Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Umami-Enriched Capsicum: In Silico Design of Glutamate Accumulation by Glutamate Decarboxylase Knockout and Glutamate Dehydrogenase Expression

Priyanca Ford

scovillesplice@gmail.com

Scoville Splice LLC https://orcid.org/0000-0003-0395-1752

Research Article

Keywords: Capsicum annuum, umami, free glutamate, glutamate decarboxylase, glutamate dehydrogenase, CRISPR-Cas9 guide design, GABA shunt, Iyengar Flavor Index

Posted Date: November 26th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11875351/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

SCOVILLESPLICE.FARM

SCOVILLE SPLICE RESEARCH SERIES

Umami-Enriched Capsicum: In Silico Design of Glutamate Accumulation by Glutamate Decarboxylase Knockout and Glutamate Dehydrogenase Expression

A computed design that lowers glutamate loss through the GABA shunt and adds glutamate synthesis in Capsicum annuum fruit placenta.

P.I. Ford

Scoville Splice LLC · ORCID: 0000-0003-0395-1752

DOI: 10.5281/zenodo.24511324 License: CC BY 4.0 **Patents:** US 64/172,827; US 64/188,197; US 64/192,062; US 64/199,105

Figshare: figshare/35015809 Harvard Dataverse: DVN/ILRDWN OSF: osf.io/wsnfz Hugging Face: capsaicinoid-data

GitHub: paper-05 Website: umami-enriched-capsicum

Package record. 91 claims traced to a file or a cited source (60 computed, 13 definition, 15 design assumption, 2 design target, 1 literature value). 17 tables, 6 figures, 41 data files. Data fingerprint `4570e79d8c5e`. Rerun every result with `bash run_all.sh`.



Pepper cross-section. Scoville Splice LLC image library.

ABSTRACT

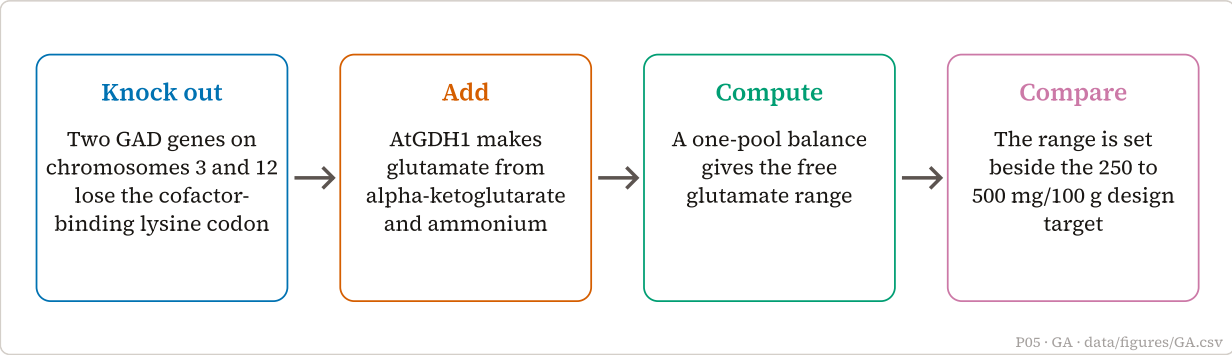
Free glutamate is a main umami compound, and the GABA shunt drains part of it in pepper placenta. This paper describes a computed design that combines two changes. The first is a knockout of two glutamate decarboxylase genes, *CaGAD1* (LOC107862809) on chromosome 3 and *CaGAD4* (LOC107850578) on chromosome 12. The second is expression of the Arabidopsis glutamate dehydrogenase AtGDH1, which makes glutamate from alpha-ketoglutarate and ammonium. In a public placenta RNA-seq series, the two genes carry 99.37 to 99.97% of the transcript of the four full-length GAD genes. Four guide RNAs were designed, two per gene, each with one perfect genomic site. Each pair deletes the codon of the cofactor-binding lysine, 476 and 595 coding bases. A one-pool glutamate balance built from stated design assumptions and a 65 mg/100 g (midpoint of 50 to 80) baseline gives 81 to 243 mg/100 g for the combined design, 70 to 108 mg/100 g for the knockout alone and 76 to 146 mg/100 g for AtGDH1 alone. The design target is 250 to 500 mg/100 g, and the balance places the combined design at 81 to 243 mg/100 g beside it. All results are computed. No experimental data are reported in this paper.

Keywords: Capsicum annuum · umami · free glutamate · glutamate decarboxylase · glutamate dehydrogenase · CRISPR-Cas9 guide design · GABA shunt · Iyengar Flavor Index

KEY POINTS

- Four guides, two per gene, delete the PLP-binding lysine codon of *CaGAD1* (LOC107862809) and *CaGAD4* (LOC107850578); each guide has 1 perfect genomic site.
- In the public placenta series the two genes carry 99.37 to 99.97% of GAD transcript.
- A one-pool balance gives 81 to 243 mg/100 g for the combined design and 70 to 108 mg/100 g for the knockout alone.
- The 250 to 500 mg/100 g design target is set beside the modelled range.
- All results are computed from stated assumptions.

Marks under numbers: computed · design target · design assumption · literature value · definition. The evidence table in Appendix A lists the source file of every one.



Graphical abstract. The design in four steps: two GAD genes are edited, AtGDH1 is added, a one-pool balance is computed, and the result is set beside the design target. Drawn from the steps file. Data: `data/figures/GA.csv`; script: `code/01_pool_model.py`.

1. Introduction

L-glutamate is a main umami compound and is sensed by the TAS1R1/TAS1R3 taste receptor (Nelson et al., 2002; Li et al., 2002). Umami is described as prolonging aftertaste and as acting together with 5'-nucleotides (Yamaguchi and Ninomiya, 2000). Free glutamate in a fruit is set by synthesis and by drains. One drain is the GABA shunt, in which glutamate decarboxylase (GAD) converts glutamate to gamma-aminobutyric acid (Fait et al., 2008). Glutamate dehydrogenase (GDH) can add glutamate from alpha-ketoglutarate and ammonium, and GDH overexpression has been reported in tomato (Kisaka and Kida, 2003; Kisaka et al., 2007).

This paper designs a combined change for the placenta of *Capsicum annuum*: a double knockout of the two GAD genes that carry the placenta transcript, and expression of AtGDH1. It reports the guide designs, the placenta expression of every GAD gene, and a one-pool glutamate balance with a design target for free glutamate. Concentrations are compared with the L-glutamate reference of the umami axis of the Iyengar Flavor Index (IFI-U; 10.5281/zenodo.24511320). The IFI scores one compound against a reference assay, so this paper uses it only as a concentration reference.

2. Methods

2.1 Computational basis

Gene models are the RefSeq annotation of the GCF_002878395.1 assembly (*Capsicum annuum* UCD10Xv1.1; Hulse-Kemp et al., 2018). Guides were scanned against the whole genome with our own C scanner (strict NGG PAM, both strands, up to four mismatches) and scored with our own implementation of the Hsu et al. (2013) MIT aggregate score. The scanner and scripts are in `code/guide_design/`. Scripts run on Python 3.9. The differential-equation version of the balance in the research record uses numpy 2.0.2 and scipy 1.13.1 with the LSODA integrator (tolerances 1e-11 and 1e-13). Chromosomes are the RefSeq sequences NC_061113.1 (chromosome 3), NC_061122.1 (chromosome 12) and the other chromosome accessions listed in `inputs/gene_names.csv`.

2.2 Expression data

The public placenta RNA-seq series is PRJNA789050: reads of *Capsicum chinense* accession SL08 at five stages (4, 24, 34, 44 and 54 days post anthesis, DPA), one library per stage. The reads were quantified with salmon against the *C. annuum* coding sequences, and the transcripts per million (TPM) of each GAD locus were summed from the per-transcript values (`inputs/gad_tpm_by_stage.tsv`).

2.3 Guide design

The PLP-binding lysine of each gene was located by alignment to tomato GAD (UniProt P54767). Every SpCas9 site in the coding exons was listed, and 78 candidates passed the filters: cut inside the window of 30 to 450 coding bases either side of the lysine codon, GC content within the range of 40 to 70%, and no TTTT or five-base run. One guide upstream and one downstream of the lysine codon was chosen for each gene from the guides with exactly one perfect genomic site. Three of the four are the highest-scoring such guide on their side. For *CaGAD4* (LOC107850578) upstream, the MIT score and the coding bases deleted are 86.4 and 595; 87.6 and 221, for the chosen guide and the highest-scoring guide respectively. Each cut position is the first plus-strand base after the cut. The lysine alignment is a global pairwise alignment (BLOSUM62, gap open -10, gap extend -0.5).

2.3a Scan and scoring

The C scanner lists every genome site with a strict NGG PAM, on both strands, with up to four mismatches to each candidate; it found 12,328 sites for 78 guides (Table S10). The MIT score of a guide is 100 / (100 + the sum of its site scores) × 100. The score of one site is the product of $(1 - w)$ over its mismatch positions, where w is the position weight (20 weights, 0 to 0.851; Table S9), times the factor $1 / (((19 - \bar{d}) / 19) \times 4 + 1)$ for sites with more than one mismatch, where $\bar{d}$ is the mean distance between mismatches, times $1 / n^2$ for n mismatches, times 100 (`code/guide_design/step4_score.py`).

2.4 Glutamate balance

Free glutamate is one pool. In wild type the pool is in steady state: synthesis equals drain, and the GABA shunt carries a share s of the drain. The double knockout removes a fraction r of shunt activity, and AtGDH1 adds a synthesis flux dJ. Relative to wild type the steady state is $(1 + dJ / J_0) / (1 - s\,r)$, where J_0 is the wild-type shunt flux divided by s. Table 3 lists every input; all are design assumptions with no source: a baseline of 65 mg/100 g (midpoint of 50 to 80), shunt share of 0.10 to 0.40, knockout fraction of 0.70 to 1.00, wild-type shunt flux of 48.3 umol/gDW/h and added flux of 80 to 150 umol/gDW/h. The balance was computed over the full ranges (`code/01_pool_model.py`); the differential-equation version in the research record gives the same values (342 values, 3.82e-10; Table S11).

2.5 Concentration comparison

Free glutamate in mg/100 g was converted with (mg/100 g × 10) / 147.13 g/mol = mM, taking 1 g of tissue as 1 mL. The concentration was divided by the IFI-U reference EC50 of L-glutamate at TAS1R1/TAS1R3, 2,500 micromolar (2.5 mM). The ratio compares a tissue concentration with the midpoint of the reference assay and says nothing about receptor response or perceived intensity.

2.6 AtGDH1 and enzyme arithmetic

The added enzyme is the Arabidopsis NADH-dependent glutamate dehydrogenase AtGDH1: At5g18170, Q43314, 411 residues, 44.5 kDa, EC 1.4.1.3 (UniProt record). The amount of enzyme in the tissue is the share of soluble protein times the total soluble protein, divided by the molecular mass. Its capacity in umol/gDW/h is that amount times kcat times the substrate saturation, [alpha-ketoglutarate] / (Km + [alpha-ketoglutarate]) × [ammonium] / (Km + [ammonium]), divided by the dry-weight fraction. The inputs are design assumptions with no source: a share of 0.005 to 0.02, total soluble protein of 15 mg/g, kcat of 28 per second, Km values of 1.8 and 2.5 mM, pools of 0.3 and 0.4; 0.8 and 1.1 mM and a dry-weight fraction of 0.12.

3. Results

3.1 The GAD genes and their placenta expression

The pepper reference carries four full-length GAD genes: *CaGAD1* (LOC107862809) on chromosome 3, *CaGAD2* (LOC107853724) on chromosome 5, *CaGAD3* (LOC107838791) on chromosome 8 and *CaGAD4* (LOC107850578) on chromosome 12. A GAD-like fragment of 116 residues (LOC107877256) lies on chromosome 11. Table 1 and Figure 2 give the placenta expression. *CaGAD1* (LOC107862809) rises from 0.1 TPM at 4 DPA to 142.4 to 260.6 TPM at 34 to 54 DPA, and *CaGAD4* (LOC107850578) falls from 150.5 TPM to 67.7 to 82.1 TPM over the same stages. *CaGAD2* (LOC107853724) and *CaGAD3* (LOC107838791) stay at 0.0 to 0.7 TPM, and the fragment at 0.0 to 2.6 TPM. At every stage *CaGAD1* (LOC107862809) and *CaGAD4* (LOC107850578) carry 99.37 to 99.97% of the transcript of the four full-length genes. The companion design paper on trait expansion (10.5281/zenodo.24511350) edits *CaGAD3* (LOC107838791) instead, which reaches at most 0.7 TPM in placenta.

Table 1. GAD loci in placenta. Placenta expression (TPM) of the five GAD loci at five stages in the public series (one library per stage). The last row is the share of the four full-length GAD genes carried by CaGAD1 (LOC107862809) and CaGAD4 (LOC107850578). Data: `data/tables/T01.csv`*; script:* `code/03_expression.py`*.*

Locus	Name in this paper	Chromosome	4 DPA	24 DPA	34 DPA	44 DPA	54 DPA
LOC107862809	CaGAD1	3	0.1	14.3	142.4	260.6	198.2
LOC107853724	CaGAD2	5	0.0	0.0	0.0	0.0	0.0
LOC107838791	CaGAD3	8	0.1	0.7	0.4	0.1	0.2
LOC107850578	CaGAD4	12	150.5	92.0	82.1	67.7	67.9
LOC107877256	GAD-like fragment	11	0.0	0.0	2.5	1.3	2.6
Share of the four full-length GAD genes carried by CaGAD1 + CaGAD4 (%)			99.91	99.37	99.80	99.97	99.92

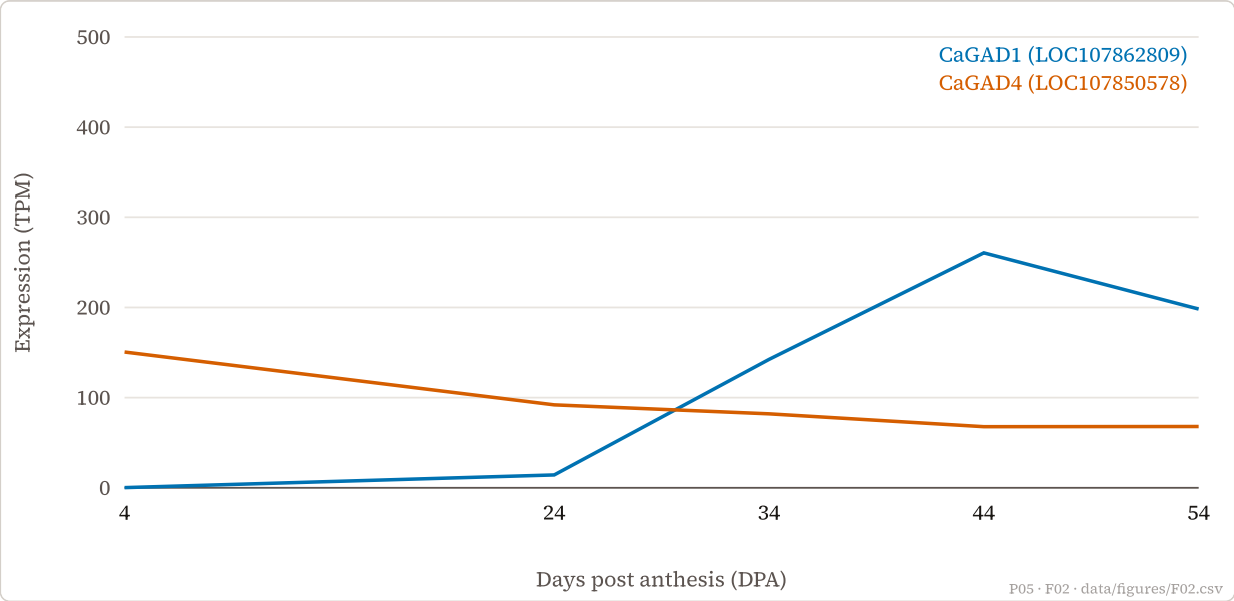


Figure 2. Placenta expression of CaGAD1 and CaGAD4. Transcripts per million in the public placenta series, one library per stage. Per-transcript quantification summed per gene. Data: `data/figures/F02.csv`; script: `code/03_expression.py`.

3.2 Guide designs

CaGAD1 (LOC107862809) and *CaGAD4* (LOC107850578) each have a PLP-binding lysine codon; the lysine is residue 278 in *CaGAD1* (LOC107862809) and residue 277 in *CaGAD4* (LOC107850578). Each coding sequence has 1,509 bases. Table 2 lists the four guides and Figure 4 shows where they cut. The four MIT scores are 86.4 to 97.4, and each guide has 1 perfect site in the genome. Each pair of cuts removes the lysine codon: *CaGAD1* (LOC107862809) loses 992 bp, 476 coding bases (31.5%) and *CaGAD4* (LOC107850578) loses 4,233 bp, 595 coding bases (39.4%). The remainders after division by three are 2 and 1, so each deletion shifts the reading frame. Of the genome sites that lie within one to three mismatches of a guide, 1 of 12. No guide has a site inside another GAD locus (4 mismatches checked). Across the 78 candidates, 60 of 78 have exactly one perfect site and the MIT scores run from 14.5 to 99.0 (Table S1, Figure 5). The scan sites by mismatch number are 100, 50, 88, 906 and 11,184 (Table S10). *CaGAD1* (LOC107862809) has 5 and 6 coding exons (Table S7). The annotated sites with one to three mismatches are in Table S6.

Table 2. The four designed guides. Each guide has exactly one perfect site in the genome. MIT score is the Hsu et al. (2013) aggregate specificity score. The cut position is the first base after the cut. Sites are counted with 0 (perfect), 1, 2 and 3 mismatches. Data: `data/tables/T02.csv`; script: `code/02_guides.py`.

Guide	Gene (RefSeq locus)	Protospacer and PAM	Strand	Genomic cut	Coding cut (exon)	GC (%)	MIT score	Sites at 0 / 1 / 2 / 3 mismatches
CaGAD1-g1	CaGAD1 (LOC107862809)	ATATGCGTTGCCGCTATTCT TGG	+	chr3:239,110,576	630 (3)	45.0	92.2	1 / 0 / 1 / 3
CaGAD1-g2	CaGAD1 (LOC107862809)	GCGGTTCGATATTGTCTCTA AGG	+	chr3:239,111,568	1106 (5)	45.0	97.4	1 / 0 / 0 / 0
CaGAD4-g1	CaGAD4 (LOC107850578)	AGCAATTATGCTTGCTGGAT TGG	+	chr12:2,170,542	401 (3)	40.0	86.4	1 / 0 / 0 / 3
CaGAD4-g2	CaGAD4 (LOC107850578)	CATAACCCAAGCGAATAAGT TGG	-	chr12:2,174,775	996 (6)	40.0	88.3	1 / 0 / 1 / 4

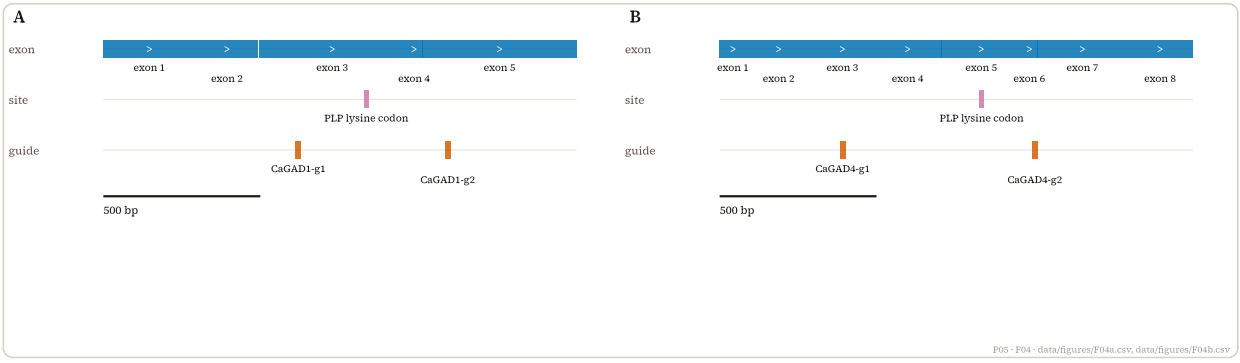


Figure 4. Guide sites in the coding sequences. A, CaGAD1 (LOC107862809). B, CaGAD4 (LOC107850578). Exons, the PLP-binding lysine codon and the four guides along the coding sequence; the scale bar is in base pairs. Positions are coding-sequence coordinates computed from the exon coordinates. Data: `data/figures/F04a.csv`, `data/figures/F04b.csv`; script: `code/02_guides.py`.

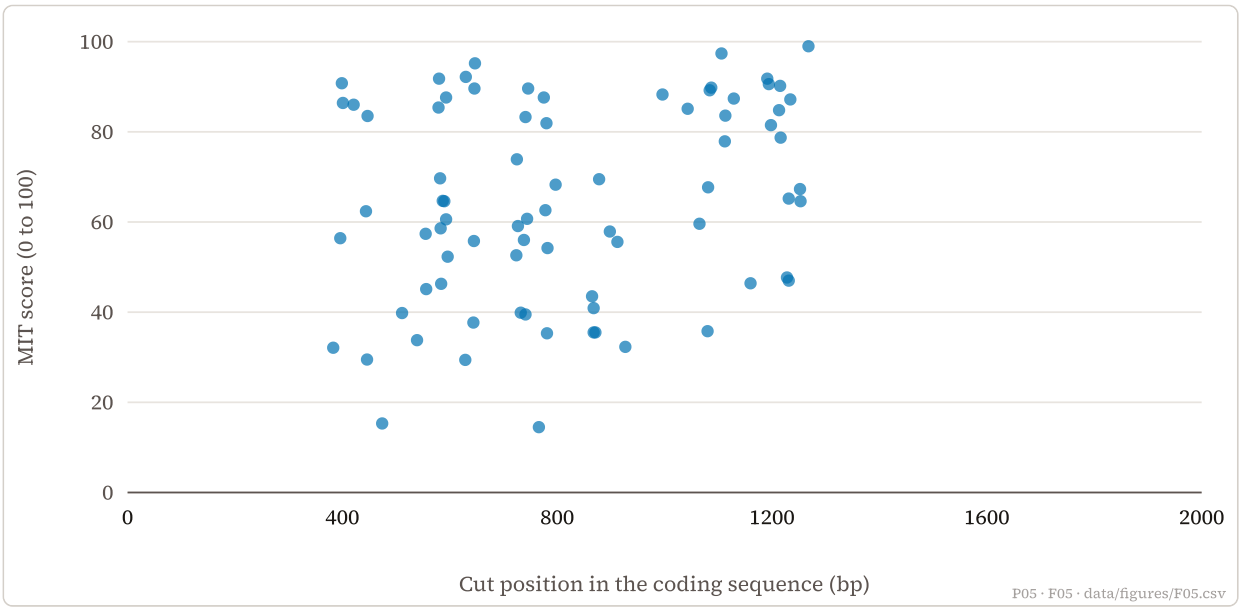


Figure 5. Specificity of the 78 candidate guides. Each point is one candidate guide: its cut position along the coding sequence and its MIT score. Scores from the whole-genome scan; both genes are plotted together. Data: `data/figures/F05.csv`; script: `code/04_supplementary.py`.

3.3 Free glutamate from the balance

Table 4 and Figure 1 give the modelled free glutamate at the 65 mg/100 g baseline. The knockout alone gives 70 to 108 mg/100 g, AtGDH1 alone gives 76 to 146 mg/100 g and the combination gives 81 to 243 mg/100 g (mid-range 132), which is 1.25 to 3.74 times the baseline. The combination is higher than either change alone over the whole input range. The design target of 250 to 500 mg/100 g is a target and is not a model result. The highest modelled value of the combined design is 243, beside the lower bound of 250. The balance also gives the added AtGDH1 flux that would reach 250: 158 umol/gDW/h with complete knockout at the highest shunt share, and 392 umol/gDW/h at the mid-range shunt share and knockout. Figure 3 shows how each input moves the combined result: the shunt share moves it most (88.0 to 192.3) and the knockout fraction least (125.7 to 138.3), around a centre of 131.7. The baseline moves it from 101.3 to 162.1 and the added flux from 116.7 to 146.6 (Table S5). Figure 6 shows the combined design as the shunt share changes; at the highest shunt share with mid-range knockout the combined result is 163.7 to 220.8 for the three added fluxes. Table S4 gives every combination of inputs.

Table 3. Inputs of the glutamate balance. Every input is a design assumption with no source; mid values are range midpoints. The 65 mg/100 g baseline is the midpoint of 50 to 80. Data: `data/tables/T03.csv` *; script:* `code/01_pool_model.py` *.*

Input	Low	Mid	High	Unit	Kind
Free glutamate of wild-type placenta (baseline)	50	65	80	mg/100 g fresh weight	design assumption
Share of glutamate loss carried by the GABA shunt (s)	0.10	0.25	0.40	fraction	design assumption
Fraction of shunt activity removed by the double knockout (r)	0.70	0.85	1.00	fraction	design assumption
Wild-type GABA-shunt flux		48.3		umol/gDW/h	design assumption
Glutamate synthesis flux added by AtGDH1 (dJ)	80	115	150	umol/gDW/h	design assumption

Table 4. Modelled free glutamate. Steady-state free glutamate from the balance over the full ranges of Table 3 at the 65 mg/100 g baseline. The mid-range column uses the mid values. Data: `data/tables/T04.csv` *; script:* `code/01_pool_model.py` *.*

Genotype	Lowest (mg/100 g)	Mid-range (mg/100 g)	Highest (mg/100 g)	Lowest (mM)	Highest (mM)
Wild type (assumed baseline)	65	65	65	4.4	4.4
GAD knockout only	70	83	108	4.8	7.4
AtGDH1 only	76	104	146	5.2	9.9
Knockout + AtGDH1	81	132	243	5.5	16.5

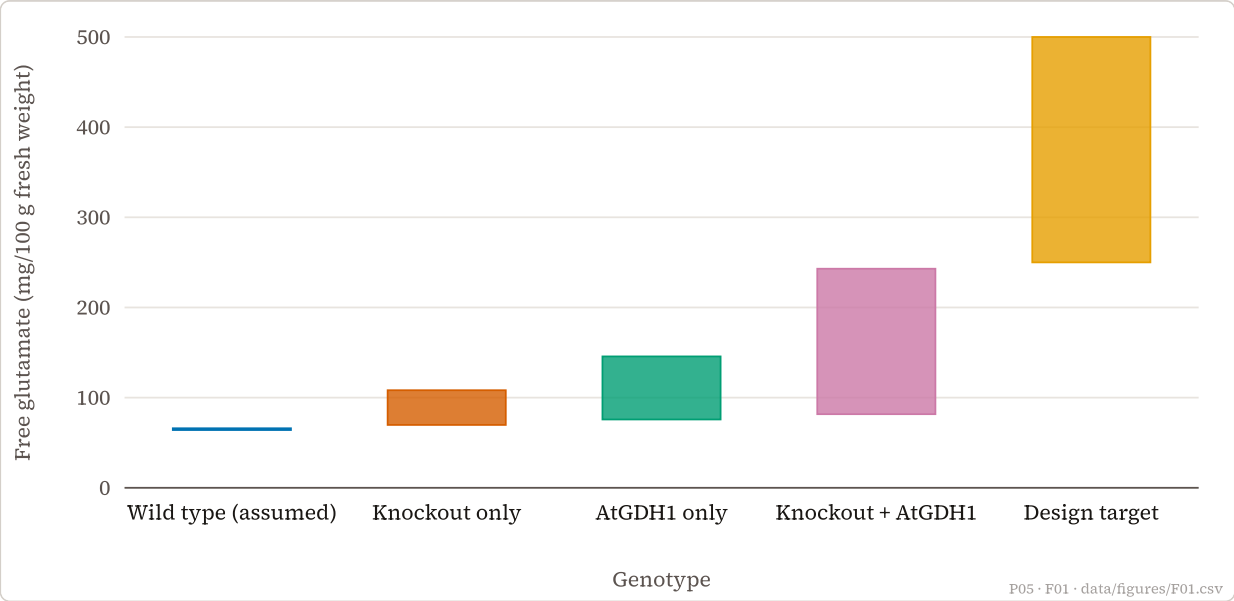


Figure 1. Modelled free glutamate by genotype. Bars span the lowest to the highest value of the glutamate balance over the input ranges of Table 3, at an assumed 65 mg/100 g baseline. The wild-type bar is a narrow marker at the baseline. The right-hand bar is the 250 to 500 mg/100 g design target, which is a target and not a model result. Closed-form steady state of a one-pool balance. Data: `data/figures/F01.csv`; script: `code/01_pool_model.py`.

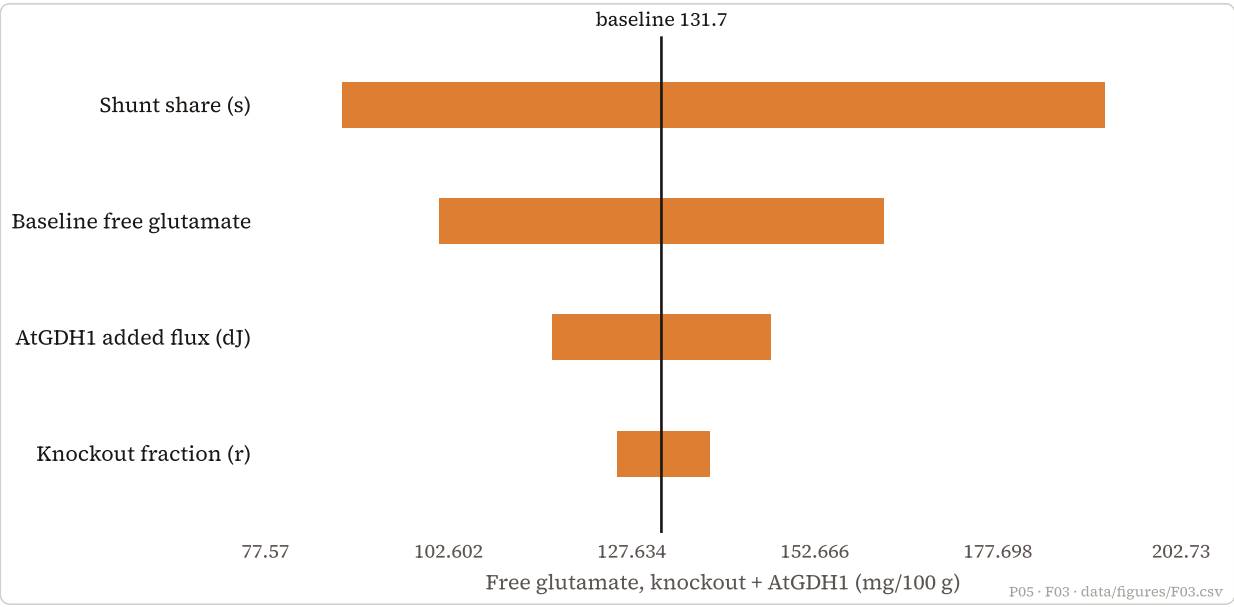


Figure 3. Sensitivity of the combined design to each input. Each bar spans the result when one input moves across its range of Table 3 and the others stay at mid-range. One-at-a-time analysis of the balance. Data: `data/figures/F03.csv`; script: `code/01_pool_model.py`.

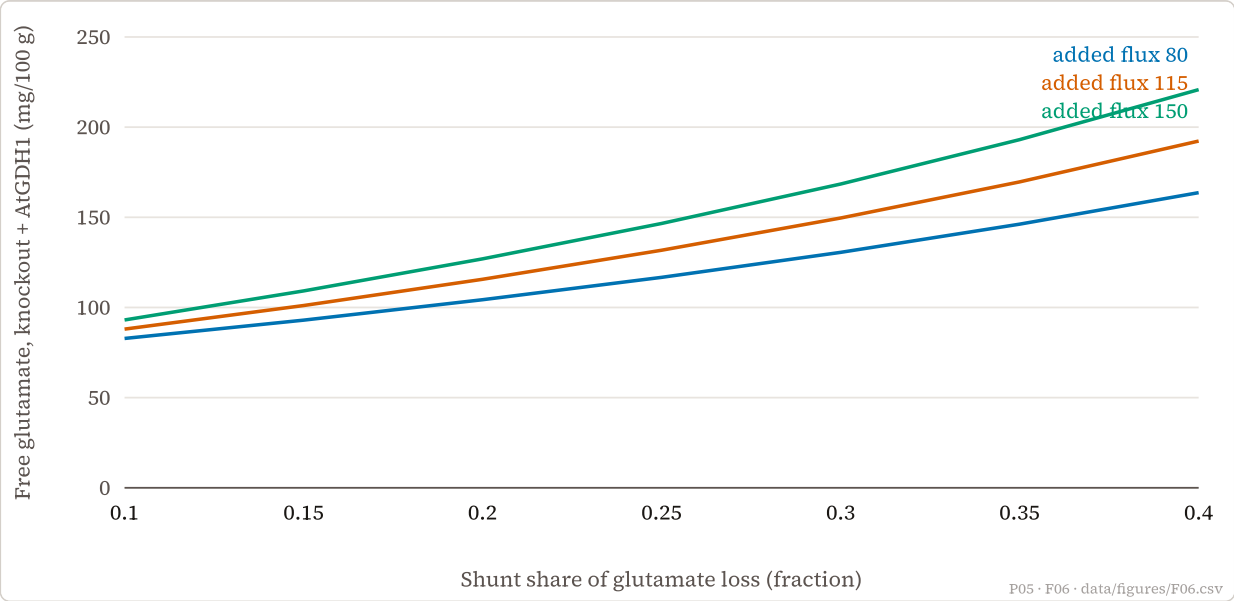


Figure 6. Combined design as the shunt share changes. Free glutamate of the combined design against the assumed shunt share, for three assumed added fluxes (umol/gDW/h) and the mid-range knockout fraction. Closed-form steady state of the one-pool balance. Data: `data/figures/F06.csv`; script: `code/04_supplementary.py`.

3.4 Comparison with the reference concentration

Table 5 converts the levels to millimolar. The baseline is 4.4 mM, ratio 1.8. The knockout alone gives 4.8 to 7.4 mM, ratio 1.9 to 2.9 and the combined design gives 5.5 to 16.5 mM, ratio 2.2 to 6.6. The design target is 17.0 to 34.0 mM, ratio 6.8 to 13.6. Levels of 100 and 150 mg/100 g (6.8 and 10.2 mM) are used as reference levels for orientation.

Table 5. Concentration comparison. Free glutamate converted to millimolar and divided by the 2.5 mM reference EC50 of L-glutamate at TAS1R1/TAS1R3. The ratio compares a concentration with the reference midpoint and is not a receptor response. Data: `data/tables/T05.csv`; script: `code/01_pool_model.py`.

Level	Free glutamate (mg/100 g)	Free glutamate (mM)	Ratio to the reference EC50
Baseline (assumed)	65	4.4	1.8
Reference level	100	6.8	2.7
Reference level	150	10.2	4.1
Knockout only (modelled range)	70-108	4.8-7.4	1.9-2.9
Knockout + AtGDH1 (modelled range)	81-243	5.5-16.5	2.2-6.6
Design target, lower bound	250	17.0	6.8
Design target, upper bound	500	34.0	13.6

3.5 Dependence on the baseline

The result scales with the assumed baseline. Table 6 gives the combined design for baselines of 50 and 80 mg/100 g: 63 to 187 and 100 to 299 mg/100 g.

Table 6. Dependence on the baseline. The combined design computed from three assumed baselines over the same input ranges. Data: `data/tables/T06.csv`*; script:* `code/01_pool_model.py`.

Baseline (mg/100 g)	Knockout + AtGDH1, lowest (mg/100 g)	Knockout + AtGDH1, highest (mg/100 g)
50	63	187
65	81	243
80	100	299

3.6 AtGDH1 amount and capacity

At 1% of soluble protein, AtGDH1 is 150 ug/g and 3.37 nmol/g (Table S8). Its capacity at the pool sizes of engineered fruit is 27.9 to 111.5 umol/gDW/h for the assumed shares of soluble protein, at wild-type pool sizes 133.0 to 532.1 and with alpha-ketoglutarate saturating at 1% 864.7. The assumed added flux of 80 to 150 overlaps the capacity at both pool sizes.

3.7 Agreement of the two routes

The closed-form balance and the differential-equation version agree: 342 values, 3.82e-10 (Table S11).

4. Limitations

The glutamate balance is a one-pool model, and every input is a design assumption with no source, so the modelled range scales with those assumptions (Table 6, Figure 3). GDH is reported to act mainly in the direction of glutamate breakdown in vivo (Labboun et al., 2009); the added AtGDH1 flux is an assumption. The expression series has one library per stage, so no error bars are attached. Guide specificity is the MIT score; on-target activity scores are not computed.

5. Data and code availability

Everything behind this paper is in the data package: the claims table, the data for every table and figure, the scripts that regenerate them, the inputs, the research record and a backup. `bash run_all.sh` reruns every step and compares each regenerated file with the stored file; `bash verify.sh` checks the checksums. The paper and data are licensed under CC BY 4.0 and the code under the MIT licence. The public input datasets are listed in `inputs/input_sources.csv`.

Public records of this paper: Zenodo 10.5281/zenodo.24511324; Figshare: figshare/35015809; Harvard Dataverse: DVN/ILRDWN; OSF: osf.io/wsnfz; Hugging Face: capsaicinoid-data; GitHub: paper-05; Website: umami-enriched-capsicum.

Intellectual Property. The engineered constructs, cultivar designs, and methodologies described in this paper are the subject of pending United States patent applications (USPTO Provisional Application Nos. 64/172,827; 64/188,197; 64/192,062; and 64/199,105).

Experimental status. All results reported in this paper are computational. Experimental work on these designs is under way at the authors' laboratory facilities in Montecito, California, and Geneva, Switzerland. Those data are not part of this paper, are shared only with commercial partners under agreement, and are planned for publication next year.

Active experiments for this design include a free amino acid baseline in placenta and pericarp (Glu, GABA, Gln, Asp, Cys), resequencing of every target locus in the lab lines, in vitro cleavage and protoplast RNP editing with the guides at all GAD loci, and sensory panel screening, training and taster-status genotyping.

6. References

1. Fait, A., Fromm, H., Walter, D., Galili, G., & Fernie, A. R. (2008). Highway or byway: the metabolic role of the GABA shunt in plants. *Trends in Plant Science, 13*(1), 14–19. doi:10.1016/j.tplants.2007.10.005
2. Ford, P.I. (2026). The Iyengar Flavor Index (IFI): A Receptor-Normalized Standard for Cross-Modal Flavor Measurement. *Scoville Splice Research Series*. doi:10.5281/zenodo.24511320
3. Ford, P.I. (2026). Trait Expansion in Capsicum: CRISPR Guide Design for Nine Trait Groups with Gene Assignment and Genome-Wide Off-Target Profiling. *Scoville Splice Research Series*. doi:10.5281/zenodo.24511350
4. Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., ... & Zhang, F. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnology, 31*(9), 827–832. doi:10.1038/nbt.2647
5. Hulse-Kemp, A. M., Maheshwari, S., Stoffel, K., et al. (2018). Reference quality assembly of the 3.5-Gb genome of *Capsicum annuum* from a single linked-read library. *Horticulture Research, 5*, 4. doi:10.1038/s41438-017-0011-0
6. Kisaka, H., & Kida, T. (2003). Transgenic tomato plant carrying a gene for NADP-dependent glutamate dehydrogenase (*gdhA*) from *Aspergillus nidulans*. *Plant Science, 164*(1), 35–42. doi:10.1016/S0168-9452(02)00325-4
7. Kisaka, H., Kida, T., & Miwa, T. (2007). Transgenic tomato plants that overexpress a gene for NADH-dependent glutamate dehydrogenase (*legdh1*). *Breeding Science, 57*(2), 101–106. doi:10.1270/jsbbs.57.101
8. Labboun, S., Tercé-Laforgue, T., Roscher, A., Bedu, M., Restivo, F. M., Velanis, C. N., ... & Hirel, B. (2009). Resolving the role of plant glutamate dehydrogenase. I. In vivo real time nuclear magnetic resonance spectroscopy experiments. *Plant & Cell Physiology, 50*(10), 1761–1773. doi:10.1093/pcp/pcp118
9. Li, X., Staszewski, L., Xu, H., Durick, K., Zoller, M., & Adler, E. (2002). Human receptors for sweet and umami taste. *Proceedings of the National Academy of Sciences, 99*(7), 4692–4696. doi:10.1073/pnas.072090199
10. Nelson, G., Chandrashekar, J., Hoon, M. A., Feng, L., Zhao, G., Ryba, N. J., & Zuker, C. S. (2002). An amino-acid taste receptor. *Nature, 416*(6877), 199–202. doi:10.1038/416199a
11. Yamaguchi, S., & Ninomiya, K. (2000). Umami and food palatability. *Journal of Nutrition, 130*(4S), 921S–926S. doi:10.1093/jn/130.4.921S

Supplementary data

The supplementary tables hold the full data behind the paper. Table S1 lists all 78 candidate guides with their cut positions, site counts and MIT scores. Table S2 gives the GAD expression to three decimals with the sums behind the share. Table S3 lists every parameter of the glutamate balance and the enzyme arithmetic with its basis. Table S4 gives the balance at every combination of the low, mid and high inputs, and Table S5 gives the sensitivity ranges. Table S6 lists the annotated genome sites with one to three mismatches to the four guides. Table S7 gives the exon coordinates. Table S8 gives the AtGDH1 amount and capacity. Table S9 gives the MIT position weights, Table S10 the scan hits by mismatch number and Table S11 the agreement of the two routes of the balance. Every table is stored as a CSV file in `data/tables/`, and the scripts that write them are in `code/`. The file map below lists every data, input, code and research-record file.

Table S1. All 78 candidate guides. Every candidate that passed the filters, with cut positions, exon, GC content, site counts at 0 to 4 mismatches and MIT score. The four chosen guides are marked. Data: `data/tables/S01.csv`; script: `code/04_supplementary.py`.

Gene (locus)	Side of the lysine	Strand	Protospacer	PAM	Chromosome	Genomic cut	Coding cut	Exon	GC (%)	Sites at 0 / 1 / 2 / 3 / 4 mismatches	MIT score	Chosen
CaGAD1 (LOC107862809)	up	+	TCTTGGTTCCACTCTGACGG	GGG	chr3	239,110,593	647	3	55.0	2 / 0 / 0 / 6 / 32	95.2	
CaGAD1 (LOC107862809)	up	+	ATATGCGTTGCCGCTATTCT	TGG	chr3	239,110,576	630	3	45.0	1 / 0 / 1 / 3 / 38	92.2	yes
CaGAD1 (LOC107862809)	up	+	TCAGAGGCAATCATGCTTGC	TGG	chr3	239,109,822	399	2	50.0	1 / 0 / 1 / 3 / 23	90.8	
CaGAD1 (LOC107862809)	up	+	TTCTTGGTTCCACTCTGACG	GGG	chr3	239,110,592	646	3	50.0	2 / 0 / 1 / 17 / 58	89.6	
CaGAD1 (LOC107862809)	up	-	AGGAGCAATAAATCCTCCAC	TGG	chr3	239,110,692	746	3	45.0	3 / 0 / 3 / 2 / 60	89.6	
CaGAD1 (LOC107862809)	up	+	CCCTGCCAAAGCAGTTGAGA	TGG	chr3	239,110,539	593	3	55.0	1 / 0 / 5 / 4 / 20	87.6	
CaGAD1 (LOC107862809)	up	+	GTCTTGCTTTCAAACGCAAA	TGG	chr3	239,109,844	421	2	40.0	1 / 0 / 0 / 4 / 121	86.0	
CaGAD1 (LOC107862809)	up	+	ATCCATGTCGATGCTGCCAG	TGG	chr3	239,110,687	741	3	55.0	2 / 1 / 0 / 2 / 35	83.3	
CaGAD1 (LOC107862809)	up	-	CCATCTCAACTGCTTTGGCA	GGG	chr3	239,110,528	582	3	50.0	1 / 0 / 3 / 4 / 57	69.7	
CaGAD1 (LOC107862809)	up	-	ATCCACCATCTCAACTGCTT	TGG	chr3	239,110,533	587	3	45.0	1 / 1 / 2 / 10 / 88	64.7	
CaGAD1 (LOC107862809)	up	-	CGGAAATCCCATTCAAGATC	TGG	chr3	239,110,724	778	3	45.0	1 / 1 / 2 / 0 / 28	62.6	
CaGAD1 (LOC107862809)	up	+	CAAGCGAGAAGAAAAGCAGA	AGG	chr3	239,109,867	444	2	45.0	1 / 0 / 1 / 12 / 311	62.4	
CaGAD1 (LOC107862809)	up	+	CATGTCGATGCTGCCAGTGG	AGG	chr3	239,110,690	744	3	60.0	2 / 1 / 0 / 6 / 45	60.7	
CaGAD1 (LOC107862809)	up	-	CTGGCAGCATCGACATGGAT	CGG	chr3	239,110,673	727	3	55.0	3 / 0 / 0 / 2 / 89	59.1	
CaGAD1 (LOC107862809)	up	-	ACCATCTCAACTGCTTTGGC	AGG	chr3	239,110,529	583	3	50.0	1 / 1 / 3 / 5 / 37	58.6	
CaGAD1 (LOC107862809)	up	+	AAGGAGGTGAAACTCAAAGA	AGG	chr3	239,110,501	555	3	40.0	1 / 0 / 5 / 25 / 257	57.4	

Gene (locus)	Side of the lysine	Strand	Protospacer	PAM	Chromosome	Genomic cut	Coding cut	Exon	GC (%)	Sites at 0 / 1 / 2 / 3 / 4 mismatches	MIT score	Chosen
CaGAD1 (LOC107862809)	up	+	ATTCTTGGTTCCACTCTGAC	GGG	chr3	239,110,591	645	3	45.0	2 / 1 / 3 / 3 / 57	55.8	
CaGAD1 (LOC107862809)	up	+	CTTGTGGCCAGATCTTGAAT	GGG	chr3	239,110,728	782	3	45.0	1 / 2 / 1 / 12 / 99	54.2	
CaGAD1 (LOC107862809)	up	+	TGCCAAAGCAGTTGAGATGG	TGG	chr3	239,110,542	596	3	50.0	1 / 1 / 1 / 19 / 127	52.3	
CaGAD1 (LOC107862809)	up	+	AGGAGGTGAAACTCAAAGAA	GGG	chr3	239,110,502	556	3	40.0	1 / 0 / 11 / 44 / 424	45.1	
CaGAD1 (LOC107862809)	up	-	CTCCACTGGCAGCATCGACA	TGG	chr3	239,110,678	732	3	60.0	2 / 2 / 0 / 4 / 8	39.9	
CaGAD1 (LOC107862809)	up	+	TATTCTTGGTTCCACTCTGA	CGG	chr3	239,110,590	644	3	40.0	2 / 1 / 3 / 9 / 165	37.7	
CaGAD1 (LOC107862809)	up	+	TCTTGTGGCCAGATCTTGAA	TGG	chr3	239,110,727	781	3	45.0	1 / 3 / 0 / 11 / 212	35.3	
CaGAD1 (LOC107862809)	up	+	CTTTGAGGTTGAATTGAAGG	AGG	chr3	239,110,485	539	3	40.0	1 / 3 / 3 / 9 / 288	33.8	
CaGAD1 (LOC107862809)	up	+	TGGTACAGTGGGTTCATCAG	AGG	chr3	239,109,806	383	2	50.0	1 / 3 / 0 / 7 / 38	32.1	
CaGAD1 (LOC107862809)	up	-	CAGAGTGGAACCAAGAATAG	CGG	chr3	239,110,575	629	3	45.0	1 / 1 / 1 / 39 / 560	29.4	
CaGAD1 (LOC107862809)	up	-	TCAAGATCTGGCCACAAGAA	AGG	chr3	239,110,712	766	3	45.0	1 / 3 / 1 / 63 / 657	14.5	
CaGAD1 (LOC107862809)	down	+	TCTGCGGGTTGTCATTCGAG	AGG	chr3	239,111,730	1268	5	55.0	2 / 0 / 0 / 1 / 8	99.0	
CaGAD1 (LOC107862809)	down	+	GCGGTTCGATATTGTCTCTA	AGG	chr3	239,111,568	1106	5	45.0	1 / 0 / 0 / 0 / 24	97.4	yes
CaGAD1 (LOC107862809)	down	-	CGATCCATCCAAATCTTCTG	AGG	chr3	239,111,653	1191	5	45.0	1 / 0 / 0 / 2 / 44	91.8	
CaGAD1 (LOC107862809)	down	+	TCTGAGCACCTCAGAAGATT	TGG	chr3	239,111,656	1194	5	45.0	1 / 0 / 1 / 3 / 26	90.6	
CaGAD1 (LOC107862809)	down	-	CTGGCGGCATTGTGTATGCT	GGG	chr3	239,111,677	1215	5	55.0	1 / 0 / 0 / 0 / 67	90.2	
CaGAD1 (LOC107862809)	down	+	ATGGAATCGCAAATATGGGG	CGG	chr3	239,111,549	1087	5	45.0	1 / 0 / 0 / 2 / 47	89.8	
CaGAD1 (LOC107862809)	down	+	CAGATGGAATCGCAAATATG	GGG	chr3	239,111,546	1084	5	40.0	1 / 0 / 0 / 6 / 57	89.2	
CaGAD1 (LOC107862809)	down	-	CTGAGAGAAAATGCTACAAC	AGG	chr3	239,111,591	1129	5	40.0	1 / 0 / 0 / 6 / 94	87.4	
CaGAD1 (LOC107862809)	down	-	ACAGCAATGTGTTCAGCATC	TGG	chr3	239,111,696	1234	5	45.0	1 / 0 / 0 / 3 / 56	87.2	
CaGAD1 (LOC107862809)	down	+	AGACATCATGGAGAATTGCC	TGG	chr3	239,111,505	1043	5	45.0	1 / 0 / 1 / 1 / 47	85.1	
CaGAD1 (LOC107862809)	down	+	GATATTGTCTCTAAGGACGT	GGG	chr3	239,111,575	1113	5	40.0	1 / 0 / 0 / 8 / 81	83.6	
CaGAD1 (LOC107862809)	down	+	AGCACCTCAGAAGATTTGGA	TGG	chr3	239,111,660	1198	5	45.0	1 / 0 / 1 / 3 / 96	81.5	
CaGAD1 (LOC107862809)	down	-	TCTGGCGGCATTGTGTATGC	TGG	chr3	239,111,678	1216	5	55.0	1 / 0 / 1 / 6 / 169	78.7	
CaGAD1 (LOC107862809)	down	+	CGATATTGTCTCTAAGGACG	TGG	chr3	239,111,574	1112	5	45.0	1 / 0 / 0 / 15 / 72	77.9	
CaGAD1 (LOC107862809)	down	+	TTGGGTGATATGGCGGAGCA	AGG	chr3	239,110,824	878	3	55.0	3 / 1 / 0 / 1 / 46	69.5	

Gene (locus)	Side of the lysine	Strand	Protospacer	PAM	Chromosome	Genomic cut	Coding cut	Exon	GC (%)	Sites at 0 / 1 / 2 / 3 / 4 mismatches	MIT score	Chosen
CaGAD1 (LOC107862809)	down	+	CTGAACACATTGCTGTTCTG	CGG	chr3	239,111,714	1252	5	45.0	1 / 1 / 0 / 2 / 82	67.3	
CaGAD1 (LOC107862809)	down	-	GCAATGTGTTCAGCATCTGG	CGG	chr3	239,111,693	1231	5	50.0	1 / 0 / 0 / 8 / 171	65.2	
CaGAD1 (LOC107862809)	down	+	TGAACACATTGCTGTTCTGC	GGG	chr3	239,111,715	1253	5	45.0	1 / 1 / 0 / 6 / 26	64.6	
CaGAD1 (LOC107862809)	down	-	ATATGAAAGACGAGCTCATC	AGG	chr3	239,110,844	898	3	40.0	1 / 1 / 0 / 1 / 103	57.9	
CaGAD1 (LOC107862809)	down	+	CAGAGACAGCAGCAAATATA	CGG	chr3	239,111,622	1160	5	40.0	1 / 0 / 4 / 47 / 410	46.4	
CaGAD1 (LOC107862809)	down	+	CTGGTGTCGGTTGGGTGATA	TGG	chr3	239,110,814	868	3	55.0	1 / 2 / 0 / 5 / 68	35.5	
CaGAD1 (LOC107862809)	down	+	GTGTCGGTTGGGTGATATGG	CGG	chr3	239,110,817	871	3	55.0	2 / 2 / 0 / 10 / 210	35.5	
CaGAD1 (LOC107862809)	down	-	AAGTAGGCTGATCAGATCCA	AGG	chr3	239,110,873	927	3	45.0	2 / 2 / 0 / 6 / 98	32.3	
CaGAD4 (LOC107850578)	up	-	ACCATTTCCACAGCTTTCTC	AGG	chr12	2,171,597	580	4	45.0	2 / 0 / 0 / 1 / 39	91.8	
CaGAD4 (LOC107850578)	up	-	CTAAAGTCCCATTCAAGGTC	TGG	chr12	2,173,170	775	5	45.0	1 / 0 / 0 / 7 / 52	87.6	
CaGAD4 (LOC107850578)	up	+	AGCAATTATGCTTGCTGGAT	TGG	chr12	2,170,542	401	3	40.0	1 / 0 / 0 / 3 / 89	86.4	yes
CaGAD4 (LOC107850578)	up	-	CCATTTCCACAGCTTTCTCA	GGG	chr12	2,171,596	579	4	45.0	2 / 0 / 0 / 7 / 56	85.4	
CaGAD4 (LOC107850578)	up	-	GCTTGTCATAGGGCTTTCCT	TGG	chr12	2,170,588	447	3	50.0	1 / 0 / 0 / 3 / 56	83.5	
CaGAD4 (LOC107850578)	up	-	GCAATCTAAAGTCCCATTCA	AGG	chr12	2,173,175	780	5	40.0	1 / 0 / 0 / 7 / 139	81.9	
CaGAD4 (LOC107850578)	up	+	GTGGGACACTCCAATTCATG	TGG	chr12	2,173,120	725	5	50.0	1 / 0 / 1 / 6 / 111	73.9	
CaGAD4 (LOC107850578)	up	+	ATGGGACTTTAGATTGCCAT	TGG	chr12	2,173,192	797	5	40.0	1 / 1 / 3 / 32 / 129	68.3	
CaGAD4 (LOC107850578)	up	+	CCCTGAGAAAGCTGTGGAAA	TGG	chr12	2,171,607	590	4	50.0	2 / 0 / 0 / 12 / 61	64.6	
CaGAD4 (LOC107850578)	up	+	TGAGAAAGCTGTGGAAATGG	TGG	chr12	2,171,610	593	4	45.0	1 / 0 / 0 / 13 / 251	60.6	
CaGAD4 (LOC107850578)	up	+	TCAGAAGCAATTATGCTTGC	TGG	chr12	2,170,537	396	3	40.0	1 / 1 / 1 / 4 / 62	56.4	
CaGAD4 (LOC107850578)	up	+	ATTCATGTGGATGCAGCAAG	TGG	chr12	2,173,133	738	5	45.0	1 / 1 / 1 / 6 / 76	56.0	
CaGAD4 (LOC107850578)	up	-	CTTGCTGCATCCACATGAAT	TGG	chr12	2,173,119	724	5	45.0	1 / 1 / 1 / 6 / 104	52.6	
CaGAD4 (LOC107850578)	up	+	GATGAACCCTGAGAAAGCTG	TGG	chr12	2,171,601	584	4	50.0	1 / 1 / 0 / 1 / 78	46.3	
CaGAD4 (LOC107850578)	up	+	TGTGTTGGGAGAAATTTGCA	AGG	chr12	2,171,528	511	4	40.0	1 / 1 / 2 / 14 / 148	39.8	
CaGAD4 (LOC107850578)	up	+	CATGTGGATGCAGCAAGTGG	TGG	chr12	2,173,136	741	5	55.0	1 / 2 / 0 / 8 / 131	39.5	
CaGAD4 (LOC107850578)	up	-	CTTGTCATAGGGCTTTCCTT	GGG	chr12	2,170,587	446	3	45.0	1 / 0 / 0 / 16 / 316	29.5	
CaGAD4 (LOC107850578)	up	+	GACAAGCCCAATATTGTTAC	TGG	chr12	2,170,615	474	3	40.0	1 / 2 / 5 / 67 / 636	15.3	

Gene (locus)	Side of the lysine	Strand	Protospacer	PAM	Chromosome	Genomic cut	Coding cut	Exon	GC (%)	Sites at 0 / 1 / 2 / 3 / 4 mismatches	MIT score	Chosen
CaGAD4 (LOC107850578)	down	-	CATAACCCAAGCGAATAAGT	TGG	chr12	2,174,775	996	6	40.0	1 / 0 / 1 / 4 / 69	88.3	yes
CaGAD4 (LOC107850578)	down	-	TCTGGTGGCATAGTGTATGC	TGG	chr12	2,175,959	1213	7	50.0	1 / 0 / 1 / 2 / 110	84.8	
CaGAD4 (LOC107850578)	down	+	GAGAAGGACTAGAAAAGACA	GGG	chr12	2,175,827	1081	7	40.0	1 / 0 / 0 / 14 / 193	67.7	
CaGAD4 (LOC107850578)	down	+	AATGCAAGGGTACTGAGAGA	AGG	chr12	2,175,811	1065	7	45.0	1 / 1 / 0 / 11 / 113	59.6	
CaGAD4 (LOC107850578)	down	-	CAGCACCAAGATAATTGATG	TGG	chr12	2,173,307	912	5	40.0	2 / 1 / 1 / 41 / 211	55.6	
CaGAD4 (LOC107850578)	down	-	GTGATGTGTTGAGCATCTGG	TGG	chr12	2,175,974	1228	7	50.0	1 / 0 / 0 / 8 / 312	47.7	
CaGAD4 (LOC107850578)	down	-	ACAGTGATGTGTTGAGCATC	TGG	chr12	2,175,977	1231	7	45.0	1 / 0 / 1 / 13 / 216	47.0	
CaGAD4 (LOC107850578)	down	+	CTGGTATTGGTTGGGTTATT	TGG	chr12	2,173,260	865	5	40.0	3 / 0 / 0 / 28 / 429	43.5	
CaGAD4 (LOC107850578)	down	+	GTATTGGTTGGGTTATTTGG	AGG	chr12	2,173,263	868	5	40.0	1 / 0 / 2 / 34 / 488	40.9	
CaGAD4 (LOC107850578)	down	+	AGAGAAGGACTAGAAAAGAC	AGG	chr12	2,175,826	1080	7	40.0	1 / 3 / 9 / 71 / 473	35.8	

Table S2. GAD expression to three decimals. TPM of the five GAD loci at five stages, the sums used for the share and the share carried by CaGAD1 and CaGAD4. Data: `data/tables/S02.csv` *; script:* `code/04_supplementary.py` *.*

Locus	Name in this paper	Chromosome accession	4 DPA	24 DPA	34 DPA	44 DPA	54 DPA
LOC107862809	CaGAD1	NC_061113.1	0.139	14.267	142.359	260.641	198.173
LOC107853724	CaGAD2	NC_061115.1	0.000	0.000	0.000	0.000	0.000
LOC107838791	CaGAD3	NC_061118.1	0.143	0.678	0.442	0.113	0.203
LOC107850578	CaGAD4	NC_061122.1	150.515	92.009	82.069	67.719	67.918
LOC107877256	GAD-like fragment	NC_061121.1	0.000	0.000	2.454	1.261	2.558
Sum of the four full-length genes			150.797	106.954	224.870	328.473	266.294
CaGAD1 + CaGAD4			150.654	106.276	224.428	328.360	266.091
Share carried by CaGAD1 + CaGAD4 (%)			99.91	99.37	99.80	99.97	99.92

Table S3. Model parameters. Every parameter used by the glutamate balance and the enzyme arithmetic, with its unit and its basis. Data: `data/tables/S03.csv` *; script:* `code/04_supplementary.py` *.*

Parameter	Value	Unit	Basis
glu_molar_mass	147.13	g/mol	definition: molar mass of L-glutamic acid
tissue_density	1.0	g/mL	design assumption: 1 g of tissue is taken as 1 mL
wt_free_glu_low	50	mg/100 g FW	design assumption (no source): low end of the baseline range used in the sensitivity check
wt_free_glu_mid	65	mg/100 g FW	design assumption (no source): baseline used for all headline results; midpoint of 50-80

Parameter	Value	Unit	Basis
wt_free_glu_high	80	mg/100 g FW	design assumption (no source): high end of the baseline range used in the sensitivity check
shunt_share_low	0.10	fraction of total glutamate drain	design assumption (no source): lowest GABA-shunt share of glutamate loss
shunt_share_mid	0.25	fraction of total glutamate drain	midpoint of the 0.10-0.40 range
shunt_share_high	0.40	fraction of total glutamate drain	design assumption (no source): highest GABA-shunt share of glutamate loss
gad_activity_reduction_low	0.70	fraction	design assumption (no source): lowest fraction of shunt flux removed by the double knockout
gad_activity_reduction_mid	0.85	fraction	midpoint of the 0.70-1.00 range
gad_activity_reduction_high	1.00	fraction	bounding case: complete loss of shunt flux
shunt_flux_wt	48.3	umol/gDW/h	design assumption (no source): wild-type GABA-shunt flux; sets the size of the added GDH1 flux relative to the wild-type glutamate supply
gdh1_added_flux_low	80	umol/gDW/h	design assumption (no source): added GDH1 flux, low (0.08 mmol/gDW/h)
gdh1_added_flux_mid	115	umol/gDW/h	midpoint of the 80-150 range
gdh1_added_flux_high	150	umol/gDW/h	design assumption (no source): added GDH1 flux, high (0.15 mmol/gDW/h)
dw_fraction	0.12	gDW per g FW	design assumption: placenta dry-weight fraction; only sets the time scale and the enzyme-arithmetic check, not the steady-state result
tsp_fraction_low	0.005	fraction of total soluble protein	design assumption (no source): AtGDH1 share of soluble protein, low
tsp_fraction_mid	0.01	fraction of total soluble protein	design assumption (no source): AtGDH1 share of soluble protein, mid
tsp_fraction_high	0.02	fraction of total soluble protein	design assumption (no source): AtGDH1 share of soluble protein, high
tsp_total	15	mg/g FW	design assumption (no source): total soluble protein of ripe placenta
gdh1_mw	44.524	kDa	UniProt Q43314 (AtGDH1, 411 aa)
gdh1_kcat	28	1/s per subunit	design assumption (no source): turnover number used for the enzyme arithmetic
gdh1_km_2og	1.8	mM	design assumption (no source): Km for 2-oxoglutarate

Parameter	Value	Unit	Basis
gdh1_km_nh4	2.5	mM	design assumption (no source): Km for ammonium
og2_pool_wt	0.8	mM	design assumption (no source): 2-oxoglutarate pool, wild type
nh4_pool_wt	1.1	mM	design assumption (no source): ammonium pool, wild type
og2_pool_gdh1_oe	0.3	mM	design assumption (no source): 2-oxoglutarate pool, engineered fruit
nh4_pool_gdh1_oe	0.4	mM	design assumption (no source): ammonium pool, engineered fruit
reference_free_glu_level_1	100	mg/100 g FW	reference level for orientation (design assumption, no source)
reference_free_glu_level_2	150	mg/100 g FW	reference level for orientation (design assumption, no source)
reference_free_glu_level_5	250	mg/100 g FW	design target, lower bound
reference_free_glu_level_6	500	mg/100 g FW	design target, upper bound

Table S4. Glutamate balance over the full input grid. The knockout-only, AtGDH1-only and combined results at every combination of the low, mid and high inputs (27 combinations) at the 65 mg/100 g baseline. Data: `data/tables/S04.csv`*; script:* `code/04_supplementary.py`*.*

Shunt share (s)	Knockout fraction (r)	Added flux (umol/gDW/h)	Knockout only (mg/100 g)	AtGDH1 only (mg/100 g)	Knockout + AtGDH1 (mg/100 g)	Knockout + AtGDH1 (mM)	Fold of wild type
0.10	0.70	80	69.9	75.8	81.5	5.5	1.25
0.10	0.70	115	69.9	80.5	86.5	5.9	1.33
0.10	0.70	150	69.9	85.2	91.6	6.2	1.41
0.10	0.85	80	71.0	75.8	82.8	5.6	1.27
0.10	0.85	115	71.0	80.5	88.0	6.0	1.35
0.10	0.85	150	71.0	85.2	93.1	6.3	1.43
0.10	1.00	80	72.2	75.8	84.2	5.7	1.30
0.10	1.00	115	72.2	80.5	89.4	6.1	1.38
0.10	1.00	150	72.2	85.2	94.7	6.4	1.46
0.25	0.70	80	78.8	91.9	111.4	7.6	1.71
0.25	0.70	115	78.8	103.7	125.7	8.5	1.93
0.25	0.70	150	78.8	115.5	140.0	9.5	2.15
0.25	0.85	80	82.5	91.9	116.7	7.9	1.80
0.25	0.85	115	82.5	103.7	131.7	8.9	2.03
0.25	0.85	150	82.5	115.5	146.6	10.0	2.26
0.25	1.00	80	86.7	91.9	122.6	8.3	1.89
0.25	1.00	115	86.7	103.7	138.3	9.4	2.13

Shunt share (s)	Knockout fraction (r)	Added flux (umol/gDW/h)	Knockout only (mg/100 g)	AtGDH1 only (mg/100 g)	Knockout + AtGDH1 (mg/100 g)	Knockout + AtGDH1 (mM)	Fold of wild type
0.25	1.00	150	86.7	115.5	154.0	10.5	2.37
0.40	0.70	80	90.3	108.1	150.1	10.2	2.31
0.40	0.70	115	90.3	126.9	176.3	12.0	2.71
0.40	0.70	150	90.3	145.7	202.4	13.8	3.11
0.40	0.85	80	98.5	108.1	163.7	11.1	2.52
0.40	0.85	115	98.5	126.9	192.3	13.1	2.96
0.40	0.85	150	98.5	145.7	220.8	15.0	3.40
0.40	1.00	80	108.3	108.1	180.1	12.2	2.77
0.40	1.00	115	108.3	126.9	211.5	14.4	3.25
0.40	1.00	150	108.3	145.7	242.9	16.5	3.74

Table S5. Sensitivity ranges. Result of the combined design when one input moves across its range and the others stay at mid-range. Data: `data/tables/S05.csv`*; script:* `code/04_supplementary.py`*.*

Input	Low value	High value	Result at low (mg/100 g)	Result at high (mg/100 g)	Span (mg/100 g)
Shunt share (s)	0.1	0.4	88.0	192.3	104.3
Baseline free glutamate (mg/100 g)	50	80	101.3	162.1	60.8
Added flux (umol/gDW/h)	80	150	116.7	146.6	29.9
Knockout fraction (r)	0.7	1	125.7	138.3	12.6

Table S6. Genome sites with 1 to 3 mismatches. Every site within one to three mismatches of the four chosen guides, with the gene at the site and whether it lies in a coding sequence or an exon. Data: `data/tables/S06.csv`*; script:* `code/04_supplementary.py`*.*

Guide	Chromosome accession	Position	Strand	Mismatches	Mismatch positions	Gene at the site	In coding sequence	In an exon
CaGAD1-g1	NC_061119.1	188628359	+	2	6, 11	LOC124887434	no	yes
CaGAD1-g1	NC_061112.1	141852895	+	3	1, 5, 16	none	no	no
CaGAD1-g1	NC_061116.1	40786874	+	3	1, 5, 16	none	no	no
CaGAD1-g1	NC_061122.1	218795341	-	3	1, 5, 16	none	no	no
CaGAD4-g1	NC_061113.1	225255720	+	3	10, 14, 20	LOC107863165	yes	yes
CaGAD4-g1	NC_061116.1	137684005	-	3	7, 10, 20	none	no	no
CaGAD4-g1	NC_061120.1	215280540	+	3	10, 15, 20	LOC107854594	no	no
CaGAD4-g2	NC_061121.1	165617410	+	2	12, 20	LOC107856685	no	no
CaGAD4-g2	NC_061111.1	129806425	+	3	6, 12, 13	none	no	no
CaGAD4-g2	NC_061114.1	148036994	-	3	2, 7, 13	none	no	no
CaGAD4-g2	NC_061115.1	220201315	+	3	8, 11, 12	LOC107871430	no	no
CaGAD4-g2	NC_061119.1	180946771	-	3	6, 12, 13	none	no	no

Table S7. Exons of CaGAD1 and CaGAD4. Genomic and coding coordinates of every exon, with its length and phase. Data: `data/tables/S07.csv`*; script:* `code/04_supplementary.py`*.*

Gene (locus)	Exon	Start	End	Length (bp)	Coding start	Coding end	Phase
CaGAD1 (LOC107862809)	1	239,108,189	239,108,482	294	1	294	0
CaGAD1 (LOC107862809)	2	239,109,718	239,109,918	201	295	495	0
CaGAD1 (LOC107862809)	3	239,110,442	239,110,910	469	496	964	0
CaGAD1 (LOC107862809)	4	239,111,040	239,111,092	53	965	1017	2
CaGAD1 (LOC107862809)	5	239,111,480	239,111,971	492	1018	1509	0
CaGAD4 (LOC107850578)	1	2,166,377	2,166,462	86	1	86	0
CaGAD4 (LOC107850578)	2	2,167,370	2,167,574	205	87	291	1
CaGAD4 (LOC107850578)	3	2,170,433	2,170,633	201	292	492	0
CaGAD4 (LOC107850578)	4	2,171,510	2,171,724	215	493	707	0
CaGAD4 (LOC107850578)	5	2,173,103	2,173,356	254	708	961	1
CaGAD4 (LOC107850578)	6	2,174,741	2,174,793	53	962	1014	2
CaGAD4 (LOC107850578)	7	2,175,761	2,176,045	285	1015	1299	0
CaGAD4 (LOC107850578)	8	2,177,307	2,177,516	210	1300	1509	0

Table S8. AtGDH1 amount and capacity. AtGDH1 amount for three shares of soluble protein and the capacity of the enzyme under three substrate conditions. Data: `data/tables/S08.csv`*; script:* `code/04_supplementary.py`*.*

AtGDH1 share of soluble protein (%)	AtGDH1 (ug/g fresh weight)	AtGDH1 (nmol/g fresh weight)	Substrate pools	Capacity (umol/gDW/h)
0.5	75.0	1.68	engineered-fruit pools	27.9
0.5	75.0	1.68	wild-type pools	133.0
0.5	75.0	1.68	alpha-ketoglutarate saturating	432.4
1	150.0	3.37	engineered-fruit pools	55.8
1	150.0	3.37	wild-type pools	266.1
1	150.0	3.37	alpha-ketoglutarate saturating	864.7
2	300.0	6.74	engineered-fruit pools	111.5
2	300.0	6.74	wild-type pools	532.1
2	300.0	6.74	alpha-ketoglutarate saturating	1729.4

Table S9. MIT position weights. The twenty position weights of the scoring formula, from the scoring script. Data: `data/tables/S09.csv`*; script:* `code/04_supplementary.py`*.*

Position from the PAM-distal end	Weight
1	0
2	0
3	0.014
4	0
5	0
6	0.395
7	0.317
8	0
9	0.389
10	0.079
11	0.445
12	0.508
13	0.613
14	0.851
15	0.732
16	0.828
17	0.615
18	0.804
19	0.685
20	0.583

Table S10. Scan hits by mismatch number. Genome sites found by the scan for the 78 candidates, by number of mismatches. Data: `data/tables/S10.csv`*; script:* `code/04_supplementary.py`*.*

Mismatches to the guide	Genome sites found by the scan
0	100
1	50
2	88

Mismatches to the guide	Genome sites found by the scan
3	906
4	11184
all	12328

Table S11. Agreement of the two routes of the balance. Comparison of the closed-form balance with the differential-equation version in the research record. Data: `data/tables/S11.csv`*; script:* `code/04_supplementary.py`*.*

Values compared	Largest relative difference	All within 1%
342	3.82e-10	yes

Supplementary data: file map

Path	Contents
`data/tables/T01.csv`	Table 1: GAD loci in placenta
`data/tables/T02.csv`	Table 2: The four designed guides
`data/tables/T03.csv`	Table 3: Inputs of the glutamate balance
`data/tables/T04.csv`	Table 4: Modelled free glutamate
`data/tables/T05.csv`	Table 5: Concentration comparison
`data/tables/T06.csv`	Table 6: Dependence on the baseline
`data/tables/S01.csv`	Table S1: All 78 candidate guides
`data/tables/S02.csv`	Table S2: GAD expression to three decimals
`data/tables/S03.csv`	Table S3: Model parameters
`data/tables/S04.csv`	Table S4: Glutamate balance over the full input grid
`data/tables/S05.csv`	Table S5: Sensitivity ranges
`data/tables/S06.csv`	Table S6: Genome sites with 1 to 3 mismatches
`data/tables/S07.csv`	Table S7: Exons of CaGAD1 and CaGAD4
`data/tables/S08.csv`	Table S8: AtGDH1 amount and capacity
`data/tables/S09.csv`	Table S9: MIT position weights
`data/tables/S10.csv`	Table S10: Scan hits by mismatch number
`data/tables/S11.csv`	Table S11: Agreement of the two routes of the balance
`data/figures/GA.csv`	Data behind the graphical abstract
`data/figures/F01.csv`	Data behind Figure 1
`data/figures/F02.csv`	Data behind Figure 2
`data/figures/F03.csv`	Data behind Figure 3
`data/figures/F04a.csv`	Data behind Figure 4
`data/figures/F04b.csv`	Data behind Figure 4
`data/figures/F05.csv`	Data behind Figure 5
`data/figures/F06.csv`	Data behind Figure 6
`data/results/model_results.csv`	Modelled free glutamate per genotype: lowest, mid-range and highest value (mg/100 g, mM, fold of wild type; rounded columns as printed)

Path	Contents
`data/results/model_required_flux.csv`	Added AtGDH1 flux needed to reach 250 mg/100 g in two input cases
`data/results/baseline_dependence.csv`	Combined-design range for baselines of 50, 65 and 80 mg/100 g
`data/results/sensitivity_combined.csv`	Combined-design result when one input moves over its range and the others stay at mid-range
`data/results/concentration_comparison.csv`	Free glutamate in mg/100 g and mM and the ratio to the reference EC50
`data/results/guide_table.csv`	The four designed guides with sequence, PAM, strand, cut position, exon, GC, MIT score and site counts
`data/results/guide_deletions.csv`	Genomic and coding bases between the two cuts of each gene, share of the coding sequence and reading-frame check
`data/results/guide_candidates_per_side.csv`	Candidate guides per gene and side, the highest-scoring guide with one perfect site and the chosen guide
`data/results/gad_share.csv`	Share of GAD transcript carried by CaGAD1 and CaGAD4 per stage
`data/figures/*`	Data behind the figures (one CSV per figure or panel)
`data/results/scan_summary.csv`	Genome sites found by the scan by mismatch number, and the number of guides scanned
`data/results/enzyme_capacity.csv`	AtGDH1 amount and capacity for three soluble-protein shares and three substrate conditions
`data/results/route_agreement.csv`	Agreement of the closed-form balance with the differential-equation version
`data/results/offtarget_sites.csv`	Genome sites with 1 to 3 mismatches to the four guides, annotated with gene and coding status
`data/wet_lab_active.csv`	Experiments active in the authors' laboratories for this paper: register identifier (WL-01 to WL-84), experiment name and stage code (S0 paperwork, S1 calibration and cheap decisive tests, S2 transient and in vitro mechanism tests, S3 stable lines, S4 tests on engineered material, S5 optional and long-lead)
`inputs/model_inputs.csv`	Every input of the glutamate balance with unit and kind
`inputs/ifi_reference_values.csv`	Umami reference axis: L-glutamate, TAS1R1/TAS1R3, EC50 2,500 micromolar
`inputs/gad_tpm_by_stage.tsv`	Placenta TPM of the five GAD loci at five stages
`inputs/gene_names.csv`	Names used in this paper with RefSeq LOC IDs, chromosomes and products
`inputs/input_sources.csv`	Public inputs (genome, RNA-seq series, UniProt entries)
`inputs/guide_design/scored_candidates.csv`	All 78 candidate guides with site counts at 0 to 4 mismatches and MIT score
`inputs/guide_design/gad_models.json`	Exon coordinates, protein sequences and PLP-lysine residues of CaGAD1 and CaGAD4
`inputs/guide_design/hits.tsv`	Every genome site with up to four mismatches to a candidate (stored scan output)
`inputs/guide_design/P54767.fasta`	Tomato glutamate decarboxylase P54767 (reference for the lysine position)
`inputs/guide_design/selected_guides.csv`	The four chosen guides
`code/01_pool_model.py`	Glutamate balance and concentration comparison
`code/02_guides.py`	Guide table, deletions and gene maps
`code/03_expression.py`	Expression table
`code/05_full_guide_scan.py`	Full-mode guide design from the genome

Path	Contents
`code/04_supplementary.py`	Supplementary tables and figure data
`code/guide_design/scan.c`	C program: scans the whole genome for sites with up to four mismatches to each candidate guide (strict NGG PAM, both strands)
`code/guide_design/step1_models.py`	Gene models of CaGAD1 and CaGAD4; PLP lysine located by alignment to tomato GAD (P54767)
`code/guide_design/step2_candidates.py`	Lists every SpCas9 site in the coding exons
`code/guide_design/step3_select.py`	Filters candidates (window 30-450 coding bases from the lysine, GC 40-70%, no TTTT or five-base run)
`code/guide_design/step4_score.py`	Counts sites by mismatch number and computes the MIT specificity score (Hsu et al. 2013)
`code/heavy/NEEDS.txt`	What full mode needs
`research_record/scripts/run_model.py`	Differential-equation route of the glutamate balance (agrees with the closed form)
`research_record/scripts/check_model.py`	Closed-form route and value-by-value agreement check
`research_record/scripts/p05_gad_loci.py`	Lists the GAD loci with chromosome, protein length and coding length
`research_record/scripts/p05_tpm_check.py`	Sums TPM per GAD gene from the per-transcript quantification files
`research_record/scripts/*_output.txt`	Output of the scripts above
`research_record/outputs/*`	Raw outputs of the glutamate balance
`research_record/second_script_rerun/*`	Independent second implementation of the guide steps, scan, scoring and model arithmetic, with its outputs

Cite this paper

Ford, P.I. (2026). Umami-Enriched Capsicum: In Silico Design of Glutamate Accumulation by Glutamate Decarboxylase Knockout and Glutamate Dehydrogenase Expression. Scoville Splice Research Series. https://doi.org/10.5281/zenodo.24511324

```
@misc{Ford2026_P05,
  author = {Ford, P.I.},
  title = {Umami-Enriched Capsicum: In Silico Design of Glutamate Accumulation by Glutamate Decarboxylase Knockout and Glutamate Dehydrogenase Expression},
  year = {2026},
  publisher = {Scoville Splice LLC},
  series = {Scoville Splice Research Series},
  doi = {10.5281/zenodo.24511324},
  url = {https://doi.org/10.5281/zenodo.24511324}
}
```

Also in `CITATION.cff`, `citation.bib` and `citation.ris` in the data package.

Related papers in this series

ID	Title	DOI
P04	The Iyengar Flavor Index (IFI): A Receptor-Normalized Standard for Cross-Modal Flavor Measurement	10.5281/zenodo.24511320
P11	The Capsaicin Design Platform: Architecture, Sixteen Design Tools and the Flavor Index Framework of a Subscription Service	10.5281/zenodo.24511346
P12	Trait Expansion in Capsicum: CRISPR Guide Design for Nine Trait Groups with Gene Assignment and Genome-Wide Off-Target Profiling	10.5281/zenodo.24511350

Appendix A. Evidence table

Claim ID	Claim	Value	Kind	Source file	Location
P05-C001	The reference genome is Capsicum annuum UCD10Xv1.1 (RefSeq assembly GCF_002878395.1)	GCF_002878395.1	definition	`inputs/input_sources.csv`	row genome
P05-C002	The placenta RNA-seq series is BioProject PRJNA789050 (C. chinense SL08)	PRJNA789050	definition	`inputs/input_sources.csv`	row RNA-seq series
P05-C003	The series has five placenta stages: 4, 24, 34, 44 and 54 days post anthesis, one library per stage	4, 24, 34, 44 and 54	definition	`inputs/gad_tpm_by_stage.tsv`	column headers
P05-C004	CaGAD1 (LOC107862809) is on chromosome 3	3	computed	`inputs/gene_names.csv`	row LOC107862809
P05-C005	CaGAD2 (LOC107853724) is on chromosome 5	5	computed	`inputs/gene_names.csv`	row LOC107853724
P05-C006	CaGAD3 (LOC107838791) is on chromosome 8	8	computed	`inputs/gene_names.csv`	row LOC107838791
P05-C007	CaGAD4 (LOC107850578) is on chromosome 12	12	computed	`inputs/gene_names.csv`	row LOC107850578
P05-C008	The GAD-like fragment LOC107877256 is on chromosome 11	11	computed	`inputs/gene_names.csv`	row LOC107877256
P05-C009	The GAD-like fragment has 116 residues	116	computed	`research_record/scripts/p05_gad_loci_output.txt`	row LOC107877256
P05-C010	CaGAD1 (LOC107862809) is at 0.1 TPM at 4 DPA	0.1	computed	`data/tables/T01.csv`	row CaGAD1, column 4 DPA
P05-C011	CaGAD1 (LOC107862809) is at 142.4 to 260.6 TPM at 34 to 54 DPA	142.4 to 260.6	computed	`data/tables/T01.csv`	row CaGAD1, columns 34 to 54 DPA
P05-C012	CaGAD4 (LOC107850578) is at 150.5 TPM at 4 DPA	150.5	computed	`data/tables/T01.csv`	row CaGAD4, column 4 DPA
P05-C013	CaGAD4 (LOC107850578) is at 67.7 to 82.1 TPM at 34 to 54 DPA	67.7 to 82.1	computed	`data/tables/T01.csv`	row CaGAD4, columns 34 to 54 DPA
P05-C014	CaGAD2 (LOC107853724) and CaGAD3 (LOC107838791) are at 0.0 to 0.7 TPM at every stage	0.0 to 0.7	computed	`data/tables/T01.csv`	rows CaGAD2 and CaGAD3

Claim ID	Claim	Value	Kind	Source file	Location
P05-C015	The GAD-like fragment is at 0.0 to 2.6 TPM	0.0 to 2.6	computed	`data/tables/T01.csv`	row GAD-like fragment
P05-C016	CaGAD1 and CaGAD4 carry 99.37 to 99.97% of the transcript of the four full-length GAD genes at every stage	99.37 to 99.97%	computed	`data/results/gad_share.csv`	column percent_carried_by_CaGAD1_plus_CaGAD4
P05-C017	CaGAD3 (LOC107838791) is at most 0.7 TPM in placenta	0.7	computed	`data/tables/T01.csv`	row CaGAD3
P05-C018	The PLP-binding lysine is residue 278 in CaGAD1 (LOC107862809), by alignment to tomato GAD (UniProt P54767)	278	computed	`inputs/guide_design/gad_models.json`	CaGAD1, plp_lys
P05-C019	The PLP-binding lysine is residue 277 in CaGAD4 (LOC107850578)	277	computed	`inputs/guide_design/gad_models.json`	CaGAD4, plp_lys
P05-C020	Each coding sequence has 1,509 bases	1,509	computed	`data/results/guide_deletions.csv`	column cds_bp
P05-C021	78 candidate guides in the coding exons pass the window, GC and run filters	78	computed	`data/results/guide_candidates_per_side.csv`	row all, column candidates
P05-C022	The window is 30 to 450 coding bases either side of the lysine codon, GC content 40 to 70%, no TTTT and no five-base run	30 to 450 coding bases	definition	`code/guide_design/step3_select.py`	line with k-450<=cc<=k-30
P05-C023	Candidate GC content is 40 to 70%	40 to 70%	definition	`code/guide_design/step3_select.py`	line with 40<=c['gc']<=70
P05-C024	The four guides have MIT scores of 86.4 to 97.4	86.4 to 97.4	computed	`data/results/guide_table.csv`	column mit_specificity
P05-C025	Each of the four guides has 1 perfect site in the genome	1	computed	`data/results/guide_table.csv`	column perfect_sites
P05-C026	The deletion between the two CaGAD1 cuts is 992 bp of genome and 476 coding bases (31.5% of the coding sequence)	992 bp, 476 coding bases (31.5%)	computed	`data/results/guide_deletions.csv`	row CaGAD1
P05-C027	The deletion between the two CaGAD4 cuts is 4,233 bp of genome and 595 coding bases (39.4% of the coding sequence)	4,233 bp, 595 coding bases (39.4%)	computed	`data/results/guide_deletions.csv`	row CaGAD4
P05-C028	The CaGAD1 coding deletion is 476 bases	476	computed	`data/results/guide_deletions.csv`	row CaGAD1, column coding_bp_between_cuts
P05-C029	The CaGAD4 coding deletion is 595 bases	595	computed	`data/results/guide_deletions.csv`	row CaGAD4, column coding_bp_between_cuts
P05-C030	The coding deletions leave remainders of 2 and 1 after division by three, so each shifts the reading frame	2 and 1	computed	`data/results/guide_deletions.csv`	column coding_bp_mod_3
P05-C031	Chosen CaGAD4 upstream guide: MIT 86.4 and 595 coding bases deleted; highest-scoring upstream guide: MIT 87.6 and 221 coding bases deleted	86.4 and 595; 87.6 and 221	computed	`data/results/guide_candidates_per_side.csv`	row CaGAD4 up

Claim ID	Claim	Value	Kind	Source file	Location
P05-C032	Of the 12 genome sites with 1 to 3 mismatches to the four guides, 1 lies in a coding sequence (LOC107863165, 3 mismatches to CaGAD4-g1)	1 of 12	computed	`research_record/second_script_rerun/p05_step2_output.txt`	section Sites with 1-3 mismatches
P05-C033	No guide has a site within 4 mismatches inside any GAD locus other than its own target	4	computed	`research_record/second_script_rerun/p05_gad_cross_output.txt`	all rows
P05-C034	The assumed baseline is 65 mg/100 g, the midpoint of 50 to 80	65 mg/100 g (midpoint of 50 to 80)	design assumption	`inputs/model_inputs.csv`	rows wt_free_glu_low, mid, high
P05-C035	The assumed baseline is 65	65	design assumption	`inputs/model_inputs.csv`	row wt_free_glu_mid
P05-C036	The GABA-shunt share of glutamate loss is assumed 0.10 to 0.40	0.10 to 0.40	design assumption	`inputs/model_inputs.csv`	rows shunt_share_low, high
P05-C037	The knockout is assumed to remove 0.70 to 1.00 of shunt activity	0.70 to 1.00	design assumption	`inputs/model_inputs.csv`	rows gad_activity_reduction_low, high
P05-C038	The wild-type shunt flux is assumed 48.3 umol/gDW/h	48.3	design assumption	`inputs/model_inputs.csv`	row shunt_flux_wt
P05-C039	The glutamate synthesis flux added by AtGDH1 is assumed 80 to 150 umol/gDW/h	80 to 150	design assumption	`inputs/model_inputs.csv`	rows gdh1_added_flux_low, high
P05-C040	The design target is 250 to 500 mg/100 g free glutamate	250 to 500 mg/100 g	design target	`inputs/model_inputs.csv`	rows reference_free_glu_level_5, 6
P05-C041	The lower bound of the design target is 250	250	design target	`inputs/model_inputs.csv`	row reference_free_glu_level_5
P05-C042	Knockout only: 70 to 108 mg/100 g	70 to 108 mg/100 g	computed	`data/results/model_results.csv`	row GAD knockout only, columns min_rounded, max_rounded
P05-C043	AtGDH1 only: 76 to 146 mg/100 g	76 to 146 mg/100 g	computed	`data/results/model_results.csv`	row AtGDH1 only, columns min_rounded, max_rounded
P05-C044	Knockout + AtGDH1: 81 to 243 mg/100 g	81 to 243 mg/100 g	computed	`data/results/model_results.csv`	row Knockout + AtGDH1, columns min_rounded, max_rounded
P05-C045	The highest modelled value of the combined design is 243	243	computed	`data/results/model_results.csv`	row Knockout + AtGDH1, column max_rounded
P05-C046	The mid-range value of the combined design is 132	132	computed	`data/results/model_results.csv`	row Knockout + AtGDH1, column mid_rounded
P05-C047	The combined design is 1.25 to 3.74 times the wild-type baseline	1.25 to 3.74	computed	`data/results/model_results.csv`	row Knockout + AtGDH1, fold columns
P05-C048	Knockout only is 1.08 to 1.67 times the baseline	1.08 to 1.67	computed	`data/results/model_results.csv`	row GAD knockout only, fold columns
P05-C049	Reaching 250 mg/100 g needs 158 umol/gDW/h of added flux with complete knockout at the highest shunt share	158	computed	`data/results/model_required_flux.csv`	row complete knockout at the highest shunt share
P05-C050	At mid-range shunt share and knockout, reaching 250 needs 392 umol/gDW/h of added flux	392	computed	`data/results/model_required_flux.csv`	row mid-range shunt share and knockout

Claim ID	Claim	Value	Kind	Source file	Location
P05-C051	Moving the shunt share over its range moves the mid-range combined result from 88.0 to 192.3 mg/100 g	88.0 to 192.3	computed	`data/results/sensitivity_combined.csv`	row Shunt share (s)
P05-C052	Moving the knockout fraction over its range moves the result from 125.7 to 138.3 mg/100 g	125.7 to 138.3	computed	`data/results/sensitivity_combined.csv`	row Knockout fraction (r)
P05-C053	The mid-range result used as the centre of the sensitivity analysis is 131.7 mg/100 g	131.7	computed	`data/results/sensitivity_combined.csv`	column baseline_result
P05-C054	A baseline of 50 mg/100 g gives 63 to 187 for the combined design	63 to 187	computed	`data/results/baseline_dependence.csv`	row 50
P05-C055	A baseline of 80 mg/100 g gives 100 to 299 for the combined design	100 to 299	computed	`data/results/baseline_dependence.csv`	row 80
P05-C056	A baseline of 50 mg/100 g	50	design assumption	`data/results/baseline_dependence.csv`	row 50
P05-C057	A baseline of 80 mg/100 g	80	design assumption	`data/results/baseline_dependence.csv`	row 80
P05-C058	The molar mass of L-glutamic acid is 147.13 g/mol	147.13	definition	`inputs/model_inputs.csv`	row glu_molar_mass
P05-C059	The IFI-U reference EC50 of L-glutamate at TAS1R1/TAS1R3 is 2,500 micromolar	2,500	definition	`inputs/ifi_reference_values.csv`	row IFI-U
P05-C060	The reference EC50 is 2.5 mM	2.5	definition	`data/results/concentration_comparison.csv`	column reference_EC50_mM
P05-C061	The 65 mg/100 g baseline is 4.4 mM, a ratio of 1.8 to the reference EC50	4.4 mM, ratio 1.8	computed	`data/results/concentration_comparison.csv`	row Baseline (assumed)
P05-C062	Knockout only is 4.8 to 7.4 mM, a ratio of 1.9 to 2.9	4.8 to 7.4 mM, ratio 1.9 to 2.9	computed	`data/results/concentration_comparison.csv`	row Knockout only
P05-C063	The combined design is 5.5 to 16.5 mM, a ratio of 2.2 to 6.6	5.5 to 16.5 mM, ratio 2.2 to 6.6	computed	`data/results/concentration_comparison.csv`	row Knockout + AtGDH1
P05-C064	The design target is 17.0 to 34.0 mM, a ratio of 6.8 to 13.6	17.0 to 34.0 mM, ratio 6.8 to 13.6	computed	`data/results/concentration_comparison.csv`	rows Design target
P05-C065	Reference levels of 100 and 150 mg/100 g are 6.8 and 10.2 mM	100 and 150 mg/100 g (6.8 and 10.2 mM)	design assumption	`data/results/concentration_comparison.csv`	rows Reference level
P05-C066	AtGDH1 is At5g18170, UniProt Q43314, 411 residues, 44.5 kDa, EC 1.4.1.3	At5g18170, Q43314, 411 residues, 44.5 kDa, EC 1.4.1.3	literature value	`literature`	UniProt Q43314 entry; Labboun et al. 2009, doi:10.1093/pcp/pcp118
P05-C067	Scripts run on Python 3.9 or newer	3.9	definition	`code/LIBRARIES.txt`	line 1
P05-C068	The differential-equation version uses numpy 2.0.2 and scipy 1.13.1	numpy 2.0.2 and scipy 1.13.1	definition	`code/LIBRARIES.txt`	line 2
P05-C069	The LSODA integration uses relative tolerance 1e-11 and absolute tolerance 1e-13	1e-11 and 1e-13	definition	`research_record/scripts/run_model.py`	solve_ivp call

Claim ID	Claim	Value	Kind	Source file	Location
P05-C070	The lysine alignment is global pairwise with BLOSUM62, gap open -10 and gap extend -0.5	BLOSUM62, gap open -10, gap extend -0.5	definition	`code/guide_design/step1_models.py`	PairwiseAligner settings
P05-C071	The MIT score uses 20 position weights from 0 to 0.851	20 weights, 0 to 0.851	definition	`data/tables/S09.csv`	column Weight
P05-C072	The scan found 12,328 genome sites within up to four mismatches of the 78 candidates	12,328	computed	`data/results/scan_summary.csv`	row all
P05-C073	The scan sites by mismatch number are 100, 50, 88, 906 and 11,184 for 0 to 4 mismatches	100, 50, 88, 906 and 11,184	computed	`data/results/scan_summary.csv`	rows 0 to 4
P05-C074	The scan covered 78 guides	78	computed	`data/results/scan_summary.csv`	row guides_scanned
P05-C075	60 of the 78 candidates have exactly one perfect genome site	60 of 78	computed	`data/results/guide_candidates_per_side.csv`	row all, column candidates_with_one_perfect_site
P05-C076	The 78 candidates have MIT scores of 14.5 to 99.0	14.5 to 99.0	computed	`data/figures/F05.csv`	column mit_score
P05-C077	CaGAD1 has 5 coding exons and CaGAD4 has 6	5 and 6	computed	`data/tables/S07.csv`	rows per gene
P05-C078	AtGDH1 is assumed to be 0.005 to 0.02 of total soluble protein	0.005 to 0.02	design assumption	`inputs/model_inputs.csv`	rows tsp_fraction_low, high
P05-C079	Total soluble protein of ripe placenta is assumed 15 mg/g fresh weight	15	design assumption	`inputs/model_inputs.csv`	row tsp_total
P05-C080	The assumed turnover number is 28 per second per subunit	28	design assumption	`inputs/model_inputs.csv`	row gdh1_kcat
P05-C081	The assumed Km values are 1.8 mM for alpha-ketoglutarate and 2.5 mM for ammonium	1.8 and 2.5	design assumption	`inputs/model_inputs.csv`	rows gdh1_km_2og, gdh1_km_nh4
P05-C082	The assumed substrate pools are 0.3 and 0.4 mM in engineered fruit and 0.8 and 1.1 mM in wild type	0.3 and 0.4; 0.8 and 1.1	design assumption	`inputs/model_inputs.csv`	rows og2_pool, nh4_pool
P05-C083	The placenta dry-weight fraction is assumed 0.12	0.12	design assumption	`inputs/model_inputs.csv`	row dw_fraction
P05-C084	At 1% of soluble protein AtGDH1 is 150 ug/g and 3.37 nmol/g fresh weight	150 ug/g and 3.37 nmol/g	computed	`data/results/enzyme_capacity.csv`	row tsp_percent 1
P05-C085	At engineered-fruit pools the capacity is 27.9 to 111.5 umol/gDW/h for 0.5 to 2% of soluble protein	27.9 to 111.5	computed	`data/results/enzyme_capacity.csv`	rows engineered-fruit pools
P05-C086	At wild-type pools the capacity is 133.0 to 532.1 umol/gDW/h	133.0 to 532.1	computed	`data/results/enzyme_capacity.csv`	rows wild-type pools
P05-C087	With alpha-ketoglutarate saturating and 1% of soluble protein the capacity is 864.7 umol/gDW/h	864.7	computed	`data/results/enzyme_capacity.csv`	row tsp_percent 1, saturating
P05-C088	The two routes of the balance agree for 342 values, with a largest relative difference of 3.82e-10	342 values, 3.82e-10	computed	`data/results/route_agreement.csv`	row 1
P05-C089	Moving the baseline over 50 to 80 moves the mid-range combined result from 101.3 to 162.1	101.3 to 162.1	computed	`data/results/sensitivity_combined.csv`	row Baseline free glutamate

Claim ID	Claim	Value	Kind	Source file	Location
P05-C090	Moving the added flux over its range moves the result from 116.7 to 146.6	116.7 to 146.6	computed	`data/results/sensitivity_combined.csv`	row AtGDH1 added flux
P05-C091	At a 0.40 shunt share and mid-range knockout the combined result is 163.7 to 220.8 for added fluxes of 80 to 150	163.7 to 220.8	computed	`data/figures/F06.csv`	row 0.40

Appendix B. Glossary, abbreviations and units

Term	Meaning	Type
GAD	Glutamate decarboxylase: the enzyme that turns glutamate into GABA (gamma-aminobutyric acid)	abbreviation
GABA shunt	The route from glutamate through GABA that drains free glutamate	term
GDH	Glutamate dehydrogenase: an enzyme that interconverts glutamate and alpha-ketoglutarate with ammonium	abbreviation
PLP	Pyridoxal 5'-phosphate: the cofactor of GAD, bound by a lysine	abbreviation
DPA	Days post anthesis: days after the flower opens	unit
TPM	Transcripts per million: a normalised measure of transcript abundance	unit
MIT score	Hsu et al. (2013) aggregate specificity score of a guide: higher means fewer likely off-target sites	term
mg/100 g	Milligrams per 100 grams of fresh weight	unit
mM	Millimolar: millimoles per litre (1 g of tissue taken as 1 mL)	unit
umol/gDW/h	Micromoles per gram dry weight per hour: the unit of metabolic flux in the balance	unit
EC50	Concentration giving half of the maximum response of a receptor assay	abbreviation
IFI-U	Umami axis of the Iyengar Flavor Index: L-glutamate at the TAS1R1/TAS1R3 receptor	term
PAM	Protospacer-adjacent motif: the NGG sequence next to the target that SpCas9 needs	abbreviation
Protospacer	The 20-nucleotide target sequence of a guide RNA	term
SpCas9	Streptococcus pyogenes Cas9 nuclease	abbreviation
TSP	Total soluble protein of the tissue	abbreviation
kcat	Turnover number: reactions per enzyme subunit per second	term
Km	Substrate concentration at half the maximum rate of an enzyme	term

Term	Meaning	Type
BLOSUM62	Amino-acid substitution matrix used for the alignment	term
LSODA	Numerical integrator for differential equations	term

Appendix C. Computational environment

Item	Value
Python	3.15.2
Platform	Darwin x86_64
Libraries	standard library only
Steps and run time	1 code/01_pool_model.py: 0.1 s; 2 code/02_guides.py: 0.1 s; 3 code/03_expression.py: 0.1 s; 4 code/04_supplementary.py: 0.1 s
Input files	11 files, SHA-256 in `code/environment.json`
Data fingerprint	`4570e79d8c5eefdea5337ba78e4319c7e01f309498237929d50fba076e3e74b8`

LEGAL NOTICES

Research use only. All results in this paper are computed designs, predictions and targets. They are not food, a dietary supplement, a drug or a consumer product, and they are not medical, dietary, agronomic, regulatory or financial advice. Concentrated capsaicinoids can cause severe pain and injury. Do not expose any person or animal to material derived from these designs.

No warranty; limitation of liability. This paper, its data and its code are provided "as is" without warranty of any kind. To the fullest extent permitted by law, Scoville Splice LLC and the authors are not liable for any claim, loss or damage arising from their use. Users are responsible for complying with all biosafety, biosecurity, plant-health, food-safety, import and export, and other laws that apply to them, and for obtaining every approval needed before any laboratory, greenhouse or field work.

Licence. The paper and the data are licensed under Creative Commons Attribution 4.0 International (CC BY 4.0); the code is licensed under the MIT licence. These licences cover copyright in the files only and grant no right under any patent, patent application, trade secret or trademark. See `legal/DISCLAIMER.txt`, `legal/TERMS_OF_USE.txt` and `legal/NOTICE.txt` in the data package.

DOI 10.5281/zenodo.24511324 · US provisional applications 64/172,827, 64/188,197, 64/192,062 and 64/199,105 · © 2026 Scoville Splice LLC and P.I. Ford · Published under CC BY 4.0 · scovillesplice.farm · ORCID: 0000-0003-0395-1752