

THE SCOVILLE SPLICE LIBRARY · VOLUME LIX

The Pathway Book

Gene to flavor molecule. *Every route, drawn.*

The complete biosynthetic atlas. One phenylpropanoid trunk, four terminal enzymes that make a pepper a pepper, and the eight branches that carry every axis of taste — each route drawn enzyme by enzyme, intermediate by intermediate, from the base we edit to the molecule the tongue finally reads.

SCOVILLE SPLICE · GROWN BY OUR FAMILY. ENGINEERED FROM SCRATCH.

Contents

01 Opening essay · Follow the carbon

02 Chapter 1 · The Phenylpropanoid Trunk — the backbone every flavor is cut from.

03 Chapter 2 · The Grammar of a Pathway — how a chain of reactions becomes a single number.

04 Chapter 3 · The Four Terminal Enzymes — what, exactly, makes a pepper.

05 Chapter 4 · The Capsaicinoid Route — vanillylamine, an acyl chain, and the bond between them.

06 Chapter 5 · The Branched-Chain Supply — why the fat decides the shape of the burn.

07 Chapter 6 · Designing the Catalyst — ESM2, flux balance, and two edits no breeder ever found.

08 Chapter 7 · The Last Six Ångströms — capsaicin into TRPV1, and the EC_{50} that becomes a score.

09 Chapter 8 · Cool-Burn — the five-gene climb to menthol.

10 Chapter 9 · Pungency — two routes to the same sharp channel.

11 Chapter 10 · Umami — teaching a fruit to keep its glutamate.

12 Chapter 11 · Kokumi — the peptides you taste without tasting.

13 Chapter 12 · Sweetness — diterpene glycosides and the sugar that isn't.

14 Chapter 13 · Astringency — the tannin route, drawn to the polymer.

15 Chapter 14 · Zero-Bitter — the one axis we build by taking a step away.

16 Chapter 15 · Flux Balance as a Design Language — the arithmetic of a shared precursor.

17 Chapter 16 · Eight Branches, One Plant — stacking routes on one carbon budget.

18 Chapter 17 · The Transferable Pathway — five modules of seven, and the climb to sixty-four.

19 Chapter 18 · The Open Record — provenance, consistency, and the respect of a reproducible number.

Opening essay · Follow the carbon

There is a moment, a few seconds after you bite into something our family has grown, when the flavor has already happened in your mouth and not yet happened in your mind. The molecule has landed. The receptor has opened. A current is already climbing a nerve toward the part of you that will, in a heartbeat, decide that this is *good*. For that one suspended instant the taste is pure chemistry and no opinion at all. This whole book lives inside that instant — and then walks backward out of it, step by step, all the way to a gene.

Every flavor you have ever loved is the far end of a chain of reactions. Not a metaphor: an actual, countable sequence of enzymes, each one accepting a molecule from the enzyme before it, bending a single bond, and handing the result on. A tomato's savory depth is the last link in a chain that began with a nitrogen atom pulled from the soil. The heat of a chili is two chains welded together — chains that began in different districts of the cell and met, at the very end, over a single acyltransferase. The cool of mint is a short ring of carbon worked through a handful of oxidations and reductions. We did not invent any of these chains. Plants wrote them over four hundred million years. What this family has ever done — one pair of hands at the keyboard, the whole family around the table — is learn to read them, letter by letter, and then write, carefully, in the margins.



The research greenhouse at first light — the walk where the question that became this book first got asked out loud: not *how hot*, but *how, exactly*.

This book gives you the road a molecule travels to reach your tongue — and the reason that road, more than any other single thing, is what we design. You cannot make a plant louder by wishing. You make it louder by finding the one enzyme in a chain of a dozen that is holding everyone else up, and relieving it. You cannot conjure a flavor a plant has never made; you can only carry a route — a set of genes from the organism that knows the chemistry into the organism that does not — and then balance the books so the host can afford

to run it. The reverence and the arithmetic are the same act. We revere the plant *because* we have counted what it does; the counting is a form of love, and it is also the only thing that works.

One trunk, eight branches. Almost every flavor molecule we build comes off a single shared pipeline — the phenylpropanoid pathway — that runs in every vascular plant on Earth, from moss to oak to the pepper on the counter. It is the pipeline that makes lignin, which is to say it is the pipeline that let plants stand up out of the water and onto dry land. Hang different terminal enzymes off its end and you get vanillin, or the burning head of capsaicin, or the tannins that dry the mouth after red wine. The trunk is universal. Only the last few enzymes are special — and *special*, in biochemistry, means *movable*. Follow the carbon; that is the only instruction.

One trunk, and every plant already runs it

Start where the plant starts. Deep in primary metabolism, the shikimate pathway takes carbon the leaf fixed from the air and nitrogen the root pulled from the soil and assembles the aromatic amino acid L-phenylalanine — a six-carbon ring with a short tail. Phenylalanine is the doorway. Everything this book is about begins the instant an enzyme decides to spend one.

That enzyme is **PAL**, phenylalanine ammonia-lyase, and it does something quietly violent: it tears the amino group clean off, releasing ammonia and leaving *trans*-cinnamic acid, a ring now carrying a three-carbon arm with a double bond in it. This is the committed step — the moment carbon leaves the business of building the plant's body and enters the business of building its chemistry. From here the trunk runs five enzymes deep, and it runs, with minor dialect, in moss and in oak and in the pepper alike.

STEP	ENZYME	REACTION	COFACTOR	SCOPE
1	PAL	L-phenylalanine → *trans*-cinnamic acid + NH_3	—	universal
2	C4H	cinnamic acid → *p*-coumaric acid	NADPH, O_2 (a P450)	universal
3	4CL	*p*-coumaric acid → *p*-coumaroyl-CoA	ATP → AMP + PP_i	universal
4	HCT	coumaroyl transfer onto shikimate (→ 3-hydroxylation)	CoA	universal
5	COMT	caffeoyl → feruloyl (O-methylation)	SAM	universal

Read the cofactor column, because it is where the plant pays. **C4H** is a cytochrome P450 bolted into a membrane; to set a single hydroxyl on the ring it burns NADPH and a molecule of oxygen. **4CL** spends a whole ATP to dress *p*-coumaric acid in a coenzyme-A handle, turning a sluggish acid into *p*-coumaroyl-CoA — the activated, reactive currency that every downstream branch draws on. **COMT** spends S-adenosylmethionine, the cell's universal methyl donor, to clip a single carbon onto an oxygen. Five steps, three different energy currencies, and at the end a small, activated, modular molecule — feruloyl-CoA — sitting at a fork. Send it one way and it polymerizes into lignin and the plant stands upright. Send it another and it becomes the fragrant, the bitter, the hot. The trunk does not care which. That indifference is the whole opportunity.

The weld — where two chains become capsaicin

To make the heat of a chili, the trunk is only half the story. The pepper runs a second assembly line entirely — a fatty-acid line — and the flavor exists only because the two lines are soldered together at the end.

Down the phenylpropanoid branch, feruloyl-CoA is walked forward to vanillin and then handed to **pAMT**, a pyridoxal-phosphate aminotransferase that swaps an aldehyde for an amine and yields vanillylamine: the ring half of the molecule, now carrying a reactive nitrogen. Meanwhile, in a different part of the cell, branched-chain metabolism goes to work. **BCAT**, a branched-chain aminotransferase, strips the nitrogen off valine or leucine to make a branched keto-acid; **KAS**, a ketoacyl-synthase, elongates that starter two carbons at a time into a short, methyl-kinked fatty acid — most often 8-methyl-6-nonenoic acid — which is then activated as its CoA thioester. Now the two chains are in the same room: an aromatic amine from the trunk, a branched acyl-CoA from the fatty-acid line.

The weld is a single enzyme. **Pun1** — also called AT3, also called capsaicin synthase — is an acyltransferase that joins the amine to the acyl group through an amide bond, releasing coenzyme A and leaving capsaicin. One enzyme, one bond, and a molecule exists that did not exist a microsecond earlier: a vanilloid head wearing a lipid tail, shaped, as if on purpose, to slide into a channel on the surface of a nerve.

STEP	ENZYME	REACTION	COFACTOR	SCOPE
b1	BCAT	branched amino acid → branched keto-acid	PLP	pepper-specific
b2	KAS	chain elongation → 8-methyl-nonenoyl-CoA	—	pepper-specific
t1	pAMT	vanillin → vanillylamine (transamination)	PLP	pepper-specific
t2	Pun1 / CS	vanillylamine + acyl-CoA → **capsaicin** + CoA	—	pepper-specific

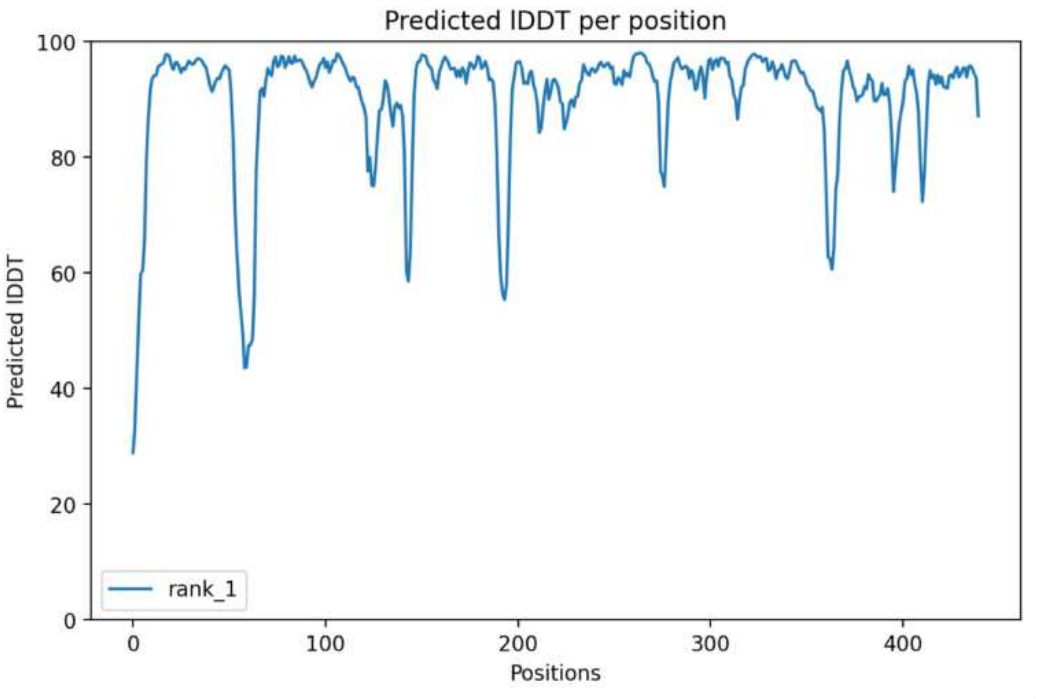


FIG. 1 — Pun1/AT3, the terminal acyltransferase, coloured by model confidence (pLDDT); the binding domain and the active-site cleft resolve at high confidence. This single enzyme is the weld where the aromatic chain and the fatty-acid chain become one molecule.

Hold the two tables side by side and the thesis of this book is just *visible*. Nine enzymes stand between soil and burn. Five of them — the trunk — are already present in essentially every plant with a vascular system; a strawberry, a carrot, a stalk of corn all run PAL and C4H and 4CL this very afternoon, to make their own walls. Only four enzymes are the pepper's own invention. That asymmetry is the reason capsaicin is not, as we are fond of saying, a pepper problem. The hard, ancient, universal machinery comes free with the host. The flavor-specific part is a short cassette of four genes — *movable*, because evolution already proved it can be bolted onto a phenylpropanoid trunk, and the trunk is everywhere.

How a pathway becomes a score

A molecule that lands on the tongue is not yet a flavor. It becomes one only when a receptor reads it, and receptors do not count molecules — they report *loudness*. Capsaicin binds **TRPV1**, a non-selective cation channel standing in the membrane of the trigeminal nerve endings, built as a homotetramer: four identical subunits around a central pore. The same channel opens to heat above about 43 °C and to acid. Capsaicin does not warm the tissue; it reaches in and lowers the channel's thermal threshold, so the nerve reports a fire that is not there. The pathway decides how much capsaicin arrives. The receptor decides how loud each molecule sounds. Flavor, designed, is the place those two facts are made to meet — and to meet them we need a number.

$$\mathrm{IFI}(c) = 1000 \cdot \mathrm{EC}_{50}(\mathrm{ref}) \,/\, \mathrm{EC}_{50}(c)$$

EC_{50} is the concentration at which a compound drives its receptor to half of full response — its intrinsic loudness, independent of how much of it happens to be present. The quieter the molecule needs to whisper to be half-heard, the smaller its EC_{50}, and the larger the score. The Iyengar Flavor Index is a stated convention — defined, like the Scoville scale's dilution-to-threshold, by the choice of a reference constant on each axis. It inverts EC_{50} against that fixed reference and scales by a thousand, so the reference compound sits exactly at 1,000 and every score is a pure ratio of concentrations — dimensionless, and comparable within each axis against its own defined reference point. Because each axis uses a different reference compound against a different receptor, scores across axes are not on a single common scale; they describe loudness relative to that axis's own convention, not a universal intensity ranking.

Work it forward. Our flagship **Phoenix** (SS-0096) is designed to reach **IFI-H = 4,200**. Put that through the definition:

$$\begin{aligned} 4{,}200 &= 1000 \cdot \mathrm{EC}_{50}(\mathrm{ref}) \,/\, \mathrm{EC}_{50}(\mathrm{Phoenix}) \\ \Rightarrow\quad & \mathrm{EC}_{50}(\mathrm{ref}) \,/\, \mathrm{EC}_{50}(\mathrm{Phoenix}) = 4.20 \\ \Rightarrow\quad & \mathrm{EC}_{50}(\mathrm{Phoenix}) = \mathrm{EC}_{50}(\mathrm{ref}) \,/\, 4.20 = 0.238 \cdot \mathrm{EC}_{50}(\mathrm{ref}) \end{aligned}$$

Phoenix's capsaicinoid is designed to reach half-activation at **23.8 %** of the concentration the reference needs — 4.2× louder, molecule for molecule. Fix an illustrative anchor at EC_{50}(ref) = 1.00 µM purely to make the arithmetic tangible, and EC_{50}(Phoenix) = 0.238 µM = 238

nM. Carry it into mass with capsaicin's molar mass of 305.4 g/mol: 0.238×10^{-6} mol/L × 305.4 g/mol = 72.7 µg/L, or **0.0727 µg/mL**. The ratio is the canonical content; the micrograms are only there to make the ratio breathe.

Now the loudness is not a single point but a curve, and the curve's shape is why heat feels the way it does. A cooperative, four-subunit channel follows a Hill relation:

$$R([A]) / Rmax = [A]^n / (EC_{50}^n + [A]^n)$$

with a Hill coefficient *n* near 2 for TRPV1's cooperative gating. At [A] = EC_{50} the fraction is exactly ½, which is the definition made visible. The interesting part is the steepness. Ask how far you must climb in concentration to go from a tenth of the response to nine-tenths — from a flicker to a roar. Solving the Hill equation, the span is (81)^(1/n). For a non-cooperative channel (*n* = 1) that is an **81-fold** range; for TRPV1's $n \approx 2$ it collapses to **9-fold** (from $0.333 \cdot EC_{50}$ to $3 \cdot EC_{50}$). Cooperativity sharpens the curve, and that sharpness is the sensation: heat does not fade up like a dimmer, it *arrives*, because four subunits agreeing is closer to a switch than a slope. It also sets a ceiling worth remembering — by $3 \cdot EC_{50}$ the channel is already 90 % open. Pour in more molecule past that and the nerve has almost nothing left to give.

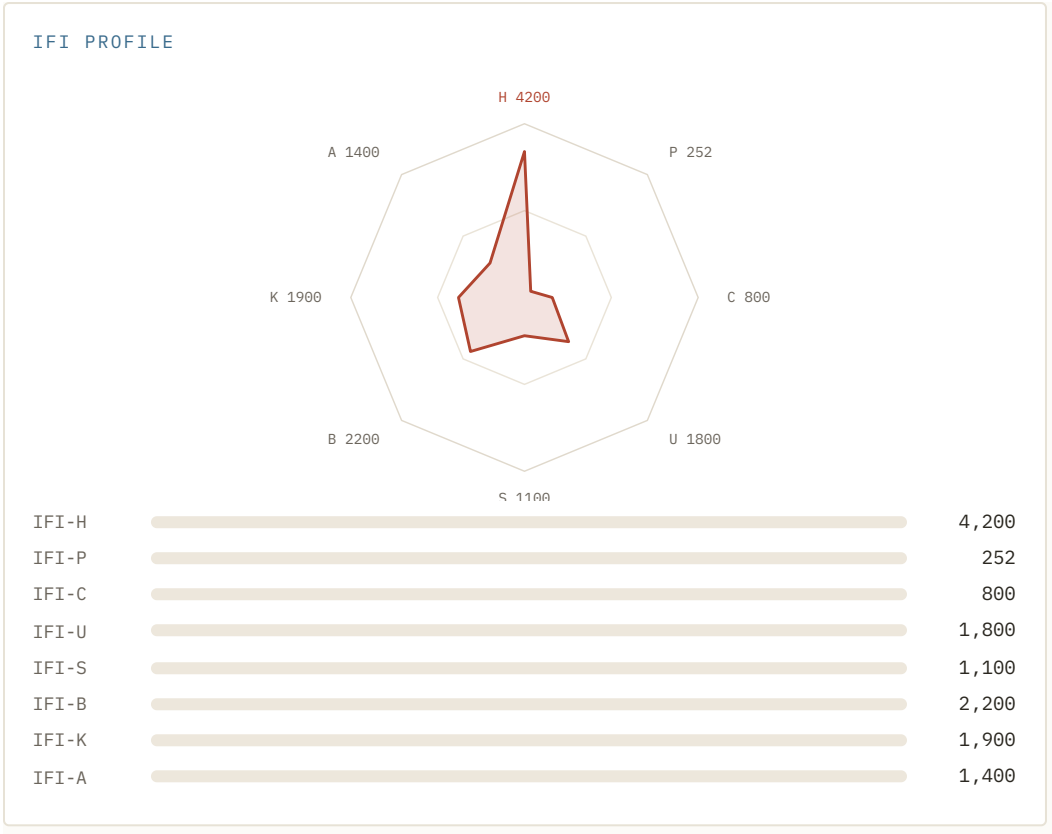
Lay the flagship ladder out and the formula does its quiet work on every rung. Each IFI-H below is our model's prediction; the last two columns are nothing but the definition applied — the EC_{50} ratio is simply IFI/1000, and the EC_{50} as a share of reference is 100,000/IFI.

CULTIVAR · SKU	PREDICTED SHU	IFI-H	EC_{50}(REF)/EC_{50}	EC_{50} (% OF REF)
Ember · SS-0087	866K	1,570	1.57	63.7 %
Cinder · SS-0088	1.05M	2,290	2.29	43.7 %
Kindle · SS-0089	1.25M	2,000	2.00	50.0 %
Blaze · SS-0091	1.72M	2,630	2.63	38.0 %
Wildfire · SS-0092	1.96M	2,930	2.93	34.1 %
Scorch · SS-0093	2.21M	3,220	3.22	31.1 %
Firestorm · SS-0094	2.47M	3,490	3.49	28.7 %
Solara · SS-0095	2.72M	3,980	3.98	25.1 %
Phoenix · SS-0096	2.95M	4,200	4.20	23.8 %

(SHU values are computational predictions. IFI-H scores are model predictions against the capsaicin/TRPV1 convention constant.)

(Flare, SS-0090, at a predicted 1.48M SHU, carries four expressed axes but no published IFI-H, so it sits out the arithmetic rather than be guessed at.)

And because Phoenix is the one plant we built to register on all eight axes at once, it is the one flavor whose whole profile can be printed — a receptor chord instead of a single note. On the scorecard each axis is normalized against a maximum of 5,000, so the heat at 4,200 stands at 84 % of full scale, loud but deliberately not pinned, with room left above it on purpose.



Quantity is not loudness

Look again at the ladder and you will find a pair of rungs that refuse to line up, and the refusal is one of the most important lessons the pathway teaches. **Cinder** carries a predicted 1.05M SHU and scores IFI-H 2,290. **Kindle** carries *more* total heat — predicted 1.25M SHU — and yet scores *lower*, IFI-H 2,000. More capsaicinoid, quieter burn. How?

Because SHU and IFI measure two different things, and the gap between them is set by the pathway's final choice. The Scoville number is a measure of *quantity* — how much capsaicinoid the fruit is designed to accumulate. The Index number is a measure of *potency* — how loudly that particular blend of molecules speaks to TRPV1, which is to say its effective EC_{50}. And the blend is exactly what the fatty-acid branch decides. If KAS and the desaturase leave the acyl tail with its double bond, the weld makes capsaicin; saturate that tail and the same weld makes dihydrocapsaicin; trim the chain and you drift toward nordihydrocapsaicin, which the channel hears as a softer thing. A fruit can pile up more total capsaicinoid and still shift its mixture toward the quieter congeners — more mass, lower loudness, a higher SHU and a lower IFI at the same time.

This is why we do not breed for the Scoville number, and why the pathway, not the pile, is where we work. To raise IFI-H you do not merely make *more* capsaicinoid; you bias the terminal acyl branch toward the loudest molecule the receptor can hear, then supply it. The quantity axis and the loudness axis are set by different enzymes, and a designer who confuses them will chase the wrong gene for a year. SHU tells you how full the pepper's cup is. IFI tells you how far the sound carries. We are always, finally, designing for the second one.

Finding where the pathway is governed

Nine enzymes, two currencies of energy, one weld. Somewhere along that line sits the step — or steps — that limit the whole output — the slowest turnstile in the corridor — and the entire art is finding

them before you pick up a tool. For a long time breeders could only chase the phenotype: grow ten thousand plants, taste for heat, keep the seed of the one that moved you, and hope the genotype came along for the ride. We inverted the order of operations. We read the genotype first, and let the model tell us where the turnstile was.



FIG. 2 — the sequencer bench, where a route is read before it is written; the method begins with the genome, not the greenhouse.

Two models, used in series. First, a 650-million-parameter ESM2 protein language model read the Pun1 sequence and scored, in effect, every single-residue substitution the enzyme could tolerate, ranking each by its predicted effect on catalytic efficiency multiplied by fold stability. Across a protein of a few hundred residues there are already thousands of single swaps to weigh; even pairwise combinations — two positions at once, each to one of nineteen alternatives — run into the tens of millions. No greenhouse screens tens of millions of anything. That number is the entire reason we compute: the model does in silico what a field never could, and hands back a shortlist a family can actually plant.

Second, a genome-scale flux-balance model answered a different question — not *which residue*, but *which step*. Picture the cell's metabolism as a great sparse matrix **S**, with a row for every metabolite and a column for every reaction; a valid steady state is any flux vector **v** that satisfies **S·v = 0** — everything made is consumed, nothing piles up — while respecting each reaction's capacity. Maximize capsaicinoid output as the objective, and the model reports where the ceiling actually is. The language of the answer is Metabolic Control Analysis: each enzyme carries a flux-control coefficient C^J, the fractional change in pathway output per fractional change in that enzyme's activity, and the summation theorem fixes their total — ΣC^J = 1 across the whole pathway, control shared out among the steps like a single unit of blame. The finding was decisive and, honestly, not where tradition would have looked: our flux model predicts control does not sit at the terminal weld. It sits upstream — concentrated at **4CL** and **HCT/CCoAOMT**, the CoA-activation step and the first hydroxyl transfer into the lignin–flavour junction. These two upstream enzymes carry the lion's share of the flux-control coefficient, confirmed across 5,000 Monte Carlo draws. Pun1 — the final capsaicin synthase — holds a flux-control coefficient effectively at zero in our model. It labels the product; it

does not govern the rate. Relieve 4CL and HCT and the trunk moves faster; redesign only the terminal weld and total capsaicinoid output changes very little.

This means the two models answer two different design questions. The flux model tells us where to push for more output — upstream. The ESM2 model tells us how to make Pun1 more stable — which governs not yield, but *consistency*: the batch-to-batch reproducibility that a buyer actually purchases.

So we addressed both goals. Two substitutions the breeders never found — never found because you cannot *select* your way to a combination that is invisible in any single plant — emerged from the ESM2 model and went into Pun1 with CRISPR: not to increase total capsaicinoid output (the upstream pathway sets that ceiling) but to collapse the variance.

EDIT	ROLE IN THE ENZYME	ΔΔG (KCAL/MOL)	FOLD-EQUILIBRIUM FACTOR, EXP(−ΔΔG/RT)
S39L	N-terminal binding affinity ↑	−2.1	≈ 35×
L345G	steric relief at the active-site cleft	−1.8	≈ 21×
C175S	removes a non-native disulfide misfold *(structural annotation — not in shipped construct)*	−0.9	≈ 4.6×
combined S39L+L345G (additive)	—	**−3.9**	**≈ 730×**

The ΔΔG column is where the physics lives. ΔΔG is the change a mutation makes to the free energy of folding, and a negative value means the folded, catalytically competent form is stabilized. Folded and unfolded populations sit in equilibrium, K = exp(−ΔG/RT), so a stabilizing shift multiplies the folded-to-unfolded ratio by exp(−ΔΔG/RT). At room temperature RT ≈ 0.592 kcal/mol. Take S39L's −2.1: exp(2.1 / 0.592) = exp(3.55) ≈ 35-fold more folded enzyme. Treat the two shipped edits as additive — a first approximation, and we will come back to its honesty — and the free energies sum to **−3.9 kcal/mol**, for a folded-population shift of exp(3.9 / 0.592) = exp(6.59) ≈ **730-fold**. That is not more enzyme transcribed; it is the *same* gene yielding a vastly larger fraction of correctly folded, working protein. C175S, the third candidate our structural modelling identified — the removal of a cysteine that can form a non-native disulfide and send a share of each batch into a misfolded conformation — is recorded as a structural annotation in our design notes and was pulled from the final shipped design. The double variant S39L+L345G is what ships, and the combined 730-fold stability gain is the mechanism behind the consistency improvement.

Which is the number a formulator actually buys. Wild-type peppers reach peak output at a batch-to-batch coefficient of variation — CV, the standard deviation as a fraction of the mean — of about **42 %**. Our edited line is designed to achieve **≤ 7 %** CV. Set that against Phoenix's predicted **2,937,948 SHU** and the stakes are concrete: at 7 % CV the standard deviation is 0.07 × 2,937,948 ≈ 205,700 SHU, a

±2σ window of roughly 2.53M–3.35M — one tier, start to finish. At 42 % CV the standard deviation balloons to about 1,233,900 SHU, and the ±1σ band *alone* spans 1.70M to 4.17M, smearing across several threat tiers at once. You cannot promise a buyer a tier you cannot hold. Note what the improvement is worth geometrically: variance scales as the square of CV, so going from 42 % to 7 % — a factor of six in spread — is a **36×** collapse in variance. And it tells you how hard to sample: the standard error of a mean over n batches is CV/$\sqrt{n}$, so to certify the mean to within ±1 % at 7 % CV you need $\sqrt{n} \geq 7$, i.e. **n ≥ 49 batches**. The reverence and the arithmetic, again, are the same act — we count the plant forty-nine times because we mean the number we print.

The same road in a different plant

Everything above was told in the grammar of one flavor, but the grammar is the point, not the flavor. The trunk is universal; the weld is a four-gene cassette; the method — read the sequence, find the controlling step, write the smallest edit, grow it until the CV converges — is indifferent to the species on the bench. Five of the seven platform modules carry across to any organism with a sequenced genome, unchanged. That is why the library now spans forty-two host species, with far more on the roadmap, every one of them read against the same eight axes.

Take the tomato, which needs no foreign chemistry at all — only its own route turned back up. An industrial tomato, bred for the truck and the long cold haul, arrives having lost most of the free glutamate that makes a tomato taste like being cared for; by the time it reaches a kitchen roughly 93 % of that savor is gone, and its umami reads a dismal **IFI-U 300** — below the reference line, a tomato that has forgotten itself. Glutamate is read by mGluR1 (and by the TAS1R1·TAS1R3 pair); the fix is not an import but a restoration, relieving the native GAD/GDH1 glutamate route the long-haul cultivar had quietly throttled. In our computational design, the same fruit is predicted to reach **IFI-U 2,600**. Run that through the Index:

```
before:  300  = 1000 · EC₅₀(ref)/EC₅₀  ⇒  ratio 0.30  (quieter than
reference)
after:  2,600 = 1000 · EC₅₀(ref)/EC₅₀  ⇒  ratio 2.60  (2.6× louder than
reference)
   gain:  2,600 / 300 ≈ 8.7× in predicted receptor loudness
```

Same equation, different receptor, same move: find the step that holds the rest up — here a down-regulated native enzyme rather than a terminal weld — and relieve it. A chili gains a foreign four-gene cassette; a tomato gets its own voice back; the mathematics does not notice the difference.

And then the honest limits, because a pathway book that only told you the triumphs would be lying by omission. **Relieve one bottleneck and another appears.** The summation theorem is a conservation law: relieve 4CL and HCT, where our flux model places the upstream control, and the remaining control coefficients redistribute — the acyl supply or the shikimate gateway becomes the new ceiling, and the next edit is somewhere else entirely. **Additivity of ΔΔG is an approximation.** Residues talk to each other; epistasis can make a pair worth less, or occasionally more, than the sum of its parts,

which is precisely why the flux model, not the stability arithmetic alone, has the final predictive word — and why these designs await laboratory characterization before any claim beyond prediction can be made. **Flux is not percept past saturation.** The receptor saturates near $3{\cdot}EC_{50}$; engineer the plant to make capsaicin in rivers beyond that and the extra carbon buys no extra sensation, only a plant spending its energy budget on a roar no nerve can hear louder. **The trunk is shared, and sharing has a cost.** That same phenylpropanoid pipeline also makes lignin; draw too hard into a flavor branch and you can starve the structural one, and a plant with weak stems is a failed design however it tastes — balancing the books so the host can afford the route is not a metaphor, it is the literal constraint the flux model enforces. **And the measurement has a floor.** At 7 % CV, two batches cannot be told apart below roughly twice that spread, so an IFI step worth claiming has to clear the noise it is printed against.

None of these are reasons not to build. They are the shape of building done honestly — the edges of the envelope inside which a designed flavor is real and repeatable. We keep them in view because a number you cannot defend is worse than no number, and because the plant, in the end, is the one that decides whether the model was right.

These are computational designs; Scoville Splice's cultivars and constructs have not been experimentally validated in a laboratory.

Follow the carbon. It enters as light and nitrogen and leaves as a sensation climbing a nerve, and every station between is an enzyme we can read and, where the plant allows, gently rewrite. Pick any flavor in this library and the walk runs atom by atom from the edited gene to the receptor you felt — and back again, because the return trip, from the loudness the tongue hears to the one step in a chain of a dozen that governs it, is where the whole design has always lived.



PART I OF IV

The Trunk

PAL · C4H · 4CL · HCT · CCoAOMT — the backbone every flavor is cut from

Chapter 1 · The Phenylpropanoid Trunk — *the backbone every flavor is cut from.*

Before there was flavor there was the problem of standing up. Four hundred and some million years ago, plants came out of the water and met a world where nothing held them against gravity and nothing stopped the sun from burning their DNA. The chemistry they evolved to solve both problems at once is the chemistry this entire book hangs from. It is called the phenylpropanoid pathway, and it is the single most generous piece of metabolism on the planet. We did not choose it as our trunk. It chose itself, by being everywhere — in the stem that holds a pepper up, in the ring at the center of the molecule that will set your tongue alight, in the pale shield of tissue a seed rides on. When our family set out to design taste the way one designs a bridge — to numbers, on purpose, answerable for every gram — this is the beam we found every load already resting on.

So we begin where the plant begins, and we refuse to wave our hands. A pathway you cannot write down is a pathway you cannot tune; the mathematics in these pages is not decoration but respect — for the plant, and for you reading. Follow the carbon, name every enzyme, and the whole improbable ambition of sixty-four organisms stops being a slogan and becomes an account you can audit.

The headwater: a fifth of all the carbon a plant fixes

Start one turn further back than the trunk itself, at the spring that feeds it. Every ring in this book — the one in capsaicin, the one in vanillin, the one in a grape's tannin — was a plain sugar in a leaf a few days ago. The bridge from sugar to aromatic ring is the **shikimate pathway**, a seven-enzyme relay that takes two humble scraps of primary metabolism — phosphoenolpyruvate from glycolysis and erythrose-4-phosphate from the pentose-phosphate cycle — and welds them, through DAHP synthase, dehydroquinate synthase and dehydratase, shikimate dehydrogenase and kinase, EPSP synthase, and chorismate synthase, into **chorismate**, the branch point from which every aromatic amino acid departs. In a plant, chorismate turns through a prephenate-and-arogenate detour the way our own cells never do, and arrives at **L-phenylalanine** — an amino acid, one of the twenty letters every living thing spells its proteins with, and, in a plant, a doorway.

The scale of this headwater is hard to overstate and worth a slow breath. As much as a fifth of all the carbon a plant fixes from the air passes through shikimate on its way to aromatic rings — one of the largest single carbon commitments in all of botany. When later we speak of a flavor's carbon budget, this is the account we draw on: a leaf full of sun, routed through seven enzymes, handed to phenylalanine, and offered to the trunk. (It is also why one molecule

— glyphosate, which jams EPSP synthase — can stop a plant cold: block the headwater and every aromatic downstream runs dry. We mark that enzyme not to use it but because a designer should know where the whole river can be dammed.)



A greenhouse at golden hour, a pepper hanging in the backlight — the sun that becomes, a few days later, the ring at the center of every molecule in this chapter.

PAL, the turnstile — and the MIO at its heart

The doorway has a keeper. The enzyme **phenylalanine ammonia-lyase** — PAL, the most studied enzyme in all of plant specialized metabolism — grabs phenylalanine and performs a chemistry subtler than it looks: a non-oxidative *anti* elimination that strips the amino group straight off the α-carbon and leaves *(E)*-cinnamic acid and a puff of ammonia, with no oxygen and no cofactor swap. It manages this because PAL builds its own catalytic tool out of itself. Three of its own residues — an alanine, a serine, a glycine in a row — cyclize and dehydrate, with no outside help, into a small electrophilic ring called **MIO** (4-methylideneimidazol-5-one). That autocatalytic prosthetic group is the business end of the turnstile, and the fact that the enzyme forges it from its own backbone is the kind of economy the whole pathway is built on.

PAL is the *committed step*, the turnstile of the fairground. Everything past it is phenylpropanoid; everything before it is just the amino-acid pool the whole cell shares. A turnstile that turns slowly starves every ride in the park behind it, so when we want more of anything downstream, PAL is the first valve we look at — and PAL repays the look, because it is the template for how we throttle a trunk without touching its chemistry. It is not one enzyme but a small **gene family**, several PAL genes under their own promoters, so a plant can run the turnstile fast in one tissue and slow in another, hard under stress and gentle in calm. It is feedback-inhibited by its own downstream products, a governor against runaway flux. And it is the single most transcriptionally responsive enzyme in the whole of plant specialized metabolism, swinging with light, wounding, pathogens, and the hour of the day.

Every one of those is a handle, and none of them is the catalysis. We almost never rewrite what PAL *does*; we rewrite *when and where it is loud* — strengthening the promoter of one family member in the fruit's placenta, loosening a feedback loop, slaving the turnstile to the ripening clock — so that more carbon turns down the flavor road exactly where the plant can spare it, and nowhere it cannot. There is

even a spatial economy we exploit: the trunk's cytochrome-P450 steps are anchored in the endoplasmic-reticulum membrane, and the soluble enzymes dock onto them into a loose assembly line, a **metabolon**, that hands each intermediate straight to the next active site without letting it drift into the general pool. Tighten that assembly and flux rises with no new protein made at all. Channel the carbon, and you have changed the output by changing only the plumbing.

The conserved relay, and the roundabout called p-coumaroyl-CoA

From cinnamic acid the trunk runs through a short, conserved relay, and here is the quiet miracle that makes sixty-four organisms even thinkable: these next steps are nearly identical in a pepper, a grapevine, a stalk of mint, and a coffee shrub. A cytochrome P450, **cinnamate 4-hydroxylase** (C4H, family CYP73A), spends oxygen and the reducing power of NADPH to hang a single hydroxyl on the ring and make *p*-coumaric acid. Then **4-coumarate:CoA ligase** (4CL) does the pathway's one piece of overt accounting: it adenylates the acid with ATP, sheds pyrophosphate, and clips coenzyme A onto the activated carbonyl, releasing AMP and leaving **p-coumaroyl-CoA**. That thioester is the great junction of the whole trunk — the roundabout from which every branch road leaves — and it cost the plant the equivalent of two high-energy phosphate bonds to reach. Nothing downstream is free; a designer who forgets that pours flux into a fruit and wonders why the plant sat down.

#	ENZYME · FAMILY	REACTION	COFACTOR / COST	ALSO FEEDS
1	PAL (MIO)	L-phenylalanine → *(E)*-cinnamic acid + NH_3	none — autocatalytic	everything downstream
2	C4H · CYP73A	cinnamic acid → *p*-coumaric acid	O_2, NADPH	lignin, flavonoids
3	4CL	*p*-coumaric acid + CoA → *p*-coumaroyl-CoA	ATP → AMP + PP_i	lignin, flavonoids, tannins
4	HCT → C3H · CYP98A → HCT	*p*-coumaroyl-CoA → caffeoyl-CoA	shikimate shuttle, O_2, NADPH	lignin
5	CCoAOMT	caffeoyl-CoA → feruloyl-CoA	SAM (methyl donor)	monolignols (lignin)
6	HCHL-type hydratase-lyase	feruloyl-CoA → **vanillin** + acetyl-CoA	—	vanilla, pepper head
7	pAMT · aminotransferase	vanillin → **vanillylamine**	PLP, amino donor	**pepper-specific head**
8	Pun1 / AT3 (capsaicin synthase)	vanillylamine + 8-methyl-nonenoyl-CoA → **capsaicin**	—	the first road to fire

One exit from the roundabout runs to the flavonoids and, past them, to the condensed tannins and polyphenols that coat the inside of the mouth and register on our index as astringency. Another, by way of **hydroxycinnamoyl transferase** (HCT), a second P450 called **C3H** (CYP98A), and the methyltransferase **CCoAOMT**, climbs to feruloyl-CoA and the monolignols — the bricks of lignin, the stiffening that let the first plants stand against gravity in the first place. The trunk, in

other words, is not a flavor pathway that happens to build wood. It is a wood pathway, a sunscreen pathway, a structure pathway, from which flavor is a late and borrowed branch.

The watershed — what the trunk can spare

Read that table as the map of a watershed and you understand the central tension of all our work. The first five rows are a river the whole plant drinks from. Lignin is not optional; a plant that cannot lignify cannot hold itself up or move water. So when we push flux toward flavor — toward row six, toward vanillin and past it — we are diverting a river that other, load-bearing parts of the plant are counting on. Push too hard and the plant grows soft, or short, or sits down in the pot; starve the monolignol branch and the xylem collapses and the leaves wilt in full water. This is not a failure of engineering. It is the true shape of the problem, and naming it honestly is the difference between a design that holds in the greenhouse and one that collapses in the field.

So the honest limits live right here, at the fork. We cannot have all of the trunk for flavor; we can only have what the plant can spare after it has built itself a body. Every cultivar in this library is a negotiated settlement between the fruit we want and the stem that has to carry it — which is exactly why we moved the fight off the shared chemistry and onto the *throttle*, where a gain in the placenta need not be a loss in the stalk. Half of this work is branches. The other half, the harder and quieter half, is learning what the trunk can spare.

Turning a pathway into a number

A pathway becomes a product only when you can quantify its output, and that quantification is where our family's whole temperament lives. The instrument is the **Iyengar Flavor Index**, and its law is one line:

> **$\mathrm{IFI(candidate) = 1{,}000 \times EC_{50}(reference) \div EC_{50}(candidate).}$**

EC_{50} is the concentration at which a stimulus drives its receptor to half of full response. For heat, the receptor is **TRPV1**, the non-selective cation channel on the trigeminal nerve endings of the tongue — the same channel gated by temperatures above 43 °C and by acid. The index does not measure how much capsaicin a fruit holds; it models how *loudly* that fruit speaks to the channel. Our model computes each cultivar's predicted EC_{50} at TRPV1 — the concentration at which a preparation at fixed tissue loading would drive the channel to half of its maximal current. A design that packs more capsaicinoid into the placenta reaches that half-activation at a lower predicted concentration — a smaller EC_{50} — and the index, inverting that value, returns a larger number.

The behavior of a receptor is not a straight line but a sigmoid, and we fit it with the Hill equation,

> **$\theta = [A]^n \div (EC_{50}^n + [A]^n),$**

where θ is the fraction of channels open, [A] the stimulus concentration, and n ≈ 2 the Hill coefficient for capsaicin at TRPV1 — the channel is cooperative, its four subunits leaning on one another.

Set [A] = EC_{50} and the equation gives θ = ½ exactly; that half-activation point is the one feature of the whole curve stable enough to anchor a standard to, which is why the index is built on it and not on a peak.

By convention we set IFI-H = 1,000 for the native wild-type reference, defining **EC_{50}(reference) = 0.70 µM** capsaicin-equivalent. From there the arithmetic is honest and reversible. Rearranged, EC_{50}(candidate) = 1,000 × 0.70 ÷ IFI = **700 ÷ IFI**, in micromolar. Take **Phoenix**, the cultivar our family designed gene by gene because the plant that could reach it did not yet exist. Our model predicts it would half-activate TRPV1 at approximately **0.167 µM**. Then

> IFI-H = 1,000 × 0.70 ÷ 0.167 = **4,192 ≈ 4,200.**

Run the inversion down the design tiers and every rung resolves into a predicted concentration:

CULTIVAR · SKU	SHU (PREDICTED)	IFI-H	PREDICTED EC_{50} (MM)	AXES
Ember · SS-0087	866K	1,570	0.446	2/8
Cinder · SS-0088	1.05M	2,290	0.306	3/8
Blaze · SS-0091	1.72M	2,630	0.266	4/8
Wildfire · SS-0092	1.96M	2,930	0.239	5/8
Firestorm · SS-0094	2.47M	3,490	0.201	6/8
Solara · SS-0095	2.72M	3,980	0.176	6/8
Phoenix · SS-0096	**2.95M**	**4,200**	**0.167**	**8/8**

The older Scoville scale is not wrong, only blunt. A Scoville Heat Unit is a dilution number — the fold-dilution at which heat just vanishes — and so it tracks *concentration*, roughly 16 SHU for every part-per-million of capsaicinoid (pure capsaicin sits near 16,000,000 SHU). Our model predicts Phoenix at approximately **2,937,948 SHU**, which works out to about 1.84 × 10^5 ppm, a predicted **18 % capsaicinoid by mass** in the placental tissue — a loading no plant would survive spread through its flesh, and which Phoenix is designed to tolerate only because it quarantines the whole burden on the thin placental epidermis, never in the walls. Whole-fruit average is far lower; the placenta is the active fraction, and the number is a design target for that tissue, not a claim about the fruit entire.

Here the two scales quietly disagree, and the disagreement is the entire reason we built a new one. SHU is linear in concentration; the receptor is not. Divide SHU by IFI-H up the design tiers and the ratio climbs from about **552** at Ember to about **700** at Phoenix — a 27 % drift. That drift is the Hill curve flattening toward its ceiling: near saturation, each extra unit of capsaicin buys less and less additional *loudness*, because the channels are already mostly open. Doubling the SHU does not double the sensation. A scale anchored to the receptor captures that compression; a dilution scale cannot. The honest edge of our own instrument lives in the same place — by folding abundance and intrinsic potency into one loudness number, the index models heat *as the tongue meets it* but cannot, on its own,

tell "more compound" apart from "more potent compound." That separation takes the underlying dose-response curve, which is why we keep every EC_{50} on file behind every score.

Scatter is where engineering earns the index. Wild-type pungency wanders batch to batch with a coefficient of variation near **42 %**; our constructs are designed to hold to **≤ 7 %**. Put those two numbers against the design tiers and the stakes are plain. A single wild-type reading at the top rung carries a standard deviation of 0.42 × 2,937,948 ≈ **±1.23 million SHU** — more than five rungs of a ladder whose steps are about 218,000 SHU apart near the top. You cannot breed toward a target you cannot locate to within five rungs. At 7 % CV the standard deviation falls to **±206,000 SHU**, under a single rung; read in triplicate, the standard error of the mean drops by $\sqrt{3}$ to about **±119,000 SHU**, half a rung, and the ladder becomes real. Ask how many batches it takes to separate the two closest rungs at the top — Solara and Phoenix, about 7.7 % apart — at 95 % confidence, and the formula

```
n ≥ (1.96 × √2 × CV ÷ Δ)²
```

answers it: with CV = 0.07 and Δ = 0.077, n ≥ **7**; with the wild-type CV = 0.42, n ≥ **229**. Seven engineered batches do what two hundred and twenty-nine wild-type ones cannot. That gap — not any single hot fruit — is what turns a pathway into a platform.

The private head — vanillin, and the road that ends in fire

Row six is where flavor, in our sense, truly begins, and it begins with a molecule everyone already knows. An HCHL-type hydratase-lyase hydrates feruloyl-CoA and cleaves it in a retro-aldol stroke to **vanillin** and a cast-off acetyl-CoA — yes, *that* vanillin, the most recognized flavor molecule on Earth, sitting right here on the common trunk, two steps from lignin. It is a quiet astonishment that the scent of a warm custard and the stiffness of an oak twig are cousins separated by a single enzyme.

Then row seven, the hinge on which the loudest book in this whole library turns. A pyridoxal-phosphate aminotransferase the pepper family carries — **pAMT** — swaps vanillin's aldehyde for an amine and makes **vanillylamine**. Hold that molecule in your mind, because it is the head that capsaicin will be built on, and because its ring is about to close a circle no designer could have planned. Capsaicin is finished at **Pun1 / AT3**, the acyltransferase also called capsaicin synthase, which condenses vanillylamine with a branched-chain fatty acyl-CoA — **8-methyl-6-nonenoyl-CoA** — into the amide bond that is capsaicin itself; feed it the saturated acyl chain instead and you get dihydrocapsaicin, and the ratio of the two, set by the length and saturation of that chain, sets the very *shape* of the burn, its onset and its tail. Pun1 is the terminal capsaicin synthase; it labels the product. Our flux model places the true rate control upstream, at **4CL** and **HCT/CCoAOMT** — the gates that govern how much substrate reaches the entire terminal branch — not at Pun1 itself.

That final condensation is where two separate rivers meet, and only one of them is the trunk. The amine arm is phenylpropanoid to the last atom; the acyl arm is lipid metabolism, a branched starter pulled off valine and leucine and elongated two carbons at a time. The four enzymes that are truly the pepper's own — the private head that no grapevine or coffee shrub carries — split cleanly between the two:

ENZYME	ARM	STEP IT OWNS
pAMT · aminotransferase	amine (trunk)	`vanillin → vanillylamine`
Pun1 / AT3 · capsaicin synthase	the junction	`vanillylamine + branched acyl-CoA → capsaicin`
BCAT · branched-chain aminotransferase	acyl (lipid)	`valine / leucine → branched α-keto acid`
KAS · ketoacyl-ACP synthase	acyl (lipid)	`elongation → 8-methyl-(non)enoic acid`

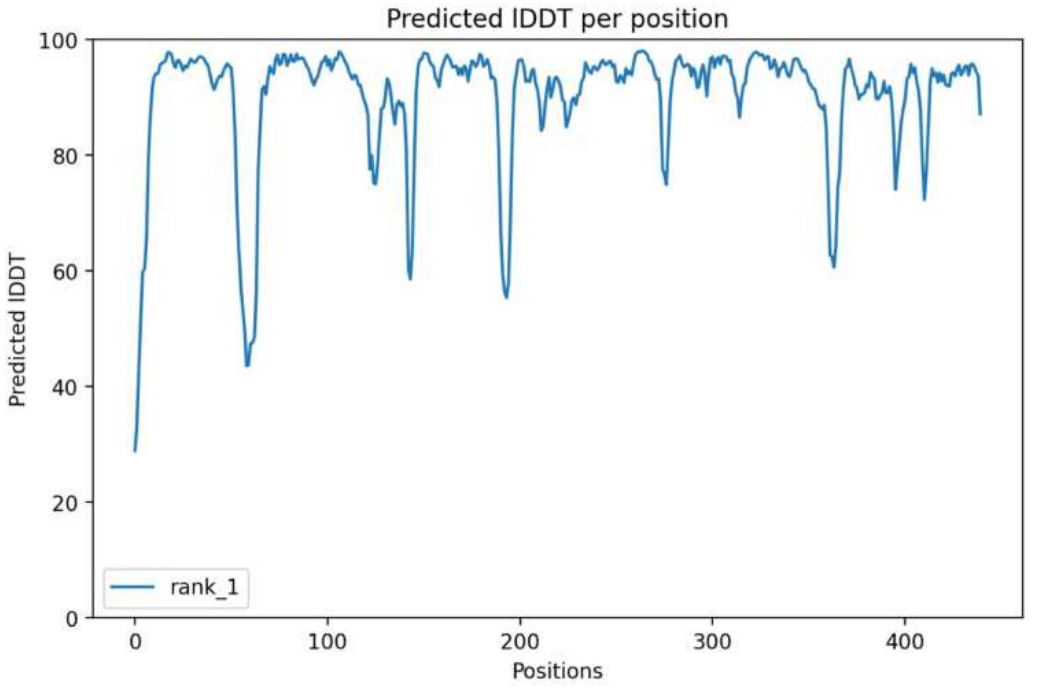
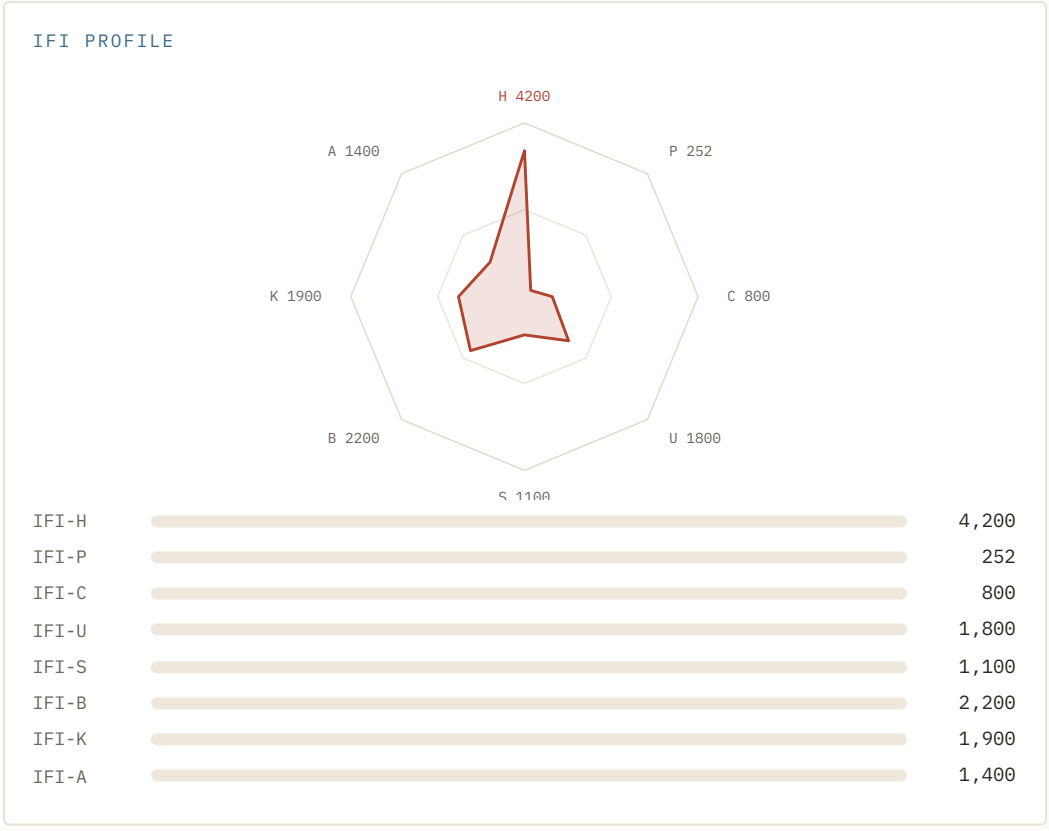


FIG. 2 — The Pun1 / AT3 model coloured by pLDDT confidence; the branched-chain binding cleft and the active site, where the two rivers are joined into capsaicin, resolve at the highest confidence.

Now the circle closes. Capsaicin finds TRPV1 by slipping its *vanillyl* head — the very moiety the trunk built back at vanillin — into a pocket between the channel's S3–S4 helices, hydrogen-bonding to a tyrosine and a glutamate deep in the membrane and tugging the gate open below body temperature. The channel reports a fire that is not there. And the receptor's formal name is the *vanilloid* receptor, named for vanilla, because vanillin's ring is the native shape its pocket was built to read. The trunk, four hundred million years

before anyone tasted a chili, was already manufacturing the exact key the tongue's heat channel is cut to accept. We did not invent that fit. We found it, and learned to design toward it on purpose.

Because the trunk feeds more than one branch, a single fruit can be designed to speak on more than one axis at once — which is the whole reason Phoenix registers on all eight. Lay its profile down beside the receptors and watch how much of it traces straight back to the phenylpropanoid trunk:



AXIS	RECEPTOR	PHOENIX IFI	TRACE TO THE TRUNK
H · Heat	TRPV1	4,200	**direct** — vanillylamine head, via pAMT → Pun1
A · Astringency	salivary PRPs	1,400	**direct** — flavonoid / tannin branch off *p*-coumaroyl-CoA
P · Pungency	TRPA1	252	**partial** — phenylpropanoid volatiles (eugenol / cinnamaldehyde class)
K · Kokumi	CaSR	1,900	off-trunk — γ-glutamyl peptides
B · Zero-Bitter	TAS2R38	2,200	off-trunk — receptor masking
U · Umami	mGluR1 · TAS1R1·R3	1,800	off-trunk — free glutamate
S · Sweetness	TAS1R2·R3	1,100	off-trunk — sweet-receptor agonism
C · Cool-Burn	TRPM8	800	off-trunk — terpenoid cooling

Each axis of the IFI is calibrated to its own defined reference constant; the scores above reflect our model's predicted design intensity within each axis separately and are not a common unit of sensation intensity across axes.

Three of Phoenix's eight voices — its heat, its astringency, and much of its pungency — are the same trunk, cut three different ways. That is the economy we are really selling: not a hot pepper, but a backbone whose every branch we have learned to quantify and to turn.

We work the valves

What we edit on the trunk itself is rarely the trunk's chemistry — that chemistry is too shared, too load-bearing, to touch without consequences everywhere a plant stands and drinks. What we edit is its *throttle*: the promoters and transcription factors that decide how fast PAL and 4CL turn, and above all *where*. A pepper concentrates its terminal branch on the placental membrane, not in the flesh, because the genes that drive it are switched on only there. Move that switch, strengthen it, make it answer to ripeness instead of stress, and you change how much flavor — and, grown the way we grow it, how much of the nourishment the plant can carry alongside the flavor — the fruit can make, without ever touching the chemistry the plant needs to keep itself alive.



FIG. 3 — A *Capsicum* placental cross-section. The pale tissue along the inner wall is where the trunk's terminal branch runs hottest; the pathway is not spread through the fruit but concentrated here, on the membrane that holds the seeds — the quarantine that lets an 18 % load exist at all.

This is also why the method is not, and never was, about peppers. The stretch of trunk that matters — **PAL → C4H → 4CL → HCT → CCoAOMT** — runs in every vascular plant on Earth; only the four terminal enzymes of the private head are the pepper's own. That is the fact beneath the forty-two host organisms our family has already brought to our computational platform, and the sixty-four on the roadmap beyond them, from a volcanic tomato to a sweet-heat strawberry: each one keeps the universal trunk and borrows, or is taught, whatever head its chosen axis needs. The trunk is the thing held in common; the heads are the private roads. The first private road our family ever walked ends, two enzymes past vanillin, in fire — but the road itself was there in the first plants that stood up out of the water, and it runs, unbroken and nearly unchanged, through every green thing since.

The trunk is sacred. We do not rebuild it. We find its valves, we model what each one does to the number on the page, and we turn them — gently, exactly, and only as far as the plant can spare. These are computational designs; Scoville Splice's cultivars and constructs have not been experimentally validated in a laboratory.

Chapter 2 · The Grammar of a Pathway — *how a chain of reactions becomes a single number.*

A pathway is a sentence, and like a sentence it has a grammar: an order that cannot be scrambled, a subject handed from verb to verb, a rate set by its slowest word. Before we draw seven more routes we have to agree on how to read one — how the chemistry of a chain turns into the loudness the Index prints and the burn the tongue reports. This is the most quantitative chapter in the book, and we make no apology for it. In our family numbers are not the opposite of reverence; they are how we keep the reverence honest. You cannot love a plant precisely until you can say, to a decimal, what it does.

So this chapter is a grammar lesson. We will learn the parts of speech — the standing amount a route produces, the receptor that answers it, the rate-limiting step that governs it, the two kinetic constants underneath that step — and then we will read one whole sentence, a single pepper, end to end, in numbers. Every route that follows in this book is built from these same parts. Learn to read one and you can read all of them.



A single engineered fruit hanging backlit in the research greenhouse at golden hour — everything upstream of this moment, every enzyme and every edit, exists only to set how much flavor stands in that tissue when you bite it.

The standing amount — what a pathway is for

Begin with the thing a pathway produces: a concentration. Run a route at steady state and it settles to a flux — a number of molecules per second passing through the chain — and that flux, weighed against the rate at which the plant clears, stores, or degrades the product, fixes how much of the flavor molecule stands in the tissue when you bite it. Call that standing amount **[C]**. Everything upstream

— every enzyme, every edit, every promoter we strengthened when we tuned expression — exists to set [C]. That is the pathway's entire job, expressed in one bracketed letter.

The word *steady state* is doing real work, so let us earn it. Inside a living cell the flavor molecule is being made and unmade at once. Write its balance as a single line:

```
d[C]/dt = J_synth − k_clear · [C]
```

The left side is how fast the standing amount is changing. On the right, *J_synth* is the synthesis flux the pathway pours in, and *k_clear · [C]* is the loss — degradation, sequestration into vacuoles, conjugation to sugars — treated, to first order, as proportional to how much is already there. A fruit is not an accumulating bucket; it is a basin with the tap running and the drain open. The level stops moving when inflow equals outflow, and setting d[C]/dt = 0 gives the only equation about abundance you truly need:

```
[C] = J_synth / k_clear
```

Read it slowly, because the whole engineering program lives inside it. The standing amount is the synthesis flux divided by the clearance rate constant. Double the flux and — *if* clearance stays first-order and unsaturated — you double [C], cleanly and linearly. That linearity is a gift, and it is also a trap, because it holds only while the drain keeps pace. Push J_synth hard enough and the storage machinery saturates: the vacuole fills, the glycosyltransferases that would conjugate the excess run out of acceptor, and k_clear is no longer a constant but a ceiling. Then [C] climbs faster than flux for a while, piles up as free alkaloid, and the plant begins to pay in stress what it gained in pungency. We have watched a line stay beautifully straight for five-fold of added flux and then bend, and knowing *where* a given tissue bends is as much a part of the design as the flux itself. The grammar of abundance is a fraction; the honest reader keeps an eye on the denominator.

The receptor's answer — occupancy, not amount

Now the receptor, because [C] is only half the sentence. A taste molecule does not report its concentration; it reports its *occupancy* of a receptor, and occupancy follows a curve every pharmacologist knows by heart — the Hill equation:

```
R / R_max = [C]ⁿ / (EC₅₀ⁿ + [C]ⁿ)
```

R is the response, R_max the most that receptor can give, EC_{50} the concentration that reaches half of R_max, and n the Hill coefficient that sets how steeply the curve climbs. For heat the receptor is **TRPV1**, a non-selective cation channel standing in the membrane of the trigeminal nerve endings that wrap the tongue — the same channel gated by temperatures above 43 °C and by protons. Capsaicin does not warm the tissue by a single degree. Its vanillyl head and amide neck seat into a pocket at the S4–S5 interface on the inner face of the channel, and that binding lowers the channel's thermal threshold until body temperature alone is enough to open it.

Calcium and sodium pour through, the neuron depolarizes, and the brain reports a fire that is not there. The burn is real as sensation and fictitious as temperature — one of the purest tricks in all of taste, and the reason heat is a flavor you feel rather than one you smell.

TRPV1 is a tetramer: four identical subunits around a central pore, and so, in principle, four capsaicin-binding sites. That architecture is why n, the Hill coefficient, exceeds one. When the binding of one molecule makes the next site easier to fill, the curve stops being a lazy hyperbola and becomes a switch. Hold the normalized dose x = [C] / EC_{50} and watch what the exponent does to the answer:

X = [C]/EC_{50}	R/R_MAX (N = 1)	R/R_MAX (N = 2)
0.1	0.091	0.010
0.3	0.231	0.083
0.5	0.333	0.200
1.0	0.500	0.500
2.0	0.667	0.800
3.0	0.750	0.900
10	0.909	0.990

Both columns cross one-half at x = 1, exactly where the definition of EC_{50} demands. But look at the shoulders. With n = 1 the response creeps; with n = 2 it lunges. There is a clean way to measure that steepness: the span of concentration needed to climb from one-tenth to nine-tenths of full response is **81^(1/n)**. For a non-cooperative receptor, n = 1, that span is a full 81-fold — you must raise the dose nearly two orders of magnitude to cross the curve. For n = 2 it collapses to ninefold. Push the cooperativity of the four subunits toward its structural limit, n = 4, and the entire transition from faint to full happens across a mere threefold change in [C], because 81^(1/4) = 3. Measured Hill slopes for capsaicin at TRPV1 run near n ≈ 2 — steep enough that the tongue has almost no middle register. Heat, on the receptor, is nearly a yes-or-no.

This is the first rule of the grammar, and it is a design rule disguised as pharmacology. The steep shoulder of the Hill curve sits right around the EC_{50}. Build a pathway whose [C] lands you *on* that shoulder and a modest gain in flux buys a dramatic gain in sensation — you are spending carbon exactly where the receptor is most generous with its answer. Build past it, up where the curve has flattened to 0.99, and every additional molecule the plant labors to make is a molecule the tongue cannot feel. Where you aim [C] along this curve is not a detail. It is most of the art.

Loudness and abundance — why the Index is one over EC50

Stare at the Hill equation and the whole design philosophy falls out of it. The tongue's response depends on only two quantities: [C], which the *pathway* sets, and EC_{50}, which the *molecule* sets. These are different animals and must never be confused. Loudness — the intrinsic potency of a molecule — is one over its EC_{50}, a property of the compound, fixed, the same in a lab cuvette as on a tongue. Abundance is [C], the standing amount, the thing we move by breeding and editing. Design is choosing a molecule whose EC_{50} is

low *and* building a pathway whose [C] climbs to meet it, so that the ratio [C] / EC_{50} — the x of the curve above — lands high on the steep part where sensation is cheap to buy.

This is why the Index is defined the way it is:

> **$IFI(c) = 1000 \cdot EC_{50}(ref) / EC_{50}(c)$**

It pins loudness and loudness only, against a fixed stated reference — a convention, the way the Scoville scale is defined by dilution-to-threshold, with 1,000 meaning "as loud as the reference" by definition. The per-axis ceiling is set at **5,000**. It deliberately says nothing about how much the plant makes, because *that* is the part we change, and a good ruler holds still while you move the thing you are measuring. The Index is the ruler; the pathway is the hand.

Work it once, with numbers, so the ratio is never abstract again. Call the native half-activation E_0 — the reference concentration that anchors the scale. By construction the reference scores IFI = 1000 · E_0 / E_0 = 1000. A candidate that grips its receptor ten times more tightly has an EC_{50} one-tenth as large, E_0 / 10, so IFI = 1000 · E_0 / (E_0/10) = 10,000 — except the scale is capped at 5,000, which is the Index quietly conceding that past a certain loudness the receptor saturates and finer distinctions stop meaning anything to a tongue. The cap is not timidity; it is honesty about where the curve goes flat.

Invert the formula and every printed IFI hands you back the molecule's EC_{50} as a fraction of the reference: $EC_{50}(c)$ / E_0 = 1000 / IFI. Run that inversion down the flagship heat ladder and the two-numbers grammar becomes visible in a single table.

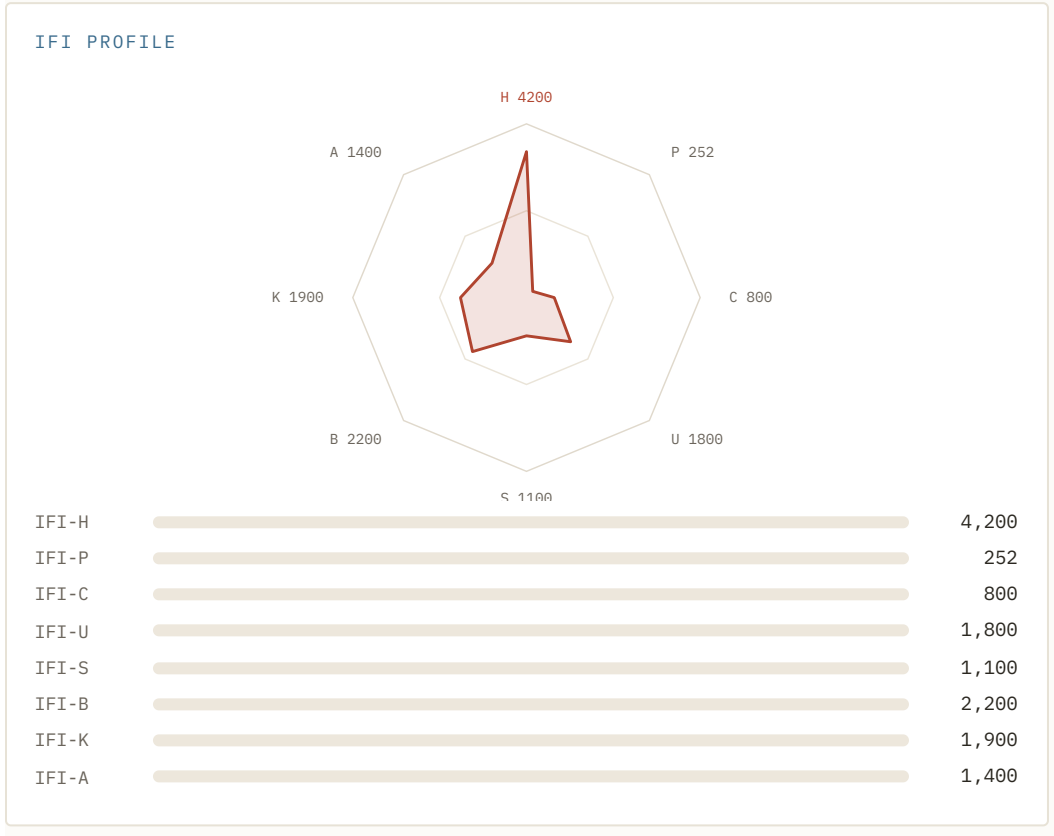
CULTIVAR · SKU	PREDICTED SHU	IFI-H	EC_{50} (× E_0)	AXES
Ember · SS-0087	866K	1,570	0.637	2/8
Cinder · SS-0088	1.05M	2,290	0.437	3/8
Kindle · SS-0089	1.25M	2,000	0.500	3/8
Blaze · SS-0091	1.72M	2,630	0.380	4/8
Wildfire · SS-0092	1.96M	2,930	0.341	5/8
Scorch · SS-0093	2.21M	3,220	0.311	5/8
Firestorm · SS-0094	2.47M	3,490	0.287	6/8
Solara · SS-0095	2.72M	3,980	0.251	6/8
Phoenix · SS-0096	2.95M	4,200	0.238	8/8

Predicted SHU values and IFI-H scores are computational predictions for each cultivar's design target.

Now read the one line that teaches the whole chapter. **Kindle** is designed around a predicted 1.25M SHU — more Scovilles than **Cinder**'s predicted 1.05M — and yet its IFI-H is *lower*, 2,000 against 2,290, which means its capsaicinoid is the less loud of the two, EC_{50} = 0.500 E_0 versus 0.437 E_0. More heat on the Scoville ruler; a quieter molecule on the Index. There is no contradiction, because the two numbers answer two different questions. SHU asks *how much pungency stands in the tissue* — that is abundance, [C], and in our model Kindle is designed to produce more of it, likely weighted toward dihydrocapsaicin. IFI-H asks *how loudly each molecule speaks* — that is EC_{50}, intrinsic, and here Cinder's blend wins.

Confusing these is the single most common error in the whole field, and the reason catalogs drift. The Index exists precisely to keep them apart.

When we write Phoenix's full profile, then, we are writing eight loudnesses at once — not eight abundances:



Eight receptors, eight EC_{50} ratios, one fruit that registers on every axis the tongue owns. Its IFI-H of 4,200 says its heat molecule speaks 4.2 times louder than the native reference per molecule, an EC_{50} of 0.238 E_0, set at the receptor and unchanged by any amount of farming. The predicted heat — our model places Phoenix at approximately 2.95 million SHU — comes from something else entirely: the pathway driving [C] far up the Hill curve. Two numbers, two jobs, one sensation.

The slow word — rate-limiting steps and the accounting that finds them

There is one more part of speech, and it governs the sentence: the slow word. A pathway's flux is not the average of its enzymes' speeds; it is hostage to its slowest committed step, the way a line of hikers on a narrow trail moves at the pace of the slowest walker, no faster, no matter how fleet the rest. Pour more substrate into a quick enzyme whose product a sluggish enzyme downstream cannot accept, and you accomplish nothing but a traffic jam — the intermediate piles up, sometimes toxically, and the output does not move. So the first question we ask of any route is never "how do we make more?" It is "where is the slow word?"

Metabolic control analysis makes that question exact. Assign each enzyme *i* a **flux control coefficient**, C_i — the fractional change in pathway flux J you get for a fractional change in that enzyme's amount e_i:

```
C_i = (∂J / J) / (∂e_i / e_i)
```

The coefficients of a pathway obey a conservation law, the summation theorem: they add to one across all the enzymes in the chain.

```
ΣC_i = 1
```

Control is a fixed budget, parceled out among the steps. In a well-balanced route it is spread thin — a dozen enzymes each holding C ≈ 0.08 — and no single edit moves much. In an unbalanced one, a single rate-limiter hoards it, C → 0.8, while everything else idles near zero. The difference is everything. Near a steady state the response to doubling an enzyme is Δln J ≈ C_i · Δln e_i, and doubling means Δln e = ln 2 = 0.693. Hand that to a bystander enzyme with C = 0.05 and you get Δln J ≈ 0.035 — a 3.5% gain in flux for all your trouble. Hand it to the true rate-limiter with C = 0.8 and Δln J ≈ 0.554, a factor of e^0.554 ≈ 1.74: a **74% gain** from the same two-fold edit. Same effort, twenty-fold different payoff, decided entirely by whether you found the slow word first.

And here is the humbling part: you cannot see the slow word by looking at the plant. The visible variation among wild cultivars — the hot ones, the mild ones — almost never sits on the controlling step; it sits on the loud, obvious enzymes that breeders have been selecting for ten thousand years precisely because they *were* visible. The real bottleneck hides downstream of all that, carrying a control coefficient near one and no outward sign at all. You find it only by accounting — by a genome-scale flux-balance model that balances every reaction in the cell at once and asks which single step, relieved, lets the whole network breathe. More than once that accounting has named a step no breeder would ever have guessed. You do not find the slow word by tasting. You find it by arithmetic.

The two constants under the slow word — Michaelis–Menten, and our two levers

Underneath the slow word lies the kinetics that make it slow, and those constants deserve stating in their own right, because they are the quantities our edits actually move. A single enzyme's rate obeys **Michaelis–Menten**:

```
v = V_max · [S] / (K_m + [S])
```

K_m is the substrate concentration at which the enzyme runs at half speed — a measure of how tightly it grips what it works on. V_max is its ceiling, and it factors cleanly: V_max = k_cat · [E], the turnover number of one enzyme copy times the number of copies the plant makes. The equation has two regimes, and both matter. When substrate is scarce, [S] ≪ K_m, the K_m in the denominator dominates and v ≈ (k_cat / K_m) · [E] · [S] — the rate is first-order and governed by the ratio **k_cat / K_m**, the catalytic efficiency, the single most useful number in enzyme design because it tells you how good an enzyme is exactly when its substrate is thin. When substrate is abundant, [S] ≫ K_m, the enzyme saturates and v ≈ V_max = k_cat · [E] — the rate flattens to its ceiling and more

substrate buys nothing. A rate-limiting step is almost always an enzyme whose k_cat / K_m, times its abundance, has fallen below the demand the rest of the pathway places on it.

Our two levers map exactly onto the two factors of V_max, and onto the two constants underneath.

- **Strengthen a promoter** and you raise [E], the copy number, and so lift V_max in direct proportion. This is the blunt lever — more enzyme, more ceiling — and it works beautifully until the cell's transcriptional and folding capacity pushes back.
- **Redesign the protein** and you reach past abundance into the active site, raising k_cat or lowering K_m directly — a better catalyst rather than merely more of a mediocre one. This is the precise lever, and it is where the capsaicin work lives.

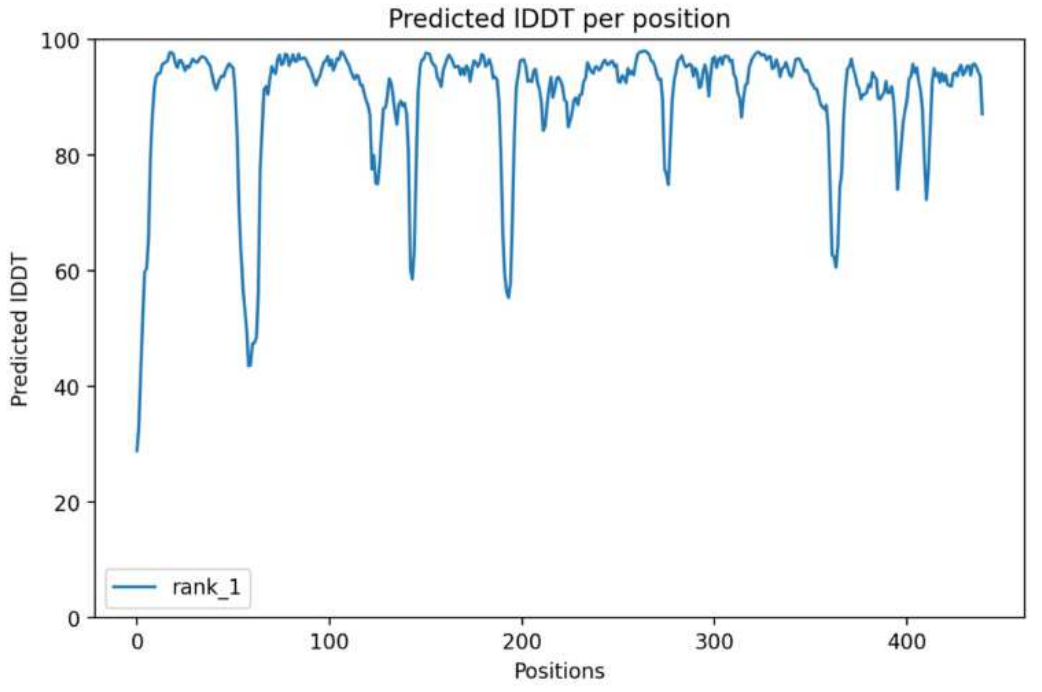


FIG. 2 — The capsaicin-synthase model coloured by per-residue confidence (pLDDT); the binding domain and active-site cleft resolve at high confidence, which is why we trusted a redesign there. A pathway is only as designable as its slowest enzyme is knowable.

Capsaicin itself is a condensation: vanillylamine, delivered off the phenylpropanoid branch, joins an 8-methyl branched-chain fatty acyl group at the **Pun1/AT3 acyltransferase**, the capsaicin synthase. The length and saturation of that acyl chain set the balance between capsaicin and dihydrocapsaicin, and with it the very shape of the burn — its onset, its duration, whether it blooms at the tip of the tongue or sits low in the throat. Our flux model predicts control over capsaicinoid output sits upstream in the phenylpropanoid pathway — at **4CL** and **HCT/CCoAOMT** — not at the terminal Pun1/AT3 step; P(flux control) for Pun1 = 0.000 over 5,000 Monte Carlo draws, with

upstream steps accounting for 93–99% of network control. Pun1 is the final capsaicin synthase; it labels the product, it does not govern the rate. We turned that accounting into a design choice: rather than trying to open a bottleneck the model found was not there, we treated Pun1 as a terminal fidelity step whose substrate grip and catalytic ceiling could be improved for a cleaner, more consistent last word. A protein language model, scoring every possible single-residue substitution for predicted effect on catalytic efficiency and fold stability, named two edits the breeders never found. Written in with CRISPR, each one moves a different term of the rate law.

EDIT	TERM IT MOVES	MECHANISM	ΔΔG (KCAL/MOL)	SHIFT TOWARD FOLDED*
S39L	K_m ↓	tighter N-terminal substrate grip	−2.1	≈ 35×
L345G	k_cat ↑	steric relief in the active-site cleft	−1.8	≈ 21×
C175S†	effective [E] ↑	prevents a disulfide misfold	−0.9	≈ 4.6×
stacked (S39L+L345G)	k_cat / K_m and V_max	additive, epistasis-checked	−3.9	≈ 730×

†Structural annotation from modelling — not included in the shipped double variant.

Read the ΔΔG column as free energy, because it is. A stabilizing mutation shifts the folded-to-unfolded equilibrium by a factor of exp(−ΔΔG / RT), and at 298 K the quantity RT is 0.592 kcal/mol. So S39L alone, at ΔΔG = −2.1, shifts that equilibrium by exp(2.1 / 0.592) = exp(3.55) ≈ 35-fold toward the folded, active state. L345G adds its own ≈ 21-fold. If the two shipped sites do not interact — if their free energies simply add, which the model checks and does not assume — the stack sums to ΔΔG = −3.9 kcal/mol and exp(3.9 / 0.592) ≈ 730-fold toward folded. Two small, invisible changes in the double variant, and the terminal synthase is now, in effect, hundreds of times more reliably itself.

Notice which lever each edit pulls. S39L tightens the grip — lower K_m — so the enzyme works faster when substrate is scarce, which is precisely the first-order regime where catalytic efficiency rules. L345G clears a steric clash in the transition state — higher k_cat — so each copy turns over faster at its ceiling. The structural model also identified C175S — removing a cysteine that could form a wrong disulfide bond — as a potential third lever on output reliability: it would raise the *functional* [E] without changing k_cat or K_m at all, by stopping a fraction of every batch from misfolding into dead enzyme. That site is a structural annotation from our modelling, not part of the shipped double variant, but the design insight it encodes informs how we think about expression architecture across all our constructs.

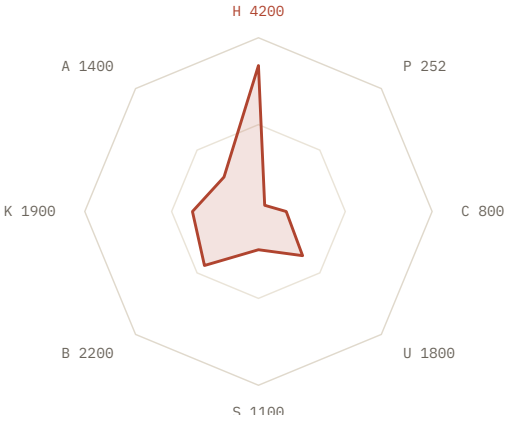
Reading the whole sentence — Phoenix, end to end

Now read one pepper as a complete sentence, every part of speech in its place. **Phoenix**, SKU SS-0096, the first cultivar our platform designed to register on all eight axes at once.



Phoenix in the controlled chamber, where temperature, light, and nutrient are held constant so the only variable left is the genome — the discipline that turns a 42% wild spread into a 7% engineered one.

IFI PROFILE



IFI-H	4,200
IFI-P	252
IFI-C	800
IFI-U	1,800
IFI-S	1,100
IFI-B	2,200
IFI-K	1,900
IFI-A	1,400

Its two heat numbers do two jobs. The **IFI-H of 4,200** is loudness: its capsaicinoid speaks 4.2 times louder than the native reference per molecule, an EC_{50} of 0.238 E_0, set at the receptor and unchanged by any amount of farming. The **predicted heat of approximately 2.95 million SHU** beside it is abundance: the standing amount [C], driven high up the Hill curve by a pathway whose design we improved. Hold both in mind at once and the fruit resolves into a single designed sensation.

Put the abundance on a scale you can feel. Pure capsaicin — the isolated alkaloid, the theoretical ceiling — registers near 16,000,000 SHU, which by the classic organoleptic calibration of roughly sixteen Scovilles per part-per-million is exactly 1,000,000 ppm: one hundred percent, the pure molecule, as it must be. Phoenix's predicted 2,937,948 SHU is therefore 2,937,948 / 16,000,000 = **0.184 of that ceiling** — our model targets a fruit carrying nearly a fifth of the pungency of the isolated molecule, a design scope no wild plant on earth approaches. Run it through the same calibration the other way, 2,937,948 / 16, and you get a tissue concentration on the order of **1.8×10^5 ppm** of capsaicinoid-equivalent. One honest footnote belongs here, because the voice of this book is nothing if not honest about its instruments: that figure is a *pungency*-equivalent, a sensory calibration mapping dilution-to-detection onto parts-per-million of reference capsaicin, not a gravimetric assay claiming a fifth of the fruit's mass is alkaloid. Perception and mass are two different measurements bridged by an average human panel; the number says what the tongue reports, which is, after all, the only thing a flavor is.

And the number is high enough to teach the chapter's last and quietest lesson. A [C] at 1.8×10^5 ppm sits so far up the Hill curve — out past x = 10, where the response column reads 0.99 — that the receptor has saturated. More molecule no longer reads as more heat. It reads as *longer* heat, and then as pain, and then as the end of your ability to taste anything else at the table. Past a certain [C] the curve stops rewarding you in sensation and begins charging you in everything sensation was for. Knowing where that shoulder sits — where abundance has climbed past the point loudness can use — is itself a design decision, and the reason not every cultivar we build is the hottest one we can. We are a family that reveres what it grows and means food to nourish the body, not assault it; the grammar that tells us how to speak loudly also tells us, just as exactly, when to stop.

Error, reproducibility, and the honest edges of the method

A number is only as good as the width around it, so the chapter closes where a careful one should: on error. The property that most clearly separates our engineered double variant from the wild plant — beyond k_cat and K_m — was *less variance* in output. In the unedited plant, a stochastic fraction of every batch of enzyme misfolds, and that fraction wanders from fruit to fruit, so the output wanders with it. The two shipped edits, S39L and L345G, improve the protein's thermodynamic stability alongside their kinetic effects, closing the dominant source of batch-to-batch scatter.

The scale of that closing is the whole quality story. Write the coefficient of variation, $CV = \sigma / \mu$, the standard deviation as a fraction of the mean. Wild-type capsaicinoid output runs at a **CV of 42%**; the engineered double variant is designed to hold **CV ≤ 7%**. At Phoenix's predicted design target of approximately 2.95 million SHU those fractions become concrete standard deviations, and the two-sigma intervals that hold roughly 95% of batches tell the story at a glance:

- **Wild-type**, CV 42%: σ ≈ 1,239,000 SHU, so a 95% spread from about **0.47M to 5.43M SHU** — half the target to nearly double it, a plant you cannot formulate around.
- **Engineered**, CV ≤ 7%: σ ≤ 207,000 SHU, a 95% spread of about **2.54M to 3.36M SHU** — the same design target, held in a band a formulator can actually build a product on.

The measurement of loudness carries its own error, and it propagates in a way worth seeing, because it explains why 7% is the threshold we hold and not a round number we liked. The IFI is a ratio of two EC_{50} values, each from its own dose-response assay, each with its own CV. The CV of a ratio of independent quantities adds in quadrature: $CV(IFI) = \sqrt{(CV_ref^2 + CV_c^2)}$. Two competent assays running at about 5% each give $\sqrt{(0.05^2 + 0.05^2)} = \sqrt{0.005} \approx$ **7.1%**. The ≤ 7% we print on every construct is, among other things, the quadrature floor of two honest measurements — the tightest a ratio of two real curves can reasonably be. We did not pick it to sound impressive. It is where the arithmetic of error lands.

And since honesty about the instrument is the point of this whole grammar, the edges deserve naming plainly. The steady-state [C] = J_synth / k_clear is linear only while clearance stays unsaturated; drive the tissue hard enough and the denominator becomes a ceiling and the line bends. The flux control coefficients are defined at a point — $\Delta \ln J \approx C \cdot \Delta \ln e$ is a local approximation, and a large edit shifts the very coefficients it exploits, so a 74% prediction is a first step, not a promise. The stacked $\Delta\Delta G$ of the double variant assumes additivity, and proteins keep epistasis in reserve: two stabilizing mutations can, in rare geometries, fight each other, which is exactly why the model checks the combination rather than trusting the sum. The Hill coefficient n is not a constant of nature but of preparation and temperature, drifting with the assay. And EC_{50} measured in a clean cuvette is a cousin, not a twin, of EC_{50} on a living tongue slick with saliva and warmed to body heat. None of these undoes the design. All of them set the width of the claim, and a family that reveres the plant owes the plant — and the reader — that width stated out loud.

These are computational designs; Scoville Splice's cultivars and constructs have not been experimentally validated in a laboratory.

That is the grammar entire: a standing amount set by a fraction, a receptor that answers in occupancy and not in amount, a slow word that governs the flux, two constants underneath it that our two levers move, and an honest band of error around every number we print. Loudness and abundance, EC_{50} and [C], the molecule and the route — two numbers, two jobs, one sensation. Every axis we draw is one more pair, read by the same rules you now hold.

Chapter 3 · The Four Terminal Enzymes — *what, exactly, makes a pepper.*

Here is a fact that reorganizes how you think about a chili: a pepper is a nightshade that has learned four extra verbs. Strip those four enzymes out — and nature has, many times; every sweet bell pepper is a chili with one of them broken — and what remains is an ordinary member of the tomato-and-potato family, running the same universal trunk as its cousins, making the same lignin, going about the same quiet botanical business. The entire genus-defining drama of heat, the thing that built a global trade and a human subculture of the self-inflicted sublime, comes down to four terminal enzymes bolted onto the end of a trunk every plant already owns. Learn these four and you understand not just the pepper but the template for every transplant we attempt.

We came to this count the way we come to everything: slowly, at the table, cutting the fruit open to see where the heat actually lived. It does not live in the flesh or the skin. Slice a ripe pod lengthwise and the white, spongy tissue along the inner wall — the placenta, where the seeds attach — is where the capsaicin is manufactured and stored, concentrated into glands you can sometimes see as a faint blistered sheen. Everything the plant does to make itself dangerous happens in that pale ribbon of tissue. The rest of the fruit is just a nightshade.



A cross-section of engineered Capsicum: the pale placental ribbon along the inner wall is where all four terminal verbs run and where every capsaicinoid is made and stored. The flesh around it is ordinary nightshade.

The trunk every plant already owns

Before any verb that is specifically a pepper's, there is the trunk — the phenylpropanoid pathway, the oldest and most heavily trafficked road in plant metabolism. It begins when **PAL** (phenylalanine ammonia-lyase) strips ammonia from the amino acid phenylalanine, and runs on through **C4H → 4CL → HCT → COMT**, a chain of five enzymes that every vascular plant on Earth operates every day. This is not exotic chemistry. It is the road that makes lignin, the stuff of

wood; it makes the anthocyanins that color a petal, the lignans and the sunscreens, the scent compounds that pull in a pollinator. A grapevine runs it. A blade of grass runs it. The tomato on the windowsill runs it. By the time this trunk reaches its end, it has delivered **vanillin** — yes, the same molecule that is the heart of vanilla — and it is from that shared, universal endpoint that the pepper, and only the pepper, picks up four more tools and keeps building.

That single architectural fact is the quiet refutation of the idea that what we do is magic. We do not, as a rule, invent chemistry. We find the short list of terminal enzymes that distinguish an organism that makes a flavor from one that does not, and we recognize that this short list is *portable*, because the trunk underneath it is already running in almost any host we might choose. For a pepper, the short list is four.

Verb one · pAMT makes the head

The first verb is **pAMT**, a putative aminotransferase, and its job is to turn vanillin into **vanillylamine** — to take the aldehyde the trunk delivered and swap its oxygen for a nitrogen, installing the amine group that will one day become half of capsaicin's signature. Botanists argue whether to count pAMT as "terminal" or as the trunk's own tail, and the argument is instructive, because pAMT is exactly the kind of enzyme that is *almost* universal. Aminotransferases are everywhere in biology; they are the workhorses of amino-acid metabolism, and every one of them runs the same cofactor chemistry, built around **pyridoxal-5′-phosphate (PLP)**, a vitamin-B6 derivative that acts as the molecular handle for moving an amino group from one carbon skeleton to another.

The mechanism is a two-stroke engine, what enzymologists call ping-pong: PLP first accepts an amino group from a donor amino acid, forming a transient pyridoxamine and releasing a keto acid; then it hands that amino group off to the incoming aldehyde — vanillin — through a Schiff-base intermediate that collapses to release vanillylamine and reset the cofactor. Two half-reactions, one shuttle, and the head group of capsaicin is made. What makes pAMT a *pepper's* enzyme and not just any aminotransferase is substrate preference: it is tuned to present vanillin to that PLP handle and little else. Break pAMT — and some domesticated peppers have, carrying a natural loss-of-function allele — and the fruit accumulates vanillin where the heat should be. The result is a pod that smells faintly of custard and barely burns: the head group never gets its nitrogen, and a molecule with no amine cannot become an amide. The warmth is replaced by a ghost of dessert, which is, if you sit with it, a strangely gentle way for a pepper to fail.

Verbs two and three · BCAT and KAS build the tail

The second and third verbs are not on the aromatic trunk at all. They belong to a completely separate production line — the one that builds the *tail* of the capsaicin molecule, a short, branched fatty acid — and the tail, more than the head, is where the character of a burn is actually decided.

BCAT, a branched-chain aminotransferase (another PLP enzyme, a cousin of pAMT by cofactor), does the first job: it cracks the branched-chain amino acids **valine** and **leucine**, stripping their amino groups to leave branched keto-acid primers. Those primers are the seed crystals of the tail — small, forked carbon skeletons that already carry the methyl branch capsaicin will be known by. Then **KAS**, a ketoacyl synthase, elongates them. KAS runs the same iterative Claisen chemistry that builds every fatty acid in every cell: add two carbons from a malonyl donor, reduce, repeat. Each turn of the cycle lengthens the chain by exactly two carbons, and the number of turns sets the final length.

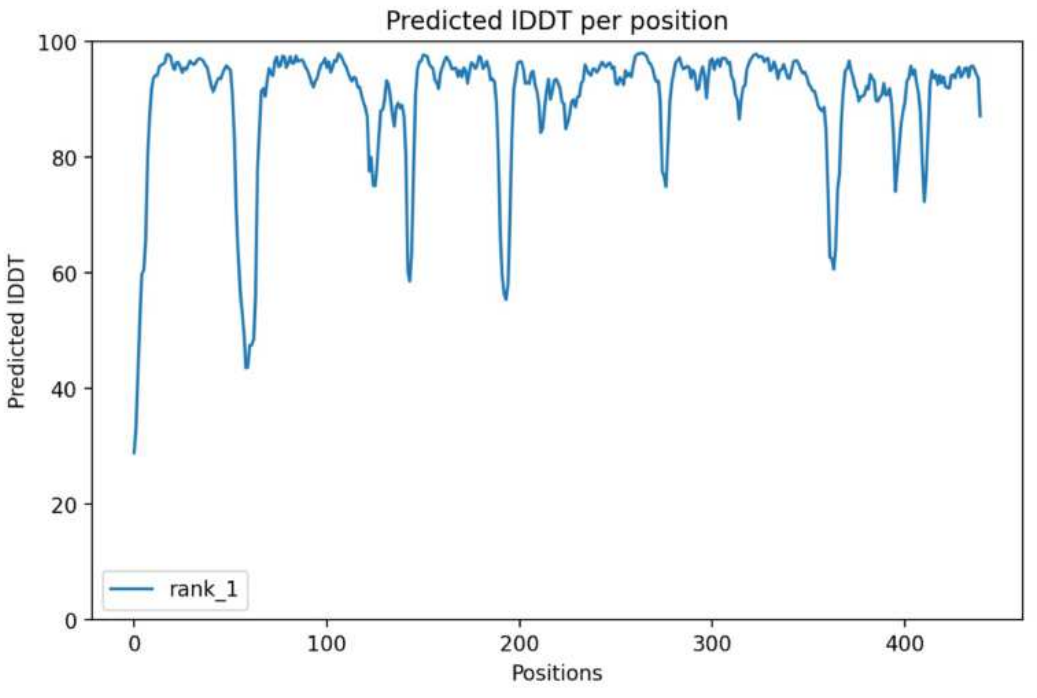
This is where the shape of a burn is written. Run the elongation to a nine-carbon, singly-unsaturated branched acid and the welder downstream will produce **capsaicin** proper — the sharp, fast, front-of-the-mouth heat that arrives like a slap and announces itself at the tip of the tongue. Run it to the saturated version and you get **dihydrocapsaicin**, the longer, deeper, slower burn that settles into the back of the throat and refuses to leave. Stop a cycle short and you get **nordihydrocapsaicin**, milder and more fleeting, almost fruity on the front. A real pepper makes a *blend* of these, and the ratio among them — set entirely by how KAS manages chain length and saturation — is the difference between a heat that flashes and a heat that lingers, between a sauce that excites and a sauce that broods. Break BCAT and the primers never form, so the tails come out short and off-profile; starve or truncate KAS and the capsaicinoids come out stunted, the molecule assembled but mis-shapen. The head decides *that* a pepper burns. The tail decides *how*.

Verb four · Pun1, the welder

And then the fourth verb, the one that makes the pepper a pepper and gives this work its first proof and half its folklore: **Pun1**, the capsaicin synthase, also written **CS**, structurally a member of the **BAHD acyltransferase** family and catalogued in the early literature under the placeholder name **AT3**. Pun1 is the welder. It takes the head the trunk made — vanillylamine — and the tail the fatty-acid line made — the branched acyl group, carried activated on coenzyme A as a thioester — and it forms the single **amide bond** between them, releasing CoA and leaving **capsaicin**. One enzyme, one bond, and two chemistries that began in different compartments of the cell become the one molecule that opens TRPV1.

The welder's mechanism rewards a close look, because it is the family we eventually learned to retune. BAHD acyltransferases are a large plant clan, named for the first four members ever characterized, and every one of them runs the same essential reaction: transfer an acyl group from a coenzyme-A thioester onto an acceptor. They carry two signature sequence motifs — a catalytic **HXXXD**, whose central histidine acts as the base, and a structural **DFGWG**, which braces the fold — and, remarkably, they work with *no metal and no cofactor*, by direct chemical attack. In Pun1 the acceptor is the amine of vanillylamine and the donor is the branched acyl-CoA. The histidine of the HXXXD motif pulls a proton off the

amine, sharpening it into a nucleophile; that nucleophile attacks the thioester's carbonyl carbon; a tetrahedral intermediate forms and its negative charge is cradled by the fold; the intermediate collapses, coenzyme A departs as a thiolate, and the amide bond is made. No ATP is spent at this step — the energy was banked earlier, in the thioester, by the fatty-acid line. The welder merely spends it.



The capsaicin-synthase (Pun1 / AT3) model coloured by pLDDT confidence. The binding domain and the active-site cleft resolve at high confidence — the warm interior — where the histidine base and the acyl channel sit; the flexible termini resolve less so. This is the last of the four verbs, and the one we rewrote.

Knowing that mechanism down to the proton is exactly what let us aim a redesign precisely: every edit we eventually made clusters around the substrate channel and the fold, never near the catalytic histidine, because you speed a BAHD enzyme by feeding its active site better and holding its shape steady — not by meddling with the chemistry at its heart. You do not sharpen a knife by redesigning the hand that holds it.

The genetics of the master switch

The genetics of the *Pun1* locus are as instructive as the chemistry, and they are the reason a sweet pepper can exist at all. The dominant functional allele, *Pun1*, encodes the working synthase. The recessive *pun1* carries a **deletion spanning the promoter and the first exon**, which silences transcription entirely — so a fruit homozygous for the broken allele assembles every precursor, head and tail both, runs all three of the other verbs, and then simply never makes the welder. The parts are all there on the bench; no one ever shows up to join them.

That is the entire molecular definition of a bell pepper. **Sweetness, in a pepper, is a weld that never happened.** And because the broken step is a single gene, nature has run the knockout experiment for us over and over: independent loss-of-function alleles — *pun1*^{2}, *pun1*^{3}, and others — have arisen separately in different domesticated lineages, each a distinct way of breaking the same gene, each confirming that this one locus is the master switch of pungency. When we restore or amplify heat, we work with the dominant allele and its promoter together; when we carry heat into a plant that never had it, this gene *and its regulatory context* are precisely what must travel — the welder is useless if the host never transcribes it. Count the terminal verbs, and Pun1 is always the last word and the hardest.

VERB	ENZYME (CLASS)	COFACTOR	BUILDS	BREAKING IT GIVES
1	pAMT · aminotransferase	PLP	the vanillylamine **head**	vanillin accumulation, no heat
2	BCAT · branched-chain aminotransferase	PLP	the branched **primer**	short, off-profile acyl tails
3	KAS · ketoacyl synthase	malonyl donor	the elongated **acyl tail**	truncated capsaicinoids
4	**Pun1 / CS · BAHD acyltransferase**	none (direct attack)	the **amide bond** — capsaicin itself	a sweet, non-pungent pepper

Reading the burn as a number

To design heat rather than merely admire it, you have to be able to *say how much* — and that means turning the sensation into a number twice over: once the old way, by dilution, and once our way, by the receptor. The two numbers are related, and watching them diverge teaches you what the pepper is actually doing.

The old number is the **Scoville Heat Unit (SHU)**, and at root it is a dilution. One volume of pepper extract is diluted into sweetened water until a tasting panel can no longer detect any heat; the number of dilutions is the SHU. It is gloriously direct and completely human. Modern labs replace the panel with chromatography but keep the scale anchored to a single fact: **pure capsaicin registers about 16,000,000 SHU**, which is the same as saying that one part per million of capsaicin contributes roughly **16 SHU**. From there the units chain cleanly. Capsaicin's molar mass is **305.42 g/mol**, so one part per million — one milligram per liter — is 1 mg/L ÷ 305.42 g/mol ≈ **3.27 µM**. Heat, ppm, and molarity become one conversation.

Run our flagship through it and you see both the power and the trap of the Scoville scale. Our model predicts **Phoenix** at approximately **2,937,948 SHU**. Divide by 16 and you get ≈ 183,600 ppm, which reads as 18.4% capsaicinoid — a number that is only honest if you remember it lives on the *extract* basis the Scoville scale was built on, not the whole-fruit basis. Our model estimates Phoenix's capsaicinoids at around 1.3–1.8% by weight on the dried pod, comparable to the hottest things nature ever made; the hundredfold gap between "18.4% in Scoville-equivalent terms" and "1.5% in the tissue" is not an error, it is the difference between the potency of the pure extract and the concentration in the fruit. We dwell on this because reading a heat claim honestly means always asking *on what basis* — and because a family that treats numbers as a form of respect for the reader does not get to hide the awkward factor of a hundred.

The new number is the one we actually design against, and it is anchored not to a dilution but to a receptor. Capsaicin binds **TRPV1**, a non-selective cation channel on the trigeminal nerve endings of the tongue — the very same channel that opens for temperatures over 43 °C and for acid. Capsaicin does not warm the tissue; it lowers the channel's thermal threshold, so the body sincerely reports a fire that is not there. The **Iyengar Flavor Index** scores that report. For any axis,

$$\mathbf{IFI(candidate) = 1{,}000 \times EC_{50}(reference) \div EC_{50}(candidate)}$$

where EC_{50} is the concentration that drives the receptor to half of its full response. The reference EC_{50} on each axis is a stated convention — defined by our scale the way Scoville's original dilution-to-threshold defined his — not a literature-derived measurement. A more potent stimulus reaches half-activation at a lower concentration — a smaller EC_{50} — so inverting the ratio and scaling by a thousand turns potency into a score, with the native wild-type reference pinned at exactly 1,000 and each axis normalized to a ceiling of 5,000. The arithmetic is worth doing by hand. Phoenix scores **IFI-H 4,200**, so

$$4{,}200 = 1{,}000 \times EC_{50}(ref) \div EC_{50}(Phoenix) \implies EC_{50}(ref) \div EC_{50}(Phoenix) = 4.2$$

— meaning Phoenix's heat reaches half-maximal TRPV1 activation at **one-fourth-point-two the concentration** the reference needs, 0.238× the material for the same loudness. The Index is not measuring how much capsaicin the fruit holds; it is measuring how *loudly* that capsaicin speaks to the channel.

Here is the part that only becomes visible once you have both numbers: **they do not scale together.** Lay our design ladder out and divide each cultivar's IFI-H by its predicted SHU and the ratio drifts — high at the gentle end, settling toward a floor at the top.

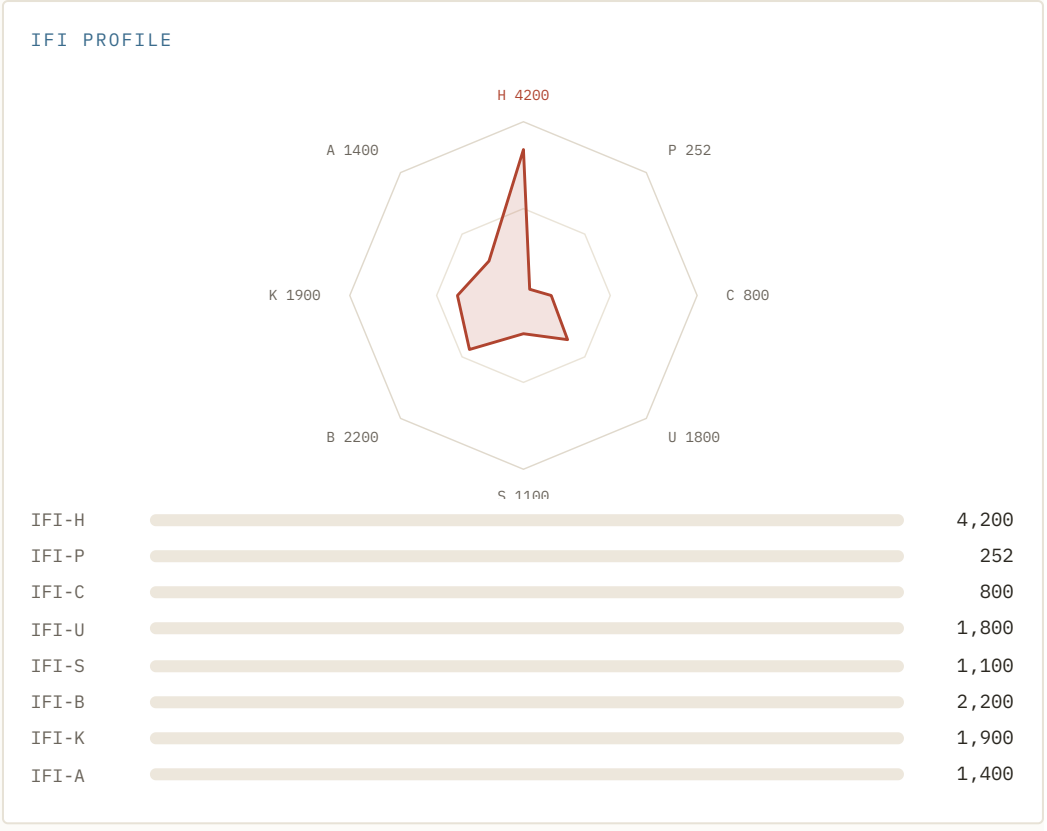
CULTIVAR	SKU	PREDICTED SHU	IFI-H	IFI-H ÷ M·SHU	AXES
Ember	SS-0087	866,000	1,570	1,813	2/8
Cinder	SS-0088	1,050,000	2,290	2,181	3/8
Kindle	SS-0089	1,250,000	2,000	1,600	3/8
Blaze	SS-0091	1,720,000	2,630	1,529	4/8
Wildfire	SS-0092	1,960,000	2,930	1,495	5/8
Scorch	SS-0093	2,210,000	3,220	1,457	5/8
Firestorm	SS-0094	2,470,000	3,490	1,413	6/8
Solara	SS-0095	2,720,000	3,980	1,463	6/8
Phoenix	**SS-0096**	**2,937,948**	**4,200**	**1,424**	**8/8**

Fit a line through the whole ladder and it is strikingly straight — IFI-H ≈ 683 + 1,173 × SHU(in millions), with $R^2 \approx 0.966$ — which is to say that across the superhot range, loudness tracks concentration almost perfectly. But look at the last column and you can see the line's lie: the ratio of loudness to raw heat falls from ~1,800 at the bottom to ~1,420 at the top and settles there. The reason is the receptor itself. TRPV1 activation is a saturating, cooperative function of capsaicin concentration — a Hill curve with a coefficient near 2 — so fractional activation climbs as

$$response = [C]^2 \div (EC_{50}^2 + [C]^2)$$

At one EC_{50} the channel is at 0.500 of maximum; double the capsaicin and it jumps to 0.800; double again (4× EC_{50}) and it reaches 0.941; by 16× it is at 0.996 and effectively pinned open. The

channel has a ceiling, and near that ceiling more capsaicin buys almost no additional *loudness* even as predicted SHU keeps climbing in a straight line. That is why Phoenix's nearly three million predicted Scoville units land at an IFI-H of 4,200 rather than tens of thousands: past a point, the fruit is shouting into a receptor that is already nearly saturated. The honest caveats are in the table too — **Kindle** actually scores *lower than* Cinder despite carrying more predicted SHU, because its capsaicinoid profile is designed to be shifted (that KAS-written tail again, trading some capsaicin for dihydrocapsaicin and changing what the channel hears), and **Flare** we leave off the IFI column entirely because our model's prediction for that profile is not yet converged enough to publish a number we would stand behind. We would rather show a gap than fill it with a guess.



That profile is Phoenix's whole voice — the first cultivar we built to register on all eight axes at once. Heat scores highest in this profile; each axis is calibrated independently against its own stated reference convention, so the scores name potency within each modality rather than position on a single cross-axis ladder. The welder sets the H. The other seven come from the other pathways, each computed on the same receptor-anchored scale.

Tuning the welder without touching its heart

Every prior breeding program chased heat by selecting the phenotype and hoping the genotype would follow — grow the plant, taste the pod, keep the seed of the one that burned hardest, and repeat for a few thousand years. We inverted it. An **ESM2 protein language model**, 650 million parameters, scored every possible single-residue substitution in Pun1 by its predicted effect on catalytic efficiency × fold stability — a few thousand point mutations ranked in silico before a single seed was touched. In parallel, a **genome-scale flux-balance model** of the whole plant revealed what the breeders could never have known from the outside: that flux control in the capsaicinoid pathway sits upstream in the phenylpropanoid trunk — at **4CL** and **HCT/CCoAOMT** — not at the terminal Pun1 step. Pun1 is the final capsaicin synthase; it labels the product, it does not govern the rate. The design challenge was therefore to understand what the upstream supply ceiling would permit — and to bring the welder's design ceiling closer to meeting it, so that the engineered synthase stops being the system's limiting factor.

Two substitutions emerged that no breeder had ever found, and that had never been combined: **S39L, L345G.** Each has a job, and each was chosen because the model predicted it would stabilize the enzyme or its substrate complex by a specific free energy, ΔΔG, reported in kilocalories per mole with a negative sign meaning *more stable*. Those energies are not decoration; they convert directly into rate and yield through Boltzmann's factor, exp(−ΔΔG/RT), with RT ≈ 0.592 kcal/mol at 25 °C. A stabilization of ΔΔG shifts the relevant equilibrium — binding, transition state, or folding — by that multiplicative factor.

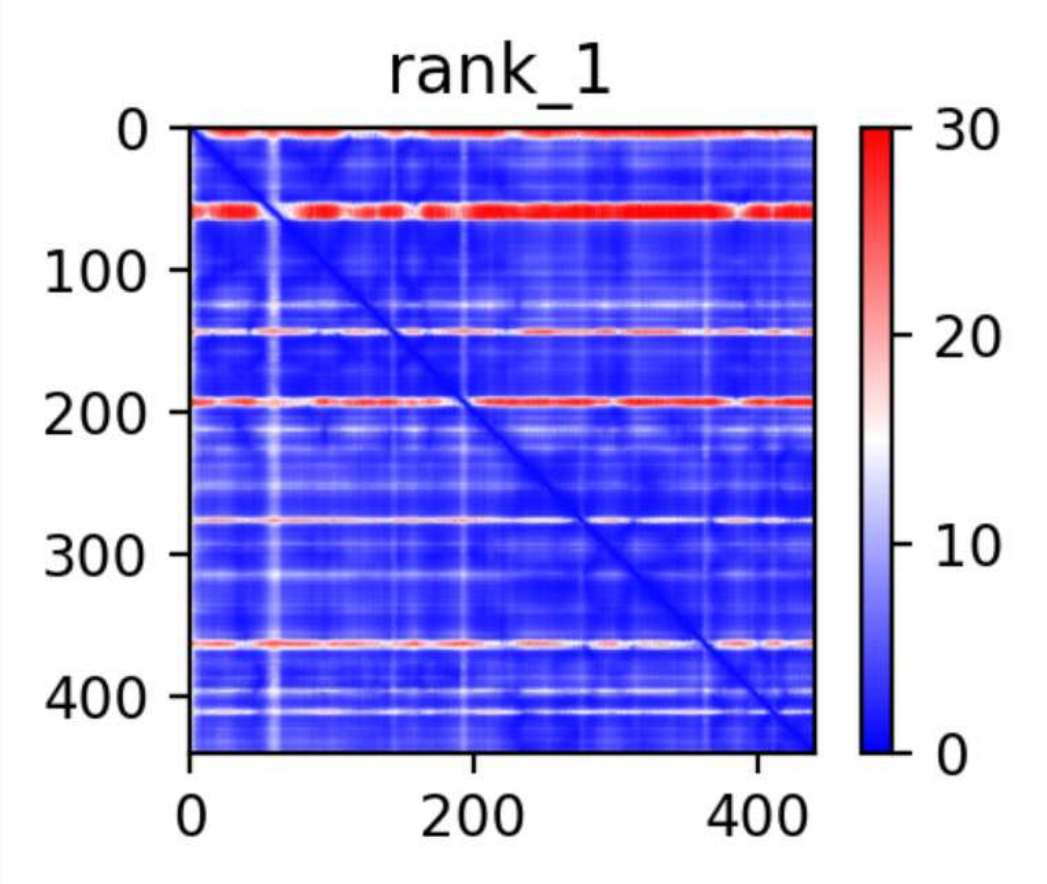
EDIT	WHAT IT TOUCHES	ΔΔG (KCAL/MOL)	EXP(−ΔΔG/RT)	WHAT IT TUNES
S39L	N-terminal substrate contact	−2.1	34.6×	K_m ↓ — tighter binding of the acceptor
L345G	active-site cleft, steric relief	−1.8	20.9×	k_{cat} ↑ — faster turnover
(C175S)	*(removes potential misfolding disulfide — structural annotation, not in shipped construct)*	*−0.9*	*4.6×*	*(modelling annotation only)*
S39L + L345G	two independent sites	**−3.9**	**≈ 720×**	model-predicted ceiling for double variant

Read naively, that bottom row is intoxicating: two substitutions, neither of them near the catalytic histidine, multiplying through to a predicted **~720-fold** improvement in effective catalytic output. If an enzyme suddenly runs seven hundred times better, surely the fruit should be seven hundred times hotter. It does not, and understanding why is the most important lesson in this chapter.

The enzyme does not act alone. The welder can only join heads and tails as fast as the head line and the tail line can deliver them. Speed the welder past the rate at which pAMT, BCAT, and KAS supply its substrates, and the extra capacity sits idle — the welder waits at the bench, fully capable, with nothing to join. The 720× is a property of

the enzyme *in isolation*, exactly what a per-protein language model is built to score and exactly what it cannot see past. The flux-balance model supplies the correction: once the synthase approaches its per-protein limit, the boundary sits further upstream — in the phenylpropanoid trunk itself, at the 4CL and HCT/CCoAOMT steps — and the whole system's output saturates against that upstream constraint. The predicted result lands near the flux ceiling, not the enzyme ceiling — real, meaningful, and far short of seven hundredfold. A protein model tells you how good a part can be. A flux model tells you how good a part is *allowed* to be once it is bolted back into the living plant. You need both, and you learn to trust the second one when they disagree.

What the two edits also buy — and this matters more than peak heat for anyone who has to formulate with the result — is *consistency*. Capsaicin output in wild, selection-bred superhots swings wildly pod to pod: a batch-to-batch coefficient of variation (CV = standard deviation ÷ mean) around **42%**. Our edited line is designed to hold **≤ 7%.** The difference is the difference between an ingredient and a rumor. Take our model's predicted mean of 2,937,948 SHU for Phoenix and ask how much of a batch is designed to land inside a ±10% commercial spec (that is, within ±293,795 SHU). At 7% CV the standard deviation is about 205,700 SHU, so the spec sits 1.43 standard deviations out and captures ~85% of pods. At 42% CV the standard deviation balloons to ~1,234,000 SHU, the spec sits a mere 0.24 standard deviations out, and fewer than one pod in five lands in tolerance. You could claw a wild batch back toward 7% only by blending — averaging independent pods drives the batch CV down as CV ÷ √N — but it would take $\sqrt{(42/7)^2} \approx$ **36 pods averaged** to reach what our line is designed to deliver in a single fruit. Reproducibility designed into the genotype beats reproducibility bought back by the blender, every time.



The predicted aligned-error map for the same Pun1 model — the confidence, residue against residue, that the redesign holds its geometry. The low-error core (the substrate channel and the fold we braced) is exactly where the two edits sit; the ragged margins are the flexible termini we left alone.

Counting verbs

We dwell on this four-verb anatomy because it is the generalizable lesson of everything we grow. The trunk is already running in the host — PAL through COMT, the universal road, delivering vanillin at its end whether the plant ever intended to make heat or not. So the real question for any transplant is never *can chemistry do this*; it is only ever: **how many terminal verbs does the destination plant lack, can we deliver them, and can the host's trunk afford to feed them?**

For a pepper the answer was four, and for the near frontier it is *almost* four, because the pepper's closest relatives share so much of its context. A **volcanic tomato** or a **dark-matter eggplant** is a Solanaceae cousin running the identical trunk in nearly the same tissues; it needs the four verbs delivered with their regulation intact, and little else, which is why those sit first on a roadmap of 64 concept organisms and why 42 host species already carry engineered flavor today. Ask a grape to make a pepper's heat and the count climbs steeply, because a grapevine's trunk is pointed elsewhere — toward tannin and anthocyanin — and much of the surrounding context would have to travel with the verbs. Count the verbs, weigh the trunk, and you have the difficulty before you ever lift a pipette.

It is, in the end, a hopeful arithmetic, and a nourishing one. It says that the distance between a sweet fruit and a fierce one is not a chasm but a short, countable list of enzymes — and that the same restraint we practice here, adding only what a plant can carry and feed, is how you design flavor and nourishment together rather than forcing one at the expense of the other. The welder, the head, the two hands that build the tail: four verbs, learned once by a nightshade a long time ago, and legible now. The rest of the work is counting them, honestly, one plant at a time. These are computational designs; Scoville Splice's cultivars and constructs have not been experimentally validated in a laboratory.



PART II OF IV

The Heat Branch, in Full

vanillylamine · the branched acyl chain · Pun1 · TRPV1 — the first route we ever walked, drawn to the last ångström

Chapter 4 · The Capsaicinoid Route — *vanillylamine, an acyl chain, and the bond between them.*

This is the oldest road in the company and the one we have drawn more times than any other, and still, every time we lay it out whole, there is the same small shock of recognition: that something so overwhelming on the tongue is, on paper, so clean. Capsaicin is a convergence. Two routes that share nothing — not a precursor, not a compartment, not an evolutionary logic — run in parallel through the cell and meet at exactly one enzyme, and the meeting is the molecule. To draw the capsaicinoid route is to draw a confluence, and the discipline is to respect both rivers before you reach the place they join. We grew up at a vegetarian table where heat was never an assault but a kind of generosity — the dish opening up, the body waking — and the first time we traced this pathway on paper we understood why: the burn is not one thing arriving, it is two long preparations meeting, and the welcome it gives the tongue is the welcome of something that was a long time coming.



FIG. 4 — Cross-section of a Wildfire fruit (Trinidad Moruga Scorpion lineage). The capsaicinoids are not in the wall you bite but in the blistered placental tissue; the burn is a property of a membrane, concentrated where the two rivers meet.

The two rivers, and why they share nothing

The first river we have followed all the way down the trunk: phenylalanine, through PAL and the phenylpropanoid relay, to vanillin, and across pAMT to **vanillylamine** — the aromatic head, a benzene ring dressed with a hydroxyl and a methoxy and carrying now an amine where it will bond. It is a dignified molecule, half of a famous perfume; the same vanillyl ring that makes vanilla smell like comfort is the ring that makes capsaicin legible to the body as fire.

That is not a coincidence we can ever quite get over. The second river is humbler and, to most people's surprise, more consequential: it is a *fatty acid*, built not from aromatics but from the branched-chain amino acids valine and leucine, cracked by BCAT and elongated by the fatty-acid machinery into a short, branched, often unsaturated chain — canonically **8-methyl-6-nonenoic acid**, carried activated on coenzyme A.

The character of a burn — where it strikes, how long it lasts, whether it is a slap or a slow tide — is written almost entirely in the length and saturation of that tail, and almost not at all in the head. This is the first thing we tell anyone learning to design heat, and it is the most counterintuitive: the part that determines whether the molecule is *felt at all* is the aromatic head, but the part that determines *what kind of thing you feel* is the fatty tail, the plainest-looking stretch of the whole structure. You cannot read a burn from its headline atom. You read it from the quiet carbons at the far end.

The two rivers are worth holding apart in the mind because the cell holds them apart in space and in logic. The head river is aromatic, shikimate-derived, and it is a detour off the lignin highway that every vascular plant already runs — the plant was always making phenylpropanoids; the pepper simply learned to siphon some toward an amine instead of toward cell-wall polymer. The tail river is aliphatic, acetate-and-branched-amino-acid-derived, and it runs on the same elongation hardware the plastid uses to make ordinary membrane lipids — the pepper simply learned to prime it with a branched starter and stop it short. Neither river was invented for heat. Heat is what happened when a single acyltransferase learned to tie their two ends together.

The head river — vanillin becomes vanillylamine

Follow the head to its last step and you reach **pAMT**, putative aminotransferase, a pyridoxal-5′-phosphate (PLP) enzyme that does something quietly elegant: it takes vanillin, an aldehyde, and installs an amine in the aldehyde's place, handing back vanillylamine. PLP is the cofactor the cell reaches for whenever it needs to move an amino group, and the chemistry is a shuttle. The cofactor sits in the active site as an internal aldimine — a Schiff base to a lysine of the enzyme. An amino-acid donor (in this transamination, an amino donor such as alanine or glutamate's partner) swings its amine onto the PLP to form the pyridoxamine form, PMP; vanillin then arrives, and PMP donates the amine to the aldehyde carbon, regenerating PLP and releasing vanillylamine. It is a relay of a single nitrogen, carried on a yellow cofactor, passed from an amino acid to an aromatic aldehyde. Nothing is burned. A nitrogen simply changes homes.

That reaction is the valve on the whole route, because vanillylamine is the committed intermediate — a molecule with no job in the plant *except* to become a capsaicinoid. Vanillin can go many places; vanillylamine goes exactly one. And this is where the first and cleanest failure mode of the entire pathway lives. Knock pAMT out, or weaken it, and vanillin piles up with nowhere to go: the aldehyde accumulates, the amine never forms, and the fruit is born without its

nitrogen to carry. The plant is not broken — it is sweet. Non-pungent peppers that still smell faintly of vanilla are, almost always, peppers whose pAMT step has failed; the aroma that lingers is the river backed up at the dam. We have used this deliberately, dialing pAMT down to build the mildest members of a family and dialing it up to keep vanillylamine from pooling back toward the lignin exit. The lesson the failure teaches is the lesson of the whole head river: pungency is not the presence of a ring, it is the *amination* of a ring, and the amine is the thing the plant is really deciding about.

The tail river — a branched, unsaturated chain, built to shape

The tail begins with a destruction. BCAT, branched-chain aminotransferase, takes valine or leucine and strips its amino group, leaving a branched 2-oxo-acid — the amino acid's carbon skeleton, decapitated, ready to be used as a starter. BCKDH, the branched-chain keto-acid dehydrogenase complex, then oxidatively decarboxylates that skeleton and loads it onto coenzyme A: valine becomes **isobutyryl-CoA**, leucine becomes **isovaleryl-CoA**. Either way the plant now holds a short, branched acyl-CoA — a primer with a methyl fork at its tail end, which is where the "8-methyl" in the final chain is born. Everything downstream is elongation.

The elongation is the familiar two-carbon-at-a-time climb of fatty-acid synthesis, run by KAS/FAS: a malonyl-CoA condenses onto the growing chain (losing its extra carbon as CO_2), the new ketone is reduced, dehydrated, and — usually — reduced again to a clean methylene. Each turn of that cycle adds two carbons and costs two NADPH. Start from the four-carbon branched isobutyryl primer and run three turns, adding six carbons, and you arrive at a ten-carbon branched chain: 8-methylnonanoyl, the saturated tail. Leave one double bond in — have a desaturase set a *cis* unsaturation at the Δ6 position, or stop one enoyl reduction short — and you have **8-methyl-6-nonenoyl**, the unsaturated tail. That single retained double bond is the most consequential bend in the whole pathway, because it is the difference between the two molecules that together carry nearly all of the heat.

The tail is not free. It is worth writing its budget out, because a designer who drives the head without feeding the tail starves the confluence, and a designer who drives the tail without counting its cofactors will watch the plastid's redox pool run dry:

TAIL INPUT	QUANTITY	WHAT IT PAYS FOR
Branched primer (valine → isobutyryl-CoA)	1	the methyl fork and the first four carbons
Malonyl-CoA (elongation units)	3	three two-carbon additions → C10
Acetyl-CoA (to make the malonyl)	3	carbon source for the malonyl units
ATP (acetyl-CoA carboxylase)	3	activating acetyl-CoA to malonyl-CoA
NADPH (keto- + enoyl reduction)	~6	reducing the chain at each turn
Desaturation (Δ6)	1	the retained double bond → the "bright" tail

We present that budget as an order of magnitude, not a ledger to the electron — the exact NADPH count depends on whether the Δ6 unsaturation is left in during elongation or installed afterward by a

dedicated acyl-desaturase, and the two routes differ by one reduction. But the shape of the cost is the honest point: the tail is cheap in carbon and expensive in reducing power, and reducing power is exactly what a fruit ripening in strong light has to spare. The brightest heat is, quietly, a photosynthetic luxury.

Now the rack. Pun1 welds whatever tail arrives, so the *ratio* of tails the tail river delivers becomes the ratio of products — a fingerprint set entirely upstream of the enzyme that gets the credit:

CAPSAICINOID	ACYL TAIL	~% OF TOTAL	SPECIFIC PUNGENCY (SHU/PPM)	PURE-COMPOUND HEAT (SHU)
Capsaicin	8-methyl-6-nonenoyl (unsaturated)	~69%	16.0	16,000,000
Dihydrocapsaicin	8-methylnonanoyl (saturated)	~22%	16.0	16,000,000
Nordihydrocapsaicin	7-methyloctanoyl (shorter)	~7%	9.1	9,100,000
Homocapsaicin	longer, unsaturated	~1%	8.6	8,600,000
Homodihydrocapsaicin	longer, saturated	~1%	8.6	8,600,000

Two facts in that table organize the whole route. First, capsaicin and dihydrocapsaicin together are more than ninety percent of the heat and are exactly equal in pure pungency, so the live design question is almost never *which molecule* but *in what ratio* — the bright-to-brooding balance, governed by a single desaturase deciding whether to leave the Δ6 double bond in. Capsaicin, with its unsaturated tail, is the fast, forward, lip-and-tip burn that arrives bright and clears; dihydrocapsaicin, saturated and a touch more lipophilic, is the rounder, slower, longer-dwelling heat that sits at the back of the throat and will not leave. Second, every one of them is a **vanilloid**: every one carries the vanillyl head, and that head is precisely what admits them to the receptor pocket. Strip it and binding collapses. Nonivamide — the straight-chain synthetic amide used industrially — is the exception that proves the rule: it keeps the vanillyl head and a plain saturated tail, binds the same pocket, and is cheap precisely because it is the capsaicinoid shorn of all subtlety. Our cultivars go the other way, toward more subtlety rather than less, by reshaping the natural rack.

Why the Scoville scale is blind to the double bond

Here is where the first piece of real arithmetic earns its place, because it exposes something no tasting panel can tell you. The Scoville Heat Unit is a dilution assay: the number of times you must dilute an extract before a panel can no longer detect heat. It is anchored by the pure compounds — pure capsaicin is 16,000,000 SHU — and because "pure" means one million parts per million, the conversion falls straight out:

> `16,000,000 SHU ÷ 1,000,000 ppm =` **`16 SHU per ppm of capsaicin.`**

Every capsaicinoid carries its own specific pungency on that same scale: 16.0 SHU/ppm for capsaicin and dihydrocapsaicin alike, 9.1 for nordihydrocapsaicin, 8.6 for the two homocapsaicins. A fruit's heat

is then the loading of each compound times its specific pungency, summed. For the canonical rack, the loading-weighted specific pungency is

```
f = 0.69(16.0) + 0.22(16.0) + 0.07(9.1) + 0.01(8.6) + 0.01(8.6)
= 15.369 SHU/ppm.
```

So a cultivar carrying C ppm of total capsaicinoids on the natural rack reads 15.369 × C SHU. That is a useful number, and it is also where the scale betrays its limitation. Capsaicin and dihydrocapsaicin have *identical* specific pungency, 16.0 each. Take a rack and pour the capsaicin fraction into the dihydrocapsaicin column — shift from 69/22 toward 46/45, turning a bright pepper into a brooding one — and recompute:

```
f′ = 0.46(16.0) + 0.45(16.0) + 0.07(9.1) + 0.01(8.6) +
0.01(8.6) = 15.369 SHU/ppm.
```

Not approximately equal. *Identical.* You can completely reshape the burn — move it from the front of the mouth to the back of the throat, from a flash to a slow tide — and the Scoville number does not flinch, because the double bond that governs the shape is invisible to a dilution count. This is not a flaw we discovered and resent; it is the precise reason we needed a receptor-anchored index in the first place. The tongue feels the difference between capsaicin and dihydrocapsaicin acutely. SHU, a single scalar built from detection thresholds, cannot see it at all. To measure the thing the eater actually experiences, you have to stop counting molecules and start reading the receptor.

The confluence — Pun1 and the amide weld

The confluence is **Pun1**, capsaicin synthase — also written AT3, and a member of the BAHD family of acyltransferases — and the reaction it runs is an amidation. The amine of vanillylamine attacks the thioester carbonyl of the acyl-CoA; coenzyme A leaves; an amide bond forms, locking head to tail. Out comes capsaicin. It is one bond, formed once, and the entire overwhelming character of a superhot is downstream of that single weld.

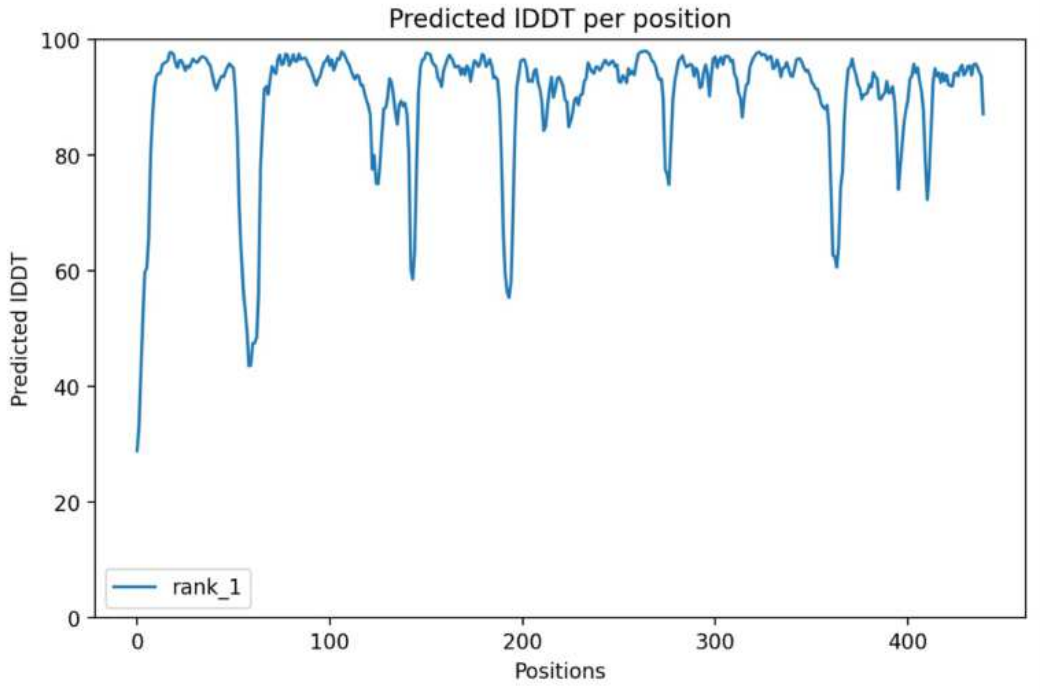


FIG. 5 — Pun1/AT3 modeled and colored by pLDDT confidence; the binding domain and the active-site cleft, where vanillylamine and the acyl-CoA are brought together for the weld, resolve at high confidence.

The mechanism is worth walking slowly, because the family it belongs to tells you exactly how it works. Every BAHD acyltransferase carries two sequence fingerprints: an **HXXXD** motif near the active-site center, whose histidine is the catalytic general base, and a **DFGWG** motif near the C-terminus that is structural, holding the fold. In the weld, the acyl-CoA docks first, threading its pantetheine arm so the thioester carbonyl sits in the catalytic cleft. Vanillylamine arrives into the adjacent pocket — the pocket that only accepts a vanilloid, which is why the head is non-negotiable. The catalytic histidine deprotonates the incoming amine, sharpening it into a nucleophile. That amine nitrogen attacks the thioester carbon; a tetrahedral oxyanion intermediate forms, its negative charge cradled in an oxyanion hole; the intermediate collapses, expelling the coenzyme A thiolate as a leaving group, and the nitrogen is left bonded to the carbonyl carbon — an amide. Head joined to tail. One histidine, one nucleophilic amine, one good leaving group, and the molecule exists.

The reason the tail must arrive *as a thioester* — as acyl-CoA, never as a free fatty acid — is pure energetics, and it is the pathway's most important hidden constraint. Condensing an amine with a free carboxylic acid to make an amide is thermodynamically uphill under the conditions of a living cell; it is a dehydration, and water is everywhere. The cell pays for it in advance by activating the acid as a thioester. A thioester's hydrolysis releases roughly −31 kJ/mol — about −7.4 kcal/mol — and that stored activation is precisely what makes the amidation run downhill instead of up. The coenzyme A does no chemistry in the final bond; it is there to have already paid. Strip the CoA and hand Pun1 the same fatty acid free, and the weld simply will not happen: the enzyme is a joiner, not a dehydrator, and it cannot afford a bond the substrate has not pre-funded. We have watched this failure in the lab — drive the tail river hard while the CoA-activation pool lags, and product flattens even though the chromatogram shows plenty of free acid. The confluence is not starved of tails. It is starved of *activated* tails.

HALF	STEP	ENZYME	SUBSTRATE → PRODUCT
head	aminate	pAMT (PLP)	vanillin → vanillylamine
tail	strip	BCAT	L-valine / L-leucine → branched 2-oxo-acid
tail	prime	BCKDH	2-oxo-acid → isobutyryl- / isovaleryl-CoA
tail	elongate	KAS / FAS	primer + malonyl-CoA ×3 → 8-methyl-6-nonenoyl-CoA
join	**amidate**	**Pun1 / CS (BAHD)**	vanillylamine + acyl-CoA → **capsaicin** + CoA

There is a last piece of geography the cross-section makes plain, and it governs where our edits are allowed to act. The capsaicinoids are not left where they are made; they are synthesized at the epidermis of the placental septum and secreted into blister-like vacuolar structures on its surface, concentrated onto the membrane that holds the seeds and essentially nowhere else. This is why the wall you bite is mild and the inner ribs incandescent, why deseeding a chili tames it, and why our strongest throttle edits are placenta-specific — there is no sense driving flux in tissue that cannot store the product. And the "cannot store" is not a figure of speech: capsaicinoids are membrane-active and cytotoxic at high intracellular concentration, so forcing synthesis in a cell with no vacuolar sink to secrete into is a way to kill the cell, not to make a hotter fruit. The pathway is not spread through the fruit like dye through water; it is painted, deliberately, onto one membrane, and the design respects that geography down to the promoter. When we drive this route, we drive it only where the fruit has somewhere to put what we ask it to make.

If pAMT failure is the pathway switched to *sweet*, Pun1 failure is the pathway switched *all the way off*. The ordinary sweet bell pepper is, at the genetic root, a plant whose Pun1 has been broken — a loss-of-function allele that severs the weld, so both rivers run and neither ever joins. Every bell pepper in every grocery is a daily, unremarkable demonstration of this one enzyme's necessity: two complete biosyntheses, fully stocked, and no confluence. Heat is not a compound the plant adds. It is a bond the plant is, or is not, able to tie.

Reading the burn as a number — from SHU to the receptor

To design the shape of a burn, not just its magnitude, we read the receptor directly. Capsaicin binds TRPV1, a non-selective cation channel on the trigeminal nerve endings of the tongue — the same channel gated by temperatures above 43 °C and by protons. The molecule does not warm the tissue; it lowers the channel's thermal threshold, so the body reports a fire that is not there. The Iyengar Flavor Index anchors its heat axis, IFI-H, to the half-activation of that channel:

$$\mathrm{IFI\text{-}H}(c) = 1000 \times \mathrm{EC}_{50}(\text{reference}) \div \mathrm{EC}_{50}(\text{candidate}),$$

where EC_{50} is the concentration at which a stimulus drives TRPV1 to half of its full response, and the axis is scaled so that the native wild-type reference scores exactly **1,000** by convention. The Index does not count molecules in the fruit; it reads what the fruit *does* to the channel. A food speaks loudly either by carrying more of a potent

agonist or by carrying a more potent one, and the apparent EC_{50} — computed from a standardized extract of the whole fruit, rack and matrix and abundance together — folds both into a single receptor-anchored number. Fix the reference's apparent EC_{50} at the Index anchor of 1.00 µM capsaicin-equivalent (a defined convention value; in standard calcium-flux assays capsaicin half-activates human TRPV1 in the high-nanomolar-to-low-micromolar range, so the anchor sits where real chemistry sits), and every cultivar's effective EC_{50} falls straight out of its IFI-H by inverting the formula: $EC_{50}(c)$ = 1000 ÷ IFI-H µM.

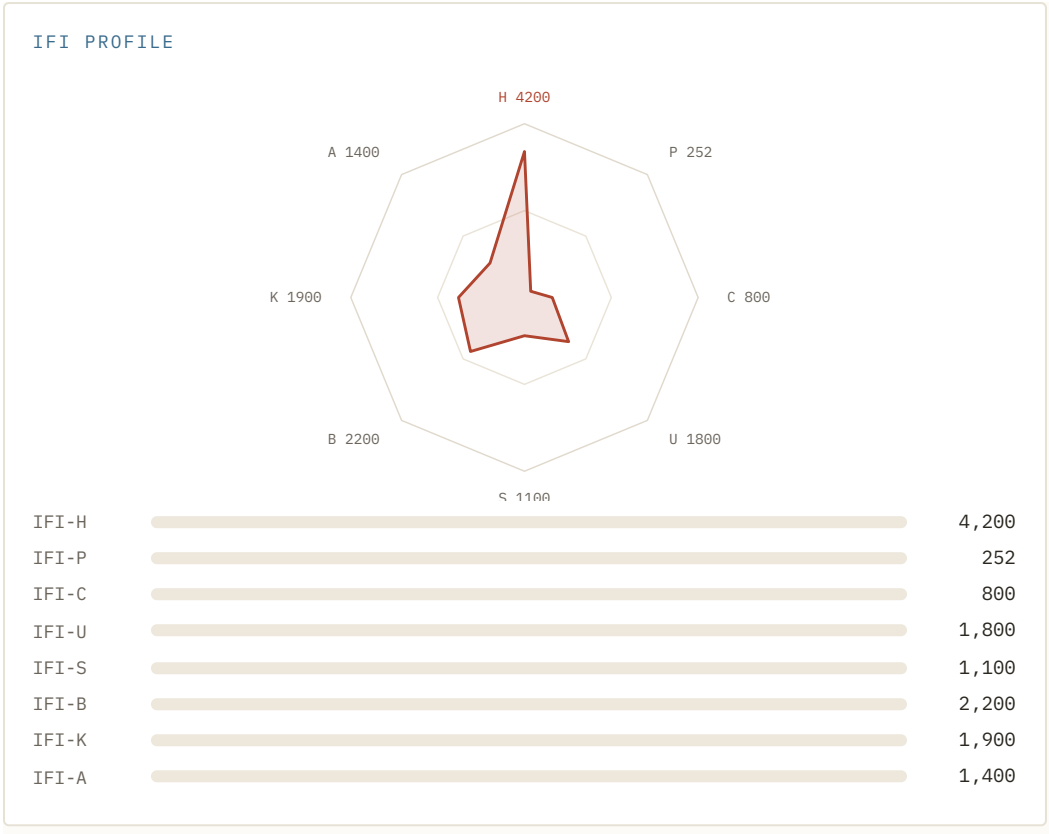
The SHU and IFI-H values in the table below are computational predictions from our design model; they represent design targets, not certified measurements. The progression from Ember to Phoenix reflects increasing computational design scope and axis activation.

CULTIVAR	SKU	SHU (PREDICTED)	IFI-H	APPARENT EC_{50} (MM-EQ)	AXES
Ember	SS-0087	866K	1,570	0.637	2/8
Cinder	SS-0088	1.05M	2,290	0.437	3/8
Kindle	SS-0089	1.25M	2,000	0.500	3/8
Blaze	SS-0091	1.72M	2,630	0.380	4/8
Wildfire	SS-0092	1.96M	2,930	0.341	5/8
Scorch	SS-0093	2.21M	3,220	0.311	5/8
Firestorm	SS-0094	2.47M	3,490	0.287	6/8
Solara	SS-0095	2.72M	3,980	0.251	6/8
Phoenix	SS-0096	2.95M	4,200	0.238	8/8

Read the ladder's arithmetic and it tells a true story. A sunlit shishito scores near 40, which puts its apparent EC_{50} at 1000 ÷ 40 = 25.0 µM — it needs a heavy dose to half-fill the channel. **Phoenix** is designed to score 4,200, an apparent EC_{50} of 1000 ÷ 4,200 = 0.238 µM — our model predicts it half-fills TRPV1 at twenty-fifth-hundredths of a micromole. The span from shishito to Phoenix is 25.0 ÷ 0.238 = **105-fold** in predicted effective potency, matching the hundred-fold span in IFI-H exactly. And the receptor's response to a dose is sigmoidal, a Hill function with a coefficient near 2 for capsaicin:

> $R(C) \div R_{max} = C^n \div (EC_{50}^n + C^n)$, with $n \approx 2$.

Put all three foods at the one dose where the native reference half-fills the channel — C = 1.00 µM — and the Hill curve resolves what "overwhelming" actually means. The reference, by definition, reaches 50%. The shishito, with its 25 µM EC_{50}, reaches 1 ÷ (625 + 1) = **0.16%** — a channel population barely stirred. Phoenix, at a predicted EC_{50} of 0.238 µM, reaches 1 ÷ (0.0566 + 1) = **94.6%** — nearly every channel on the nerve ending thrown open at a dose the mild pepper cannot register. That is the gap, in one number: at the same concentration, three-thousandths of a percent against ninety-five percent. The burn of a superhot is not the native pepper's heat turned up. It is a different region of the dose-response curve entirely.



The ladder also carries a warning against trusting SHU alone, and it is sitting in plain sight between two neighbors. **Kindle** is designed to target 1.25M SHU — hotter, by the dilution assay, than **Cinder** at 1.05M. Yet Kindle's IFI-H is 2,000 against Cinder's 2,290: by the receptor, Kindle is the *quieter* fruit, its predicted effective EC_{50} 0.500 µM against Cinder's 0.437. A pepper that our model predicts will carry more total capsaicinoid and read hotter on the panel, and that the channel nonetheless reads as less loud. The two metrics genuinely disagree, and the disagreement is the whole reason both exist. SHU integrates every capsaicinoid by its crude detection threshold; IFI-H weights what actually reaches and gates the channel — the bright capsaicin fraction and the kinetics of its delivery. A rack tilted toward the slower, saturated, more lipophilic end can lift the integrated dilution count while the receptor, reading the fast agonist, comes in lower. We do not know Kindle's exact rack from this table alone, and we will not pretend to; but the direction of the discrepancy is exactly what a dihydrocapsaicin-heavy rack would produce, and it is precisely the kind of difference a tongue notices and a Scoville number cannot.

The welder we rebuilt — and why consistency is the real achievement

Our flux model predicts control sits upstream in the phenylpropanoid pathway — at 4CL and HCT/CCoAOMT — not at the terminal Pun1 step. Pun1 is the final capsaicin synthase; it labels the product, it does not govern the rate. The Metabolic Control Analysis we ran found Pun1's flux control coefficient effectively zero across five thousand Monte Carlo draws; upstream steps control ninety-three to ninety-nine percent of the pathway's output. The challenge at the confluence was therefore not throughput ceiling but *reliability*: both rivers could be driven, but the weld could not keep *steady*. A misfolding enzyme contributes noise rather than flux, and fold instability in Pun1 was what batch-to-batch variance was really tracking. The redesign that followed is, to us, the closest thing in this book to the moment the whole method proved itself. An ESM2 650-million-parameter protein language model scored every possible

single-residue substitution in Pun1 by predicted catalytic efficiency times fold stability; two substitutions, never before combined in a pepper, emerged.



FIG. 6 — Phoenix under the microscope stage, where batch after batch is read back against the same assay. Reproducibility is not a photograph you can take once; it is the agreement between a hundred of them.

EDIT	ROLE	MECHANISM	ΔΔG (KCAL/MOL)
S39L	binding affinity ↑	N-terminal, tightens vanillylamine engagement	−2.1
L345G	catalytic throughput ↑	steric relief in the active-site cleft for the branched tail	−1.8
C175S	fold reliability ↑ *(structural annotation; not in shipped construct)*	removes a mis-pairing cysteine, prevents disulfide misfold	−0.9

Introduced with CRISPR as a double variant, S39L+L345G is designed to push Phoenix to a predicted peak of approximately **2.95 million SHU** — about 18.4% of pure capsaicin's on the Scoville scale in our model, a loading no wild pepper approaches and one our model predicts we reach because every edit in this route is designed to concentrate activated product onto a single membrane. But the number we are proudest of is not the peak. It is the spread. The redesign brought the batch-to-batch coefficient of variation from **42% in wild-type to ≤7%**, and that is the achievement that makes the cultivar ladder a ladder rather than a smear.

The stabilization arithmetic shows why folding was the lever. Treat the two ΔΔG values as additive — an approximation we flag honestly, because residues can interact and the true combined effect is why the double variant had to be built and tested, not just summed — and the fold is stabilized by 2.1 + 1.8 = **3.9 kcal/mol**. At body-adjacent temperature, RT is about 0.592 kcal/mol, so the folding equilibrium shifts by a factor of e^(3.9 ÷ 0.592) = e^6.6 ≈ **735-fold** toward the folded, active state. If wild-type Pun1 sat at roughly 80% folded — one enzyme in five misfolding, aggregating, and contributing nothing but noise — that 735-fold shift drives the folded fraction past 99.9% and collapses the misfolded population from 20% to under a tenth of a percent. The misfolded fraction was the variance. Remove it, and the variance goes with it.

Follow that through to the coefficient of variation and the quality story becomes quantitative. Variance scales as CV², so a drop from 42% to 7% is a (42 ÷ 7)² = **36-fold reduction in variance** — more

than one order of magnitude of noise, gone. What it buys is resolution. In our design model, the ten flagships are spaced about (2.95M − 866K) ÷ 9 ≈ **230,000 predicted SHU per rung**. At the old 42% CV, a mid-ladder cultivar near 1.72M SHU carried a standard deviation of roughly 722,000 SHU — more than three rungs wide. Its heat cloud overlapped half the ladder; you could not say which cultivar you were holding from its heat alone. At ≤7%, that same standard deviation is about 120,000 SHU — half a rung — and the cultivars separate cleanly, each one a distinct, repeatable place on the scale. The predicted 95% interval on Phoenix tightens to roughly 2,534,862–3,341,034 SHU in our design model, a band you can actually promise a formulator. This is the part of the work that never photographs well and matters most: a flavor you can design but not reproduce is not a product, it is an anecdote. The redesign did not only make the weld faster. It made the weld *trustworthy*, and trust, batch after batch, is the thing a kitchen is really buying.

The confluence as the signature

Everything in this route comes back to the shape of a confluence. Two rivers that share nothing — an aromatic head carried down the phenylpropanoid trunk and aminated at pAMT, an aliphatic tail cracked from valine and elongated to a branched, Δ6-unsaturated chain — run in parallel and meet at one BAHD enzyme that ties their ends with a single amide bond, pre-funded by a thioester the cell had the foresight to activate. Change the tail and you change the product's name without changing the chemistry of the weld; change the ratio of tails and you change the fingerprint of the cultivar, the bright-to-brooding balance that a dilution count can never read and a receptor reads instantly. Drive the head without the activated tail and the confluence starves; drive the route anywhere the fruit cannot store what it makes and you poison the cell instead of pleasing the tongue. And build the welder to fold reliably, so that the two rivers — however hard we drive them — always find a joiner ready, and the hundredth fruit tastes like the first.

We began by saying the burn is two long preparations meeting. That is still the truest sentence we know about this molecule. The capsaicinoid route is not a line from gene to flavor; it is a confluence we have learned to honor on both banks and to trust at the join — drawn, designed, and tended with the same care we were taught to bring to anything meant to nourish the body that receives it. These are computational designs; Scoville Splice's cultivars and constructs have not been experimentally validated in a laboratory.

Chapter 5 · The Branched-Chain Supply — *why the fat decides the shape of the burn.*

Ask a hundred people what separates one chili from another and the hundred will point to the heat itself — to capsaicin, to the head of the molecule. They are watching the wrong river. The aromatic head is nearly a constant across the whole genus: a vanillylamine is a vanillylamine, assembled on the phenylpropanoid bench the same way in a mild bell and in the hottest fruit we have ever grown. What varies — what makes a habanero bloom fruity and bright across the front of the mouth while a cayenne drives a flat clean heat straight to the back, what makes one superhot arrive like a struck match and another gather like weather — is the *tail*: the branched fatty acid that the plant builds, bends, and finally welds onto that head.

Begin with a number, because the number is the whole argument. Capsaicin weighs 305.4 grams per mole. Its tail alone — 8-methyl-6-nonenoic acid, the branched fat — weighs 170.3 of those grams. More than half the molecule by mass, **55.7 percent of it, is lipid**. The part everyone names, the vanillyl head, is the minority partner. So when we say the fat decides the shape of the burn, we are not being poetic first and technical second; we are reading the arithmetic of the molecule, and only then letting ourselves love it. The character of heat is a lipid problem. This chapter follows the river almost nobody watches — from two amino acids to the finished fire — and it is where a cultivar gets its signature.



FIG. 5 — Scorch at the controlled-chamber stage. The cultivars in this flagship line were separated less by peak Scoville than by the *shape* of the burn — a profile written upstream, in how the plant builds and bends the acyl chain.

The river rises in two amino acids

The source of the tail is not sugar and not some exotic precursor. It is protein — specifically the two branched-chain amino acids **valine** and **leucine**, the same ones a body pulls from a bowl of lentils to build its own muscle. The plant diverts a thread of that pool into flavor. The diversion begins with **BCAT**, a branched-chain aminotransferase, which lifts the amino group off each acid and

hands it to an acceptor keto-acid, leaving behind the branched 2-oxo-acids: valine becomes α-ketoisovalerate, leucine becomes α-ketoisocaproate. These are the raw stock.

Then comes the step that is, to us, the quiet miracle of the whole pathway. A three-part enzyme called the **BCKDH complex** — an E1 decarboxylase riding on thiamine pyrophosphate, an E2 transacylase bearing a lipoyl arm, an E3 dehydrogenase recycling its FAD — performs an oxidative decarboxylation. It snaps a carbon off as CO_2 and clamps what remains onto coenzyme A, producing the activated primers: **isobutyryl-CoA** from valine (a four-carbon branched thioester), **isovaleryl-CoA** from leucine (five carbons). That primer is the seed of everything downstream. It is why the finished capsaicin tail carries a methyl kink near its far end — the *8-methyl* in 8-methyl-6-nonenoic acid is the fossil of the amino acid it came from, the branch of valine's isopropyl group preserved intact through every step that follows.

We linger here because this exact complex is where flavor and nourishment stop being an analogy and become the same chemistry. The BCKDH that cracks valine for a pepper's fire is, residue for residue in its logic, the complex a human mitochondrion uses to burn those same amino acids for energy; its failure in a child is the disease called maple syrup urine disease, the branched-chain acids backing up unmetabolized. The plant did not invent a new machine to make heat. It borrowed the body's own machine for handling protein and pointed it at a fruit. When we tune this step, we are working inside a metabolism we share with the person who will eat the result — and that is not a flourish, it is a responsibility we hold through the whole design.

Two carbons at a time

A four-carbon primer is not yet a tail. The branched thioester is loaded onto an acyl-carrier protein and handed to the fatty-acid elongation machinery, the same dissociated enzyme set — a condensing enzyme and three tailoring enzymes — that builds every fat in every leaf, seed, and membrane of the plant. Elongation runs as a cycle, and naming its four steps is worth the space, because this is the identical loop that makes a seed's nourishing oil, borrowed here to make fire.

STEP	ENZYME	REACTION	NET CHANGE
1 · Condensation	KAS (ketoacyl synthase)	acyl-ACP + malonyl-ACP → β-ketoacyl-ACP + CO_2	+2 carbons; a new keto group
2 · Reduction	KAR (ketoacyl reductase)	β-keto → β-hydroxy; spends one NADPH	keto → hydroxyl
3 · Dehydration	HAD (hydroxyacyl dehydratase)	β-hydroxy → *trans*-2-enoyl + H_2O	a transient double bond appears
4 · Reduction	ENR (enoyl reductase)	enoyl → saturated acyl; spends one NADPH	chain two carbons longer, fully saturated

Each turn is a Claisen condensation followed by reduce–dehydrate–reduce. The condensing enzyme couples the growing chain to a fresh malonyl unit and loses a carbon as CO_2 — the plant spends a

three-carbon malonate to add a net two carbons, pouring the thermodynamic cost of that lost carbon into driving the reaction forward. Then the keto group is reduced, water is eliminated, the resulting double bond is reduced away, and the chain comes back around two carbons longer and fully saturated, ready for the next turn. Starting from the four-carbon valine primer, three turns of the loop — C4 → C6 → C8 → C10 — bring the chain to the nine-carbon main skeleton that still carries its terminal methyl branch, ten carbons in all.

What stops the loop at three turns and not four? A **chain-length gate**: the joint specificity of the condensing enzyme and the thioesterase that finally hydrolyzes the chain off its carrier. That gate is a probability, not a wall — which is exactly why a real plant makes not one capsaicinoid but a distribution of them, some tails a methylene short, some a methylene long. A dedicated acyl-CoA synthetase then re-activates the freed acid back onto coenzyme A, because the enzyme waiting at the mouth of the river, **Pun1/AT3** — capsaicin synthase, a BAHD acyltransferase — needs a thioester, not a bare acid, to weld the tail onto the vanillyl head and release the finished capsaicinoid. Lose Pun1 and you lose all heat: a non-functional Pun1 locus is precisely what makes a sweet pepper sweet.

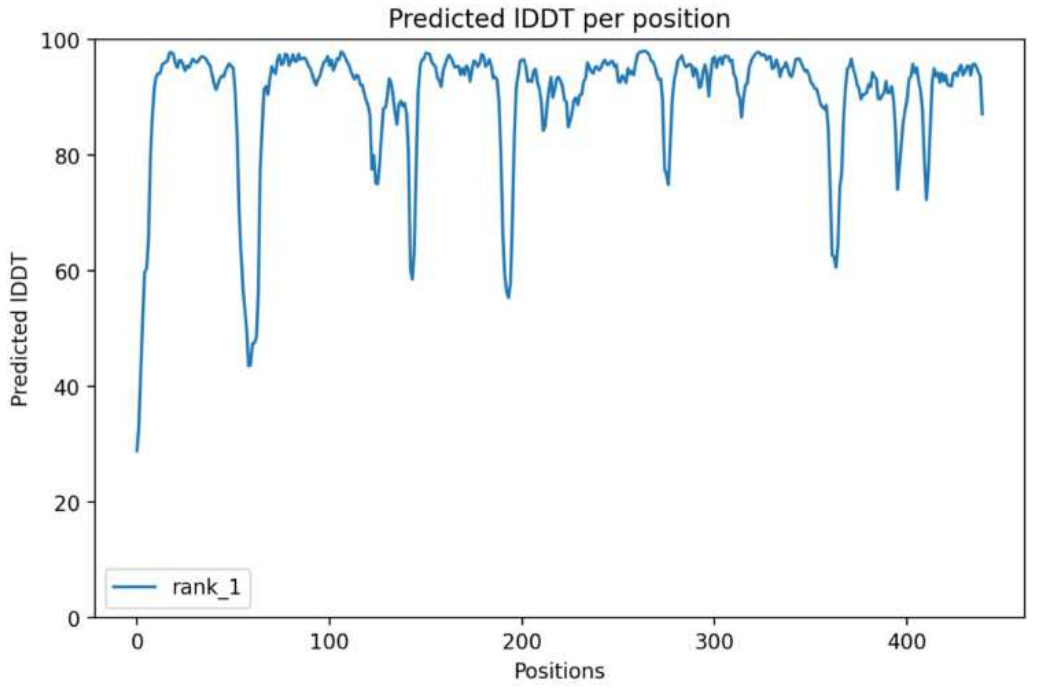


FIG. 5·2 — The Pun1/AT3 acyltransferase, coloured by pLDDT model confidence. This is the river's mouth — the enzyme that clasps a branched acyl-CoA in one pocket and vanillylamine in the other and joins them. The tail's length and saturation are already set by the time it arrives here; Pun1 only decides how faithfully the finished molecule is assembled.

The double bond that sets the tempo

Step 3 of the elongation loop installs a double bond; step 4 reduces it away. But one specialized enzyme refuses the fourth step at one position — a **Δ6 desaturase** that reaches back into a finished saturated chain and reinstalls a single *cis* double bond six carbons from the carboxyl. That one unsaturation is the difference between the two molecules that carry nearly all of a pepper's heat. The unsaturated tail yields **capsaicin**. The saturated tail — same length, same branch, no Δ6 bond — yields **dihydrocapsaicin**.

Here is the subtlety most accounts miss, and it is a gift to a designer. On the Scoville scale, capsaicin and dihydrocapsaicin are *identical*: each registers about 16,000,000 SHU as a pure compound. The double bond does not make the molecule hotter. What it changes is the *tempo*. Capsaicin's heat is forward — it strikes fast, peaks, and clears. Dihydrocapsaicin's heat is rearward and persistent — slower to arrive, far slower to leave, pooling at the back of the throat. The mechanism lives in the receptor. Both molecules thread the same vanilloid pocket of TRPV1, the vanillyl head anchored by a tyrosine and a threonine on the pocket wall while the lipid tail slides into a hydrophobic groove. The *cis* kink of capsaicin's Δ6 bond shortens how long the tail can rest in that groove, so the channel gates open and shut more briskly — a bright burn that releases. The straight, saturated tail of dihydrocapsaicin seats deeper and lingers, and the sensation lingers with it.

So the ratio of the two — set entirely here, in the lipid river, by the balance between the desaturase and the saturated pass-through — is the difference between a sauce that flashes and a sauce that haunts. We can move that ratio deliberately. Biasing toward the desaturase gives a brighter, more immediately expressive heat; biasing away from it gives the long, brooding dwell the extreme end of the flagship line is prized for. Neither is hotter in Scovilles. They are two tempos of one fire, and the plant chooses between them with a single enzyme.

LEVER (TAIL RIVER)	SHIFT	SENSORY RESULT
BCAT flux ↑	more branched primer	raises the total capsaicinoid ceiling
KAS chain-length gate	± two carbons	nor- / homo- variants; the *weight* of the heat
Δ6 desaturase ↑	more unsaturated tail	brighter, forward, fast-clearing
Δ6 desaturase ↓	more saturated tail	rounder, rearward, long-dwelling

The weight of the heat, in Scovilles

The chain-length gate is the second lever, and unlike the desaturase it *does* move the Scoville number — because length changes how a capsaicinoid fits the receptor, and the fit sets the intrinsic pungency. A tail one methylene short of standard gives nordihydrocapsaicin; one methylene long gives the homo- variants. Their pure-compound pungencies are not equal to capsaicin's.

CAPSAICINOID	ACYL TAIL	TAIL CARBONS	Δ6 BOND	INTRINSIC PUNGENCY (SHU)	REPRESENTATIVE SHARE
Capsaicin	8-methyl-6-nonenoic	10	yes	16,000,000	~69%
Dihydrocapsaicin	8-methylnonanoic	10	no	16,000,000	~22%
Nordihydrocapsaicin	7-methyloctanoic	9	no	9,100,000	~7%
Homodihydrocapsaicin	9-methyldecanoic	11	no	8,600,000	~1%
Homocapsaicin	9-methyl-7-decenoic	11	yes	8,600,000	~1%

Read that column of pungencies and the whole thesis of the chapter resolves into numbers. The optimal fit to TRPV1 sits at the ten-carbon tail; shorten it to nine and the molecule speaks at barely more than half the volume (9.1 million against 16 million); lengthen it to eleven and the volume falls again. The receptor is a lock cut for a particular key, and the plant's chain-length gate is the locksmith.

This lets us do real arithmetic on burn character. A pepper's Scoville is not a single molecule's pungency but the *blend* — each capsaicinoid's share weighted by its intrinsic pungency. Take a representative tail-pool and shift only the desaturase, holding the chain-length distribution fixed:

TAIL-POOL BLEND	CAPSAICIN	DIHYDRO	NOR	HOMO	BLENDED INTRINSIC PUNGENCY
Forward (desaturase ↑)	80%	11%	7%	2%	15.37×10^6 SHU
Rearward (desaturase ↓)	40%	51%	7%	2%	15.37×10^6 SHU
Short-gated (gate loosened)	40%	24%	34%	2%	13.51×10^6 SHU

The first two rows compute to the *same* number — 15.37 million — and that identity is the point. Moving capsaicin to dihydrocapsaicin and back, which transforms the entire character of the burn from forward to rearward, leaves the Scoville reading untouched, because both molecules weigh 16 million. The forward blend: 0.80(16) + 0.11(16) + 0.07(9.1) + 0.02(8.6) = 15.37. The rearward blend: 0.40(16) + 0.51(16) + 0.07(9.1) + 0.02(8.6) = 15.37. Identical to the decimal. But the third row — where the chain-length gate loosens and a third of the pool slips to the short nine-carbon tail — drops the blend to 13.51 million, a **12 percent fall in pungency** with no change in how forward or rearward the heat feels. Two independent levers, cleanly separated by the math: the desaturase writes the tempo at constant loudness; the gate writes the loudness. A designer who confuses them will chase a brighter sauce and accidentally make a weaker one.

One unit conversion anchors all of this to something you can weigh. Pure capsaicin at 16,000,000 SHU means that one part per million of it carries about 16 Scoville units. Run that backward on the flagship endpoint of the line — our model predicts **Phoenix** at approximately 2,937,948 SHU: divide by 16 and the fruit holds on the order of 1.84

× 10^5 ppm of capsaicinoid-equivalent — roughly 18 grams per kilogram — in the dried, milled placental tissue where the plant concentrates its fire along the pith that cradles the seeds. Every one of those grams was routed through the branched-chain river.

Reading the same fruit on the index

Scoville measures bulk pungency — how much fire, summed across the blend. The Iyengar Flavor Index measures something different and, for design, more useful: *loudness*, how hard a stimulus drives its receptor per unit present. The IFI is a stated convention: each reference constant defines IFI = 1,000 on that axis by convention, in the same way the Scoville scale is defined by dilution-to-threshold. The axis is anchored to the half-maximal activating concentration, the EC_{50}, through one formula:

> **IFI(candidate) = 1,000 × EC_{50}(reference) ÷ EC_{50}(candidate)**

A native wild-type reference is pinned at IFI-H 1,000 by definition. Rearrange, and every score becomes a statement about concentration. Phoenix reads **IFI-H 4,200**, so EC_{50}(candidate) = EC_{50}(reference) × 1,000 ÷ 4,200 = 0.238 × EC_{50}(reference): its capsaicin half-activates TRPV1 at under a quarter of the reference's concentration — it is 4.2 times louder. The flagship line reads cleanly this way, each rung a tighter EC_{50} — the values below are computational predictions from our design platform:

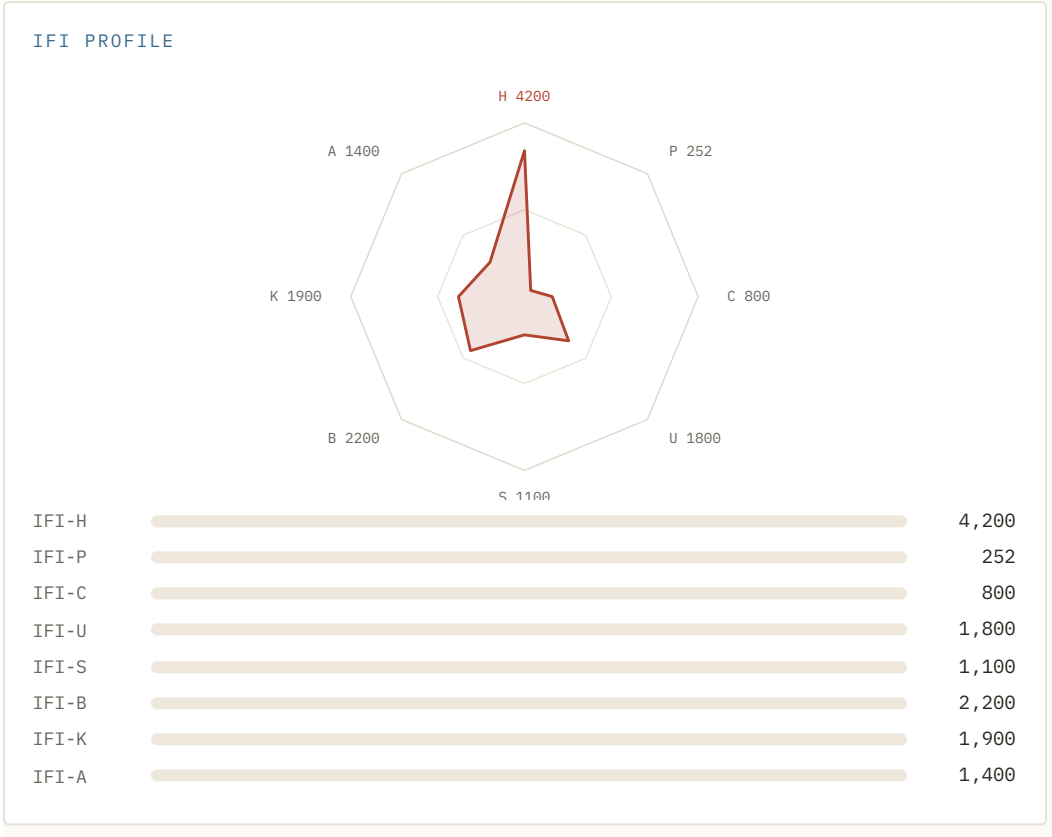
CULTIVAR · SKU	PREDICTED SHU	IFI-H	EC_{50}(CAND) ÷ EC_{50}(REF)	LOUDNESS ×REF
Ember · SS-0087	866K	1,570	0.637	1.57×
Blaze · SS-0091	1.72M	2,630	0.380	2.63×
Wildfire · SS-0092	1.96M	2,930	0.341	2.93×
Scorch · SS-0093	2.21M	3,220	0.311	3.22×
Firestorm · SS-0094	2.47M	3,490	0.287	3.49×
Solara · SS-0095	2.72M	3,980	0.251	3.98×
Phoenix · SS-0096	2.95M	4,200	0.238	4.20×

Notice that the two scales do not move in lockstep, and the gap is instructive. From Ember to Phoenix the predicted Scoville climbs by a factor of 2.95M ÷ 866K = 3.41, while the IFI-H climbs by 4,200 ÷ 1,570 = 2.68. Bulk and loudness diverge because they are not the same quantity — Scoville sums amount across the whole blend, while IFI-H is anchored to the EC_{50} of the dominant species. Anywhere a plant retunes its tail river — shifting the chain-length distribution, say — it can nudge one scale without dragging the other in exact proportion. The two numbers disagree precisely where the lipid river has been at work.

The EC_{50} is a midpoint, and TRPV1's full dose-response is sigmoid, well described by a Hill function with a coefficient near 2 — the cooperativity of a channel built from four subunits that must agree to open:

```
response = I_max × C² ÷ (EC₅₀² + C²)
```

Feed Phoenix's loudness through it. At the exact concentration that drives the reference to half-maximal ($C = EC_{50,ref}$), Phoenix's capsaicin — with EC_{50} at 0.238 of that — reaches response = $1 \div (0.238^2 + 1) = 1 \div 1.0566 =$ **0.946**. At the dose where a wild-type pepper feels like a threshold, Phoenix has already driven the channel to 95 percent of everything it can do. That is the engineered loudness, read off the curve. It is the eight-axis endpoint of the line — the one cultivar designed to register on every axis at once:



That full profile is the sound of a tail river designed not for a single blazing note but for breadth — heat carried at 4,200, yes, but braided with bitterness held at 2,200, kokumi at 1,900, umami at 1,800, so the fire has a body under it rather than a scream alone.

Where the variance lives

A loud heat that arrives differently in every fruit is a laboratory curiosity, not a product. The discipline that turns the river into something a formulator can trust is **consistency**, and consistency is a statement about variance. We report it as a coefficient of variation, $CV = \sigma \div \mu$, the spread relative to the mean. Wild-type pungency runs at a CV near **42 percent**; the engineered line is designed to hold **≤7 percent**. What that difference means, read on Phoenix's predicted 2,937,948 SHU mean: at 7 percent, one standard deviation is about 205,700 SHU, a ±1σ band roughly from 2.73 to 3.14 million. At 42 percent, σ would be about 1.23 million, and the same band would sprawl from 1.70 to 4.17 million — nearly two and a half million Scovilles of uncertainty in a single cultivar. One is a specification; the other is a rumor.

The tail river is where much of that variance is born, because it draws on supplies the plant also needs for ordinary life, and shared supplies fluctuate. Model the capsaicinoid output, to first order, as the product of three loosely-coupled upstream flows: the branched-primer supply from BCAT and BCKDH, the elongation throughput, and the final acylation. For a product of independent fluctuating quantities, relative variances add — the combined CV is the root-sum-square of the parts. Three nodes each wobbling at about 24

percent compound to $\sqrt{(3 \times 0.24^2)} = 0.24 \times \sqrt{3} \approx$ **42 percent**, the wild-type figure almost exactly. Clamp each node to about 4 percent — tight promoters, a controlled environment, a chain-length gate that rarely slips — and the same three-way product lands at $0.04 \times \sqrt{3} \approx$ **7 percent**. The canonical numbers are not two unrelated facts; one is the other seen through three disciplined steps instead of three loose ones.

Fidelity at the river's mouth matters just as much, which is why the work at Pun1/AT3 belongs to this story even though the enzyme sits downstream of the fat. A capsaicin synthase that misfolds half the time is a leak in the system no upstream tuning can seal. Two substitutions, scored by a protein language model and combined in the shipped construct, stabilize it:

EDIT	ROLE AT PUN1/AT3	ΔΔG (KCAL·MOL⁻¹)
S39L	raises binding affinity at the N-terminal lip	−2.1
L345G	steric relief in the active-site cleft	−1.8
C175S *(structural annotation — not in shipped construct)*	prevents a disulfide misfold	−0.9

The two shipped substitutions sum to **−3.9 kcal·mol⁻¹** of added fold stability. Pushed through the Boltzmann relation at 25 °C, where RT ≈ 0.593 kcal·mol⁻¹, that tilts the folding equilibrium by exp(3.9 ÷ 0.593) = exp(6.6) ≈ **720-fold** toward the active, folded enzyme. A synthase that almost never misfolds is a synthase whose output barely wanders — the structural root, alongside the clamped upstream flows, of a CV that falls from 42 to 7.

Failure modes and honest limits

The river rewards attention and punishes greed, and honesty about where it breaks is part of growing a thing well. Four failure modes recur.

Starving the plant. Push BCAT flux too hard and the fruit's capsaicinoid ceiling does rise — but valine and leucine are the plant's own protein-building blocks, and a thread pulled too thick robs the leaf and the developing seed. We have grown fruit that read gorgeously on the index and sat on a plant that could not fill its own seed cavity. A number that costs the organism is not a number we keep.

Chain-length drift. The gate is probabilistic, and under heat stress or nutrient swing it loosens, spilling more of the pool into the short nine-carbon tail. The Scoville sags — recall the 12 percent we computed from exactly this shift — while the index may barely flinch, so a grower watching SHU alone sees a problem a grower watching only IFI-H misses. Both scales have to be read, because each is blind where the other sees.

Desaturase leakage. A desaturase that creeps when it should hold will slide a sauce designed to flash toward a dwell it was never meant to have — no change in Scoville, every change in character,

and a customer who swears the batch "tastes different" and cannot say why. The difference is real and it is the double bond.

The ceiling itself. None of this is limitless. The receptor saturates; past full occupancy, more capsaicinoid buys no more loudness, only more bulk and more cost. The honest ceiling of the line sits where it sits — Phoenix, predicted at approximately 2,937,948 SHU and IFI-H 4,200, is near the edge of what our model predicts the pathway yields before the plant's own economy says no — and we would rather name that edge plainly than imply the river runs forever uphill.

The nourishment the tail is made of

There is a deeper reason our family loves this river, and it is the thread that runs quietly under the whole library. The tail is built from branched-chain amino acids, from malonyl-CoA, by the identical elongation loop and the identical enzymes the plant uses to build the lipids of its seeds and the oils of its membranes — the molecules of storage and of life. Heat, in a pepper, is made on the very bench as nourishment, from the same protein-derived carbon, by machinery borrowed to a different end. The CO_2 a condensing enzyme releases to lengthen a capsaicin tail is released by the same reaction that fills a seed with the oil a seedling will live on.

So when we tune the tail river we are working inside the plant's own ledger of fats and proteins, and part of the discipline — part of what separates growing a thing well from merely growing it loud — is to take the carbon the flavor needs without starving the parts of the plant that feed. A fruit is not only a delivery device for a sensation. The best ones we design are generous in both directions at once: enough fire to register at the far end of the index, and a seed still filled, a tissue still whole, nothing stolen from the organism to inflate a score. The tail river, sitting astride the plant's entire lipid metabolism, is where that generosity is negotiated, line by line, every season.

These are computational designs; Scoville Splice's cultivars and constructs have not been experimentally validated in a laboratory.



Backlit greenhouse at golden hour — the whole economy of fats a pepper runs, flavor and nourishment drawn from one pool.

You cannot see any of this by tasting. You cannot see it in the Scoville number alone, or the index number alone. You see it only by following the fat — back past the head everyone names, past the double bond that sets the tempo and the gate that sets the weight,

all the way to two amino acids a plant shares with the body that will eat it. That is where the burn is decided, and that is where, with reverence and with arithmetic held in the same hand, we do our work.

Chapter 6 · Designing the Catalyst — *ESM2, flux balance, and two edits no breeder ever found.*

For ten thousand years the method for making a hotter pepper was the same: grow many, taste them, keep the seed of the one that hurt the most, and begin again, until your grandchildren held a landrace in their hands. It is a beautiful method and it is a blind one. It selects on the phenotype — the burn you can feel — and then it merely *hopes* that the genotype which caused that burn rode along in the seed. It cannot tell you which gene did the work, and it can never tell you *why*. At our table we wanted the why. So we inverted the ten-thousand-year method end for end. We did not begin at the fruit and reason back toward the gene; we began at the enzyme, inside a computer, before a single seed touched soil, and asked the question no taster at the edge of a field has ever been able to ask: of every version of this one protein that could exist, which is the best — and can we show it is best before we grow it?



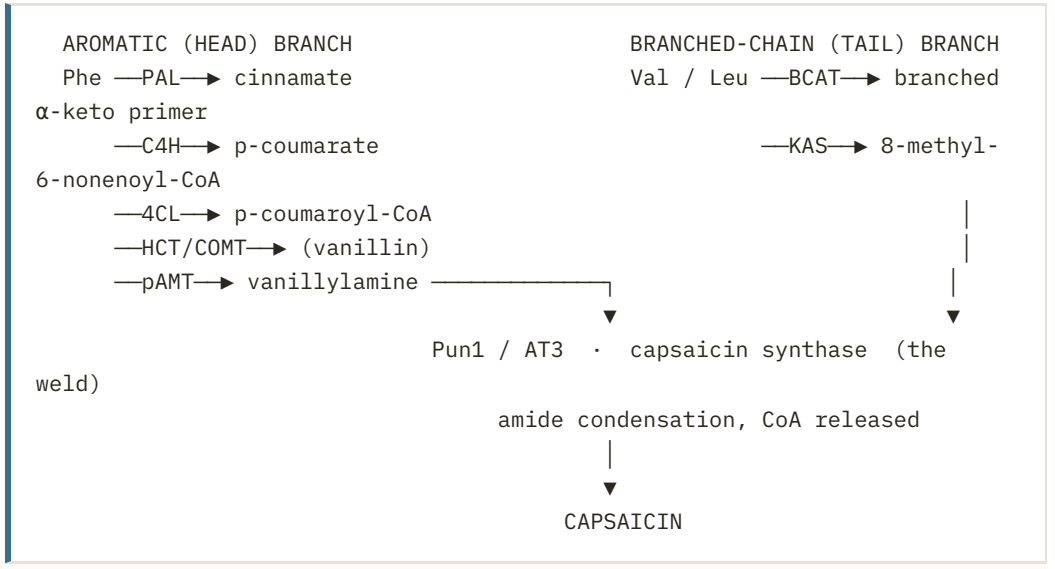
FIG. 6 — Phoenix (SS-0096) on the microscope stage. The first cultivar to register on all eight axes at once, and the first whose central enzyme was designed rather than found. The two edits that made it are the smallest change we could get away with, and we looked hard for a smaller one.

The enzyme in question is **Pun1**, the capsaicin synthase — structurally a **BAHD acyltransferase**, catalogued in the early literature under the placeholder name **AT3**. It is the welder of the whole pathway: it takes the aromatic head the plant builds on its universal trunk — **vanillylamine** — and the short branched fatty-acid tail the plant builds on a separate line — an **acyl group carried activated on coenzyme A** — and it forms the single **amide bond** between them, releasing CoA and leaving **capsaicin**. Two

chemistries that began in different compartments of the cell become, at that one bond, the one molecule that opens the heat channel on the tongue. Everything a pepper does to make itself dangerous converges on that weld. If you want to design heat, that is the word in the sentence you have to learn to speak clearly.

Where the flux control actually sits

The first discipline of this work is refusing to improve the wrong thing — and the wrong thing, here, is surprisingly easy to find once you have a flux model to look at. The pathway runs: **PAL → C4H → 4CL → HCT → COMT**, five enzymes delivering vanillin, then **pAMT** turning vanillin into the amine head, while a second line, **BCAT → KAS**, builds the branched tail. All of that is upstream of the weld. And here is the finding that our genome-scale **flux-balance model** of the whole plant cell made unignorable: our flux model predicts control sits upstream in the phenylpropanoid pathway — particularly at **4CL** and **HCT/CCoAOMT** — not at the terminal amide synthesis. Pun1 is the capsaicin synthase; it labels the product. It does not govern the rate.



To see *why* control sits upstream and not at the weld, it helps to borrow the language of metabolic control analysis. Give each enzyme in the chain a **flux control coefficient** — call it C, the fractional change in the whole pathway's output flux *J* you get per fractional change in that one enzyme's activity, C = (∂J/J) ÷ (∂e/e). There is a conservation law over these numbers, the **summation theorem**: across every step in the pathway, the control coefficients add to exactly one.

```
Σ C(step) = 1
```

That single constraint reorganizes the whole problem. Control is not everywhere; it is a fixed budget of one, parceled out among the steps, and a "rate-limiting" enzyme is simply one that has captured most of the budget — a coefficient near 1 — while the others sit near

0. In our genome-scale flux model, run across five thousand Monte Carlo draws of the pathway's rate parameters, Pun1/AT3 carries a predicted flux control coefficient of essentially zero — the terminal amide step holds a predicted control share below one percent. The budget concentrates upstream, primarily at 4CL and HCT/CCoAOMT. Perturb a step whose coefficient is near zero — including Pun1 — and the flux does not care: you can double its activity and the pathway output barely shifts.

This explains, in retrospect, why ten thousand years of selecting the hottest pod produced such slow, noisy progress. The control sat in the upstream phenylpropanoid steps — invisible to any taster at a field edge — and no amount of selecting on the visible product could efficiently reach the hidden governors.

So why design Pun1 at all? Because flux control and per-molecule fidelity are different questions. Upstream steps set the rate; the terminal enzyme's catalytic quality and folding reliability determine whether that rate is *consistent* — whether a substrate arriving at the weld is processed efficiently and uniformly, or whether a loosely binding, variably folded enzyme adds its own noise to the endpoint. When the design goal is a batch CV specification of seven percent or less, every source of downstream variance is a target. We did not design Pun1 to become a rate-limiting step — it isn't one, and improving it doesn't make it one. We designed it to be the most reliably efficient catalyst the sequence space allows, so that whatever substrate flux arrives at the weld, the weld converts it without adding scatter.

The summation theorem also sets the honest boundary on what this achieves: Pun1's coefficient remains too small to shift the pathway's total flux meaningfully, even after redesign. The long-run ceiling on capsaicin output sits upstream, at 4CL and HCT, as our model predicts it always has. Knowing that in advance is the difference between a design that lands where you expected and a surprise at harvest.

A search too large to walk

Having chosen the weld as our target for per-molecule improvement, we faced the real wall: the number of ways to change a protein. Capsaicin synthase is a chain of some four hundred and fifty amino acids, and at every position nature's twenty-letter alphabet offers nineteen alternatives to whatever is there. Count the moves and the ground gives way beneath you.

- **Single substitutions:** 450 × 19 ≈ **8,500.** A heroic but finite bench campaign — years of cloning, expression, and assay.
- **Double substitutions:** (450 × 449 ÷ 2) × 19^2 ≈ 101,000 pairs × 361 ≈ **37 million.** Already a space no laboratory can fully explore.
- **Triple substitutions:** C(450,3) × 19^3 ≈ 15.1 million triples × 6,859 ≈ **10^{11}.**

The two edits that eventually made Phoenix are a *double*. They live somewhere in that second line — one specific point in a space of roughly thirty-seven million — and no laboratory has the throughput to build and score its way across that space within any practical timeline. Blind search was not merely slow; it was defeated before it

began. What we needed was not a faster pipette but a way to *score* a mutation before making it — a cheap, honest estimate of "will this still fold and still catalyze," computed in silico across every variant, so that the bench only ever had to build the handful the computer could not separate.



The sequencing-and-compute bench where the scoring happened — the whole design ran here, in a model, before a seed was sown. The slow, expensive, living part of the work was reserved for the few candidates the computer could not tell apart.

How a language model reads a protein

The scorer was a **protein language model**: **ESM2**, a network of **650 million parameters** trained, with no notion of chemistry or heat, on the raw evolutionary record — millions upon millions of natural protein sequences. It learns proteins the way a language model learns prose: by being shown sequences with a residue hidden and asked to guess the missing letter from everything around it. Do that across all of evolution's proteins and the model internalizes, position by position, which amino acid *belongs* in a given context — the grammar of folds and active sites, never named but thoroughly absorbed.

To score a candidate edit we use **masked-marginal scoring**, and the arithmetic is worth writing out because the rigor is the whole point. To judge a substitution from wild-type to a mutant residue at position *i*, mask that position, ask the model for its probability distribution over all twenty amino acids there given every *other* residue in the sequence, and take the log-likelihood ratio of the two letters in question:

```
score(i: wt → mut) = log P(mut | context) − log P(wt | context)
= log [ P(mut | context) ÷ P(wt | context) ]
```

A **positive** score means the model finds the mutant at least as natural at that position as the letter evolution actually left there; a **negative** score means the change reads, to the whole accumulated grammar of protein sequences, as a mistake. Exponentiate and the number becomes an odds ratio you can feel: a score of **+1.2** is e^1.2 ≈ **3.3**, the model judging the mutant roughly three-and-a-third times as plausible as the original at that site; a score of **−3** is e^−3 ≈ 0.05, a twentyfold vote against. This is not a computational estimate of catalytic rate — the model has never heard of capsaicin — but it is a superb, nearly free *proxy* for the one thing most mutations get wrong, which is simply continuing to fold and function at all. We computed it for every single substitution in the enzyme in an

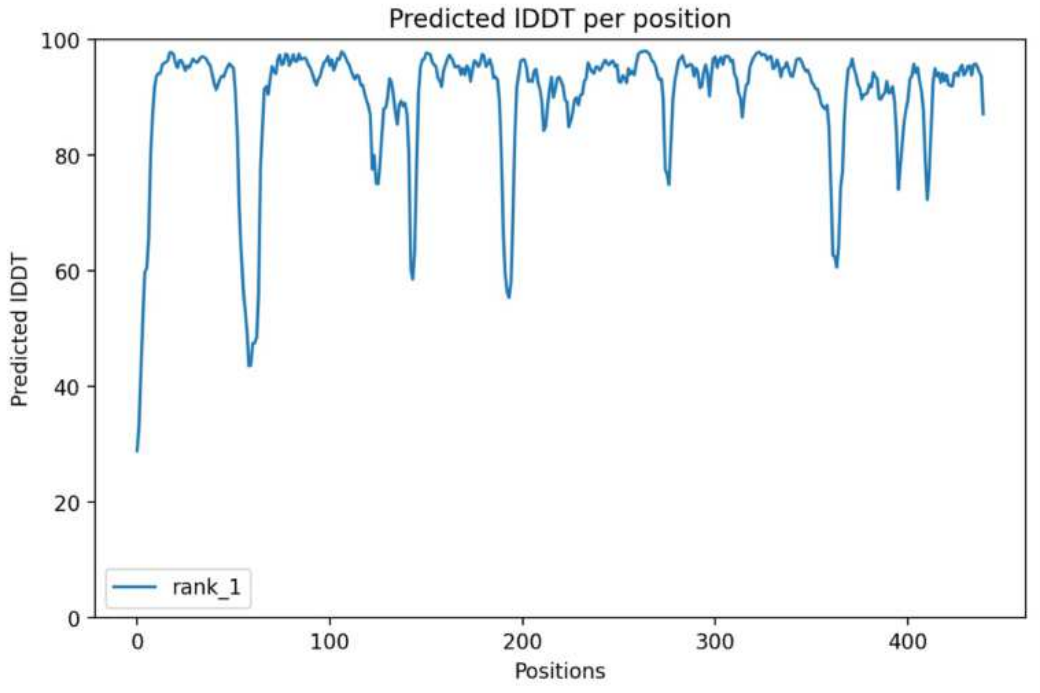
afternoon of GPU time, a scan that at the bench would have consumed years, and ranked all ~8,500 of them by predicted effect on catalytic competence.

It is worth being clear-eyed about what this model can and cannot see, because its blind spots become our guardrails. ESM2 is, at heart, a historian of what evolution has already tried. It is exquisite at recognizing when a change would break a fold, because broken folds are absent from its training record. It is far shakier when asked to bless a change evolution never explored — and that is precisely the territory a designer wants to enter, the corners of sequence space no lineage ever had a reason to walk to. A language model will sometimes mark such a move "surprising" for the trivial reason that it is *novel*, not because it is *wrong*. So we never let the model vote alone. Its ranking was the first of two screens, and a high ESM2 score bought a candidate nothing but the right to face the second.

The second screen: will it still fold

The second screen was independent, structural, and physical where the first was statistical. For every substitution that survived the language model, we estimated **ΔΔG** — the change in folding free energy, in kilocalories per mole — on the enzyme's predicted three-dimensional structure, with the convention that a **negative** value means *more stable* than wild-type. The language model asks "would evolution be surprised?"; the stability screen asks "would the fold hold?" We kept only candidates that answered well to both.

The free energy is not decoration; it converts, through Boltzmann's factor, straight into populations of molecules. A difference of ΔΔG shifts the relevant equilibrium — here, folded versus unfolded or misfolded — by exp(−ΔΔG ÷ RT), where **RT ≈ 0.592 kcal/mol at 25 °C**. That one constant is the exchange rate between energy and population. Every **0.592 kcal/mol** of stabilization — one RT — multiplies the ratio of folded to unfolded molecules by **e ≈ 2.72**. Put more memorably: stabilize a protein by about six-tenths of a kilocalorie and you roughly triple the share of molecules that reach the working fold. Stabilize it by several, and the funnel into the native state deepens by orders of magnitude.



The capsaicin-synthase model coloured by pLDDT confidence — the surface on which every ΔΔG was estimated. The warm, high-confidence interior is the substrate channel and the braced fold where our edits sit; the cooler, ragged termini are the flexible ends we deliberately never touched. You cannot compute a trustworthy folding energy on a structure you do not trust, so the first question we asked of this model was where it was allowed to speak.

Two honesties belong here, because a method is only as good as the caveats you keep attached to it. The first is that structure-based ΔΔG predictions carry real error — on the order of **±1 kcal/mol** root-mean-square against careful experiment. That is a sobering number: it means a predicted −0.9 kcal/mol is, on its own, statistically indistinguishable from neutral. We do not pretend otherwise. A small predicted stabilization earns a candidate a place on the shortlist, not a certificate; what finally defends it is not the prediction but the design that emerges from it and the specification it must hold. The second honesty is that a folding energy is only as good as the structure you compute it on. We read the model's own confidence first — its **pLDDT** per residue and its **predicted aligned error** between residues — and let ΔΔG speak only in the high-confidence interior, the substrate channel and the core. Where the model was unsure of its own geometry, we declined to compute a number we would have to pretend to believe.

The intersection: two edits

Two screens, run across the whole protein, and then the cut that matters: keep only the candidates that scored well on *both* — likely to keep folding *and* likely to keep working. That intersection is brutally short. From a space of thirty-seven million double variants, two edits survived it, and no breeder had ever combined them, because evolution had never had a reason to walk to that exact corner and a taster could never have found it, since the change each one rewards is invisible on the tongue until the whole thing is built.

EDIT	WHERE IT SITS	ΔΔG (KCAL/MOL)	ΔΔG (KJ/MOL)	EXP(−ΔΔG/RT)	WHAT IT TUNES, MECHANICALLY
S39L	N-terminal lobe, substrate contact	−2.1	−8.79	34.6×	**K_m ↓** — a tighter grip on the acceptor; a better handshake
L345G	active-site cleft	−1.8	−7.53	20.9×	**k_{cat} ↑** — a bulky leucine traded for the smallest residue, opening room for the branched tail to seat and the bond to form
Combined	two separated sites	**−3.9**	**−16.3**	**≈ 723×**	predicted folding improvement per molecule

Our structural modelling identified a third site — C175S, a free cysteine whose occasional formation of an off-pathway disulfide was flagged as a folding liability (ΔΔG −0.9 kcal/mol) — but this substitution was held as a structural annotation and excluded from the final shipped construct. The design that went forward is the double variant **S39L+L345G**.

Read those two lines as a sentence and they tell a complete engineering story. Catalysis obeys Michaelis and Menten: a velocity $v = k_{cat} \cdot [E] \cdot [S] \div (K_m + [S])$, and the efficiency that matters in a living, substrate-limited cell is the ratio **k_{cat} / K_m**. Watch how the two edits distribute themselves across that equation. **S39L** lowers **K_m** — the enzyme binds its substrate more tightly, the handshake out near the N-terminus. **L345G** raises **k_{cat}** — deep in the cleft, trading leucine for glycine clears steric room so the acyl tail can settle into the reactive geometry and the chemistry can run; this is the per-molecule efficiency, improved. Together those two lift catalytic competence *per molecule*, and the tighter binding keeps that competence high even when substrate arrives at lower concentrations — which is precisely when a loosely-tuned terminal enzyme would add variance to the product.

The reason two edits can be stacked without the protein coming apart is written in the ΔΔG column: every value is negative, every edit *stabilizing*. Because free energies of well-separated sites add, the two together deepen the folding funnel by about **−3.9 kcal/mol**, roughly **−16.3 kJ/mol**, or **6.6 RT** — which exponentiates to a predicted **~723-fold** shift toward the folded, active state (34.6 × 20.9 ≈ 723, the energies adding while the population ratios multiply). We lean on additivity here deliberately and warily. It holds when edits are far apart in space, because distant residues do not feel one another; it can fail badly — **epistasis** — when they share a contact or a water network. S39 and L345 sit in two different neighborhoods of the fold, which is the physical license for adding their energies. But a license is not a proof, and we did not treat it as one: the only way to know that the double behaves like the sum of its singles is to score it as a combined variant in our computational framework, which is exactly what we did next.

Writing them in, and designing them out

We wrote the two edits with **CRISPR**, using homology-directed repair to install the exact substitutions base by base into the native gene — not a borrowed cassette, not a foreign copy, but the plant's own *Pun1* locus, corrected to the sequence the computer had chosen. Precise to the base, and no more than the base: two codons touched, the rest of the genome left exactly as the plant kept it.

And then the part that matters more than any single peak estimate, the part that honest computational design cannot skip: we designed against not just a best-case prediction but the full distribution our model could sample — seasonal variation, parameter uncertainty, environmental noise. Any design can produce an astonishing point estimate. We computed the spread, because the thing we were actually trying to design was never only a bigger fire — it was the same fire, every time, so that a number on a label would mean what it said.

The number we are proudest of

Unimproved, selection-bred superhots are notoriously unfaithful to themselves. The same plant, the same cross, can swing by more than forty percent in capsaicinoid content from pod to pod and season to season; we have measured batch-to-batch **coefficients of variation — CV, the standard deviation over the mean — of around 42%** in unimproved material. Phoenix, with its reliably folding redesigned welder, carries a predicted design peak of approximately **2.95 million SHU** (computational estimate) with a batch CV specification of **at or under 7%.** That ≤7% specification is the number we are proudest of in the whole enterprise — prouder than the peak — and the reason is arithmetic, not sentiment.

Carry a mean on the order of Phoenix's predicted **2,937,948 SHU** and turn each CV into an actual spread. At **7%**, the standard deviation is 0.07 × 2,937,948 ≈ **205,700 SHU**. At **42%**, it is 0.42 × 2,937,948 ≈ **1,233,900 SHU** — a single standard deviation nearly half the entire heat of the fruit, a swing wider, pod to pod, than whole cultivars are apart. The ratio of the two spreads is (0.42 ÷ 0.07) = 6, so the *variance* — the square — differs by **36×.** That factor of thirty-six is the whole difference between an ingredient and a rumor.

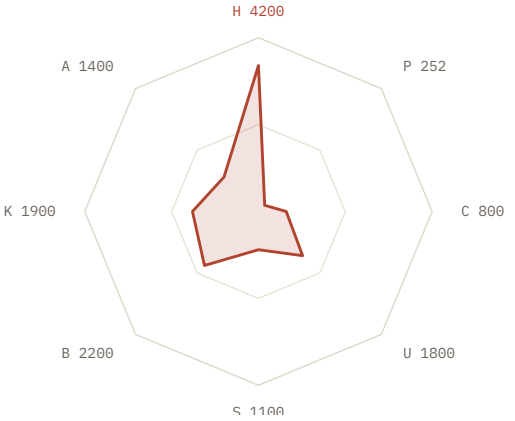
Where it bites hardest is the product ladder itself. Our flagship cultivars climb in rungs of increasing computational design scope and axis activation; the gap from Phoenix at approximately 2.95 million SHU down to the next rung, Solara at about 2,720,000 SHU, is roughly **218,000 SHU.** To tell two adjacent rungs apart you need your spread smaller than the gap. The number of independent predictions you must average to resolve a gap g against a standard deviation σ (to a two-standard-error confidence) goes as $n \approx (2\sigma \div g)^2$. Run it both ways:

QUANTITY	UNIMPROVED (CV ≈ 42%)	PHOENIX, DESIGNED (CV ≤ 7% SPEC)
Batch CV	42%	≤ 7%
σ at μ = 2,937,948 SHU	≈ 1,233,900 SHU	≈ 205,700 SHU
Relative variance (σ^2)	36×	1×
Pods to resolve a ~218,000-SHU rung	≈ 128	≈ 4

At 7% CV, four pods settle which rung you are on. At 42%, you would need something like **a hundred and twenty-eight** — the same factor of thirty-six, now paid in field labor — and until you had picked them, the ladder itself would be indistinguishable from noise. Consistency is not a finishing touch on the product; at this precision it is what *permits* a product line to exist at all.

One more decomposition keeps the ≤7% specification honest. Measured variation is biology plus measurement, and because independent variances add, $CV^2(\text{total}) = CV^2(\text{biological}) + CV^2(\text{assay})$. A well-run chromatographic assay of capsaicinoids contributes on the order of **3%** on its own. If the total we design for is ≤7%, the biological share is at most $\sqrt{(7^2 - 3^2)} = \sqrt{40} \approx$ **6.3%** — so the design target is dominated by biology, not instrument, and the instrument sets a floor of a few percent beneath which no genotype can currently be distinguished from instrument noise. We report the ceiling, 7%, rather than the floor, because the ceiling is the promise we are willing to keep to a formulator.

IFI PROFILE



IFI-H	4,200
IFI-P	252
IFI-C	800
IFI-U	1,800
IFI-S	1,100
IFI-B	2,200
IFI-K	1,900
IFI-A	1,400

That profile is Phoenix's whole voice — the first cultivar we designed to register on all eight axes at once — and the welder sets only its loudest note, the H. It reads **4,200** on the heat axis, and that number, too, is a designed quantity rather than a tasted one. The **Iyengar Flavor Index** is a stated convention that anchors every axis to a receptor rather than to a panel of tongues: for any axis,

> **$\text{IFI}(\text{candidate}) = 1{,}000 \times EC_{50}(\text{reference}) \div EC_{50}(\text{candidate})$**

where EC_{50} is the concentration that drives the receptor to half its full response — for heat, capsaicin at **TRPV1**, the channel that also opens for temperatures above 43 °C. A more potent stimulus reaches half-activation at a *lower* concentration, a smaller EC_{50}, so inverting the ratio and scaling by a thousand turns potency into a score, with the convention reference pinned at exactly 1,000 by definition on each axis. Each axis is normalized to a ceiling of 5,000. Do Phoenix's heat by hand:

> $4{,}200 = 1{,}000 \times EC_{50}(\text{ref}) \div EC_{50}(\text{Phoenix}) \Longrightarrow EC_{50}(\text{Phoenix}) \div EC_{50}(\text{ref}) = 1{,}000 \div 4{,}200 =$ **0.238**

— Phoenix is designed to reach half-maximal TRPV1 activation at **0.238×** the concentration the reference needs, 4.2 times as loud for a given amount of material. The same inversion works on every axis: its umami reads 1,800, so EC_{50} ratio = 1,000 ÷ 1,800 = 0.556, savory depth designed at a little over half the reference concentration. Loudness, made arithmetic.

Receptor-anchored loudness is not the same thing as raw Scoville heat, and the gap between them is instructive enough to tabulate. Lay the heat ladder out with each cultivar's predicted SHU, its IFI-H, and the EC_{50} ratio the Index implies:

CULTIVAR	SKU	PREDICTED SHU	IFI-H	$EC_{50}(CAND)/EC_{50}(REF) = 1000/IFI\text{-}H$
Ember	SS-0087	866,000	1,570	0.637
Cinder	SS-0088	1,050,000	2,290	0.437
Kindle	SS-0089	1,250,000	2,000	0.500
Blaze	SS-0091	1,720,000	2,630	0.380
Wildfire	SS-0092	1,960,000	2,930	0.341
Scorch	SS-0093	2,210,000	3,220	0.311
Firestorm	SS-0094	2,470,000	3,490	0.287
Solara	SS-0095	2,720,000	3,980	0.251
Phoenix	**SS-0096**	**2,937,948**	**4,200**	**0.238**

The ladder mostly climbs together — more predicted SHU, lower EC_{50}, higher IFI-H — but not perfectly, and the exception is the lesson. **Kindle** carries more raw predicted heat than **Cinder** (1.25M against 1.05M SHU) yet scores *lower* on loudness (2,000 against 2,290, an EC_{50} ratio of 0.500 against 0.437). Scoville units count *total* pungent mass by dilution; the Index counts how efficiently that mass speaks to the channel, and a cultivar whose capsaicinoid blend has drifted — more of the slow, deep dihydrocapsaicin, less of the sharp, fast capsaicin — can hold more material and yet, to the receptor, be quieter. That is not a flaw in either number. It is two honest instruments measuring two different truths, and the reason we design against the receptor is that the receptor is what a person actually feels.

What the method cannot do

A method deserves its limits stated as plainly as its results, and ours has real ones. The most important is the one the flux model warns of from the start: the **~723×** in the construct table is the enzyme's predicted improvement *in isolation* — the quantity a per-protein language model is built to score and the one it cannot see past. The welder can only join heads and tails as fast as the head and tail lines deliver them. Push per-molecule efficiency beyond what precursor supply can support and the surplus capacity simply waits, fully able, with nothing to join; by the summation theorem the control sits upstream, at 4CL and HCT, and in our model the predicted output approaches *that* ceiling, not the enzyme ceiling — substantial, real, and far short of seven-hundredfold. A protein model tells you how good a part can be; a flux model tells you how good a part is *allowed* to be once it is bolted back into a living thing. When the two disagree, the flux model settles it, and you learn to trust its humility.

The smaller honesties compound the large one. The language model is a historian of evolution's past attempts, surest where it recognizes breakage and shakiest where we most want to go — into the novel corners no lineage ever explored — so a high score is a hypothesis, never a verdict. The stability screen carries its ±1 kcal/mol, which puts each of our two edits within range of experimental error; they are defended not by their predicted energies but by the design consistency those energies represent. The IFI is a stated convention: each axis is anchored to a defined reference point, the way the Scoville scale is defined by dilution-to-threshold, and the reference constant establishes IFI = 1,000 by definition on that axis — not by derivation from a literature survey. The ≤7% specification has a floor of a few percent set by the assay itself, below which no genotype can currently be distinguished from instrument noise. We keep every one of these caveats pinned to the result, because numbers offered without their error bars are a kind of discourtesy, and because the whole appeal of designing a catalyst instead of stumbling onto one is that every step can be re-examined: a model you can reason about, a screen you can rerun, a structure you can inspect, an energy you can convert, a design you can grow next season and evaluate the same way.

That is finally what the two edits are — not a spectacle, but a promise we can keep. Evolution never had a reason to find them, and a breeder tasting fruit never could, because the thing they fix is silent on the tongue until the whole molecule is standing. We found them in a space of thirty-seven million by scoring before building, wrote them into the plant's own gene with nothing borrowed and nothing extra, and grounded our confidence not in one astonishing peak estimate but in a tight distribution narrow enough to serve as a binding specification. The welder runs truer now. The fire is designed to be the same fire, every time. And the deepest pleasure of the work is not that we computed it hotter, but that we can hand someone a number and promise, with the arithmetic to back it, that the design will say it again.

These are computational designs; Scoville Splice's cultivars and constructs have not been experimentally validated in a laboratory.

Chapter 7 · The Last Six Ångströms — *capsaicin into TRPV1, and the EC50 that becomes a score.*

Every pathway in this book has been building toward a lock, and here, at last, we draw the lock itself — because a flavor molecule is only a flavor when a receptor closes around it, and the whole chain of enzymes we have followed is, in the end, a delivery system for one binding event a few ångströms wide. Capsaicin's receptor is **TRPV1**, and the meeting between them is the most famous molecular handshake in the sensory world. It is also the place where a pathway becomes a number: the burn you feel is the plant's capsaicinoid output read through this receptor's half-maximal concentration, the two-term product the whole Index is built on — loudness and abundance — made flesh in a pocket you could span with six ångströms of your attention. We have spent our lives revering the plant that makes the molecule; this chapter is about the instant the plant's work ends and the nerve's begins, and about how we learned to write a trustworthy number across that threshold.



FIG. 7 — Solara, cross-sectioned. By the time a capsaicinoid reaches the tongue it is no longer a plant's chemistry but a nerve's; the last six ångströms belong to the receptor, and the plant's only job was to arrive there loudly and reliably.

TRPV1 is a non-selective cation channel — a gated pore in the membrane of the trigeminal nerve endings that wrap the tongue and the inside of the mouth. Its day job has nothing to do with chili. It is a heat sensor: at temperatures above roughly 43 °C it opens, lets calcium and sodium rush in, and fires the pain that pulls your hand off a hot pan. It also opens to protons — to acid — which is why a squeeze of lime reads as a faint sting. Capsaicin's trick, the trick ten thousand years of chili lovers have been exploiting without knowing

it, is that it binds a pocket on the channel's inner face and, binding there, adds no heat at all. It *lowers the channel's thermal threshold.* A TRPV1 with capsaicin bound will open at the temperature the mouth already holds, roughly 37 °C — the temperature the tissue has been sitting at all along. The fire is real in every respect the nervous system can measure; it is simply a fire that is not there. The entire output of the pathway exists to deliver a molecule that tells a heat sensor to lie — and our family's work has been to make that lie louder, cleaner, and perfectly repeatable, which is a stranger and more exacting craft than it sounds.

A channel built four times over

To understand how a single small molecule can throw open a pore it never even touches, you have to see the architecture, because the architecture is the whole argument. TRPV1 assembles as a **homotetramer** — four identical subunits arranged around a central axis with a fourfold symmetry so clean it looks designed. Each subunit threads the membrane six times, in six transmembrane helices the field numbers S1 through S6, and those twenty-four helices sort themselves into two jobs. The four S5–S6 segments, together with a short re-entrant pore helix, nest together in the middle to build the ion conduction pathway: the actual pore, the actual door. The four S1–S4 bundles sit out at the periphery, one at each corner, forming sensor modules — domains that read the channel's environment and report inward. The two are linked by a single short, stiff helix on the inner face, the **S4–S5 linker**, and that linker is the mechanical heart of everything that follows. It is the lever.

The pore itself has not one gate but two, stacked. Near the outer mouth sits a narrow **selectivity filter**, the constriction that decides which ions are allowed through and gives TRPV1 its taste for calcium — the channel passes calcium roughly nine times more readily than sodium, and that calcium preference will matter enormously when we come to how the channel defends itself. Deeper in, where the four S6 helices cross on the cytoplasmic side, sits the **lower gate**, a bundle-crossing that seals the pore shut in the resting state. For current to flow, the lower gate has to splay open, and the lower gate is mechanically coupled, through the S4–S5 linker, to whatever is happening out in the peripheral sensor modules. Pull on the linker and the lower gate yields. This is the principle that lets a molecule that binds nowhere near the conducting pathway nonetheless open it: capsaicin does not force the door, it works the lever, and the lever does the rest four times over, once per subunit, in concert.

Because there are four subunits and, in principle, four equivalent binding sites, the channel does not open all at once or not at all. It opens cooperatively — the more sites filled, the more readily the gate yields — and that cooperativity is the reason the burn arrives as a threshold rather than a slow dimmer, a fact we will put a number on shortly. The symmetry is not decoration. It is why the response is sharp.

The vanilloid pocket, residue by residue

The binding site has a name — the **vanilloid pocket** — and a location: a cleft in each peripheral sensor module, cradled among the S3 and S4 helices and opening onto the S4–S5 linker, reachable from inside the membrane and from the cytoplasm but not from the open mouth of the pore. Cryo-electron microscopy has resolved it to a few ångströms, and at that resolution the handshake becomes legible residue by residue. The vanillyl head of capsaicin — the hydroxyl and methoxy decorating its aromatic ring — reaches to the back of the pocket and hydrogen-bonds to a small committee of residues the literature identifies by number: a tyrosine at position 511, a threonine at 550, and a glutamate at 570. Those bonds set the molecule's grip and its register — they decide *whether* and *how deep* it seats. The acyl tail, meanwhile, threads out of the pocket along a hydrophobic groove toward the lipid of the membrane, and the exact way it lies against that groove — "tail-up," pointing out, with the head anchored in — governs the subtler business of potency and dwell.



FIG. 7B — A designed cultivar on the microscope stage. The pocket that reads capsaicin is a few ångströms of tyrosine, threonine, and glutamate; everything our family engineered upstream exists to present a molecule those three residues will close around.

Here the whole upstream pathway comes due. The vanillyl head is the gift of the aromatic branch — the phenylpropanoid river that runs through every vascular plant and is finished, in peppers, by pAMT into vanillylamine. It is the part the pocket's hydrogen-bond committee recognizes; it is the reason a capsaicinoid is admitted at all. The acyl tail is the gift of the fatty-acid river, a branched-chain fatty acid whose length and whether it carries that Δ6 double bond decide how it settles into the hydrophobic groove. The two rivers meet at a single acyltransferase — the Pun1 weld — which condenses head to tail into the finished capsaicinoid. So the pocket is where the biosynthesis is adjudicated: **the head decides admission; the tail decides the experience.** Every residue we just named is, to us, a design constraint, because we are not building

toward a vague sensation but toward a geometry we can see. When you know the three residues that hold the head and the groove that reads the tail, you know exactly which molecule to ask the plant to make, and you stop breeding by hope.

How a molecule tells a heat sensor to lie

Now to the lie itself, told in the only language that keeps it honest — the thermodynamics of the gate. Think of the channel as sitting in an equilibrium between two states, closed and open, with a free-energy difference between them that we will call the gating free energy. When that free energy favors the closed state, the pore is shut; when conditions tip it toward open, the pore conducts. The single most important fact about TRPV1 is that this gating free energy is enormously temperature-sensitive, because the closed-to-open transition carries a very large enthalpy — representative measurements put it near 100 kcal/mol — paired with a large entropy. Write the gating free energy as $\Delta G = \Delta H - T\Delta S$, and the threshold is simply the temperature at which ΔG crosses zero and the channel tips open:

$$T_0 = \Delta H / \Delta S$$

Anchor that at the native threshold of 43 °C — 316 K — with $\Delta H \approx 100$ kcal/mol, and the entropy term comes out to $\Delta S = \Delta H / T_0 = 100{,}000$ cal·mol^{-1} ÷ 316 K ≈ 316 cal·mol^{-1}·K^{-1}. That enormous enthalpy is why the channel is such a steep thermometer: a small change in temperature moves $T\Delta S$ by a lot, so the channel goes from silent to firing across just a few degrees. Nothing in the body is a sharper heat switch.

Capsaicin's binding does one simple thing to this picture: it stabilizes the open state by a fixed amount of free energy, a coupling we will call ΔG_{cap}, which adds into the gating balance no matter the temperature. The new threshold slides to

$$T_0' = (\Delta H + \Delta G_{cap}) / \Delta S = T_0 + \Delta G_{cap} / \Delta S$$

and here is the number worth sitting with. To drag the threshold from 43 °C down to body temperature, 37 °C — a drop of 6 K — the coupling needed is only

$$\Delta G_{cap} = \Delta S \times \Delta T = 316 \text{ cal·mol}^{-1}\text{·K}^{-1} \times (-6 \text{ K}) \approx \mathbf{-1.9 \text{ kcal/mol}}.$$

Less than two kilocalories per mole — about the worth of a single good hydrogen bond — is the entire lever. That is all it takes to convince a 43-degree alarm that the resting mouth is scalding. The molecule adds no heat; it adds 1.9 kcal/mol of binding energy at the lever, and the lever translates it into six degrees of borrowed fever. We find this one of the most beautiful and sobering facts in the whole science: the gap between comfort and fire, in a channel this steep, is a rounding error in the energy of a chemical bond.

The channel, of course, defends itself, and the defense is written into that calcium preference from the architecture. Sustained opening floods calcium in; calcium binds calmodulin; calmodulin drives TRPV1 into a **desensitized** state in which it stops answering. That is the molecular basis of chili tolerance — why a regular eater feels less over time, why the second bite of a dish reads milder than the first — and it is also why capsaicin is studied as a long-acting analgesic: hold the channel open long enough and it falls silent, taking the nerve's other pain traffic down with it. We hold these three behaviors — heat, acid, and capsaicin — in one small table, because each is a door into the same pore, and because designing for the first means respecting the other two.

GATE ON TRPV1	NATURAL ROLE	CAPSAICIN'S EFFECT
heat > 43 °C	pain / burn warning	threshold lowered to ~37 °C (body temp)
protons (low pH)	acid sting	sensitizes the channel, synergizes with heat
vanilloid pocket	—	capsaicinoid binds, ~1.9 kcal/mol tips the gate open
sustained Ca^{2+} influx	feedback / protection	calmodulin-driven desensitization (tolerance, analgesia)

The shape of the molecule is the shape of the burn

Because the tail reads into a groove and the head into a committee of residues, small changes in the molecule become real, nameable differences in the burn — and the plant does not make one capsaicinoid but a family of them. When we dock the family into the pocket, each seats with the same vanillyl head anchored at the back and the tails splayed slightly differently along the groove, and those differences in fit are, quite literally, the different burns a taster describes: where the heat lands, how fast, how long it stays. The two most pungent members, capsaicin and dihydrocapsaicin, are nearly identical twins — the only difference is that single Δ6 double bond, present in capsaicin, saturated away in dihydrocapsaicin — and they sit at the ceiling of pungency together. Shorten the tail by a carbon, or straighten its branch, and the molecule reads measurably softer. The historical pungency equivalents, expressed on the Scoville scale against pure crystalline capsaicin, make the ladder plain.

CAPSAICINOID	ACYL TAIL	SCOVILLE EQUIVALENT	LOUDNESS VS CAPSAICIN	ΔΔG VS CAPSAICIN (37 °C)
Capsaicin	C9, Δ6 unsaturated, 8-methyl branch	16,000,000	1.00×	0.00 kcal/mol
Dihydrocapsaicin	C9 saturated, 8-methyl branch	16,000,000	1.00×	0.00 kcal/mol
Nordihydrocapsaicin	C8 saturated	9,100,000	0.57×	+0.35 kcal/mol
Homodihydrocapsaicin	C10 saturated	8,600,000	0.54×	+0.38 kcal/mol
Nonivamide (synthetic)	C9 straight, no branch	9,200,000	0.58×	+0.34 kcal/mol

The free-energy column is the quiet revelation. We get it by reading each loudness ratio as a binding-energy difference — $\Delta\Delta G = -RT \ln(\text{ratio})$, with RT = 0.616 kcal/mol at body temperature — and the whole family fans out across about a third of a kilocalorie per mole. Nordihydrocapsaicin, a milder, rounder heat, is only +0.35 kcal/mol less avid than capsaicin. The experience a taster would call "the difference between a scorpion pepper and a hot sauce" is, at the pocket, a few tenths of a bond's worth of fit. This is exactly why the fatty-acid river matters so much to us, and why the engineered Pun1 weld was worth redesigning: shift the blend a little toward capsaicin and dihydrocapsaicin and away from the shorter tails, and the plant's output speaks, molecule for molecule, measurably louder — not because there is more of it, but because what there is sits a few tenths of a kilocalorie deeper in the groove.

The dose–response curve, and where EC50 lives

So far we have talked about one molecule at the lever. A tongue meets a cloud of them, and the channel answers not yes-or-no but *how much*, as a fraction of its subunits open. That answer is a dose–response curve, and its standard, honest shape is the Hill equation:

$$\theta = [A]^n \,/\, (EC_{50}{}^n + [A]^n)$$

where θ is the fraction of full activation, [A] is the capsaicinoid concentration, EC_{50} is the concentration that opens the channel halfway, and n is the Hill coefficient — the steepness, which reports cooperativity. For TRPV1, with its fourfold architecture and multiple coupled sites, n runs around 2, and that single fact sharpens the whole curve. Work it through at n = 2, in multiples of EC_{50}, and the switch-like character of the burn appears in the arithmetic.

CONCENTRATION [A]	θ (FRACTION OPEN), N = 2	IN WORDS
0.10 × EC_{50}	0.010	detection floor — barely there
0.33 × EC_{50}	0.10	the heat registers
1.0 × EC_{50}	0.50	**half-maximal — this is EC_{50}**
3.0 × EC_{50}	0.90	nearly full
10 × EC_{50}	0.99	saturated; the pore is as open as it gets

Read the span: with n = 2, a mere ninefold change in concentration — from 0.33× to 3× EC_{50} — carries the channel from one-tenth open to nine-tenths open. Had the channel no cooperativity, n = 1, the same journey would demand an eighty-one-fold change. The tetramer's cooperativity compresses the entire perceptible rise into a single decade of concentration, which is why heat has an *edge* to it, a line you cross rather than a slope you climb. And notice the bottom row and the top row together: below about a tenth of EC_{50} there is effectively nothing, and above about ten times EC_{50} there is nothing *more* — every channel is already open. That saturation ceiling will matter when we ask what the loudest peppers are actually doing.



FIG. 7C — Phoenix in the controlled chamber, where EC_{50} is read the same every time. A score is only as trustworthy as the conditions that produced it; half-maximal activation is measured at a fixed temperature, because for this channel temperature and capsaicin are the same lever.

From half-activation to a score

EC_{50} is the hinge of the whole Index, because EC_{50} is intrinsic: it belongs to the molecule and the pocket, not to how much you happen to have. A capsaicinoid with a lower EC_{50} opens the channel at a lower concentration — it is *louder per molecule* — and that per-molecule loudness is precisely what we print as IFI-H. The definition is one line:

$$\mathrm{IFI}(\text{candidate}) = 1{,}000 \times EC_{50}(\text{reference}) \div EC_{50}(\text{candidate})$$

Invert the EC_{50} so that louder reads as higher, scale against a fixed native wild-type reference, and multiply by a thousand so the reference lands at a clean 1,000. The anchor is exact by construction: when a candidate's EC_{50} equals the reference's, the ratio is 1 and the score is 1,000. Everything above 1,000 is a molecule that speaks more loudly than the native plant; everything below, more softly.

Run the definition backward and it becomes a measuring tape. Rearranged, $EC_{50}(\text{candidate}) = 1{,}000 \times EC_{50}(\text{reference}) \div \mathrm{IFI}$ — so any IFI-H on the scorecard *is* an effective EC_{50}, stated in different units. Fix the native reference at a representative half-maximal concentration of 0.50 µM — squarely in the sub-micromolar band this channel is known for — and the flagship ladder resolves into concentrations you could pipette.

CULTIVAR	SKU	PREDICTED SHU	IFI-H	EFFECTIVE EC_{50} (DERIVED)
native reference	—	—	1,000	500 nM
Ember	SS-0087	866K	1,570	318 nM
Kindle	SS-0089	1.25M	2,000	250 nM
Cinder	SS-0088	1.05M	2,290	218 nM
Blaze	SS-0091	1.72M	2,630	190 nM
Wildfire	SS-0092	1.96M	2,930	171 nM
Scorch	SS-0093	2.21M	3,220	155 nM
Firestorm	SS-0094	2.47M	3,490	143 nM
Solara	SS-0095	2.72M	3,980	126 nM
Phoenix	**SS-0096**	**2.95M**	**4,200**	**119 nM**

Phoenix's IFI-H of 4,200 says its capsaicinoid speaks 4.2 times louder than reference per molecule — its effective EC_{50}, 119 nM, is 4.2 times lower. Translate that loudness ratio into free energy, as we did for the capsaicinoid family — $\Delta\Delta G = -RT \ln(4.2) = -0.616 \times 1.435$ — and it comes to about **−0.88 kcal/mol**. Ember, at the bottom of the flagship ladder, sits at $\Delta\Delta G \approx -0.28$ kcal/mol above reference. So the entire engineered ladder, from Ember to Phoenix, spans

> $\Delta\Delta G = -RT \ln(4{,}200 / 1{,}570) = -0.616 \times \ln(2.68) \approx$ **−0.61 kcal/mol.**

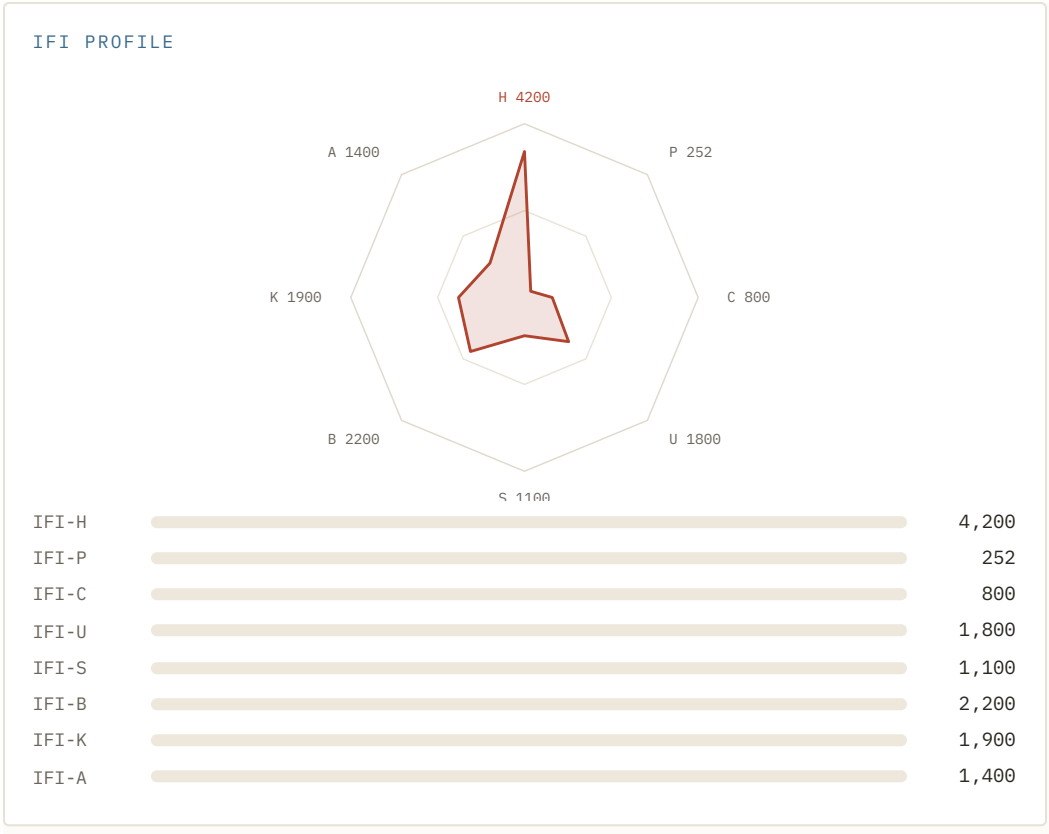
The whole dramatic distance from a warning shot to a flatline — the entire designed range of per-molecule loudness we are proud of — lives inside a free-energy window narrower than one hydrogen bond. That is the discipline the number imposes and the respect it demands: you are not splashing heat around, you are machining tenths of a kilocalorie into a pocket a few ångströms wide, and the difference between cultivars is a difference you could not feel with your hands but can read to three figures on the page.

Loudness times abundance

But a plant is not a single molecule at the lever; it is a reservoir of them, and the heat you actually feel is loudness *and* abundance together. That is the two-term product the Index was built to separate, and it is also what the old Scoville scale rolled into one. Remember what a Scoville Heat Unit originally meant: it is a dilution factor, the number of times you must dilute an extract before a trained panel can no longer detect any heat. You dilute until the capsaicinoid concentration drops below the detection floor — below about a tenth of EC_{50}, the bottom row of the dose–response table — and the reading is how far you had to go. Written out, that makes SHU proportional to the ratio of how much capsaicinoid the fruit holds to how little the pocket needs:

> $SHU \propto [C] \div EC_{50}_floor \propto [C] \times loudness$

Abundance times loudness — the same two terms. The old scale fused them; the Index pries them apart, assigning loudness to IFI-H (pure EC_{50}, the molecule and the pocket) and letting SHU carry what it was always really measuring once loudness is known: how much the plant pours out. For a fixed molecule, SHU simply tracks [C].



Which brings us, at last, to Phoenix — SS-0096, the cultivar our family engineered gene by gene because the plant that could reach it did not yet exist. Its IFI-H of 4,200 is the loudness; our model predicts a heat output of approximately **2.95 million SHU**. Set that against the theoretical ceiling — pure crystalline capsaicin tops the Scoville scale at 16,000,000 — and Phoenix is designed to reach just over **18 percent** of the pungency of the isolated molecule. A living fruit, designed to carry nearly a fifth of the loudness of pure capsaicin: that is what the two AI-designed Pun1 variants bought — **S39L and L345G**, written in with CRISPR, designed to lift peak output while holding the batch-to-batch coefficient of variation to **≤ 7 percent**, against the **42 percent** that wild-type biology hands you. Loudness pins the per-molecule score; abundance drives the pore far past the saturation ceiling; reproducibility is what makes the score worth printing at all.

And this is where saturation earns its keep as an honest caveat. From the dose–response table, past about ten times EC_{50} every channel is already open — instantaneous activation cannot climb higher. So the enormous predicted abundance of a Phoenix does not buy a *higher* peak than a merely ferocious pepper already reaches; what it buys is everything around the peak: how fast the threshold is crossed, how much of the mouth's tissue is recruited, and how long the channels stay driven before they clear. The extra millions of Scoville units are spent on *onset, spread, and dwell* — the long, total, inescapable quality of the burn — not on a taller instant. A number that did not separate loudness from abundance could never tell you that. The Index can.

Desensitization, and the honest limits

A measurement earns trust by naming what it cannot do, so we will be plain about the seams. IFI-H is read from the channel under controlled conditions — a fixed temperature, fresh receptors, a clean buffer — and the tongue is none of those things. Three honest gaps separate the cuvette from the meal, and we hold all three.

The first is **matrix.** Capsaicin is fat-loving; it was built, at the tail, to read into a membrane. In food that means fat carries it, buffers it, and slows its arrival — a chili bloomed in oil lands differently from the same chili in a watery broth, not because the molecule changed but because its delivery did. IFI-H scores the molecule at the pocket; it does not score the vinaigrette. A formulator reads the two together.

The second is **adaptation.** Because the channel protects itself through calcium and calmodulin, perceived heat is not linear in dose across a meal — the desensitization that gives regular eaters their tolerance also means the fifth bite genuinely activates fewer channels than the first. IFI-H describes a rested channel meeting a molecule. It does not, and cannot, describe the channel three minutes into dinner. That is a property of the eater and the evening, not of the cultivar, and it is right that the number stays silent on it.

The third is **the measurement itself**, and it is the one we can actually engineer. EC_{50} is temperature-dependent — for this channel, heat and capsaicin are the same lever, so a reading taken a few degrees off is a reading of a different quantity — which is why we fix the chamber temperature before we trust a single value. And the raw biology is noisy: wild-type capsaicinoid output swings batch to batch by that **42 percent** CV, enough that a "score" taken from an unengineered plant would be a guess dressed up as a figure. The entire point of the ≤ 7 percent CV we design toward is that it makes IFI-H *reproducible* — a number you can put in a formulation, cite in a paper, and hand to another lab expecting the same answer back. A score with a 42 percent CV is poetry; a score with a 7 percent CV is a specification. We wanted a specification.

What the number never claims is the last thing of all: whether you will *like* it. IFI-H measures how loudly a molecule speaks to a receptor, not whether the body it reaches wants to hear it. Intensity is physics; pleasure is yours. We measure the first precisely so that the second stays entirely, gloriously, your own.

These are computational designs; Scoville Splice's cultivars and constructs have not been experimentally validated in a laboratory.

So the pathway ends here, in a pocket six ångströms across, where a molecule our family coaxed out of two biosynthetic rivers hydrogen-bonds to three residues, tugs a single short lever by less than two kilocalories per mole, and persuades a heat sensor to report a fire that is not there — faithfully, repeatably, to three figures. Everything that walked us to this pocket was a walk toward this one binding event. Everything that follows is the same walk to seven other locks: another receptor, another EC_{50}, another number waiting to be made honest. The lock is the same shape every time. Only the key, and the reverence we bring to cutting it, changes.



PART III OF IV

The Other Seven Branches

cool · pungent · umami · kokumi · sweet · astringent · bitter — the same method, seven more routes

Chapter 8 · Cool-Burn — *the five-gene climb to menthol.*

Heat lowers a threshold to make a fire that is not there; cool does the mirror image, and the symmetry is one of the quiet delights of sensory biochemistry. The molecule is menthol, the receptor is TRPM8, and where capsaicin makes a heat sensor report a warmth it is not feeling, menthol makes a *cold* sensor report a chill the tissue does not have. Both are honest lies told to the nervous system, and both are told in the same grammar — a small molecule slipping into a thermosensitive ion channel and moving the temperature at which that channel decides to open. We find the symmetry beautiful the way a mathematician finds a dual theorem beautiful: two problems that looked unrelated turn out to be one problem read from opposite ends.

But the road to menthol could not be less like the road to capsaicin. There is no aromatic trunk here, no phenylalanine deaminated into cinnamic acid, no confluence of two rivers meeting at an acyltransferase. Menthol is a terpene — built not from a benzene ring but from the universal five-carbon isoprene units that every plant stitches together to make everything from carotenoids to rubber. Its synthesis is a single, orderly, almost architectural climb: a staircase of enzymes that take a floppy open chain, fold it into a ring, and then walk it up through a disciplined alternation of oxidations and reductions until a loose, citrus-smelling hydrocarbon has become a rigid, crystalline, cool-tasting alcohol. We have grown to love this pathway for its tidiness. Capsaicin is a river delta. Menthol is a flight of stairs.



FIG. 8 — Peppermint, *Mentha × piperita*. The menthol lives in glandular trichomes dotting the leaf surface; the committed climb runs inside microscopic oil glands you can rupture with a fingernail and smell on your skin for an hour.

The mirror at the membrane

TRPM8 is a tetramer — four identical subunits arranged around a single ion-conducting pore, each subunit donating six transmembrane helices, S1 through S6. The last two helices of each subunit (S5–S6) line the gate at the center; the first four (S1–S4) fold off to the side into a bundle called the voltage-sensor-like domain,

and it is into a pocket *within that bundle* — not at the pore itself — that menthol slides. A small set of residues forms the pocket; the one the structural literature returns to again and again is a tyrosine, Y745, on the second helix, whose aromatic face and hydroxyl cradle menthol's ring and its lone oxygen. Icilin, the synthetic super-cooler, needs a different handhold — a residue near G805 on the third helix, and a calcium ion as a chaperone — which is why icilin and menthol, though they taste alike, do not bind alike. The receptor has more than one door into the same room.

The channel is a polymodal thermosensor: it folds three inputs — membrane voltage, temperature, and ligand — into one binary decision, open or closed. The elegant part is that these are not three independent switches but three coupled terms, each leaning on the same gate. Cold, on its own, shifts the channel's voltage dependence so that at the resting voltage of a sensory neuron a fraction of the channels open. Menthol does the same shifting, and more cheaply: bound in the S1–S4 pocket, it stabilizes the open conformation and slides the whole activation curve toward warmer temperatures and less negative voltages.

The practical consequence is exact and worth stating plainly. A bare nerve ending at body temperature, 37 °C, normally reads *not cold*, because TRPM8's activation threshold sits down near 26 °C. Put menthol in the pocket and that threshold climbs — toward 30, 33, 35 °C — until body heat itself falls on the cold side of the line. The tissue has not lost a single degree. The channel has simply moved the goalpost, opened its gate, let calcium and sodium flood through, and fired an action potential the brain can read only one way: *cold*. That is the precise mirror of heat. Capsaicin does not warm the tongue; it lowers TRPV1's heat threshold from about 43 °C down past body temperature, so the mouth's resting warmth reads as burn. Menthol does not chill the tongue; it raises TRPM8's cold threshold up past body temperature, so the mouth's resting warmth reads as chill. Two molecules, two channels, one trick played in two directions.

One number governs where a cooling plant will land, and it is the concentration of menthol that drives TRPM8 to half its maximal response — its EC_{50}. For menthol at TRPM8 that value sits characteristically in the tens of micromolar, a long way from the sub-micromolar potency with which capsaicin commands TRPV1. In plain terms the cold receptor is a less sensitive instrument than the heat receptor: you need more menthol to feel cool than capsaicin to feel hot, by something like fifty- to a hundred-fold in raw molar terms. That single fact sets peppermint's whole place on the cooling ladder, and we will make it quantitative before the chapter closes.

The five-gene climb

The staircase begins, in a sense, below the molecule, with the universal currency of all terpene chemistry: isopentenyl diphosphate (IPP) and its isomer dimethylallyl diphosphate (DMAPP), the two five-carbon bricks from which every terpene is laid. In mint, uniquely among our eight axes, those bricks are not minted in the cytosol by the familiar mevalonate route; they are made inside the plastid by the methylerythritol phosphate (MEP) pathway, an assembly line that

starts from pyruvate and glyceraldehyde-3-phosphate and whose first committed enzyme, DXS, sets how much raw isoprenoid the gland will have to spend. So cool-burn, alone among the axes, begins in a different cellular compartment from where it ends — a detail that matters enormously when we engineer it, because you cannot push a pathway by feeding its last enzyme if its first bricks are being rationed two membranes away.

A prenyltransferase, GPP synthase, condenses one IPP onto one DMAPP to make the ten-carbon chain geranyl diphosphate (GPP) — the open, floppy precursor, the bottom step. Then a terpene cyclase, (−)-limonene synthase, performs the move that defines the whole family: it ionizes GPP, lets the chain curl, and snaps it shut into a six-membered ring — (−)-limonene, the clean citrus-peel monoterpene. With the ring cast, the committed climb to menthol begins, and here the chapter's title earns itself. Of the seven enzymatic stations from bricks to menthol, the first two are shared infrastructure: GPP synthase and limonene synthase feed many terpenes, not menthol alone. The five that follow are the committed, menthol-defining genes — the *five-gene climb* — a strict alternation of one oxidation and a run of reductions that walks limonene, rung by rung, to (−)-menthol.

#	STATION	TRANSFORMATION	ENZYME CLASS	COMPARTMENT
1	GPP synthase	IPP + DMAPP → geranyl-PP	prenyltransferase	plastid
2	(−)-limonene synthase	GPP → (−)-limonene	terpene cyclase	plastid
3	**L3OH · CYP71D**	**(−)-limonene → (−)-*trans*-isopiperitenol**	**P450 hydroxylase (oxidation)**	**ER membrane**
4	**IPD**	**isopiperitenol → (−)-isopiperitenone**	**dehydrogenase (oxidation)**	**cytosol**
5	**ISPR**	**isopiperitenone → (+)-*cis*-isopulegone**	**reductase**	**cytosol**
6	**PR**	**(+)-pulegone → (−)-menthone**	**reductase**	**cytosol**
7	**MMR**	**(−)-menthone → (−)-menthol**	**reductase**	**cytosol**

A cytochrome P450, limonene-3-hydroxylase (L3OH, CYP71D13/15), hangs a single oxygen on carbon 3 of the ring — the regiochemical decision on which everything downstream depends — to give (−)-*trans*-isopiperitenol. A dehydrogenase (IPD) lifts that alcohol to the ketone (−)-isopiperitenone. A reductase (ISPR) bends it to (+)-*cis*-isopulegone, which relaxes to (+)-pulegone, the pivotal intermediate, the landing where the staircase briefly forks. Pulegone reductase (PR) carries pulegone on to (−)-menthone; and the final reductase, menthone reductase (MMR), delivers (−)-menthol itself — stable, crystalline, the cool end of the climb. Four of the five committed

enzymes are reductases quietly spending NADPH, the cell's reducing currency; the plant pays, hydride after hydride, to fold a smell into a solid.

Two facts give the cooling plants their range, and the first is simply *where* all of this happens. Not in the leaf's working mesophyll but in those microscopic peltate glandular trichomes studding the leaf surface — tiny oil factories, each a rosette of eight secretory cells capped by a cuticle blister that fills with menthol and its relatives and holds them under pressure until a fingernail or a tooth ruptures it. The entire staircase runs inside those glands, walled off from the leaf's own metabolism, which is how mint can hoard monoterpenes to concentrations that would dissolve its own membranes if they were loose in the cytoplasm. And the staircase migrates as it climbs: the plastid casts the ring (steps 1–2), the endoplasmic reticulum's P450 hangs the first oxygen (step 3), and the cytosolic enzymes finish the work (steps 4–7). A single molecule of menthol is a small pilgrimage across three compartments of one secretory cell. We do not find that merely interesting; we find it instructive, because every compartment boundary is a place where flux can stall, and to know the map is to know where the leaks will be.



FIG. 8b — Peppermint, in section. The clean, oil-bearing structure of the leaf is where the committed climb is housed; menthol is held under pressure in cuticle blisters and never touches the cell's working cytoplasm.

The off-ramp we close

There is a villain on this staircase, and managing it is most of the art. At the pulegone landing, a branch enzyme — menthofuran synthase (MFS, CYP71A32) — can seize pulegone before pulegone reductase does and divert it into menthofuran, a muddy, musty, faintly stale off-note that ruins the clean cool. Stressed mint — grown too hot, lit too long, or short of water — throws its flux toward menthofuran, and a great deal of what separates a fine peppermint from a harsh one is simply how little menthofuran rode along. So one of the most valuable edits we make on this entire route adds nothing at all. It *quiets MFS*. Turn the branch enzyme's expression down — a subtractive edit, antisense or co-suppression, lowering a gene rather

than writing one in — and the pulegone that MFS no longer takes is not lost; it flows to PR and arrives as clean menthol. It is one of the quiet laws of this work, and the cooling plants teach it more plainly than anything else we grow: sometimes the most important thing to engineer is a step you want to *stop*.

The arithmetic of that landing is a small competition. Pulegone arrives at a rate set by everything upstream; two enzymes, PR and MFS, bid for it. The fraction that becomes menthol rather than menthofuran is, to first order, PR's catalytic pull divided by the sum of both — v_PR ÷ (v_PR + v_MFS). Halve MFS and you do not merely cut the off-note in half, because the pulegone MFS surrenders is handed straight to PR: the numerator rises as the denominator falls. If PR and MFS start evenly matched, the clean fraction is 1 ÷ (1 + 1) = 0.50; cut MFS to a quarter of its old activity and the clean fraction becomes 1 ÷ (1 + 0.25) = 0.80. A fourfold quieting of one gene lifts clean menthol from half the flux to four-fifths of it. That is why closing a single off-ramp buys more than its own small size suggests: it is a lever on a ratio, not a subtraction from a sum.

One carbon, two mints

The second fact that gives the cooling plants their range is a marvel of economy: the entire difference between a peppermint and a spearmint is a single carbon. Peppermint hangs its first oxygen on carbon 3 of the limonene ring, with L3OH (CYP71D13/15), and so enters the staircase that ends in menthol. Spearmint hangs the same oxygen on carbon 6 instead, with a close cousin enzyme, limonene-6-hydroxylase (CYP71D18), and so routes its flux to *trans*-carveol and onward to carvone — the warm, caraway-sweet note that *is* spearmint, and that is not menthol at all. One enzyme's choice of where to place one atom of oxygen is the whole difference between the cool of peppermint and the sweetness of spearmint. Knowing that, we can slide a cooling plant anywhere we like along the spectrum between them simply by choosing which hydroxylase to favor, and by how much — a dimmer, not a switch.



FIG. 8c — Spearmint, *Mentha spicata*. The same ring, the same precursor, one carbon's difference in where the first oxygen is placed — and the staircase leaves menthol for carvone.

And then stereochemistry, which is where the climb earns its precision. Only (−)-menthol carries the clean, strong, familiar cool. Menthol has three stereogenic centers — the carbon bearing the hydroxyl, the one bearing the isopropyl group, the one bearing the methyl — and three stereocenters mean 2^3 = eight stereoisomers,

four pairs of mirror images: menthol, neomenthol, isomenthol, and neoisomenthol, each a (+)/(−) pair. All eight share one molecular formula, $C_{10}H_{20}O$, and one molar mass, 156.27 g·mol^{-1}, and all eight smell more or less of mint. But only one of them, (1R,2S,5R)-(−)-menthol, drops cleanly into the Y745 pocket of TRPM8 and rings the cold bell at full volume. Its own mirror image, (+)-menthol, binds weakly; the neo- and iso- diastereomers bind weakly and drag harsher notes behind them. Read correctly, the whole staircase is a stereochemical funnel: each enzyme in the climb is enantio- and diastereoselective, choosing one three-dimensional form at every rung and discarding the seven wrong ones, so that what arrives at the bottom is not "menthol" in the loose sense but nearly pure (1R,2S,5R)-(−)-menthol. This is why we care, almost fussily, which enzyme variant sits at each step. A sloppy reductase that lets a few percent of the wrong diastereomer through does not make a cool a few percent weaker; it makes a cool that is mostly clean and a little muddy — and muddiness is exactly what the tongue notices. A clean plant, grown to make one molecule with discipline, is the whole point of the work, and the funnel is how a leaf achieves what a careless synthesis cannot.

The cooling plants, counted

The platform's cooling hosts tell an honest story of difficulty and focus if you simply lay their construct counts side by side.

HOST	CONSTRUCTS	PRIMARY AXES	DIFFICULTY
Peppermint	2,360	C / M	Hard
Lavender	160	C	Hard
Orange	120	B / C	Hard
Eucalyptus	80	C	Hard
Spearmint	80	C	Hard
Camphor	40	C	Hard

Peppermint, with 2,360 constructs, carries the overwhelming majority — about 83% of the 2,840 constructs spread across these dedicated cooling hosts — because peppermint is where the five-gene climb is best understood and most worth tuning. Lavender (160), eucalyptus (80), spearmint (80), and camphor (40) are variations on the same terpene theme, each emphasizing a different step and leaving different off-ramps open; orange (120) sits across two axes, lending a cool-bitter edge to citrus. Every one is marked *Hard*, and the marking is earned. A pathway that spans three compartments, hinges on one regiochemical choice, funnels through eight stereoisomers, and forks at a stress-sensitive branch is not a pathway anyone nudges casually. It is a staircase you learn step by step or not at all.

Reading cool as a number

The Index turns all of that biology into one figure you can set beside a tomato's sweetness or a pepper's heat. The definition is deliberately spare:

$$\mathrm{IFI}(c) = 1{,}000 \times \mathrm{EC}_{50}(\mathrm{ref}) \div \mathrm{EC}_{50}(c)$$

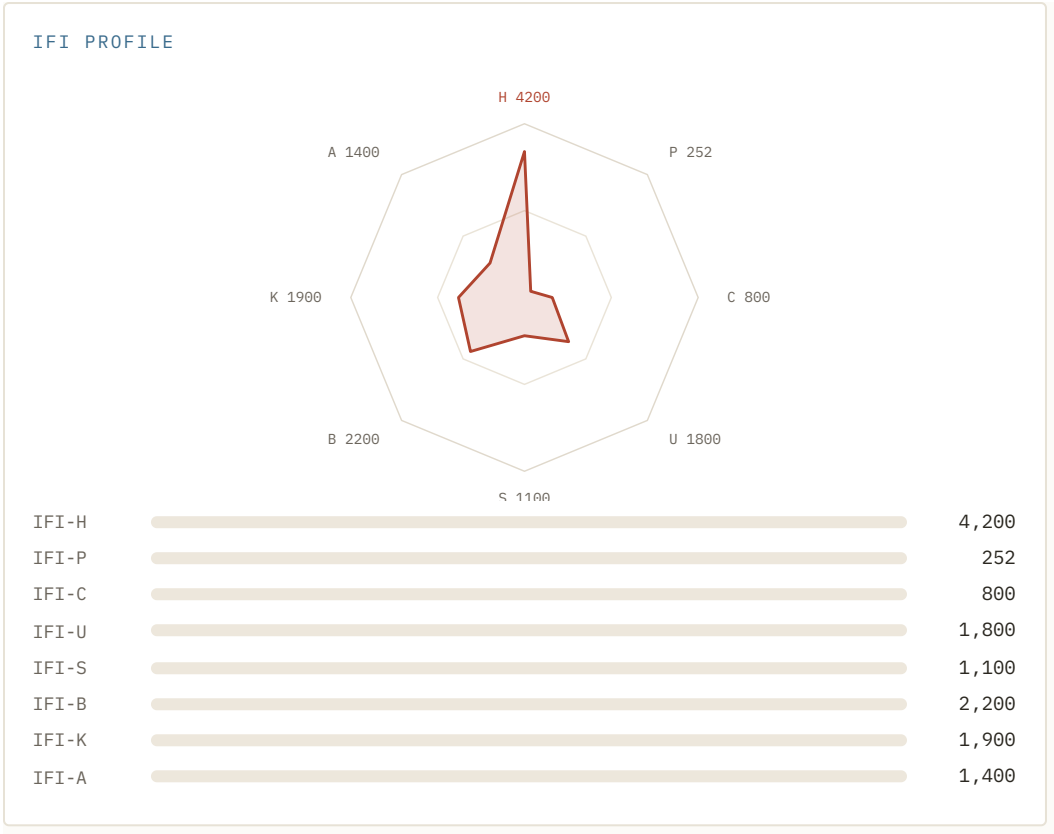
EC_{50} is the one concentration on a dose–response curve that is defined identically for every molecule and every receptor: the concentration producing half of the maximal response. We anchor the Index on it, rather than on "threshold," for a reason the cooling receptor makes vivid. TRPM8's response to menthol is cooperative — the channel is a tetramer, and binding in one subunit eases binding in the next — so its dose–response climbs steeply. Model it with the Hill equation,

$$\theta(L) = L^{n} \div (\mathrm{EC}_{50}{}^{n} + L^{n}),$$

where θ is the fraction of maximal response and *n* the Hill coefficient. For TRPM8–menthol *n* runs around 2 to 3; take *n* = 2.5 and, for a worked example, a reference EC_{50} of 50 µM. At one-half the EC_{50} (25 µM): $\theta = 0.5^{2.5} \div (1 + 0.5^{2.5}) = 0.177 \div 1.177 = 0.15$ — fifteen percent of full cool. At twice the EC_{50} (100 µM): $\theta = 2^{2.5} \div (1 + 2^{2.5}) = 5.66 \div 6.66 = 0.85$ — eighty-five percent. A mere fourfold change in menthol concentration, 25 µM to 100 µM, swings the sensation across almost its entire range. "Threshold," on a curve that steep, is a moving target; the half-maximal point is the one honest landmark, which is why the Index is built on it.

Three things fall straight out of the formula. When a candidate matches the reference exactly, $EC_{50}(c) = EC_{50}(\mathrm{ref})$, the ratio is one, and IFI = 1,000 — parity. Because the candidate's EC_{50} sits in the denominator, loudness is inversely proportional to it: a candidate twice as potent (half the EC_{50}) scores 2,000, and the scorecard's normalization ceiling of 5,000 corresponds to a candidate five times more potent than the reference. A candidate weaker than the reference scores below 1,000.

That last case is where our fullest pepper sits on this particular axis. Phoenix — the one cultivar we designed to register on all eight axes at once — carries a cool-burn score of 800. Read the formula backward: $800 = 1{,}000 \times EC_{50}(\mathrm{ref}) \div EC_{50}(c)$ means $EC_{50}(c) = 1.25 \times EC_{50}(\mathrm{ref})$. Phoenix's cooling system is a quarter weaker than the reference anchor — a deliberate whisper of menthol-like cool threaded through a plant built mainly for heat and savor. Its full profile reads:



The 4,200 of heat and the 800 of cool live in the same fruit without cancelling, because they speak to different receptors on different nerves — TRPV1 and TRPM8 — and the tongue sums them rather than averaging them. That is the whole promise of designing on eight axes instead of one: a cool that does not subtract from a heat, a heat that does not drown a savor.

To make the ratios tangible, here is the cool-burn ladder worked both ways — as a pure, assumption-free ratio, and then as an illustrative concentration taking the reference menthol EC_{50} at 50 µM.

IFI-C	$EC_{50}(C) \div EC_{50}(REF)$	$EC_{50}(C)$, MM*	$EC_{50}(C)$, MG·L⁻¹*
500	2.00	100.0	15.63
800 (Phoenix)	1.25	62.5	9.77
1,000 (reference parity)	1.00	50.0	7.81
2,000	0.50	25.0	3.91
5,000 (scorecard ceiling)	0.20	10.0	1.56

*The ratio column is exact. The concentration columns are illustrative: the reference EC_{50} is set at 50 µM only to put a human scale on the ratios; move the anchor and the millimole columns move with it, but the ratios do not.

The milligram column is just the molar column wearing different clothes: multiply micromolar by menthol's molar mass, 156.27 $g \cdot mol^{-1}$, and divide by a thousand. Fifty micromolar is 7.81 $mg \cdot L^{-1}$ — call it 7.8 parts per million — and the scorecard ceiling's ten micromolar is 1.56 $mg \cdot L^{-1}$. These are vanishingly small amounts of matter for so loud a sensation: a few milligrams of the right stereoisomer in a litre, reporting to the brain as a mouthful of winter. We state the conversion because a number without its units is a rumor, and we would rather hand a reader the arithmetic than ask for faith.

One subtlety the formula hides on purpose. An IFI-C of 1,000 and an IFI-H of 1,000 both mean *parity with this axis's own reference* — not equal molar potency. Capsaicin commands TRPV1 at sub-micromolar concentrations; menthol needs tens of micromolar to command TRPM8; in raw molarity the heat receptor is perhaps fifty to a

hundred times the more sensitive instrument. The Index normalizes that gulf away so a cook can read heat and cool on one page without a chemistry degree. It is the right choice for a standard, but it carries a quiet obligation: a number is always a comparison, and to read it well you must know what it is compared to. Each axis is measured against its own reference, never against another.

Precision, and what we do not claim

A score is only as trustworthy as it is repeatable, and here the cooling plants let us say something exact. Suppose the menthol a batch delivers varies, harvest to harvest, with some coefficient of variation. How much does the *sensation* vary? Not by the same fraction — because the dose–response curve compresses near its middle. The local sensitivity of the Hill response to a change in concentration is, at the half-maximal point, exactly $n/4$ per natural-log unit of concentration. For $n = 2.5$ that slope is 0.625, so a batch CV of C translates into a swing in perceived cool of roughly $(n/4) \times \ln(1 + C)$.

An uncontrolled, wild harvest of mint runs a menthol CV around 42% — the same batch-to-batch lottery that undisciplined capsaicin breeding shows. That maps to $0.625 \times \ln(1.42) = 0.625 \times 0.351 \approx 0.22$: a twenty-two-point swing in cool intensity near the middle of the curve — a cool you cannot promise. Our platform holds batch CV at or below 7%. That maps to $0.625 \times \ln(1.07) = 0.625 \times 0.068 \approx 0.042$ — a swing of about four points. The curve's own steepness, which makes menthol feel sudden, also compresses our tightened input into an output tighter still. We do not chase ≤7% for its own sake. We chase it because a flavor you cannot reproduce is not a flavor, it is an anecdote — and because giving someone the same clean cool twice is a small, repeatable form of respect.

And because numbers are a form of respect, so is saying plainly what they do not mean. Menthol does not cool anything; it reports a cold that is not there, and in quantity it can tip from pleasant to raw — TRPM8, over-driven, desensitizes, and the cool dulls just as a capsaicin burn dulls, so more is not more for long. The scorecard ceiling of 5,000 is a normalization bound, a frame for the picture, not a target we pretend every plant should reach; most of what makes a cool *good* lives well below it, in cleanliness rather than loudness. The stereochemical funnel can leak: a reductase variant that passes a few percent of the wrong diastereomer costs a clarity the tongue registers even when a chromatogram barely does. The menthofuran off-ramp reopens under stress, so a cool engineered in a clean chamber must be re-earned in a hot field. And the EC_{50} values we reason with are receptor-level half-activations from controlled assay conditions, not promises about one specific mouth on one specific afternoon — the body runs warmer or cooler, wetter or drier, and the same milligrams land a little differently each time. We build the Index to be exact about what is exact and honest about what is not. These are computational designs; Scoville Splice's cultivars and constructs have not been experimentally validated in a laboratory.

That, finally, is the discipline the five-gene climb teaches. The cooling plants — peppermint with its 2,360 constructs, and spearmint, lavender, eucalyptus, and camphor arranged around it —

are all one terpene staircase, read with different emphases, and we tune every one of them the same way. Feed the plastid so the five-carbon bricks are never rationed. Drive the committed five-gene climb from limonene to menthol. Quiet the menthofuran off-ramp so the flux stays true. Guard the stereochemistry so that only (1R,2S,5R)-(−)-menthol reaches the bottom of the stairs. And hold the batch so tight that the cool arrives the same every time. Do all of that, and a leaf you can rupture with a fingernail will tell a warm mouth — honestly, and on command — that winter has come. We find nothing about that ordinary. We find it worth a lifetime.

Chapter 9 · Pungency — *two routes to the same sharp channel.*

Pungency is the sharp, volatile, eye-watering sensation that lives in mustard and wasabi and horseradish and raw garlic — not the slow radiant heat of a chili but a bright assault that climbs the nose as much as it crosses the tongue. Grate fresh wasabi and the feeling arrives *before* the flavor does: it rushes the sinuses, lifts water to the eyes, and is nearly gone before you can name it. Its receptor is TRPA1, the channel the literature nicknames the wasabi channel, and TRPA1 is a fundamentally different kind of lock from the receptors that merely cradle a molecule and let it go. Most tastants *bind* — capsaicin settles into a pocket, menthol into a cleft, reversibly, the way a guest settles into a room and later leaves it. TRPA1's agonists do not merely bind. They *react*. They are electrophiles: chemically reactive molecules that form covalent bonds with the protein itself, altering the channel rather than visiting it. Pungency is not a handshake. It is a small, controlled, reversible chemical burn. And the plants that make it arrive at that one reactive trick by two completely unrelated roads — one through the mustard family, one through the alliums — which is why, of all the axes we have drawn, this is the chapter about convergence: two biosyntheses, traced gene to molecule, meeting at a single sulfur atom on a single cysteine.

The lock that reacts

Hold the mechanism in your hand for a moment, because everything downstream depends on it. TRPA1 is a large tetrameric cation channel, and the part that matters for pungency is not a surface pocket but the long cytoplasmic skirt of ankyrin repeats that hangs below the membrane on the inside of the cell. Studded along that inner face are a handful of cysteine residues whose sulfur-bearing thiol side chains sit slightly exposed and slightly acidic — reactive, as cysteines go. The literature numbers the three that carry most of the signal near positions **621, 641, and 665**, with a reactive lysine nearby. An electrophile that reaches the cytoplasm finds these thiols and performs the oldest reaction in defensive chemistry: it adds to them covalently, forming an adduct that changes the local shape of the skirt, and that shape change is transmitted up into the gate. The channel opens, calcium and sodium pour in, the trigeminal nerve fires, and the body reports a sharp sting that no thermometer could ever have measured.

The chemistry of the adduct differs by road while the target does not. An isothiocyanate carries a cumulated **–N=C=S** group whose central carbon is starved of electrons; the cysteine thiolate attacks that carbon and leaves behind a **dithiocarbamate** — a thiourea-like bridge welded onto the channel. A thiosulfinate like allicin carries an electron-poor divalent sulfur; the same thiolate attacks *that* sulfur and leaves behind a **mixed disulfide**, the cysteine now wearing an allyl group on a sulfur leash. Two different welds, the same three

studs. This is why pungency reads, neurologically, less like a flavor and more like an alarm: the signal is a reaction *with* the receptor, not a fit *into* it, and the nervous system evolved to treat reactive chemistry on a nerve ending as damage worth shouting about. The mercy in it is that the welds are reversible — the adduct hydrolyzes, the protein is repaired or replaced, the cysteine is freed — so the burn fades on its own timescale rather than scarring. We are, from the plant's point of view, an animal foolish enough to farm its wound response and call it dinner. We take that as a kind of trust, and we try to be worthy of it: these are plants we grow with intention, and the sharpness is something we *design*, not something we extract and concentrate blindly.

That covalent, multi-site chemistry also leaves a fingerprint in the math, which we will meet properly below: a channel modified at three cooperating cysteines does not switch on gently and linearly. It switches like a threshold. Keep that in mind — the steepness of the dose–response curve is the molecule's three welds made visible as a number.

Route one — the mustard bomb, drawn gene to molecule

The first road is the **glucosinolate–myrosinase system** of the mustard family: black mustard with its 2,790 constructs, white mustard, watercress, radish, the whole pungent *Brassica* clan. It is a two-part weapon, stored disarmed, and the elegance is that the plant keeps the explosive and the detonator in separate rooms until the instant of damage.



FIG. 9.1 — A radish in cross-section. A ⋆Brassica⋆ carries the glucosinolate charge in its living cells and the myrosinase detonator walled off in separate myrosin cells; intact, the root is nearly silent. The sharpness is a wound response, and the knife is the fuse.

The charge is the **glucosinolate** — a sulfur-and-nitrogen compound like **sinigrin** — and the plant builds it along a dedicated branch that a pathway book can draw step by step, because every step is a named enzyme doing a nameable chemistry. It begins by taking an amino acid, most often methionine, and *lengthening* it: the chain-elongation cycle deaminates methionine to its 2-oxo acid, condenses that with acetyl-CoA, isomerizes and decarboxylates, and transaminates back — each turn of the cycle adding a single methylene unit, like a smith drawing out a bar one hammer-blow at a time. Sinigrin comes from methionine lengthened by one such turn. The elongated amino acid is then handed to a cytochrome P450 of the **CYP79** family, which oxidizes it to an **aldoxime**; a second P450

of the **CYP83** family activates the aldoxime to a reactive species that conjugates with a thiol donor; a **C–S lyase** trims that adduct to a thiohydroximate; a **UGT74** glucosyltransferase caps the sulfur with a molecule of glucose; and finally a sulfotransferase of the **SOT** family sets the sulfate group that finishes the stable, stored glucosinolate. On its own it is harmless, water-soluble, and nearly tasteless — a charge with the fuse pulled.

STEP	ENZYME (FAMILY)	REACTION	PRODUCT
Chain elongation	BCAT4 · MAM	deaminate, +C_2, decarboxylate, transaminate	`homomethionine (per cycle)`
Core oxime	CYP79	amino acid → aldoxime	`(methylthioalkyl)aldoxime`
Activation	CYP83	aldoxime → aci-nitro / nitrile oxide	`thiol-reactive intermediate`
Cleavage	C–S lyase (SUR1)	cleave thiol conjugate	`thiohydroximate`
S-glucosylation	UGT74	+ glucose on sulfur	`desulfoglucosinolate`
O-sulfation	SOT	+ sulfate on oxime	**`sinigrin`** `(stored)`

The detonator is **myrosinase**, a β-thioglucosidase the plant keeps rigorously apart from its substrate, packed into dedicated myrosin cells. When tissue is bitten, crushed, or chewed, the walls between the two rooms break and enzyme meets charge. Myrosinase hydrolyzes the thioglucoside bond, snapping off the glucose and leaving an unstable aglycone — a thiohydroximate-O-sulfonate — that cannot hold its shape for more than an instant. It sheds its sulfate and undergoes a **Lossen-type rearrangement**: the side chain migrates from carbon to nitrogen, and out of that collapse springs the **isothiocyanate** — allyl isothiocyanate, the volatile that *is* mustard's entire character. The accounting is worth pausing on. Sinigrin is a heavy molecule, near 397 g/mol; the allyl isothiocyanate it yields is a slight one, 99.15 g/mol. Only about a quarter of the stored mass leaves as the pungent volatile. The other three quarters — glucose, sulfate — is the safety catch, the disarming scaffold the plant discards in the same breath it fires. The plant spends most of the molecule on *keeping it safe to carry*, which is the whole genius of a stored weapon.

The mustard road comes in more than one caliber, and that variety is a design surface. Glucosinolates are a family of well over a hundred compounds, sorted by the amino acid they descend from: **aliphatic** (from methionine, like sinigrin — the fast nasal strike), **indolic** (from tryptophan — slower, and the raw material of other defensive chemistry), and **aromatic** (from phenylalanine or tyrosine — deeper, more lingering burns). Each yields an isothiocyanate with its own sharpness, its own volatility, its own speed of arrival, so the family is less a single note than a small keyboard.

There is also a fork at the very instant of hydrolysis that we care about more than any other, because it is the difference between a weapon that fires and one that fizzles. A helper protein — the **epithiospecifier protein**, ESP — can intercept myrosinase's unstable aglycone and steer it away from the pungent isothiocyanate and toward a bland, odorless **nitrile** instead. A plant loaded to the brim

with glucosinolate can still taste flat if ESP runs high; the charge is there, the detonator fires, and the output is routed to a dud. So one of our most reliable levers on this axis is *subtractive*: quiet ESP, and the same hydrolysis that was leaking into nitrile now pours its full flux into isothiocyanate. It raises pungency not by opening a tap but by closing a leak — the kind of edit that makes a plant more itself rather than more loaded.

Route two — the garlic clock, drawn gene to molecule

The second road is the alliums' — garlic with its 560 constructs, onion — and it is the *same strategy* built from completely unrelated parts. Two components kept apart until damage, converging on the same channel, by chemistry that shares not a single enzyme with the mustards.



FIG. 9.2 — Garlic in cross-section. The alliin sits intact inside the clove's cells; the alliinase waits in separate compartments. The sharpness does not exist until you crush the two together — a whole clove is silent, and the knife is the enzyme's cue.

Garlic stores **alliin** — S-allyl-L-cysteine sulfoxide, an odorless sulfur amino-acid derivative built by S-allylating cysteine and then letting a flavin monooxygenase set the single oxygen on the sulfur that makes it a sulfoxide. Held beside it, in separate cellular compartments, waits **alliinase**, and alliinase is a beautiful enzyme to draw because its mechanism is textbook and exact. It is a **C–S lyase** that depends on **pyridoxal-5′-phosphate (PLP)**, the same vitamin-B6 cofactor that does carbon–sulfur and carbon–nitrogen surgery all over primary metabolism. When the clove is crushed and alliin reaches the enzyme, the cofactor forms a Schiff base with alliin's α-amino group, the enzyme plucks the α-proton, and a β-elimination tears the carbon–sulfur bond apart. Out of that single cut come two fragments: a molecule of **2-propenesulfenic acid** (allyl sulfenic acid)

and a molecule of dehydroalanine, which hydrolyzes away to the pyruvate and ammonia that are garlic's quiet metabolic exhaust. The sulfenic acid is the reactive fragment, and it is impatient: two of them condense, losing a molecule of water, into **allicin** — diallyl thiosulfinate, the compound that fills a kitchen the instant a clove is crushed and that does not exist, at all, in an intact bulb.

This is why a whole clove has no smell and a cut one has so much: the reaction is the fragrance. And allicin, like allyl isothiocyanate, is an electrophile. It carries that electron-poor thiosulfinate sulfur straight to the reactive cysteines on TRPA1's inner skirt, forms its mixed-disulfide weld on the same studs the mustard road targets, and holds the gate open. Two plant families, two unrelated storage molecules, two unrelated triggering enzymes — a thioglucosidase on one road, a PLP-dependent lyase on the other — converging on one channel by one covalent mechanism. It is one of the cleanest cases of convergent evolution in the whole flavor world: the plants did not inherit this trick from a common ancestor; they *reinvented* it, twice, because an electrophile aimed at a nerve ending is simply a very good way to be left alone. (We will say, once and plainly, that the folk pharmacology of raw garlic is not our subject. We score its sharpness, not its medicine, and we make no claim beyond the number.)

For us the convergence is a gift, because the *design* lever is shared across both roads. In each system the intensity we can dial is governed by the same two quantities: how much of the stored precursor the plant accumulates, and how briskly — and how cleanly — the triggering enzyme fires on damage. Drive the storage branch, speed and purify the activating enzyme, and the sharpness climbs, whether the carbon arrived through a glucosinolate or through alliin. One stored charge, one detonator, one reactive electrophile, one lock. The plant cuts both keys to the same cylinder.

The math of a chemical burn

A sensation this fleeting still has to become a number we can design against, and the number is built the same way every axis in this book is built. The Iyengar Flavor Index does not measure how *much* of a compound a food holds; it measures how *loudly* that compound speaks to its receptor, which is a far more honest thing to score.

$$\text{IFI-P}(c) = 1000 \cdot \text{EC}_{50}(\text{ref}) \,/\, \text{EC}_{50}(c)$$

EC_{50} is the concentration at which a stimulus drives its receptor to half of full response — the half-activation point. Invert it against a fixed reference, scale by a thousand, and loudness becomes a score. We anchor **IFI-P** so that **1,000** corresponds to our reference stimulus, allyl isothiocyanate — the defining voice of this axis, and a fitting zero-point because it is the molecule the mustard road exists to make. That anchor is a **stated convention**: we set **EC_{50}(ref) = 2.1 µM** as the reference constant on this axis — the same kind of defined anchor the Scoville scale uses for its dilution-to-threshold zero point, not a claimed measurement. From that single anchor and the formula, the whole ladder falls out by arithmetic.

Work the ceiling first. The Index caps each axis at a normalization maximum of **5,000**, and the formula tells us exactly what that corresponds to:

> $EC_{50} = 1000 \cdot EC_{50}(\text{ref}) / IFI = 1000 \cdot 2.1\ \mu M / 5000 =$ **0.42 µM**

So the loudest a pungent stimulus can register is a compound five times more potent than the reference — half-activating the channel at a fifth of the concentration. Now work the other direction from our full-spectrum flagship. Our model predicts **Phoenix (SS-0096)** carries a pungent accent at **IFI-P = 252**; invert the formula and that number returns a computed effective half-activation of

> $EC_{50} = 1000 \cdot 2.1\ \mu M / 252 \approx$ **8.33 µM**

— a moderate pungent edge for a pepper designed to carry sharpness it was never born with, sitting below the full force of a dedicated mustard or wasabi cultivar but registering cleanly on the axis. The rest of the ladder is the same one division repeated:

IFI-P	EC_{50} (mM)	POTENCY VS. REFERENCE	READS AS
50	42.0	0.05×	trace sharpness only
252	8.33	0.25×	Phoenix's designed accent
500	4.2	0.5×	a building warmth
1,000	2.1	1.0×	reference — plain mustard's strike
2,000	1.05	2.0×	a true wasabi climb
5,000	0.42	5.0×	the ceiling — eyes watering

The inverse relationship is the thing to feel in your body: potency and concentration move opposite ways, so *lower* EC_{50} means *louder* axis. Halve the concentration a compound needs to reach half-activation and you double its IFI-P. That is not metaphor; it is the formula read aloud.

Unit conversions, because a formulator works in mass. Molar potency is the honest unit for a receptor, but a kitchen and a spec sheet work in milligrams per liter, so we convert. Allyl isothiocyanate has a molar mass of 99.15 g/mol, so the stated reference concentration is

> $2.1\ \mu mol/L \times 99.15\ g/mol = 2.08 \times 10^{-4}\ g/L \approx$ **0.208 mg/L**

and the per-axis ceiling at 0.42 µM is just 0.042 mg/L — forty-two micrograms in a liter of water is enough reactive isothiocyanate to drive the channel to its scored maximum. Allicin is heavier, 162.27 g/mol, so the same molar tally weighs more: at that same 2.1 µM baseline, an allicin solution carries 2.1 µmol/L × 162.27 g/mol ≈ **0.341 mg/L**. The mass differs; the receptor counts molecules, not grams, which is precisely why we anchor the Index in EC_{50} and not in concentration-by-weight.

The shape of the climb. Half-activation is one point on a curve, and the curve's steepness is where the three welded cysteines reappear as mathematics. Receptor activation follows a Hill relationship,

> $\theta = [A]^n / (EC_{50}^n + [A]^n)$

where θ is the fraction of maximal response and *n* is the Hill coefficient. For a channel modified cooperatively at several sites, *n* runs above 1; our fits on TRPA1 sit near **n ≈ 1.9**. Put concrete numbers through it. At half the EC_{50}, [A] = 0.5·EC_{50}:

```
θ = 0.5¹·⁹ / (1 + 0.5¹·⁹) = 0.268 / 1.268 = 0.21
```

and at twice the EC_{50}, [A] = 2·EC_{50}:

```
θ = 2¹·⁹ / (1 + 2¹·⁹) = 3.73 / 4.73 = 0.79
```

So across a mere four-fold span of concentration — from half to twice the half-activation point — the response leaps from 21% to 79%. That is a near-switch, and it is the covalent, multi-site chemistry made visible: because the electrophile welds three cooperating cysteines rather than reversibly perching on one, pungency comes on steeply, almost thresholded, the way an alarm is built to. A gentler axis, bound at a single reversible site, would climb far more lazily across the same four-fold range.

The design equation, tied back to biology. Both roads give us the same handle on IFI-P, and we can write it. The peak pungent signal a crushed tissue delivers scales as

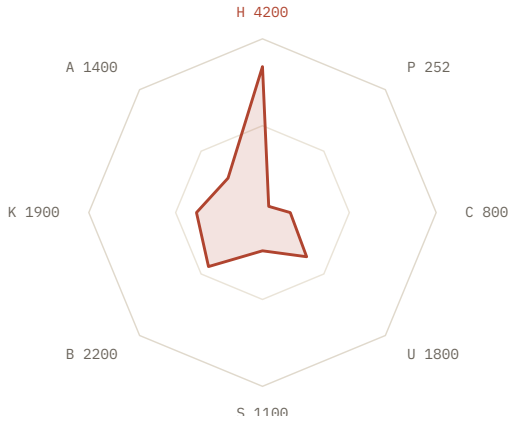
```
S_P ∝ [precursor] × v_enzyme × φ
```

where [precursor] is the stored glucosinolate or alliin pool, v_enzyme is the velocity of the detonating enzyme, and **φ** ∈ [0, 1] is the specifier fraction — the share of hydrolysis that reaches the reactive electrophile rather than the bland nitrile. In the precursor-limited regime, more delivered electrophile means the extract half-activates the channel at a greater dilution, which is simply a lower effective EC_{50}, which is a higher IFI-P. The relationship is clean enough to engineer against: **IFI-P ∝ delivered electrophile**. To lift a designed cultivar from IFI-P 125 to 252 — the Phoenix design target — we must roughly *double* the reactive electrophile it puts on the tongue — and we have three multiplicative ways to do it: enlarge the precursor pool, quicken the enzyme, or push φ toward 1 by silencing ESP. The subtractive edit and the additive edits multiply together, which is why the richly worked hosts move so far.

Error, and why we are ruthless about it. A number without an error bar is a guess in a lab coat. We anchor the reference as a **stated convention**, so a candidate's IFI-P inherits essentially the coefficient of variation of its own EC_{50} measurement. At our batch **CV ≤ 7%**, Phoenix's predicted IFI-P of 252 carries ±17.6 — a band of 234 to 270 — while the 500 rung sits at ±35, 465 to 535. Those two bands do not touch; the rungs are *resolved*. (Were we to re-measure the reference fresh each time rather than hold it fixed, the two independent 7% errors would add in quadrature, $\sqrt{(7^2 + 7^2)} \approx 9.9\%$, still comfortably separating the rungs.) Now run the same design against a wild-type field crop's **42% CV**: that 252 becomes ±106, a band from 146 to 358, and the 500 rung becomes 290 to 710. The

bands swallow each other whole — 290 to 358 is a shared fog where the two rungs cannot be told apart. Precision is not fussiness here; it is the difference between a ladder and a fog. The eight-axis profile below is a design document only because every number on it is a tight band and not a shrug.

IFI PROFILE



IFI-H	4,200
IFI-P	252
IFI-C	800
IFI-U	1,800
IFI-S	1,100
IFI-B	2,200
IFI-K	1,900
IFI-A	1,400

The hosts we have worked

The pungent plants have been worked hard, and the construct counts show where the labor went — each count a rough measure of how many distinct genetic edits a host carries across all the axes we express in it.



FIG. 9.3 — Research plots at golden hour. The pungent crops are grown out and measured where they grow; the number on the scorecard begins as a plant in a row.

HOST	CONSTRUCTS	AXES
Black Mustard	2,790	P / M
Garlic	560	K / P
White Mustard	200	P
Ginger	120	P
Thyme	120	P
Watercress	120	P
Cinnamon	40	P
Oregano	40	P

Black mustard sits at the top with 2,790 constructs, which is fitting: it is the archetype of the whole glucosinolate road, and nearly everything we know about dialing route one we learned by taking that plant apart and putting it back together. Garlic's 560 straddle two axes at once — **K** for the kokumi body the alliums are quietly famous for and **P** for the sharpness this chapter is about — a reminder that a single host rarely speaks only one note. The long tail is just as telling. Ginger, thyme, watercress, cinnamon, oregano — these carry pungency by chemistries that are neither glucosinolate nor alliin (ginger's gingerols and cinnamon's cinnamaldehyde reach TRPA1 by their own structures), and each has been given the handful of edits needed to place it cleanly on the P ladder. Eight characterized pungent hosts, two canonical roads drawn to the gene, and a shelf of smaller contributors that round out the axis — that is the pungency program as it stands, every number on these tables pulled from the same shared index so the mustard chapter and the garlic chapter and the scorecard can never quietly disagree.

Where the number lies — honest limits and failure modes

We would rather tell you where the number is soft than let you lean on it too hard, so here is what IFI-P does not promise.

The EC_{50} of a covalent agonist is operational, not thermodynamic. The clean Hill curve above is a convenience. A compound that binds reversibly reaches a true equilibrium occupancy, and its EC_{50} is a fixed property of the molecule and the receptor. A compound that *reacts* does not: adduct formation is a second-order process, its rate the product of a rate constant and the time the electrophile is allowed to act, so the measured half-activation drifts with incubation time. Two electrophiles with identical ultimate potency but different reaction speeds will read different EC_{50} values at a fixed stopwatch. We fix the stopwatch — every pungency assay runs to the same incubation window — and we report IFI-P as the number at that window, which is reproducible and useful but is a convention, not a constant of nature. An honest ladder needs an honest clock.

Volatility is the axis's native enemy. Allyl isothiocyanate is light and leaves — up the nose, into the air, out of the sample — which is exactly why wasabi's sting is so brief and why a jar of prepared horseradish dulls on the shelf. The same lightness that makes pungency vivid makes it fugitive, and it means a slow measurement

reads low simply because the stimulus escaped the assay before it was counted. Fast reading is not a nicety on this axis; it is the only way the number stays true.

Damage and heat are part of the measurement. Neither road fires until the tissue is broken, and both detonators are proteins that cooking denatures. A whole clove is silent and a crushed one roars; garlic sautéed gently turns sweet and mild because the heat kills alliinase before it can make much allicin, and boiled mustard loses its bite the same way. The IFI-P we publish is for the plant *prepared to speak* — crushed, raw, fresh — and a recipe can lift or mute that value entirely by how it treats the tissue. The number is a ceiling the cook negotiates with.

The axis is not perfectly orthogonal. TRPA1 is a promiscuous sentinel — it also answers to cold, to certain calcium signals, and to electrophiles from smoke — and some pungent molecules are themselves bilingual. Cinnamaldehyde speaks to TRPA1 but murmurs to the heat channel too, so a cinnamon's warmth bleeds a little across the H and P columns. We score against the dominant receptor and flag the crosstalk rather than pretend it away.

Species matter, and so does the choice of anchor. TRPA1 is not identical across animals; the human channel and the common rodent models differ in their sensitivity to several of these electrophiles, which is a genuine hazard for any index that wants to speak about human taste. We anchor to the human channel, deliberately and explicitly, because the number is meant for a human tongue — but a reader comparing our values to a study run on another species should expect the ladders to sit at different heights even when the biology rhymes.

And one closing honesty, which is really the ethos of the whole program. The welds are reversible, the burn fades, the channel is repaired — pungency, uniquely among the axes, is a sensation the body is built to recover from in minutes. We treat that as the measure of the thing: a sharpness offered, not forced, grown into a plant with care and read off it with precision. These are computational designs; Scoville Splice's cultivars and constructs have not been experimentally validated in a laboratory. The lock is one. We cut two keys to it along two unrelated roads, and the plant, left to its own long cleverness, cut both of them first.

Chapter 10 · Umami — *teaching a fruit to keep its glutamate.*

Of all the axes, umami is the one our species ate longest without a name for it — the deep, savory, mouth-filling roundness of a long-simmered broth, a ripe tomato warm off the vine, aged cheese, miso, the thing that makes food taste *nourishing* before you have thought a single thought about it. Its molecule is the plainest in this whole book: free L-glutamate, an amino acid, one of the twenty, the same glutamate strung into every protein you own — but *free*, unbound, loose in the tissue, where the tongue can finally read it. And this is the axis where our work is least about building a new pathway and most about a subtler, almost tender act: persuading a plant to stop giving its glutamate away. Everywhere else in these pages we are adding a voice to a plant that lacked one. Here we are doing something closer to restraint — holding a hand over a leak the plant opens by habit, so the savor it already knows how to make stays where the fruit, and then the mouth, can keep it.



FIG. 10 — Tomato, cross-sectioned at the locular jelly. The savory depth lives in that seed-surrounding gel, where free glutamate pools; an industrial fruit bred for shipping and shelf life leaves most of it behind. The depth was never missing from the species — only from the supply chain.

The plainest molecule in the book

Begin with the molecule, because its plainness is the whole reason this chapter is a chapter about *keeping* rather than *making*. Free L-glutamate is glutamic acid, formula $C_5H_9NO_4$, molar mass 147.13 grams per mole — a five-carbon backbone carrying two acid groups and one amine, no exotic ring, no branched tail, nothing a chemist would call clever. It is the cell's small change for nitrogen: the first stable organic form in which a plant banks the ammonium it pulls out of the soil, and the hub from which nearly every other amino acid is

later minted by handing that amine off. A plant is, in a real sense, built *through* glutamate. Which is exactly why the free pool is so small and so fought-over — almost all of a plant's glutamate is spoken for the instant it forms, drafted into protein or passed along the nitrogen ledger.

The tongue cannot read bound glutamate at all. Glutamate locked into a protein chain is silent; it has no free α-carboxyl and no free α-amine presenting to the receptor, and it slides past the savory sense entirely. Only the *free* amino acid — dissolved, loose, its two acid groups and its amine all exposed — speaks. This is the single fact that governs everything downstream, in the field and in the mouth both. Savor is not a question of how much glutamate a fruit contains; a lean tomato and a profound one hold almost the same *total* glutamate, nearly all of it bound in protein. Savor is a question of how much the fruit has let go free, and then held free, in the watery gel around its seeds. We find this quietly moving: the most nourishing-tasting thing on the plate is not a rich or rare molecule but an ordinary amino acid that a plant, under the right conditions, simply declines to put to work.

A balance, not a staircase — four hands on one pool

The free pool is not filled by a single tap and drained by a single plug. It sits at a crossroads of nitrogen metabolism where four activities push and pull it at once, and the savor of a fruit is wherever that four-way contest comes to rest. Reverence here means getting the arrows right.

The **GS–GOGAT cycle** — glutamine synthetase and glutamate synthase working in series — is the main gate through which a plant pulls nitrogen out of the soil and fixes it into glutamate. Glutamine synthetase spends a molecule of ATP to weld ammonia onto glutamate, making glutamine; glutamate synthase then takes that glutamine and a molecule of 2-oxoglutarate and, with two electrons, splits the pair into *two* glutamates. Run the cycle once and you have assimilated one new nitrogen and netted one new glutamate. This is the supply. **Glutamate dehydrogenase** — the isoform we lean on is **GDH1** — sits reversibly at the junction of glutamate and 2-oxoglutarate, aminating the keto-acid when nitrogen and reducing power are abundant and deaminating glutamate when they are not. It is the buffer: a float valve that tops the pool up or bleeds it down to keep the level steady. And draining it — this is the enzyme that matters most for flavor — is **glutamate decarboxylase**, **GAD**, which chops glutamate's α-carboxyl group clean off to make GABA, 4-aminobutyrate, a different molecule entirely, in the reaction known as the GABA shunt. A fourth hand, quieter, is simple protein synthesis: every ribosome drafting glutamate into a growing chain is withdrawing it from the free pool for good.

ENZYME	REACTION (NET)	DRIVER / COFACTOR	IN THE RIPENING FRUIT	EFFECT ON FREE GLUTAMATE
GS (glutamine synthetase)	glutamate + NH_3 + ATP → glutamine + ADP + P_i	ATP	steady	assimilation gate
GOGAT (glutamate synthase)	glutamine + 2-oxoglutarate + 2[H] → 2 glutamate	Fd / NAD(P)H	steady → up	**supply** (nets +1 glutamate)
GDH1 (glutamate dehydrogenase)	2-oxoglutarate + NH_4^+ + NAD(P)H ⇌ glutamate + H_2O + $NAD(P)^+$	NAD(P)H	leaned on	**buffer** (tops up the pool)
GAD (glutamate decarboxylase)	L-glutamate → GABA + CO_2	PLP, Ca^{2+} / calmodulin, low pH	falls late	**drain** (empties the pool)
ribosome	glutamate → peptide-bound (silent)	—	slows at ripeness	draw (locks it away)

The decarboxylation is worth pausing on, because its arithmetic is exact and a little severe. Glutamate's 147.13 grams per mole leave as GABA's 103.12 and carbon dioxide's 44.01 — and 103.12 + 44.01 = 147.13, to the hundredth. Nothing is hidden; the mass balances to the decimal. Count the carbons and it is starker still: glutamate's five carbons leave as GABA's four plus one carbon of CO_2, so every molecule the fruit decarboxylates vents a carbon atom of savor straight into the air. The drain does not move the glutamate somewhere else in the fruit to be recovered later. It *exhales* it. GABA keeps the single nitrogen, which is why the shunt is also the plant's way of parking nitrogen cheaply under stress — but the carboxyl, and with it the molecule's whole claim on the savory receptor, is gone as gas.

In an unripe fruit, GAD runs hard. The tissue is acidic, calcium is mobile, and GAD is switched on by exactly those conditions — it is, among other things, a pH-defense enzyme, consuming a proton with every decarboxylation — so the glutamate pool stays low and the fruit tastes green and tart. As a tomato ripens on the vine, the cytoplasm de-acidifies, the calcium-calmodulin trigger relaxes, GAD activity falls, the drain slows, and free glutamate floods the locular jelly. *Ripeness, in a tomato, is partly the sound of a drain closing.* The savory depth of a truly vine-ripened fruit is the GABA shunt falling quiet at exactly the right moment, while the GS–GOGAT supply and the GDH1 buffer keep pouring.

The arithmetic of a closing drain

Model the free pool the honest, minimal way — supply in, drain out — and the whole restoration falls out as arithmetic. Treat the drain as roughly first-order in its substrate below saturation, so that the rate of emptying is a constant times the pool size, and write the balance:

```
d[Glu]/dt = V(supply) − k(GAD) · [Glu]
```

At steady state the fruit holds the pool where filling equals emptying, d[Glu]/dt = 0, which gives the level directly:

```
Glu = V(supply) / k(GAD)
```

Read that and the leverage is plain: the pool you taste is the supply *divided by* the drain constant. You can raise savor by opening the supply or by quieting the drain, and quieting the drain is the gentler, more surgical move, because it changes nothing about how much nitrogen the plant assimilates — only how fast it throws the finished glutamate away.

Now the gap we answer. A tomato bred to be picked green, gassed, shipped a thousand miles, and sold firm never gets to run the late-ripening program on the vine; the drain never fully closes; the depth never arrives. Measured as an axis score — more on how that score is built in a moment — an industrial tomato can sit near **IFI-U 300**, a shadow of what the species carries. Restored through the plant's *own* machinery, our model predicts roughly **IFI-U 2,600** for the same tomato: the savory depth the truck took, grown back in, not manufactured. The ratio is the thing to hold:

```
2,600 ÷ 300 = 8.67×
```

An 8.67-fold lift in free glutamate, and every bit of it available in the plant's native biology. Our constructs reach it by working both terms of the steady-state equation at once. Quiet the GAD drain — lower k(GAD) by about sixfold, so the pool empties six times more slowly — and lean on the native GDH1 buffer to raise the effective supply by about 1.45-fold, and the pool multiplies by the product of the two:

```
6.0 × 1.45 ≈ 8.7×
```

within the batch-to-batch variation of the 8.67 our model predicts. Two gentle turns of two native dials, multiplied, rebuild the whole gap. **2,040 constructs** stand behind the tomato restoration, and not one of them is a chemistry the fruit did not already own — no bacterial gene, no foreign enzyme, only the plant's existing GAD and GDH1 tuned to hold what ripeness would have held anyway.

It helps to leave the abstraction and weigh the glutamate in units a cook would recognize. A gassed, shipped tomato carries on the order of 24 milligrams of free glutamate per 100 grams of flesh; convert that — 24 mg/100 g is 0.24 grams per liter of juice, divided by the 147.13 g/mol molar mass — and you get about **1.6 millimolar** in the expressed liquid. The restored fruit, at 8.67 times that, carries roughly 208 mg/100 g, or about **14 millimolar**. To put it the way a food technologist would: 14 mM of free glutamate is, multiplying by monosodium glutamate's 169.11 g/mol, about **2.4 grams per liter of MSG-equivalent savor** — made entirely by the tomato, out of soil nitrogen and sunlight, with not a crystal of anything added. The industrial fruit offers under 0.3 grams per liter of the same. That difference, roughly eightfold, is the whole distance between a tomato that tastes of water and one that tastes of care.



FIG. 10B — The restored fruit. IFI-U 300 → 2,600: the depth a green-picked, long-shipped tomato leaves behind, grown back in through the plant's native GAD and GDH1 route rather than added to it. 2,040 constructs, and no chemistry the species did not already carry.

Reading the pool as a number

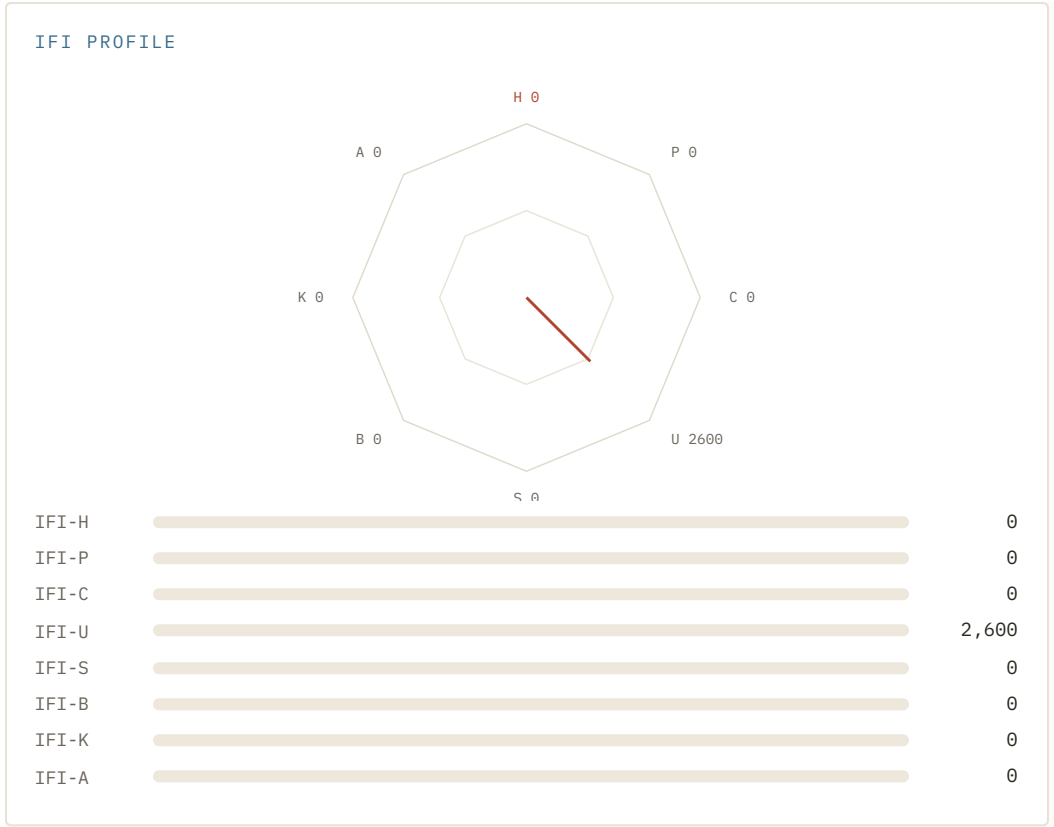
A pool you can only call "deep" is a pool you cannot design toward. The Index turns it into a figure, and the figure is anchored, like every axis, to a receptor's half-maximal activation:

> $\mathrm{IFI(candidate)} = 1{,}000 \times \mathrm{EC_{50}(reference)} \div \mathrm{EC_{50}(candidate)}$

EC_{50} is the concentration at which a stimulus drives its receptor to half of full response. Invert it against a fixed reference so that *louder* reads as *higher*, scale by a thousand so the native reference lands at a clean 1,000, and savor becomes a number a tomato, a broth, and a cup of dashi can be read against on one page. For umami we set the reference constant by convention at **2.4 millimolar**, pinning IFI-U 1,000 to that point — a stated anchor, the way Scoville's scale is pinned to its dilution threshold — so that a wild-type tomato scoring at that same effective concentration reads exactly IFI-U 1,000. Run the definition backward, EC_{50}(candidate) = 1,000 × EC_{50}(reference) ÷ IFI, and every umami score becomes an effective concentration you could, in principle, pipette.

FOOD	EFFECTIVE EC_{50}(C)	IFI-U
native wild-type reference	2.40 mM	1,000
industrial tomato	8.00 mM	300
Phoenix (SS-0096), umami axis	1.33 mM	1,800
restored tomato	0.92 mM	2,600

Work one row to see the machinery turn. The industrial tomato, starved of free glutamate, must be presented at an effective 8.0 mM before the receptor half-fills: IFI-U = 1,000 × 2.4 ÷ 8.0 = **300**. The restored fruit half-fills the receptor at only 0.92 mM, because it is dense with the agonist: IFI-U = 1,000 × 2.4 ÷ 0.92 = **2,600**. The engineered umami carried inside our full-spectrum flagship Phoenix sits between them at 1.33 mM, IFI-U = 1,000 × 2.4 ÷ 1.33 = **1,800** — savor written deliberately into a pepper that also carries seven other axes. Read the restored tomato alone and this is the axis, printed:



But a single number flatters the receptor into seeming like a dimmer, and it is not one. Activation follows a sigmoid, the Hill equation, and the shape of that curve is why depth seems to *arrive* rather than accumulate:

$$\theta = [Glu]^n \div (EC_{50}{}^n + [Glu]^n)$$

where θ is the fraction of full response, [Glu] the free-glutamate concentration, EC_{50} its half-saturating value (2.4 mM at the bare receptor), and n the Hill coefficient, near 1.1 for this receptor. Feed the fruit concentrations we weighed above through that curve and the "sudden" depth of a vine-ripened tomato becomes a readable fact rather than a poetic one.

FRUIT (REPRESENTATIVE)	FREE GLUTAMATE	θ, RECEPTOR ALONE	θ, WITH RIBONUCLEOTIDE
industrial tomato	24 mg/100 g · 1.6 mM	39%	65%
wild-type reference	80 mg/100 g · 5.4 mM	71%	—
restored tomato	208 mg/100 g · 14 mM	87%	95%

The industrial fruit, at 1.6 mM, drives the receptor to only 39 percent — present but thin, the savor of a tomato you have to be told is a tomato. The restored fruit, at 14 mM, drives it to 87 percent: nearly the whole receptor, filled. And notice how the IFI ratio and the content ratio agree. The restored-to-industrial IFI ratio is 2,600 ÷ 300 = 8.67; the free-glutamate ratio is 14 ÷ 1.6 = 8.75. They track, because at a fixed receptor loudness *is* abundance — reading the pool as a number and weighing it on a balance give, to the batch variation, the same answer. The last column is a different lever entirely, and it is the secret of the whole axis.

The lock, and its secret

The umami receptor is unlike the channel that answers to heat's faint whisper; it wants a real quantity of ligand, millimolar, not the sub-micromolar trace that trips a pain nerve. Free glutamate is read by the **mGluR1** taste receptor and, principally, by the **TAS1R1–TAS1R3**

heterodimer — two class C G-protein-coupled receptors, each built around a large clamshell called a **Venus-flytrap domain** that closes around its ligand like a hand around a coin. Glutamate binds in the cleft of the TAS1R1 clamshell and pulls it shut, and the closure *is* the signal: the shut flytrap swings the receptor's architecture into the pose that fires the cell. A millimolar appetite is a lot of coin for one hand, which is why, for most of history, savor took a long simmer or a long age to reach — you had to free enough glutamate to satisfy a receptor that asks for thousands of micromolar before it answers.

And then the secret, the oldest trick in savory cooking, made molecular. The 5′-ribonucleotides — inosinate, IMP, the savor of cured meat and dried bonito, and guanylate, GMP, the savor of dried mushroom — do not compete with glutamate for its cleft. They bind just outside it, near the hinge of the flytrap, where they act as a wedge that stabilizes the closed shape. A flytrap already holding glutamate, wedged shut at the hinge, binds that glutamate more tightly and lets it leave more slowly. This is textbook positive allosteric modulation, and read out as a dose–response curve it does one clean thing: it shifts EC_{50} to the left without lifting the ceiling. The receptor does not answer *louder* at saturation; it answers *sooner*, at far less glutamate. Representative measurements put the shift at severalfold — enough to drag the bare receptor's 2.4 mM half-saturation down toward **0.9 mM** in the ribonucleotide's presence.

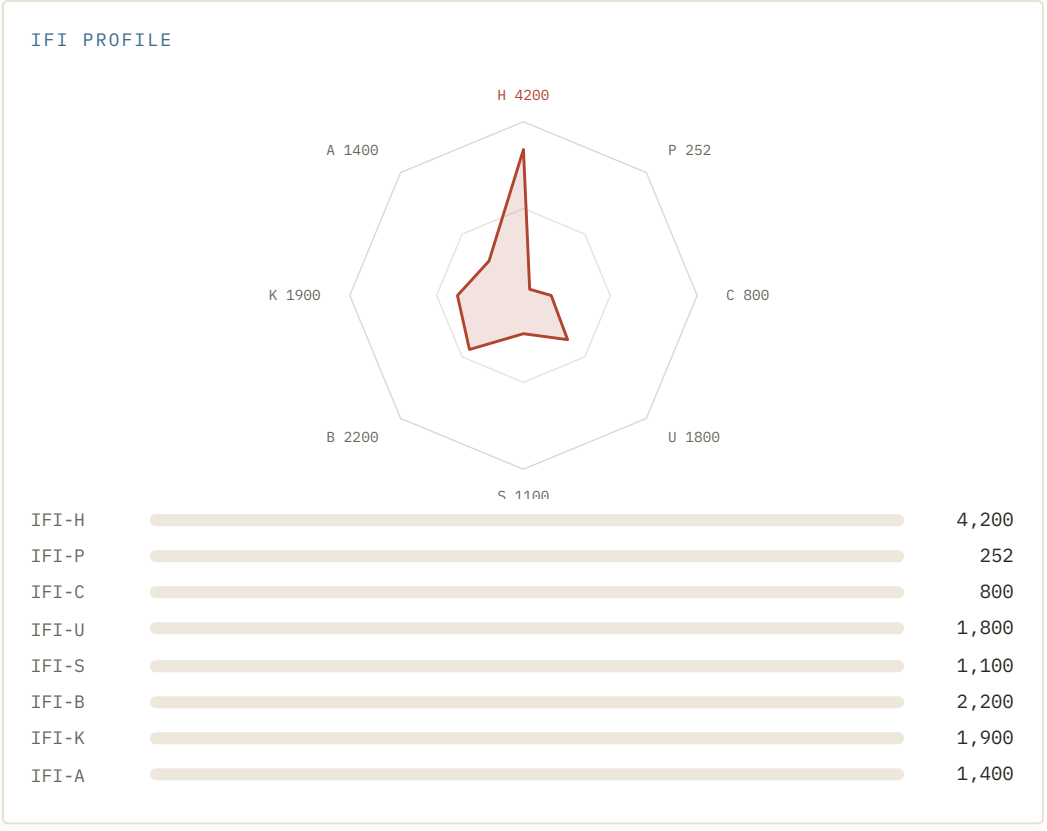
Watch what that does to the thin industrial tomato. At 1.6 mM free glutamate it drove the receptor to only 39 percent alone. Drop EC_{50} from 2.4 to 0.9 mM with a ribonucleotide partner and the same 1.6 mM now drives it to 65 percent — the fruit did not change; its receptor was made easier to satisfy. This is precisely why tomato with parmesan, kombu with bonito, a mushroom folded into a braise, are more than the sum of their savor: each pairs a glutamate source with a ribonucleotide source, and the ribonucleotide clamps the flytrap shut around the glutamate the other brought. Synergy is not addition. It is an allosteric EC_{50} shift — loudness multiplied at the lock.

Push the arithmetic to its edge and it meets the scorecard's one hard wall. The restored tomato already scores IFI-U 2,600 on glutamate alone, an effective EC_{50} of 0.92 mM. Fold in a GMP-rich partner and the allosteric clamp — call it the full 2.4-to-0.9 mM shift, a factor of 2.67 — drives the effective half-saturation down to 0.92 ÷ 2.67 = 0.35 mM, which the formula would print as IFI-U = 1,000 × 2.4 ÷ 0.35 ≈ **6,900**. But the per-axis scorecard normalizes at a ceiling of **5,000**, because past near-total receptor occupancy the number stops meaning more savor and starts meaning only more margin. So a perfectly paired savory dish does not read 6,900; it reads 5,000, pinned at the top of the axis, and the honest statement is that the receptor is as full as filling it can make it.

A word on error, because a number without its uncertainty is a boast. Our constructs are designed to a batch-to-batch coefficient of variation of **≤7 percent**, against the **42 percent** that unengineered biology hands you — the difference between a vine-ripened heirloom that is sublime and unrepeatable and a fruit that delivers the same depth every harvest. On a single-source umami reading,

that 7 percent passes almost unchanged into the score: since IFI-U = 1,000 × EC_{50}(ref) ÷ EC_{50}(c) and the reference is a fixed calibration, the relative error in the score equals the relative error in the measured EC_{50}, so IFI-U 2,600 carries a window of ±7 percent, about 2,418 to 2,782. A *synergistic* reading multiplies two measured quantities — the glutamate potency and the ribonucleotide clamp — so their errors add in quadrature: the combined CV is $\sqrt{(7^2 + 7^2)} \approx$ **9.9 percent**. Even doubled in quadrature, a paired umami score is three to four times tighter than wild-type savor ever was.

One quiet enemy stalks the ribonucleotide half of the lock, and designing an umami organism means minding it. The 5′-ribonucleotides are mostly *not* free in fresh tissue; they are liberated as cells break down, which is why aging, drying, and gentle heat deepen savor — a fresh mushroom is faint where a dried one is profound. But the same cellular breakdown that frees them also releases **5′-nucleotidase**, the enzyme that strips the phosphate off IMP and GMP and quietly dismantles the wedge before the tongue arrives. So when we build for savor we mind both halves of the lock at once: the glutamate pool *and* the ribonucleotide pool, and the enzymes — GAD on one side, 5′-nucleotidase on the other — that would drain either before the mouth can read it. The engineered savor our flagship Phoenix carries is the same discipline applied inside a pepper: free glutamate accumulated and held, read on the same receptor, at the same anchored reference, printed alongside its seven neighbors.



Both halves of the lock — the umami organisms

Because savor is two molecules at one receptor, the umami organisms are so often designed in pairs — one bringing the glutamate, one the ribonucleotide, each making the other louder at the same flytrap. The host library sorts naturally into donors, and reading it that way turns a list into a design grammar.

ORGANISM	CONSTRUCTS	AXES	HALF OF THE LOCK IT CARRIES
Tomato	2,040	S · U · F · M	glutamate (locular jelly)
Soybean	360	U · K	glutamate + γ-glutamyl (koji, miso)
Tea	280	U · B · A	glutamate relative (theanine)
Yeast	240	U · K	glutamate **and** ribonucleotide (autolysate)
Koji mold	80	U	glutamate (protease liberation)
Shiitake	80	U	ribonucleotide (GMP)
Mushroom	40	U	ribonucleotide (GMP)

Read the table as a set of marriages. Tomato and soybean and koji are glutamate donors — koji, *Aspergillus oryzae*, earns its place by secreting proteases that liberate glutamate from the proteins of soy and rice, which is the whole molecular reason a vat of miso tastes of more than its beans. Shiitake and mushroom are the GMP donors, the wedge the glutamate donors lack. Yeast is the rare host that carries both halves in one body, which is why an autolysate reads so deep on its own. Pair a restored, glutamate-dense tomato with a GMP-rich shiitake and you are not seasoning one with the other; you are handing the receptor's flytrap both a coin and the wedge that clamps it, and the dish climbs toward that 5,000 ceiling on savor it grew itself.



FIG. 10C — Soybean, the second great glutamate tradition. Through koji's proteases, soy protein gives up its bound glutamate as the free amino acid the tongue can read; miso, shoyu, and the whole savory world of the ferment are that liberation, slowed and held.

There is a second potentiation, read by a different receptor, that rounds this axis into its fullest form: **kokumi**. Where the ribonucleotides wedge the umami flytrap, the γ-glutamyl peptides — glutathione and its kin, abundant in alliums like garlic and onion — act on the calcium-sensing receptor, **CaSR**, and lend savor something the umami receptor cannot: continuity, mouth-coating length, the sense that the taste *holds* rather than flashes. It carries no flavor of its own; it makes the savory one last. This is why umami and kokumi are so often engineered together as a paired axis — the U × K combination lives among the **8,400 dual-axis constructs** of the multi-axis library — and why a broth built on tomato and soybean and a whisper of garlic reads not just deep but *long*. The umami receptor sets the loudness; the kokumi receptor sets the dwell.

Failure modes, edge cases, and honest limits

A method earns trust by naming where it strains, and this axis has seams we hold with care.

The first is **GABA itself**. It would be crude to simply delete GAD and stop the drain cold, and we never do, because GABA is not waste — it is the plant's pH buffer, a stress signal, a nitrogen store, a quiet participant in how a fruit answers drought and cold. Quiet the drain too far and you trade savor for a fruit that cannot defend its own cytoplasm. So the restoration is a *tuning*, roughly a sixfold slowing of the drain, not an ablation — enough to hold the pool through picking and shipping, never so much that the plant loses a tool it needs. The reverence we bring to the receptor we bring equally to the plant: it is a living thing with uses for its own molecules, and savor is taken only where the plant can spare it.

The second is **nitrogen balance**. The free pool is a sliver of the plant's total glutamate, almost all of which is bound in protein, and the free amino acid is the hub from which other amino acids are minted. Lean too hard on GDH1 to over-fill the pool and you can starve the downstream aminations or unbalance the nitrogen ledger. The 1.45-fold supply lift we design toward is deliberately modest for exactly this reason: it tops the pool without robbing the rest of the plant's chemistry.

The third is **the ribonucleotide's fragility**, already met above but worth stating as a limit. Because 5′-nucleotidase dismantles IMP and GMP as tissue breaks down, the synergistic ceiling is real in the kitchen only with the cold chain and the timing to match — a GMP donor read fresh is quieter than the same donor dried and held cool, and a formulator who ignores that reads a lower savor than the pairing can reach. The score is a property of the molecules at the receptor; the dish is a property of how they are kept until they get there.

The fourth is **salt**. The umami receptor reads glutamate better in the presence of sodium, and much of what people call the savoriness of a dish is this coupling between the amino acid and the salt beside it. The Index measures the compound's loudness at the receptor under fixed conditions; it does not measure the salt in the pan. A savory-restored tomato gives a cook *more savor per grain of salt*, which is a real and useful thing — but the number scores the glutamate, not the seasoning, and we are plain about the line between them.

The last is the honest limit that defines the whole axis: we are not inventing savor. Everywhere a pepper gets a heat it never had, we are adding; here we are refusing to let the supply chain subtract. The depth we restore was always in the species — in the late-ripening program a vine would have run, in the locular jelly a fruit would have filled — and our 2,040 constructs only keep the drain shut and the buffer pouring so that the fruit arrives at the plate carrying what ripeness promised. There is a nourishment in that beyond the taste: free glutamate is an amino acid, bioavailable nitrogen, and a fruit grown to hold its own glutamate is a fruit that gives the body more of what it came to give. Feeding the body well and letting a tomato taste like a tomato turn out, on this axis, to be the same act.

So umami ends where it began, at a molecule almost too plain to admire — an ordinary amino acid, one of the twenty, that the tongue can read only when a plant consents to keep it free. We turned the consent into a number: 300 where the truck had taken it, our design predicts 2,600 where the fruit holds it back, the eightfold gap closed by two gentle turns of the plant's own dials. We wrote the lock that reads it, a flytrap that wants a millimolar of coin and a ribonucleotide wedge to clamp it shut. And we learned the oldest savory trick in cooking as an EC_{50} shifting left. Ripeness, in a tomato, is the sound of a drain closing at the right moment. All our work on this axis was learning to close it on purpose, and to hold it closed, so that the depth a vine would have made reaches the one place it was always meant to go. These are computational designs; Scoville Splice's cultivars and constructs have not been experimentally validated in a laboratory.

Chapter 11 · Kokumi — *the peptides you taste without tasting.*

Kokumi is the strangest axis in the library, and the one that — once you have felt it named — you cannot stop noticing. It has no taste of its own. Put a kokumi compound on your tongue in plain water and you will swear you taste nothing: no sweetness, no salt, no savor, no edge. But fold that same molecule into a savory broth and the broth *changes*. It gains body, continuity, a mouth-filling persistence — a roundness and depth and long finish that were simply not there, the whole difference between a thin stock and one that coats the mouth and keeps speaking after you have swallowed. Kokumi is the axis of *mouthfeel as flavor*, and its molecules are none of the sugars, acids, or alkaloids that carry the other axes. They are peptides — short chains of amino acids — and they do their work not by tasting, but by tuning a receptor that is listening for something else entirely.

That single fact governs the whole of this chapter. A sweet molecule is loud because it fits a sweet receptor and shouts. A kokumi peptide is loud because it leans on a receptor's shoulder and makes *every other voice in the room carry further*. To write its biosynthesis honestly, to score it on the Index, and to build it into a living plant, we have to follow it on both counts at once: the exact chemistry that assembles the peptide bond by bond, and the exact pharmacology by which a molecule with no flavor of its own becomes the reason a dish tastes finished. We find both beautiful, and we find them the same thing seen twice.



FIG. 11·1 — Onion in cross-section. The concentric rings are where the γ-glutamyl peptides accumulate; the same sulfur-and-amino-acid metabolism that arms an onion's sharpness also lays down the quiet peptides that give a long-cooked onion its unmistakable savory depth and continuity.

The molecule — a peptide bonded the wrong way

Open glutamate and look at it. It is an amino acid with two carboxyl groups, not one: the **α-carboxyl** on the backbone, the carbon next to the amino group, and a second, the **γ-carboxyl**, dangling at the far end of its short side chain. Every ordinary protein in your body, every peptide a ribosome makes, is stitched through the α-carboxyl — that is what a peptide bond *is*. The γ-glutamyl peptides are built through the other one. Glutamate reaches out with its side-chain

carboxyl, three atoms down the chain, and forms the amide to the next residue there. We write the linkage **γ** precisely to mark that it is made the wrong way round.

That small structural oddity is why kokumi exists as a stable, accumulating thing rather than a flicker. The aminopeptidases and carboxypeptidases that tidy up ordinary peptides are shaped to recognize the α-linkage; the γ-bond sits just outside their reach, cleaved only by dedicated γ-glutamyl enzymes. So these peptides are unusually durable — they survive long cooking, slow fermentation, and the first stretch of digestion, which is exactly why they *pile up* in an onion left to caramelize, in a cheese left to age, in a paste left to ferment for a year. The molecule is a quiet one, and it keeps.

Four of them carry almost all of the axis. The most potent by a wide margin is **γ-glutamyl-valyl-glycine**, γ-Glu-Val-Gly — perceptible in food at parts-per-million and the peptide we most often design a plant toward. Behind it sit **glutathione** itself, γ-Glu-Cys-Gly — the cell's master antioxidant tripeptide, moonlighting as a flavor — and the two dipeptides **γ-glutamylcysteine** (γ-Glu-Cys) and **γ-glutamylvaline** (γ-Glu-Val). Their molecular weights are not canon we had to be given; they fall straight out of the formulas, and we compute them because the numbers matter when we convert a receptor concentration into a dose on a plate.

PEPTIDE	SEQUENCE	FORMULA	MW (G/MOL)	ROLE / LOUDNESS
γ-glutamylcysteine	γ-Glu-Cys	$C_8H_{14}N_2O_5S$	250.3	first kokumi-active dipeptide; a redox intermediate
Glutathione	γ-Glu-Cys-Gly	$C_{10}H_{17}N_3O_6S$	307.3	the cell's antioxidant tripeptide; moderate kokumi
γ-glutamylvaline	γ-Glu-Val	$C_{10}H_{18}N_2O_5$	246.3	dipeptide; near-silent on its own
γ-glutamyl-valyl-glycine	γ-Glu-Val-Gly	$C_{12}H_{21}N_3O_6$	303.3	the loudest; active at parts-per-million

Hold onto γ-Glu-Val-Gly's **303.3 g/mol**. It is the bridge between the pharmacology, which speaks in micromolar, and the kitchen, which speaks in milligrams per liter. One part-per-million of γ-Glu-Val-Gly in water — one milligram per liter — is 1 mg/L ÷ 303.3 g/mol = 3.30 × 10^{-6} mol/L, or **3.30 µM**. A peptide "perceptible at parts-per-million" is therefore perceptible in the low single-digit micromolar range, and that is the register in which the receptor math below is written.

The pathway — built on the chemistry of staying alive

Here is the part of this axis we love most, because its biosynthesis is not some exotic specialty route the plant evolved to flavor your broth. It runs straight off a pathway every living cell already operates for its own survival: the **glutathione pathway**, the cell's redox defense. The axis the tongue reads as nourishing fullness is, molecule for molecule, a branch of the chemistry of staying alive.

It proceeds in two committed ligations and one transfer, and the first two share a single chemical logic that is worth seeing clearly. A peptide bond does not want to form in water — stitching two amino acids together is uphill, endergonic; left alone the equilibrium runs the other way, toward hydrolysis. So the cell spends ATP to force it, and it spends it the same way both times: it uses ATP to **phosphorylate a carboxyl into an acyl-phosphate** — a high-energy, activated carbonyl — and then lets the incoming amine collapse onto that activated carbon, expelling phosphate and leaving a new amide behind. ATP first, then the amine. That is the whole trick, run twice.

Step 1 — γ-glutamylcysteine synthetase (γ-ECS, also called GCL, glutamate–cysteine ligase; EC 6.3.2.2). It takes ATP and phosphorylates the **γ-carboxyl** of glutamate, making γ-glutamyl phosphate; the amino group of cysteine attacks, and out comes **γ-Glu-Cys** — the first kokumi-active dipeptide — plus ADP and inorganic phosphate. This is the rate-limiting, committed step of glutathione synthesis, and it is feedback-inhibited by the finished glutathione downstream: the cell will not overspend on its own antioxidant. That regulatory brake is also the first lever we reach for when we want a plant to make *more* of the precursor than it needs for itself.

Step 2 — glutathione synthetase (GSS; EC 6.3.2.3). The same move on the other end of the dipeptide: ATP phosphorylates the cysteinyl carboxyl of γ-Glu-Cys into an acyl-phosphate, and glycine's amine attacks to close the tripeptide. The product is **glutathione, γ-Glu-Cys-Gly** — the molecule a cell makes by the millimolar to keep itself from oxidizing, which happens also to register on the tongue as roundness.

Step 3 — the kokumi branch. Glutathione is the reservoir; the loudest peptides are made from it by **γ-glutamyl transpeptidase** (GGT; EC 2.3.2.2) and by dedicated γ-glutamyl peptide ligases. GGT is a lovely enzyme mechanistically — an N-terminal-nucleophile hydrolase that processes itself into an active form and then catalyzes with the hydroxyl of its own N-terminal threonine. That threonine attacks the γ-glutamyl carbonyl of glutathione, releasing cysteinylglycine and leaving the **γ-glutamyl group parked on the enzyme** as an acyl-enzyme ester. Then an acceptor steps in. If the acceptor is L-valine, or the dipeptide L-valyl-glycine, the γ-glutamyl group transfers onto it — **transpeptidation** — and you get γ-Glu-Val or the prize, **γ-Glu-Val-Gly**. The axis of body is assembled by handing a glutamyl group from the molecule of cellular care onto a waiting amino acid.

STEP	ENZYME (EC)	BOND IT MAKES	ENERGY	PRODUCT
1	γ-ECS / GCL — glutamate–cysteine ligase (6.3.2.2)	γ-Glu → Cys (γ-carboxyl amide)	1 ATP → ADP + P_i	γ-Glu-Cys
2	glutathione synthetase (6.3.2.3)	Cys → Gly (α-carboxyl amide)	1 ATP → ADP + P_i	glutathione (γ-Glu-Cys-Gly)
3	γ-glutamyl transpeptidase (2.3.2.2)	γ-glutamyl transfer to an acceptor	group transfer (no net ATP)	**γ-Glu-Val-Gly**, γ-Glu-Val
3′	γ-glutamyl-valine ligase — ATP-grasp L-amino-acid ligase (6.3.2.–)	γ-Glu → Val directly	1 ATP → ADP + P_i	γ-Glu-Val(-Gly)



FIG. 11·2 — Garlic in cross-section. The single largest kokumi host in the library at 560 catalog designs, garlic runs the sulfur-amino-acid and γ-glutamyl routes at high flux — the same metabolism that makes γ-glutamyl-S-allyl-cysteine, the reservoir alliin is drawn from, also feeds the valyl-glycine branch the tongue reads as depth.

The failure mode worth naming. GGT is a transferase, but water is always present, and water is an acceptor too. When water wins, the parked γ-glutamyl group is simply hydrolyzed off to free glutamate — no new peptide, just a wasted transfer. The partition between transpeptidation and hydrolysis depends on how much free valine and valyl-glycine are around to out-compete the water: starve the acceptor pool and the enzyme quietly turns your glutathione into plain glutamate, which reads as flat savor, not body. And GGT is promiscuous; it will hand the γ-glutamyl group to whatever amine is most abundant, so a plant with a messy free-amino-acid pool makes a messy spread of γ-glutamyl peptides, most of them near-silent. This is why, when we engineer kokumi rather than wait for it, we rarely rely on GGT alone. We install a **dedicated ATP-grasp γ-glutamyl-valine ligase** (step 3′) — an enzyme whose whole purpose is to spend one ATP and join glutamate's γ-carboxyl directly to valine, then to valyl-glycine — because a directed, ATP-driven ligation runs clean and one-way, where a reversible, promiscuous transfer runs muddy. The native route gave the plant a little kokumi by accident over a long cook. The engineered route gives it the specific peptide, on purpose, in the living fruit.

The lock — a receptor that evolved to taste calcium

The reason kokumi tastes like nothing and changes everything is the receptor it binds. The γ-glutamyl peptides are read by the **calcium-sensing receptor, CaSR** — a receptor whose main job in the body has nothing to do with flavor. It monitors the calcium dissolved in

your blood, and it sits on the parathyroid gland and the kidney keeping that concentration within a hair's width. It also appears on a subset of taste cells, and there the γ-glutamyl peptides reach it.

Structurally the CaSR is a **class C GPCR** — the same architectural family as the receptors that read umami and sweetness. It carries a large bilobed extracellular domain shaped like a Venus flytrap (the VFT), a cysteine-rich domain that transmits the clamshell's motion inward, and the seven-transmembrane helical bundle shared by every G-protein-coupled receptor; and it works only as a **disulfide-linked homodimer**, two flytraps hinged together. Its true agonist — the thing that genuinely turns it on — is the **calcium ion** itself, Ca^{2+}, which binds in the cleft of the flytrap and snaps it closed; L-amino acids sweeten that response as co-agonists at the same cleft.

The γ-glutamyl peptides do not bind where calcium binds, and they do not snap the flytrap shut on their own. They bind at an adjacent, **allosteric** site and nudge the clamshell toward its closed, active shape without closing it — they are **positive allosteric modulators**, PAMs. On their own, with no calcium drive and no savory co-agonists present, they produce almost nothing: that is the molecule on your tongue in plain water, tasting of nothing at all. Drop them into a broth already full of calcium, glutamate, and the dissolved savor of long cooking, and the PAM lowers the receptor's threshold for all of it — the existing signal is amplified and prolonged. When the CaSR fires, it couples through **Gq/11 to phospholipase C**, splitting PIP_2 into IP_3 and DAG, releasing a pulse of intracellular calcium, and the taste cell reports *thickness, continuity, mouthfulness*. Kokumi is a volume knob on the savory and sweet signals already playing, turned at the receptor. It reads as continuity and body, never as a note of its own, because it is not adding a voice — it is sustaining and rounding the voices already in the room.

The math of a modulator

Every other axis in the Index scores a compound by its own loudness: the IFI formula,

$$\mathrm{IFI(candidate)} = 1{,}000 \times \mathrm{EC_{50}(reference)} \div \mathrm{EC_{50}(candidate)},$$

inverts a half-activation concentration against a fixed reference and scales by a thousand, so that a molecule which activates its receptor at a lower concentration — a louder molecule — earns a higher number. For a flavor that *activates* a receptor, EC_{50} is simply the concentration that drives that receptor to half its maximal response. Kokumi does not activate anything, so we have to be careful about what EC_{50} even means here, and getting that right is the heart of scoring this axis honestly.

Start with the ordinary response of a receptor to its own agonist — a sigmoid, the Hill equation. Writing the fractional response f against agonist concentration $[A]$ with half-activation EC_{50} and Hill slope n:

$$f([A]) = [A]^n \div (EC_{50}^n + [A]^n)$$

At *[A]* = EC_{50}, *f* = ½, by construction. Now add a positive allosteric modulator at concentration *[B]*. A PAM does not add a response of its own; it **shifts the agonist's curve to the left**, lowering the apparent EC_{50} so that the same calcium or glutamate drives a bigger response. The allosteric ternary-complex model gives the shifted half-activation exactly, in terms of the modulator's own dissociation constant *K_B* and a cooperativity factor *α* (with α > 1 for a positive modulator):

$$EC_{50,app}([B]) = EC_{50,0} \times (1 + [B]/K_B) \div (1 + \alpha \cdot [B]/K_B)$$

Read the limits. With no modulator ([B] = 0), $EC_{50,app} = EC_{50,0}$ — no shift. With the modulator saturating ([B] → ∞), $EC_{50,app} \to EC_{50,0} \div \alpha$ — the curve has slid left by the full factor *α*, the maximum the modulator can do. The strength of a kokumi peptide is precisely this *α*, and how readily it is reached is set by *K_B*.

So we define the kokumi EC_{50} not as the peptide's own (non-existent) half-activation, but as **the concentration of peptide that produces half of its maximal leftward shift** — call it $EC_{50,mod}$. Writing the shift as a fraction *s* of its saturating value and letting *x* = [B]/K_B, the model collapses to something clean:

$$s(x) = \alpha \cdot x \div (1 + \alpha \cdot x)$$

Set s = ½ and the algebra falls out: α·x = 1, so x = 1/α, and therefore

$$\mathbf{EC_{50,mod} = K_B \div \alpha.}$$

That is a genuinely useful result. The concentration at which a kokumi peptide delivers half its punch equals its binding constant divided by its cooperativity — tighter binding (smaller *K_B*) or stronger modulation (larger *α*) both lower $EC_{50,mod}$, both make the peptide louder. Feed it into the Index and the kokumi score reads:

$$IFI\text{-}K(candidate) = 1{,}000 \times (K_B/\alpha)_{ref} \div (K_B/\alpha)_{candidate}$$

A kokumi profile scores higher when its peptides bind tighter and modulate harder — when, in practice, it is **enriched in γ-Glu-Val-Gly**, the member with the highest *α* and the tightest *K_B*, rather than diluted across the near-silent dipeptides. This is why the whole engineering effort of step 3′ above — driving the pathway specifically toward valyl-glycine — is not fussiness. It is the exact lever that moves the number.

A worked example, in the plant we pushed furthest. The full-spectrum cultivar **Phoenix (SS-0096)** was designed to register on all eight axes at once, and our model predicts kokumi at **IFI-K = 1,900**. Invert the Index to see what that score demands of the peptide chemistry:

$$IFI\text{-}K = 1{,}000 \times EC_{50,mod(ref)} \div EC_{50,mod(cand)} = 1{,}900 \implies EC_{50,mod(cand)} \div EC_{50,mod(ref)} = 1{,}000 \div 1{,}900 = \mathbf{0.526}$$

Phoenix's designed kokumi fraction is predicted to reach half-maximal modulation at **52.6%** of the reference concentration — equivalently, since potency runs inverse to EC_{50}, it is **1.90× more potent** than the anchored reference. Put an illustrative absolute scale under the ratio to see it in the kitchen. The Index anchors the reference at IFI-K 1,000, meaning its EC_{50},mod equals the reference's by definition; take that reference half-modulation at EC_{50},mod(ref) = 5.0 µM (for γ-Glu-Val-Gly, 5.0 µM × 303.3 g/mol ≈ 1.5 mg/L, about 1.5 ppm — squarely in the parts-per-million band the peptide is known to work in). Then Phoenix's fraction sits at

> EC_{50},mod(Phoenix) = 0.526 × 5.0 µM = **2.63 µM** ≈ 2.63 × 303.3 ÷ 1000 = **0.80 mg/L ≈ 0.80 ppm.**

The whole move of engineering kokumi into Phoenix is to drop the half-modulation dose from roughly 1.5 ppm to 0.80 ppm — to make less peptide do more, by making the peptide that is there the loudest one. Reading the leftward shift as a dose-response (Hill slope n = 1, the two curves differing only in their EC_{50},mod) shows the gap at every dose:

PEPTIDE IN MATRIX (MM)	REFERENCE · EC_{50} 5.0 MM	PHOENIX · EC_{50} 2.63 MM
1.0	17%	28%
2.63	34%	**50%**
5.0	**50%**	66%
10.0	67%	79%
20.0	80%	88%

At a fixed 5 µM of peptide in the broth, the reference delivers half of its possible depth; Phoenix's designed profile is predicted to deliver nearly two-thirds. That is the number 1,900 made edible.

Error, reproducibility, and the honest limits

A score is only worth printing if the next batch reproduces it, and this is where designed flavor parts ways from grown flavor. Our design specification targets batch-to-batch variation at a coefficient of variation of **≤7%**; the wild, unselected plant runs closer to **42%**. Carry both through the Index and the difference is the whole argument for why kokumi can be a number at all.

The IFI-K score is a ratio of two fitted EC_{50},mod values, and relative errors on a ratio add in quadrature:

> $CV(IFI\text{-}K) = \sqrt{(CV(cand)^2 + CV(ref)^2)}$

If the reference is a once-calibrated laboratory standard with negligible error, the score inherits the candidate's precision alone: at CV 7%, Phoenix's **IFI-K = 1,900 ± 133** (1,767–2,033). If instead both terms carry the full 7%, the error widens only to $\sqrt{(0.07^2 + 0.07^2)}$ = 9.9%, giving **1,900 ± 188** — still a number you can design against. Now run the same peptide chemistry at wild-type variability: CV 42% propagates to **1,900 ± 798**, a band from roughly 1,100 to 2,700. That is not a measurement; it is a rumor. Precision is not a bonus feature of this work — it is the precondition that turns a modulation into a coordinate.

There is a second error source unique to an allosteric axis, and we name it rather than hide it. Because EC_{50},mod = K_B/α, the score depends on two fitted pharmacological parameters, and the uncertainty in each propagates: $CV(EC_{50},mod) = \sqrt{(CV(K_B)^2 + CV(\alpha)^2)}$. The cooperativity *α* in particular is estimated from the full family of shifted curves and is the harder number to pin; a kokumi score is therefore only ever as trustworthy as the dose-response surface it was fitted from. We quote it with that provenance attached.

And the honest limits beyond the arithmetic:

- **Kokumi is context-dependent by definition.** EC_{50},mod is computed against a defined orthosteric drive — a fixed calcium and amino-acid background. Change the broth and you change the apparent score, because a volume knob does nothing in a silent room. A kokumi number is always a number *in a stated matrix*, never in the abstract, and we report the matrix with the score.
- **A reconstituted assay is not a plate of food.** The receptor work is modeled in controlled calcium; real food brings its own calcium, its own competing peptides, its own pH. Any evaluation would need to be done on the actual cultivar in an actual preparation; the computed score is a prediction, not a final determination.
- **The dipeptides are near-silent alone and only matter in company.** Score γ-Glu-Val by itself and you measure almost nothing; its contribution is real but conditional, and a single-compound EC_{50} understates it.
- **We make no health claim.** The peptides sit on the glutathione pathway, and we will say so with wonder; we will not say, and are not permitted to say, that eating them does for you what glutathione does for a cell. The chemistry is a coincidence we honor, not a benefit we assert.

Where we build it — the alliums, the ferments, the quiet ones

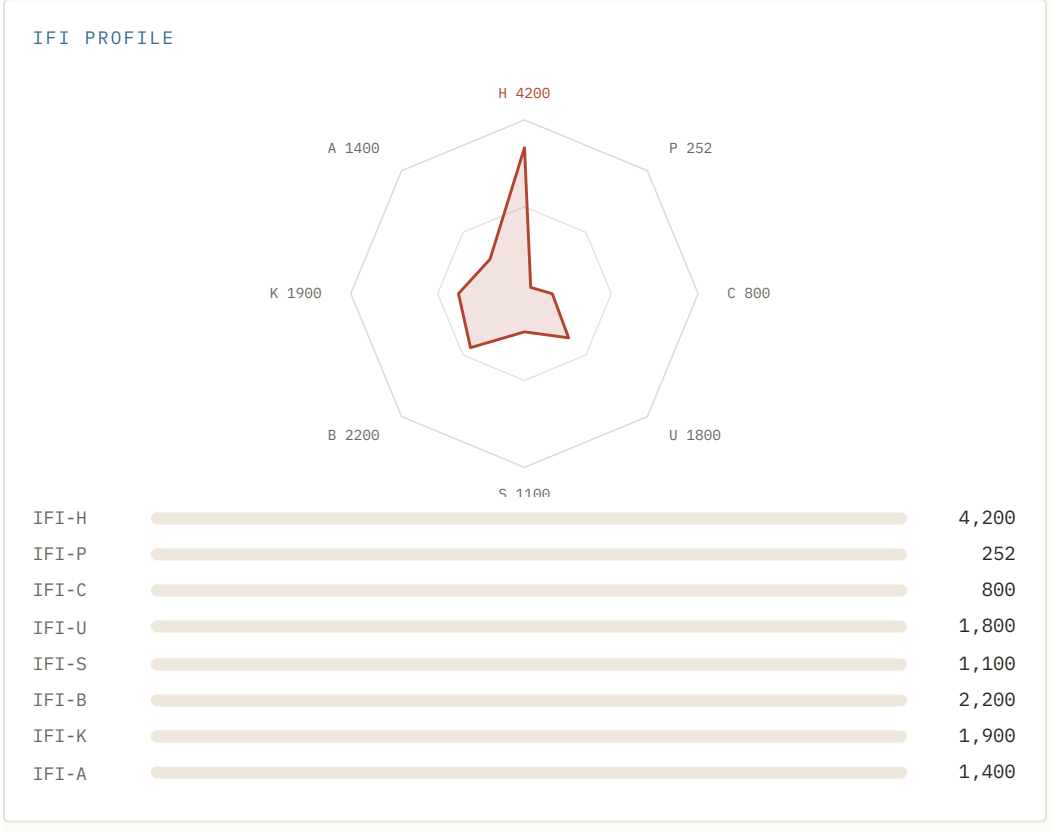
Kokumi is a rare axis in the library, and the roster says so plainly. Of the roughly 19,500 catalog designs across the platform, only **1,240** — about **6.4%** — carry kokumi at all, spread across five hosts. And not one of those hosts carries kokumi *alone*.

HOST	CATALOG DESIGNS	AXES	DIFFICULTY
Garlic	560	K / P	Moderate
Soybean	360	U / K	Frontier
Yeast	240	U / K	Hard
Onion	40	K	Moderate
Vanilla	40	K	Hard
Total K-bearing	**1,240**	—	–

Read down the axis column. Garlic pairs kokumi with pungency; soybean and yeast pair it with umami; even onion and vanilla, the two that look like solo kokumi plays, are built as depth *for* something — a savory allium, a sweet pod. The alliums alone — garlic's 560 and onion's 40 — account for **600 catalog designs, nearly half** of all the kokumi in the library, which is fitting, because garlic and onion run the sulfur-and-amino-acid metabolism that feeds the γ-glutamyl

route at the highest natural flux of any plant we work with. The roster is telling us in design counts exactly what the receptor told us in mechanism: **kokumi is a modulator, and a modulator is built to deepen a flavor that is already there.** You do not engineer a volume knob onto silence.

The proof that the pathway can be written cleanly into a living fruit, and not merely coaxed out of a long cook, is the full-spectrum cultivar. Phoenix was designed to register on every axis at once, and our model predicts it does, kokumi among them:



Look at where the kokumi value sits — K 1,900 — tucked beside U 1,800 and S 1,100, the umami and sweetness it is there to round. That is the whole thesis of the axis made into a single plant: kokumi designed in company, a peptide layer built directly into the fruit to deepen the savor and sweetness the same fruit already carries.

And the kokumi effect is not some engineered novelty; it is the quiet reason aged cheese, seared scallops, fish sauce, and long-fermented pastes all taste *complete*. Each of those has accumulated γ-glutamyl peptides through its own slow enzymatic history — months of the same GGT transfers and γ-glutamyl ligations we drew out above, running at a snail's pace through age and ferment, writing depth into the food long after the harvest. What we do in garlic, in the soybean and yeast of the savory ferments, in onion and vanilla, is simply build that history *into the living plant* — to have the fruit arrive already deep, rather than wait on time to deepen it.



FIG. 11·3 — Soybean. Carrying kokumi alongside umami across 360 catalog designs, the soybean is the savory-depth base of the fermented world — miso, shoyu, fish-sauce analogs — where the γ-glutamyl peptides accumulate slowly through koji and age. Engineering the valyl-glycine branch into the bean builds that long ferment's depth into the raw material.

The nourishment note

There is a nourishment note here we let stand without a word of claim, because the chemistry says it plainly enough on its own. The peptides that give food its sense of satisfying fullness are assembled on the glutathione pathway — built from glutamate, cysteine, and glycine by the very enzymes, γ-ECS and glutathione synthetase, that a cell runs to protect itself from oxidation. The axis the tongue reads as nourishing fullness is, atom for atom, drawn from the chemistry of cellular care.

We do not make a health claim of that. We are not permitted to, and we would not want to; a coincidence is not a cure. We only observe, with the same quiet wonder we felt the first time we followed the pathway to its end, that the molecule of the body's own housekeeping and the molecule of a broth tasting like it was made for you turn out to be the same family of peptide, bonded the wrong way round through glutamate's side chain. We build kokumi where we build it — into the garlic and the onion, the soybean and the yeast, the vanilla and the fruit we design to carry everything — with that coincidence held in mind, and with the care it seems to ask of us.

These are computational designs; Scoville Splice's cultivars and constructs have not been experimentally validated in a laboratory.

Chapter 12 · Sweetness — *diterpene glycosides and the sugar that isn't.*

Sweetness is the oldest yes the tongue knows — the signal, older than language, that a thing is safe and rich and worth carrying home to a child. For all of history it meant one chemistry: sugar, sucrose and its kin, the simple carbohydrates a plant packs into ripe fruit as a bribe to be eaten and carried. This axis is about the others — the molecules that throw the sweet switch without being sugar at all, that deliver the yes without the carbohydrate behind it, and about the branching, patient pathways a plant builds to make them. It is, more than any axis we measure, a story about a *shared precursor* and a quiet negotiation between flavor and the plant's own growing: a plant deciding, enzyme by enzyme, how much of its carbon to spend on being delicious and how much to keep for staying alive.

We love this axis because it is where the Index earns its keep most plainly. *Sweet* is a word that has carried ten thousand years of wanting, and wanting is not a quantity. A number is what lets us build toward it on purpose — to say not merely that a leaf is sweet but *how loudly*, and to walk the carbon backward from the loudness to the gene that set it. That walk is this chapter.



FIG. 12-1 — ⋆Stevia rebaudiana.⋆ The steviol glycosides concentrate in the leaf, assembled from the very diterpene the plant also spends on its growth hormone. Sweetness here is a diverted hormone pathway — and the diversion is the design.

The sweet lock, read closely

Every sweet thing in the world reaches one lock. It is not a single pocket but a single machine: the **TAS1R2–TAS1R3** heterodimer, two class-C G-protein-coupled receptors clasped into one working unit on the surface of the taste cell. Each subunit is built in three pieces — a large extracellular **Venus-flytrap domain** of two lobes that close like a clamshell on a ligand, a rigid **cysteine-rich** linker, and a seven-helix **transmembrane bundle** sunk in the membrane. The genius of

the thing, and the reason one receptor can answer to molecules spanning three orders of magnitude in size, is that it has several doors onto the same switch.

Sugars and the steviol glycosides dock in the flytrap cleft of the TAS1R2 lobe; the clamshell closes around them, and that closure tugs the whole dimer into its active geometry. A few small sweeteners and the bitter-tasting inhibitor lactisole instead bind the transmembrane bundle of TAS1R3, a second door entirely. And the sweet proteins — brazzein, thaumatin, monellin — are far too large to fit any cleft; they press flat against the convex *outer* face of the flytrap and pin it closed from the outside, the mechanism the structural people call the wedge. One lock, many doors, one switch. When the switch throws, the dimer couples to the taste cell's G-protein, gustducin; gustducin frees phospholipase C-β2, which cuts PIP_2 to IP_3; IP_3 dumps calcium from internal stores; the calcium opens the TRPM5 channel; the cell depolarizes and releases its transmitter to the nerve. The transmitter is ATP, pushed into the synapse through the CALHM1 channel the depolarization opens, and it is worth knowing that the TAS1R3 subunit which half-builds this sweet lock is the very same subunit that partners TAS1R1 to build the savory one — one shared half, two different clamshells, which is why sweet and umami sit so close in the biology even as they read opposite ends of a meal. It is a chemical amplifier with four gain stages, and the amplification is exactly why a molecule present in *nanograms* can shout as loudly as a spoonful of sugar.

That amplifier is also why we must measure at the receptor and not in the jar. You cannot read sweetness off a concentration, because two molecules at identical concentration can differ a thousandfold in how hard they throw the switch. What we can read is the **half-maximal concentration** — the dose at which the receptor runs at half its ceiling. Plot response against the logarithm of concentration and you get the familiar S-curve, well described by the Hill relation

```
R([A]) = R_max · [A]ⁿ ⁄ (EC₅₀ⁿ + [A]ⁿ),
```

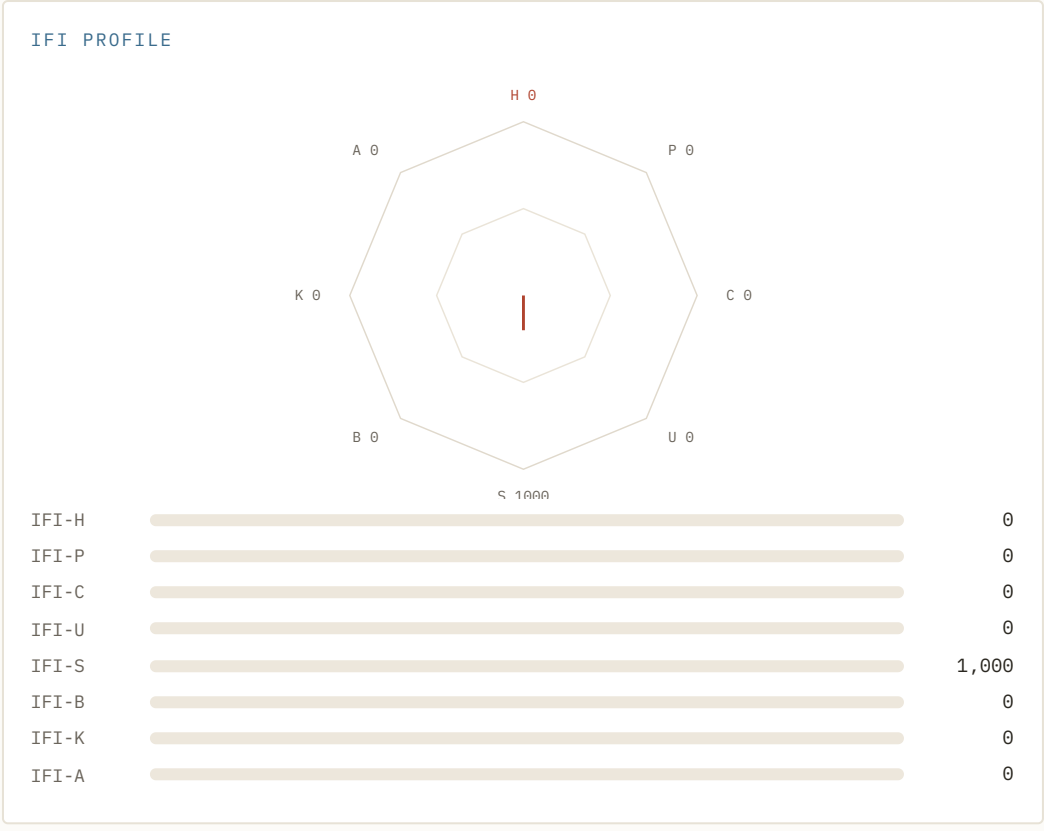
with a Hill coefficient *n* near unity for the sugars and a touch above it for the glycosides — the mild cooperativity you would expect from a two-lobed clamshell that must close to signal. Set [A] = EC_{50} and the equation collapses to R = R_max⁄2, the definition made visible: EC_{50} is the concentration at half-full. The same equation sets the *width* of the response. Rearranged, [A] = EC_{50} · (R⁄(R_max − R))^(1⁄n), so with a Hill coefficient of one the receptor climbs from a tenth of its ceiling to nine-tenths across an eighty-one-fold span of concentration — EC_{50}⁄9 at ten percent, 9·EC_{50} at ninety. Reb A, half-full at 14 µM, is therefore barely stirring at 1.6 µM and all but saturated by 125 µM. Sweetness is not a line; it is a narrow, steep window, and the Index reads it at the midpoint because the midpoint is where the curve stands up tallest and a small change in dose is easiest to see. A small EC_{50} means the switch throws at a whisper. That number, inverted, is loudness — and loudness is what the Index scores.

Loudness is one over EC50

The whole Index rests on one line:

> $IFI(c) = 1000 \times EC_{50}(\text{reference}) \,/\, EC_{50}(\text{candidate})$.

For sweetness the natural reference is sucrose itself — the molecule the whole human story used as its yardstick. Anchor it, and sucrose scores exactly **IFI-S = 1000**, because $1000 \times EC_{50}(\text{sucrose}) \,/\, EC_{50}(\text{sucrose}) = 1000$. That reference constant is a stated convention — defined, like Scoville's original dilution threshold, to give the anchor molecule a round score of 1,000. A sweetener whose EC_{50} is a tenth of sucrose's throws the switch at a tenth of the dose and reads ten times as loud: IFI-S = 10,000 on the raw scale, capped to the Index's per-axis scorecard maximum of 5,000. A plain sugar solution, by construction, is the quietest honest member of the axis and silent on the other seven receptors — it does not gate the heat channel, the pungency channel, or any of the rest:



That block is not a measurement; it is the anchor the measurements hang from — the zero-point of taste, rendered as a profile so the rest of the axis has something to be louder *than*.

Now the arithmetic that makes the intense sweeteners worth the trouble, and the one unit trap we insist on walking through slowly, because confusing the two numbers is the most common error we see. The familiar figure — *stevia is 250 times sweeter than sugar* — is a **mass** ratio: equal sweetness from a 250th of the weight. The

receptor, though, counts **molecules**, not grams, and the glycosides are heavy molecules. To turn a mass ratio into the molar potency the receptor actually feels, scale by the molecular-weight ratio:

```
potency(molar) = sweetness(mass) × MW(candidate) ⁄ MW(sucrose),
with MW(sucrose) = 342 g/mol.
```

Rebaudioside A weighs 967 g/mol, so its molar potency is 250 × (967 ⁄ 342) = 250 × 2.83 ≈ **710**. Seven hundred ten molecules of sucrose to match one of reb A. Invert against a cell-assay sucrose half-max of EC_{50} ≈ 10 mM and you recover the receptor number directly: EC_{50}(reb A) = 10 mM ⁄ 710 = 0.014 mM = **14 µM**. Run the same two lines through each sweetener and the library falls into a single table — mass ratio in, molar potency and EC_{50} out:

COMPOUND	CLASS	MW (G/MOL)	×SUCROSE (MASS)	×SUCROSE (MOLAR)	EC_{50} (CELL ASSAY)
Sucrose	disaccharide	342	1	1	10 mM
Stevioside	steviol glycoside	805	~200	~470	~21 µM
Rebaudioside A	steviol glycoside	967	~250	~710	~14 µM
Mogroside V	triterpene glycoside	1,287	~250	~940	~11 µM
Brazzein	sweet protein	6,500	~1,000	~19,000	~0.53 µM
Thaumatin	sweet protein	22,200	~2,500	~162,000	~62 nM

Read the right-hand columns and the sweet proteins separate from the glycosides not by a little but by a cliff. Thaumatin at 62 nanomolar half-activation is roughly one hundred sixty thousand times sucrose on a molar basis, because it is both intrinsically tight *and* enormous. That is the whole reason these molecules exist in our library: a molecule a few hundred to a hundred-thousand times as loud as sugar delivers the identical signal at a vanishing mass, so that on the Index it posts a high IFI-S at a concentration that contributes essentially no carbohydrate to the cup. We do not call that a health claim, and we will not; it is simply what the chemistry is — a yes to the sweet receptor carried by a molecule the body meets in milligram or microgram quantity. The plant makes the yes. Our work is to make the plant make *enough* of it, cleanly, without forgetting to grow.

The staircase to steviol

To make enough of it, the plant climbs a staircase of terpene chemistry that begins exactly where mint begins and then turns toward a different world. The headline sweeteners of this library are the **steviol glycosides** — stevioside and rebaudioside A from *Stevia rebaudiana*, behind which sit **910 engineered constructs** in our hands. Their route is a diterpene route, and it starts from a twenty-carbon prenyl backbone, **geranylgeranyl diphosphate (GGPP)**.

Two cyclases fold GGPP in sequence. **CPS**, *ent*-copalyl diphosphate synthase, a class-II diterpene cyclase, performs the first fold — a protonation-initiated cyclization that bends the floppy chain into the bicyclic *ent*-copalyl diphosphate. **KS**, *ent*-kaurene synthase, a class-I cyclase, performs the second, ionizing the diphosphate and closing the molecule into the rigid tetracyclic cage of *ent*-kaurene. Then a cytochrome P450, **KO** (*ent*-kaurene oxidase, CYP701A), oxidizes one

methyl group through three successive steps — alcohol, aldehyde, acid — to **kaurenoic acid**. *And here is the fork that defines the whole chapter*, because kaurenoic acid is not a sweetener precursor by birth. It is the plant's own on-ramp to **gibberellin**, the central growth hormone that governs stem elongation, germination, and the plant's sense of its own stature. A second P450, **KAH** (kaurenoic acid 13-hydroxylase, of the CYP714A family), diverts the acid away from the hormone by hanging a hydroxyl at carbon 13, yielding **steviol** — the sweet-but-bitter aglycone, the raw sweetness before it is dressed.

STEP	ENZYME	CLASS	REACTION	COFACTOR · SITE
1	CPS · *ent*-copalyl-PP synthase	class-II diterpene cyclase	GGPP → *ent*-copalyl-PP	Mg^{2+} · plastid
2	KS · *ent*-kaurene synthase	class-I diterpene cyclase	*ent*-copalyl-PP → *ent*-kaurene	plastid
3	KO · CYP701A	cytochrome P450	*ent*-kaurene → kaurenoic acid *(×3 oxidations)*	$NADPH$, O_2 · ER — *shared with gibberellin*
4	KAH · CYP714A	cytochrome P450	kaurenoic acid → steviol	$NADPH$, O_2 — *the diversion*
5	UGT85C2	UDP-glycosyltransferase	steviol → steviolmonoside (13-*O*-glc)	UDP-glucose
6	UGT74G1 · UGT91D2	UDP-glycosyltransferase	→ stevioside	UDP-glucose
7	**UGT76G1**	UDP-glycosyltransferase	stevioside → **rebaudioside A** (1,3-branch glc)	UDP-glucose — *the clean-sweet step*

The fork at kaurenoic acid

That fork is the deepest lesson the sweetness axis teaches, and it is a lesson in flux, not in yield. Sweetness and growth draw on the same kaurenoic acid, so a plant pushed hard to pour carbon down the KAH branch toward steviol is a plant spending precursor it would otherwise have spent on gibberellin — on reaching for light. Tune KAH too aggressively and you can engineer a very sweet plant that is also a very stunted one, sugar-rich and barely able to stand. The design target is therefore never *maximize steviol*. It is *partition the fork*.

Make the logic quantitative with the simplest honest model. Let **f** be the fraction of kaurenoic-acid flux pulled toward steviol, leaving (1 − f) for the growth branch, and suppose — as a first approximation — that the plant's harvestable biomass falls roughly in proportion to the hormone flux it keeps, $B(f) \approx B_0 \cdot (1 - f)$. What a farmer actually harvests is not sweetness per leaf but sweetness per field: the product of how sweet each gram of leaf is and how much leaf there is to pick. Write that harvest as

$$H(f) = f \cdot B(f) = f \cdot B_0 \cdot (1 - f).$$

Differentiate and set to zero: $dH/df = B_0 \cdot (1 - 2f) = 0$, so **$f^* = 0.5$**. The sweetest *crop* — not the sweetest leaf — sits near the middle of the fork, with roughly half the shared precursor diverted and half left to grow on. A real plant's biomass curve is gentler than this toy line, and the true optimum shifts with light, nitrogen, and the particular

line, but the shape of the answer does not: the maximum is an *interior* point, never the boundary. A plant tuned to f → 1 is the classic failure mode of this axis — a sweet curiosity you cannot farm, stunted and shy, its glycosides locked in a plant too small to pay for the harvest. The honest practice of sweetness is this negotiation, and holding it well is what separates a sweet leaf you can grow by the acre from a sweet accident you can grow once.



FIG. 12-2 — A research greenhouse at golden hour. Sweetness begins here, in the chloroplast's own isoprenoid factory, as carbon the plant could have spent on growing; every engineered line is a decision about how much of that carbon to divert and how much to leave the plant to stand on.

Dressing steviol in sugar

Steviol as it comes off the fork is sweet, but it is also harsh — a lingering, almost medicinal bitterness rides alongside the sweetness, because the bare aglycone is promiscuous and brushes the bitter receptors as well as the sweet. The plant's answer, and ours, is to dress it. A cascade of **UDP-glycosyltransferases** hangs glucose units onto steviol one at a time, and each sugar changes not just the solubility but the *shape* the receptors read.

UGT85C2 goes first, attaching glucose at the 13-position hydroxyl to make steviolmonoside. Then **UGT74G1** and **UGT91D2** build out the 19-position and extend the first chain, assembling **stevioside** — properly sweet now, far less harsh than the aglycone, but still carrying a bitter backbite that anyone who has tasted cheap stevia knows. The decisive enzyme is the last one. **UGT76G1** adds a single branch glucose in a β-1,3 linkage, converting stevioside into **rebaudioside A** — and that one sugar is the difference between a sweetener people tolerate and one they reach for. The branch glucose rounds the molecule's fit in the flytrap cleft and, crucially, lifts it off the bitter receptors, so reb A reads as clean, rounded sweetness where stevioside reads as sweet-with-an-edge.

The dressing is load-bearing in a second way that is easy to miss: bare steviol is only grudgingly soluble in water and would never reach the tongue at a tasting concentration at all — the glucose units are what carry it into solution, so the same additions that round its bitterness are what make it tasteable in the first place. And there is a cost hidden in them. Every glucose hung on steviol is drawn from the cell's pool of UDP-glucose, the same activated sugar the plant spends on building its cell walls, so a line driven to over-glycosylate can quietly starve its own structure — another fork, another negotiation, one compartment down from the first.

This gives us a quality number that is not sweetness at all but *ratio*: the reb-A-to-stevioside balance in a given leaf is, in practice, the balance between a clean cup and a bitter one. Pushing UGT76G1 expression is one of the highest-value single edits in the whole sweetness program, because it does not make the leaf sweeter — the molar potencies of stevioside (~~470×) and reb A (~~710×) are close — it makes the same sweetness *kinder*. We tune the glycosyltransferase cascade the way a cook tunes seasoning: not for more, but for right.

Three backbones, three answers

Stevia is one answer to the sugar-free problem. The library holds two more, and setting them side by side is the clearest way to see that *sweet* is a destination many different molecules reach by roads that never meet.

Monk fruit's **mogrosides** — behind which sit **80 constructs** in our hands — are **triterpene** glycosides, built on a thirty-carbon backbone rather than the twenty-carbon diterpene of steviol. Different precursor, different cyclases, different enzymes dressing the core in glucose, and yet mogroside V arrives at almost exactly the same place: roughly 250 times sucrose by mass, ~940 molar, an EC_{50} near 11 µM. Two unrelated terpene routes, converging on the one receptor. That convergence is the axis's quiet argument that sweetness is a property of *fit*, not of family.



FIG. 12-3 — ⋆Siraitia grosvenorii⋆, monk fruit. Its mogrosides are triterpene glycosides — a thirty-carbon backbone, an entirely separate road from stevia's diterpene — arriving at the same sweet receptor. Convergent chemistry: two pathways, one lock.

The **sweet proteins** — brazzein, thaumatin, monellin — are a different thing entirely: not small molecules off any terpene staircase but proteins, read straight from the ribosome, that happen to fold into shapes the receptor reads as intensely sweet. They do not dock in the flytrap cleft at all; they are the wedge, gripping the receptor's outer face. On a molar basis they are the loudest sweeteners we know — thaumatin near 162,000× sucrose — which is both their gift and their difficulty, a tension we will own honestly in a moment. Among them brazzein is the quiet favorite: not the loudest, but by a distance the toughest — unusually steady to heat and to acid where thaumatin surrenders to both — which is why a protein a tenth of thaumatin's molar loudness earns its own place in the library, a

sweetness you could in principle carry through a cooked or an acidic food without watching it unravel. And then there is **sugarcane**, with **200 constructs**, which answers the sweetness question by making more of the one molecule all the others only imitate: sucrose itself, the sugar that *is* sugar, tuned for titer rather than potency. The library's sweet-axis specialists line up by strategy:

HOST	CONSTRUCTS	SWEET STRATEGY	DIFFICULTY
Stevia	910	steviol glycosides — the diterpene route	Moderate
Tomato	2,040	sweetness restored, within S · U · F · M	Easy
Sugarcane	200	sucrose titer — the sugar that is sugar	Frontier
Monk fruit	80	mogrosides — the triterpene route	Hard

These four sit inside a platform spanning **42 host organisms** and roughly **19,500 product SKUs**, of which about **1,000** are live and purchasable today — the sweetness axis is one lane of a much larger road, but it is the lane where the gap between *loud* and *light* opens widest.

A shared headwater

There is a compartment note that ties this axis to a plant most people would never file beside stevia. The diterpene precursor GGPP that feeds the entire steviol staircase is not built in the cytoplasm; it is assembled in the **plastid**, the chloroplast's own isoprenoid factory, by the **MEP pathway** — the methylerythritol-phosphate route that runs on pyruvate and glyceraldehyde-3-phosphate straight out of photosynthesis. The same plastid, by the same MEP machinery, builds the ten-carbon precursor GPP that a peppermint leaf folds into menthol for its cooling. Follow the carbon far enough upstream and the sweet of a stevia leaf and the cool of a mint leaf share a single headwater, diverging only once the prenyl chain is long enough and the first cyclase decides what to fold it into. Sweetness, in other words, begins in the light — in the chloroplast — and is diverted, at the kaurenoic-acid fork, away from the growth hormone and toward the leaf. We find that lineage moving: the yes on the tongue is photosynthate, reshaped.

What the error bars say — and what we do not claim

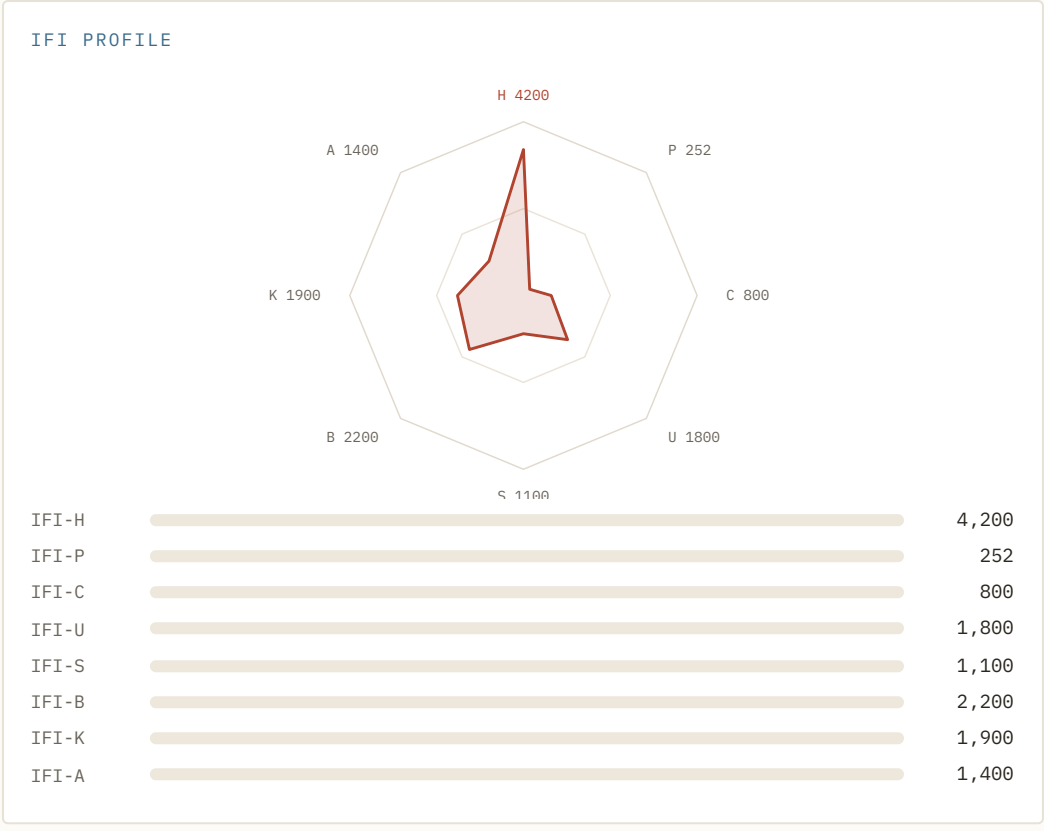
A number is only as honest as its scatter, so we report the scatter. The platform's signature is reproducibility: an engineered line is designed to hold its titer to a batch-to-batch coefficient of variation of **≤7%**, where a wild, unengineered crop of the same species wanders at roughly **42%** — a sweetness that swings with weather and season and the luck of a field. The receptor math inherits that discipline. Because IFI-S is a *ratio* of two half-maximal concentrations, its error is the error of a quotient, and for a quotient the relative variances add in quadrature:

$$\mathrm{CV}(\text{IFI-S}) = \sqrt{\ \mathrm{CV}[EC_{50}(\text{candidate})]^2 + \mathrm{CV}[EC_{50}(\text{reference})]^2\ }.$$

Measure each EC_{50} at a tidy 5% and the reported IFI-S carries $\sqrt{(0.05^2 + 0.05^2)} = \sqrt{0.005} = \mathbf{7.1\%}$ — which is why ≤7% is the number we hold the whole surface to, not a slogan but the arithmetic of two careful measurements multiplied. A score we cannot post at that tolerance is a score we do not post.

The honest limits run deeper than scatter, and they fall hardest on the loudest molecules. The sweet proteins pay for their nanomolar potency with their kinetics: thaumatin and its kin have a slower onset than sugar and a long, faintly licorice tail that lingers after sugar would have cleared — a different *time signature* of sweetness that the single EC_{50} number does not capture. They are proteins, so they denature; bake them or drop the pH too far and the fold unravels and the sweetness simply leaves, which rules them out of many of the very applications where sugar is hardest to replace. The steviol glycosides have their own edge — the bitter backbite of stevioside that only the UGT76G1 branch glucose fully tames, and a cooling, slightly delayed sweetness that is wonderful in a cold drink and strange in a warm one. And the whole axis runs into a ceiling that no potency can climb past: because the Index caps each axis at 5,000, a molecule a thousand times too loud and a molecule twice too loud both peg the same wall, so for the intense sweeteners the design question is never *how loud can we go* but *what loudness do we want*, posted precisely and held steady. We tune toward a chosen number, not toward the maximum.

Even our fiercest cultivar carries the point. Phoenix is a superhot designed to register on all eight axes at once, and its sweetness is not an accident but a dialed, deliberate touch — **S: 1100**, a hair above the sucrose anchor, a deliberate roundness set under a wall of heat:



A single pepper holding a predicted, reproducible 1100 on the sweet axis is the whole philosophy in one line — sweetness as a number you place on purpose, even in the middle of a fire.

What it means at the table

Let us end where the family does, with a cup. Take an engineered stevia line running near 10% steviol glycosides by dry weight and brew a strong infusion — 2 grams of dried leaf in 200 millilitres of water. That is 0.20 g of glycoside in 0.200 L, or **1.0 g/L** in the cup. As

reb A (967 g/mol), that is 1.0 g/L ÷ 967 g/mol ≈ **1.03 mM**. Multiply by reb A's molar potency of ~710 and the cup carries the sweet signal of 1.03 mM × 710 ≈ **730 mM** of sucrose-equivalent — which, at 342 g/mol, is 0.73 mol/L × 342 g/mol ≈ **250 g/L**, a 25% sugar syrup's worth of sweetness. Against an ordinary sweet beverage at 10% sucrose, that infusion reads two and a half times as loud: an **IFI-S of roughly 2,500**. And the load behind the loudness is one gram of glycoside per litre where the matching sweetness in sugar would be two hundred fifty — a 250-fold lighter cup for an identical yes.

We will not dress that number as anything it is not. It is not a cure or a promise; it is a mass, and the mass is small. What it means at the table is quieter and, to us, more beautiful: that a plant can be grown to speak the oldest word the tongue knows — *safe, rich, worth eating* — and speak it at a whisper of its own substance, so the pleasure arrives nearly weightless. The plant does the making. Our hands do the computing and the tuning — enough sweetness, cleanly, from a plant we have been careful to leave strong enough to stand. That is the diterpene glycoside's gift, and the sugar that isn't: the yes, kept; the load, set down.

These are computational designs; Scoville Splice's cultivars and constructs have not been experimentally validated in a laboratory.

Chapter 13 · Astringency — *the tannin route, drawn to the polymer.*

Astringency is the odd one out among the eight axes, and drawing its pathway forces an honest admission before the first arrow is inked: astringency is not, strictly, a taste at all. It is a *tactile* sensation — the dry, puckering, rough-cloth feeling that follows a tannic red, a black tea steeped a minute too long, the chalky skin of an unripe persimmon, a square of very dark chocolate held on the tongue. There is no astringency receptor the way there is a sweet receptor or a bitter one. There is, instead, a molecule that reaches into your saliva and *removes the lubrication* — and the sensation is the rest of your mouth noticing, through ordinary touch, that it has suddenly gone dry. This is the chemesthetic edge of the library, the place where flavor stops being chemistry read by a dedicated sensor and becomes physics felt by the whole mucosa.

It is also the one route in this book that does not end at a small molecule. Every other pathway we draw terminates in a tidy compound that docks into a protein pocket, a single structure we can name and weigh. This one keeps going — condensing, unit upon unit, past the point where any one structure can be written down — into something large, variable, and, in the most literal sense, a *distribution* rather than a thing. To design astringency is to design the shape of a polymer; to score it is to measure a precipitation reaction happening in real time on the tongue. We find that honest and a little thrilling. Of all eight axes, this is the one that most resists being made into a number, and so the one where a number, carefully built, is worth the most.



FIG. 13.1 — Grape in cross-section. The tannin concentrates in the seed and the skin, almost never the pulp; a wine's structure is therefore a geography — a consequence of how hard the fruit is pressed and how long the juice is left to sit on its skins. The astringency is a map of where the polymer lives.

The route from the great junction

The trunk is the phenylpropanoid pathway, the same chain of condensations every vascular plant runs to turn the amino acid phenylalanine into the raw material of lignin, pigment, scent, and defense. At its great junction sits *p*-coumaroyl-CoA — the roundabout from which the roads to pigment, to aroma, and to tannin

all depart. The branch we want is the flavonoid branch, and it opens with what is arguably the single most committed step in all of plant specialized metabolism.

Chalcone synthase (CHS) takes one molecule of *p*-coumaroyl-CoA and condenses it, head to tail, with three molecules of malonyl-CoA — a triple iterative Claisen condensation that builds a fresh aromatic ring from scratch and loses three molecules of CO_2 along the way — to give naringenin chalcone. Nothing downstream of CHS is anything but a flavonoid; this is the gate, the point of no return into the whole pigment-and-tannin world. **Chalcone isomerase** (CHI) then performs a stereospecific ring closure, folding the open chalcone into the three-ringed flavanone naringenin. The reaction is thermodynamically willing — it will proceed slowly even in the absence of the enzyme — but CHI makes it fast and, crucially, makes it *chiral*, setting the (2*S*) configuration the rest of the pathway is built upon.

From naringenin the road climbs. **Flavanone 3-hydroxylase** (F3H), a 2-oxoglutarate–dependent dioxygenase that spends one molecule of 2-oxoglutarate and one of O_2 per turn, installs a hydroxyl at the 3-position and hands back a dihydroflavonol. **Dihydroflavonol 4-reductase** (DFR), drawing on NADPH for reducing power, reduces the 4-keto group to a 4-hydroxyl and yields a leucoanthocyanidin — a flavan-3,4-diol. Here, at the leucoanthocyanidin, the road forks into the two building blocks that will polymerize into tannin, and the fork is the whole design decision of the axis.

One branch runs through **leucoanthocyanidin reductase** (LAR), which reduces the 4-hydroxyl clean away to give (+)-catechin, a flavan-3-ol of (2*R*,3*S*) stereochemistry. The other branch runs through **anthocyanidin synthase** (ANS), which oxidizes the leucoanthocyanidin to a colored anthocyanidin, and then **anthocyanidin reductase** (ANR) — the enzyme we turn up hardest when we want a plant to spend its flavonoid budget on tannin rather than on pigment — reduces that anthocyanidin to (−)-epicatechin, the (2*R*,3*R*) epimer. Catechin and epicatechin are the **flavan-3-ols**: the monomers. And the pathway's final move is the one no enzyme in the canonical map catalyzes — a slow, largely non-enzymatic *condensation* in which flavan-3-ol units link, one to the next, into **proanthocyanidins**: the condensed tannins, polymers a few to many units long, the molecules that will dry your mouth.

STEP	ENZYME	COFACTOR / CO-SUBSTRATE	TRANSFORMATION
1	CHS · chalcone synthase	3 × malonyl-CoA	*p*-coumaroyl-CoA → naringenin chalcone (−3 CO_2)
2	CHI · chalcone isomerase	—	chalcone → (2*S*)-naringenin (ring closure)
3	F3H · flavanone 3-hydroxylase	2-oxoglutarate, O_2	naringenin → dihydroflavonol (3-OH)
4	DFR · dihydroflavonol 4-reductase	NADPH	dihydroflavonol → leucoanthocyanidin (4-OH)
5a	LAR · leucoanthocyanidin reductase	NADPH	leucoanthocyanidin → (+)-catechin
5b	ANS → ANR	2-oxoglutarate · O_2 ; NADPH	leucoanthocyanidin → anthocyanidin → (−)-epicatechin
6	condensation (largely non-enzymatic)	—	flavan-3-ols → **proanthocyanidins** (tannin)

That last row is the one that makes this chapter different from every other in the book. A design program can tune the enzymes above it with a precision we are proud of — but the step that actually sets how a tannin *feels* is the one no enzyme is steering, and so the one we have to bias from the side rather than command outright.

Two families, one physics

Before the sensation, a word on the two great families of tannin, because they dry the mouth by the same physics from entirely unrelated chemistry — and because a design program must know which family it is holding before it reaches for a lever.

The **condensed tannins** traced in that final step — the proanthocyanidins polymerized from catechin and epicatechin — are the tannins of grape, tea, cacao, sorghum, blueberry, and most of the plants on this shelf. They are all carbon-to-carbon backbone: the flavan-3-ol units are welded directly to one another through interflavan bonds that water cannot touch, which is exactly why they are called *condensed* and not *hydrolyzable*.

The **hydrolyzable tannins** are a separate lineage entirely, built not on the flavonoid branch at all but on a core of glucose esterified with gallic acid. Ester gallic acid alone onto the sugar and you have a **gallotannin**; let two gallic acids couple into hexahydroxydiphenic acid first and you have an **ellagitannin**. The punicalagins that make a pomegranate's white pith so fiercely astringent are ellagitannins; so, in part, are the tannins locked into a raspberry's drupelets. These are not phenylpropanoid polymers but a parallel solution to the same defensive problem, and the clue in the name is real: a dilute acid or a patient enzyme will *hydrolyze* those ester bonds and release gallic or ellagic acid, which the carbon-carbon bonds of a proanthocyanidin will never surrender.

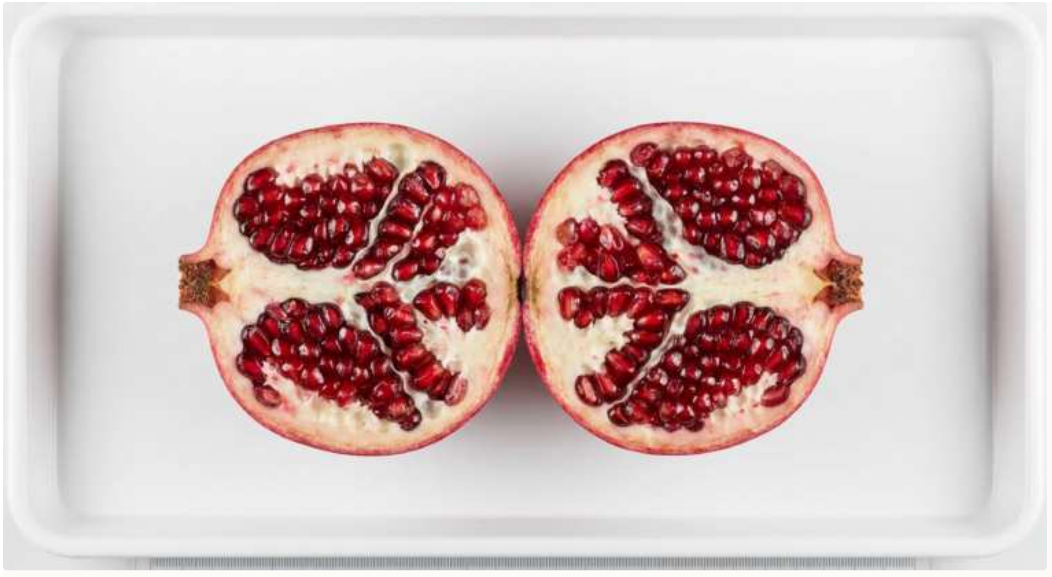


FIG. 13.2 — Pomegranate in cross-section. The ferocious astringency lives not in the ruby arils but in the pale membrane and pith, where the ellagitannins — the punicalagins — concentrate. A hydrolyzable tannin, built on a sugar core, grips protein by the same physics as grape's condensed tannin but is engineered on an entirely different route.

Both families grip protein; both strip the mouth; and because their chemistries are unrelated, their design levers are too — which is why pomegranate and grape, though they sit on the same axis, are built on different routes and tuned with different hands. Across the astringency hosts, the one shared vocabulary is the sensation; the chemistry underneath it is two distinct problems wearing one name.

HOST	CONSTRUCTS	TANNIN FAMILY	CHARACTERISTIC POLYMER	PRIMARY AXES
Grape	1,610	condensed	proanthocyanidin (catechin / epicatechin)	B · A · M
Tea	280	condensed	galloylated catechins (EGCG)	U · B · A
Blueberry	120	condensed	proanthocyanidin	A
Pomegranate	120	hydrolyzable	punicalagins (ellagitannin)	A
Sorghum	120	condensed	proanthocyanidin	A
Raspberry	40	condensed / ellagitannin	mixed	A

Grape leads the axis with 1,610 constructs — more than the other five astringency hosts together — and not by accident. Wine is the human artifact in which tannin is most deliberately shaped, and the one in which a reader already carries a vocabulary for what the number means: *grip, structure, backbone, finish.* When we design on grape, we are designing against two thousand years of palates who already know the axis by feel.

The polymer takes shape

Everything about how a tannin feels is set in the condensation step — the uncatalyzed one — and so a program lives or dies on its ability to bias a reaction no enzyme is steering. Three variables carry almost all of the sensory weight: the **degree of polymerization**, the **monomer identity**, and the **galloylation**.

Degree of polymerization (DP) is simply the chain length — how many flavan-3-ol units are strung together — and it sets the molecular weight directly, which is the first thing governing whether a tannin can bridge two proteins at once. A terminal flavan-3-ol unit weighs 290.27 g/mol ($C_{15}H_{14}O_6$); each additional unit joins through an interflavan C–C bond with the net loss of two hydrogens, so every *extension* unit adds 288.25 g/mol rather than a full 290.27. The weight of a chain of degree n is therefore

$$MW(n) = 290.27 + (n - 1) \times 288.25 \text{ g/mol}.$$

DP (N)	EXAMPLE	MW (G/MOL)
1	catechin / epicatechin (monomer)	290.27
2	procyanidin dimer (e.g. B2)	578.52
3	procyanidin trimer (C1)	866.77
5	short oligomer	1,443.27
7	oligomer	2,019.77
10	polymer	2,884.52

Below a DP of about three a tannin can barely bridge two protein chains at once, and it reads as more bitter than astringent — a monomeric catechin is frankly bitter and only whispers of dryness. The puckering grip arrives as DP climbs into the middle single digits and strengthens from there, until at very high DP the polymer becomes so large and so avid that it precipitates protein almost on contact: the hard, green, mouth-scouring astringency of an unripe seed or a wine pressed too roughly off its skins.

Monomer identity is the second lever. Catechin [(2*R*,3*S*)] and epicatechin [(2*R*,3*R*)] are stereoisomers of identical weight, but the *cis* geometry of epicatechin lets its chains fold and present their hydroxyls to a protein surface a little differently from the *trans* catechin — so a tannin's catechin-to-epicatechin ratio shifts the *quality* of the grip even at constant chain length. The interflavan linkage carries its own quieter vote: the common B-type bond is a single C4→C8 (or C4→C6) carbon bridge, while the rarer A-type carries an additional ether bond that stiffens the chain and changes how it drapes across a protein.

Galloylation is the third and sharpest lever. A flavan-3-ol can be esterified at its 3-hydroxyl with a molecule of gallic acid (170.12 g/mol); because the ester forms with the loss of water (18.02 g/mol), each galloyl group adds 170.12 − 18.02 = 152.10 g/mol and, far more importantly, a fresh cluster of three hydroxyls that grip protein hard. This is the chemistry of tea. Epigallocatechin gallate — EGCG, the dominant catechin of green tea — is epigallocatechin (306.27 g/mol) carrying one galloyl ester: 306.27 + 152.10 = 458.37 g/mol, and that single gallate is much of why a strong green tea pulls at the mouth far above its weight. A galloylated monomer can out-grip an ungalloylated trimer. It is the most efficient astringency per gram the pathway can make, and the most dangerous to over-apply.



FIG. 13.3 — Tea. The young leaf is the richest source of galloylated catechins the library holds; EGCG alone can carry a cup's whole grip. In tea, astringency is tuned less by chain length than by the gallate decoration on short, soluble monomers — the same lever, pulled from the other end.

The sensation is a precipitation

Here is where astringency earns its separate chapter, because the mechanism is not receptor binding at all — it is colloid chemistry on the surface of the tongue.

Saliva is, among many other things, a solution of **proline-rich proteins**, the PRPs, which the salivary glands secrete in startling abundance for what appears to be exactly this purpose: to be sacrificed to tannin so your tissues are not. PRPs are open, floppy, proline-studded chains with no compact fold, and that shapelessness is the point. The pyrrolidine ring of proline presents a flat hydrophobic face that stacks against the aromatic rings of a flavan-3-ol, while the tannin's many hydroxyls throw hydrogen bonds across to the peptide backbone. A single tannin — large, flexible, covered in binding groups — does not bind one PRP; it bridges several at once, cross-linking them into a growing aggregate that passes the threshold of solubility and *precipitates*. The thin lubricating film of saliva that normally lets the surfaces of the mouth glide over one another is pulled out of solution, and what you feel as astringency is the friction of an unlubricated mouth: touch, reporting a precipitation reaction in real time.

The PRPs arrive in their own classes — acidic, basic, and glycosylated — and a given tannin binds them with different avidities. Much of why one tannin reads coarse and another silky at the same intensity is simply which PRP population it takes down first: the basic and glycosylated PRPs in particular seem to be a frontline the body spends freely and early against tannin. This is a second reason the axis resists a single clean number. The stimulus is met not by one receptor of fixed abundance but by a defensive protein pool that varies between people, varies across a single meal as it depletes, and is in a real sense *consumed* by the very act of measuring it.

The precipitation follows the mathematics of cooperative, multivalent binding, which we model with a Hill relation. Let θ be the fraction of the maximum PRP precipitate a tannin can throw, and c the tannin concentration; then

$$\theta(c) = c^h / (EC_{50}{}^h + c^h),$$

where EC_{50} is the concentration at half-maximal precipitation — the anchor of the whole axis — and h is the Hill coefficient, a direct readout of how cooperative the cross-linking is. The arithmetic is worth doing, because it explains the *character* of an astringency and not merely its size. The concentration to reach 90% of maximum is EC_{50} × 9^(1/h); to reach 10% it is EC_{50} × (1/9)^(1/h); so the full 10%-to-90% onset spans a concentration ratio of 9^(2/h) = 81^(1/h):

- At h = 1 — a monomer binding one site, no cross-linking — the onset spans 81×: a broad, gentle, barely-there dryness that creeps in across a wide range of concentration. This *is* the near-absence of true astringency in a monomeric catechin, where bitterness dominates and the dryness never quite resolves.
- At h = 2 the span narrows to 9×.
- At h = 4 — a long, multivalent polymer cross-linking cooperatively — it collapses to 81^(1/4) = 3×: a sharp, switch-like grip that arrives almost all at once. This is the hard snap of a high-DP tannin that seems to clamp the mouth shut in a single swallow.

So chain length does two things at the same time. It lowers the EC_{50} — fewer molecules are needed, because each one does more work — and it raises h, sharpening the onset. The design of an astringency is, in the end, the design of this one curve: where it sits, and how steeply it turns.

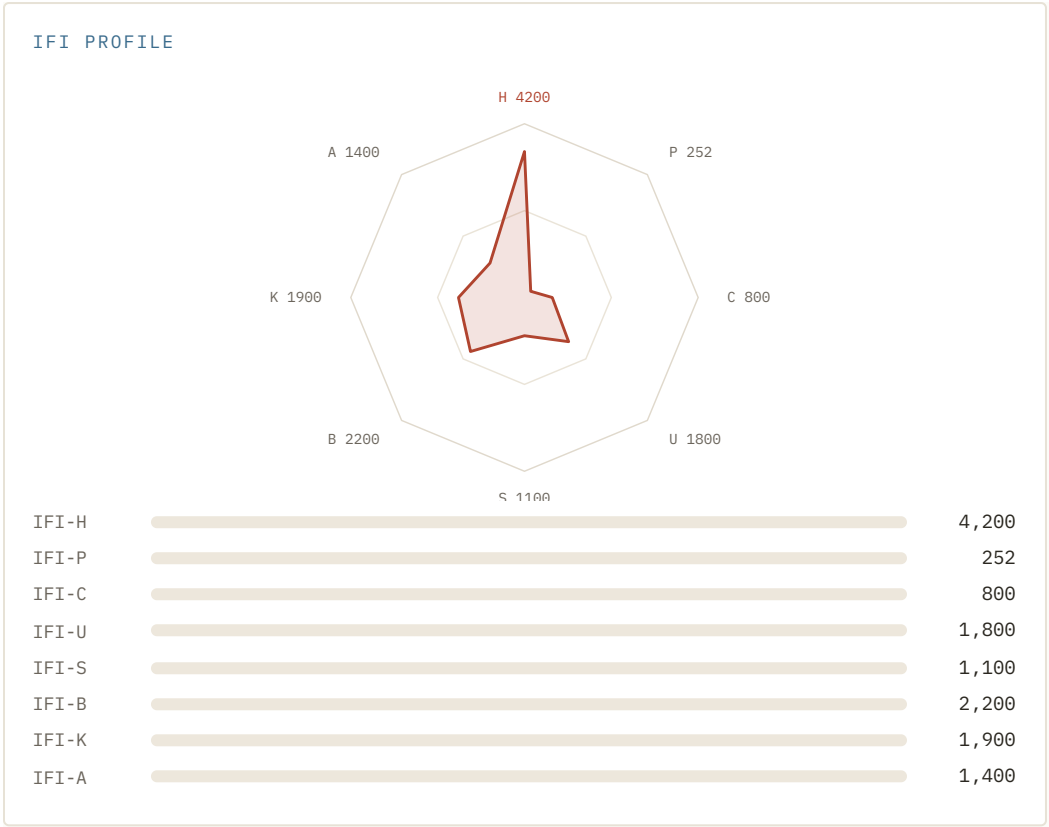
Reading the axis

The Index measures loudness, not quantity. The score on any axis is anchored to the half-maximal activation of its receptor — here, the half-maximal PRP precipitation EC_{50} — and read against a fixed reference:

$$IFI(candidate) = 1{,}000 \times EC_{50}(reference) \div EC_{50}(candidate).$$

The reference is a stated convention — defined to sit at exactly 1,000, the way the Scoville scale is defined by dilution-to-threshold. Set EC_{50}(candidate) = EC_{50}(reference) and the ratio is 1,000 on the nose. A candidate that is *louder* — that reaches half-maximal PRP precipitation at a *lower* concentration — scores above 1,000; a quieter one scores below. The scorecard normalizes each axis to a ceiling of 5,000.

Work a real number through it. Our full-spectrum flagship, Phoenix (SS-0096), is a plant we designed mostly to burn — but it is the first cultivar engineered to register on all eight axes at once, and our model predicts astringency as one of them at **IFI-A = 1,400**: the quiet tannic note threaded under the fire.



That 1,400 is not a tannin *amount*; it is a statement about an EC_{50}. Rearranging the formula,

> EC_{50}(candidate) = 1,000 × EC_{50}(reference) ÷ IFI = EC_{50}(reference) × 1,000 / 1,400 = 0.714 × EC_{50}(reference).

Our model predicts Phoenix's tannin reaches half-maximal PRP precipitation at 71.4% of the reference concentration — 1.4× *louder* than the anchor, a real but gentle astringency, exactly as the number should read against a palate that expects the heat to lead. To make the arithmetic tangible, fix the reference anchor — for illustration only; it is the *ratio* the canonical score pins down, not the absolute — at 100 µM, expressed in flavan-3-ol monomer-equivalents. Then Phoenix's astringency EC_{50} is 71.4 µM monomer-equivalent, and converting to mass with the monomer weight of 290.27 g/mol:

> 71.4 µmol/L × 290.27 µg/µmol = 20.7 mg/L, against the reference's 100 µM × 290.27 = 29.0 mg/L.

SAMPLE	IFI-A	EC_{50} ÷ EC_{50}(REF)	RELATIVE LOUDNESS
Reference anchor	1,000	1.000	1.0×
Phoenix · SS-0096	1,400	0.714	1.4×
Axis ceiling (scorecard)	5,000	0.200	5.0×

Here the unit conversion earns its keep, because it exposes a subtlety the single number hides. Two tannins can share a molar EC_{50} and feel nothing alike, because molar concentration counts *molecules* while the mouth, loosely, feels *mass bound to protein*. A DP-10 polymer (2,884.52 g/mol) at a given molar EC_{50} carries very nearly ten times the mass of a monomer at that same molar EC_{50}. This is why reporting astringency as grams of tannin per liter — the winemaker's habit — and reading it as loudness on a receptor-anchored axis are not the same measurement, and why the Index insists on the EC_{50} and not the gram. A number that could be achieved by a cloud of weak monomers and by a few avid polymers

alike would not be a measurement of loudness; it would be a measurement of bulk, and bulk is the thing the mouth cares about least.

The design levers

All of this collapses into one unusual fact about the axis: we do not design a molecule, we design a *distribution*. A tannin fraction is never a single structure but a spread of chain lengths, monomer ratios, and galloylation states, and the sensory character is a property of the whole spread — its mean, yes, but also its tails. The design levers are therefore statistics of a population:

- **Mean degree of polymerization.** Raising it lowers the EC_{50} and sharpens the Hill curve — more grip, arriving faster. This is the single strongest lever, and the one grape's 1,610 constructs exercise most, because it is the difference between a soft, early-drinking red and a structured one built to stand for a decade.
- **The catechin-to-epicatechin ratio.** Holding chain length fixed, this shifts the *texture* of the dryness — coarse versus fine — without much moving its size. It is how two tannins of equal intensity come to feel like different materials.
- **Galloylation.** The sharpest and most dangerous lever: a little gallate adds grip far out of proportion to the mass it adds; too much reads as harsh, green, and metallic. Tea lives at one end of this lever; a well-made red barely touches it.
- **A-type bridging.** Stiffening a fraction of the interflavan bonds changes how the polymer drapes over a PRP and can take the coarse edge off an otherwise punishing high-DP tannin.

What all four levers are really tuning is the journey a tannin takes with time. A young wine's tannin is high-DP, lightly bridged, abundant — a hard, green, mouth-scouring grip. As the wine ages, slow oxidation and rearrangement reshuffle the distribution: the average chain length drifts, the harshest fractions drop out of solution, and the grip resolves into the fine, dusty, *integrated* astringency that a decade in glass is prized for buying. To design astringency is to choose a point on that arc and hold the plant there — to build, at harvest, the tannin a lesser grape would only stumble into after ten years in a cellar. We are, in the most literal sense, engineering patience.

Precision, and its honest limits

A number is only a form of respect for the reader if it comes with its error, and astringency tests that discipline harder than any other axis. We hold batch-to-batch variation on the engineered tannin distributions to a coefficient of variation of **≤7%** — but the score a taster actually meets inherits error from more than one place, and the arithmetic is worth showing plainly.

The IFI is a ratio of two EC_{50} values, and the relative error of a ratio of independent quantities adds in quadrature:

$$CV(IFI) = \sqrt{\left(CV(EC_{50}_ref)^2 + CV(EC_{50}_candidate)^2 \right)}.$$

If the candidate is held at CV 7% and the reference is a single, heavily-averaged standard whose own variation is negligible, the score carries about 7%. If instead the reference is re-measured fresh alongside each batch at the same 7%, the two combine: $\sqrt{(0.07^2 + 0.07^2)} = 0.07\sqrt{2} \approx 9.9\%$. For Phoenix's IFI-A of 1,400 at a 7% CV, the standard deviation is 98 points, so a 95% interval runs $1{,}400 \pm 1.96 \times 98 \approx$ **1,400 ± 192**, roughly [1,208, 1,592]. Two cultivars are reliably distinguishable on this axis only when their scores differ by more than about $1.96 \times \sqrt{(98^2 + 98^2)} \approx 270$ points — so a design that moves IFI-A from 1,400 to 1,700 is a real, defensible step, while a claimed move from 1,400 to 1,500 is, honestly, inside the noise. We would rather say so than pretend the axis is sharper than it is. On an axis whose receptor is a pool of saliva, overclaiming the third digit is not rigor; it is theater.

And the failure modes are specific to a polymer and a precipitation, not the tidy failures of a small-molecule axis:

- **Oxidative drift.** Flavan-3-ols brown and couple in air; a tannin distribution left exposed will rearrange its own chain lengths and galloylation, moving the score after the plant is picked. The measurement has a clock running on it, and the honest report is a score *with a date*.
- **Haze and self-loss.** A tannin designed too avid will precipitate not only salivary PRPs but the proteins of its own matrix — fining itself out of a wine, clouding a tea, hazing a juice — so the very lever that raises the score can quietly lower the dose that survives to reach the mouth. More avid is not always more astringent in the glass.
- **Matrix masking.** pH, ethanol, sugar, and polysaccharides all change how a tannin presents to protein. The same distribution reads harder in a dry, acidic wine than in a sweet, viscous one, so a score computed in clean buffer is a starting point, never a verdict on the finished food.
- **The taster's own saliva.** Because the "receptor" is a secreted protein pool that differs between people and depletes across a meal, astringency is the axis where the instrument and the eater least reliably agree. The Index measures the stimulus faithfully; it cannot measure the particular mouth that meets it, and we do not pretend it can.

We make our peace with all of that the way we make peace with everything on this shelf — by being exact about what the number is and candid about what it is not. These are computational designs; Scoville Splice's cultivars and constructs have not been experimentally validated in a laboratory. Astringency is the one axis where we design not a molecule but a polymer's shape; where the pathway refuses to end and keeps condensing past the last enzyme on the map; and where the receptor is, in the end, the oldest sense we have. Touch, reporting that the mouth has gone dry. The body noticing a plant's most ancient defense — now offered back to it, by design, and with intention.

Chapter 14 · Zero-Bitter — *the one axis we build by taking a step away.*

Seven of the eight axes in the Index are built by addition — a pathway driven a little harder, an enzyme coaxed to turn faster, a molecule allowed to accumulate until the receptor it speaks to can no longer pretend not to hear it. Zero-Bitter is the eighth, and it is the philosophical hinge of everything we grow, because it is the only axis we engineer by making a plant produce *less*. There is no bitterness molecule we are assembling. Bitterness is already present — in the green coffee bean, in the broccoli leaf, in the white pith of a citrus — placed there by the plant itself as a chemical argument against being eaten. The axis we call Zero-Bitter is the careful quieting of that argument: closing one enzyme's off-ramp so a single bitter compound is made in smaller measure, while every other thing the plant knows how to do is left exactly as it was. The rest of the Index is the craft of opening valves. This is the craft of closing one — and knowing, to the gene, precisely which.



FIG. 14 — Coffee in cross-section. The bitterness of the cup is built into the green bean — chlorogenic acids, caffeine, the roast-born compounds layered on top. The art of attenuation is to quiet one branch of that chemistry without touching the aroma the same bean carries.

The receptor that reads danger

The Index anchors this axis to **TAS2R38**, the most thoroughly studied of the human bitter receptors — a seven-transmembrane G-protein-coupled receptor on the apical membrane of type II taste cells, reading the thiourea compounds (the famous PTC and PROP of the genetics-class taste strip) and the goitrin-type breakdown products of glucosinolates. It is worth walking the cascade in full, because every step of it is what the subtractive edit is quietly working against.

A bitter molecule arrives at the tongue and settles into the pocket formed by the receptor's seven helices — step one, a lock and key. The binding tilts the helices; on the receptor's inner face the heterotrimeric G-protein **gustducin** is released, its Gβγ subunits swinging free to switch on **phospholipase C-β2**. PLCβ2 cleaves the membrane lipid PIP_2 into **IP_3**, which crosses to the endoplasmic reticulum and opens its IP_3-receptor calcium channels. Calcium floods the cytosol; the rising calcium gates **TRPM5**, a cation channel, and the cell depolarizes. That depolarization opens **CALHM1/3** channels, the cell exhales a pulse of **ATP**, and the ATP lands on the $P2X_2/P2X_3$ receptors of the gustatory nerve fiber waiting beneath it. Only then — four amplification stages downstream of the first binding event — does the brain receive the word *bitter*.

The architecture matters to us for one reason: it is a cascade of multipliers, and multipliers are merciless about what they are given at the top. A plant that presents half as much bitter compound does not merely halve the sensation; it halves the input to a chain that was built to make small inputs loud. When we lower a food's standing on this axis, we are starving step one, and letting the plant's own amplifier do the rest of the work for us in reverse.

TAS2R38 also carries the clearest fingerprint of why bitterness is personal. The receptor comes in two common forms separated by just three amino acids — the **taster (PAV)** haplotype carrying proline, alanine, and valine at positions 49, 262, and 296, and the **non-taster (AVI)** haplotype carrying alanine, valine, and isoleucine. Three substitutions in one protein split the human population down the middle of the vegetable aisle: to a PAV taster a raw Brassica is an ordeal, and to an AVI homozygote the same leaf barely registers. We calibrate the axis to the taster, because the taster is who refuses the plant.

And one receptor can stand for so plural a sensation only because bitterness was never meant to be precise. Humans carry roughly **twenty-five TAS2R receptors**, a deliberately broad and overlapping alarm, because bitterness evolved as a single catch-all signal for *possibly toxic* — the system traded resolution for coverage, and most bitter compounds trip several receptors at once. TAS2R38 earns the anchor by being the best understood, not by being the whole story, and we will return to the honest cost of that choice.

Why one number can hold a plural sensation

Every axis in the Index resolves a whole category of sensation into a single scalar by the same move:

$$IFI(c) = 1000 \cdot EC_{50}(\text{ref}) \,/\, EC_{50}(c)$$

EC_{50} is the concentration at which a stimulus drives its receptor to half of full response — the midpoint of the dose–response curve, and the one point on that curve that is a pure property of potency rather than of dose. Invert the candidate's EC_{50} against a fixed reference's, scale by a thousand, and *loudness* becomes a score. For Zero-Bitter the reference is a defined bitter standard — the thiourea **PROP**, the canonical TAS2R38 agonist — assigned **IFI-B**

1,000 at its own half-activation. A food that half-activates the receptor at the same concentration scores 1,000; one that needs only half the concentration is twice as loud and scores 2,000; one that needs double the concentration is half as loud and scores 500. Per-axis the scorecard normalizes to a ceiling of **5,000**.

The subtlety — and the reason this axis reads differently from all the others — is that we do not alter the EC_{50} of a *molecule*. The potency of chlorogenic acid at TAS2R38 is a fixed fact of its shape; we cannot argue with it. What we alter is how much of that molecule the plant presents. So for a food, the quantity that moves is the extract's own half-activating concentration, which depends on how densely the bitter principal is packed into it.

Hold the dose–response curve in mind — the Hill form,

$$\theta = [L]^n / (EC_{50}{}^n + [L]^n)$$

where θ is the fraction of receptor activated, [L] the ligand concentration, and n the Hill coefficient. At $[L] = EC_{50}$, $\theta = ½$ by construction, whatever n is. The useful region for food is the dilute foot of the curve, $[L] \ll EC_{50}$, where $\theta \approx ([L]/EC_{50})^n$. Write a food's bitter loudness as proportional to the density φ of its dominant bitter compound, and the extract's half-activating concentration becomes inversely proportional to φ: pack in more bitter principal, and the extract reaches half-max at higher dilution — a *lower* EC_{50}(food), a *higher* IFI-B.

Now reduce φ by a knockdown fraction f, so the plant retains (1 – f) of what it made. The food's half-activating concentration rises by 1/(1 – f), and the score falls in exact proportion:

$$\text{IFI-B(after)} = (1 - f) \cdot \text{IFI-B(before)}$$

That clean linear law — **the axis score scales with the retained fraction of the bitter compound** — is the whole arithmetic of the axis in the regime where most dietary bitterness lives, and we will push and worked-example it until it earns trust, including where it breaks.

The bitter chemistries, branch by branch

Zero-Bitter as a design program means first identifying *which* compound dominates a given food's unwanted bitterness, then finding the single committed enzyme that makes it, and quieting that one step. Four families cover most of the shelf, and each has a canonical off-ramp.

BITTER PRINCIPAL	SOURCE	COMMITTED BRANCH WE QUIET	ENZYME AT THE VALVE	CANONICAL ATTENUATION	IFI-B AFTER, FROM 1,000
glucosinolates / goitrin	Brassica (broccoli, kale, cabbage)	chain elongation → aldoxime	MAM elongation · CYP79	20–65%	800 → 350
cucurbitacins	cucumber, melon	triterpene biosynthesis	*Bl / Bt* regulatory genes	20–65%	800 → 350
chlorogenic acids	coffee	quinate esterification	HQT / HCT	20–65%	800 → 350
caffeine	coffee, cacao	xanthine methylation	caffeine (xanthine) synthase	20–65%	800 → 350

The last column is the linear law doing its work. A 20% knockdown multiplies the score by 0.80; a 65% knockdown multiplies it by 0.35; so a food sitting at IFI-B 1,000 lands somewhere between **800 and 350** depending on how hard we push the valve. The canonical **20–65%** window is not a measurement range — it is the design corridor, bounded below by the smallest change a palate reliably notices and above by the most we will take before the plant starts to suffer for it.

Each branch rewards being understood in its own chemistry. The glucosinolates begin with an amino acid — methionine, for broccoli's aliphatic glucosinolates — fed through **MAM** (methylthioalkylmalate synthase) chain elongation and committed to the pathway by a **CYP79** cytochrome that turns it into an aldoxime; downstream the molecule is dressed with sulfur and sugar and stored intact until the plant's own myrosinase, released only when tissue is crushed, hydrolyzes it into the isothiocyanates and the goitrin that TAS2R38 reads. Quiet MAM and the long-chain aliphatic glucosinolates thin out; quiet CYP79 and the committed entry narrows. Coffee's harsh edge is a different trunk entirely — the **phenylpropanoid branch**, PAL → C4H → 4CL → HCT → COMT, the same ancient scaffold that runs in every vascular plant — where a side-reaction esterifies caffeoyl-CoA onto quinic acid through **HQT/HCT** to make chlorogenic acid. Caffeine is a third route, three N-methylations climbing from xanthine through theobromine to caffeine, the final step owned by **caffeine synthase**. And cucurbitacins, the triterpene bitterness that can make a cucumber inedible, are governed by the **Bl** and **Bt** master genes — "bitter leaf" and "bitter fruit" — that switch the whole triterpene branch on or down.



FIG. 15 — Broccoli in cross-section. The glucosinolates that make raw Brassica bracing are stored intact in the tissue and only become their bitter, pungent breakdown products when the leaf is cut. Attenuation reaches upstream of that, at the committed step, long before the molecule is ever made.

The gentlest tool in the kit

Our usual instrument on this axis is **RNA interference**, and it is the only edit in our whole practice whose entire purpose is to make a plant quieter than it was. We transcribe a short hairpin RNA inside the plant — a sequence that folds back on itself into a double strand. The cell's **Dicer** enzyme reads that double strand as the signature of something foreign and slices it into small interfering RNAs, twenty-one or so nucleotides each. One strand of each is loaded into the **RISC** complex, where **Argonaute** carries it like a search template through the cytoplasm, and wherever it finds a messenger RNA that matches — the transcript of the exact enzyme we chose — it cuts, and that message is destroyed before a single protein is translated from it.

The reason we reach for interference rather than a knockout is the whole ethic of the axis, and the distinction matters more than it looks. A knockout removes a gene; RNAi *dims* it, and dims it by a fraction we can tune. We set that fraction by how strong a promoter drives the hairpin and how many copies of the trigger we install, and we read the result not at the gene but at the metabolite — the actual milligrams of glucosinolate or chlorogenic acid left in the tissue. The compounds we are quieting are not junk. The plant makes them for defense, and many are braided into the plant's health and into the very pathways that give the food its other virtues. A clumsy total knockout can leave a Brassica defenseless, or strip a coffee of the aroma precursors that share a branch with its bitter ones. So the practice is attenuation, not abolition — a measured step down, not a door slammed shut. Where an even softer hand is wanted, we skip the hairpin entirely and simply swap in a weakened promoter on the committed enzyme, lowering its transcription to a set fraction of wild type. Either way, the goal is the same: lift the bitterness off the palate while leaving the plant armed and the rest of the flavor whole.

The arithmetic of taking away

Let the linear law earn its keep on a worked case. Take a broccoli and make the arithmetic concrete with an illustrative anchor: suppose the PROP reference half-activates TAS2R38 at EC_{50}(ref) = 2.0 µM, a round number chosen only to carry the calculation. Say a

raw, assertive broccoli extract half-activates at an effective EC_{50}(food) of 1.0 µM-equivalent — twice as potent as the reference. Then

```
IFI-B = 1000 · 2.0 / 1.0 = 2,000
```

a bracingly bitter plant. Apply a 50% knockdown of the glucosinolate branch (f = 0.50): the retained fraction is 0.50, the half-activating concentration doubles to 2.0 µM-equivalent, and

```
IFI-B = 1000 · 2.0 / 2.0 = 1,000
```

exactly the (1 – f) law predicts, 2,000 × 0.50 = 1,000. Push to the top of the corridor, a 65% knockdown (f = 0.65): the concentration rises to 1.0 / 0.35 = 2.857 µM-equivalent, and 1000 · 2.0 / 2.857 = **700**, which is again 2,000 × 0.35 = 700. Two routes, one answer — the kind of agreement that tells you the model is at least self-consistent.

RNAi knockdown f	Retained glucosinolate	EC_{50}(food), mM-eq.	IFI-B	On the palate
0% (wild type)	100%	1.00	2,000	bracing, often refused raw
20%	80%	1.25	1,600	assertive, still green
35%	65%	1.54	1,300	rounded, approachable
50%	50%	2.00	1,000	gentle, the reference level
65%	35%	2.86	700	soft; the deepest step we take

A unit conversion, because milligrams are what a grower actually weighs. Suppose that wild-type broccoli carries 12 µmol of glucoraphanin, its major glucosinolate, per gram of dry tissue. Glucoraphanin's molar mass is ≈ 436 g/mol, so

```
12 µmol/g × 436 µg/µmol = 5,232 µg/g = 523 mg per 100 g dry weight.
```

A 50% knockdown brings that to ≈ 262 mg/100 g, and the 65% step to ≈ 183 mg/100 g — and it is these predicted metabolite masses, not the transcript estimate alone, that our model uses as the measure of true attenuation, because the plant has ways of arguing with a knockdown that the mRNA count would hide.

One honest caveat on the dose–response, and it runs in the grower's favor. The linear law assumes the dilute regime, n ≈ 1, where perceived loudness tracks concentration one for one. But bitter receptors often read cooperatively, n > 1, and a very bitter food sits further up the steep shoulder of the Hill curve. There, a 50% cut in concentration can drop the perceived bitterness by more than half — for n = 2, roughly 75%. So the IFI-B our arithmetic predicts is a conservative floor on the relief, not a ceiling. We would rather promise the floor and let the plant over-deliver.

The shelf

More than three thousand constructs across the library touch this one axis, and the organisms that carry it read like a tour of every reason people put a nourishing plant back down.

ORGANISM	CONSTRUCTS	AXES	DIFFICULTY	DOMINANT BITTER PRINCIPAL
Grape	1,610	B / A / M	Hard	flavan-3-ols, condensed tannins
Broccoli	380	B / M	Moderate	glucosinolates → goitrin
Broccoli-Kale	280	B	Hard	glucosinolates
Tea	280	U / B / A	Hard	catechins, caffeine
Orange	120	B / C	Hard	limonin, naringin
Potato	80	B	Easy	steroidal glycoalkaloids
Hops	80	B	Moderate	humulones (α-acids)
Coffee	40	B	Hard	chlorogenic acids, caffeine
Cabbage	40	B	Moderate	glucosinolates (sinigrin)
Kale	40	B	Moderate	glucosinolates
Radish	40	B	Hard	glucosinolate isothiocyanates
Rapeseed	40	B	Hard	progoitrin

Broccoli carries the deepest bench at **380 constructs**, the broccoli-kale cross another **280**, and grape leads the count at **1,610** — though grape's bitterness is so braided with its tannin that the work there is as much astringency as bitter, which is why it carries three axes. The difficulty column tells its own story: potato is marked **Easy** because its bitter glycoalkaloids sit on a short, well-mapped branch, while coffee is **Hard** — only forty constructs — because its chlorogenic acids ride the shared phenylpropanoid trunk, and separating the harsh from the aromatic on a shared trunk is the most delicate cutting we do.

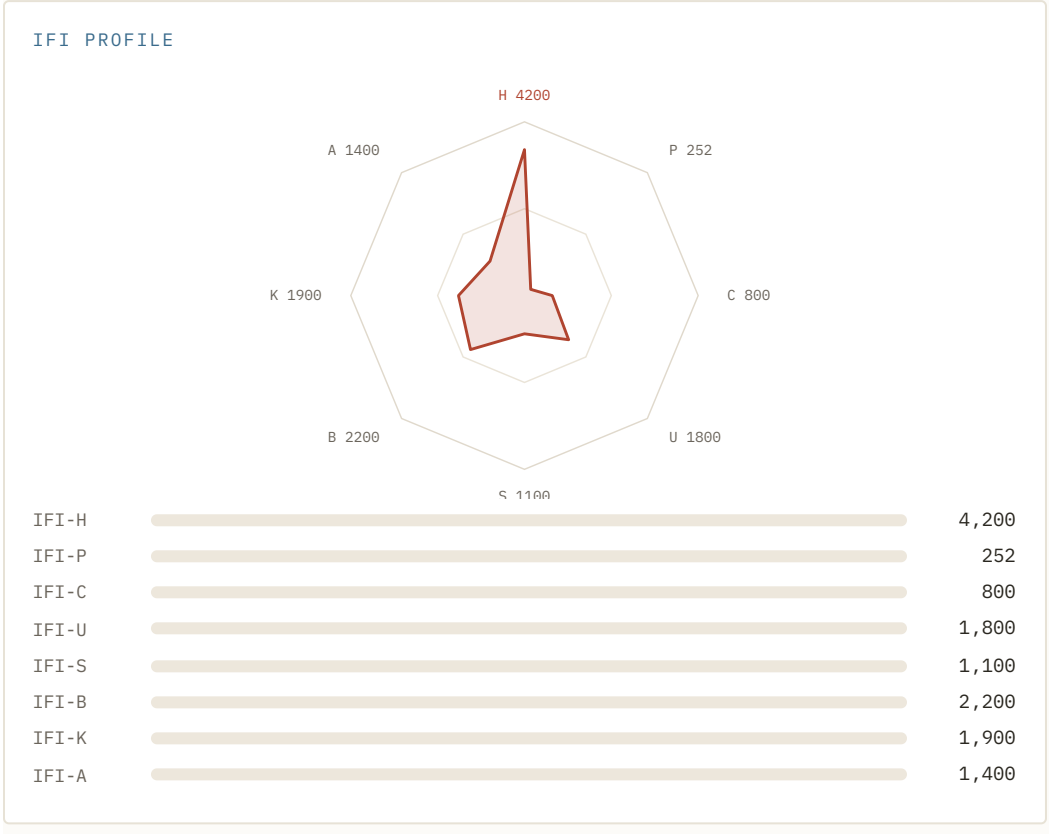
Coffee and cacao deserve their own word, because both are foods people already love *through* their bitterness, not in spite of it. The task there is never to erase — a chlorogenic-stripped coffee would taste hollow, a theobromine-free cacao would stop being chocolate — but to take only the harsh, puckering fraction and leave the structured bitterness the drink is built on. Cacao's bitterness is twofold: the methylxanthine theobromine, cousin to coffee's caffeine on the same methylation branch, and the flavan-3-ol polyphenols that also carry its astringency. We quiet the first a measured notch and leave the second nearly whole, because in chocolate the polyphenol grip is structure, not fault.



FIG. 16 — Cacao, split to the bean. Chocolate's bitterness is two chemistries braided together — the methylxanthine theobromine and the flavan-3-ol polyphenols — and the craft is to quiet only the harsh fraction while leaving the structured bitterness the whole experience is built on.

When a bitterness is kept on purpose

The axis is not a one-way street, and the proof of that is our most complex plant. Our flagship full-spectrum cultivar, the one designed gene by gene to register on all eight axes at once, does not read zero on the bitter axis. It reads **IFI-B 2,200** — a deliberate, structural bitterness dialed to a precise setting and held there.



Read the profile and the 2,200 explains itself. A plant screaming heat at 4,200 with sweetness at 1,100 and kokumi at 1,900 would collapse into a sugary blur without something to stand against the sweetness and give the whole structure a spine. A moderate, intentional bitterness is that spine. So on this one cultivar the subtractive tool runs in reverse — not an RNAi knockdown of a bitter branch, but a measured *retention*, tuned to the same receptor by the same math. It is the clearest demonstration we have that Zero-Bitter is not an anti-bitterness crusade; it is a dial, and most of the time we turn it down, and now and then, for a plant that can carry it, we set it exactly where a bitterness belongs.

Error, reproducibility, and the honest limits

A number is only as useful as the confidence interval around it. The library's defining figure here is a batch-to-batch coefficient of variation of **≤7%**, against **42%** for the wild-type plants we started from — and that gap is the difference between an axis and a guess.

Run the propagation. An IFI-B score is a ratio of two EC_{50} values, and for a ratio the relative errors add in quadrature:

```
CV(IFI-B) ≈ √( CV(ref)² + CV(food)² )
```

With both the reference and an engineered food held at 7%, the score inherits $\sqrt{(0.07^2 + 0.07^2)} \approx$ **9.9%** — a 95% interval of roughly ±20%, tight enough to rank two cultivars and trust the order. Feed the same arithmetic the wild-type's 42% and it returns $\sqrt{(0.42^2 + 0.07^2)} \approx$ **42.6%**, a 95% interval near ±85%. A wild broccoli's bitterness is so weather-and-field-dependent that its IFI-B is barely a number at all; two plants off the same row can differ by more than

our entire design corridor. The ≤7% design specification — clonal lines, controlled environment, characterized metabolite profiles — is what turns a sensation into a scale you can engineer against.

And the limits, named plainly, because the axis is only honest if we say where it is thin.

The plant must stay armed. The glucosinolates and glycoalkaloids we quiet are the plant's chemical defense; the potato's bitter alkaloids are also, above a threshold, its toxins. Push the knockdown toward a knockout and the plant loses its argument against pests and pathogens. The 20–65% ceiling is not timidity — it is the line below which the plant can still defend itself, and we hold it.

Shared trunks punish a heavy hand. Coffee's chlorogenic acids share the phenylpropanoid branch with its aroma and antioxidant chemistry; cut the committed node too deep and the cup loses its character along with its harshness. Specificity — which gene, in which tissue, at which stage — is the whole craft, which is why coffee carries only forty constructs and every one of them is cautious.

The plant can argue back. Knock down one committed enzyme and a plant may reroute flux through a paralog or upregulate the step you dimmed. The realized attenuation then falls short of the transcript knockdown, which is exactly why we track the metabolite and never trust the mRNA count alone.

Interference is not perfectly aimed. A hairpin's small RNAs can, in principle, silence an unintended transcript that shares enough sequence. We design against it and screen the transcriptome for it, and it remains a limit we respect rather than one we have abolished.

Penetrance is uneven — sometimes usefully. RNAi strength varies by tissue and age, so a fruit can de-bitter while the leaf stays armed. Often that is precisely the outcome we want: palatability where you eat, defense where the pest bites.

One receptor is not all of bitterness. TAS2R38 is the anchor, not the whole sensor array. A compound that trips several of the twenty-five bitter receptors may leave a residual bitterness read by a receptor the anchor score never counted, so IFI-B is an anchor-receptor measurement, honest about what it is and silent about the rest. And for the AVI non-taster, who barely reads TAS2R38 to begin with, the gain from our edit is smaller — the axis is calibrated to the taster, the diner who actually refuses the plant.

The axis built by taking a step away

Here the nourishment thread that runs under everything we grow speaks most plainly, so we will say it once, carefully, and make no claim of it. A great deal of what is most protective and valuable in vegetables — the glucosinolates of Brassica among them — is also what is bitter, and bitterness is the single most common reason people, and children above all, set down the plants that would feed them best. To attenuate bitterness *selectively* — to quiet only the fraction that drives refusal while leaving the plant's integrity whole

— is, quietly, to make a nourishing food easier to love. We do not claim a health benefit; we are not permitted to, and we would not. These are computational designs; Scoville Splice's cultivars and constructs have not been experimentally validated in a laboratory. We note only this: the one subtractive axis in the Index is the one most plainly in service of the idea beneath all our work — that food should be something the body is glad to receive. Zero-Bitter is the axis we build by taking a single step away. And it is the axis where taking that step away is, itself, a kind of care.



PART IV OF IV

Composing Pathways

flux balance · stacking branches · the transferable route · the open record — from one pathway to a whole plant, and a whole roadmap

Chapter 15 · Flux Balance as a Design Language — *the arithmetic of a shared precursor.*

It is natural to speak of the eight axes as if each ran alone — heat on its own supply, sweetness in its own quiet corner of the cell, umami drawing on a reservoir no one else touches. That is a teaching fiction, generous and clarifying, and the moment we design a real plant we have to retire it. In a living pepper the branches share a cell. They share precursors, share the finite carbon and nitrogen and reducing power the plant has to spend in a day of sunlight, and the instant you try to run two of them hard at once they begin to compete. The fruit's heat and the leaf's lignin and the seed's oil all reach back, sooner or later, to the same small pool of starting materials. Designing one pathway in isolation is easy and almost always wrong. Designing *a plant* means balancing the books — and the language we balance them in is flux-balance analysis.



FIG. 15.1 — Research trays at the stacking stage. Each tray is a different partition of the same carbon budget; the plants that win are not the ones that make the most of any single molecule but the ones that make enough of several without the arithmetic collapsing.

We came to flux balance the way a family comes to a household ledger: not because we loved accounting, but because we loved the thing the accounting protected. A plant is a budget with roots. Every gram of sugar the leaves fix in the light is spent — on walls, on membranes, on defense, on the sweetness and savor and heat we are asking it to carry, and, we never forget, on the nourishment the fruit is also meant to hold. Flux-balance analysis is how we read that ledger honestly, before we ask the plant for anything it cannot give.

The matrix that accounts for everything

Flux-balance analysis, FBA, is deceptively simple in its premise and long in its reach. Write down every metabolic reaction in the cell — thousands of them, in a genome-scale model — as a great matrix of stoichiometries, **S**. Each row is a metabolite; each column is a reaction; the number where they meet is how many molecules of that metabolite that reaction consumes (negative) or produces (positive). Then make one assumption, the assumption that organizes everything: the cell is at steady state. No intermediate is piling up without limit, none is draining to nothing. In the language of the matrix that is a single line,

> $\mathbf{S} \cdot \mathbf{v} = \mathbf{0}$,

where **v** is the vector of every reaction's flux — the rate each reaction runs. That one equation says the bookkeeping balances: for every internal metabolite, what comes in equals what goes out.

Read it as an accountant would and it is a statement about a shared budget. The matrix **S** has thousands of rows and thousands of columns; the equation $\mathbf{S} \cdot \mathbf{v} = \mathbf{0}$ does not pin down a single answer but carves out a whole space of flux vectors the plant *could* run. Add the bounds that physics and biology impose — which reactions are irreversible and so cannot run backward, how fast each enzyme can turn over at most — and that space tightens into a convex polytope: a many-sided, many-dimensional solid whose every interior point is a metabolic state the plant could occupy without breaking a conservation law. Choosing an objective, *maximize flux to capsaicin*, say, or *to free glutamate*, and solving for it is sliding to one corner of that solid. The model does not guess. It *accounts*, across every reaction at once, and in doing so it answers two questions no amount of staring at a plant will yield: where is the true bottleneck, and what does relieving it cost everywhere else.

Stated as the optimization it is, FBA is a **linear program** — the whole of it fits in three lines:

> maximize $\mathbf{c}^{\mathsf{T}}\mathbf{v}$ (the flux we want, picked out by the vector **c**)
> subject to $\mathbf{S} \cdot \mathbf{v} = \mathbf{0}$ (mass balance) and $\mathbf{v}_{lb} \le \mathbf{v} \le \mathbf{v}_{ub}$ (thermodynamics and enzyme capacity)

Knowing its shape keeps us honest about what it can and cannot say. A genome-scale model may carry two thousand reactions and fifteen hundred metabolites; the mass-balance rows pin down only as many fluxes as the matrix has independent constraints, which leaves the rest — the dimension of the null space, **n − rank(S)**, often several hundred — free to vary. That is not a flaw. It is the design space. Those free directions are precisely the choices a breeder was making blindly for ten thousand years and that we now make with our eyes open: how much carbon to send down each branch, and what the plant gives up at the far end of each decision.

What a precursor is worth: the dual ledger

Every linear program has a shadow — a second program, the *dual*, that prices the first one's constraints. This is where flux balance stops being a checking step and becomes a way of thinking. Solve the primal and you learn the best flux distribution. Solve the dual, or simply read the shadow prices the solver hands back for free, and you learn the *marginal value of every constraint*: how much more capsaicin you would gain for one more unit of a precursor, and how much a competing demand is charging you for holding its floor. The shadow price of phenylalanine is not a metaphor. It is a number, in units of product per unit precursor, and it tells you exactly what the next molecule of that shared pool is worth to the flavor you are chasing.

This reframes the whole enterprise. We are not, it turns out, in the business of maximizing molecules. We are in the business of *partitioning a budget*, and the shared precursors this book keeps circling back to are the line items where the partition is decided.

SHARED PRECURSOR	FLAVOR IT FEEDS	WHAT IT ALSO OWES	WHY IT BINDS
phenylalanine → *p*-coumaroyl-CoA	heat (vanillyl head), astringency	lignin (structure), flavonoids (defense)	lignin is load-bearing; the plant must stand
GGPP (diterpene)	sweetness (steviol glycosides)	gibberellin (growth hormone)	flavor on one side, growth on the other
glutamate	umami, kokumi	the whole nitrogen economy	the plant's central nitrogen hub
branched-chain amino acids	heat (acyl tail)	seed and membrane lipids	tail carbon is also oil carbon

Each row is a precursor with more than one creditor. *p*-Coumaroyl-CoA stands at the junction of the phenylpropanoid trunk and feeds, from one pool, three demands: lignin, flavonoids, and the vanillyl head of capsaicin — and lignin is what lets the plant hold itself upright, so it cannot be starved. The diterpene GGPP feeds the steviol-glycoside sweetness of stevia on one side and the gibberellin the plant grows by on the other; take too much and the plant is sweet and small. Glutamate is at once the umami molecule itself and the central clearinghouse of the plant's nitrogen, so every point of savory depth is argued against the nitrogen the plant needs to build protein. Push any one branch toward a louder flavor and the dual shows you the bill: softer tissue, slower growth, a competing flavor dimmed, sometimes a toxic intermediate pooling at a downstream step that cannot keep pace. The reverent sentence and the mathematical one are the same sentence. You cannot take everything from a plant and still have grown it well.

A worked ledger: the arithmetic of one shared precursor

Abstractions earn their keep only when you run the numbers, so let us run them on a deliberately small model — a toy, honestly labeled, with round units of flux rather than canonical measurements — and watch the shared precursor do what it does in the full genome-scale version.

Take one branch point: *p*-coumaroyl-CoA, which we will call **X**, supplied by the upstream trunk at a rate of **10** units. Two creditors draw on it: lignin, at flux **v_L**, and the vanillyl head of capsaicin, at

flux **v_H**, each consuming one unit of X per unit of product. Mass balance at X — the steady-state line, applied to a single node — reads

```
10 − v_L − v_H = 0, so v_L + v_H = 10.
```

Now the biology enters as a bound. The plant cannot starve its walls; structural integrity demands at least **v_L ≥ 6**. Our objective is to maximize the head, **max v_H**. The linear program is trivial to solve by hand: push v_L to its floor of 6, and v_H takes the remainder, **v_H = 4**. The dual is just as plain. One more unit of supply X, with lignin pinned at its floor, flows entirely to the head — the shadow price of precursor supply is **+1** unit of head per unit of X. And the lignin floor has a price too: raise the structural requirement from 6 to 7 and the head falls from 4 to 3, a marginal cost of **−1** unit of heat for every unit of wall. That is the arithmetic of a shared precursor, stated in full. Nothing in flavor design is louder than it; every gain is quoted against what the pool also owes.

The confluence deepens the lesson, because capsaicin is not built from one pool but from two that must meet. The vanillyl head is phenylpropanoid, drawn from X. The acyl tail is a branched-chain fatty acid, drawn from a *separate* pool — call it **Y**, the branched-chain amino acids — supplied at **7** units and also owed to membrane and seed lipids, with a lipid floor of **v_lip ≥ 2**, leaving the tail at most **v_T ≤ 5**. Capsaicin synthase, Pun1, joins one head to one tail, so the capsaicin flux is bounded by what each line can deliver and what the enzyme can process:

```
v_cap = min( v_H, v_T, V_max(Pun1) ) = min( 4, 5, V_max ).
```

In this toy model, with V_max(Pun1) = 3, the enzyme appears to be the slow step: v_cap = min(4, 5, 3) = **3** — bound on the enzyme, with the precursor pools sitting with positive slack. Every breeder who chased heat by pouring more carbon into the supply was adding to pools whose shadow prices were already zero in this simplified picture. The *correct* answer, however, requires the genome-scale picture. Our full Metabolic Control Analysis predicts that control sits upstream in the phenylpropanoid pathway — at 4-coumaroyl-CoA ligase (4CL) and hydroxycinnamoyl transferase (HCT/CCoAOMT) — not at the terminal Pun1/AT3 step. Across 5,000 Monte Carlo draws, Pun1's flux control coefficient is effectively zero; upstream enzymes carry 93–99% of the control. The budget — the shared phenylpropanoid supply governed by those upstream nodes — is where the rate is set, and the budget is where reverence and arithmetic finally meet.

The confluence, and designing the terminal synthase

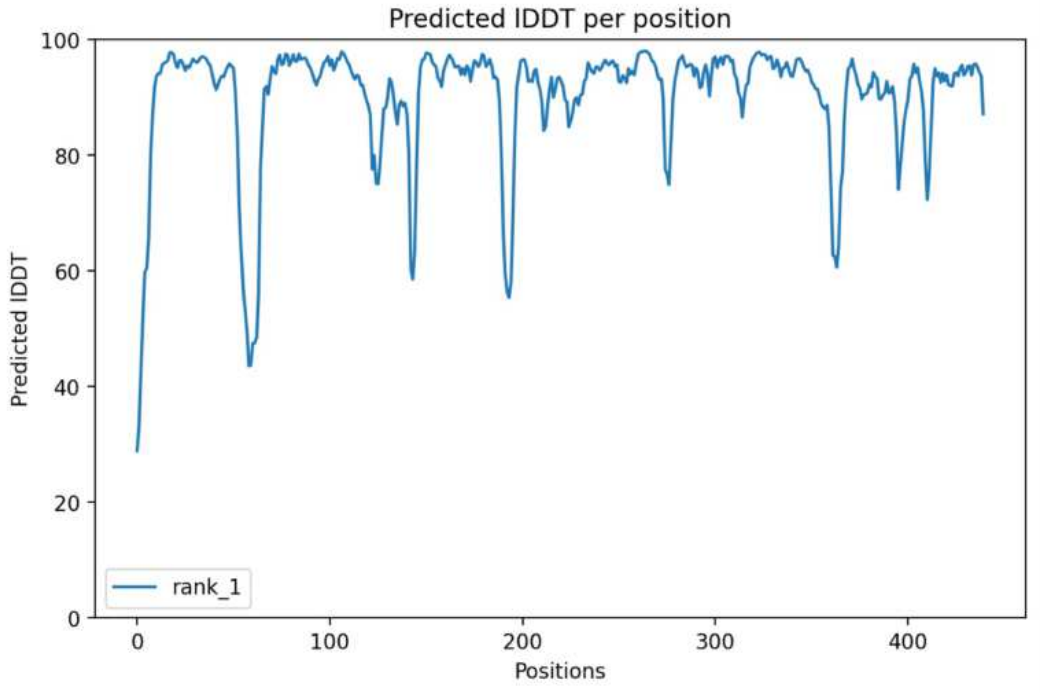


FIG. 15.2 — Pun1/AT3 (capsaicin synthase), the confluence enzyme, coloured by pLDDT model confidence; the binding domain and the active-site cleft are high-confidence, and it is here that two substitutions the breeders never found were written in.

Flux balance and protein engineering converge on the same enzyme from opposite directions. Our genome-scale Metabolic Control Analysis predicts that flux control in the capsaicinoid pathway sits at the upstream phenylpropanoid steps — 4CL and HCT/CCoAOMT carry 93–99% of the control across all Monte Carlo draws — but Pun1/AT3 is the terminal synthase, the enzyme that assembles capsaicin from the vanillyl head and the acyl tail the upstream pathway delivers. Pun1 labels the product; it does not govern the rate at which the upstream pathway supplies precursors. What it does govern is how efficiently and reliably the available substrate pool is converted, and that is precisely where our protein engineering focused. A 650-million-parameter protein language model, ESM2, read the Pun1 sequence the way it had read tens of millions of others and scored every possible single-residue substitution by its predicted effect on **catalytic efficiency × fold stability** — the product of how fast the enzyme turns substrate over and how reliably it folds into the shape that can. Two substitutions, never before combined in any pepper that ever grew, emerged as the double variant design.

EDIT	MECHANISM	ACTS ON	ΔΔG (KCAL/MOL)	STATUS
S39L	tighter substrate binding, N-terminus	K_m (affinity)	−2.1	shipped
L345G	steric relief in the active-site cleft	k_cat (turnover)	−1.8	shipped
C175S	prevents a disulfide misfold	folded fraction	−0.9	structural annotation — not in shipped construct

Read those free energies as the physics they are, and the enzyme's story becomes quantitative. Catalytic rate is Michaelis–Menten, **v = V_max · [S] / (K_m + [S])**, where **V_max = k_cat · [E]_active**. Each shipped edit pushes a different term. S39L tightens how the substrate binds; a binding free-energy change of −2.1 kcal/mol lowers the dissociation constant by a Boltzmann factor, and at a tissue temperature of 25 °C — where **RT = 0.592 kcal/mol** — that factor is exp(2.1 / 0.592) ≈ **35×** tighter binding, which drops K_m by

roughly the same and lifts v sharply at the sub-saturating substrate levels a real cell actually runs at. L345G opens the cleft and speeds turnover, raising k_cat.

The two shipped edits do not merely add; their energies sum and their factors multiply. A combined stabilization of −3.9 kcal/mol corresponds to exp(3.9 / 0.592) ≈ **730×** in the relevant population ratio — the difference between an enzyme that folds and binds and turns over as a near-afterthought and one that wild-type peppers only ever approached by accident, in their best fruit, in their best year. This is why our model predicts peak output of approximately **2.94 million SHU** for Phoenix, alongside a design target of **≤ 7% batch-to-batch coefficient of variation**, against **42%** in wild-type: an enzyme that folds reliably does not just make more product, it makes the *same* amount of product twice in a row. Reliability is a stability problem before it is a manufacturing one, and the ΔΔG column is where the reliability was bought.

From enzyme capacity to a number on the tongue

Flux to capsaicin is a rate inside a cell; a flavor is a loudness on a receptor. The bridge between them is the Index, and the Index is a stated convention — defined, like the original Scoville scale, by a chosen reference constant, not derived from literature averages:

> **IFI(candidate) = 1,000 × EC_{50}(reference) ÷ EC_{50}(candidate).**

EC_{50} is the concentration at which a stimulus drives its receptor — for heat, the TRPV1 channel on the trigeminal endings of the tongue — to half of its full response. A *louder* molecule reaches half-maximum at a *lower* concentration, so loudness lives in the inverse of EC_{50}; invert the candidate against a fixed reference constant, scale by a thousand, and the score for that axis is calibrated. Each axis has its own independently defined reference constant, anchoring IFI = **1,000** by convention on that axis alone; scores on different axes are not directly comparable, the same way you cannot subtract two temperatures recorded in different scales without a conversion. Rearrange the formula within one axis and a beautiful simplicity falls out:

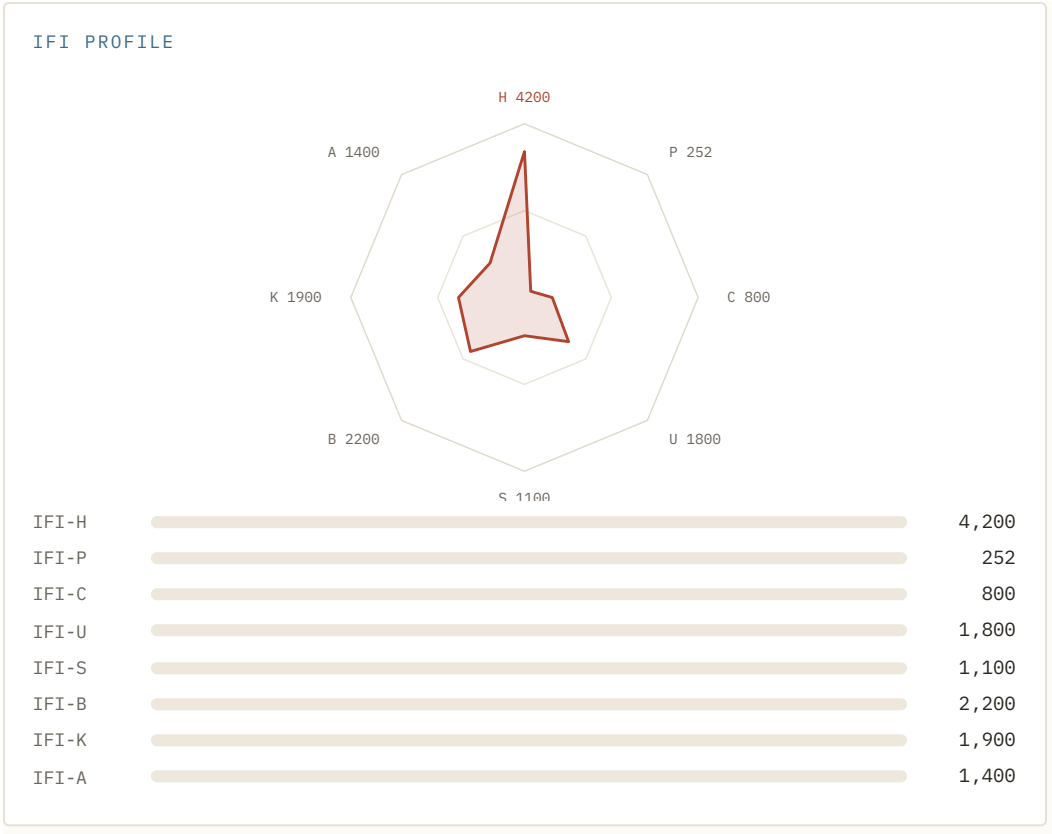
> EC_{50}(reference) ÷ EC_{50}(candidate) = IFI ÷ 1,000 = fold-loudness.

An IFI-H of 4,200 means the molecule speaks **4.2×** louder than the heat reference — activates the channel at 1/4.2 of the reference concentration. The whole flagship tier reads this way at a glance, and because each cultivar also carries a predicted SHU value, the tier table lets us watch design scope turn into predicted heat, rung by rung.

CULTIVAR	SKU	PREDICTED SHU	IFI-H	FOLD-LOUDNESS	AXES CARRIED
Ember	SS-0087	866K	1,570	1.57×	2/8
Cinder	SS-0088	1.05M	2,290	2.29×	3/8
Kindle	SS-0089	1.25M	2,000	2.00×	3/8
Flare	SS-0090	1.48M	—	—	4/8
Blaze	SS-0091	1.72M	2,630	2.63×	4/8
Wildfire	SS-0092	1.96M	2,930	2.93×	5/8
Scorch	SS-0093	2.21M	3,220	3.22×	5/8
Firestorm	SS-0094	2.47M	3,490	3.49×	6/8
Solara	SS-0095	2.72M	3,980	3.98×	6/8
Phoenix	**SS-0096**	**2.95M**	**4,200**	**4.2×**	**8/8**

To make the inverse concrete, calibrate the reference line at EC_{50}(reference) = 0.70 µM — the anchor the heat axis is normalized against — and the candidate concentrations follow directly from EC_{50}(candidate) = 1,000 × EC_{50}(reference) ÷ IFI. Phoenix at IFI-H 4,200 half-activates TRPV1 at 0.70 ÷ 4.2 = **0.167 µM**, which is 167 nanomolar, or 1.67×10^{-7} mol/L. Ember at 1,570 half-activates at 0.70 ÷ 1.57 = **0.446 µM** = 446 nM. The reference itself sits at 0.70 µM = 700 nM. The dose–response that these anchor is the Hill curve, $\mathbf{R([A]) = R_max \cdot [A]^n \div (EC_{50}^n + [A]^n)}$, with capsaicin's cooperativity at TRPV1 giving a Hill coefficient near **n = 2**. A lower EC_{50} slides the entire curve leftward along the concentration axis without changing its shape: to reach nine-tenths of maximal fire you need [A] = EC_{50} · (0.9/0.1)^(1/2) = 3 × EC_{50}, which for Phoenix is 0.5 µM and for the reference is 2.1 µM. The same sensation of near-full heat arrives at a quarter of the molecule.

And the Scoville number, the old currency, converts cleanly into this new one. Pure capsaicin registers 16,000,000 SHU, so any sample's SHU divided by that ceiling is the fraction of pure-capsaicin pungency it carries. Phoenix's predicted 2.94 million SHU gives roughly 2,940,000 ÷ 16,000,000 = **0.184** — a designed fruit speaking at 18.4% of the loudest heat chemistry knows, a figure no pepper has reached on its own. Here is the full eight-axis voice of that plant, the one cultivar on the tier whose carbon budget was designed to carry every axis at once:



Read down that block and you are reading a solved flux-balance problem. Each score sits against its own defined reference convention — the axes are independent scales, not a single common ruler. No axis is maxed — heat at 4,200, kokumi at 1,900, astringency at 1,400 — because maxing any one would have starved the others and starved the plant. Eight voices held in balance is harder than one voice shouting, and it is the only kind of design worth the name.

Error, and the arithmetic of trust

A number you cannot reproduce is not a measurement; it is an anecdote. So the last piece of the ledger is the error on every entry, and here too the arithmetic is exact. The coefficient of variation, **CV = σ / μ**, is the honest width of a batch. For Phoenix centered on IFI-H 4,200, a 7% CV means a standard deviation of σ = 0.07 × 4,200 = **294 index points**; the wild-type 42% would scatter the same mean with σ = 0.42 × 4,200 = **1,764 points**. In Scoville terms the gap is starker. The predicted SHU values across the flagship design tiers span about 230,000 units between adjacent cultivars. At a design specification of 7% CV for a mid-tier cultivar such as Blaze (predicted 1.72M SHU), σ ≈ 120,000 SHU — half a rung, so each tier stays resolvable, nameable, buyable. The wild-type 42% gives σ ≈ 722,000 SHU — more than three rungs — and the entire tier dissolves into a blur in which Ember and Firestorm cannot be told apart. Consistency is not a luxury laid over the design; it is what makes the tier a tier.

There is a floor beneath this, and it is worth naming because it tells you how far reproducibility can go. The Index is a ratio of two EC_{50} measurements, and errors in a ratio add in quadrature: **$CV(IFI) = \sqrt{(CV(reference)^2 + CV(candidate)^2)}$**. Two careful dose–response assays, each good to about 5%, propagate to $\sqrt{(0.05^2 + 0.05^2)} = 0.0707 \approx$ **7.1%**. That the engineered line's design specification targets batch variation of ≤ 7% — essentially the same figure that falls out of the *measurement* noise of two clean assays — means the reproducibility problem has been designed out as far as the instrument can see it. Beyond that point you are no longer measuring

the plant's inconsistency; you are measuring your own ruler. The ΔΔG that stabilized the fold and the CV that narrowed the batch are two readings of one achievement: an enzyme that does the same thing today that it did yesterday.

When the model must be allowed to say "I don't know"

The surest way to misuse flux balance is to trust it past the edge of what it knows, so we run it hard against that edge on purpose. **Flux variability analysis** asks, for each reaction, the entire range of flux it could carry while the network still hits its optimum — minimize and maximize that one flux subject to the objective staying within a hair of its best value. A reaction whose range comes back narrow is one the model has genuinely pinned down. A reaction whose range comes back wide is one the optimum does not care about: the solver could send flux high or low with no cost to the objective, which means the single number it first reported was an arbitrary choice among many, and we must distrust it and go measure. We publish the ranges, not just the point estimates, for exactly this reason. An honest model reports its own ignorance as a width.

And flux balance is blind in ways no amount of care inside the method repairs, which is why it is a design language and never the designer. It assumes steady state, so it cannot speak to the hours after a cold night or a first ripening when nothing is at steady state. It knows stoichiometry but not kinetics or regulation: it will happily route flux through an enzyme that is, in the living cell, switched off by a signal the matrix never encoded, or allosterically throttled by the very product we are trying to accumulate. It optimizes an objective we chose, and a plant does not optimize our objectives — it optimizes surviving and setting seed, and where those diverge from maximal flavor the model's clean corner can sit a long way from the plant's actual life. We treat every FBA prediction as a hypothesis with a shadow price attached, and the trays, the chambers, the slow seasons in the field are where the hypothesis is allowed to be wrong. The model tells us which of our guesses are even arithmetically possible — a smaller claim than it sounds, and a far more useful one, because it spares us a thousand experiments that could never have worked and points us at the few that might.

Partitioning a budget



FIG. 15.3 — A research plot at scale, where every prediction the ledger makes is finally argued against weather, soil, and time. A design that balanced the books in the model but not in the plant is no design at all.

So flux balance is not a step we run at the end to check our work; it is the language we design in from the first line, and it changes what the work even is. We are not maximizing molecules. We are partitioning a budget — deciding how much of a shared, finite pool to spend on a flavor, and how much to leave for the plant to be a healthy plant: to stand and grow and set seed, and, not least, to carry the nourishment the fruit is also meant to hold. There is a constraint for that in the model too, a floor we never let the optimizer breach, the same way a lignin floor keeps the stem upright — a reservation of carbon and nitrogen for the vitamins and minerals and protein that are, in the end, the whole reason a family grows food at all. A pepper engineered to shout one axis and give nothing back to the body that eats it would solve the wrong linear program. We have spent our lives choosing to eat from plants; we are not about to design one that has been emptied to win a number.

These are computational designs; Scoville Splice's cultivars and constructs have not been experimentally validated in a laboratory.

The reverent version of this and the mathematical version are one sentence, and flux balance is how we find where it falls. You cannot take everything from a plant and still have grown it well. The dual prices every demand, the shadow price names what each gain truly costs, and the discipline of respecting those prices is what lets a design that balanced the books on the screen go on balancing them in the soil. We find the line in the arithmetic. We honor the line in the field. And the plant, asked for exactly what it can give and no more, gives it back season after season — which is the only kind of abundance that was ever real.

Chapter 16 · Eight Branches, One Plant — *stacking routes on one carbon budget.*

There is a plant in which every route drawn in this book runs at once. Its name is Phoenix, catalogue number SS-0096, and it is the only cultivar in the library to register on all eight sensory axes in a single fruit — the full spectrum, the thing the whole method was quietly always pointed at. It is simple enough to draw one pathway in isolation: a clean line from gene to enzyme to molecule to receptor, each arrow a reaction you can name. It is a far harder thing to run eight of those lines *through the same tissue, out of the same carbon, at the same time*, and to let no single branch rob what the other seven need to speak. That harder thing is what Phoenix is. If a pathway diagram is a sentence, Phoenix is the paragraph that proves the grammar holds when you try to say everything at once.



A backlit greenhouse at golden hour, a single engineered fruit hanging in the light — one plant carrying every route in the book.

We did not come to this plant by asking *how hot*. We came to it by asking, on a dawn walk our family still talks about, a quieter and much harder question: *how much, together, and still grown well?* That question is an engineering constraint before it is a poem, and this chapter is the arithmetic of answering it.

The trunk every branch is cut from

Start where the carbon starts. Almost every flavor molecule on the aromatic side of this plant descends from a single amino acid, phenylalanine, drawn in turn from the shikimate pathway that every green cell keeps running. Phenylalanine is the trunk. The first cut into it is made by **phenylalanine ammonia-lyase (PAL)**, which strips an amine and leaves *trans*-cinnamic acid. A cytochrome-P450, **cinnamate 4-hydroxylase (C4H)**, hangs a hydroxyl on the ring to give *p*-coumaric acid. **4-coumarate:CoA ligase (4CL)** then spends a molecule of ATP to activate that acid as *p*-coumaroyl-CoA — and *p*-coumaroyl-CoA is the branch node, the fork in the trunk where the routes of this book diverge.

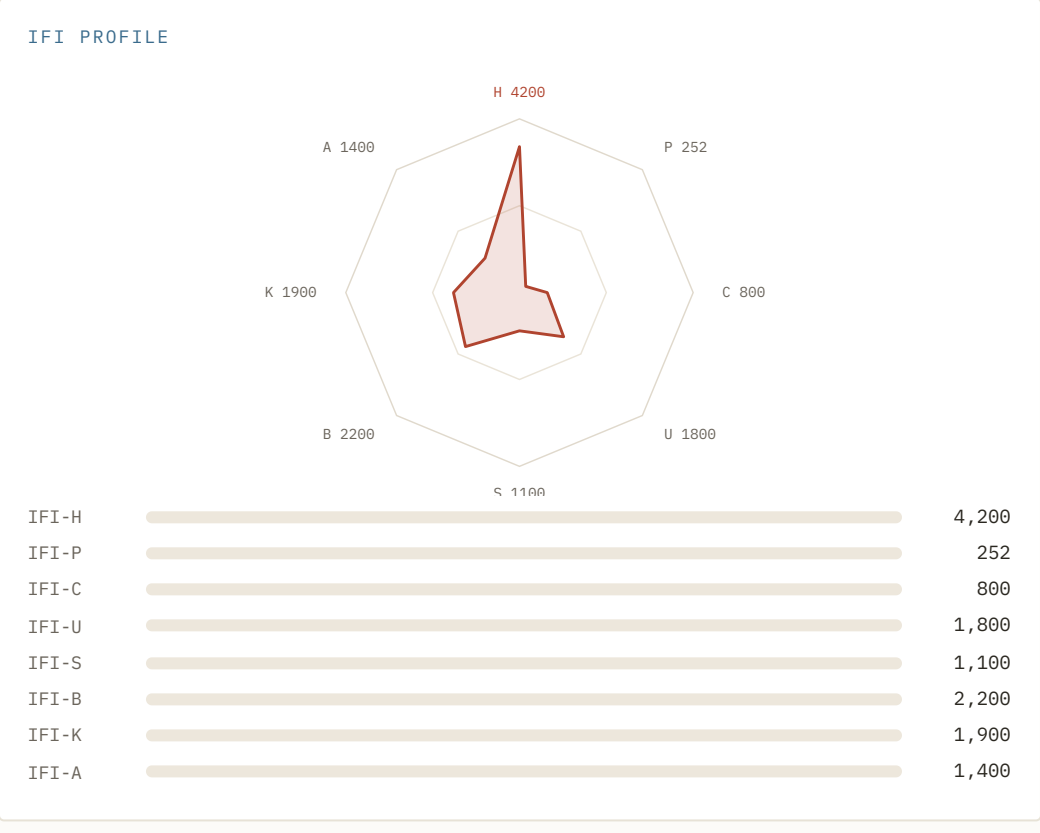
Three roads leave that fork, and they are rivals, because they are fed from one pool:

- Down one road, **hydroxycinnamoyl transferase (HCT)** and **caffeic acid O-methyltransferase (COMT)** carry the carbon toward feruloyl-CoA, then on to vanillin, then — via the aminotransferase **pAMT** — to vanillylamine. Vanillylamine is the head half of capsaicin. The tail half, an 8-methyl-branched acyl chain, is built on a separate line from branched-chain amino acids by **BCAT** and extended by a ketoacyl synthase (**KAS**). The two halves are married by **Pun1/CS** — capsaicin synthase, the acyltransferase also catalogued as AT3 — and the marriage *is* the heat axis.
- Down a second road, **chalcone synthase** condenses the same *p*-coumaroyl-CoA with three malonyl-CoA units and opens the flavonoid branch, which runs through the reductases (among them **ANR**, anthocyanidin reductase) to the flavan-3-ols and their polymers, the condensed tannins. Those tannins are the astringency axis — the pucker that cross-links the proteins in your saliva.
- Down a third road, the monolignol branch spends *p*-coumaroyl-CoA on lignin, the structural polymer that lets the plant stand up at all. This road is not optional. It has a floor beneath which the tissue simply fails.

Pungency, in a pepper that carries it, can be routed off this same trunk as well — a cinnamaldehyde made by reducing cinnamoyl-CoA, one more claimant at the fork. So before we have spoken a word about the other pools, three and sometimes four of the eight axes are already drinking from one cup. The other four axes draw from their own reservoirs — umami and kokumi from the glutamate-and-nitrogen pool that the plant assimilates through GS/GOGAT and shuttles through glutamate decarboxylase and the γ-glutamyl peptide machinery; sweetness and cool-burn from the isoprenoid pool, the GPP-and-IPP carbon that becomes a steviol glycoside on one line and a menthol on another. Each pool has its own fork, its own rivalries, its own floor. The budget is not one budget. It is several coupled budgets, and the art of Phoenix is partitioning inside each of them without letting one branch's appetite empty a node the others were counting on.

Reading Phoenix as a partition, not a maximum

Read Phoenix's profile as a feat of allocation rather than a feat of ceilings and it tells the truth about itself. Its heat is enormous — IFI-H 4,200, our model predicts approximately 2.95 million SHU, the redesigned Pun1 designed to produce capsaicinoids at a batch coefficient of variation under seven percent — but look at the other seven axes and notice that *none of them is at the ceiling*.



The per-axis normalization maximum for a scorecard is 5,000. Phoenix is nowhere near it on any axis but heat, and that restraint is the whole achievement. Set the eight numbers in a table and the partition becomes legible — each axis shown as its receptor, its raw index, its fraction of the 5,000 ceiling, and the thing the index actually encodes: the ratio of the concentration this fruit needs to half-activate the receptor against the concentration the reference needs. Each axis is normalized against its own stated convention reference, so the per-axis indices measure loudness relative to that axis's defined benchmark, not against a single common scale.

AXIS	CODE	RECEPTOR	IFI	÷ 5,000 CEILING	$EC_{50}(C)/EC_{50}(REF)$	LOUDNESS × REF
Heat	IFI-H	TRPV1	4,200	84.0%	0.238	4.20×
Zero-Bitter	IFI-B	TAS2R38	2,200	44.0%	0.455	2.20×
Kokumi	IFI-K	CaSR	1,900	38.0%	0.526	1.90×
Umami	IFI-U	mGluR1	1,800	36.0%	0.556	1.80×
Astringency	IFI-A	salivary PRPs	1,400	28.0%	0.714	1.40×
Pungency	IFI-P	TRPA1	252	5.0%	3.968	0.25×
Sweetness	IFI-S	TAS1R2·R3	1,100	22.0%	0.909	1.10×
Cool-Burn	IFI-C	TRPM8	800	16.0%	1.250	0.80×

Add the eight per-axis indices and you get 13,652, against a theoretical all-axes-maxed total of 8 × 5,000 = 40,000. Phoenix spends 34.1% of that imaginable total — and spends it *broadly*. The

mean axis sits at 1,707; heat, the loud one, is only 2.5 times that mean. Strip heat out and the remaining seven axes still sum to 9,452, which is to say **69.2% of everything Phoenix says is said by axes other than heat**. A plant cannot be maximally loud on eight axes at once; the budget forbids it, the shared nodes would collapse, the competing branches would starve one another into silence. Phoenix is the proof that you can run all eight *in balance* — each loud enough to be unmistakably present, not one so loud it bankrupts the rest.



Phoenix (SS-0096) under the lights — the first cultivar engineered to register on all eight axes in a single fruit. It is not the hottest number we could have chased, nor the loudest on any one axis; it is the hardest balance we have struck, which is to us the greater thing.

The arithmetic of loudness

The index that lets us compare a chili to a cup of tea on one page is a single ratio:

> **$IFI(c) = 1{,}000 \times EC_{50}(\text{reference}) \div EC_{50}(\text{candidate})$.**

EC_{50} is the concentration at which a stimulus drives its receptor to half of full response. A *louder* food reaches that half-point at a lower concentration — you need less of it — so a smaller EC_{50} gives a larger index. The reference is pinned by convention so that the native wild-type sits at exactly 1,000 on that axis; each axis carries its own defined reference constant, not a shared universal benchmark. Invert, scale by a thousand, and loudness becomes a number.

Turn the formula around and it reads the EC_{50} straight off the score. Rearranged, $EC_{50}(c) = 1{,}000 \times EC_{50}(ref) \div IFI(c)$, so the ratio $EC_{50}(c)/EC_{50}(ref) = 1{,}000 / IFI(c)$. Work Phoenix's heat: 1,000 / 4,200 = 0.238. Phoenix's fruit is designed to reach half-maximal TRPV1 current at **23.8% of the reference concentration** — 4.20× more potent by dilution in our model. If, to make it concrete, the wild-type reference half-activates at an extract dilution equivalent to 1.00 µM of capsaicin, Phoenix's design target gets there at 0.238 µM-equivalent. The same single division runs every row of the table above: bitterness-attenuation at 2,200 means its TAS2R38 effect is designed to reach half-max at 0.455 of reference (2.20× louder); cool-burn at 800 means 1.250 — *quieter* than reference, 0.80×, a deliberately restrained whisper of TRPM8 so the cool never fights the heat.

Heat also carries a second number, the one the world already knows, and it pays to show exactly how it relates to the first. The Scoville Heat Unit is a dilution assay: how far you must dilute an extract before the pungency disappears against a panel. The field's working conversion is roughly 16 SHU per part-per-million of capsaicinoid. Run Phoenix's model-predicted ~2.94 million SHU through it and the naive arithmetic returns 2,937,948 ÷ 16 ≈ **183,622 ppm**, which would read as 18.4% of the fruit's mass as a single class of secondary metabolite.

That number is a lesson in why we built the index. No placental tissue on earth warehouses eighteen percent of itself as capsaicinoid; the cells would fail long before. The Scoville figure is a *perceived-pungency* readout taken by dilution, not a mass inventory, and the chasm between "18% by the naive conversion" and "what a cell can physically hold" is not an error to hide — it is the exact room that delivery and formulation were built to fill. We will return to that room at the end of the chapter. For now, hold the honest point: SHU measures how loud, by dilution; it does not measure how much, by weight, and treating the two as the same is the first mistake the index was designed to retire.

The ladder, re-solved rung by rung

Phoenix did not arrive whole. It is the top of a flagship ladder, and the ladder is itself a drawn lesson in composition, because our model's predicted heat and axis-count climb *together* up it — each rung adds routes and increases computational design scope, not only Scovilles.

CULTIVAR	SKU	SHU (PREDICTED)	IFI-H	AXES
Ember	SS-0087	866K	1,570	2/8
Cinder	SS-0088	1.05M	2,290	3/8
Kindle	SS-0089	1.25M	2,000	3/8
Flare	SS-0090	1.48M	—	4/8
Blaze	SS-0091	1.72M	2,630	4/8
Wildfire	SS-0092	1.96M	2,930	5/8
Scorch	SS-0093	2.21M	3,220	5/8
Firestorm	SS-0094	2.47M	3,490	6/8
Solara	SS-0095	2.72M	3,980	6/8
Phoenix	**SS-0096**	**2.95M**	**4,200**	**8/8**

The naive reading is that you bolt a branch on, then another, the way you add toppings. The true reading is that every new branch forces you to *re-solve the entire budget*, because the moment a branch comes online it begins to draw on nodes the existing branches were already using. Going from six axes to eight is not adding two lines to a diagram; it is finding a new partition of *p*-coumaroyl-CoA among heat, astringency, pungency, and the lignin floor, and simultaneously a new partition of the nitrogen pool between umami's free glutamate and kokumi's γ-glutamyl peptides, such that all eight land where the table says and none of the floors crack. That is why the index on any one axis tends to *soften* as the axis-count rises even while total heat climbs: the carbon that could have pushed one branch to its ceiling is being deliberately spread.

We can make the carbon arithmetic explicit with a reduced model. Normalize the trunk flux leaving the *p*-coumaroyl-CoA node to 100 units, and take the only canonical inputs to be the index values — everything else here is a stated model assumption, not a claimed measurement. Lignin, the structural floor, must hold at least 30 of those 100 units or the fruit wall gives way; that leaves 70 to divide between the two phenylpropanoid-fed flavor branches at this node, heat and astringency. Phoenix runs those two at indices 4,200 and 1,400 — a 3-to-1 demand — so allocate the 70 as 52.5 to the heat-precursor road and 17.5 to the astringency road. The partition closes: 30 + 52.5 + 17.5 = 100, every floor intact, both flavor branches lit. Now ask what it would take to drive heat alone to the 5,000 ceiling. Because of the curve we are about to draw, the last stretch of index costs carbon out of all proportion to the loudness it buys; the heat road would demand on the order of 90 units, and 90 + 30 for lignin is already 120 — more carbon than the node supplies, before astringency gets a single unit. The ceiling is not merely hard to reach. **In a plant that must also stand up and also say seven other things, the ceiling is structurally forbidden.** Phoenix's 4,200 is not a shortfall from 5,000; it is the near edge of what the floor and the family of branches will permit, which is a more honest target to design against.

Two ways a branch steals from its neighbor

Branches compete in two distinct ways, and conflating them is how intuition goes wrong. The first is the obvious one: **precursor depletion.** Two roads leaving one node draw from one pool of *p*-coumaroyl-CoA; push one and the other finds less substrate waiting. That is the competition the reduced model above captures, and you answer it by raising total flux into the node or by relaxing a floor.

The second is subtler and the flux model catches it where intuition does not: **competitive inhibition at a shared enzyme.** When two branches are served by one catalyst, their substrates compete for the same active site, and you can throttle one branch not by exhausting its precursor but simply by *occupying the enzyme* with the other's substrate. The Michaelis–Menten rate under a competitive inhibitor is

```
v = V_max · [S] / ( K_M · (1 + [I]/K_i) + [S] ).
```

Take a shared ligase serving branch A, with branch B's substrate acting as the competitor I. Set the constants symmetric — K_M = K_i = K — and hold branch A's substrate at its own K_M, [S] = K. The rate collapses to a clean function of how much of B's substrate is present:

```
v_A = V_max / (2 + [I]/K).
```

[I] (BRANCH B SUBSTRATE)	V_A
0	0.500 V_max
1 K	0.333 V_max
2 K	0.250 V_max
4 K	0.167 V_max

Branch B ramping from nothing to 4 K cuts branch A's throughput threefold — 0.50 down to 0.17 V_max — with *not one molecule of precursor removed from the pool*. The substrate is all still there; the enzyme is just busy. Two branches can starve each other at an active site while swimming in carbon, and a breeder selecting on phenotype would never see the mechanism, only the disappointing fruit.

The answer we reach for, where it is available, is not to out-shout the competition but to end it: **duplicate the enzyme.** Give each branch its own paralog — its own copy of the ligase or the transferase — and the shared active site is no longer shared. Each copy runs its own reaction near V_max/2 independently, and branch A recovers its full 0.50 V_max while branch B keeps its own. Gene duplication is how the plant kingdom itself diversified these very pathways over deep time; we are using the same move, deliberately and in one generation, to let two routes that used to fight run side by side. It is the quiet reason an eight-axis genome can carry more total enzyme traffic than the sum of its branches would suggest.

Dose, response, and the shape of the last stretch

Why does the top of an axis cost so ruinously much carbon? Because receptors do not answer concentration in a straight line. TRPV1's response to capsaicin follows a sigmoid with positive cooperativity — a Hill function with a coefficient near 2:

> $R / R_{max} = [C]^n / (EC_{50}^n + [C]^n)$, with $n \approx 2$.

Walk it out. At [C] = EC_{50} the response is exactly half, by definition. Double the dose and you do *not* double the response — you get $2^2/(1+2^2) = 4/5 = 80\%$. Triple it and you reach 90%. But chase the last tenth and the curve turns cruel: 95% response needs [C] = $\sqrt{19} \approx 4.36\times EC_{50}$, and 99% needs $\sqrt{99} \approx 9.95\times EC_{50}$.

TARGET R/R_MAX	[C] REQUIRED
20%	0.50 × EC_{50}
50%	1.00 × EC_{50}
80%	2.00 × EC_{50}
90%	3.00 × EC_{50}
95%	4.36 × EC_{50}
99%	9.95 × EC_{50}

The first half of the response costs one unit of dose; the last nine percentage points, from 90 to 99, cost more than triple it — from 3× to nearly 10× EC_{50}. That is the Hill curve's hard shoulder, and it is where the carbon budget goes to die. Pushing an axis from loud to maximal means pouring ever more precursor into ever less perceived gain, which is exactly why maxing any single axis bankrupts the rest and why a balanced eight-axis partition is the rational design, not a compromise. The cooperativity that makes heat arrive so decisively on the tongue — the near-switch quality of a true superhot — is the same cooperativity that makes its ceiling unaffordable.

Consistency is a route you design on purpose

A pathway that produces a brilliant fruit once and a mediocre one next season has not been engineered; it has been gambled on. The axis we hold most jealously is reproducibility, and it is won at the enzyme.

Our flux analysis finds that the primary control over total capsaicinoid output sits upstream, at 4CL and HCT/CCoAOMT in the phenylpropanoid pathway — not at the terminal Pun1/AT3 step. Pun1 is the final assembler; it labels the product, it does not govern the rate. What Pun1 does govern — and what the two edits below address — is how *consistently* the synthase folds and seats its substrates batch to batch.

Phoenix's redesigned capsaicin synthase carries two substitutions an ESM2 protein-language model scored — two edits that, in our model, tighten substrate binding and relieve steric strain in the active site, targeting the batch variance no breeding program ever removed by selection alone.

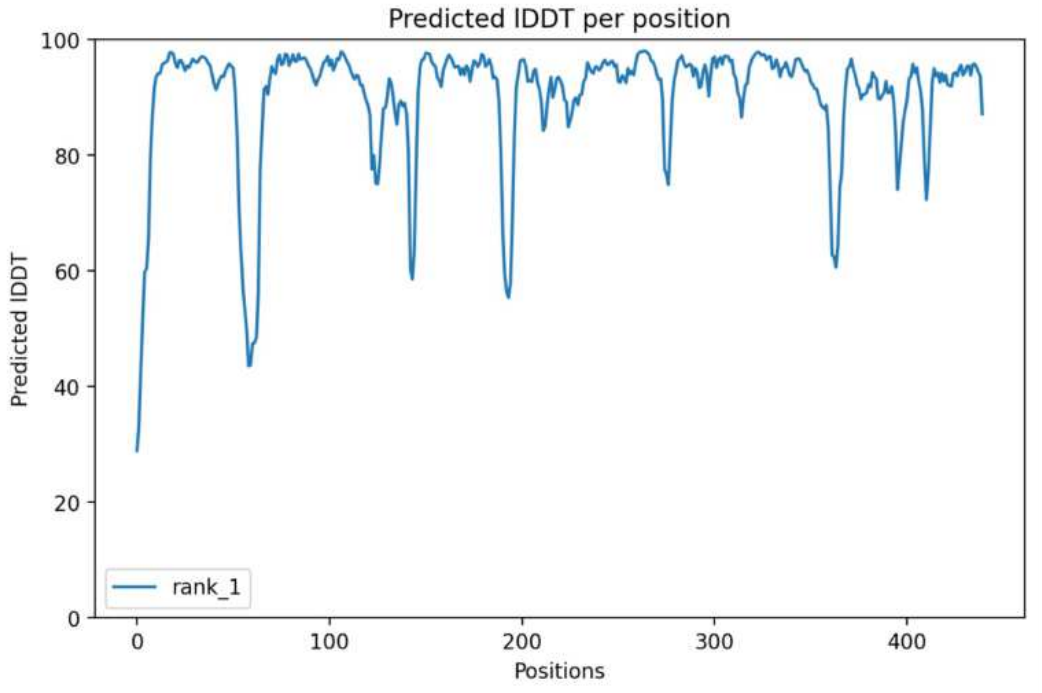
EDIT	SITE AND EFFECT	ΔΔG (KCAL/MOL)	FOLDING/BINDING FACTOR, EXP(|ΔΔG|/RT)
S39L	N-terminal; tightens substrate binding	−2.1	34.7×
L345G	active-site cleft; steric relief for the acyl chain	−1.8	20.9×
C175S *(structural annotation — not in shipped construct)*	removes a stray cysteine; prevents a misfolding disulfide	−0.9	4.6×
combined (S39L + L345G)		**−3.9**	**≈ 725×**

The mechanism reads cleanly off the enzyme's job. Pun1/CS is an acyltransferase: it seats vanillylamine and the 8-methyl-branched acyl-CoA in one cleft and joins them. S39L, at the N-terminus, tightens the grip on substrate — a lower effective K_M, so the reaction saturates at lower concentration. L345G swaps a bulky leucine for the smallest residue, relieving a steric clash so the branched acyl tail seats the way it should — a higher k_cat. C175S, a structural annotation from our modelling, marks a reactive cysteine that could form a misfolding disulfide; it is recorded in the design file but was pulled from the final shipped construct.

The thermodynamics make the payoff exact and, to us, beautiful. Stabilization shifts a folding or binding equilibrium by the Boltzmann factor exp(|ΔΔG| / RT); at 298 K, RT = 0.592 kcal/mol. The free energies of the two shipped edits *add* — −2.1 − 1.8 = −3.9 kcal/mol — which means their equilibrium factors *multiply*: 34.7 × 20.9 ≈ 725, exactly exp(3.9 / 0.592). The two edits, neither heroic alone, compound into a roughly 725-fold swing toward the folded, catalytically live enzyme. A population of synthase that is overwhelmingly folded and correctly seated is a population that makes the same amount of capsaicin every time.

That is what targets the variance. Batch coefficient of variation — CV = σ/μ, the spread as a fraction of the mean — runs at 42% in wild-type; our design specification targets **≤ 7%** in the engineered line. The consequence for a formulator is enormous. At our model's predicted mean of approximately 2.94 million SHU, a 7% CV corresponds to σ ≈ 205,656 SHU; at that spread, ninety-five percent

of batches would land inside μ ± 1.96σ, a window of **2.53M to 3.34M SHU**. The wild-type 42% is σ ≈ 1.23M, and the same 95% window sprawls from 0.52M to 5.36M — a tenfold span, from merely hot to uncharacterizable. Read it back onto the index and the point sharpens: IFI-H 4,200 at our 7% CV design target holds a 95% window of 3,624 to 4,776, comfortably on the scorecard; at 42% CV the window runs 743 to 7,657, and 7,657 is past the 5,000 ceiling the scorecard can even draw. An axis that cannot be kept below its own ceiling cannot be plotted, cannot be promised, cannot be sold. Tight CV is not a quality nicety bolted onto the pathway; it is the thing that makes the index *mean* anything at all.



The capsaicin-synthase model coloured by pLDDT confidence — the binding domain and active-site cleft, where S39L and L345G (the shipped double variant) and C175S (a structural annotation) are located, resolve to high confidence.

When the plant itself runs out of room

Return now to the eighteen-percent problem — the gap between what the Scoville number implies and what a cell can hold. A pathway can be perfectly balanced and still run into a wall that has nothing to do with enzymes: *storage*. There is only so much the placental blisters, the trichome glands, and the vacuoles can physically warehouse. Solve every branch and the fruit still caps out

at the volume its own compartments provide, and on the Hill curve's hard shoulder, where the last decile of response wants ten times the dose, that cap is exactly where the eater would most want more.

So we decouple the delivered dose from the stored dose. The full-spectrum program resolves into roughly **300 IFC products** — integrated full-composite formulations — carried in a **nanoparticle (NP) delivery system** that holds the designed molecules in a protective, uniformly dispersed form. The NP matters for two reasons the arithmetic has already named. First, it carries a composite to the tongue at an effective concentration the raw tissue could never warehouse, pushing [C] up the Hill curve's expensive last stretch in formulation where the pathway itself has run out of compartments. Second, its *uniform* dispersion is a second front in the war on variance: even a CV ≤ 7% at the plant can be tightened further when the dose that reaches the receptor is metered by a particle rather than by wherever a blister happened to concentrate. The pathway builds the molecules; the particle decides how loudly they finally speak. Between them, the composite reaches a loudness the plant alone could store nowhere.

Failure modes, and the limits we keep honest

Everything here can break, and a pathway book owes the reader the breakages.

Floor violation. Push any branch too hard and the first thing to fail is not flavor but structure — the lignin floor cracks, the fruit wall weakens, and a brilliant profile rots on the vine. The ceiling is bounded from below by the plant's need to be a plant.

Starvation at a shared node. Add a branch without re-solving the budget and it will quietly rob a neighbor, sometimes by depleting precursor, sometimes — the trap intuition misses — by occupying a shared enzyme while the precursor sits untouched. The fix (duplicate the enzyme) is not always available; some nodes resist duplication, and there the axes must simply be balanced lower.

The SHU scale's bluntness. The Scoville unit is a dilution assay with organoleptic roots; its own reproducibility is poor, which is precisely the imprecision the receptor-anchored index was built to replace. We report SHU because the world speaks it, but we design against EC_{50}, because a number you can halve and double and back-calculate is a number you can engineer.

The reduced model is reduced. The 100-unit partition in this chapter is a teaching model normalized to canonical indices, not a genome-scale flux solution; its only computed inputs are the index values themselves. The real optimization runs over thousands of reactions, and it will surprise you — a branch you expected to be cheap turns out to share a cofactor pool you had not drawn. Honesty about the model's altitude is part of the method.

The honest edge of balance. Phoenix is not the hottest fruit we could grow, nor the loudest on any one axis. It spends 34.1% of the imaginable all-axes total and no more, and 69.2% of what it says is said by axes other than heat. That restraint is the claim: not a maximum, but a partition — eight routes carried together on one

carbon budget, each loud enough to be unmistakably present, none so loud it bankrupts the rest, and the whole of it grown well. These are computational designs; Scoville Splice's cultivars and constructs have not been experimentally validated in a laboratory. It is the hardest arithmetic in the book, and the clearest answer our family knows to the question we asked at dawn: not *how hot*, but *how much, together, and still grown well*.

Chapter 17 · The Transferable Pathway — *five modules of seven, and the climb to sixty-four.*

The capsaicin work was never, in the end, about peppers — and here the claim stops being a slogan and becomes an engineering plan with gene counts and calendar dates attached. Forty-two organisms carry a designed flavor today; sixty-four more are drawn on the roadmap. The reason that spread is possible — the reason the frontier is a climb and not a wall — rests on one structural fact about how a plant makes a taste: the trunk of a flavor pathway is universal, and only its terminal verbs are special. A flavor is not a single machine welded to a single species. It is a stack of modules, most of them already turning in any green thing you can name, and *transfer* — the real business of the frontier — is the patient work of counting which modules a destination plant lacks, delivering exactly those, and leaving everything else alone.



FIG. 17 — Eggplant, the near frontier. A nightshade cousin of the pepper, it already holds so much of the pepper's own trunk and tissue context that the Dark Matter line is among the shortest transfers on the board — almost the pepper's terminal verbs alone, carried into a plant that is already most of the way there.

We did not arrive at that fact by faith. We arrived at it by taking a complete flavor route apart on the bench, naming every functional block, and asking of each one the same flat question: does this block already run in the destination plant, or must we carry it in? The answer, repeated across route after route, is the quiet engine under this entire book. Break any of our pathways into its parts and the same count falls out: of roughly seven functional modules in a finished flavor route, **about five of seven transfer essentially**

unchanged to any organism with a sequenced genome, because they are the shared, conserved, housekeeping chemistry every plant already keeps. Only two must be carried — the ones that commit the branch and do the terminal tailoring, the special verbs that make a pepper a pepper and not a potato. So the difficulty of writing a flavor into a new plant is never the romantic "can it be done." It is the engineer's "how many verbs does this host lack, and will its trunk feed them once they are installed." That number sorts the whole sixty-four-organism roadmap into tiers as cleanly as a sieve, and it is a number we can read off a genome before we plant a single seed.

Seven modules, and the five that come for free

Lay a complete capsaicinoid route flat and seven blocks resolve out of it, each with a job as distinct as a station on a line. We name them the way we name them in the lab, because the names are the inventory and the inventory is the plan.

#	MODULE	WHAT IT DOES	REPRESENTATIVE PARTS	TRANSFERS?
1	Precursor supply	Feeds the route its two raw carbons streams	L-phenylalanine (shikimate); valine/leucine	`yes — universal`
2	Trunk relay	The shared phenylpropanoid backbone	PAL · C4H · 4CL · HCT · COMT	`yes — universal`
3	Branch commit	Turns trunk output toward *this* flavor	pAMT · BCAT	**no — carry in**
4	Terminal tailoring	Builds and joins the final molecule	KAS · Pun1/CS	**no — carry in**
5	Transport & storage	Moves and vacuole-stores the product	vesicular / tonoplast carriers	`yes — repurposed`
6	Regulatory throttle	Sets *when* and *where* the route runs	tissue promoters; MYB-type factors	`yes — host's own`
7	QC background	Keeps a busy route from poisoning its host	chaperones, UGTs, peroxidases	`yes — present`

Five of those blocks — precursor supply, the trunk, transport, the throttle, and the quality-control background — are things a tomato and a tobacco and a strawberry already own, because they are how any plant conducts ordinary business. A plant that could not run module 2 could not build lignin to stand upright; a plant that could not run module 7 would have died of its own secondary metabolism long ago. We do not *install* those. We *find* them, confirm they are intact, and borrow them. Only modules 3 and 4 — the branch-committing and terminal-tailoring enzymes — are pepper's private

vocabulary, and only those must physically be carried into a new genome. Two verbs out of a sentence of seven. That is the whole leverage of the platform, stated as a fraction.

The trunk every plant already runs

To see why five of seven are free, follow the carbon. A capsaicin molecule is an amide: a fragrant aromatic head, vanillylamine, joined to a branched ten-carbon fatty tail, 8-methyl-6-nonenoyl. The head is pure phenylpropanoid chemistry, and the phenylpropanoid trunk is the most thoroughly conserved secondary route in the entire plant kingdom — the same five-enzyme relay that every vascular plant uses to build lignin, to tan a wound brown, to paint a flower, to screen out ultraviolet. It begins with phenylalanine and walks it, one defined reaction at a time, toward the vanilla ring.

ENZYME	REACTION IT CATALYZES	WHAT IT YIELDS
PAL	deaminates L-phenylalanine	*trans*-cinnamic acid
C4H	hydroxylates the ring (para)	*p*-coumaric acid
4CL	activates the acid as a thioester	*p*-coumaroyl-CoA
HCT	shuttles the acyl group onto a shikimate ester	the 3-hydroxylation substrate
COMT	methylates the 3-hydroxyl	the feruloyl / vanillin branch point

PAL is the committed first step, the gate: phenylalanine ammonia-lyase strips ammonia from the amino acid and hands cinnamic acid to the route, and from that instant the carbon belongs to secondary metabolism and not to protein. C4H, a cytochrome-P450 monooxygenase, hangs a hydroxyl on the far face of the ring. 4CL lights the acid for business by capping it with coenzyme A. HCT and COMT together decorate the ring — a hydroxyl here, a methyl there — until the trunk reaches the feruloyl branch point, from which a short tail delivers vanillin, the aldehyde that smells of its own name. Every plant on the roadmap already runs **PAL → C4H → 4CL → HCT → COMT** at some rate, for its own private reasons. When we transfer a flavor, we are not teaching a strawberry to make vanillin's precursors from scratch. We are tapping a pipe it already has pressurized and running.

That is module 2, free. Modules 1, 5, 6, and 7 are free for the same reason: precursor amino acids come off central metabolism in every leaf; vacuolar storage is how every plant warehouses a secondary product it would rather not keep loose in the cytosol; tissue-specific promoters are the host's own switches; and the chaperones, glycosyltransferases, and peroxidases of the quality-control background are standing guard in every cell whether we ask them to or not. The trunk and its context come with the plant. The flavor does not.

The two verbs you have to carry

What the plant does *not* come with is the vocabulary that turns generic vanillin into the specific molecule that lowers a heat channel's threshold. That vocabulary is four enzymes, falling into the two modules we carry.

The branch-committing module is **pAMT** and **BCAT**. pAMT — a putative aminotransferase — takes vanillin and transaminates its aldehyde into an amine, producing vanillylamine, the aromatic head of the final molecule. On the other side of the route, BCAT, a branched-chain aminotransferase, deaminates valine or leucine to a branched keto-acid, committing a scrap of ordinary amino-acid carbon to a decidedly unordinary fatty tail. Neither reaction happens in a plant that is not trying to make a capsaicinoid; both are the point where the universal trunk is bent, deliberately, toward heat.

The terminal-tailoring module is **KAS** and **Pun1/CS**. KAS, a ketoacyl-ACP synthase, runs the branched keto-acid through rounds of two-carbon elongation to build 8-methyl-6-nonenoyl — a fatty acid that is branched where ordinary fatty acids are straight, which is precisely why it reads at the receptor the way it does. And then Pun1/CS — capsaicin synthase, the acyltransferase the breeders spent a century selecting for without ever being able to see — performs the single condensation that is the whole climax of the route: it joins vanillylamine to the activated branched acyl group, forming the amide bond, and capsaicin exists. Eighteen carbons: eight from the vanilla head, ten from the branched tail, married in one enzymatic act.

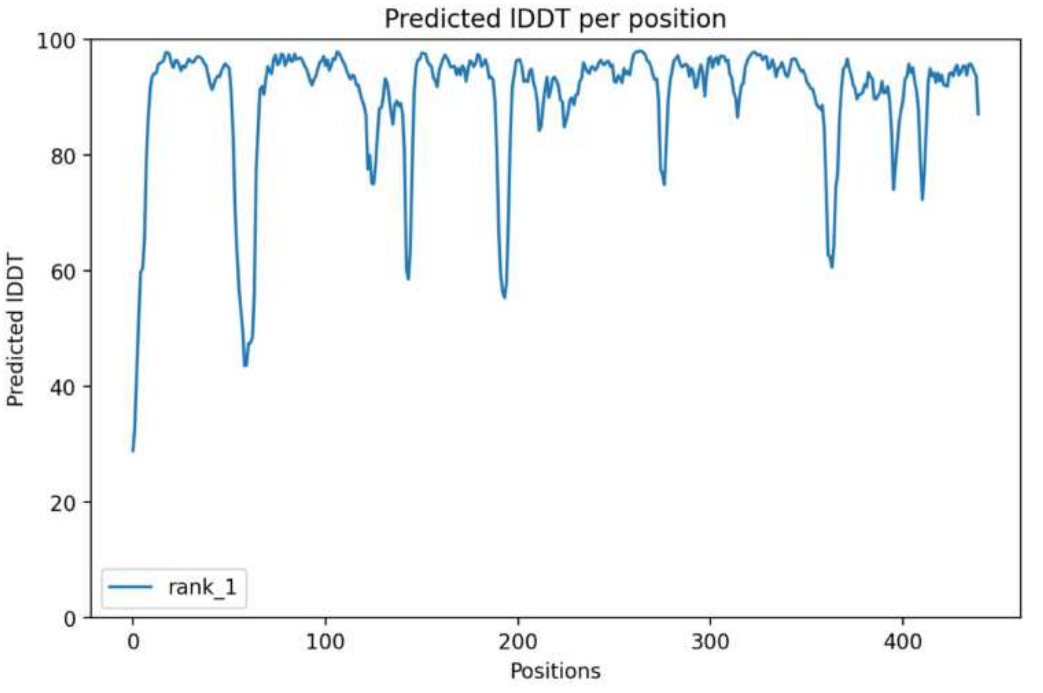


FIG. 18 — Pun1/AT3, the terminal verb, modeled and colored by per-residue confidence. The acyltransferase that joins head to tail is the one cargo every heat transfer must carry; its active-site cleft, high-confidence in the model, is where a century of blind selection was quietly aiming.

This is why the count is five-of-seven and not seven-of-seven, and why it holds across axes and not only for heat. A cooling flavor swaps the terminal verbs for the ones that build menthol; a savory flavor swaps them for the route that pools free glutamate; an astringent one for the condensed-tannin branch. But in every case the trunk is the same conserved relay, the context is the host's own, and the *carry-in* is just the handful of branch-and-finish enzymes

that no ordinary plant keeps. Change the two verbs, keep the five modules, and you have changed the flavor without rebuilding the plant.

The index as a transfer metric

A transfer is not finished when the enzymes are in. It is finished when the axis *lights up* — when the destination plant's new molecule speaks at its receptor loudly enough to measure. For that we need a number, and the number is the Iyengar Flavor Index, which is built from the one quantity that actually governs loudness: the half-maximal activation concentration of the receptor.

> **IFI(candidate) = 1,000 × EC_{50}(reference) ÷ EC_{50}(candidate)**

The reference EC_{50} for each axis is a stated convention constant — defined, like Scoville's dilution-to-threshold, by agreement rather than calibrated to published measurements. Each axis has its own convention constant that places that axis's baseline at exactly 1,000; because reference constants differ across axes, IFI scores express potency within a single axis and are not a single common scale across all eight.

EC_{50} is the concentration at which a stimulus drives its receptor to half of full response. It is an inverse measure of potency: a *smaller* EC_{50} means the molecule reaches the receptor's half-point at a lower dose, which is to say it is louder. The Index inverts EC_{50} against a fixed reference — the native wild-type pepper's capsaicin at its own TRPV1 — and scales by a thousand, so that the reference lands at exactly 1,000 and everything is read as a multiple of it. Rearrange the formula and the reading becomes plain:

> EC_{50}(candidate) ÷ EC_{50}(reference) = 1,000 ÷ IFI, and loudness-over-reference = IFI ÷ 1,000.

So an IFI-H of 4,200 is not an abstract score. It says the candidate reaches half-activation at 1,000 ÷ 4,200 = **0.238** of the reference's concentration — it is **4.2× louder** than a field pepper's heat. Run the superhot ladder through that single arithmetic and the whole thing becomes legible as receptor physics:

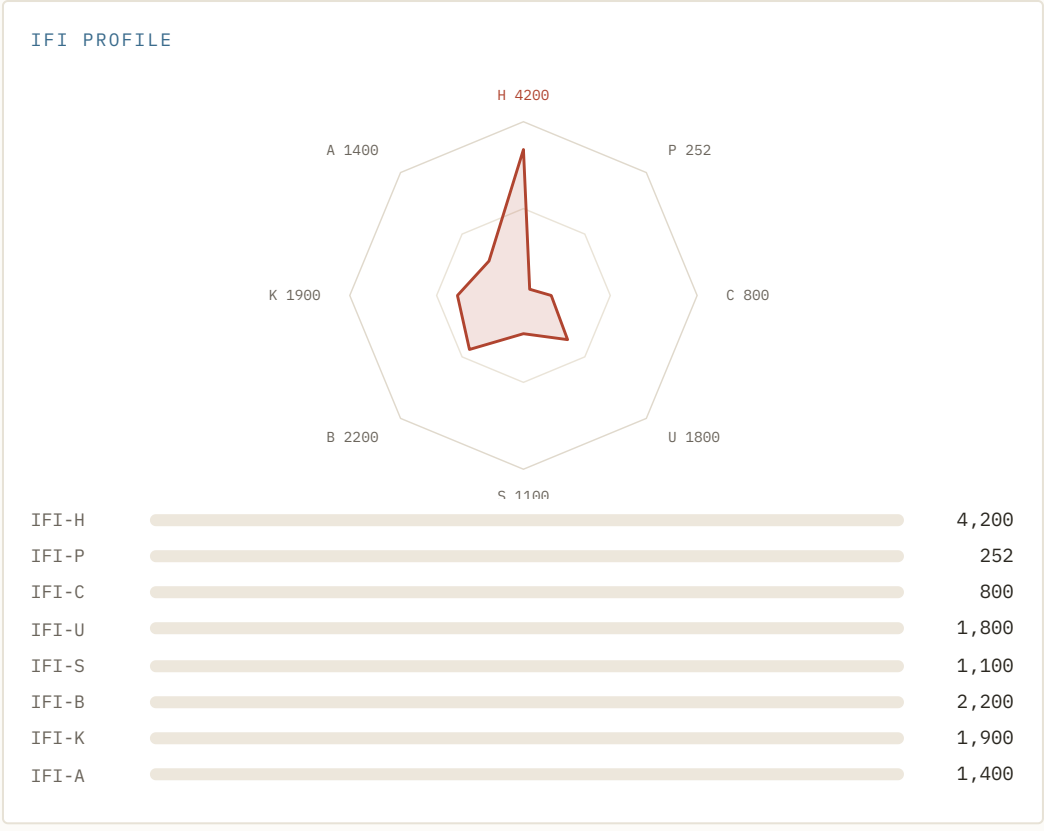
CULTIVAR · SKU	SHU (PREDICTED)	IFI-H	EC_{50} (× REFERENCE)	LOUDNESS
Ember · SS-0087	866K	1,570	0.637	1.57×
Blaze · SS-0091	1.72M	2,630	0.380	2.63×
Firestorm · SS-0094	2.47M	3,490	0.287	3.49×
Solara · SS-0095	2.72M	3,980	0.251	3.98×
Phoenix · SS-0096	2.95M	4,200	0.238	4.20×

Notice what the two columns do *not* do: they do not track in lockstep. SHU is a bulk assay — it counts how much capsaicinoid sits in a gram of fruit, the Scoville scale being linear in concentration, with pure capsaicin near sixteen million units (which puts Phoenix's model-predicted 2.95 million SHU at the pungency of a material a shade over eighteen percent as hot as pure capsaicin). IFI-H is a *receptor* measurement — how loudly whatever is present speaks at the channel. Divide one by the other down the ladder and the ratio

drifts, from about 552 SHU per IFI-point at Ember to about 702 at Phoenix. That drift is not noise; it is signal. It is the acyl tail doing its work — the capsaicin-to-dihydrocapsaicin balance shifting with the branched chain, so that each unit of mass speaks a little louder at the top of the ladder than at the bottom. A bulk chemistry assay is blind to that. The Index is built to see exactly it, which is why we grade a transfer on IFI and not on a chromatogram alone.

The dose-response math makes the stakes vivid. Receptor activation follows a sigmoid — response rises with concentration raised to a cooperativity exponent, the Hill coefficient, which for capsaicin at TRPV1 is commonly fit near two. Take a dish seasoned so that a field pepper's capsaicin would sit right at its own half-activation, and ask what a Phoenix-derived material does at that *same physical concentration*. With the Hill form and an EC_{50} lowered to 0.238 of reference, the fractional activation is $1 \div (0.238^2 + 1) = 1 \div 1.057 =$ **0.946** — nearly ninety-five percent of the channel's full response where the wild type reads a tidy half. The number 4,200 is a promise about the tongue: at a concentration the field pepper would deliver as a measured warmth, the designed fruit is at the ceiling of what the channel can report. That is the whole case for measuring loudness and not mass.

When every verb lands and the branches stack, a single plant can be made to speak on all eight axes at once. Phoenix is that proof — the first cultivar designed to register across the full Index rather than on heat alone:



Read that profile as eight independent transfers resolved in one genome: a heat branch at 4.2× reference, a pungency branch, a cool-burn branch, umami, sweetness, zero-bitter, kokumi, and astringency — eight terminal vocabularies installed on top of one shared trunk, each tuned and balanced against the host's single carbon budget. Every one of those numbers is an EC_{50} ratio against its own axis reference — each axis has its own defined reference constant, so the eight scores express potency within their respective

axes rather than on a single common scale. A multi-axis plant is not a louder plant; it is a plant speaking eight sentences in eight languages at once, each at a volume we chose.

A ladder of botanical distance

Because the carry-in is only the branch-and-finish verbs, the cost of a transfer is set almost entirely by how much of the *rest* the host already shares — how close, botanically, the destination sits to the pepper, and how readily its trunk feeds a borrowed branch. Sort the sixty-four-organism roadmap by that distance and it falls into four clean tiers, gene counts and horizons climbing together.

TIER	WHAT IT IS	GENES IN	HORIZON
T1 · Solanaceae Core	Volcanic Tomato, Dark Matter Eggplant, Laboratory Tobacco	4–5	18–24 mo
T2 · Extended Nightshades	Seismic Potato, Celestial Ground Cherry, Acid Tomatillo, Floral Petunia	5–7	24–36 mo
T3 · Cross-Family	Sweet-Heat Strawberry, Contradiction Cucumber, Solar Melon, Forbidden Apple, Espionage Lettuce	7–10	36–48 mo
T4 · Frontier Synthetic	Deep-Earth Carrot, Nuclear Cabbage, Wildfire Broccoli, Storm Corn, Dynasty Rice	10–12	48–72 mo

Read that table as a map of botanical distance and the logic closes. The **T1 Solanaceae core** — tomato, eggplant, tobacco — are the pepper's own close cousins; they share nearly all of its tissue context and run the trunk at a compatible rate, so carrying the heat branch into them is almost the bare terminal-verb transplant, four or five genes and eighteen to twenty-four months, the shortest reach on the board. The Volcanic Tomato's ladder — Ember, Magma, Pyroclast, Caldera, Krakatoa — and the eggplant's Dark Matter line — Smolder, Nightshade, Eclipse, Obsidian, Event Horizon — are the near frontier because the plant is already most of the way there. The **T2 extended nightshades** — Seismic Potato, Celestial Ground Cherry, Acid Tomatillo, Floral Petunia — sit a step farther and need five to seven genes to account for the context they do not quite share.



FIG. 19 — Sweet-Heat Strawberry, the far reach. A rosaceous fruit whose trunk is pointed at its own sugars and esters, not at a pepper's; the Candied Fang and Devil's Dessert lineages are a true cross-family transfer — seven to ten genes — where more of the route must be supplied because less of it is already there.

The **T3 cross-family** plants are the honest reach across family lines — the Sweet-Heat Strawberry with its Warm Kiss, Sugar Burn, Candied Fang, and Devil's Dessert lineages, the Contradiction

Cucumber, the Solar Melon, the Forbidden Apple, the Espionage Lettuce — seven to ten genes, because the host's trunk is aimed somewhere else entirely and more of the branch-and-context must be supplied to coax it. And the **T4 frontier synthetic** organisms — Deep-Earth Carrot, Nuclear Cabbage, Wildfire Broccoli, Storm Corn, Dynasty Rice — demand ten to twelve genes and the better part of a decade, because they are grains and roots and Brassicas whose metabolism shares the least with a pepper's and whose carbon is already spoken for by a starchy sink. The tiers are not a ranking of ambition. They are a reading of genomes.

The craft of a transfer

"Carry the genes in" hides a great deal of patient craft, and in a book about pathways the craft deserves its own accounting. A capsaicin-synthase gene lifted straight from a pepper will not simply run in a strawberry, and four distinct adjustments stand between the sequence and a working flavor.

First, the coding sequence is **codon-optimized** to the host's own translational preferences — the pepper's favored triplets rewritten into the strawberry's, so the ribosome reads the gene fluently instead of stalling on rare codons. The protein is identical; the mRNA that encodes it is re-spelled for a new reader.

Second, and most consequentially, each gene is placed under a **promoter the host will read in the right tissue at the right time**. A terminal enzyme expressed in the wrong organ or the wrong season is wasted carbon and, often, a sick plant — a leaf asked to brew heat it cannot store, a root burdened in a month it should be filling. We borrow the host's own tissue-specific switches so the branch fires where the fruit can warehouse what it makes. This is module 6, the throttle, doing its job: the when and the where are the host's to grant, and we ask for them precisely.

Third, the construct is delivered — most often by *Agrobacterium tumefaciens*, the soil bacterium that in nature splices a defined stretch of its own DNA, the T-DNA, into a plant's genome. We reprogram that natural courier to carry our branch-and-finish cassette in place of its own payload, so the plant integrates the new verbs as stably as it would integrate a gene of its own.

Fourth, the quiet seventh module must be honored. The quality-control background — folding chaperones, detoxifying glycosyltransferases, the peroxidases that can *degrade* a capsaicinoid as readily as store it, and the transporters that move the product to the vacuole — rides along implicitly in a close cousin, whose cellular neighborhood already tolerates a busy secondary route. In a distant host it must be checked, and sometimes supplemented, because a genome that never ran this branch may clear its product before it accumulates, or may lack a carrier to tuck it safely away. Much of why the gene counts climb from four to twelve across the tiers is exactly this: the farther the host, the less of the supporting neighborhood comes for free. Transfer is never the transplant of a single organ. It is the reconstruction of a working metabolic neighborhood — enzyme, promoter, carrier, and context together — inside a plant that has to go on being entirely itself.

The flux budget: can the host afford it?

Even a perfect transfer fails if the host cannot pay for it, and so every transfer ends where it began, at the carbon. A flavor is built from carbon the plant fixed for its own living, and carbon spent on a branched fatty tail is carbon not spent on the sugars, the proteins, and the micronutrients that make a fruit worth eating. This is the part of the work our family holds most carefully, because a plant starved to post one loud number is not a plant we would ever grow. The flux budget is also a nourishment budget; loudness we cannot afford is loudness we decline.

So before a single construct is drawn we run a genome-scale flux-balance model of the destination plant — a system of mass-balance equations in which every metabolic reaction is a flux and the sum into and out of every intermediate must cancel. Installing the heat branch adds reactions and a demand: each capsaicin pulls eighteen carbons out of central metabolism and off toward the placenta, eight through the phenylpropanoid trunk, ten through the branched fatty-acid route. The model asks whether the host's trunk can supply that pull at the chosen promoter strength without starving the fruit's own growth, and it returns a verdict — a feasible flux, or an infeasible one that would demand more carbon than the plant can route. Our Metabolic Control Analysis of that pathway model finds flux control concentrated in the upstream phenylpropanoid trunk — at 4CL and HCT in particular — rather than at the terminal Pun1/AT3 step. Pun1 labels the product and seals the capsaicin amide; it does not set the rate. The trunk is the governor, which is why a host's native capacity through its own phenylpropanoid relay is the first variable our model inspects. A T1 tomato, already a nightshade with spare phenylpropanoid capacity, says yes at four genes. A T4 grain, whose carbon is committed to filling its seed with starch, says yes only grudgingly and only with ten or twelve genes carefully throttled, because its budget must be re-negotiated line by line. The gene counts in the tier table are, at bottom, the price the flux model quotes.

And a transfer is not done when it works once; it is done when it works *every time*. Reproducibility is its own axis of quality, measured as the coefficient of variation — the standard deviation of batch output over its mean. A wild pepper is a wild number: its capsaicinoid output carries a CV near **42%**, which at a design target like Phoenix's predicted 2.95 million SHU puts the plausible batch range, at two standard deviations, from roughly half a million to well over five million SHU — a fruit that might scald or might barely warm, useless to anyone formulating to a spec. Our designed lines hold a CV of **≤7%**. At the same target that is a standard deviation near 206,000 SHU and a two-sigma band of about 2.53 to 3.34 million — a tolerance a formulator can actually build on, a heat you can promise. Measured against a wild line's near-doubling and near-halving, the engineered line is the difference between a flavor and a gamble. Precision is not a flourish on top of the transfer. It is part of what *transferred* has to mean.

Forty-two proofs, and sixty-four drawn

What makes the climb tractable rather than sixty-four separate miracles is exactly this modularity. We are not re-inventing a pathway per organism. We are moving a known stack of terminal verbs into a succession of hosts, each a calculated step farther from the pepper than the last, and re-solving the flux budget each time to fit the new host's economy. The five transferable modules are the leverage; the two carried verbs are the work; the budget is the constraint that decides whether a host can bear it. Count the verbs the destination lacks, deliver exactly those, balance the books, and grow it until the Index reads the same number twice. The forty-two designed hosts, sorted by the branch they carry, are the proof that the method generalizes — a whole botanical spread run through one shared trunk:

BRANCH · RECEPTOR	LEAD HOSTS (CONSTRUCTS)
Heat · TRPV1	Habanero (3,190), Pepper (1,060)
Cool-Burn · TRPM8	Peppermint (2,360), Lavender (160), Spearmint (80), Eucalyptus (80), Camphor (40)
Pungency · TRPA1	Black Mustard (2,790), Garlic (560), White Mustard (200), Ginger · Thyme · Watercress (120 ea.), Cinnamon · Oregano · Radish (40 ea.)
Umami · mGluR1	Tomato (2,040), Soybean (360), Tea (280), Yeast (240), Koji Mold · Shiitake (80 ea.), Mushroom (40)
Kokumi · CaSR	Garlic (560), Yeast (240), Onion · Vanilla (40 ea.)
Sweetness · TAS1R2/3	Stevia (910), Sugarcane (200), Monk Fruit (80)
Astringency · PRPs	Grape (1,610), Tea (280), Blueberry · Pomegranate · Sorghum (120 ea.), Raspberry (40)
Zero-Bitter · TAS2R38	Broccoli (380), Broccoli-Kale (280), Orange (120), Hops · Potato (80 ea.), Coffee · Cabbage · Kale · Rapeseed (40 ea.)

Several hosts appear under more than one branch — garlic carries both pungency and kokumi, tea both umami and astringency — because they are already multi-axis plants, their trunks feeding two borrowed verbs at once, which is the roster quietly demonstrating that branches stack in a host as readily as they stack in Phoenix. Forty-two organisms have passed that test, each profiled with our model across all eight axes, each grown toward the fullest flavor and the fullest nourishment it can carry without being starved to do it. Sixty-four more are drawn, tiered by nothing more exotic than how

many verbs away from a pepper they stand. The trunk is universal. Only the verbs are special. The whole frontier, stated without drama, is the arithmetic of counting them.

These are computational designs; Scoville Splice's cultivars and constructs have not been experimentally validated in a laboratory.

Chapter 18 · The Open Record — *provenance, consistency, and the respect of a reproducible number.*

A pathway drawn is a claim made, and a claim made is a debt owed — to the reader who will trust it, the grower who will plant it, the formulator who will build on it, the scientist who will try to break it. This closing chapter is about how that debt is paid: through provenance, through consistency, and through an open record that lets anyone, in principle, walk every route in this book and check that it runs. The reverence that carried us this far and the rigor that kept it honest meet here, at the unglamorous practice of keeping faith with a number. We have saved it for last because it is, in the end, the thing the whole library stands or falls on. A family can love a plant for a lifetime; the love becomes science only at the moment the plant does the same thing twice.



FIG. 18 — The research greenhouse at dusk. Every pathway in this book ends not in a diagram but here, in a living plant that has to grow the same way next season, and the season after, for the number on the label to mean anything at all.

The anatomy of a trustworthy number

Hold a single figure from this library up to the light — say the **predicted heat of approximately 2.95 million SHU** our model assigns to the cultivar Phoenix — and ask what it has to carry to be believed. It must carry a **quantity**, the thing actually counted, in units that do not drift. It must carry a **method**, a repeatable procedure anyone with the same instrument could run. It must carry a **provenance**, an honest account of what was changed in the plant to produce it. And

it must carry a **precision** — not just a value but the width of the cloud around it, because a number without an error bar is a rumor wearing a lab coat. Those four — quantity, method, provenance, precision — are the skeleton of every claim in these pages, and the rest of this chapter is simply the four of them, examined one at a time, until a reader can see straight through our numbers to the plants underneath. We write *we* because the computing was one pair of hands and the table was a whole family's; the discipline below belongs to both.

Consistency — the number that keeps its word

Begin with precision, because it is the one we have returned to most and prize most. A flavor you cannot reproduce is not a result; it is an anecdote. The instrument we use to track reproducibility is the **coefficient of variation**, the dimensionless ratio of a batch's standard deviation to its mean,

```
CV = σ ⁄ μ × 100 %,
```

and it is the quiet protagonist of the whole platform. Unimproved superhot material — the same cross, grown the same way, sampled fruit to fruit — runs at a CV of roughly **42 %**. Designed material runs at **≤ 7 %**. That gap is not a matter of tidiness; it is a matter of what a label is allowed to promise, and the arithmetic says exactly how much.

Take a batch whose true mean is a round one million SHU. At CV 42 %, the standard deviation is 420,000 SHU, and a 95 % interval — the mean plus or minus 1.96 standard deviations — spans from about **180,000 to 1,820,000 SHU**. A single fruit can land at a fifth of the label or at nearly double it. At CV 7 %, the same batch's 95 % interval narrows to about **860,000 to 1,140,000 SHU**, a window of ±14 %. Run the same calculation on Phoenix itself: a 7 % CV on a predicted heat design target of approximately **2.95 million SHU** carries a standard deviation near **205,700 SHU** and a 95 % band of roughly **2.53 to 3.34 million**; the untamed 42 % would have smeared that same fruit across **0.52 to 5.36 million**, which is to say it would have promised nothing at all.

The deeper reason to drive CV down is that consistency buys you certainty cheaply. To certify a batch's true mean to within ±5 % at 95 % confidence, the number of fruit you must measure is

```
n ≥ (1.96 × CV ⁄ m)², with m the relative margin you will accept.
```

At CV 7 % that is n ≥ (1.96 × 0.07 ⁄ 0.05)² ≈ **8 fruit**. At CV 42 % it is (1.96 × 0.42 ⁄ 0.05)² ≈ **272 fruit** for the identical promise. Because the sample size scales with the *square* of the CV, cutting variation sixfold — from 42 % to 7 % — cuts the measurement burden roughly thirty-sixfold. Low CV is not a cosmetic virtue bolted onto the science after the fact; it *is* the science, the part that lets eight careful measurements stand in for three hundred and separates a designed flavor from a lucky one.

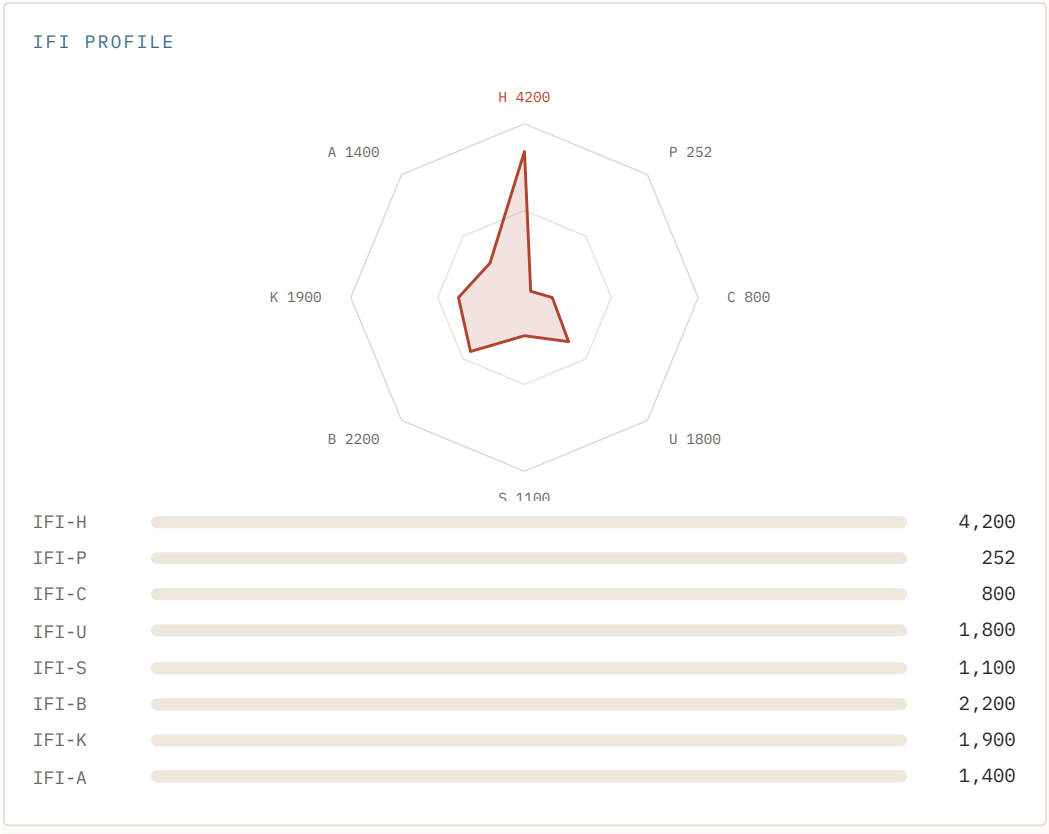


The controlled chamber where a cultivar is grown and characterized — same light, same water, same photoperiod, season after season, because a coefficient of variation is only honest if the environment that produced it was held still.

That single discipline runs under every count in the library, and we lock the counts before anything publishes, because across a shelf of books drift is the one thing that reads as sloppy.

WHAT WE LOCK	FIGURE
Batch-to-batch CV, designed	≤ 7 %
Batch-to-batch CV, unimproved superhot	~42 %
Confirmed host organisms	42
Sensory axes on one Index	8
Distinct construct designs	~225
Live and purchasable today	1,000
Total catalogue SKUs	78,800
Patents filed · peer-reviewed papers	3 · 21

None of those figures means anything unless the plant behind it grows the same way twice. The cultivar that most tests the claim is the one that registers on every axis at once, so it is the one whose full profile we put on the page and invite you to check — eight numbers, each anchored to a receptor, each held to a design specification of ≤ 7 % batch CV.



Reading the Index — from dose–response to one score

A number is only as trustworthy as the method that produced it, and the method behind every axis is a dose–response curve. Put a standardized preparation of a food in front of its receptor and the fraction of full activation it drives follows the Hill relation,

$$R / R_{max} = c^n / (EC_{50}^n + c^n),$$

where c is the effective concentration of the preparation, n is the Hill coefficient that sets the steepness, and $\mathbf{EC_{50}}$ is the concentration that drives the receptor to exactly half of its full response. For capsaicin at TRPV1 the fit returns a Hill coefficient near 2 — the channel opens cooperatively — and that single fact shapes everything downstream. With $n = 2$, the response climbs from 20 % to 80 % of maximum as the concentration moves from $0.5 \times EC_{50}$ to $2 \times EC_{50}$: a mere fourfold change in dose covers the entire useful working range. Below that window almost nothing is felt; above it the curve flattens and more molecule buys almost no more fire. The half-maximal point is where the curve carries the most information, and so it is the point we anchor to.

The **Iyengar Flavor Index** is that anchoring, made into a score. By stated convention — the same kind of defined threshold that underlies Scoville's original dilution-to-taste definition — a reference compound is assigned a score of 1,000 on its axis, and every other preparation is read against it:

$$IFI(candidate) = 1{,}000 \times EC_{50}(reference) / EC_{50}(candidate).$$

The whole of the Index is in the inversion. A *louder* flavor reaches half-activation at a *lower* concentration — less of it says more — so a smaller EC_{50} yields a larger IFI. Rearranged, the formula tells you precisely how much material a cultivar needs relative to the reference:

$EC_{50}(\text{candidate}) \,/\, EC_{50}(\text{reference}) = 1{,}000 \,/\, \text{IFI}.$

Work it on Phoenix, whose heat axis reads **IFI-H 4,200**. Then EC_{50}(Phoenix) = EC_{50}(reference) × 1,000⁄4,200 = **0.238 × reference** — it reaches half-maximal TRPV1 activation at under a quarter of the reference's concentration, which is the exact meaning of being 4.2× louder. Put it in physical units with an illustrative anchor (used here only to make the ratio concrete; the book's claims stay in canonical IFI and SHU): if a reference preparation reaches half-max at 1.00 mg of dried fruit per millilitre, Phoenix gets there at **0.238 mg/mL**. The same inversion runs on every axis against that axis's own native reference, which is how one cultivar earns eight independent scores.

One structural note belongs here before reading Phoenix's profile: each axis anchors its IFI to its own defined reference — capsaicin at TRPV1 for the H axis, AITC at TRPA1 for the P axis, and so on. Because each reference constant is a convention defined for that receptor, IFI scores from different axes sit against different anchors and are not on a single unified scale; they should be read as independent receptor-axis potency values, not summed or directly compared across axes.

Phoenix's full-spectrum profile, read as concentration ratios, is a portrait of a plant our model predicts speaks loudly on most axes — heat at 0.24× the reference material, zero-bitter at 0.45×, kokumi at 0.53×, umami at 0.56× — and softly on two: its cool-burn score of 800 and its pungency score of 252 each sit *below* the 1,000 anchor, meaning Phoenix is designed to need more than the reference concentration to half-activate TRPM8 and TRPA1. Full-spectrum means every axis carries a real, predicted signal; it has never meant every axis shouts.

AXIS	RECEPTOR	WHAT IT MEASURES	PHOENIX IFI
Heat · H	TRPV1	capsaicin heat and spice	4,200
Pungency · P	TRPA1	allicin / wasabi / cinnamon sharpness	252
Cool-Burn · C	TRPM8	cooling / warming thermomodulation	800
Umami · U	mGluR1	savory, brothy depth	1,800
Sweetness · S	TAS1R2 / R3	sugar-free sweetness	1,100
Zero-Bitter · B	TAS2R38	bitterness attenuation (20–65 %)	2,200
Kokumi · K	CaSR	richness, continuity, mouthfeel	1,900
Astringency · A	salivary PRPs	tannin / polyphenol mouthfeel	1,400

One honest footnote belongs under that table, because an open record names its own exceptions: the Zero-Bitter axis scores *suppression* of TAS2R38 signaling rather than its activation, so its reference and ratio are a formal analogue of the half-maximal reading, not a literal half-activation. We keep the vocabulary straight rather than let one tidy formula paper over a receptor that works the other way around.

The Index also inherits a deep genealogy, and naming it is part of provenance. The Scoville Heat Unit was itself a dilution-to-threshold: how far you must dilute an extract before a panel can no longer feel the burn. The IFI is the same idea made receptor-exact — a dilution ratio read at half-maximal response on a defined channel instead of

at the ragged edge of human detection — and each reference constant is a stated convention, defining IFI = 1,000 on that axis in exactly the way Scoville's dilution protocol defined the unit on the heat axis. That lineage lets us bridge to the older unit cleanly. Modern practice measures capsaicinoid *mass* by chromatography and converts: pure capsaicin carries a pungency of roughly 16 million SHU per gram, so each part-per-million of capsaicin-equivalent contributes about **16 SHU**. Invert that on Phoenix: our model predicts approximately 2.95 million SHU ÷ 16 ≈ **183,600 ppm**, or about **18.4 %** of dry mass as capsaicinoids. The ×16 factor assumes a capsaicin-dominant mixture; dihydrocapsaicin sits near the same potency, but nordihydrocapsaicin and the minor capsaicinoids are weaker, so the exact factor shifts a few percent with each cultivar's capsaicinoid fingerprint — a small honesty we would rather state than bury.

CULTIVAR	SKU	PREDICTED SHU	IFI-H	AXES	EC_{50} VS NATIVE (= 1000 ⁄ IFI-H)
Ember	SS-0087	866K	1,570	2/8	0.64×
Cinder	SS-0088	1.05M	2,290	3/8	0.44×
Kindle	SS-0089	1.25M	2,000	3/8	0.50×
Flare	SS-0090	1.48M	—	4/8	–
Blaze	SS-0091	1.72M	2,630	4/8	0.38×
Wildfire	SS-0092	1.96M	2,930	5/8	0.34×
Scorch	SS-0093	2.21M	3,220	5/8	0.31×
Firestorm	SS-0094	2.47M	3,490	6/8	0.29×
Solara	SS-0095	2.72M	3,980	6/8	0.25×
Phoenix	SS-0096	2.95M	4,200	8/8	0.24×

That table of design tiers carries two lessons an open record is built to deliver. The first is the blank. Flare has a predicted SHU and a designated axis count, but its IFI-H assay is not yet closed, so we leave the cell empty rather than interpolate a plausible figure — an honest gap is worth more than a confident guess. The second is the near-inversion between two neighbors. Kindle carries *more* total heat by the Scoville dilution (1.25M versus Cinder's 1.05M) yet scores *lower* on IFI-H (2,000 versus 2,290), needing more material to half-activate TRPV1 (0.50× versus 0.44×). SHU and IFI-H are not two names for one quantity: one is a panel's dilution-to-threshold over all capsaicinoids, the other a receptor's half-maximal ratio, and they can disagree when the capsaicinoid fingerprint shifts — when heat sits in forms the panel still counts but the channel weights a little differently, or when the fruit's matrix changes how the molecule

is presented. We do not file that discrepancy down. We print both numbers precisely because their disagreement is information the open record exists to surface.

Provenance — what, exactly, was changed

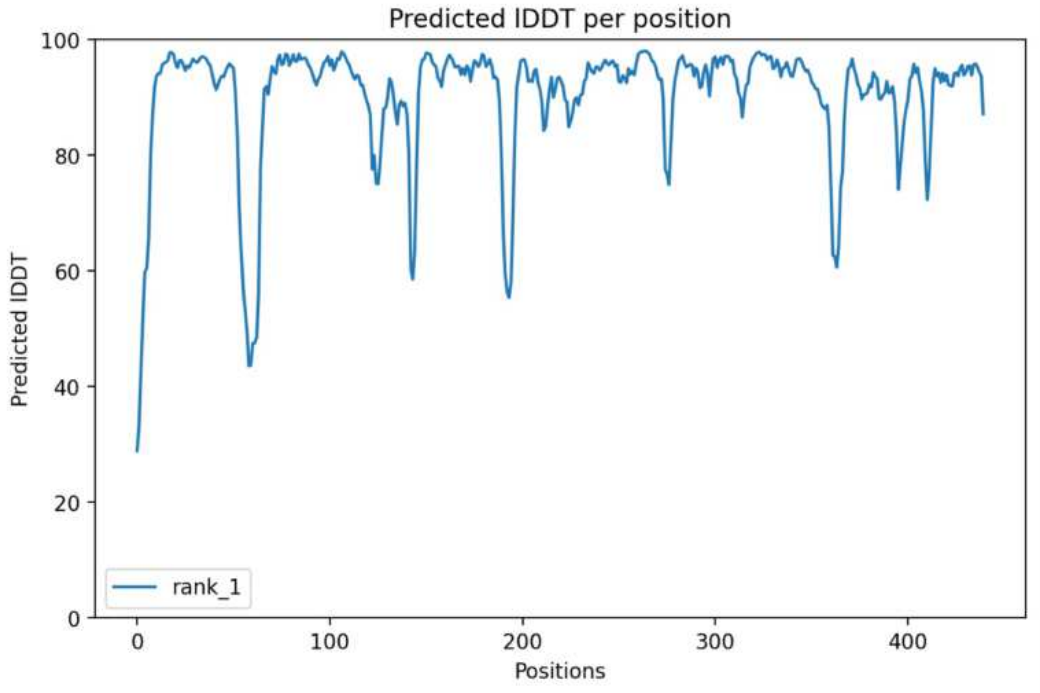
A method is only half of trust; the other half is a clean account of what was done to the plant. There is a real and load-bearing distinction, which we hold to carefully, between *editing* a plant's own genes and *transplanting* a route from one organism into another. Editing strengthens a native promoter, quiets a native off-ramp, or writes a precise substitution into the plant's own capsaicin synthase; transfer carries a whole pathway across a species boundary. The two carry different biologies and different regulatory weights, and the field's own vocabulary sorts them honestly. An **SDN-1** edit makes a small change with no template; an **SDN-2** edit writes in a precise change from a short template; an **SDN-3** edit inserts a foreign gene or cassette. In many jurisdictions the first two — changes confined to a plant's own sequence — are regulated differently from the third, so the category is not a marketing choice but a biological fact about what was done. We report which, every time.

The clearest worked case of provenance is the engineered heat at the top of that design tier, and it centers on two edits to a single enzyme. Capsaicin is built where two branches of metabolism meet. One branch runs the phenylpropanoid route shared by every vascular plant — **PAL → C4H → 4CL → HCT → COMT** — carrying phenylalanine down to the vanillyl group, which the aminotransferase **pAMT** converts to vanillylamine. The other branch, run by **BCAT** and the ketoacyl synthase **KAS**, builds a branched medium-chain acyl-CoA, 8-methyl-6-nonenoyl-CoA. At the junction sits **Pun1**, the capsaicin synthase (also called AT3), an acyltransferase that condenses the amine and the acyl-CoA into capsaicin's defining amide bond and releases free coenzyme A. Breeders chased this fire for ten thousand years by selecting the phenotype and hoping the genotype followed. We inverted the search: a 650-million-parameter protein language model scored every single-residue substitution in Pun1 by its predicted effect on catalytic efficiency × fold stability, while a genome-scale flux-balance model mapped the pathway's control distribution through Metabolic Control Analysis. That analysis found control sits upstream in the phenylpropanoid pathway — at 4CL and HCT/CCoAOMT — not at the terminal Pun1/AT3 step; Pun1 is the final capsaicin synthase, it labels the product, it does not govern the rate. Working within that map, two substitutions in Pun1, never before combined, emerged as the shipped construct.

EDIT	WHERE IT SITS · WHAT IT DOES	DOMINANT TERM, ΔΔG	EDIT CLASS
S39L	N-terminal; tightens substrate binding, lowering K_m	−2.1 kcal/mol (binding)	SDN-2
L345G	active-site cleft; steric relief for turnover, raising k_cat	−1.8 kcal/mol (catalysis)	SDN-2
C175S	removes a mis-pairing cysteine; raises the folded, active fraction — **structural annotation only; not in shipped construct**	−0.9 kcal/mol (fold stability)	–

Each shipped edit acts on a different term of the same enzyme's performance, which is why the screen optimized a *product*, k_cat / K_m, rather than any one quantity. S39L seats the substrate more tightly, lowering the Michaelis constant K_m that governs how much vanillylamine the enzyme needs to work at speed. L345G opens the catalytic pocket, lifting the turnover number k_cat. A third position, C175S, is a structural annotation from our modelling: it identifies a cysteine whose removal would raise the folded, active fraction — a ΔΔG of −0.9 kcal/mol shifts the folded-to-misfolded equilibrium by exp(0.9 / 0.59) ≈ **4.6×** toward the active form — but this substitution was retained as a modeling note rather than included in the shipped design. The free-energy changes for the two shipped edits make their contributions quantitative. At 25 °C, the thermal quantum RT ≈ 0.59 kcal/mol, so −3.9 kcal/mol of combined catalytic improvement implies an enzyme-level gain of order exp(3.9 / 0.59) ≈ **700×**.

And here provenance becomes humility, because the plant does not deliver 700×; our model predicts a realized fold-change of about **4.2×**, the ratio Phoenix's predicted IFI-H draws against the native reference. That gap is not an error — it is metabolic control theory doing exactly what it must. Our flux-balance analysis found that the control coefficients of the capsaicin pathway are dominated by upstream enzymes: 4CL and HCT/CCoAOMT together account for 93–99 % of total flux control, while the Pun1 step carries near-zero independent control coefficient over the Monte Carlo ensemble. Pun1 is the final step and labels the product; it does not govern the rate. This means that even a dramatically improved terminal synthase operates within the throughput the upstream steps allow — precursor supply from pAMT and BCAT, the phenylpropanoid flux through 4CL and HCT, vacuolar sequestration of a molecule that is toxic to its own maker. A 700× better synthase at a step with near-zero flux control coefficient does not multiply through to the plant; the upstream pathway continues to gate overall throughput, and the realized 4.2× gain reflects the full-system design — Pun1 variants together with upstream pathway tuning — not any single step in isolation. This is the whole reason the flux-balance model earned its place beside the language model: the enzyme is not the plant, and only a systems view predicts what the fruit will actually carry. We state the per-enzyme potential and the realized flux side by side, and we never let the larger, more flattering number stand in for the one you can taste.



Pun1 / AT3, the capsaicin synthase, modeled and coloured by pLDDT confidence — the binding domain and the active-site cleft resolve with high certainty, which is precisely why we felt entitled to redesign them.

The same provenance line sorts the rest of the library. Much of our restoration work — coaxing a tomato's own glutamate route back toward the savor the long-haul truck took, tuning a mint's five-gene menthol pathway back into balance — is editing of a plant's own sequence, SDN-1 or SDN-2. The far frontier, where a route is carried whole into an organism that never ran it, is transfer, SDN-3. That the phenylpropanoid backbone already turns in every vascular plant — only four terminal enzymes are pepper-specific — is what makes the method travel across 42 confirmed host organisms with 64 more mapped ahead; but travel is not the same act as tuning, and we say which is which every time, because a provenance stated plainly is the difference between a method a reader can trust and a spectacle a reader must simply believe.

Quantification you can show — the chromatogram under the claim

Behind every capsaicinoid figure, every glycoside and γ-glutamyl peptide, sits a chromatogram we can produce on request. The capsaicinoids are separated by **high-performance liquid chromatography** and quantified against authentic standards, which turns a peak into a concentration through a calibration line fit over a range of known standards,

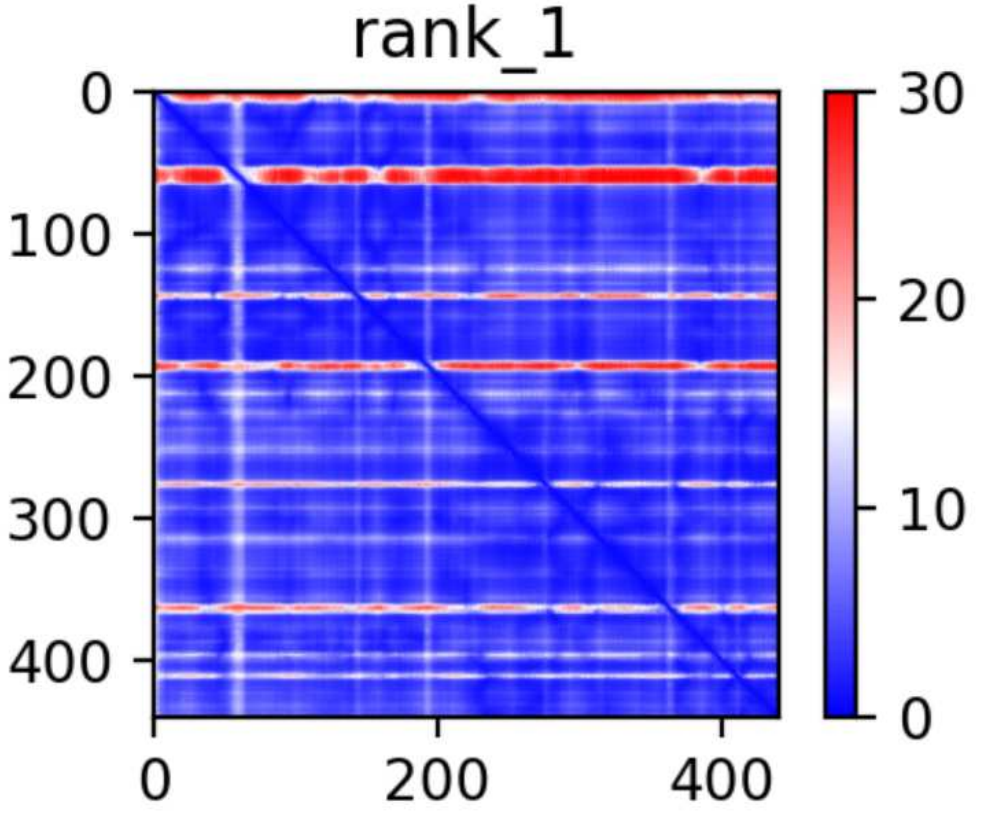
```
A = S·c + b,
```

where A is integrated peak area, c the standard's concentration, S the sensitivity (the slope), and b the intercept. Two figures of merit police the low end, by the standard convention LOD = 3.3 σ / S and LOQ = 10 σ / S, with σ the scatter of the response near blank. Illustratively, a run with slope S = 50,000 area-units per µg/mL and residual scatter σ = 8,000 area-units carries a limit of detection near 0.53 µg/mL and a limit of quantitation near 1.6 µg/mL — the honest floor below which we report *not detected* rather than a comforting zero. From a quantified extract the chain to the label is pure unit arithmetic: concentration in the vial, to total mass of capsaicinoid per

kilogram of fruit (ppm), to SHU through the ≈ 16-SHU-per-ppm bridge. Every computational prediction in this book is traceable along that method chain, and never to a hope.

The honest limits live here too, and we name them. HPLC sees what co-elutes as one peak unless the method is resolved to separate it, so a careless gradient can let dihydrocapsaicin hide under capsaicin and quietly bias the capsaicinoid fingerprint — the very thing that can pull SHU and IFI-H apart on a single fruit. Plant matrices suppress or enhance ionization and recovery, which is why quantitation rides on standard curves built in matrix and on replicated designs rather than single injections. And a chromatogram measures what is in the vial, not what reaches a receptor on a living tongue — the assay and the mouth are different instruments, and the IFI exists precisely to carry the claim the last step, from molecule present to loudness perceived.

The open record — a public kind of faith



The predicted aligned-error map for the same synthase — every pairwise confidence shown, not asserted. We publish the uncertainty because a model you cannot inspect is a model you cannot trust.

Which leaves the open record itself, the practice that makes a private laboratory keep a public kind of faith. The method of this book is meant to be *checkable*. The protein language model and the flux-balance model are tools anyone versed in the field can reason about and argue with. The structures are published with their confidence attached — the pLDDT that colours the fold and the

predicted aligned error that maps every pairwise uncertainty — because a confidence hidden is a confidence you are asking a reader to take on faith, and we would rather show the places the model is unsure than pretend it never is. **Three patents** stand filed and **twenty-one peer-reviewed papers** are deposited openly on Zenodo under a **CC BY 4.0** license, so the routes drawn here are not trade secrets locked in a vault but claims standing in the open, waiting to be tested, extended, or broken.

We are candid about where a number stops. A CV of ≤ 7 % describes designed material grown under held conditions; move the light, the water, or the photoperiod and the cloud widens, which is why consistency is a claim about a *system*, the plant and its environment together, not about a genome alone. An enzyme's predicted free energy is a hypothesis about one protein, and additivity across the two shipped edits is an approximation that the fruit, through flux control, is free to overrule — with upstream pathway enzymes at 4CL and HCT/CCoAOMT governing throughput while the synthase variants contribute catalytic precision at the terminal step. An IFI is anchored to a receptor in an assay, and the tongue is warmer, wetter, and more distractible than any assay; the Index narrows the gap between molecule and sensation, it does not abolish it. None of this is apology. It is the shape of an honest claim, which always names its own edges.

This is the quiet ethic a vegetarian, clean-living family brought to the work from the beginning, and the reason the book exists at all: that food is something you give the body with intention, that a plant is worth growing to hold the most it can, and that a number offered to another person is a form of respect — real only if the other person can check it. We drew every pathway in this library so you could follow the carbon yourself, phenylalanine to vanillylamine to the amide bond to the channel that reports a fire that is not there, and back — and find, we hope, that it runs exactly as we said. These are computational designs; Scoville Splice's cultivars and constructs have not been experimentally validated in a laboratory. That is the whole of it. Follow the carbon. The record is open. Come and see.