

BRAZZEIN VARIANTS

TECHNICAL FIELD

5 The present disclosure relates broadly to protein variants. In particular, the present disclosure relates to brazzein variants.

BACKGROUND

10 The dramatic rise in obesity and diabetes in recent decades has seen the widespread use of artificial sweeteners in food and drinks as sugar replacements. These artificial sweeteners regulate blood glucose level and provide a means to reduce calorie consumption whilst maintaining food palatability with its sweet taste. Emerging evidence, however, points to increased risks in cancer and cardiovascular diseases, and these
15 sweeteners have been shown to contribute to other adverse health effects such as gastrointestinal issues. Thus, highlighting the need for other sweeteners.

 Sweet proteins of natural origin have the potential to replace these artificial sweeteners due to their intensely sweet nature and low risk safety profile. Unlike sucrose, sweet proteins do not trigger a demand for insulin in diabetic patients. So far, seven
20 vastly different sweet-tasting proteins have been discovered from plants located in tropical rainforests. These are brazzein, thaumatin, monelin, curculin, mabinlin, miraculin, and pentadin.

 Brazzein is a relatively small, sweet protein comprising of 54 amino acids and possesses an intense sweetness that is 500 to 2000 times over sucrose. Originally
25 isolated from the fruit of the west African plant, *Pentadiplandra brazzeana* Bailon, brazzein's heat and pH stability make it an ideal system for application in the biotechnology and food processing industries. It is recently found that brazzein interacts with the human sweet taste receptor TAS1R2/TAS1R3 to bring about sweet taste perception in humans and are not associated with adverse health effects.

30 While brazzein is a promising alternative for sugars and artificial sweeteners, natural production of brazzein can be expensive. Therefore, to meet the demand for an alternative for sugars and artificial sweeteners, there is a need to provide a brazzein variant.

SUMMARY

In one aspect, there is provided a brazzein variant comprising

X₁X₂KCKX₆X₇YX₉NYPX₁₃X₁₄KCX₁₇X₁₈AX₂₀QCNYDCKLDKHARX₃₄GECFYDEX₄

5 2RNX₄₅X₄₆CX₄₈CDYCX₅₃YX₅₅ (SEQ ID NO: 1),

wherein X₁, X₂₀, and X₄₇ are each independently an amino acid with polar uncharged side chain, X₂ is an amino acid with a negatively charged side chain, X₆ is a glutamine (E) or an amino acid with positively charged side chains, X₇, X₁₃, X₁₈, X₄₅, and X₄₈ are each independently an amino acid with hydrophobic side chain, X₁₇ is a lysine (K) or an amino acid with polar uncharged side chain, X₁₄ is an arginine (R) or an amino acid with polar uncharged side chain, X₉ is an arginine (R) or an amino acid with negatively charged side chain, X₄₂ is an amino acid with positively charged side chain, X₃₄ is an alanine or an amino acid with polar uncharged side chain, X₄₆ is an amino acid with polar uncharged side chain, X₅₃ is a glutamic acid, a glutamine, or a lysine, and X₅₅ is a proline, an arginine, a threonine, or no residue.

In some examples, the variant has at least 85% sequence identity to a wild type brazzein.

In some examples, the variant has 4 to 8 amino acids differences from a wild type brazzein.

20 In some examples, the variant retains substantially the same structure as wild type brazzein.

In some examples, the variant has the same secondary structure as wild type brazzein.

In some examples, the variant is capable of binding to a taste receptor.

25 In some examples, the variant is capable of binding to a human sweet taste receptor.

In some examples, the variant comprises

X₁ is a glutamine (Q) or a serine (S),

X₂ is an aspartic acid (D) or a glutamic acid (E),

30 X₆ is a glutamine (Q), a lysine (K), or an arginine (R),

X₇ is valine (V), methionine (M), or isoleucine (I)

X₉ is a glutamic acid (E) or an arginine (R),

X₁₃ is isoleucine (I) or valine (V),

X₁₄ is an asparagine (N), a serine (S), or an arginine (R),

35 X₁₇ is a lysine (K) or a glutamine (Q),

X₁₈ is a leucine (L) or isoleucine (I),

X₂₀ is a serine (S) or an asparagine (N),

X₃₄ is a serine (S), an alanine (A), or a threonine (T),

X₄₂ is an arginine (R) or a lysine (K),

5 X₄₅ is a methionine (M), leucine (L), or valine (V),

X₄₆ is a glutamine (Q) or a threonine (T)

X₄₈ is an isoleucine (I), valine (V), or leucine (L),

X₅₃ is independently a glutamic acid, a glutamine, or a lysine, and

X₅₅ is independently a proline, an arginine, a threonine, or no residue.

10 In some examples, the variant comprises one or more of substitutions selected from the group consisting of K6Q, E9R, S14R, Q17K, S34A, E53Q, or E53K.

In some examples, the variant comprises no residue at position 55.

In some examples, the variant comprises

i. K6Q, and Q17K,

15 ii. E9R,

iii. S34A, and E53Q,

iv. S14R, and an addition of 55R, or

v. E53K, and an addition of 55T.

20 In some examples, the variant comprises the sequence selected from the group consisting of

QDKCKQVYENYPINKCKLASQCNYDCKLDKHARGECFYDERRNMQCICDY
CEY (SEQ ID NO: 3),

QEKCKKIYRNYPVSKCQLANQCNYDCKLDKHARTGECFYDEKRNLQCVCDY
CEY (SEQ ID NO: 4),

25 QDKCKKMYENYPVSKCQLANQCNYDCKLDKHARAGECFYDEKRNVCICDY
CQY (SEQ ID NO: 5),

SDKCKKVYENYPIRKCQLANQCNYDCKLDKHARGECFYDEKRNLQCICDYC
EYR (SEQ ID NO: 6), and

30 QDKCKRVYENYPISKQIANQCNYDCKLDKHARGECFYDEKRNLTCLCDYC
KYT (SEQ ID NO: 7).

In another aspect, there is provided a polynucleotide and/or a vector encoding a brazzein variant as described herein.

In yet another aspect, there is provided a cell comprising a vector comprising the polynucleotide as described herein.

In yet another aspect, there is provided a method of producing a brazzein variant, comprising expressing a polynucleotide and/or a vector encoding a brazzein variant as described herein in an expression system.

5 DESCRIPTION OF EMBODIMENTS

Brazzein is a promising alternative for sugars and artificial sweeteners.

Here, the inventors demonstrate the use of protein language models to design new diverse brazzein homologs where a thermostable and potentially sweeter homolog was obtained. The wildtype protein sequence of brazzein is projected into the embedding
10 space, following which a random walk was performed to explore the sequences around the vicinity of the wildtype protein. Representative sequences in the local embedding space of the wildtype brazzein are obtained and expressed.

Therefore, exemplary, non-limiting embodiments of sweeter brazzein variants are disclosed hereinafter.

15 In one aspect, there is provided a brazzein variant comprising
X₁X₂KCKX₆X₇YX₉NYPX₁₃X₁₄KCX₁₇X₁₈AX₂₀QCNYDCKLDKHARX₃₄GECFYDEX₄
2RX₄₅X₄₆CX₄₈CDYCX₅₃YX₅₅ (SEQ ID NO: 1),
wherein
X₁, X₂₀, and X₄₇ are each independently an amino acid with polar uncharged side
20 chain,
X₂ is an amino acid with a negatively charged side chain,
X₆ is a glutamine (E) or an amino acid with positively charged side chains,
X₇, X₁₃, X₁₈, X₄₅, and X₄₈ are each independently an amino acid with hydrophobic
side chain,
25 X₁₇ is a lysine (K) or an amino acid with polar uncharged side chain,
X₁₄ is an arginine (R) or an amino acid with polar uncharged side chain,
X₉ is an arginine (R) or an amino acid with negatively charged side chain,
X₄₂ is an amino acid with positively charged side chain,
X₃₄ is an alanine or an amino acid with polar uncharged side chain,
30 X₄₆ is an amino acid with polar uncharged side chain,
X₅₃ is a glutamic acid, a glutamine, or a lysine, and
X₅₅ is a proline, an arginine, a threonine, or no residue.

The term “variant” as used herein may be used interchangeably with the term
35 “homolog” and is used to refer to a protein that retains the same function and structure

as its wild type but has an amino acid sequence that differs from that of the wild-type protein. In some examples, the variant may have improved function and/or properties. For example, a brazzein variant or brazzein homolog refers to brazzein having an amino acid sequence that differs from that of the wild-type protein. The brazzein variant is
5 capable of interacting with sweet receptors and may have improved binding and/or thermostability. In some examples, the variant is not a naturally occurring variant. In some examples, the variant is an isolated variant. In some examples, the variant is a synthetic variant that does not occur naturally.

In some examples, the variant may comprise naturally occurring amino acids
10 such as arginine (arg, R), histidine (his, H), lysine (lys, K), aspartic acid (asp, D), glutamic acid (glu, E), serine (ser, S), threonine (thr, T), asparagine (asn, N), glutamine (gln, Q), cysteine (cys, C), glycine (gly, G), proline (pro, P), alanine (ala, A), valine (val, V), isoleucine (ile, L), leucine (leu, L), methionine (met, M), phenylalanine (phe, F), tyrosine (tyr, Y), and tryptophan (trp, W).

15 In some examples, amino acid with polar uncharged side chain may include, but is not limited to, serine (ser, S), threonine (thr, T), asparagine (asn, N), and glutamine (gln, Q).

In some examples, amino acid with electrically charged side chain may include negatively charged side chain amino acids and/or positively charged side chain amino
20 acids.

In some examples, amino acid with positively charged side chain amino acids may include arginine (arg, R), histidine (his, H), and lysine (lys, K).

In some examples, amino acids with negatively charged side chain amino acids may include aspartic acid (asp, D), and glutamic acid (Glu, E).

25 In some examples, amino acid may be a special amino acid such as cysteine (cys, C), selenocysteine (sec, U), glycine (gly, G), and proline (pro, P).

In some examples, amino acid may be a hydrophobic side chain amino acid such as, but is not limited to, alanine (ala, A), valine (val, V), isoleucine (ile, I), leucine (leu, L), methionine (met, M), phenylalanine (phe, F), tyrosine (tyr, Y), and tryptophan (trp, W).

30 As known in the art, a wild type brazzein comprises the core sequence QDKCKKVYENYPVSKCQLANQCNYDCKLDKHARSGECFYDEKRNLQCICDYCEY (SEQ ID NO: 2). The amino acid sequence and structure of brazzein is described in detail in Caldwell, J., Abildgaard, F., Džakula, Ž. *et al.* Solution structure of the thermostable sweet-tasting protein brazzein. *Nat Struct Mol Biol* **5**, 427–431 (1998).

<https://doi.org/10.1038/nsb0698-427>, the content of which is thereby incorporated herein.

In some examples, the brazzein variant may comprise one or more mutation at positions 1, 2, 6, 7, 9, 13, 14, 17, 18, 20, 34, 42, 45, 46, 48, 53, and/or 55.

5 As used herein, the term "mutation" refers to alterations in amino acids (or gene sequences encoding for the amino acid sequences). Mutations may include one or more alterations known in the art. In some examples, the mutation may be a substitution, a deletion, or an insertion/addition. In some examples, the mutation is a substitution and/or a deletion.

10 In some examples, the variant may have at least 65%, 70%, 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity with the wild type protein (such as a wild type brazzein). In some examples, the variant may comprise at least 65%, 70%, 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence
15 identity with the sequences as described herein. In some examples, the variant has at least 85% sequence identity to a wild type brazzein.

In some examples, the variant may comprise at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, or 18 amino acids difference from the wild type protein (such as a wild type brazzein) and/or with the sequences as described herein. In some
20 examples, the variant has 4 to 8 amino acids difference from a wild type brazzein. In some examples, the variant has 8 amino acids difference from a wild type brazzein (e.g. V23). In some examples, the variant has 6 amino acids difference from a wild type brazzein (e.g. V22). In some examples, the variant has 4 amino acids difference from a wild type brazzein (e.g. V21 and V24). In some examples, the variant has 5 amino acids
25 difference from a wild type brazzein (e.g. V25).

In some examples, the variant may comprise one mutation, two mutations, three mutations, four mutations, five mutations, six mutations, seven mutations, eight mutations, nine mutations, ten mutations, 11 mutations, 12 mutations, 13 mutations, 14 mutations, 15 mutations, 16 mutations, 17 mutations, 18 mutations, 19 mutations, or
30 more.

In some examples, the variant may comprise a mutation at positions 1, 2, 6, 7, 9, 13, 14, 17, 18, 20, 34, 42, 45, 46, 48, 53, and/or 55.

In some examples, the variant may comprise a mutation that is a substitution at positions 1, 2, 6, 7, 9, 13, 14, 17, 18, 20, 34, 42, 45, 46, 48, 53, and 55.

In some examples, the variant may further comprise a mutation at positions 1, 2, 7, 13, 18, 20, 42, 45, 46, and 48. In some examples, the mutation at positions 1, 2, 6, 7, 9, 13, 14, 17, 18, 20, 34, 42, 45, 46, and 48 is a substitution of the amino acid residue with a natural amino acid of the same group as the amino acid residue of the wild type

5 protein.

As used herein, a variant may be represented by referring to the substituted amino acid positions, which characterise the variant. Substitutions are herein indicated by providing the wild type amino acid residue, followed by the position number, followed by the substituted amino acid residue to be substituted.

10 In some examples, the mutation is a substitution with a natural amino acid known in the art. In some examples, the substitution may be a substitution with a synthetic amino acid known in the art.

In some examples, the variant may further comprise mutations known in the art that supports for increase thermostability.

15 In some examples, the variant may further comprise one or more mutations at the binding site of a sweet receptor.

In some examples, the variant retains substantially the same structure as wild type brazzein.

20 In some examples, the variant has substantially the same (or similar) secondary, tertiary, and/or quaternary structures as wild type brazzein, which allows the variant to have the same (or improved) biological function as the wild type brazzein.

In some examples, the variant has the same secondary structure as wild type brazzein.

25 In some examples, the variant may comprise a beta pleated sheet at positions 5 to 7, 34 to 39, and 45 to 50. In some examples, the variant may comprise an alpha helix structure at positions 12 to 17, and 20 to 31.

In some examples, the variant is capable of binding to a taste receptor.

30 In some examples, the variant is capable of binding (or binds to) to a taste receptor. In some examples, the taste receptor is a taste receptor of a mammal. In some examples, the taste receptor is a taste receptor of a human. In some examples, the taste receptor is a sweet taste receptor. In some examples, the taste receptor is a human sweet taste receptor such as TAS1R2/TAS1R3.

In some examples, the variant is capable of binding to a human sweet taste receptor.

35 In some examples, the variant may comprise

- X₁ is a glutamine (Q) or a serine (S),
 X₂ is an aspartic acid (D) or a glutamic acid (E),
 X₆ is a glutamine (Q), a lysine (K), or an arginine (R),
 X₇ is valine (V), methionine (M), or isoleucine (I)
 5 X₉ is a glutamic acid (E) or an arginine (R),
 X₁₃ is isoleucine (I) or valine (V),
 X₁₄ is an asparagine (N), a serine (S), or an arginine (R),
 X₁₇ is a lysine (K) or a glutamine (Q),
 X₁₈ is a leucine (L) or isoleucine (I),
 10 X₂₀ is a serine (S) or an asparagine (N),
 X₃₄ is a serine (S), an alanine (A), or a threonine (T),
 X₄₂ is an arginine (R) or a lysine (K),
 X₄₅ is a methionine (M), leucine (L), or valine (V),
 X₄₆ is a glutamine (Q) or a threonine (T)
 15 X₄₈ is an isoleucine (I), valine (V), or leucine (L),
 X₅₃ is independently a glutamic acid, a glutamine, or a lysine, and
 X₅₅ is independently a proline, an arginine, a threonine, or no residue.

In some examples, the variant comprises one or more of substitutions selected from the group consisting of K6Q, E9R, S14R, Q17K, S34A, E53Q, E53K.

- 20 In some examples, the variant comprises an addition of 55P, 55R, or 55T.

In some examples, the variant comprises

- i. K6Q, and Q17K,
- ii. E9R,
- iii. S34A, and E53Q,
- 25 iv. S14R, and an addition of 55R, or
- v. E53K, and an addition of 55T.

In some examples, the variant comprises the sequence selected from the group consisting of

- QDKCKQVYENYPINKCKLASQCNYDCKLDKHARGECFYDERRNMQCICDY
 30 CEY (SEQ ID NO: 3),
 QEKCKKIYRNYPVSKCQLANQCNYDCKLDKHARTGECFYDEKRNLCVCDDY
 CEY (SEQ ID NO: 4),
 QDKCKKMYENYPVSKCQLANQCNYDCKLDKHARAGECFYDEKRNVCICDY
 CQY (SEQ ID NO: 5),

SDKCKKVYENYPPIRKQCLANQCNYDCKLDKHARSGECFYDEKRNLCICDYC
EYR (SEQ ID NO: 6), and

QDKCKRVYENYPISKQCIANQCNYDCKLDKHARSGECFYDEKRNLTCLCDYC
KYT (SEQ ID NO: 7).

5 Without wishing to be bound by theory, the inventors of the present disclosure found that amino acid residues at positions 29-33, 36, 39-43, and C-terminus are involved in sweetness.

 In some examples, the brazzein variant may further comprise other mutations known in the art. For example, the brazzein variant as disclosed herein may further
10 comprise other mutations such as, but is not limited to, D29, H31, K5, D2, E41, and the like.

 In another aspect, there is provided a polynucleotide and/or a vector encoding a brazzein variant as described herein.

 In yet another aspect, there is provided a cell comprising a vector comprising the
15 polynucleotide as described herein.

 In some examples, the host cell is a bacterial cell, a yeast cell, and/or a mammalian cell. In some examples, the host cell is a generally recognised as safe bacterium. In some examples, the host cell is a *Lactococcus lactis*.

 To improve production, the inventors also provide a faster and significantly more
20 productive protocol for protein purification of brazzein in generally recognized as safe (GRAS) bacteria such as *Lactococcus lactis* (*L. lactis*).

 In this disclosure, protein folding algorithms are used to design brazzein variants for production in a generally recognized as safe (GRAS) host organism, including a sweeter and thermostable biosimilar brazzein variant. To improve production, the
25 inventors also optimized biochemical production and purification methods for brazzein expression in a GRAS microbe.

 Therefore, in yet another aspect, there is provided a method of producing a brazzein variant, comprising expressing a polynucleotide and/or a vector encoding a brazzein variant as described herein in an expression system.

30 In some examples, the expression system may be one or more comprising bacterial, yeast, and/or fungi expression system. In some examples, the expression system may be a bacterial expression system such as, but not limited to a GRAS (generally regarded as safe) bacterial system. In some examples, the expression system may be a yeast expression system such as, but is not limited to, *Saccharomyces cerevisiae*, *Pichia pastoris*, *Hansulae polymorpha*, *Kluyveromyces lactis*, and the like. In
35

some examples, the expression system may be a fungi expression system (for example filamentous fungi), such as, but is not limited to, *Aspergillus*, *Trichoderma*, and the like.

In some examples, the expression system is a bacterial expression system *Lactococcus lactis*.

5 In some examples, the method further comprises the steps of extracting the expressed brazzein variant. In some examples, the extracting of the expressed brazzein variant may be through sonication or through heat lysis. Advantageously, the inventors of the present disclosure found that heat lysis provides better yield than sonication. Therefore, in lieu of sonication, the sample may only be treated through heat lysis. In
10 some examples, heat lysis may comprise boiling at around 90 to 100 °C, or about 91, about 92, about 93, about 94, about 95, about 96, about 97, about 98, about 99, 100, about 101, about 102, about 103, about 104, about 105, about 106, about 107, about 108, about 109, or about 110 °C. In some examples, heat lysis may be performed for about 5 to 30 mins, or about 5 mins, about 10 mins, about 15 mins, about 10 mins, about
15 20 mins, about 25 mins, or about 30 mins.

In some examples, the method further comprises boiling the sample for 1 to 6 hours, or about 3.5 hours, 4 hours, 4.5 hours, or 5 hours.

In some examples, there is provided a method of sweetening a product comprising providing the brazzein variant as described herein.

20 The term "and/or", e.g., "X and/or Y" is understood to mean either "X and Y" or "X or Y" and should be taken to provide explicit support for both meanings or for either meaning.

Further, in the description herein, the word "substantially" whenever used is understood to include, but not restricted to, "entirely" or "completely" and the like. In
25 addition, terms such as "comprising", "comprise", and the like whenever used, are intended to be non-restricting descriptive language in that they broadly include elements/components recited after such terms, in addition to other components not explicitly recited. For example, when "comprising" is used, reference to a "one" feature is also intended to be a reference to "at least one" of that feature. Terms such as
30 "consisting", "consist", and the like, may in the appropriate context, be considered as a subset of terms such as "comprising", "comprise", and the like. Therefore, in embodiments disclosed herein using the terms such as "comprising", "comprise", and the like, it will be appreciated that these embodiments provide teaching for corresponding embodiments using terms such as "consisting", "consist", and the like. Further, terms
35 such as "about", "approximately" and the like whenever used, typically means a

reasonable variation, for example a variation of +/- 5% of the disclosed value, or a variance of 4% of the disclosed value, or a variance of 3% of the disclosed value, a variance of 2% of the disclosed value or a variance of 1% of the disclosed value.

Furthermore, in the description herein, certain values may be disclosed in a range. The values showing the end points of a range are intended to illustrate a preferred range. Whenever a range has been described, it is intended that the range covers and teaches all possible sub-ranges as well as individual numerical values within that range. That is, the end points of a range should not be interpreted as inflexible limitations. For example, a description of a range of 1% to 5% is intended to have specifically disclosed sub-ranges 1% to 2%, 1% to 3%, 1% to 4%, 2% to 3% etc., as well as individually, values within that range such as 1%, 2%, 3%, 4% and 5%. It is to be appreciated that the individual numerical values within the range also include integers, fractions and decimals. Furthermore, whenever a range has been described, it is also intended that the range covers and teaches values of up to 2 additional decimal places or significant figures (where appropriate) from the shown numerical end points. For example, a description of a range of 1% to 5% is intended to have specifically disclosed the ranges 1.00% to 5.00% and also 1.0% to 5.0% and all their intermediate values (such as 1.01%, 1.02% ... 4.98%, 4.99%, 5.00% and 1.1%, 1.2% ... 4.8%, 4.9%, 5.0% etc.) spanning the ranges. The intention of the above specific disclosure is applicable to any depth/breadth of a range.

Additionally, when describing some embodiments, the disclosure may have disclosed a method and/or process as a particular sequence of steps. However, unless otherwise required, it will be appreciated that the method or process should not be limited to the particular sequence of steps disclosed. Other sequences of steps may be possible. The particular order of the steps disclosed herein should not be construed as undue limitations. Unless otherwise required, a method and/or process disclosed herein should not be limited to the steps being carried out in the order written. The sequence of steps may be varied and still remain within the scope of the disclosure.

Furthermore, it will be appreciated that while the present disclosure provides embodiments having one or more of the features/characteristics discussed herein, one or more of these features/characteristics may also be disclaimed in other alternative embodiments and the present disclosure provides support for such disclaimers and these associated alternative embodiments.

BRIEF DESCRIPTION OF FIGURES

Example embodiments of the disclosure will be better understood and readily apparent to one of ordinary skill in the art from the following discussions and if applicable, in conjunction with the figures. It should be appreciated that other modifications to the variant may be made without deviating from the scope of the invention. Example embodiments are not necessarily mutually exclusive as some may be combined with one or more embodiments to form new exemplary embodiments. The example embodiments should not be construed as limiting the scope of the disclosure.

Fig. 1 shows the overall strategy of designing proteins by exploring the latent space of Large Language Models trained on protein sequences. The process of generating a library of biosimilar proteins begins with the selection of an initial native protein with the desired sweetness profile.

Fig. 2 shows the sequences and data on the sequences of the AI-derived mutants. (A) shows the sequence alignment of mutants/variants (V21-25) against wild type. Identical residues are highlighted in grey. Similar residues are coloured grey and residues that are not similar are circled. Structural elements of wild type brazzein are also shown (PDB: 4HE7). The numbering starts from Q after the N-terminal M. (B) shows the comparison of sweet taste receptor responses using calcium mobilization responses of various mutants and wild-type brazzein expressed in *E. coli* using the luminescence-based readout sweet taste receptor assay. Percent calcium mobilization is calculated against the maximum assay response from 15 mM of sucralose. Data is interpolated and averaged from non-linear fits of experimentally derived data of at least two independent assay runs at a standardized protein concentration of 0.1 mg/mL. Error bars are S.D.

Fig. 3 shows the expression and processing of brazzein from generally regarded as safe (GRAS) *L. lactis*. (A) shows purification workflow (B) shows a representative protein gel of purified wild type brazzein with L: Novex pre-stained ladder, lane 1: sonicated (Fig. 3A – method 1), lane 2: heat lysis (95 °C for 10 mins, Fig 3A – method 2), lane 3: heat lysis, purified and boiled 4 hours at 95 °C (Fig. 3A – method 3) and its equivalent for V23 (lanes 4-6). Invitrogen Novex 16% Tricine gel was used with Tricine SDS Running buffer. The gel was stained with Coomassie blue and imaged. (C) shows total determined protein concentration of His-tagged purified brazzein with two lysis methods (Fig 3A – methods 1 and 2). Data are from 3 replicates over 3 independent runs. Error bars are S.D. Paired t-test of protein yields between two lysis methods reveal statistical significance ($p < 0.05$).

Fig. 4 shows sweet potencies using calcium mobilization responses of wild type brazzein and the V23 mutant, expressed in *L. lactis* and subject to two different lysis methods, and a high heat treatment at 95 °C for 4 hours. Samples were tested using the fluorescence-based sweet taste receptor assay. Thaumatin is shown for scale of sweet taste receptor assay response at various concentrations. Data is interpolated and averaged from non-linear fits of experimentally derived data at a single protein concentration point of 68 µg/mL, from at least six experimental replicates. Error bars are S.D. One asterisk (*) indicates $p = 0.0138$ (one-way ANOVA). Four asterisks (****) indicate $p < 0.0001$.

EXPERIMENTAL DATA

Materials & Methods

Plasmid construction

Escherichia coli (*E. coli*) codon-optimized DNA sequences for the His-tagged brazzein constructs were synthesized in a pET24a(+) vector from Twist Biosciences (Singapore). The constructs were transformed into *E. coli* Omnimax for sequencing and into *E. coli* BL21(DE3) for protein expression. *L. lactis* codon-optimized DNA sequences for the His-tagged brazzein constructs were synthesized from Twist Biosciences. The fragments were cloned into pNZ8148 vector via Gibson assembly. The constructs were transformed into *L. lactis* NZ9000 for sequencing and protein expression. The amino acid sequences of the final constructs can be found in Table S1.

Table 1. EC₅₀ values of wild type brazzein and various mutants, expressed in mg/mL and µM, subjected to three different lysis method. EC₅₀ values, expressed in µM, of wild type brazzein and AI-designed V23 mutants, subjected to three different lysis methods. The samples were tested using the fluorescence based assay.

EC ₅₀ (µM)	WT (heat lyse)	V23 (heat lysed)	WT (heat lysed + boil)	V23 (heat lysed + boil)	WT (sonicate)	V23 (sonicated)
curve 1	15.1	10.3	25.6	3.2	1.9	1.3
curve 2	20.5	11.7	22.9	4.9	5.2	8.8
curve 3	16	12.9	22.1	3.5	1.6	5.6

curve 4	25.6	8.3	14.7	5.3	11.9	2.1
curve 5	17.4	8.3	11	3.1	2.4	1.7
curve 6	23.3	12.5	12.5	5.2	10.1	1.1
Mean and S.D.	19.6 ± 4.2	10.6 ± 2.1	18.1 ± 6.2	4.2 ± 1.1	5.5 ± 4.5	3.4 ± 3.1

The EC₅₀ values were generated using the four-parameter logarithmic regression equation in Prism 8 (GraphPad) software with the following constraints applied: bottom asymptote > 0.25; 0 < top asymptote < 2.

5 *E. coli* BL21(DE3) protein expression and purification

- Single colonies from transformed *E. coli* BL21(DE3) were inoculated in LB Broth containing 50 µg/mL kanamycin for overnight culture at 37 °C, with shaking (200 rpm). The overnight cultures were transferred into 300 mL Terrific Broth containing 50 µg/mL kanamycin at 37 °C, 200 rpm. When the optical density at 600nm (OD₆₀₀) reached 0.4 – 0.6, protein expression was induced with 1 mM of isopropyl β-D-1-thiogalactopyranoside (IPTG), and incubated overnight at 30 °C, 200 rpm. The cultures were subsequently harvested by centrifugation at 8000 g at 4 °C (10 min), and the pellets were freeze-thawed, resuspended in BugBuster Protein Extraction Reagent (Merck, Cat. No. 70584) and then incubated at room temperature (15 min).
- 15 The resulting lysate was then centrifuged at 18000 g for 20 min at 4 °C. The supernatant was incubated with PureCube 100 INDIGO Ni-Agarose resin (Cube-biotech, Cat. No. 75110) for 1 h at room temperature, and the protein-bound resin was washed with 50 mM sodium phosphate buffer pH 7.4, 500 mM sodium chloride and 20 mM imidazole. The bound protein was eluted with 50 mM sodium phosphate
- 20 buffer pH 7.4, 500 mM sodium chloride and 500 mM imidazole. The eluate was buffer-exchanged and concentrated with Hank's Balanced Salt Solution (HBSS) containing 20 mM HEPES at pH 7.0.

L. lactis NZ9000 protein expression and purification

- 25 Single colonies from the transformed *L. lactis* NZ9000 were inoculated in M17 Broth (0.5% glucose, 10 µg/mL chloramphenicol) and incubated overnight at 30 °C without shaking. The overnight cultures were inoculated in 2 L 2x M17 Broth (2% glucose, 10 µg/mL chloramphenicol) to OD₆₀₀ 0.1, and incubated at 30 °C. When OD₆₀₀ reached 1.0, protein expression was induced with 50 ng/mL nisin at 30 °C for 3 h.

The cultures were centrifuged at 8000 *g* for 10 min at 4 °C. The resulting pellets were freeze-thawed, resuspended in 50mM sodium phosphate buffer pH 7.4, 300 mM sodium chloride, 10 mM imidazole and 0.03% Triton X-100, and then incubated at room temperature for 15 min. The resuspended pellets were either sonicated 4 times for 10 s at 10 s intervals on ice or boiled at 95 °C for 10 min. The resulting lysate was then centrifuged at 18000 *g* for 20 min at 4 °C. The supernatant was incubated with PureCube 100 INDIGO Ni-Agarose resin for 1 hat room temperature. The protein-bound resin was washed with 50 mM sodium phosphate buffer pH 7.4, 500 mM sodium chloride and 20 mM imidazole. The bound protein was eluted with 50 mM sodium phosphate buffer pH 7.4, 500 mM sodium chloride and 500 mM imidazole. The eluate was buffer-exchanged and concentrated with HBSS containing 20 mM HEPES pH 7.0. For thermostability testing (Fig 2A-3), the samples in HBSS-HEPES buffer were boiled at 95 °C for 4 h then rapidly cooled to 4 °C.

15

Sweet taste receptor luminescence assay

HEK 293T (ATCC) cells were maintained at 37 °C in a humidified atmosphere of 5% CO₂ and cultured in high-glucose Dulbecco's modified Eagle's medium (DMEM; Gibco) supplemented with 10% (v/v) heat-inactivated Fetal Bovine Serum (FBS; Biowest) and 1% (v/v) penicillin-streptomycin (Gibco). White 384-well tissue culture plates (Greiner) were coated with Poly-D-Lysine (PDL; Sigma) at a final concentration of 1 mg/mL. These coated plates were then seeded with 293T cells at a density of 20,000 cells per well and incubated overnight.

The cells were transiently transfected with two plasmids - a multigene CMV-promoter based expression vector containing the genes for the sweet taste receptor (TAS1R2/TAS1R3) and the chimeric Gα16-gust44gene. The second plasmid expression vector contains the gene for the apophotoprotein, mitochondrial-targeted (mt)-clytin II. These two plasmids were transfected at a ratio of 20 ng:20 ng per well using ViaFect (Promega), employing a transfection agent to plasmid ratio of 3:1 μL:μg. A complete media change was performed 6 hours post-transfection with low-glucose DMEM (Gibco) supplemented with 10% (v/v) heat-inactivated FBS (Biowest) and 1% (v/v) penicillin-streptomycin (Gibco).

After an overnight incubation at standard cell culture conditions, the transfected cells were loaded with Coelenterazine F (AAT Bioquest) to a final concentration of 10 μ M, in low-glucose media and assay buffer (1 $\times$ HBSS assay buffer with 20 mM HEPES at pH 7.0). The assay plate was then incubated at 27 $^{\circ}$ C in the dark for 4 hours. The

5 assay was performed using the luminescence mode of the Fluorescent Imaging Plate Reader (FLIPRETETRA, Molecular Devices) controlled by the ScreenWorks software (version 4.0.0.30, Molecular Devices). During the run, a baseline read was captured for 10 seconds before 25 μ L of test ligand prepared to a two times

10 plate. The kinetic data was acquired for a further 100 seconds to record the responses of each well to added test sample. The well responses were exported as area under the curve (AUC) values and the data were plotted using the four-parameter logarithmic regression equation using Prism 8 (GraphPad) software. Unless otherwise stated, data reported were derived from at least two independent

15 experiments, performed in duplicates. For this study, two reference sweeteners, sucralose, and the sweet protein thaumatin were use for comparison to our brazzein test samples.

Sweet taste receptor fluorescence assay

20 The Adenovirus 293AD (Cell Biolabs, Inc) cells were maintained under similar cell culture conditions as 293T cells. Cells were seeded to a density of 12,000 cells per well in black 384-well tissue culture plates (Greiner) and grown overnight.

25 A multigene CMV-promoter based expression vector containing the genes for the sweet taste receptor (TAS1R2/TAS1R3) and the chimeric G α 16-gust44gene was transiently transfection into 293AD cells at 25 ng per well using Viafect reagent. After 6 h post-transfection, the growth media was removed and replaced with low-glucose DMEM (Gibco) supplemented with 10% (v/v) heat-inactivated FBS (Biowest) and 1%

30 (v/v) penicillin-streptomycin (Gibco). The following day, the transfected cells were loaded with Calcium 6 (Molecular Devices) fluorescent dye. The assay plate was first incubated at 37 $^{\circ}$ C in a humidified incubator with 5% CO₂ for 2 h, followed by a further 30 min, on the lab bench for equilibration at *room temperature*.

- The assay was performed using the fluorescence mode of the FLIPR-TETRA. The fluorescence intensity is directly correlated to the amount of intracellular calcium that is released into the cytoplasm in response to ligand-mediated activation of the sweet taste receptor, which in turn is regarded as a measure of receptor activation.
- 5 Changes in calcium membrane potential were measured over time with an excitation at 470–495 nm and measurement of emission at 515–575 nm. A baseline measurement read was taken every second for 10 s prior to addition of sweetener or test sample, where further measurement reads were acquired for 310 s.
- 10 Emission fluorescence values were converted to response (max) over baseline (min) values using the ScreenWorks software (version 4.0.0.30, Molecular Devices), and the data was plotted using the four-parameter logarithmic regression equation using Prism 8 (GraphPad) software. The sweet taste response from the compound can be evaluated for its potency towards the sweet taste receptor, expressed as EC₅₀, which
- 15 is the concentration of molecule required to a give half-maximal response in the sweet taste receptor assay.

Protein Library Sequence Characterization

- The protein library contains 5 orthogonal sequences that capture different variations
- 20 using Brazzein as the natural wildtype reference (Table 2).

Table 2. Sequences of the representative sequences captured within the sweet protein library

>Brather malv21	QDKCKKMYENYPVSKCQLANQCNYDCKLDKHAR AGECFYDEKRNVCICDYCQY	SEQ ID NO: 3
>Brather malv22	QEKCKKIYRNYPVSKCQLANQCNYDCKLDKHART GECFYDEKRNLQCVCDYCEY	SEQ ID NO: 4
>Brather malv23	QDKCKQVYENYPINKCKLASQCNYDCKLDKHARS GECFYDERRNMQCICDYCEY	SEQ ID NO: 5
>Brather malv24	SDKCKKVYENYPISKCQLANQCNYDCKLDKHARS GECFYDEKRNLQCICDYCEYR	SEQ ID NO: 6
>Brather malv25	QDKCKRVYENYPISKQIANQCNYDCKLDKHARS GECFYDEKRNLCLCDYCKYT	SEQ ID NO: 7

Table 3. EC₅₀* of *L. lactis* purified variants based on a sweet taste receptor (TAS1R2/R3) fluorescent assay.

ID	EC ₅₀
WT Brazzein	52.2 μM
V23	39.1 μM
V23 H31R/E41A	24.2 μM
V23	31.9 μM
V23 H31R/E41A	16.8 μM

The variants were purified via a boil lysis of the cells, however, these have not undergone further heat treatment. (*EC₅₀ is the concentration of molecule required to give half-maximal response in the sweet taste receptor.)

H31R/E41A mutations are derived from Lee, J.-W., Cha, J.-E., Jo, H.-J., & Kong, K.-H. (2013). Multiple mutations of the critical amino acid residues for the sweetness of the sweet-tasting protein, brazzein. *Food Chemistry*, 138(2–3), 1370–1373. <https://doi.org/10.1016/j.foodchem.2012.10.140>

Results

Generation of computationally fold-able sweet protein variants

Using pretrained language models (SeqVec, UniRep, CPCprot), the protein sequence of Brazzein is embedded into the latent space of these models. It has previously been shown that the latent space close to the protein of interest contains sequences that preserve both the structural and functional properties of the original protein. Using this property, the inventors of the present disclosure explore the latent space surrounding brazzein with a simple adaptive walk that introduces mutations in the latent space that simultaneously optimizes for many of the desired properties: 1) divergence in sequence, 2) thermostability and 3) solubility. The directed evolution of the sequence along different walks in the latent space allows the inventors to

construct a library of high confidence sequences as leads for downstream characterization (Fig 1).

Scaling up throughput of sweetness measurement and comparison

- 5 In this study, the inventors employed the use of cell-based sweet taste receptor assays to rapidly and systematically assess the relative sweetness of brazzein variants. Similar receptor-based assays have been routinely used in studies of sweet taste reception and sweetener molecules optimization (Riedel et al., 2017). Human sweet taste receptors, along with their signalling components are heterologously
- 10 expressed in cultured mammalian HEK cells and shown to respond to a wide array of sweeteners. This approach measures calcium mobilization in response to sweet taste receptor activation by sweeteners such as sweet proteins like brazzein, carbohydrate sweeteners like sucrose, and both natural and synthetic sweetener molecules such as sucralose and stevioside. This allows rapid and increased
- 15 throughput of screening which otherwise would not be possible with a human sensory panel.

- Five brazzein variants were generated computationally, with a range of 5-8 mutations, including deletions (Fig 2A). As an initial assay, the variants were
- 20 expressed in *E. coli* and purified via affinity tag pull-down before screening for sweet taste receptor response and compared against wild type brazzein as a control. As no optimization for purification of brazzein was accomplished in this initial set of screening experiments, the samples expressed in *E. coli* had a high percentage of impurity (Fig 5), which contributes to non-specific signals in the assay readout.
- 25 Consequently, the inventors cannot wholly attribute specific sweet taste responses to make accurate comparisons between the *E. coli* expressed samples. This initial screening mutant dataset suggested a trend where the V23 variant could potentially be sweeter than WT (Fig 2B). To further investigate this, the inventors focused on producing high purity samples of the V23 variant alongside WT brazzein.

30

Bioprocessing optimization in a *L. lactis*

- To optimize expression and purity of brazzein, the inventors turned to GRAS *L. lactis* NZ9000. Although brazzein expression from *L. lactis* was sufficient to increase
- 35 purity, its yields are significantly low; <0.1mg/L. Subsequently, the inventors also

exploited brazzein's thermostability to establish a heat-based purification protocol. While both heat lysis and sonication effectively rupture bacteria cells to release intracellular proteins, heat lysis is useful when working with heat-stable proteins. In addition, heating may improve purity by denaturing all non-heat-stable proteins.

5 Increases in yield could also potentially occur via heat denaturation of proteases which would otherwise degrade the protein of interest. Others have previously used a two-hour heat treatment at 80 °C as a second purification step after ammonium precipitation to successfully increase the purity of brazzein expressed in transgenic tobacco leaves. Hence, the inventors hypothesize that heating can be used in lieu

10 of mechanical lysis (sonication) to lyse and purify the expressed brazzein. Heat-based lysis of cell pellets was performed by boiling cell pellets for 95 °C, 10 minutes (Fig 3A). In our observations, purity of the samples increased with heat lysis protocol (Fig 3B). More importantly, there was also a significant 10-fold increase in brazzein yield between the two protocols (Fig 3C). Although this study did not explore further,

15 the inventors expect that yields can be further improved through optimization of fermentation conditions compared with heat lysis-mediated purification.

Characterization of AI-derived V23 variant

Using products from the bioprocessing outputs, the inventors examined the

20 sweetness response of sonicated and heat lysed products of WT and V23 variants. Thermostability was also further tested by boiling the brazzein samples for a further 4 hours at 95 °C (Fig 3A – method 3). In comparison with WT's sonicated, heat lysed and boiled products, V23 equivalents were more sensitive in the sweet taste receptor assay (Fig 4, Table1). Overall, heat treated products are less sensitive in the taste

25 receptor assay compared to sonicated products, which is not surprising since it is expected that protein folding can be disrupted in the presence of heat. However, V23 in all three observations are more sensitive compared to WT, indicating that V23 is potentially sweeter than wild type brazzein and maintains this potency with heat treatment. Interestingly, the samples that were boiled for 4 hours had slightly high

30 calcium response compared to 10 min lysis, suggesting that perhaps further heating removed residual heat labile proteins including slightly misfolded brazzein proteins. Even though there seems to be a drop in sweet potency of products from 10 mins heat lysis and sonication protocols, prior examples of heat treatment protocols at mild conditions, for example, 80 °C, 2 hours, imply that further refinement of the heat

treatment could be utilized to minimize impact on sweetness potency while preserving yields.

The present inventors demonstrated here the application of zero-shot computation design to create a thermostable and potentially sweeter brazzein homolog, V23. To characterize this, the inventors established a workflow which includes quality control assays that quantifies sweetness of proteins, allowing for the accurate characterization and comparison of brazzein mutants, and a more productive brazzein purification protocol from GRAS *L. lactis*.

AI-powered protein design allows the inventors to quickly evaluate a large number of protein sequences which in turn allows the inventors to identify highly optimized candidates that would be hard to uncover with conventional methods. Without *a prior* input of brazzein -specific data, the inventors were able to design high order mutants (5-8 mutations), where a screen of 5 mutants uncovered one better than wild-type brazzein. Based on the wild-type brazzein, multiple studies have previously investigated mutants to identify critical regions that are important for sweetness. Results of these experiments suggest that residues 29-33 and 39-43, plus residue 36 between these stretches, as well as the C-terminus are involved in the sweetness and that charge plays an important role in its interaction with the sweet taste receptor. Interestingly, the mutations generated computationally in this study were regions distant from previously observed regions of interest, suggesting that there is more room to further increase the sweetness of brazzein beyond the traditionally probed regions and vice versa, the computationally design variant can be further optimized with combination of known mutations.

References

1. Alley, E. C., Khimulya, G., Biswas, S., AlQuraishi, M., & Church, G. M. (2019). Unified rational protein engineering with sequence-based deep representation learning. *Nature Methods*, 16(12), 1315–1322. <https://doi.org/10.1038/s41592-019-0598-1>
2. Assadi-Porter, F. M., Aceti, D. J., Cheng, H., & Markley, J. L. (2000). Efficient production of recombinant brazzein, a small, heat-stable, sweet-tasting protein of plant origin. *Archives of Biochemistry and Biophysics*, 376(2), 252–258. <https://doi.org/10.1006/abbi.2000.1725>

3. Assadi-Porter, F. M., Aceti, D. J., & Markley, J. L. (2000). Sweetness determinant sites of brazzein, a small, heat-stable, sweet-tasting protein. *Archives of Biochemistry and Biophysics*, 376(2), 259–265. <https://doi.org/10.1006/abbi.2000.1726>
- 5 4. Bueno-Hernández, N., Vázquez-Frías, R., Abreu Y Abreu, A. T., Almeda-Valdés, P., Barajas-Nava, L. A., Carmona-Sánchez, R. I., Chávez-Sáenz, J., Consuelo-Sánchez, A., Espinosa-Flores, A. J., Hernández-Rosiles, V., Hernández-Vez, G., Icaza-Chávez, M. E., Noble-Lugo, A., Romo-Romo, A., Ruiz-Margaín, A., Valdovinos-Díaz, M. A., & Zárate-Mondragón, F. E. (2019). Review of the scientific evidence and technical opinion on noncaloric sweetener consumption in gastrointestinal diseases. *Revista de Gastroenterología de Mexico (English)*, 84(4), 492–510. <https://doi.org/10.1016/j.rgmx.2019.08.001>
- 10 5. Burgess, C., O'connell-Motherway, M., Sybesma, W., Hugenholtz, J., & van Sinderen, D. (2004). Riboflavin production in *Lactococcus lactis*: potential for in situ production of vitamin-enriched foods. *Applied and Environmental Microbiology*, 70(10), 5769–5777. <https://doi.org/10.1128/AEM.70.10.5769-5777.2004>
- 15 6. Berlec, A., & Strukelj, B. (2009). Large increase in brazzein expression achieved by changing the plasmid /strain combination of the NICE system in *Lactococcus lactis*. *Letters in Applied Microbiology*, 48(6), 750–755. <https://doi.org/10.1111/j.1472-765X.2009.02608.x>
- 20 7. Berlec, A., Tompa, G., Slapar, N., Fonović, U. P., Rogelj, I., & Štrukelj, B. (2008). Optimization of fermentation conditions for the expression of sweet-tasting protein brazzein in *Lactococcus lactis*. *Letters in Applied Microbiology*, 46(2), 227–231. <https://doi.org/10.1111/j.1472-765X.2007.02297.x>
- 25 8. Caldwell, J. E., Abildgaard, F., Dzakula, Z., Ming, D., Hellekant, G., & Markley, J. L. (1998). Solution structure of the thermostable sweet-tasting protein brazzein. *Nature Structural Biology*, 5(6), 427–431. <https://doi.org/10.1038/nsb0698-427>
- 30 9. Choi, H.-E., Lee, J.-I., Jo, S.-Y., Chae, Y.-C., Lee, J.-H., Sun, H.-J., Ko, K., Hong, S., & Kong, K.-H. (2022). Functional expression of the sweet-tasting protein brazzein in transgenic tobacco. *Food Science and Technology*, 42. <https://doi.org/10.1590/fst.40521>

10. Debras, C., Chazelas, E., Sellem, L., Porcher, R., Druet-Pecollo, N., Esseddik, Y., de Edelenyi, F. S., Agaësse, C., de Sa, A., Lutchia, R., Fezeu, L. K., Julia, C., Kesse-Guyot, E., Allès, B., Galan, P., Hercberg, S., Deschasaux-Tanguy, M., Huybrechts, I., Srour, B., & Touvier, M. (2022). Artificial sweeteners and risk of cardiovascular diseases: results from the prospective NutriNet-Santé cohort. *BMJ*, e071204. <https://doi.org/10.1136/bmj-2022-071204>
11. Debras, C., Chazelas, E., Srour, B., Druet-Pecollo, N., Esseddik, Y., Szabo de Edelenyi, F., Agaësse, C., de Sa, A., Lutchia, R., Gigandet, S., Huybrechts, I., Julia, C., Kesse-Guyot, E., Allès, B., Andreeva, V. A., Galan, P., Hercberg, S., Deschasaux-Tanguy, M., & Touvier, M. (2022). Artificial sweeteners and cancer risk: Results from the NutriNet-Santé population-based cohort study. *PLoS Medicine*, 19(3), e1003950. <https://doi.org/10.1371/journal.pmed.1003950>
12. Gardner, C., Wylie-Rosett, J., Gidding, S. S., Steffen, L. M., Johnson, R. K., Reader, D., & Lichtenstein, A. H. (2012). Nonnutritive Sweeteners: Current Use and Health Perspectives. *Circulation*, 126(4), 509–519. <https://doi.org/10.1161/CIR.0b013e31825c42ee>
13. Ghanavati, P., Khalifeh, K., & Jafarian, V. (2016). Structural features and activity of Brazzein and its mutants upon substitution of a surfaced exposed alanine. *Biochimie*, 131, 20–28. <https://doi.org/10.1016/j.biochi.2016.09.006>
14. Gnanavel, M., & Serva Peddha, M. (2011). Identification of novel sweet protein for nutritional applications. *Bioinformation*, 7(3), 112–114. <https://doi.org/10.6026/97320630007112>
15. Gu, W., Xia, Q., Yao, J., Fu, S., Guo, J., & Hu, X. (2015). Recombinant expressions of sweet plant protein mabinlin II in Escherichia coli and food-grade Lactococcus lactis. *World Journal of Microbiology & Biotechnology*, 31(4), 557–567. <https://doi.org/10.1007/s11274-015-1809-2>
16. Heinzinger, M., Elnaggar, A., Wang, Y., Dallago, C., Nechaev, D., Matthes, F., & Rost, B. (2019). Modeling aspects of the language of life through transfer-learning protein sequences. *BMC Bioinformatics*, 20(1), 723. <https://doi.org/10.1186/s12859-019-3220-8>
17. Hugenholtz, J., Sybesma, W., Groot, M. N., Wisselink, W., Ladero, V., Burgess, K., van Sinderen, D., Piard, J.-C., Eggink, G., Smid, E. J., Savoy, G., Sesma, F., Jansen, T., Hols, P., & Kleerebezem, M. (2002). Metabolic

- engineering of lactic acid bacteria for the production of nutraceuticals. *Antonie van Leeuwenhoek*, 82(1–4), 217–235.
18. Jin, Z., Danilova, V., Assadi-Porter, F. M., Aceti, D. J., Markley, J. L., & Hellekant, G. (2003). Critical regions for the sweetness of brazzein. *FEBS Letters*, 544(1–3), 33–37. [https://doi.org/10.1016/s0014-5793\(03\)00383-1](https://doi.org/10.1016/s0014-5793(03)00383-1)
19. Jin, Z., Danilova, V., Assadi-Porter, F. M., Markley, J. L., & Hellekant, G. (2003). Monkey electrophysiological and human psychophysical responses to mutants of the sweet protein brazzein: delineating brazzein sweetness. *Chemical Senses*, 28(6), 491–498. <https://doi.org/10.1093/chemse/28.6.491>
20. Joseph, J. A., Akkermans, S., Nimmegeers, P., & van Impe, J. F. M. (2019). Bioproduction of the Recombinant Sweet Protein Thaumatin: Current State of the Art and Perspectives. *Frontiers in Microbiology*, 10, 695. <https://doi.org/10.3389/fmicb.2019.00695>
21. Kalthoff, C. (2003). A novel strategy for the purification of recombinantly expressed unstructured protein domains. *Journal of Chromatography B*, 786(1–2), 247–254. [https://doi.org/10.1016/S1570-0232\(02\)00908-X](https://doi.org/10.1016/S1570-0232(02)00908-X)
22. Kim, T.-Y., Woo, E.-J., & Yoon, T.-S. (2022). Binding mode of brazzein to the taste receptor based on crystal structure and docking simulation. *Biochemical and Biophysical Research Communications*, 592, 119–124. <https://doi.org/10.1016/j.bbrc.2022.01.004>
23. Lee, J.-W., Cha, J.-E., Jo, H.-J., & Kong, K.-H. (2013). Multiple mutations of the critical amino acid residues for the sweetness of the sweet-tasting protein, brazzein. *Food Chemistry*, 138(2–3), 1370–1373. <https://doi.org/10.1016/j.foodchem.2012.10.140>
24. Lim, J.-K., Jang, J.-C., Kong, J.-N., Kim, M.-C., & Kong, K.-H. (2016). Importance of Glu53 in the C-terminal region of brazzein, a sweet-tasting protein. *Journal of the Science of Food and Agriculture*, 96(9), 3202–3206. <https://doi.org/10.1002/jsfa.7501>
25. Linares, D. M., Kok, J., & Poolman, B. (2010). Genome sequences of *Lactococcus lactis* MG1363 (revised) and NZ9000 and comparative physiological studies. *Journal of Bacteriology*, 192(21), 5806–5812. <https://doi.org/10.1128/JB.00533-10>
26. Liu, B., Jiang, H., Wang, H., & Yang, L. (2021). Removal of the N-terminal methionine improves the sweetness of the recombinant expressed sweet-

- tasting protein brazzein and its mutants in *Escherichia coli*. *Journal of Food Biochemistry*, 45(3), e13354. <https://doi.org/10.1111/jfbc.13354>
27. Lu, A. X., Zhang, H., Ghassemi, M., & Moses, A. (2020). Self-Supervised Contrastive Learning of Protein Representations By Mutual Information Maximization. *BioRxiv*, 2020.09.04.283929. <https://doi.org/10.1101/2020.09.04.283929>
28. Lu, R., Li, X., Hu, J., Zhang, Y., Wang, Y., & Jin, L. (2022). Expression of a triple mutational des-pGlu brazzein in transgenic mouse milk. *FEBS Open Bio*, 12(7), 1336–1343. <https://doi.org/10.1002/2211-5463.13411>
29. Ming, D., & Hellekant, G. (1994). Brazzein, a new high-potency thermostable sweet protein from *Pentadiplandra brazzeana* B. *FEBS Letters*, 355(1), 106–108. [https://doi.org/10.1016/0014-5793\(94\)01184-2](https://doi.org/10.1016/0014-5793(94)01184-2)
30. Rega, M. F., di Monaco, R., Leone, S., Donnarumma, F., Spadaccini, R., Cavella, S., & Picone, D. (2015). Design of sweet protein based sweeteners: hints from structure-function relationships. *Food Chemistry*, 173, 1179–1186. <https://doi.org/10.1016/j.foodchem.2014.10.151>
31. Riedel, K., Sombroek, D., Fiedler, B., Siems, K., & Krohn, M. (2017). Human cell-based taste perception - a bittersweet job for industry. *Natural Product Reports*, 34(5), 484–495. <https://doi.org/10.1039/c6np00123h>
32. Suez, J., Korem, T., Zeevi, D., Zilberman-Schapira, G., Thaiss, C. A., Maza, O., Israeli, D., Zmora, N., Gilad, S., Weinberger, A., Kuperman, Y., Harmelin, A., Kolodkin-Gal, I., Shapiro, H., Halpern, Z., Segal, E., & Elinav, E. (2014). Artificial sweeteners induce glucose intolerance by altering the gut microbiota. *Nature*, 514(7521), 181–186. <https://doi.org/10.1038/nature13793>

APPLICATIONS

Embodiments of brazzein homologs or variants disclosed herein provide an alternative sweetener.

Advantageously, the brazzein homologs or variants as described herein has superior high thermostability where boiling of purified brazzein samples did not significantly reduce or ameliorate their ability to activate sweet taste receptors.

More advantageously, the brazzein homologs or variants as described herein have the desired properties of 1) divergence in sequence, 2) thermostability and/or 3) solubility. In some examples, the brazzein variants or homologs are found to be sweeter than wild type brazzein.

5 Even more advantageously, the brazzein homologs or variants as described herein can be produced at low costs.

The present disclosure also provides improve production and purification methods for brazzein expression.

10 It will be appreciated by a person skilled in the art that other variations and/or modifications may be made to the embodiments disclosed herein without departing from the spirit or scope of the disclosure as broadly described. For example, in the description herein, features of different exemplary embodiments may be mixed, combined, interchanged, incorporated, adopted, modified, included etc. or the like across different exemplary embodiments. The present embodiments are, therefore,
15 to be considered in all respects to be illustrative and not restrictive.

CLAIMS

1. A brazzein variant comprising
X₁X₂KCKX₆X₇YX₉NYPX₁₃X₁₄KCX₁₇X₁₈AX₂₀QCNYDCKLDKHARX₃₄GECFYDEX₄
5 2RNX₄₅X₄₆CX₄₈CDYCX₅₃YX₅₅ (SEQ ID NO: 1),
wherein
X₁, X₂₀, and X₄₇ are each independently an amino acid with polar uncharged side chain,
X₂ is an amino acid with a negatively charged side chain,
10 X₆ is a glutamine (E) or an amino acid with positively charged side chains,
X₇, X₁₃, X₁₈, X₄₅, and X₄₈ are each independently an amino acid with hydrophobic side chain,
X₁₇ is a lysine (K) or an amino acid with polar uncharged side chain,
X₁₄ is an arginine (R) or an amino acid with polar uncharged side chain,
15 X₉ is an arginine (R) or an amino acid with negatively charged side chain,
X₄₂ is an amino acid with positively charged side chain,
X₃₄ is an alanine or an amino acid with polar uncharged side chain,
X₄₆ is an amino acid with polar uncharged side chain,
X₅₃ is a glutamic acid, a glutamine, or a lysine, and
20 X₅₅ is a proline, an arginine, a threonine, or no residue.
2. The brazzein variant of claim 1, wherein the variant has at least 85% sequence identity to a wild type brazzein.
- 25 3. The brazzein variant of claim 1 or claim 2, wherein the variant has 4 to 8 amino acids differences from a wild type brazzein.
4. The brazzein variant of any one of the preceding claims, wherein the variant retains substantially the same structure as wild type brazzein.
- 30 5. The brazzein variant of any one of the preceding claims, wherein the variant has the same secondary structure as wild type brazzein.
6. The brazzein variant of any one of the preceding claims, wherein the variant is
35 capable of binding to a taste receptor.

7. The brazzein variant of any one of the preceding claims, wherein the variant is capable of binding to a human sweet taste receptor.
- 5 8. The brazzein variant of any one of the preceding claims, wherein the variant comprises
- X₁ is a glutamine (Q) or a serine (S),
 X₂ is an aspartic acid (D) or a glutamic acid (E),
 X₆ is a glutamine (Q), a lysine (K), or an arginine (R),
 10 X₇ is valine (V), methionine (M), or isoleucine (I)
 X₉ is a glutamic acid (E) or an arginine (R),
 X₁₃ is isoleucine (I) or valine (V),
 X₁₄ is an asparagine (N), a serine (S), or an arginine (R),
 X₁₇ is a lysine (K) or a glutamine (Q),
 15 X₁₈ is a leucine (L) or isoleucine (I),
 X₂₀ is a serine (S) or an asparagine (N),
 X₃₄ is a serine (S), an alanine (A), or a threonine (T),
 X₄₂ is an arginine (R) or a lysine (K),
 X₄₅ is a methionine (M), leucine (L), or valine (V),
 20 X₄₆ is a glutamine (Q) or a threonine (T)
 X₄₈ is an isoleucine (I), valine (V), or leucine (L),
 X₅₃ is independently a glutamic acid, a glutamine, or a lysine, and
 X₅₅ is independently a proline, an arginine, a threonine, or no residue.
- 25 9. The brazzein variant of any one of the preceding claims, wherein the variant comprises one or more of substitutions selected from the group consisting of K6Q, E9R, S14R, Q17K, S34A, E53Q, or E53K.
10. The brazzein variant of any one of the preceding claims, wherein the variant
 30 comprises no residue at position 55.
11. The brazzein variant of any one of the preceding claims, wherein the variant comprises
- i. K6Q, and Q17K,
 35 ii. E9R,

- iii. S34A, and E53Q,
- iv. S14R, and an addition of 55R, or
- v. E53K, and an addition of 55T.

- 5 12. The brazzein variant of any one of the preceding claims, wherein the variant comprises the sequence selected from the group consisting of

QDKCKQVYENYPINKCKLASQCNYDCKLDKHARSGECFYDERRNMQCICDY
CEY (SEQ ID NO: 3),

- 10 CEY (SEQ ID NO: 4),

QDKCKKMYENYPVSKCQLANQCNYDCKLDKHARAGECFYDEKRNVCICDY
CQY (SEQ ID NO: 5),

SDKCKKVYENYPIRKCQLANQCNYDCKLDKHARSGECFYDEKRNLCICDYC
EYR (SEQ ID NO: 6), and

- 15 QDKCKRVYENYPISKQIANQCNYDCKLDKHARSGECFYDEKRNLTCLCDYC
KYT (SEQ ID NO: 7).

13. A polynucleotide and/or a vector encoding a brazzein variant of any one of the preceding claims.

20

14. A cell comprising a vector comprising the polynucleotide of claim 13.

15. A method of producing a brazzein variant, comprising
expressing a polynucleotide and/or a vector encoding a brazzein variant of claims

- 25 1 to 14 in an expression system.