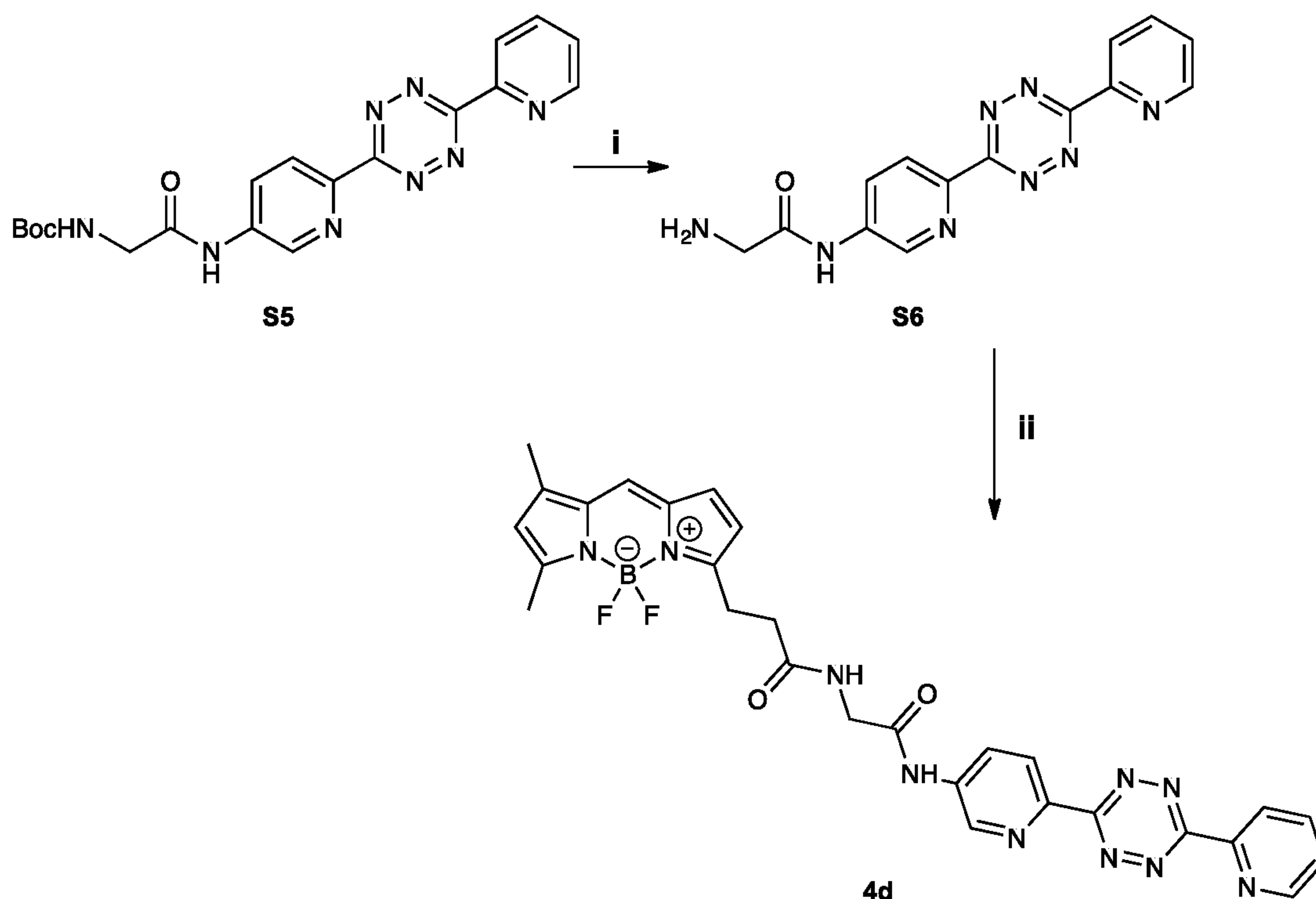
Example 4 - Genetic encoding of 3 in *D. melanogaster*

20 We demonstrated that **3** can be site specifically incorporated into proteins in *D. melanogaster*. To achieve this, we used flies containing the PylRS/tRNA_{CUA} pair (with the tRNA expressed ubiquitously from a U6 promoter and UAS-PylRS expression directed to ovaries using a nos-vp16-GAL4 driver), and a dual luciferase reporter bearing an amber codon between firefly and renilla luciferase.²⁰ We observe a strong
 25 luciferase signal that is dependent on the addition of **1** or **3**, and the dual luciferase signal is larger with **3**. These experiments demonstrate that **3** is taken up by flies and is more efficiently incorporated *in vivo* in response to an amber codon than **1** (**Fig. 10a**), a known excellent substrate for PylRS. **3** may be supplied by feeding food

supplemented with amino acid **3** at 10mM. In additional experiments, we demonstrated by western blot the efficient incorporation of **3** into a GFP-TAG-mCherry-HA construct (**Supplementary Fig. 15**) expressed in ovaries²⁰ (**Fig. 10b**), and the specific fluorescent labelling of the incorporated amino acid with **4g** (**Fig.**

5 **10c**).

Example 5 - Synthesis of Tetrazine-BODIPY FL 4d



10 **Scheme 3.** Synthesis of tetrazine-biotin **4d**. *Reagents and conditions:* i. HCl dioxane, RT, 100% yield; ii. Bodipy-FL-NHS ester, Hünig's base, DMF, RT.

i. **S6**

Boc-protected Tetrazine **S6** was synthesized using the procedure reported earlier⁶. 4M HCl in dioxane (500 µL, 2.0 mmol) was added to a stirring solution of Tetrazine **S5** (8 mg, 0.02 mmol) in DCM (500 µL). The reaction was carried out for 2 h at room temperature and subsequently the solvent was removed under reduced pressure to yield primary amine hydrochloride **S6** as a pink solid (6mg, 0.02 mmol, 100%). The compound was directly used in the next step without any further purification.

15

ii. **4d**

20 BODIPY FL succinimidyl ester (5mg, 0.013 mmol, Life technologies) and Hünig's base (50 µl, 2.8 mmol) were added to the solution of Tetrazine-amine **S2** (6mg, 0.02 mmol) in dry DMF (1 mL). The reaction mixture was stirred at room temperature for 16 h. The

reaction mixture was diluted with 4ml of water and the product was purified by semi-preparative reverse phase HPLC using a gradient from 10% to 90% of buffer B in buffer A (buffer A: H₂O; bufferB: acetonitrile). The identity and purity of the tetrazine-BODIPY FL conjugate **4d** was confirmed by LC-MS. ESI-MS: [M-H]⁻, calcd. 581.38, found 581.2.

SUMMARY OF EXAMPLES 1 to 5

We have characterized the synthesis of, and the genetically encoded, site-specific incorporation of a cyclopropene containing amino acid **3**, and demonstrated the quantitative labelling of **3**, with tetrazine probes, in proteins expressed in *E. coli*, mammalian cells and *D. melanogaster*, thereby showing the widespread utility and industrial application of the present invention.

Supplementary References to Examples 1 to 5

1. Gautier, A. et al. Genetically Encoded Photocontrol of Protein Localization in Mammalian Cells. *Journal of the American Chemical Society* **132**, 4086-4088 (2010).
2. Karp, N.A., Kreil, D.P. & Lilley, K.S. Determining a significant change in protein expression with DeCyder during a pair-wise comparison using two-dimensional difference gel electrophoresis. *Proteomics* **4**, 1421-1432 (2004).
3. Karp, N.A. & Lilley, K.S. Design and analysis issues in quantitative proteomics studies. *Proteomics* **7 Suppl 1**, 42-50 (2007).
4. Lilley, K.S. in Current Protocols in Protein Science (John Wiley & Sons, Inc., 2001).
5. Von Stetina, J.R., Lafever, K.S., Rubin, M. & Drummond-Barbosa, D. A Genetic Screen for Dominant Enhancers of the Cell-Cycle Regulator alpha-Endosulfine Identifies Matrimony as a Strong Functional Interactor in Drosophila. *G3 (Bethesda)* **1**, 607-613 (2011).
6. Lang, K. et al. Genetically encoded norbornene directs site-specific cellular protein labelling via a rapid bioorthogonal reaction. *Nat Chem* **4**, 298-304 (2012).

Example 6 - Dual Labelling of Proteins

The ability to attach two distinct molecules to programmed sites in proteins will facilitate a variety of applications including FRET^{1,2} to study protein structure, conformation and dynamics. Several approaches for doubly labeling proteins have been reported. One approach relies on the installation of one unnatural amino acid that is specifically labeled in combination with cysteine thiol labeling, but this approach is generally limited to proteins that do not contain free thiols.^{3,4} Chemical ligation approaches can be combined with the genetic encoding of a single unnatural amino acid for protein labeling,⁵ but this may limit the size and/or sites that may be labeled. Perhaps the most generally applicable approach for protein double labelling is based on the genetic incorporation of two distinct amino acids in response to two distinct codons introduced at user defined sites in the gene of interest.

An ideal strategy for dual labeling requires i) the efficient, cellular, incorporation of two distinct unnatural amino acids into a protein that can be labelled in mutually orthogonal reactions, and ii) the development of mutually orthogonal reactions that allow the simultaneous addition of two molecules to the protein for rapid, quantitative labelling of the protein in aqueous media at physiological pH, temperature and pressure.

Scheme A (Figure 14) shows concerted, rapid, one-pot quantitative dual labelling of proteins in aqueous medium at physiological pH and temperature. (a) Unnatural amino acids and fluorophores used in this example. (b) Concerted labeling at an encoded terminal alkyne and an encoded cyclopropene via mutually orthogonal cycloadditions.

The cellular, genetically directed incorporation of two distinct unnatural amino acids into proteins has been demonstrated in response to an amber and quadruplet codon,⁶ two distinct stop codons,^{7,8} or two distinct quadruplet codons.⁹ We previously demonstrated the evolution of an orthogonal ribosome (ribo-Q1) that efficiently reads quadruplet codons and amber codons on orthogonal mRNA using cognate extended anticodon tRNAs or amber suppressors respectively.⁶ We demonstrated that the pyrrolysyl-tRNA synthetase/tRNA pair and synthetically evolved derivatives of the *M*/TyrRS/tRNA pair are mutually orthogonal in their aminoacylation specificity and can be used to direct the incorporation of pairs of unnatural amino acids in response to amber and quadruplet codons.⁶ We recently described several major advances in this system, including the evolution of a series of quadruplet decoding tRNAs based on the pyrrolysyl-tRNA synthetase (PylRS)/tRNA pair that efficiently direct the incorporation of unnatural amino acids in response to quadruplet codons using the evolved orthogonal translation machinery.⁹ We demonstrated the very efficient incorporation of a matrix of pairs of unnatural amino acids using the evolved PylRS/tRNA^{UACU} pair and derivatives of the *M*/TyrRS/tRNA^{CUA} pair with orthogonal messages bearing TAG and AGTA codons and ribo-Q1.⁹

A limited range of chemistries have been investigated for the double labeling of proteins containing pairs of unnatural amino acids. The incorporation of azide- and alkyne-containing amino acids, and their non-quantitative labeling with alkyne and azide based fluorophores has been reported⁷, but this is not ideal for double labeling of proteins; if the encoded azide and alkyne are in proximity they can react to form a triazole in the protein, a strategy which allows genetically directed protein stapling,⁶ but precludes labeling with probes. Moreover, an efficient one-pot reaction is not feasible because of the reaction between azide- and alkyne- bearing probes with each other. The incorporation of ketone and azide containing amino acids has been reported,^{8,10} which allows one-pot reaction of the encoded ketone with alpha effect nucleophiles, and the azides with alkyne probes.¹⁰ However this approach is problematic because encoded azides are subject to reduction in many proteins when expressed in *E. coli*,^{8,11} which will prevents quantitative labeling. Moreover, ketone labeling with alpha effect nucleophiles is very slow (rate constant approximately $10^{-4} \text{ M}^{-1}\text{s}^{-1}$) and the reaction is optimal at pH4-5.5,¹² which limits its utility for many proteins that are denatured or precipitate when kept for long periods under acidic conditions. We recently genetically installed a deactivated tetrazine containing amino acid¹³ and a norbornene containing amino acid¹⁴⁻¹⁶ into proteins using our optimized orthogonal translation system.⁹ Because the rate of inverse electron demand Diels Alder reaction between the deactivated tetrazine and norbornene is very slow, but the tetrazine can react with bicyclononyne based probes and the norbornene can react with activated tetrazine probes we were able to use this approach to specifically and quantitatively double label proteins.⁹ While this approach has the advantage of proceeding in aqueous media at physiological pH, temperature and pressure; it does require sequential labeling steps (to avoid inverse electron Demand reactions between probes), each of which takes several hours, with purification between steps. All approaches reported to date for doubly labeling proteins at genetically encoded unnatural amino acids take tens of hours to days to reach completion.

An ideal approach to double label proteins would allow rapid one-pot labeling of genetically installed bio-orthogonal functional groups, proceed rapidly in aqueous media at physiological pH, temperature and pressure and be implemented simply by adding the labeling reagents to a recombinant protein bearing the site specifically incorporated

bioorthogonal groups. A promising pair of mutually orthogonal reactions for one-pot labeling under aqueous conditions at physiological pH are the Cu(I)-catalysed 3+2 cycloaddition between azides and terminal alkynes,¹⁷ and the inverse electron demand Diels Alder reaction of a strained alkenes and a tetrazine¹⁸⁻²³(**Figure 11**). The reaction of strained alkynes and azides can also be orthogonal to strained alkene tetrazine reactions, but since tetrazines react with strained alkynes this approach requires careful tuning of the rate constants for each reaction.²⁴ No combination of 3+2 cycloaddition and inverse electron demand Diels Alder reaction has been demonstrated for protein labelling.

We demonstrated in examples 1 to 5 that a 1,3 disubstituted cyclopropene containing amino acid, **2** (referred to as **3** in examples 1 to 5 and elsewhere in this document), can be efficiently and site specifically incorporated into proteins using the PylRS/tRNA_{CUA} pair.²⁵ This amino acid, unlike the 3,3 disubstituted cyclopropene incorporated for photoclick reactions,²⁶ reacts with tetrazines^{19,27} with on-protein rate constants of $27 \text{ M}^{-1}\text{s}^{-1}$.²⁵ Here we demonstrate the efficient genetic encoding of a terminal alkyne containing amino acid **1** and a cyclopropene containing amino acid **2** into a single protein and their rapid, quantitative, one-pot labeling with azide and tetrazine probes (**Figure 11**). This work provides the first approach to the concerted double labeling of proteins in a one-pot process under aqueous conditions, at physiological pH, and provides a step change in the speed of double labeling, from days in previous work to 30 minutes in the approach reported here.

Proteins containing either **1** or **2** were overexpressed to examine the specificity of the orthogonality of the proposed labeling reactions. A fusion protein of glutathione-S-transferase and calmodulin (GST-CaM) with amino acid **1** at position 1 in calmodulin was expressed from cells containing ribo-Q1 (an evolved orthogonal ribosome^{6,28,29}), *O-gst-cam*_{ITAG} (a fusion gene between *glutathione-S-transferase* (*gst*) and *calmodulin* (*cam*) on an orthogonal message³⁰ in which the first codon of *cam* is replaced with a TAG codon), and *M/PrpRS*/tRNA_{CUA} (a synthetase/tRNA pair developed for incorporating **1** in response to the TAG codon)³¹ grown in the presence of **1** (4 mM). The GST tag was subsequently removed by cleavage using thrombin at an engineered thrombin-cleavage site between GST and CaM. CaM₁ (CaM containing **1** at position 1, ~100 pmole) was labelled with the azide containing fluorophore **3** (2 nmole), in a Cu (I)-catalysed click reaction. The reaction was quantitative as judged by both the quantitative shift of the fluorescently labelled protein by SDS-PAGE and electrospray ionization mass spectrometry (ESI-MS) (**Figure 11a**).

The cyclopropene containing amino acid, **2**, was site specifically incorporated at position 40 of calmodulin. The modified protein was expressed in cells bearing the PylRS/tRNA_{CUA} (that efficiently directs the site specific incorporation of **2**),²⁵ ribo-Q1, and *O-gst-cam*_{40TAG} grown in the presence of **2** (1 mM). CaM₂₄₀ (~100 pmol) (obtained after thrombin cleavage of the GST tag) was labelled with the tetrazine containing fluorophore **4** (2 nmole). The reaction was quantitative as judged by both the quantitative shift of the fluorescently labelled protein by SDS-PAGE and electrospray ionization mass spectrometry (ESI-MS) (**Figure 11b**). CaM₂₄₀ was not labeled with **3** under the conditions that led to quantitative labeling of CaM₁ with **3** (**Figure 11a**). Similarly, CaM₁ was not labeled with **4** under conditions where CaM₂₄₀ was quantitatively labeled with **4**. These experiments demonstrate that the two labeling reagents react quantitatively with their target amino acid, but do not react with their non-targeted unnatural amino acid in proteins.

Next we investigated labeling **1** and **2** within the same protein. We site-specifically incorporated **1** and **2** at positions 1 and 40 of calmodulin to produce CaM₁₂₄₀ (**Figure 12**). We directed the incorporation of amino acid **1** with an *M/PrpRS*/tRNA_{CUA} pair and the incorporation of amino acid **2** with the evolved PylRS/tRNA_{UACU} pair, which efficiently decodes the quadruplet AGTA codon on orthogonal messages using ribo-Q1.⁹ Unnatural amino acids were incorporated in response to UAG and AGTA codons at positions 1 and 40 in calmodulin, within a GST-calmodulin gene on an orthogonal message (*O-gst-cam*_{ITAG-40AGTA}). Expression of full-length GST-CaM₁₂₄₀ was dependent on the addition of amino acids **1** and **2** to *E. coli*, and ESI-MS demonstrated the genetically directed incorporation of amino acids **1** and **2** (**Figure 12c**). The yield of full length GST-CaM₁₂₄₀ was ~2 mg per L of culture.

To determine the time required to quantitatively label CaM₁₂₄₀ with azide **3** or tetrazine **4** we incubated 100 pmol of CaM₁₂₄₀ with 2 nmol of either **3** or **4** and followed each reaction

by both mobility shift on SDS-PAGE and fluorescent imaging upon labeling (**Figure 12b**). These experiments demonstrate that fluorophore labeling is complete in 30 minutes.

Next we investigated the labeling of CaM1₂₄₀ with both **3** and **4** (**Figure 13**). We first tested the addition of **4** (2 nmol) to CaM1₂₄₀ (100 pmol) followed by purification to remove free **4**, and subsequent labelling with **3** (2 nmol) (**Figure 13a** lane 4). This led to efficient double labelling as judged by SDS-PAGE mobility shift and fluorescence imaging. Next we performed sequential labeling without purification by incubating CaM1₂₄₀ with **4** for 30 minutes and then adding **3** and click reagents and incubating further for 30 min (**Figure 13a** lane 5). This also led to efficient double labelling as judged by SDS-PAGE mobility shift and fluorescence imaging. Finally, we simultaneously added **4** (2 nmol), **3** (2 nmol) and click reagents to CaM1₂₄₀ (100 pmol) and incubated for 30 minutes. (**Figure 13a** lane 6). This again led to efficient double labelling as judged by SDS-PAGE mobility shift and fluorescence imaging. In all doubly labeled proteins we observe a decrease in the BODIPY-FL fluorescence relative to the singly labeled control upon excitation at 488 nm (compare lanes 4, 5, and 6 to lane 3 in **Figure 13a**), consistent with in gel Förster resonance energy transfer (FRET) between BODIPY-FL and BODIPY-TMR-X. ESI-MS further demonstrates that this concerted, one-pot protocol leads to genetically directed efficient, rapid and quantitative double labeling of proteins.

In summary, in this example we show an efficient and rapid protocol for expressing recombinant proteins bearing a site specifically incorporated alkyne and a site specifically incorporated cyclopropene. We demonstrate that the inverse electron demand Diels Alder reaction of an encoded 1,3 disubstituted cyclopropene and tetrazine probe, and the 3+2 cycloaddition reaction of the encoded alkyne and azide probe are mutually orthogonal to each other and to the functional groups in proteins. By combining the genetic encoding of an alkyne and a cyclopropene in a single protein and labelling with the mutually orthogonal reactions we demonstrate the concerted, one-pot rapid double labeling of a protein in aqueous media at physiological pH and temperature. This strategy has utility for doubly labeling proteins for a variety of studies and applications, and may be extended to the double labeling of diverse molecules in diverse cells and organisms.

Note on example 6: The chemical designations in example 6 and in the corresponding figures (drawings) discussed in example 6 are self-contained and apply only to example 6. Discussion of chemical designations in the rest of this document are consistent with the exception of example 6. For example, the skilled reader will immediately appreciate that compound **2** of example 6 corresponds to compound **3** in the rest of this document (i.e. the exemplary cyclopropene amino acid of the invention). Compounds **3** and **4** of example 6 are tetrazine compounds.

REFERENCES TO EXAMPLE 6

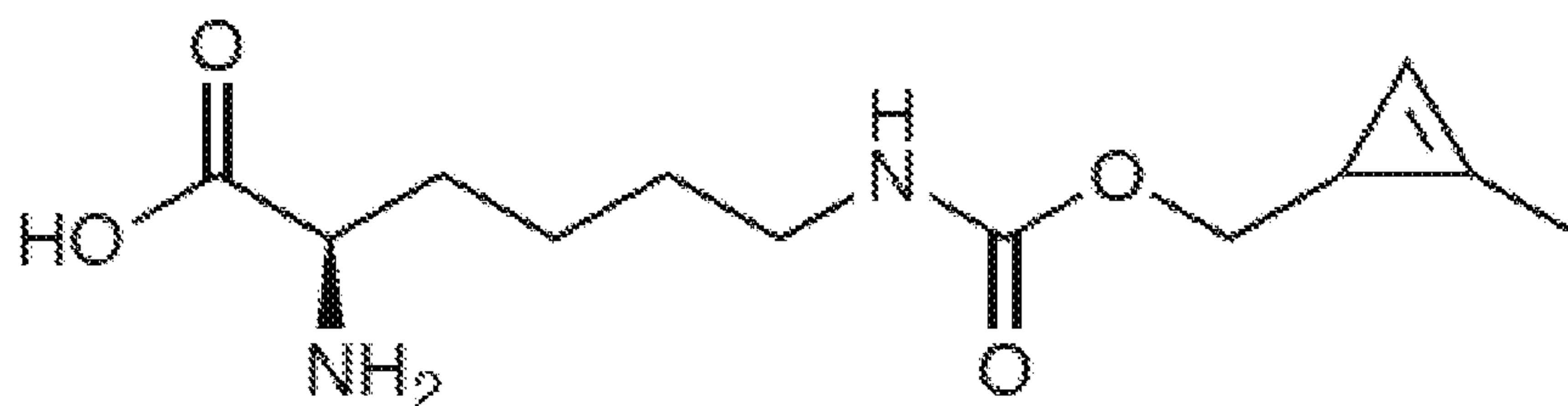
- (1) Zhang, J.; Campbell, R. E.; Ting, A. Y.; Tsien, R. Y. *Nature Reviews Molecular Cell Biology* **2002**, 3, 906.
- (2) Kajihara, D.; Abe, R.; Iijima, I.; Komiyama, C.; Sisido, M.; Hohsaka, T. *Nat Methods* **2006**, 3, 923.
- (3) Brustad, E. M.; Lemke, E. A.; Schultz, P. G.; Deniz, A. A. *J Am Chem Soc* **2008**, 130, 17664.
- (4) Nguyen, D. P.; Elliott, T.; Holt, M.; Muir, T. W.; Chin, J. W. *J Am Chem Soc* **2011**, 133, 11418.
- (5) Wissner, R. F.; Batjargal, S.; Fadzen, C. M.; Petersson, E. J. *J Am Chem Soc* **2013**, 135, 6529.
- (6) Neumann, H.; Wang, K.; Davis, L.; Garcia-Alai, M.; Chin, J. W. *Nature* **2010**, 464, 441.
- (7) Wan, W.; Huang, Y.; Wang, Z.; Russell, W. K.; Pai, P. J.; Russell, D. H.; Liu, W. R. *Angew Chem Int Ed Engl* **2010**, 49, 3211.
- (8) Chatterjee, A.; Sun, S. B.; Furman, J. L.; Xiao, H.; Schultz, P. G. *Biochemistry* **2013**.
- (9) Wang, K.; Sachdeva, A.; Cox, D. J.; Wilf, N. W.; Wallace, S.; Mehl, R. A.; Chin, J. W. *submitted*.
- (10) Wu, B.; Wang, Z.; Huang, Y.; Liu, W. R. *ChemBioChem : a European journal of chemical biology* **2012**, 13, 1405.
- (11) Sasmal, P. K.; Carregal-Romero, S.; Han, A. A.; Streu, C. N.; Lin, Z.; Namikawa, K.; Elliott, S. L.; Koster, R. W.; Parak, W. J.; Meggers, E. *ChemBioChem* **2012**, 13, 1116.
- (12) Rotenberg, S. A.; Calogeropoulou, T.; Jaworski, J. S.; Weinstein, I. B.; Rideout, D. *Proceedings of the National Academy of Sciences of the United States of America* **1991**, 88, 2490.
- (13) Seitchik, J. L.; Peeler, J. C.; Taylor, M. T.; Blackman, M. L.; Rhoads, T. W.; Cooley, R. B.; Refakis, C.; Fox, J. M.; Mehl, R. A. *J Am Chem Soc* **2012**, 134, 2898.
- (14) Lang, K.; Davis, L.; Torres-Kolbus, J.; Chou, C.; Deiters, A.; Chin, J. W. *Nat Chem* **2012**, 4, 298.
- (15) Plass, T.; Milles, S.; Koehler, C.; Szymański, J.; Mueller, R.; Wießler, M.; Schultz, C.; Lemke, E. A. *Angewandte Chemie International Edition* **2012**, 51, 4166.

- (16) Kaya, E.; Vrabel, M.; Deiml, C.; Prill, S.; Fluxa, V. S.; Carell, T. *Angewandte Chemie International Edition* **2012**, *51*, 4466.
- (17) Wang, Q.; Chan, T. R.; Hilgraf, R.; Fokin, V. V.; Sharpless, K. B.; Finn, M. G. *J Am Chem Soc* **2003**, *125*, 3192.
- 5 (18) Devaraj, N. K.; Weissleder, R. *Accounts of Chemical Research* **2011**, *44*, 816.
- (19) Yang, J.; Šečkutė, J.; Cole, C. M.; Devaraj, N. K. *Angewandte Chemie International Edition* **2012**, *51*, 7476.
- (20) Blackman, M. L.; Royzen, M.; Fox, J. M. *J Am Chem Soc* **2008**, *130*, 13518.
- (21) Lang, K.; Davis, L.; Wallace, S.; Mahesh, M.; Cox, D. J.; Blackman, M. L.; Fox, J. M.; Chin, J. W. *J Am Chem Soc* **2012**, *134*, 10317.
- 10 (22) Borrmann, A.; Milles, S.; Plass, T.; Dommerholt, J.; Verkade, J. M. M.; Wießler, M.; Schultz, C.; van Hest, J. C. M.; van Delft, F. L.; Lemke, E. A. *ChemBioChem* **2012**, *13*, 2094.
- (23) Schoch, J.; Staudt, M.; Samanta, A.; Wiessler, M.; Jaschke, A. *Bioconjug Chem* **2012**, *23*, 1382.
- (24) Karver, M. R.; Weissleder, R.; Hilderbrand, S. A. *Angew Chem Int Ed Engl* **2012**, *51*, 920.
- (25) Bianco, A.; Elliott, T. S.; Townsley, F. M.; Pisa, R.; Davis, L.; Elsässer, S. J.; Ernst, R. J.; Lang, K.; Sachdeva, A.; Chin, J. W. *Under Review*.
- 15 (26) Yu, Z.; Pan, Y.; Wang, Z.; Wang, J.; Lin, Q. *Angewandte Chemie International Edition* **2012**, *51*, 10600.
- (27) Kamber, D. N.; Nazarova, L. A.; Liang, Y.; Lopez, S. A.; Patterson, D. M.; Shih, H. W.; Houk, K. N.; Prescher, J. A. *J Am Chem Soc* **2013**, *135*, 13680.
- (28) Wang, K.; Schmied, W. H.; Chin, J. W. *Angew Chem Int Ed Engl* **2012**, *51*, 2288.
- 20 (29) Wang, K.; Neumann, H.; Peak-Chew, S. Y.; Chin, J. W. *Nature biotechnology* **2007**, *25*, 770.
- (30) Rackham, O.; Chin, J. W. *Nature chemical biology* **2005**, *1*, 159.
- (31) Deiters, A.; Schultz, P. G. *Bioorganic & Medicinal Chemistry Letters* **2005**, *15*, 1521.

CLAIMS

1. A polypeptide comprising an amino acid having a cyclopropene group wherein said cyclopropene group is joined to the amino acid via a carbamate group.
- 5 2. A polypeptide according to claim 1 wherein said cyclopropene group is a 1,3-disubstituted cyclopropene.
3. A polypeptide according to claim 2 wherein said cyclopropene is a
10 1,3-dimethylcyclopropene.
4. A polypeptide according to any of claims 1 to 3 wherein said cyclopropene group is present as a residue of a lysine amino acid.
- 15 5. A polypeptide according to any of claims 1 to 4 further comprising a tetrazine compound linked to said cyclopropene group.
6. An amino acid comprising cyclopropene wherein said cyclopropene group is joined to the amino acid moiety via a carbamate group.
- 20 7. An amino acid according to claim 6 wherein said cyclopropene is a 1,3-disubstituted cyclopropene.
8. An amino acid according to claim 7 wherein said cyclopropene is a
25 1,3-dimethylcyclopropene.
9. An amino acid according to any of claims 6 to 8 wherein said amino acid is a lysine amino acid.
- 30 10. An amino acid according to claim 9 which comprises *N*^ε-[((2-methylcycloprop-2-en-1-yl)methoxy)carbonyl]-l-lysine.

11. An amino acid according to claim 10 which consists of



12. A method of producing a polypeptide comprising a cyclopropene group wherein said cyclopropene group is joined to the amino acid moiety via a carbamate group, said method comprising genetically incorporating an amino acid comprising a cyclopropene group joined to the amino acid moiety via a carbamate group, into a polypeptide.

13. A method according to claim 12 wherein producing the polypeptide comprises

(i) providing a nucleic acid encoding the polypeptide which nucleic acid comprises an orthogonal codon encoding the amino acid having a cyclopropene group;

(ii) translating said nucleic acid in the presence of an orthogonal tRNA synthetase/tRNA pair capable of recognising said orthogonal codon and incorporating said amino acid having a cyclopropene group into the polypeptide chain.

15

14. A method according to claim 12 or claim 13 wherein said orthogonal codon comprises an amber codon (TAG), said tRNA comprises *MbtRNA*_{CUA} and said tRNA synthetase comprises *MbPyIRS*; or wherein said orthogonal codon comprises an amber codon (TAG), said tRNA comprises *MmtRNA*_{CUA} and said tRNA synthetase comprises *MmPyIRS*.

20

15. A method according to any of claims 12 to 14 wherein said amino acid comprising a cyclopropene group is an amino acid according to any of claims 6 to 11.

16. A method of producing a polypeptide comprising a tetrazine group, said method comprising providing a polypeptide according to any of claims 1 to 4, contacting said polypeptide with a tetrazine compound, and incubating to allow joining of the tetrazine to the cyclopropene group by an inverse electron demand Diels-Alder cycloaddition reaction.

30

17. A method according to claim 16 wherein said reaction is allowed to proceed for 10 minutes or less, preferably for 1 minute or less, preferably for 30 seconds or less.

18. A polypeptide according to any of claims 1 to 5 wherein said polypeptide comprises two or more amino acids each having a cyclopropene group, wherein each said cyclopropene group is joined to each said amino acid via a carbamate group.

5 19. A polypeptide according to claim 18 wherein said polypeptide comprises four amino acids each having a cyclopropene group.

20. An antibody drug conjugate (ADC) comprising a polypeptide according to any of claims 1 to 5, 18 or 19.

10

21. A compound, polypeptide or method substantially as described herein.

22. A compound, polypeptide or method substantially as described herein with reference to the accompanying drawings

15

1 / 20

FIG. 1

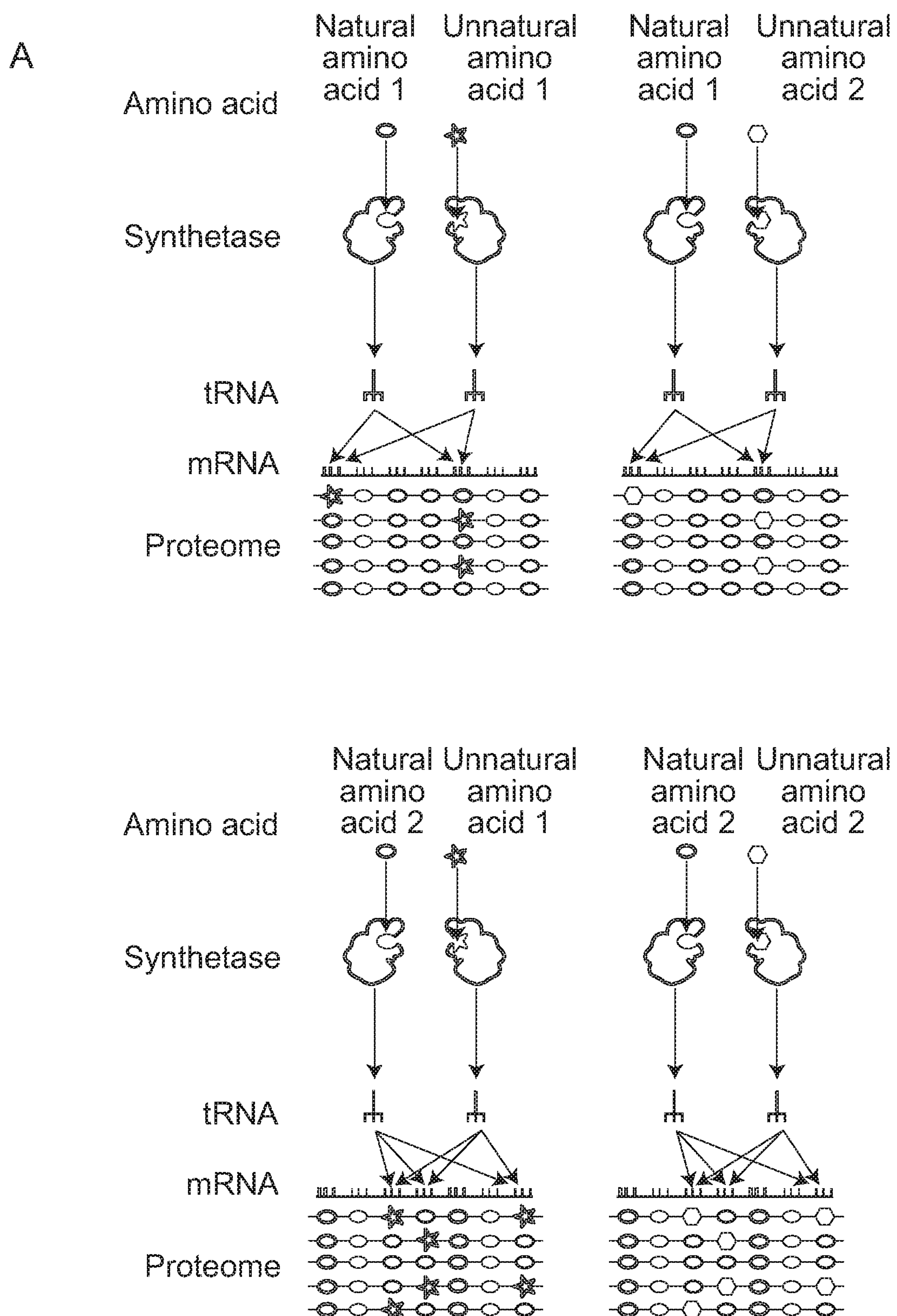
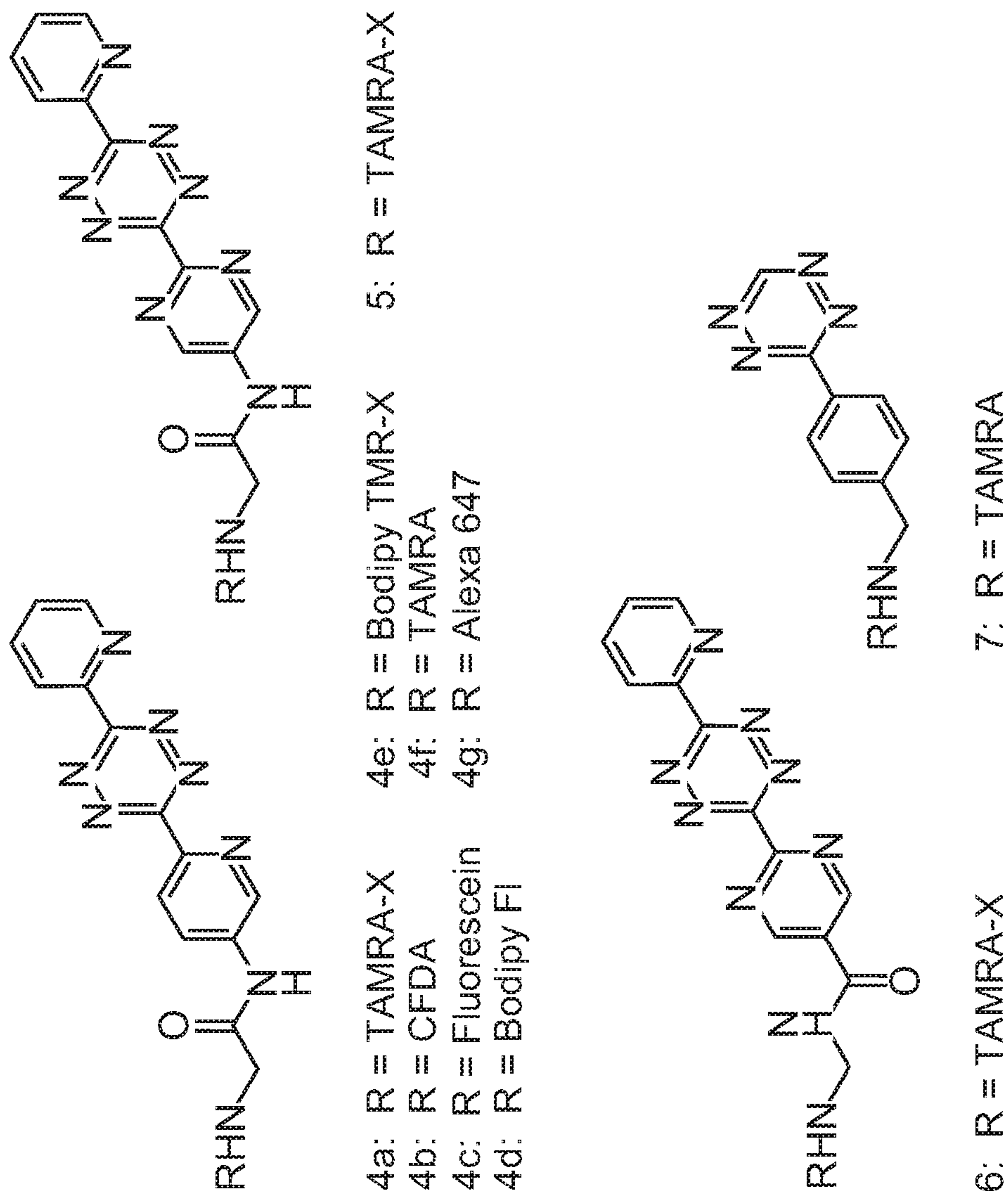
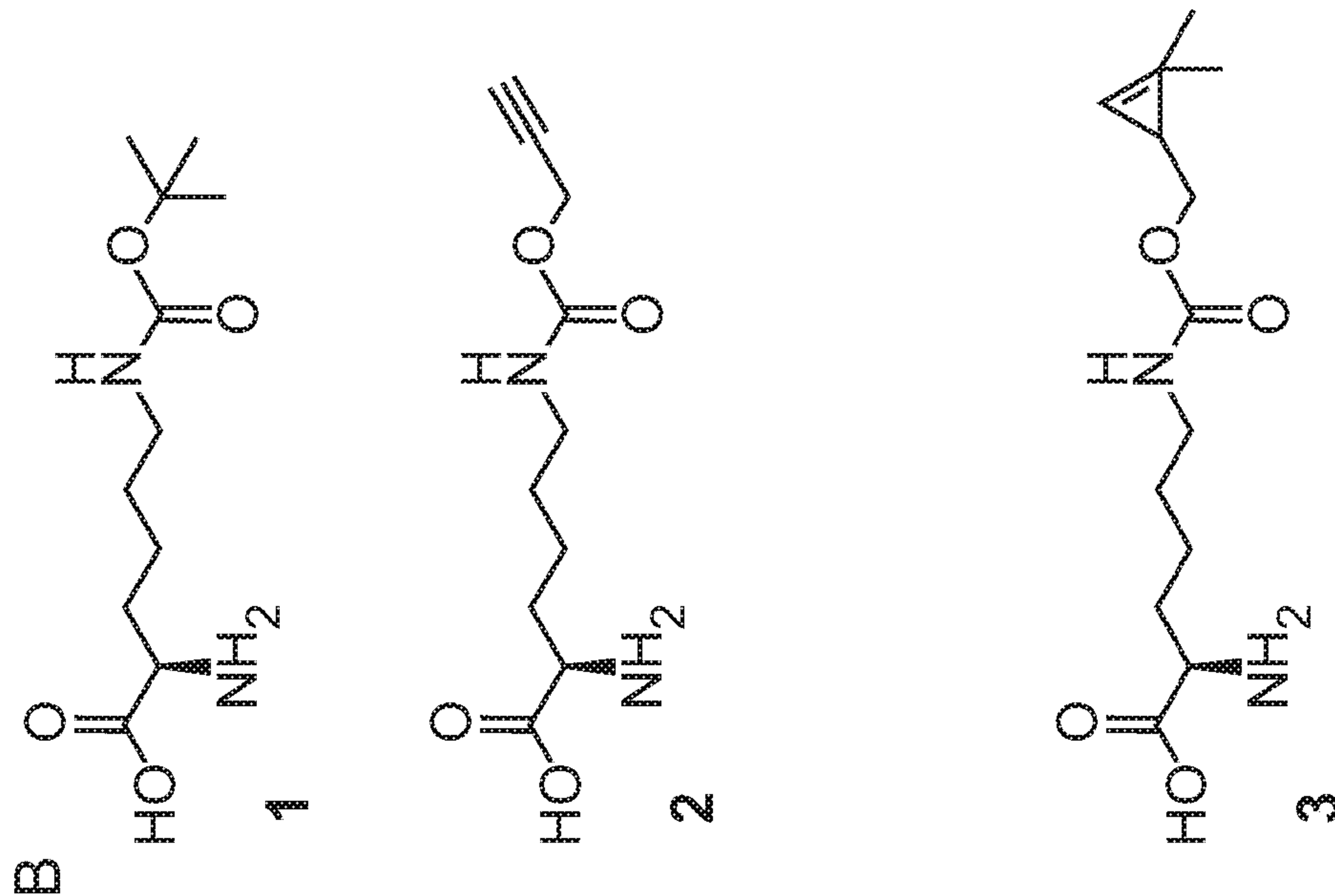


FIG. 1 continued



3 / 20

FIG. 2

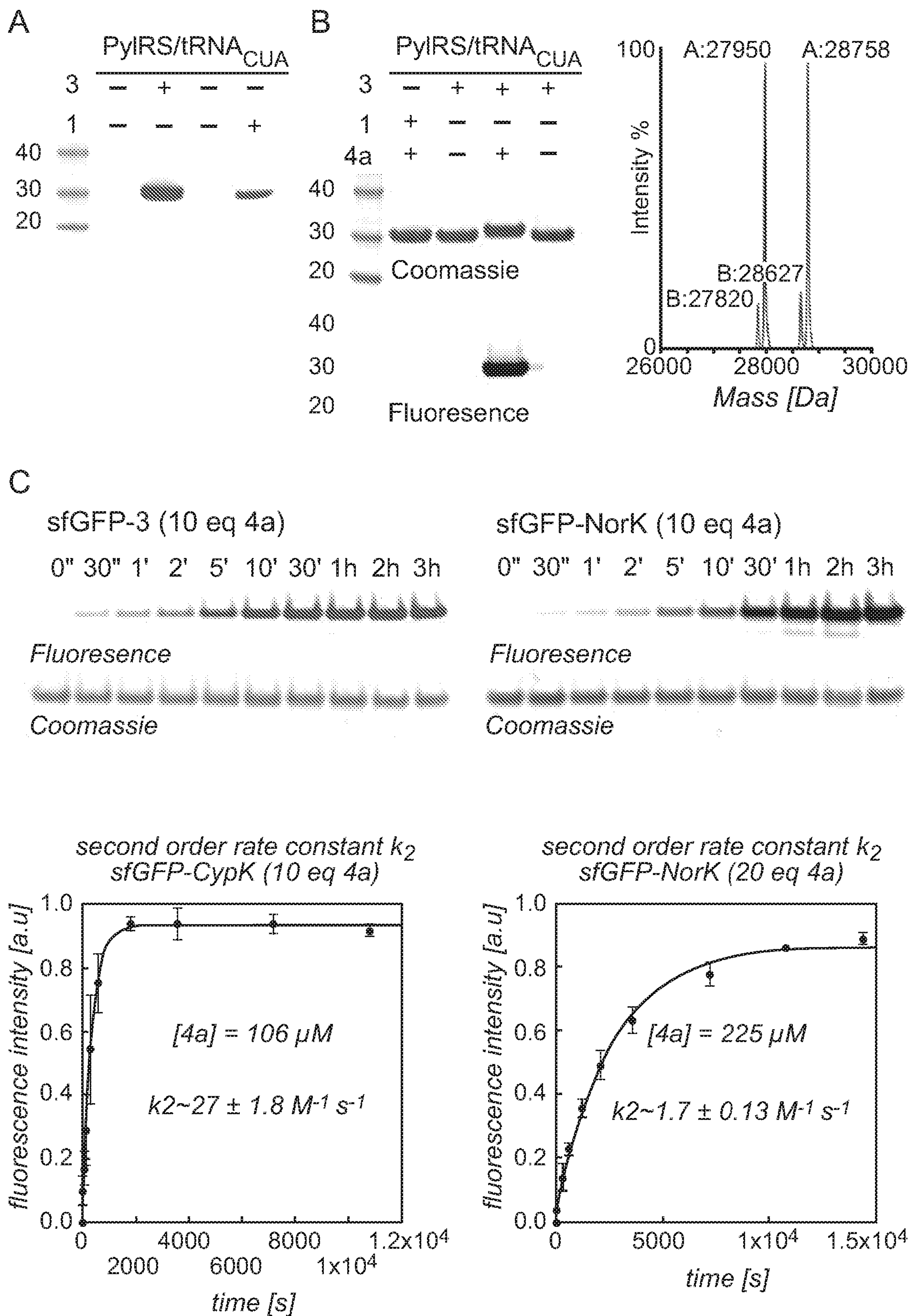


FIG. 3

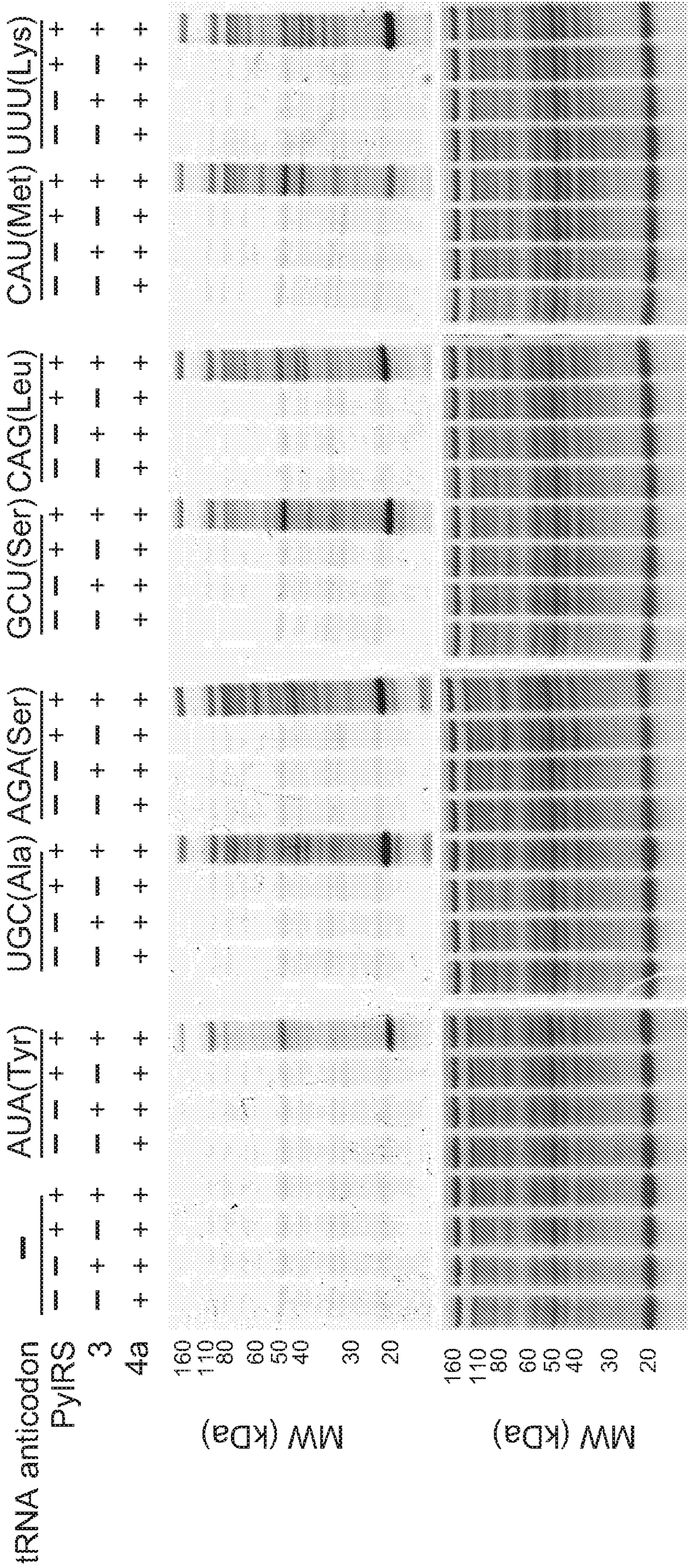
Name	Sequence	Notes
FMT17	CATGTAGATCGAATGGACTATAAATCC GTTCAGCCGGG	Mutagenesis primers for D.m.PylT
FMT18	CCCGGCTGAACGGATTTATAGTCCATT CGATCTACATG	CTA to ATA (Tyr)
FMT19	CATGTAGATCGAATGGACTTGCAATCC GTTCAGCCGGGTTAG	Mutagenesis primers for D.m.PylT
FMT20	CTAACCCGGCTGAACGGATTGCAAGTC CATTCGATCTACATG	CTA to TGC (Ala)
FMT21	GTAGATCGAATGGACTAGAAATCCGTT CAGCCGGG	Mutagenesis primers for D.m.PylT
FMT22	CCCGGCTGAACGGATTTCTAGTCCATT CGATCTAC	CTA to AGA (Ser)
FMT23	CATGTAGATCGAATGGACTGCTAATCC GTTCAGCCGGGTTAG	Mutagenesis primers for D.m.PylT
FMT24	CTAACCCGGCTGAACGGATTAGCAGTC CATTCGATCTACATG	CTA to GCT (Ser)
FMT25	GTAGATCGAATGGACTTAAAATCCGTT CAGCCGGG	Mutagenesis primers for D.m.PylT
FMT26	CCCGGCTGAACGGATTTTAAGTCCATT CGATCTAC	CTA to TAA (Leu)
FMT27	GTAGATCGAATGGACTCAGAATCCGTT CAGCCGGG	Mutagenesis primers for D.m.PylT
FMT28	CCCGGCTGAACGGATTCTGAGTCCATT CGATCTAC	CTA to CAG (Leu)
FMT29	GTAGATCGAATGGACTCATAATCCGTT CAGCCGGG	Mutagenesis primers for D.m.PylT
FMT30	CCCGGCTGAACGGATTATGAGTCCATT CGATCTAC	CTA to CAT (Met)
FMT31	CATGTAGATCGAATGGACTTTTAATCC GTTCAGCCGGGTTAG	Mutagenesis primers for D.m.PylT
FMT32	CTAACCCGGCTGAACGGATTAAAAGTC CATTCGATCTACATG	CTA to TTT (Lys)
MAM01	GTAGATCGAATGGACTCATAATCCGTT CAGCCGGG	Forward mutagenesis primer for CTA to CAT (Met)
MAM02	CCCGGCTGAACGGATTATGAGTCCATT CGATCTAC	Reverse mutagenesis primer for CTA to CAT (Met)

5 / 20

FIG. 3 continued

Name	Sequence	Notes
MAM03	GATCGAATGGACTCACAATCCGTTTCAG CCGG	Forward mutagenesis primer for CTA to CAC (Val)
MAM04	CCGGCTGAACGGATTGTGAGTCCATTC GATC	Reverse mutagenesis primer for CTA to CAC (Val)
MAM05	GATCGAATGGACTCTTAATCCGTTTCAG CCG	Forward mutagenesis primer for CTA to CTT (Lys)
MAM06	CGGCTGAACGGATTAAGAGTCCATTCG ATC	Reverse mutagenesis primer for CTA to CTT (Lys)
MAM07	GATCGAATGGACTCTCAATCCGTTTCAG CCG	Forward mutagenesis primer for CTA to CTC (Glu)
MAM08	CGGCTGAACGGATTGAGAGTCCATTCG ATC	Reverse mutagenesis primer for CTA to CTC (Glu)
MAM09	GTAGATCGAATGGACTGTAAATCCGTT CAGCCG	Forward mutagenesis primer for CTA to GTA (Tyr)
MAM10	CGGCTGAACGGATTACAGTCCATTCG ATCTAC	Reverse mutagenesis primer for CTA to GTA (Tyr)
MAM11	GATCGAATGGACTCCAAATCCGTTTCAG CCG	Forward mutagenesis primer for CTA to CCA (Trp)
MAM12	CGGCTGAACGGATTGGAGTCCATTCG ATC	Reverse mutagenesis primer for CTA to CCA (Trp)

FIG.4
A



7 / 20

FIG. 4

B

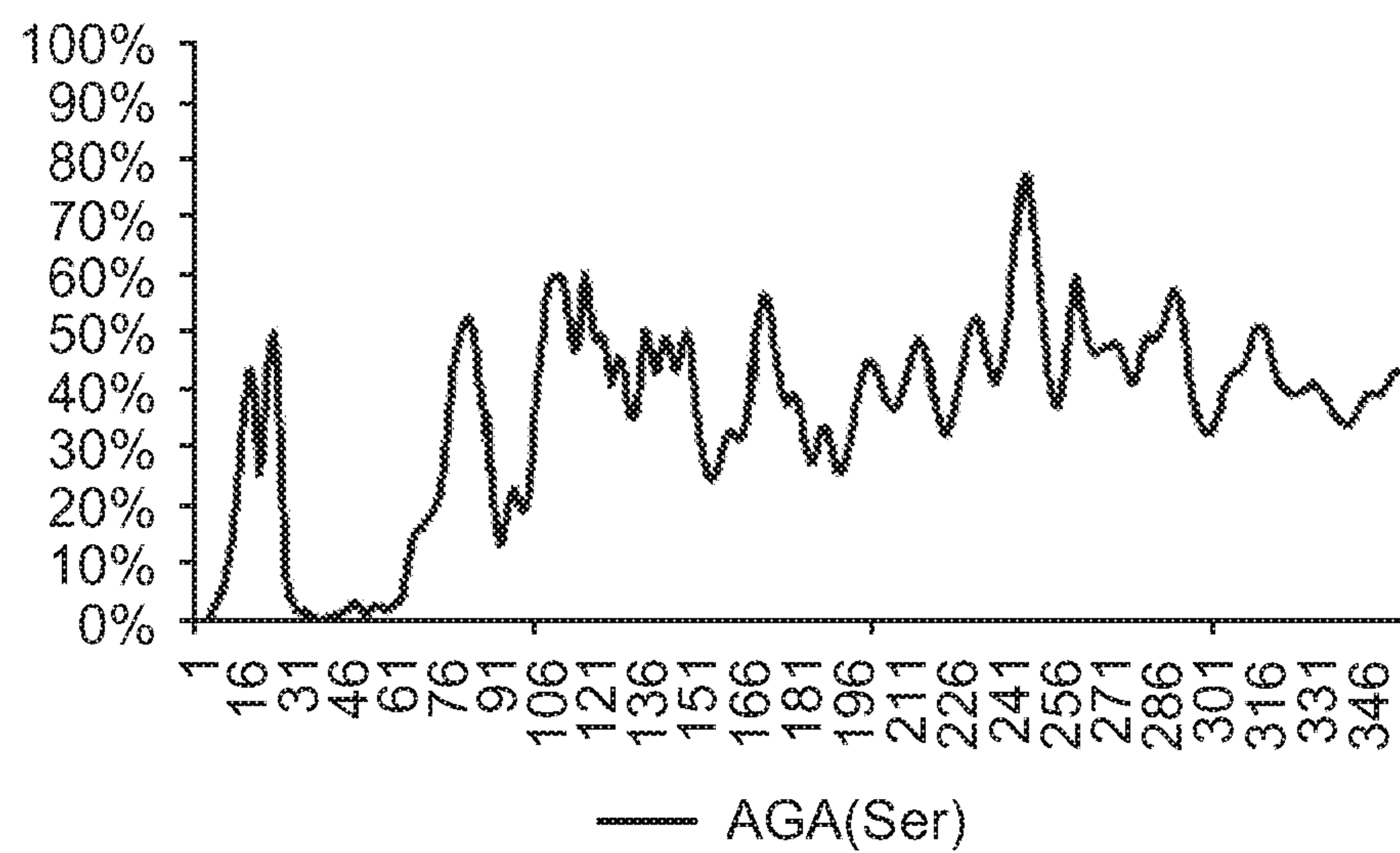
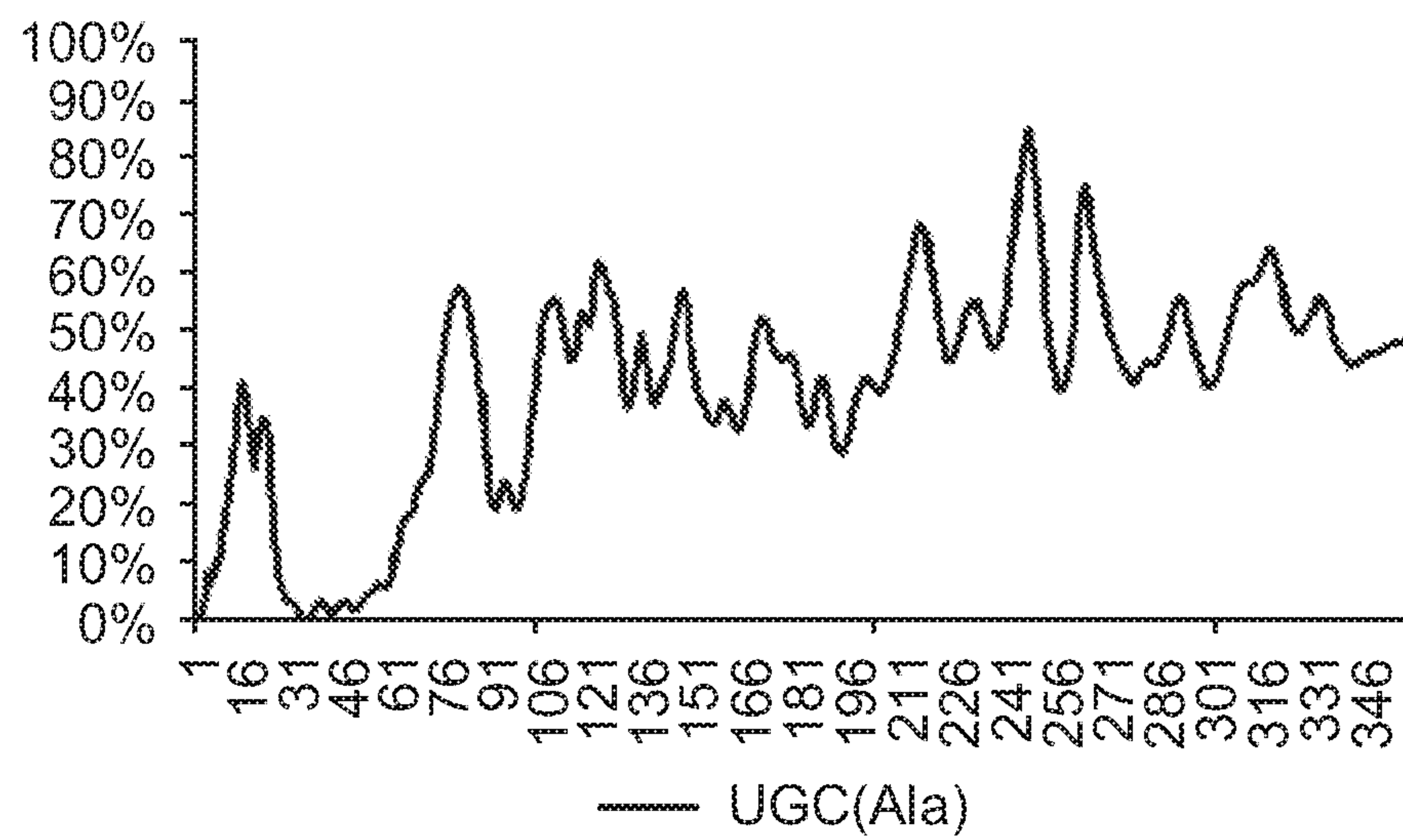
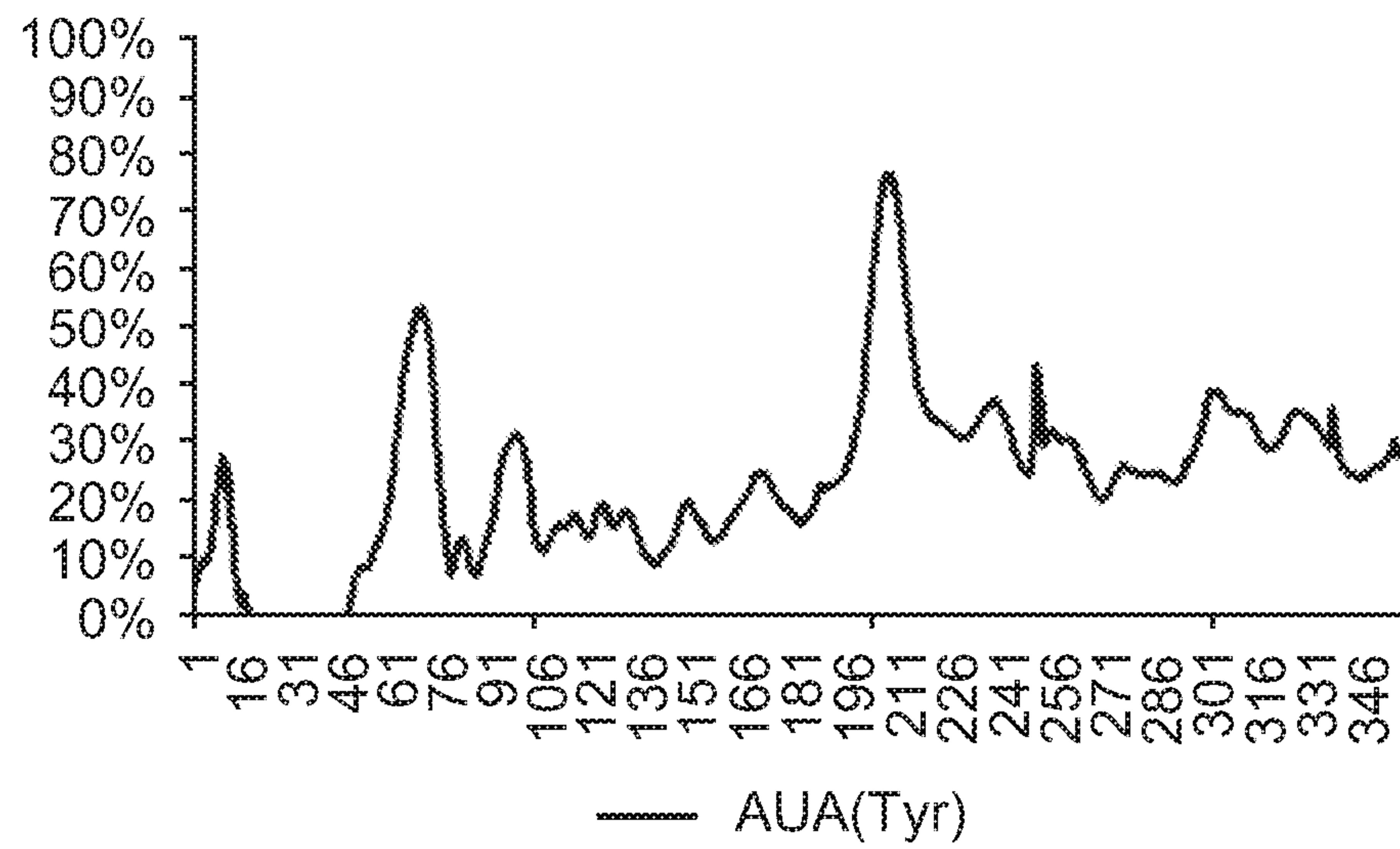
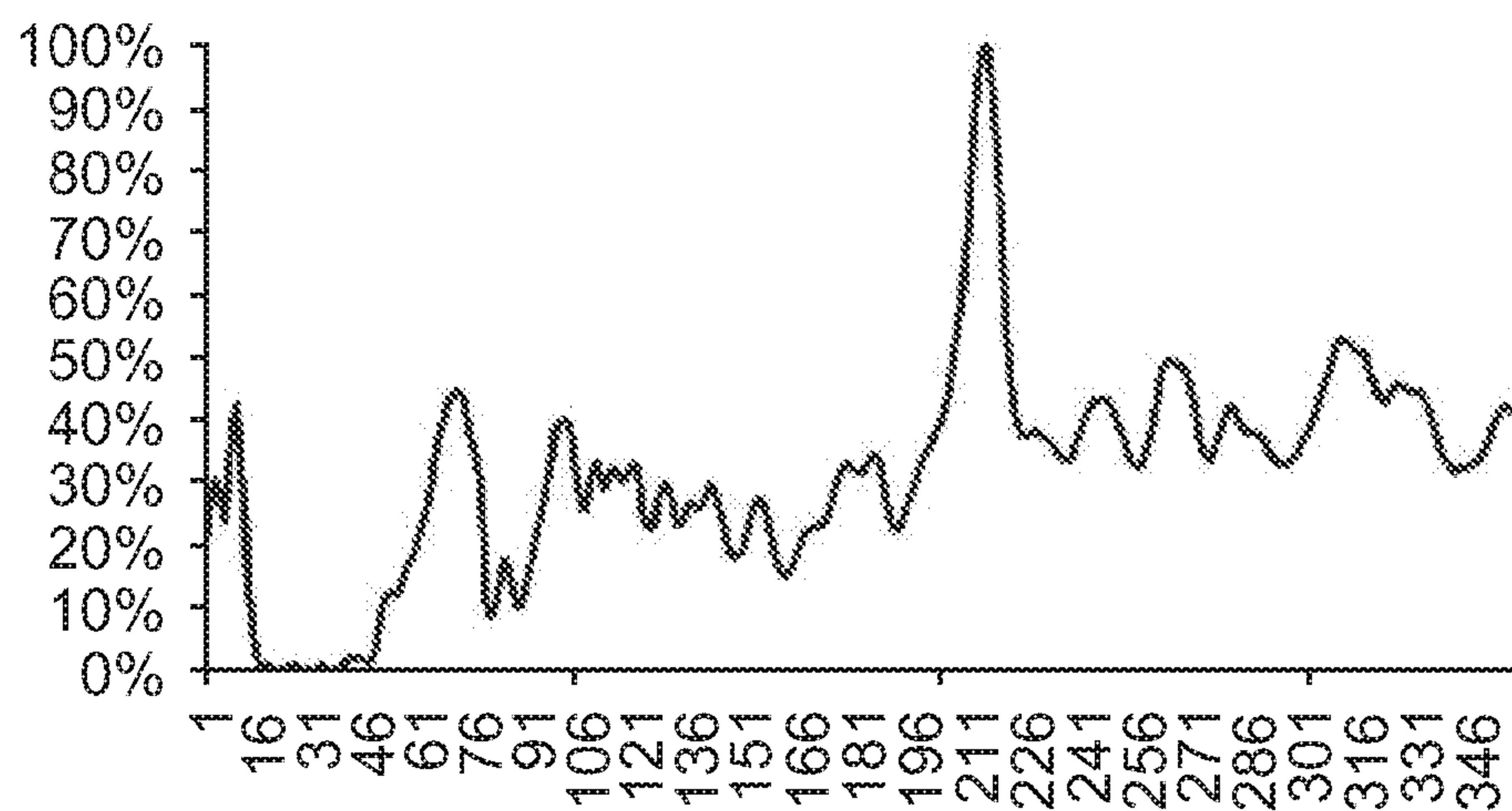


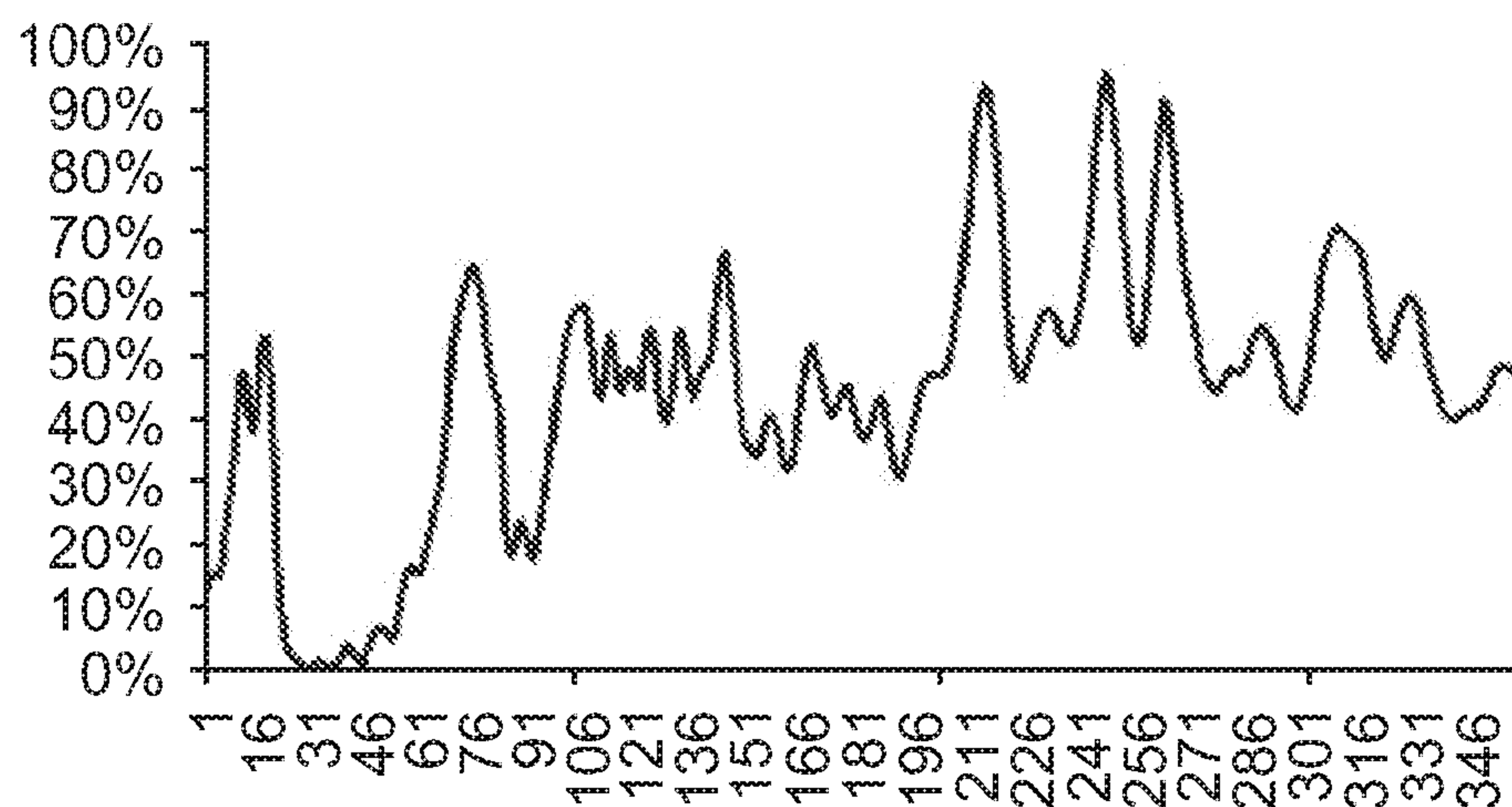
FIG. 4 continued

8 / 20

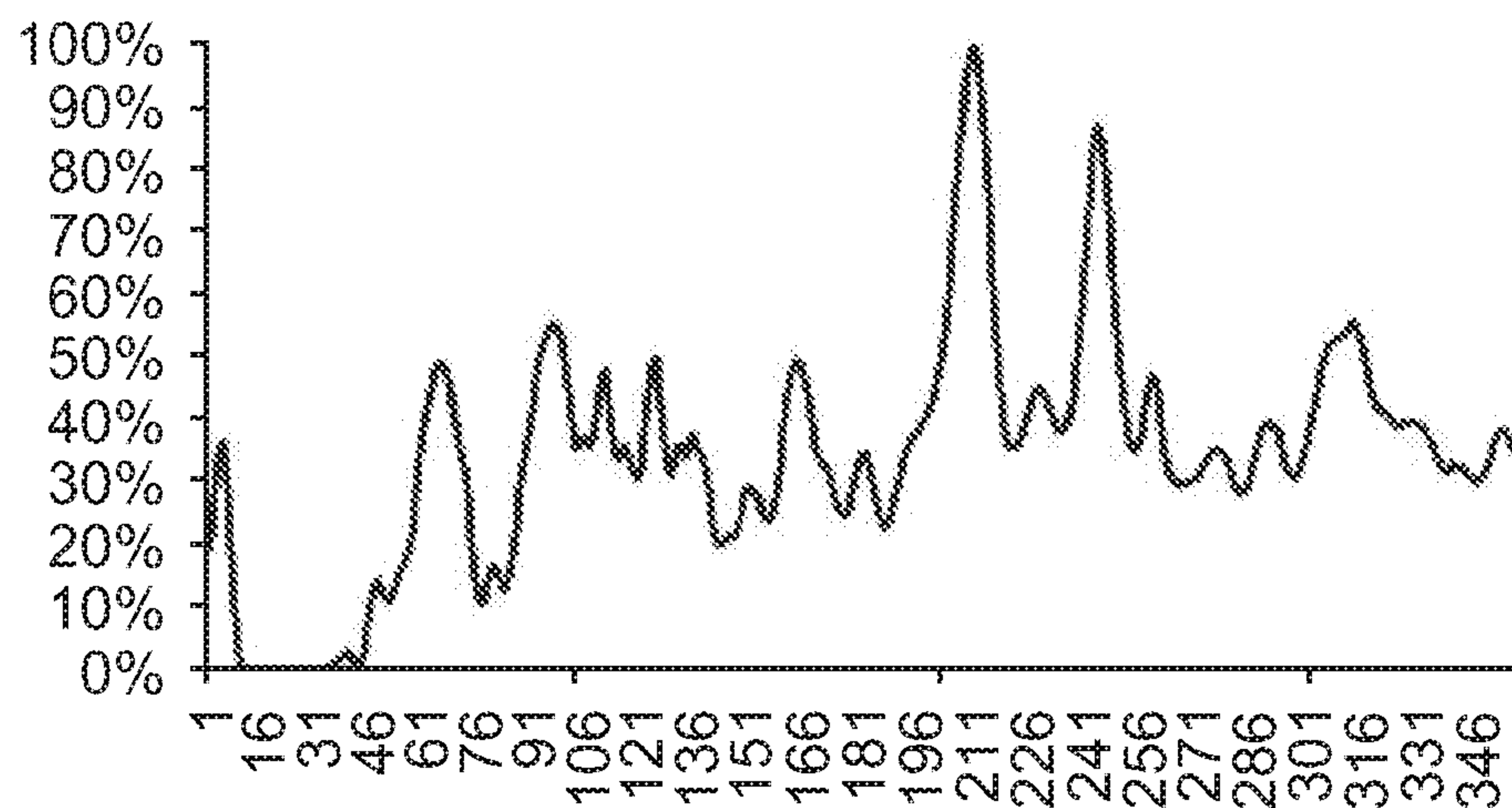
B



— GCU(Ser)



— CAG(Leu)



— CAU(Met)

9 / 20

FIG. 4 continued

B

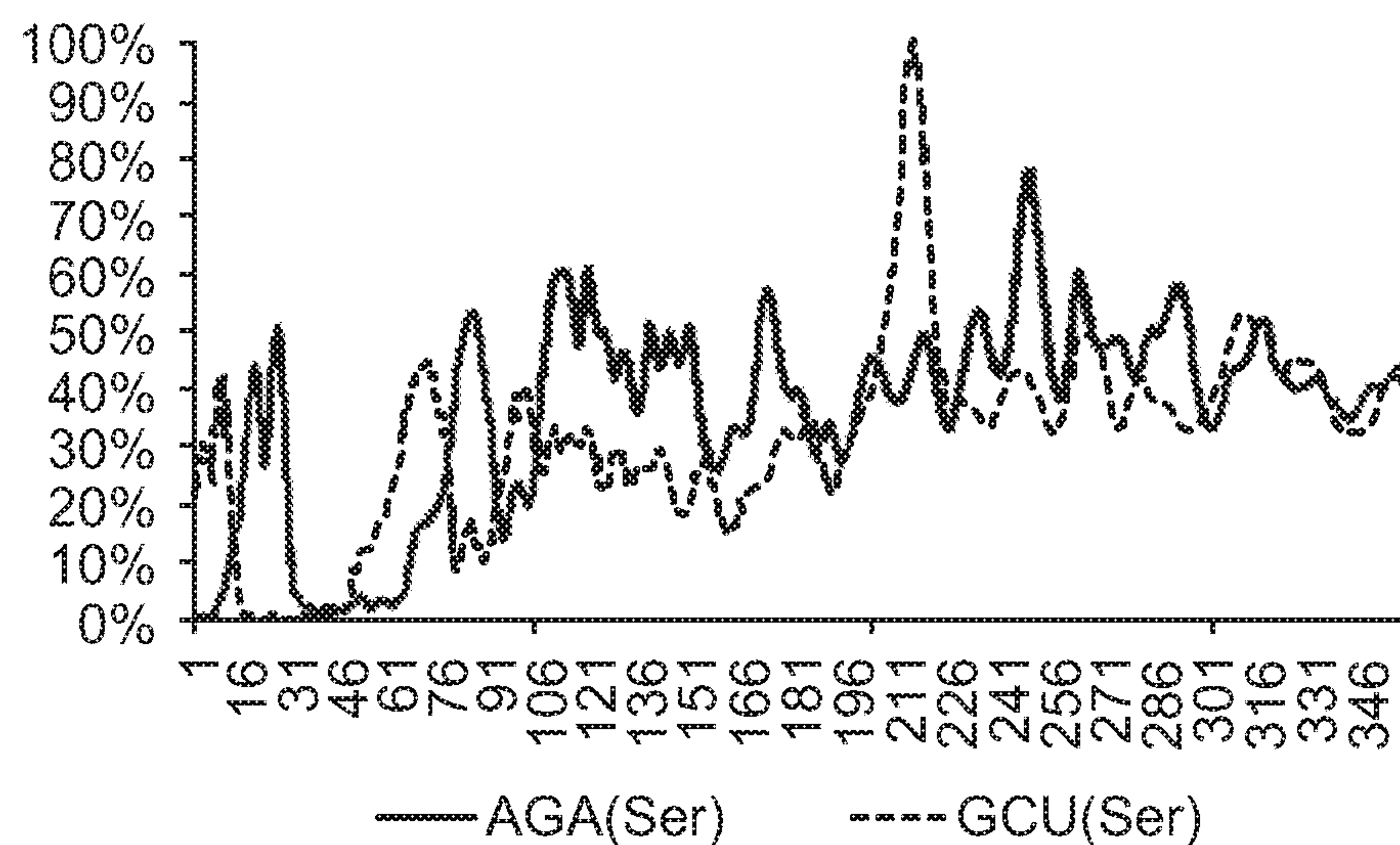
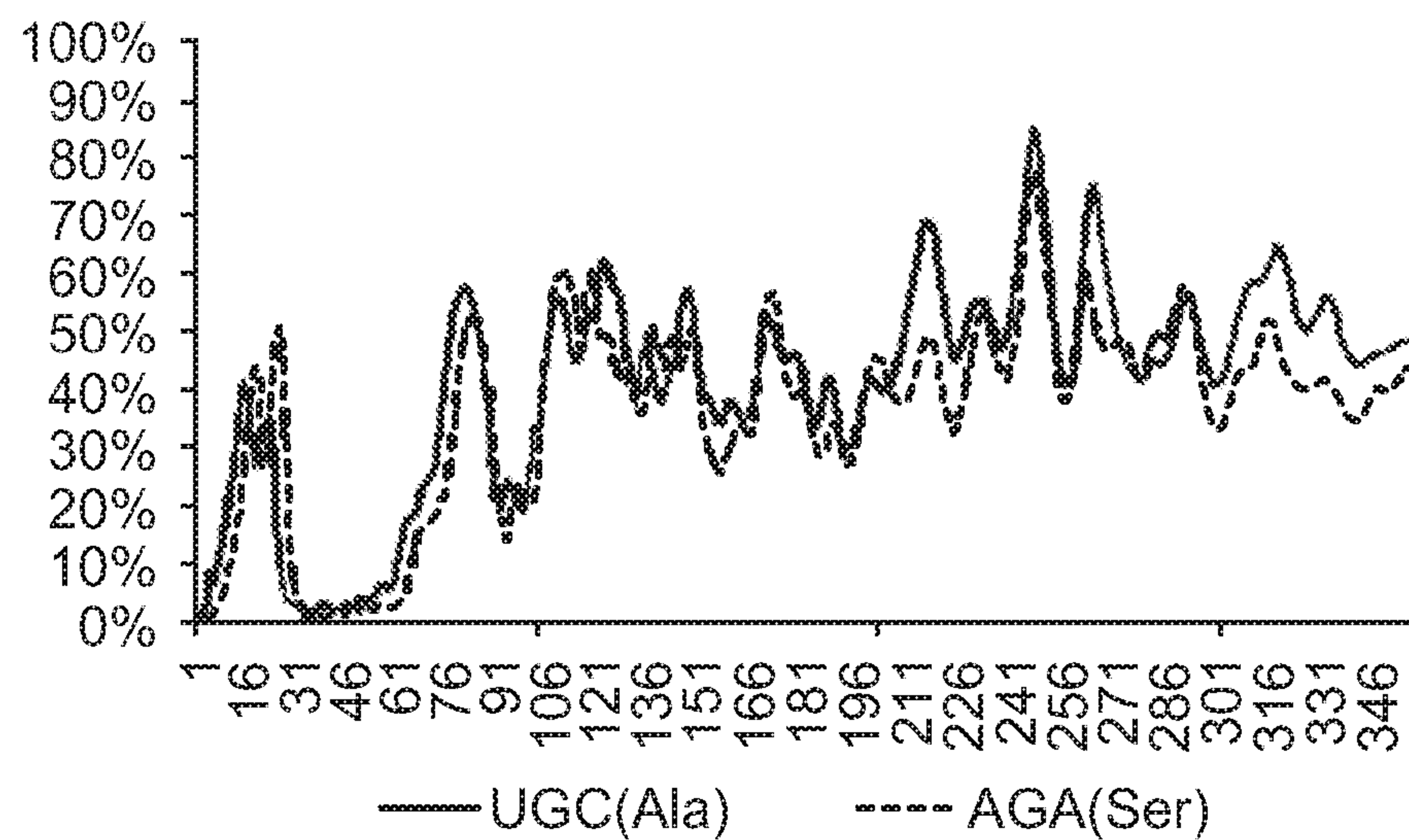
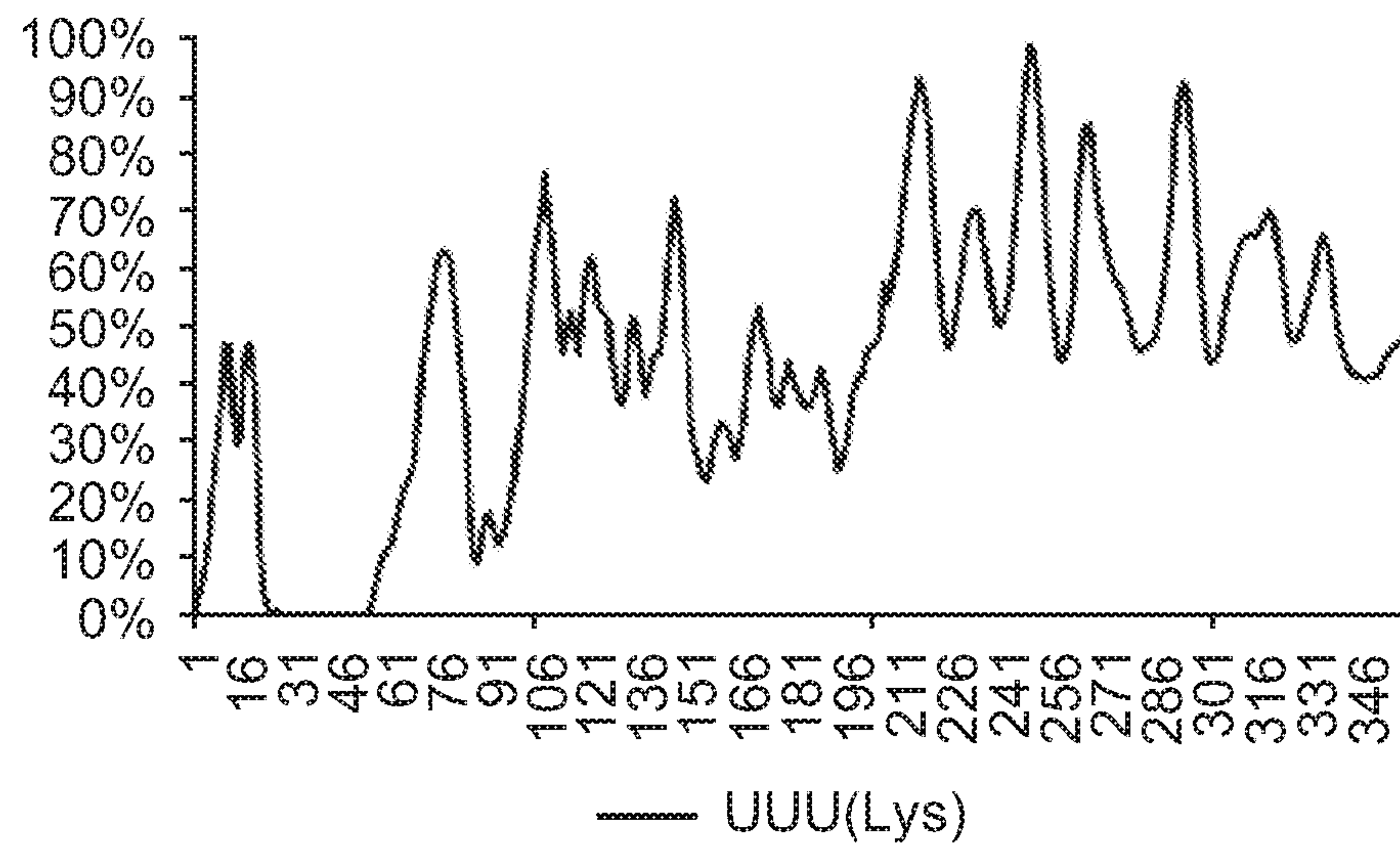


FIG. 5

10 / 20

100

Intensity %

A:19510

B:19620

19000

19500

20000

Mass [Da]

FIG. 6

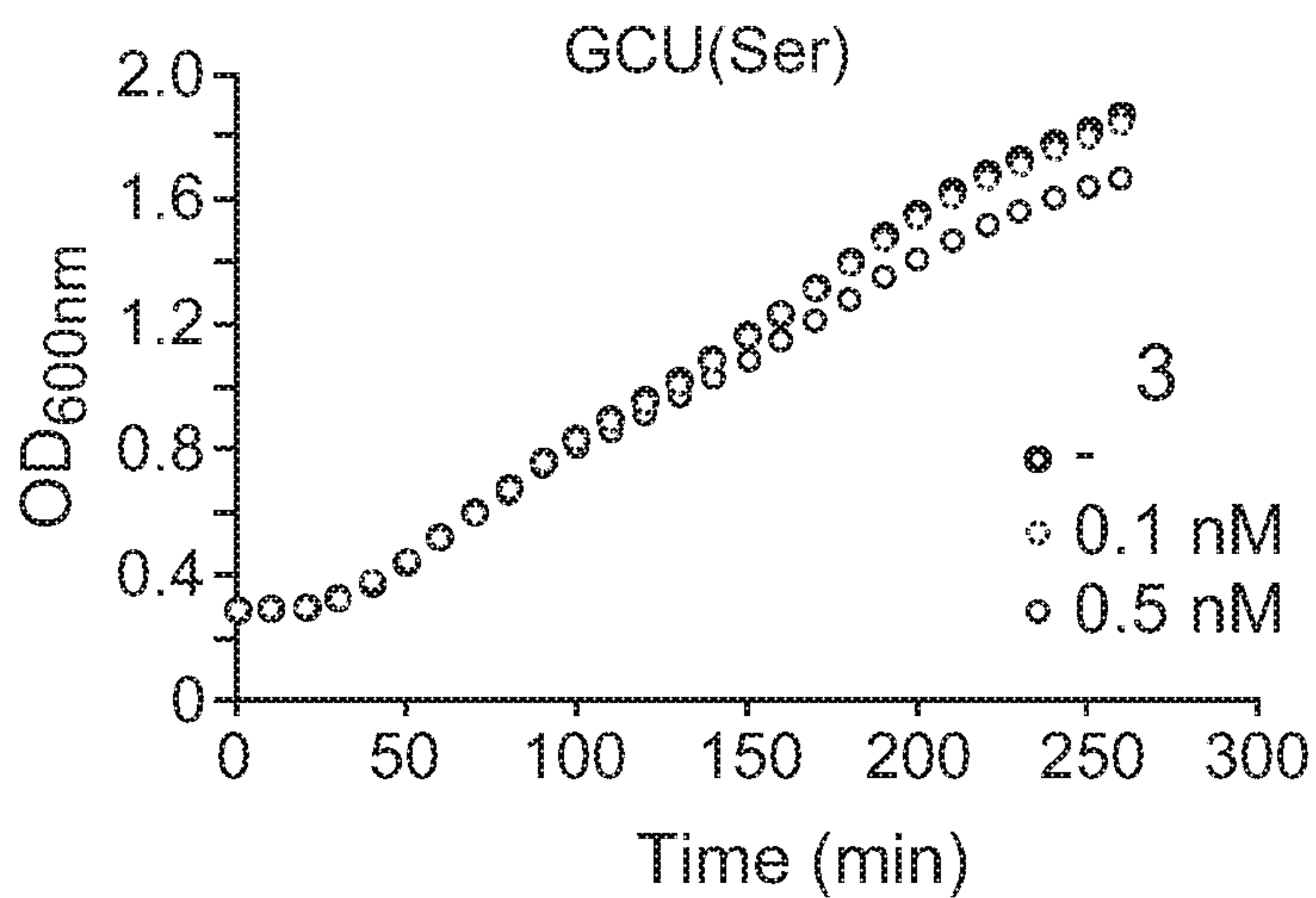
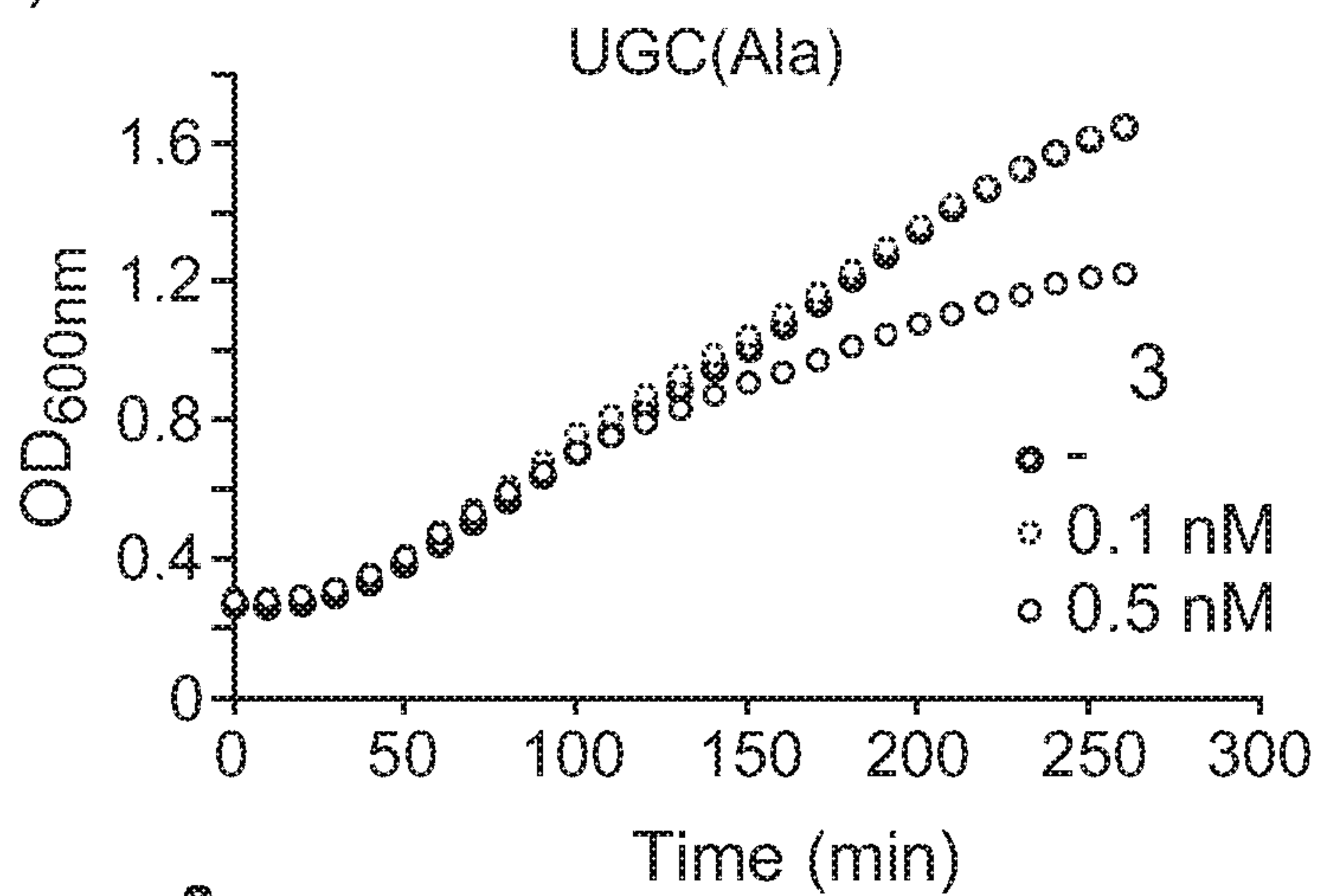
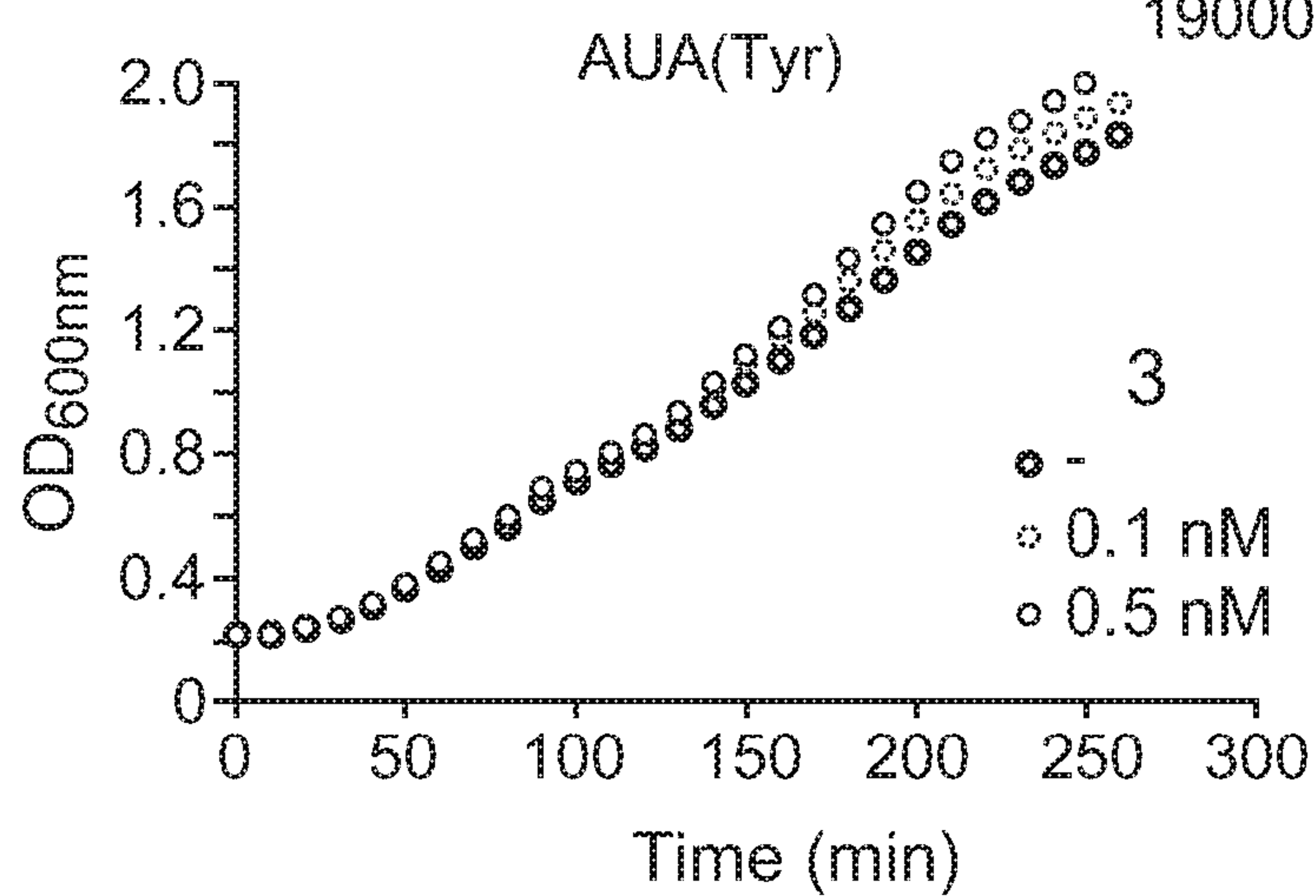


FIG. 6 continued

11 / 20

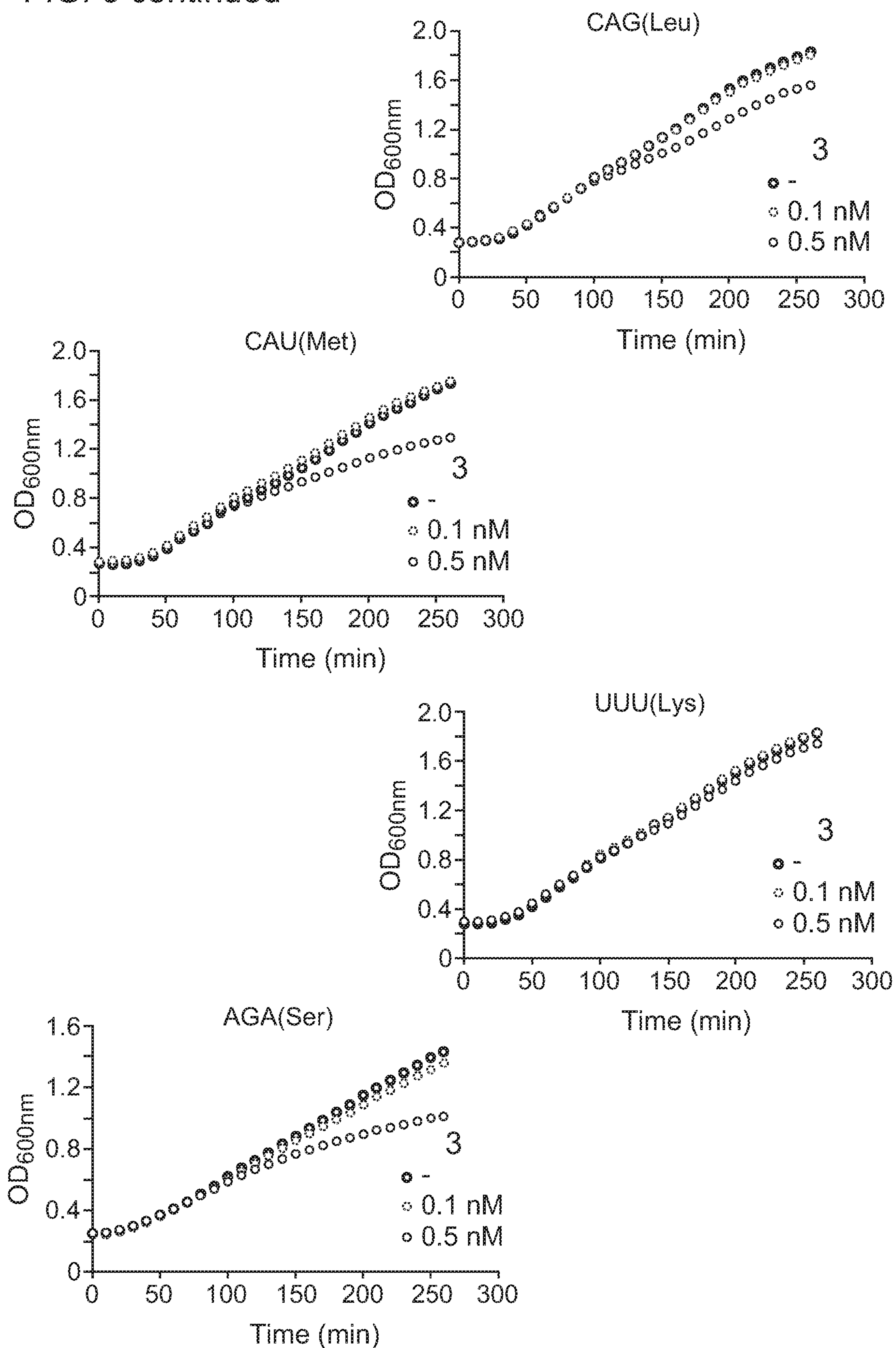


FIG. 7

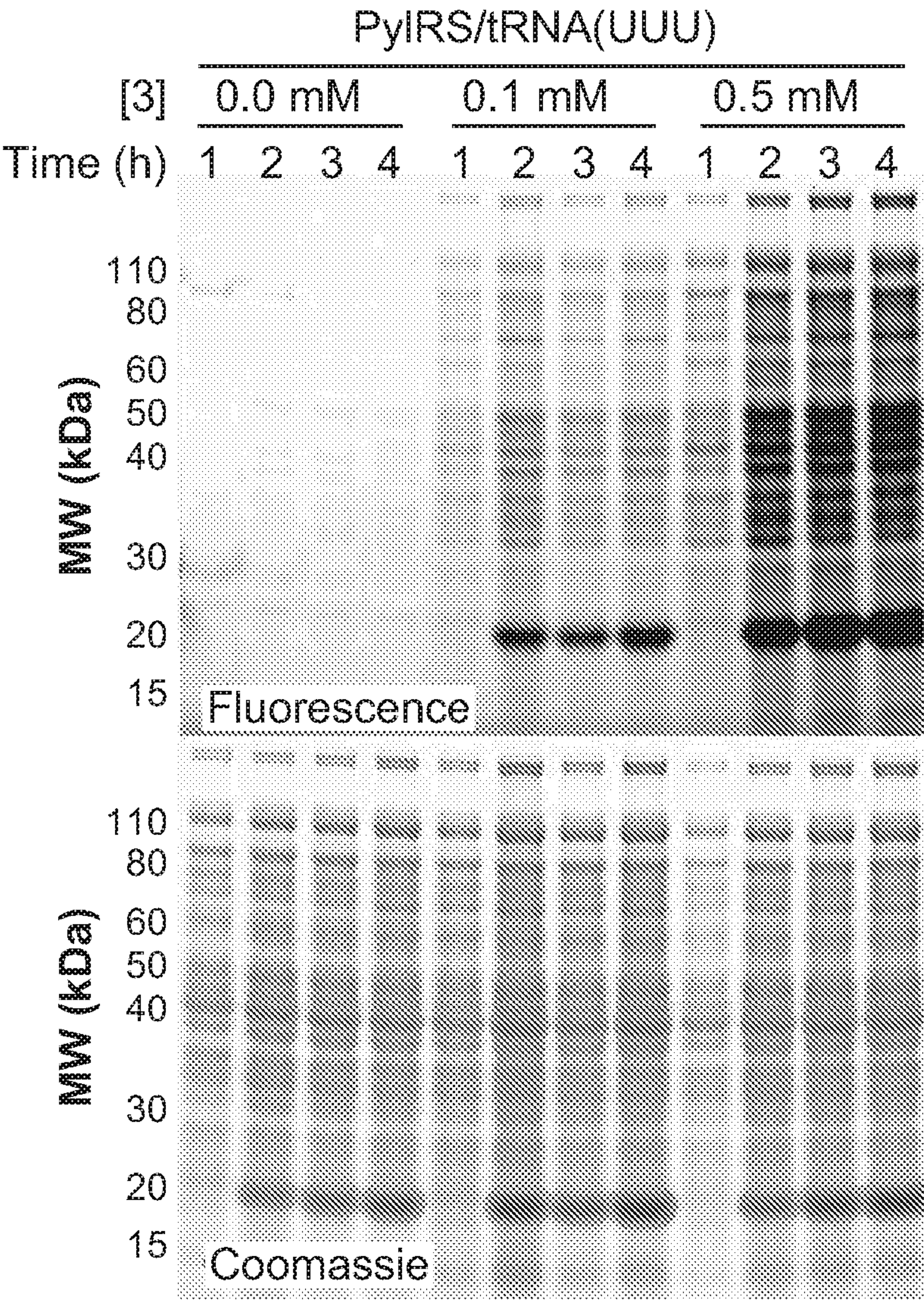


FIG. 8

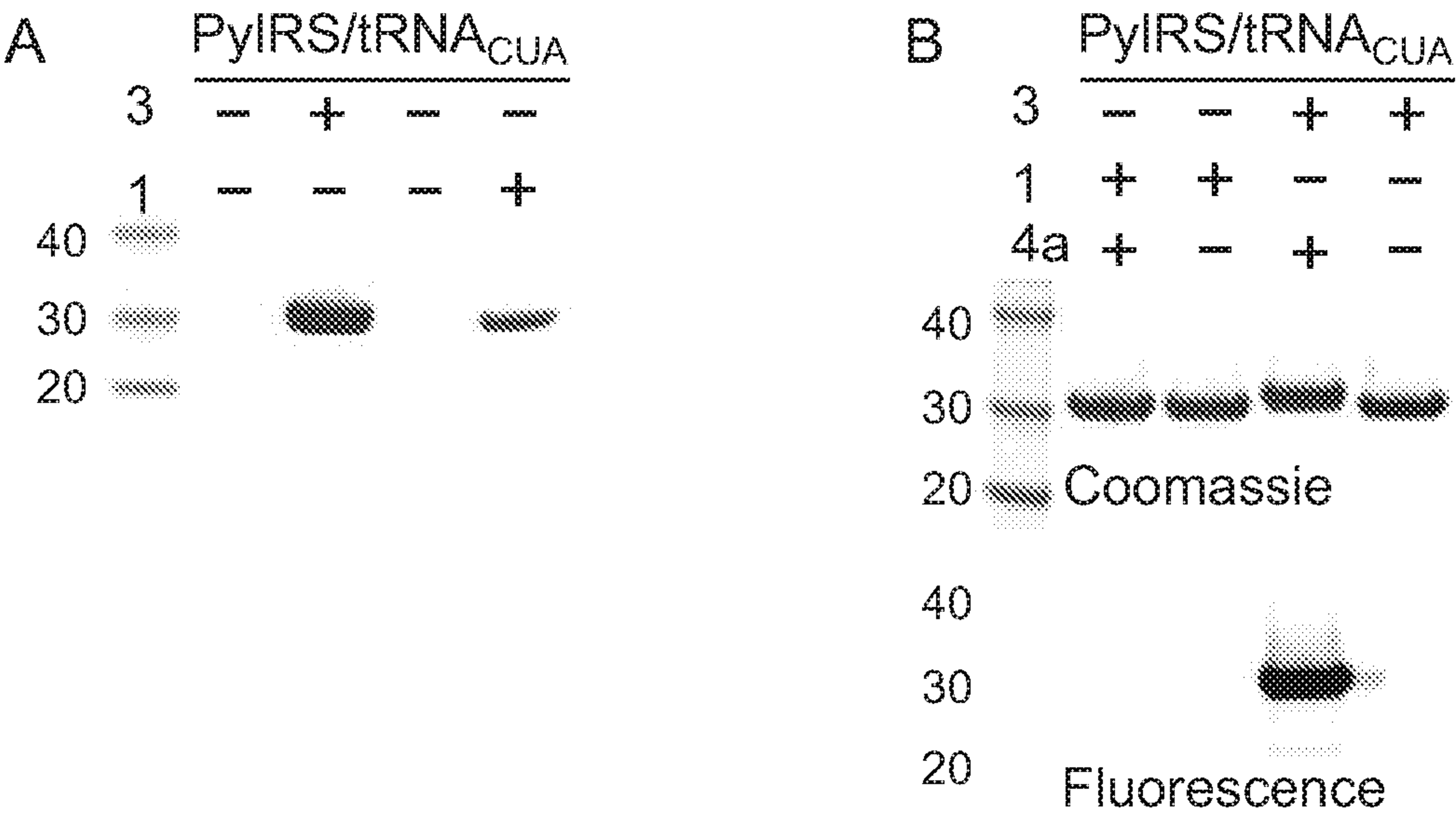
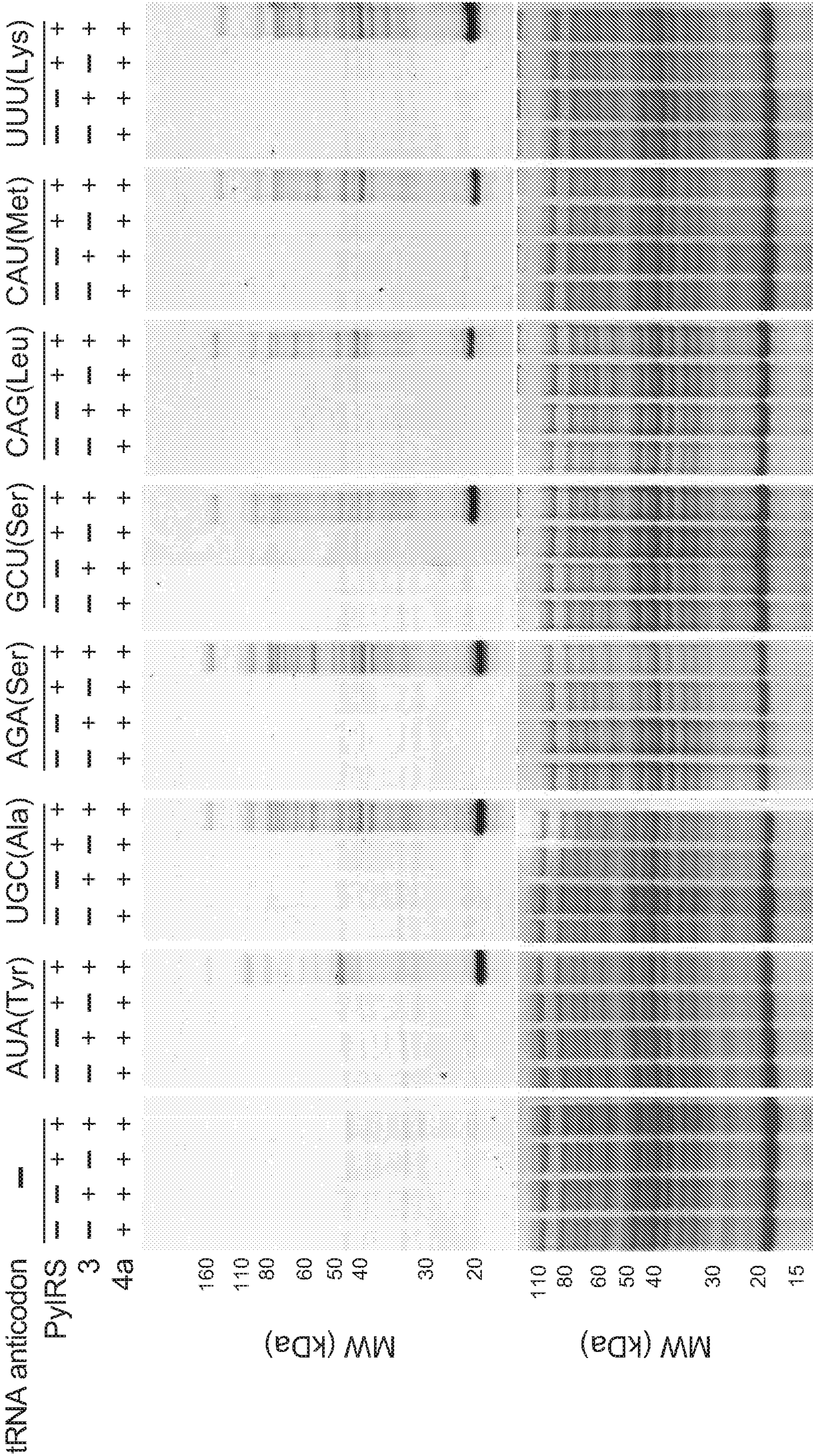


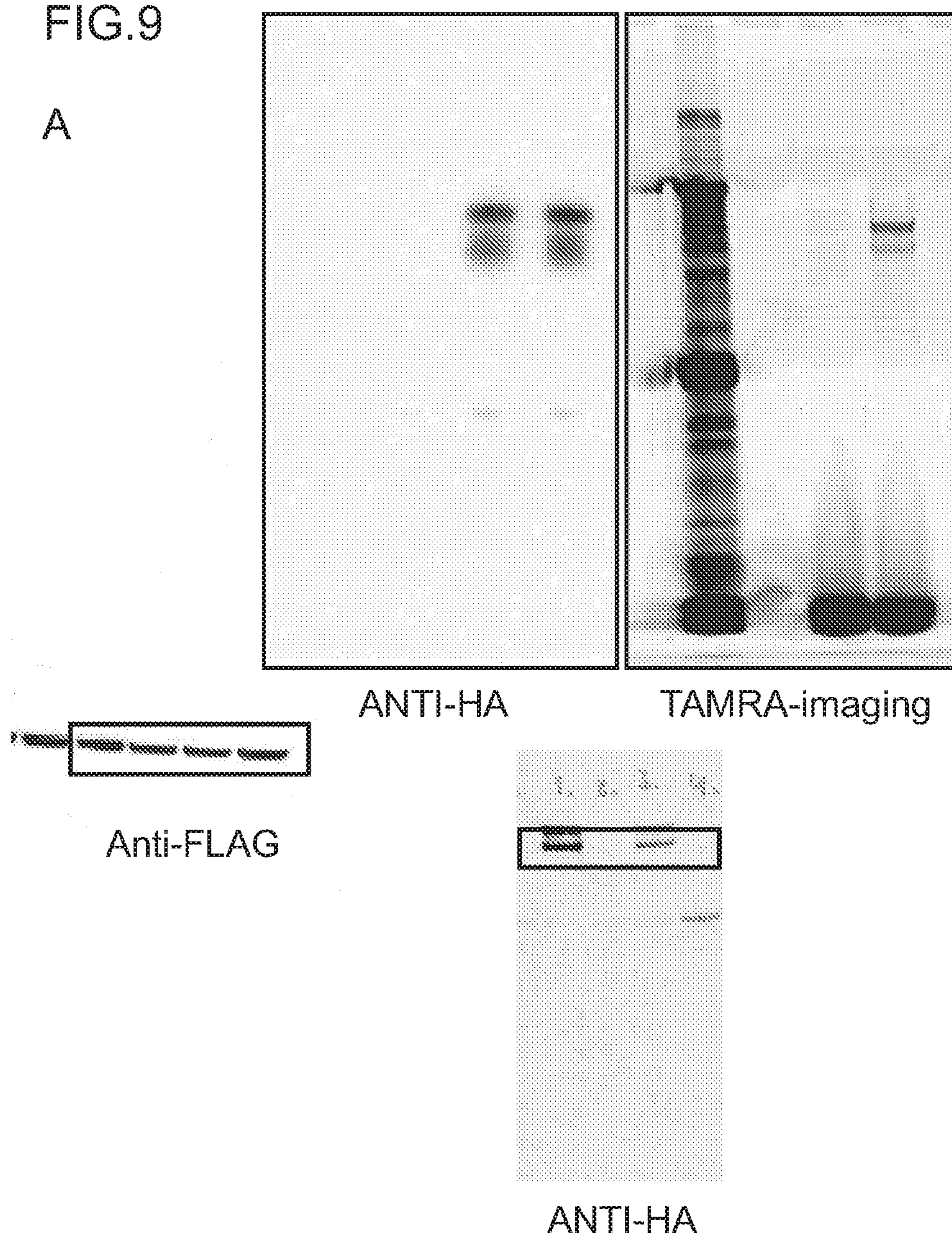
FIG.8
C



14 / 20

FIG.9

A



B

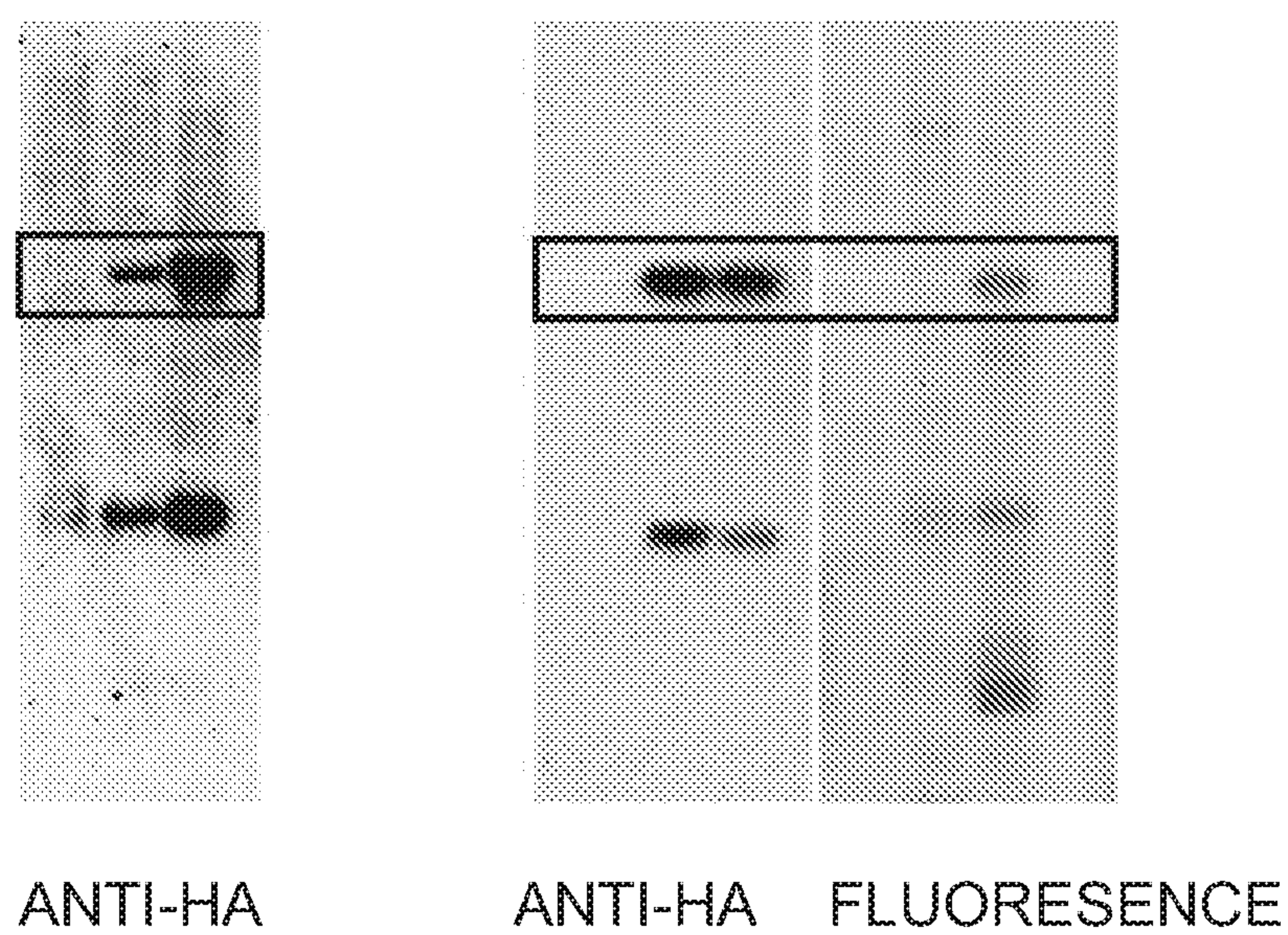


FIG.10

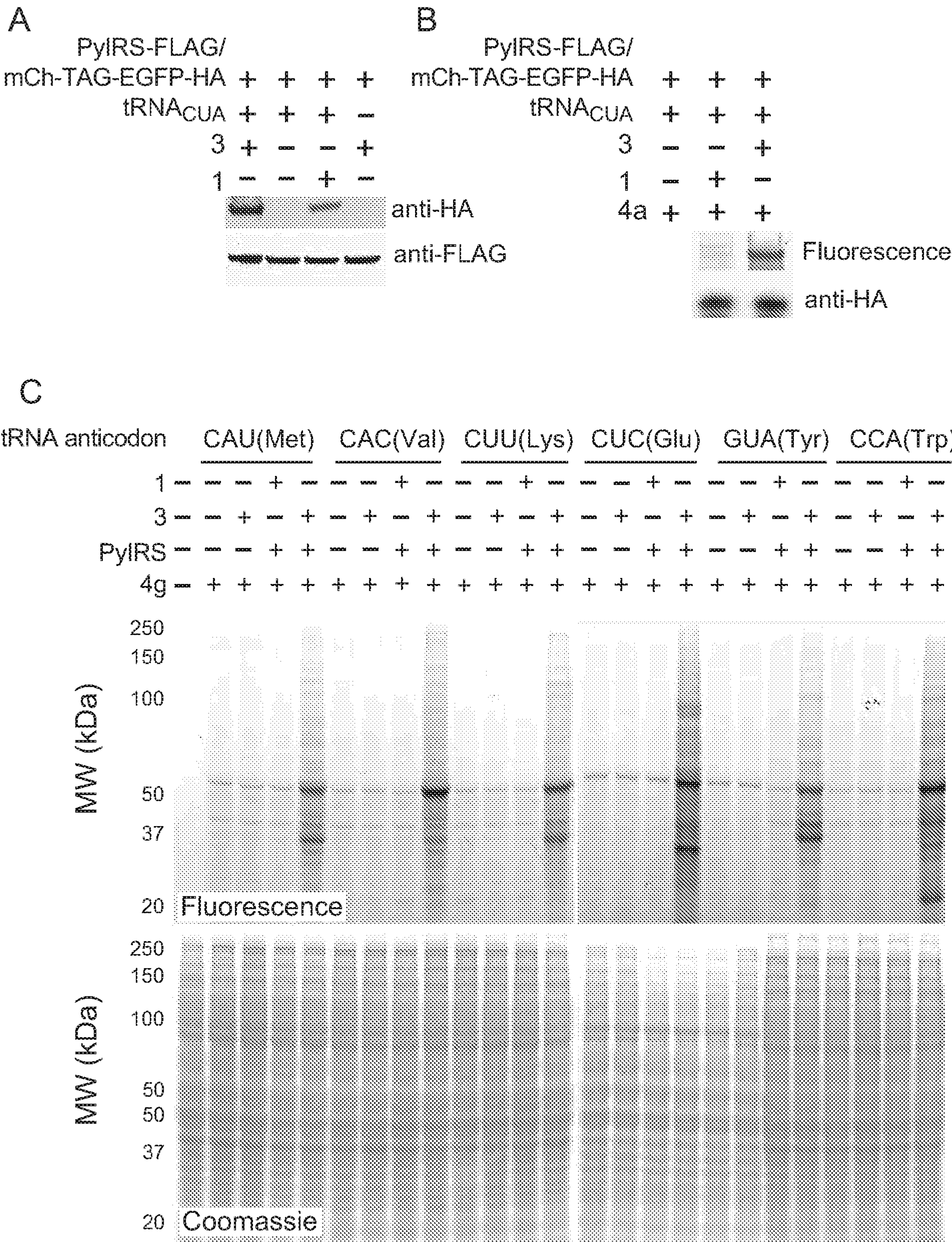


FIG.11

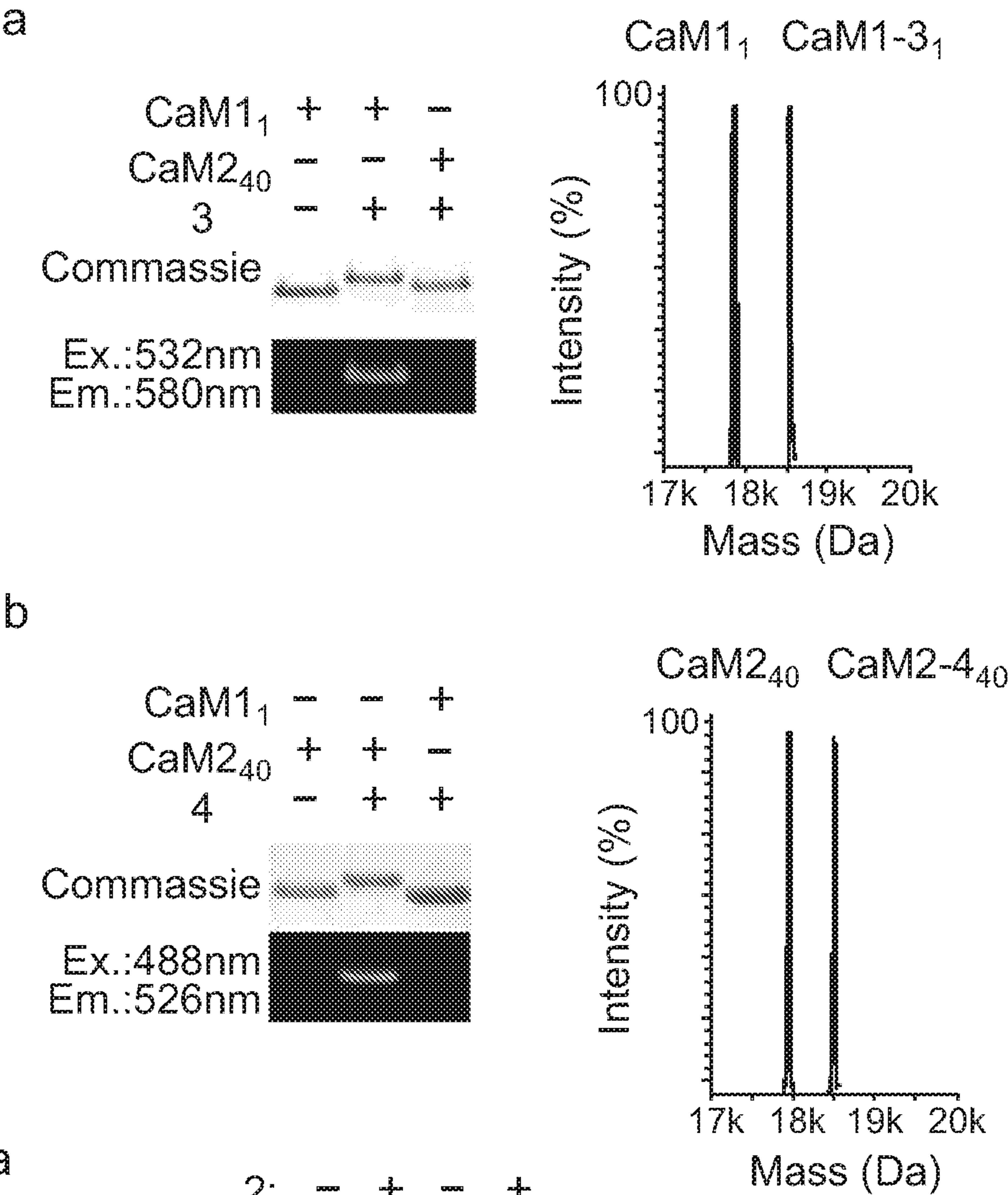


FIG.12

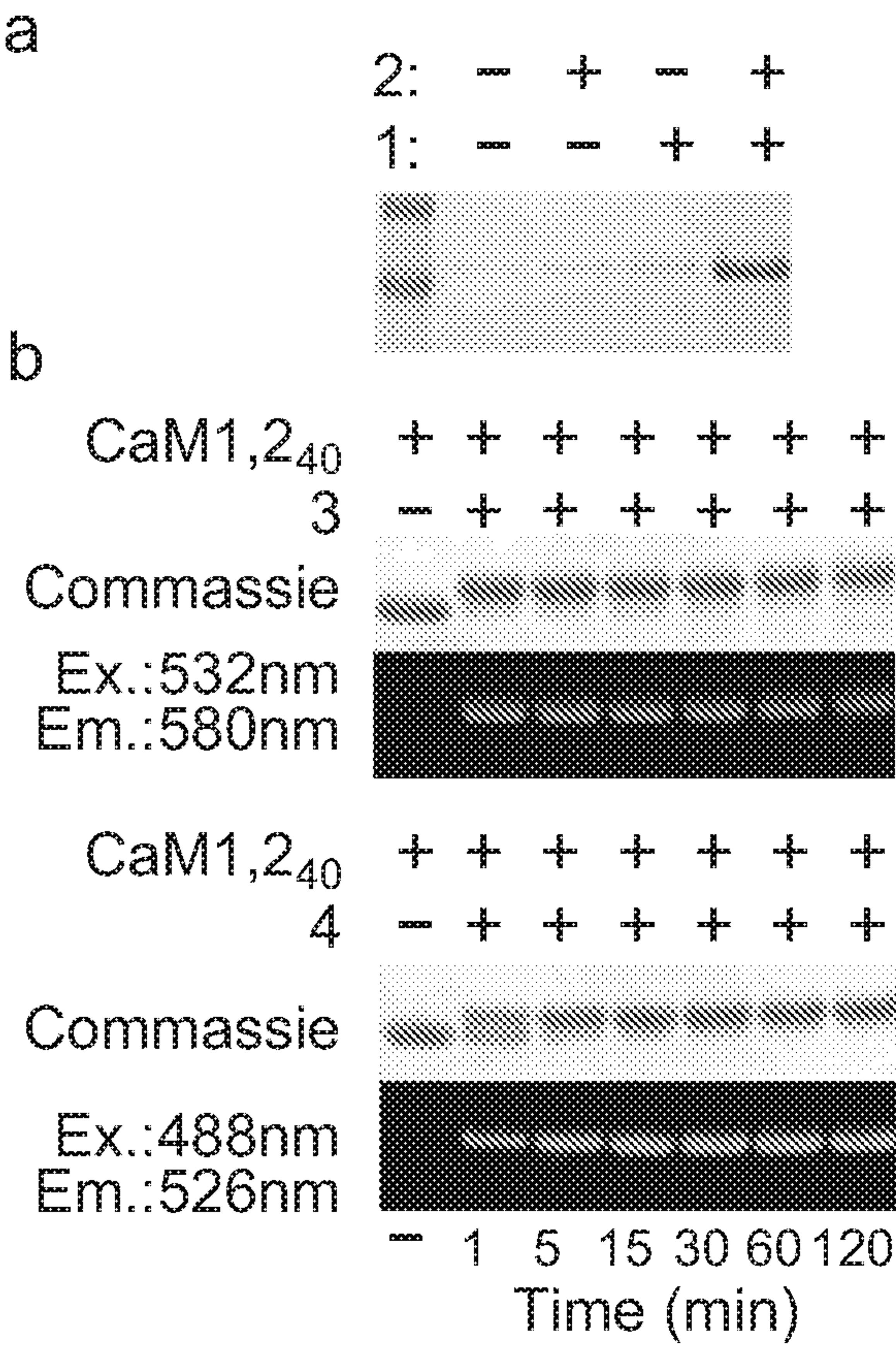
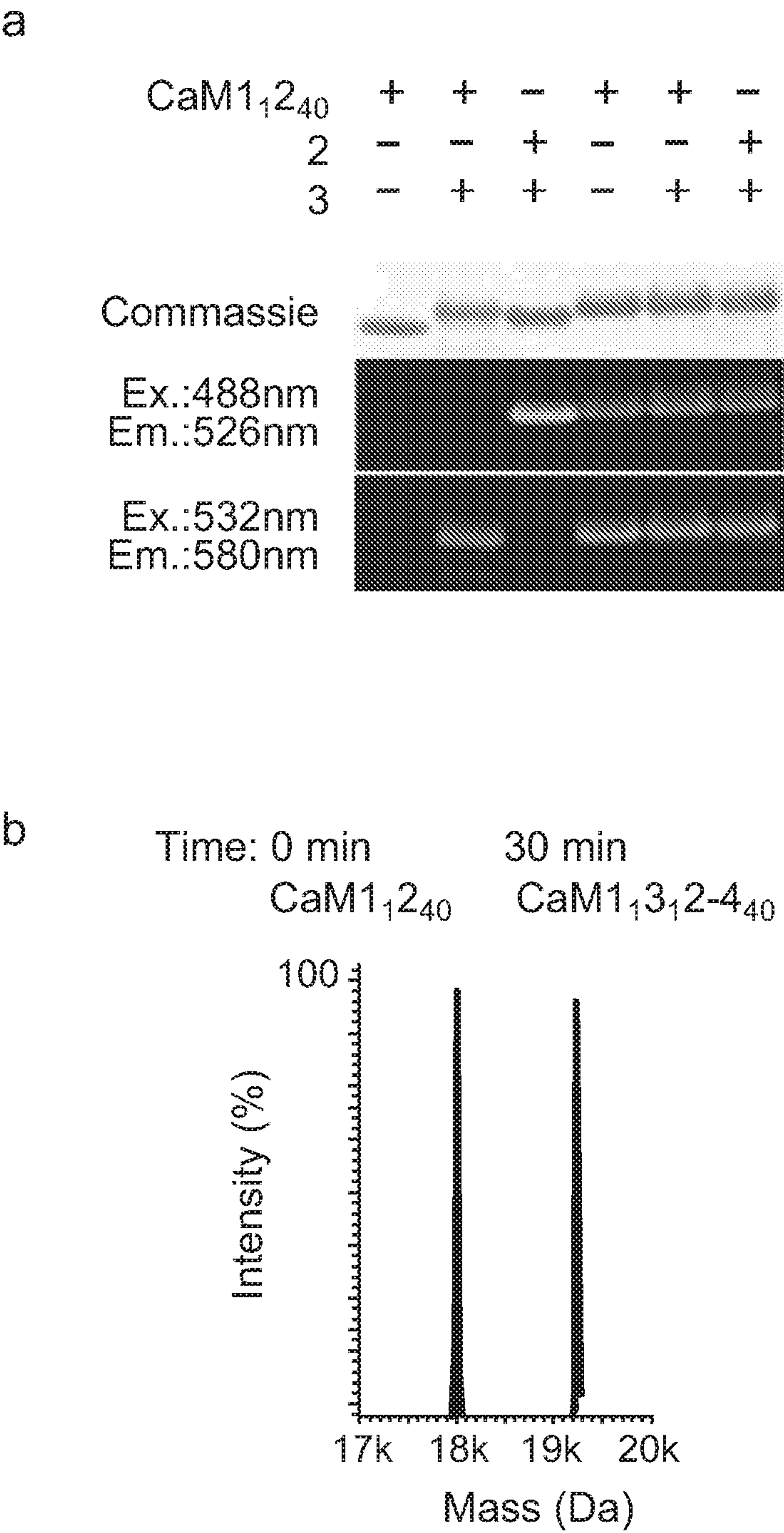
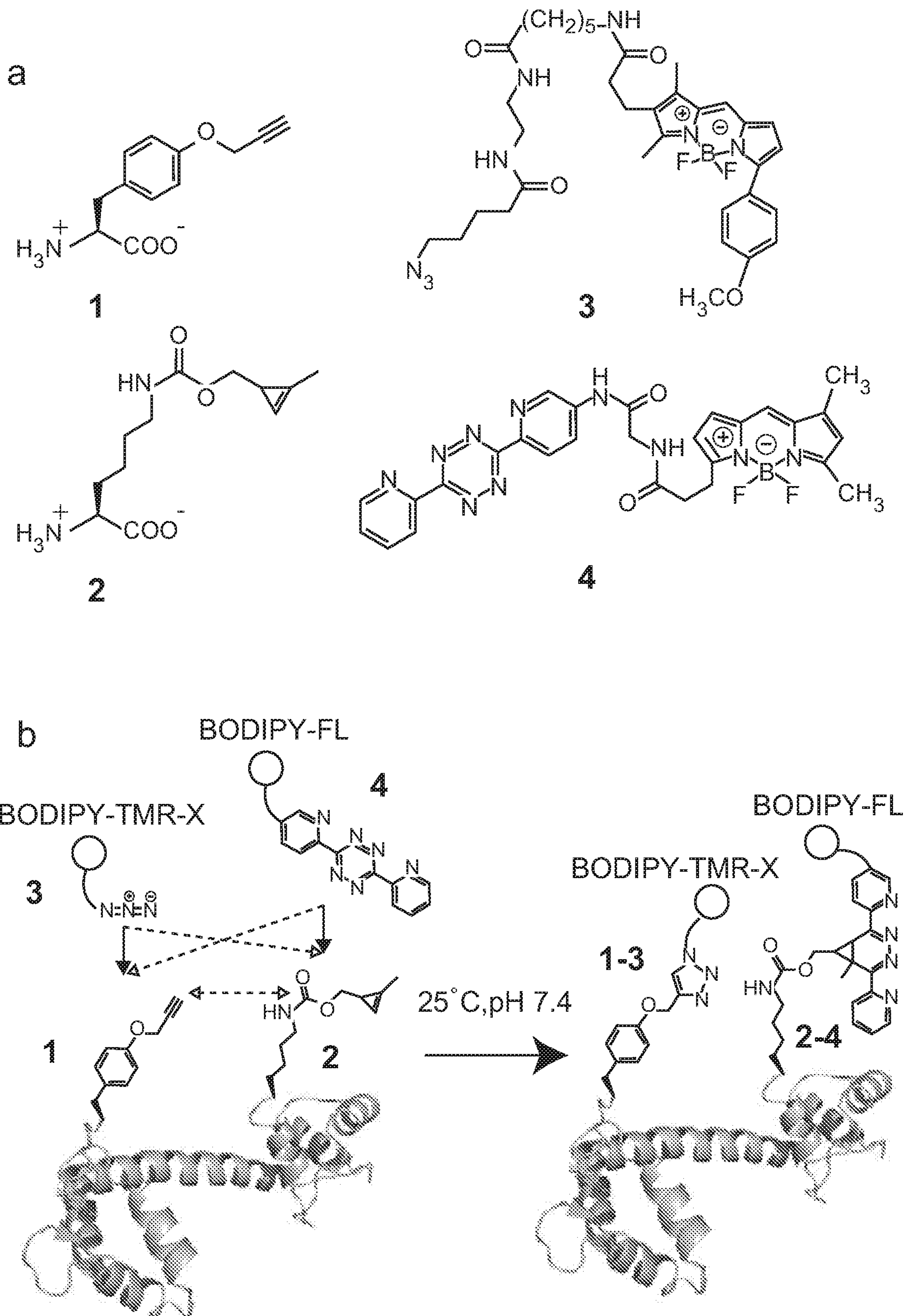


FIG.13



18 / 20

FIG. 14



19 / 20

FIG. 15

atggtgtccaagggcgaggagctgtttaccggcgtggtgcccatttctggtggagctggat
 M V S K G E E L F T G V V P I L V E L D
 ggcgacgtgaacggccacaagttcagcgtgtccggcgagggcgagggcgacgccacctat
 G D V N G H K F S V S G E G E G D A T Y
 ggaaagctgaccctgaagttcatctgcaccaccggcaagctgcccgtgccatggccaacc
 G K L T L K F I C T T G K L P V P W P T
 ctcgtgaccaccctgacctatggcgtgcagtgcttcagccgctaccccgatcacatgaag
 L V T T L T Y G V Q C F S R Y P D H M K
 cagcacgattttcttcaagagcgccatgcccgaggggtacgtgcaggagcgcaccatcttt
 Q H D F F K S A M P E G Y V Q E R T I F
 ttcaaggatgacggcaactacaagacccgcgccgaagtgaagttcgagggcgataccctc
 F K D D G N Y K T R A E V K F E G D T L
 gtgaaccgcatcgagctgaagggcatcgatttcaaggaggatggaaacatcctggggccac
 V N R I E L K G I D F K E D G N I L G H
 aagctggagtacaactacaacagccacaacgtgtacatcatggccgacaagcagaagaac
 K L E Y N Y N S H N V Y I M A D K Q K N
 ggcataaaagtgaacttcaagatccgccacaacatcgaggatggcagcgtgcagctggcc
 G I K V N F K I R H N I E D G S V Q L A
 gatcactaccagcagaacaccccaatcgggcgacggcccagtgctgctgcccgataaccat
 D H Y Q Q N T P I G D G P V L L P D N H
 tacctgagcaccagagcgccctgagcaaggaccccaacgagaagcgcgaccacatggtg
 Y L S T Q S A L S K D P N E K R D H M V
 ctgctggagtttgtgaccgcccggcgccattaccctggggcatggatgagctgtacaagcgc
 L L E F V T A A G I T L G M D E L Y K R
 gccagatccacgatctggtgtaggtgggaggatcgacccgcgtgtcgaagggcgaggag
 A Q I H D L V - V G G S T R V S K G E E
 gataacatggccatcatcaaggagttcatgcgcttcaaggtgcacatggaggggtccgtg
 D N M A I I K E F M R F K V H M E G S V
 aatggacacgagttcgagattgagggcgagggcgagggacgcccataatgaggggaacccag
 N G H E F E I E G E G E G R P Y E G T Q
 accgccaagctgaaagtgaccaagggcggaacccctgcccttcgcctgggatattctgagc
 T A K L K V T K G G P L P F A W D I L S
 cccagtttatgtacggcagcaaggcctacgtgaagcaccgccgatatcccgattac
 P Q F M Y G S K A Y V K H P A D I P D Y
 ctgaagctgagcttcccagaggggttcaagtgggagcgcgtgatgaatttcgaggacggc
 L K L S F P E G F K W E R V M N F E D G
 ggagtcgtgaccgtgaccaggatagcagtttgcaggatggcgagttcatctacaaagtg
 G V V T V T Q D S S L Q D G E F I Y K V
 aagctgcgcggcaccaacttcccgtccgatggcccagtgatgcagaagaagaccatgggc
 K L R G T N F P S D G P V M Q K K T M G

20 / 20

FIG. 15 continued

tgggaggccagcagcgagcgcgcgtatccagaggatggcgccctgaagggcgagatcaag
 W E A S S E R M Y P E D G A L K G E I K
 cagcgccctgaagctgaaggatggcgggccactacgatgccgaagtgaagaccacctacaag
 Q R L K L K D G G H Y D A E V K T T Y K
 gccagaagccggtgcagctgccaggcgcctacaatgtgaacatcaagctggatatcacc
 A K K P V Q L P G A Y N V N I K L D I T
 tcccacaacgaggactacaccatcgtggagcagtatgagcgcgccgagggccgccatagt
 S H N E D Y T I V E Q Y E R A E G R H S
 accggcggaatggacgagctgtataagatgtaccctacgatgtgcccgattacgccgag
 T G G M D E L Y K M Y P Y D V P D Y A E
 cagaagctgatctccgaggaggacctgcaccatcaccaccaccacggaagtggcagcggc
 Q K L I S E E D L H H H H H H G S G S G
 tccccaaagaagaagcgcaaggtgtaa
 S P K K K R K V -

FIG. 16

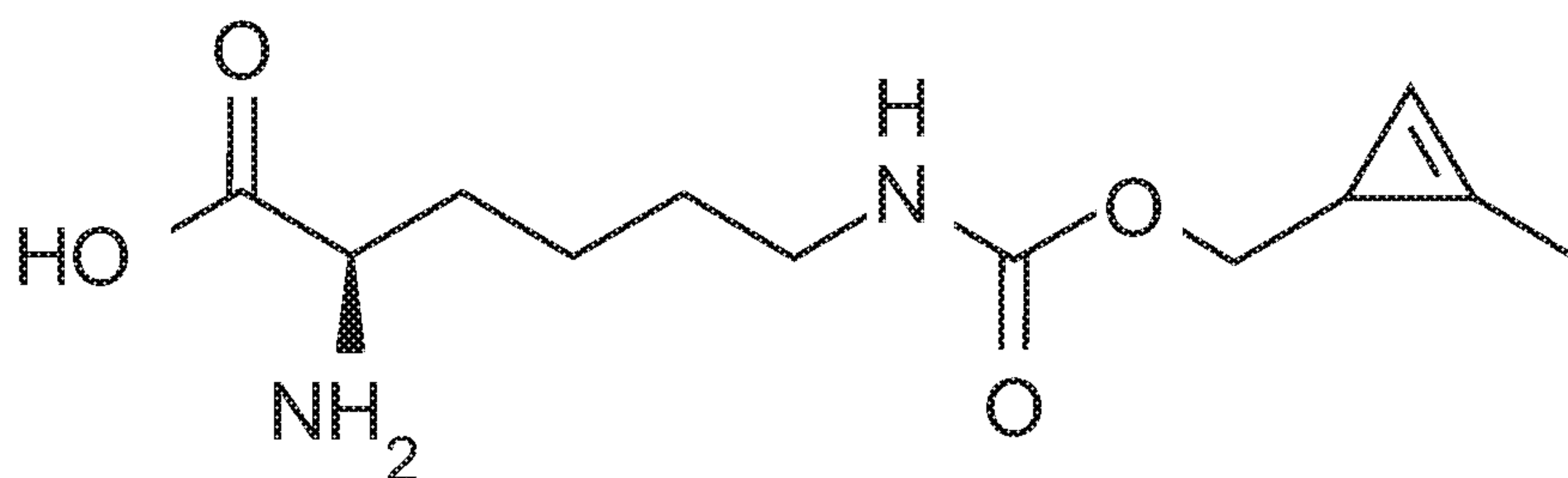


FIG. 1

A

