

D3 Chat-Log Analysis, Part 2: 10/22–10/23 (r2 results → r3 → r4 design)

Source: `margarita_d3_PAP_r1-r2 (2).md`. Same session as Part 1, extended from 10/21 07:16 to 10/23 07:27 UTC.

`L####` refs point into the updated log. Data checks use `d3_PAP_r1-r2-3-4_261023_IGOR-export.csv`.

0. Shape of the new portion

	Part 1 (10/20–10/21)	Part 2 (10/22–10/23)	Full session
Events	429	223	652
Messages (Margarita / IGOR)	103 / 147	66 / 80	169 / 227
Rounds covered	r2 design, first plate failure, redo set-up	r2 redo results → r3 design → r3 results → r4 design	r0–r4

The arc. 1. **10/22, r2 redo results.** IGOR found a systematic pattern of missing wells and excluded 4 outliers. The model retrained, and the r2 results overturned IGOR's own earlier "low-phosphate peak" hypothesis. 2. **r3 design.** IGOR designed the r3 plate (10/21 run), including a paired test of 1 vs 3 nM DNA. 3. **10/23, r3 results.** Yields were 35–60% lower than r2. IGOR separated a PAP-specific loss from an edge effect in Row L. 4. **r4 design.** Margarita and IGOR designed r4 with a protein (PMix) titration, tighter bounds and new stocks. 5. **Context overflow.** Around 04:00 UTC on 10/23 the session hit IGOR's context limit, and four messages went unanswered until a compaction hotfix was added by hand.

What changed in how she used IGOR. The plate-review work from Part 1 continued. New in this part: - **Diagnosing failed or odd runs** (r2 missing wells, the r3 PAP drop); - **Using the model's updated view to steer the next round**, including tighter bounds and coupled protein constraints; - **Pushing back on IGOR far more often**: "stop dropping conditions", "????", "you're still waiting for me to ask you to continue".

1. Where she got value (continues V1–V12)

V13. r2 QC: finding a systematic well-loss pattern, then salvaging the plate

- **Problem / goal:** "Any outliers in these data? ... I want to make sure there are no systematic issues in this dataset" (L14427).
- **Method / how:** IGOR used three tools on the r1020 data: `get_outliers` (the outlier test), `get_control_diagnostics` (replicate noise) and a well-by-well comparison against the planned 56-well plate map.
- **What happened:**
- **Missing wells:** IGOR found that only 26 of 56 wells had been imported. All of Row H and the centre columns of Rows G–J were missing, so several conditions had only one surviving well (L15093).

- **Outliers:** it flagged two clear replicate dropouts (L9: 0.71 vs 3.25–3.40 μM ; E6: 1.56 vs 2.67–2.81 μM) and one suspicious single well (K14). After retraining, the outlier test also flagged I18.
- **Plate alignment:** the controls lined up with earlier plates (CP0 3.16–3.30 vs 2.87–3.08 μM ; CKonly 2.10–2.32 vs 2.34–2.44 μM).
- **Result:** 4 wells were excluded and 22 of the 26 imported wells were kept for training (the story in V12).
- **What we learned:**
 - The engine's outlier test plus the agent's comparison against the plan caught both well-level failures and a plate-level dispensing pattern.
 - Note that the 22 of 26 counts imported wells; roughly 30 more were lost before import.

V14. Closing the loop: the model corrected its own hypothesis

- **Problem / goal:** retrain on r2 and decide where to look next.
- **Method / how:** IGOR retrained the model and compared the updated predictions with the earlier suggestions (L16024–L17030).
- **What happened:**
 - **The top suggestion failed:** before r2, the top suggestion (GES-10: low phosphate, high pyruvate) was predicted at 4.04 μM with a 56.2% hit probability. It measured 2.89 μM . After retraining, its hit probability fell to 7.7×10^{-12} .
 - **The best recipe firmed up:** the best measured recipe (Yadav_Val18_CP20) rose to a 60–69% hit probability.
 - **Hypothesis replaced:** IGOR said plainly that r2 "disproved [the] low-phosphate hypothesis", which was the basis for its own Part-1 picks (Expts 1–4). A new ridge emerged: 35–48 mM K-phosphate, 18–35 mM pyruvate, about 16 mM Mg, 18–20 mM CP.
- **What we learned:**
 - This is the Bayesian loop working as designed: a wrong prediction became information, and the model updated.
 - It is a stronger story than "IGOR picked winners". The agent-designed and engine picks from Part 1 were partly wrong, and the system corrected course in one round.

V15. Diagnosing the r3 plate-wide yield drop with pathway controls

- **Problem / goal:** "not many conditions are higher than CP/CK pathway alone ... everything looks lower yield overall. Help me assess this run" (L19494).
- **Method / how:** IGOR compared identical recipes across r1020 and r1021, separating three groups:
 - CP/CK only (CKonly: no PAP substrates);
 - PAP only (CP = 0);
 - PAP + CP.

It also checked within-plate position and the dispensed protein amounts. - **What happened** (verified against the export, wells with 1 nM DNA):

Pathway	r1020 median	r1021 median	Change
CP/CK only	2.28 μM	1.90 μM	–17%
PAP only (CP = 0)	3.06 μM	1.30 μM	–58%

Pathway	r1020 median	r1021 median	Change
PAP + CP	2.87 μ M (best 3.99)	2.21 μ M (best 2.53)	−23%

- **Reader ruled out:** a reader or calibration shift would have lowered CKonly by the same amount, so IGOR ruled it out.
- **Row L edge effect:** IGOR found that every Row L well read 15–35% below its replicates.
- **Dispense note:** it later surfaced Margarita's own plate-map note: "changed the MM reload pressure halfway thru ... saw some liquid on the bottom of the head" (L24669).
- **Plate offset handled:** IGOR noted that the categorical DNA ID input would absorb the r3 plate offset, and recommended making Run name a model input.
- **What we learned:**
- The pathway controls Margarita insisted on keeping turned a confusing plate into a specific diagnosis: the loss is in the PAP module, not the system as a whole.
- **Caveat:** IGOR attributed the loss to enzyme inactivation ("smoking gun"). But every r1021 well also used a new DNA prep (10/20, 60 nM), and Fig. 4a showed that the plasmid effect is specific to PAP. The DNA prep is an equally plausible cause and was not weighed (see F15).

V16. A clean, decisive DNA-dose result

- **Problem / goal:** decide whether more template DNA helps; Margarita wanted to test 3 nM on a few conditions.
- **Method / how:** IGOR designed matched 1 nM vs 3 nM pairs across five pathway conditions on the same plate.
- **What happened:** 3 nM DNA cut yield by 37–48% and shortened reaction lifetime by about 1 h in every pair: CKonly, CP0, ctr, Mg16 and SY-V18-c20 (L20117).
- **What we learned:**
- A paired, within-plate design gave an unambiguous answer in a single round, and DNA was fixed at 1 nM for r4.
- This is the cleanest experimental result in the log, and it holds despite the plate-wide PAP drop because every comparison is within the plate.

V17. Turning the optimizer around: feasible, focused suggestions

- **Problem / goal:** too many suggestions were unprintable or extreme. Earlier batches had 3 of 5 at 70–100 mM CP or 5 mM pyruvate (L17009).
- **Method / how:**
- IGOR set `max_similarity = 0.9`, a fix it had proposed in Part 1 (F8).
- It proposed tightening the bounds (Mg 14–22, CP 20–32, K-phosphate 30–55, pyruvate 12–38 mM).
- It added constraints that lock the four proteins to move together, since they come from one tube (L23130).
- **What happened:**
- All 5 suggestions in the next batch fitted the 2 μ L well, with predictions of $3.5\text{--}3.7 \pm 0.2\text{--}0.3 \mu\text{M}$.
- After the protein bounds opened, GES-18 reached a 79.7% hit probability, the highest in the workspace.
- **What we learned:**

- Choosing the bounds is where the scientist's knowledge and the engine meet. Once the bounds reflected what the bench could actually make, the suggestions became usable.
- **Caveat:** this same batch also exposed engine defects (F14).

V18. Designing experiments that separate mechanisms

- **Problem / goal:** test whether protein (PMix) level limits yield, which was IGOR's leading explanation for r3.
- **Method / how:** IGOR proposed a 0.8×/1.0×/1.2× titration of the protein tube. It is run on the CP/CK-only control as well as on PAP conditions, so the effect of the base protein mix can be separated from that of the PAP enzymes.
- **What happened:**
- **Agreed design:** the r4 plate includes the titration on both controls, plus 1.2× on the top recipes.
- **Model scope:** IGOR explained that the model had only seen protein levels of 1.74–1.91 mg/mL, so it cannot predict outside that range. That is why the titration has to be measured.
- **What we learned:** IGOR designed controls that make the result interpretable, not just optimization points.

V19. Making the plan physically runnable (continues V1)

- **Problem / goal:** fit 20 conditions × 3 replicates into 2 µL wells, given the 9.8 nL droplet size, one protein tube and the reader's block geometry.
- **Method / how:**
- IGOR moved the Mg and CP base amounts ("floors") into the master mix.
- It re-optimized the stocks (Mg 500 mM, pyruvate 1200 mM, K-phosphate 1500 mM).
- It diluted the DNA stock so 1 nM is 3–4 droplets rather than 2, after Margarita raised that concern.
- It generated the plate map.
- **What happened:**
- **Kept the control runnable:** IGOR caught that putting the K-phosphate and pyruvate floors into the master mix would make the CP/CK-only control impossible to dispense. Margarita adopted the fix.
- **Plate map checked:** the final map has 20 conditions, each at n = 3, with no condition repeated in a row or column and a randomized ladder in the flanking columns (verified independently).
- **Errors caught:** IGOR caught a 1 nM vs 3 nM DNA mismatch on the reagents tab, duplicate rows, and a stray backtick character that would have broken the instrument file.
- **What we learned:** hardware-aware design is now the default mode of collaboration, but the arithmetic still runs in prose and has slipped several times (F12, F17).

V20. Turning an outside document into a plate decision

- **Problem / goal:** decide what to take from a 10-experiment portfolio PDF and a later engineering summary.
- **Method / how:** IGOR mapped each proposed run to droplet counts, the volume budget, and the conditions already planned.
- **What happened:**
- Two runs were identical after droplet rounding.
- Two overflowed the well by 250–300 nL, because they came from an old goal of a completion time of 7 h or more.

- Three fitted and were new.
- IGOR advised deferring all of them, and Margarita agreed: "if not, I'll wait to run them next time."
- **What we learned:** IGOR is an effective filter for design proposals from other sources.

2. Where it fell short (continues F1–F10)

#	Issue	Evidence	Why it matters
F11	The session broke at the context limit. Four messages (03:59–04:15, including two "try again") got no response until a compaction event was added by hand at 04:55.	L22462–L22481	A long, high-value session is exactly how this product is meant to be used, and it hit a hard wall. Automatic summarizing of older context needs to ship.
F12	Recipes changed silently. GES-13_Pmix1.2x was written as 48 mM K-phosphate in the instrument file while the conditions sheet said 55 ("????", L25263). In one CSV, the benchmark SY-V18-c20 was changed from 35 to 38 mM K-phosphate (L24220).		Any change to a recipe must be shown and approved. This breaks the "every write is proposed and approved" principle.
F13	Imposing a limit the user didn't set, and reversing when challenged. IGOR kept cutting the plate to 19 conditions after she said she could run more ("stop dropping conditions ... I'm not limited to 19", L23071). It reversed on the 0.8× protein condition for GES-5, on 1000 vs 2000 mM K-phosphate, and on GES-18 (keep, then drop, within two messages).		This is the same pattern as F2. Stated constraints should persist in workspace memory, and recommendations should hold steady unless new evidence appears.
F14	The optimizer returned suggestions that broke active constraints. GES-19 and GES-21 violated the protein-coupling constraints. Two other suggestions (GES-14, GES-15) duplicated other suggestions by switching ribosome lot, a category nobody had fixed, which got them past the diversity filter. GES-12, generated with DNA fixed at 3 nM, went to extreme corners.	L22373, L24092	This is an engine bug: the optimizer must enforce constraints. Categories should default to the values the user set, and the diversity filter should ignore category-only differences.
F15	Diagnosing too confidently, again. IGOR called r3 a "smoking gun" of enzyme inactivation and leaned on a model sensitivity to protein level (+5 to 7 μM per mg/mL), even though protein level had varied only through dispense jitter. That is the same confound it correctly flagged for ribosomes in Part 1 (V7). It did not weigh the new DNA prep, and it projected that fresh enzymes would exceed 4.0 μM without support.	L20078, L22611	Present these as competing hypotheses with a test for each. The r4 titration is the right test for protein level; the DNA prep needs one too.
F16	A promise it couldn't keep. IGOR said it would continue automatically after each card approval; it didn't ("you're still waiting for me to ask you to continue", L20433). It recovered by batching 7 cards.		Either support auto-continue after an approval in the app, or never promise it.
F17	Plate layout and arithmetic in prose, again. Visible "Wait—" corrections on the plate map (a row with 8 reaction cells, repeated columns) and on volume budgets.	L24669–L25154	Reinforces F3: plate-map generation and volume checks belong in deterministic tools.
F18	Exclusion calls shifted. IGOR first said to keep I18 as "real biology", then excluded it after retraining; it was the only measurement of HI-CP-lowPi. It excluded all of Row L, including wells only about 15% low; Margarita's plate-map note later supported the edge effect.	L15126, L16036, L20887	Exclusion rules should be explicit and consistent, and should be logged with the reason.

#	Issue	Evidence	Why it matters
F19	The misattribution persisted. IGOR kept describing r1007 as "the Master Mix flatline/inhibition issue" (L18524), so the uncorrected memory convention from V2 kept being reused.		Shows the cost of saving an unsupported conclusion to memory (see V2, Fig. 4a).

Positive notes on interaction: - When asked, IGOR switched to Simplified Technical English and saved that preference to memory (L20241). - Margarita's corrections were consistently right and were taken up quickly: the evaporation and Mg-floor constraints, droplet feasibility, the 55 vs 48 mismatch, and conditions she wanted kept.

3. What to highlight (additions to Part 1, section 4)

D3 devnote - The $r2 \rightarrow r3 \rightarrow r4$ sequence, as it happened: - r2: GES-10 failed, the model updated, and the low-phosphate hypothesis was retired. - r3: the DNA-dose result was clean; the plate-wide PAP drop is reported openly along with its candidate causes. - r4: built on the r3 findings (DNA fixed at 1 nM, the protein titration, tightened bounds). - The pathway-control diagnosis (V15): it shows why including CP/CK-only and PAP-only controls on every plate matters.

Customer case study - "One round to an unambiguous answer": 3 nM DNA cut yield by 37–48% across all five pathways (V16). - "Feasible by construction": after the bounds were tightened, 5 of 5 suggestions fitted the well, compared with 2 of 5 before (V17).

Investors / Anthropic - The loop corrects itself (V14): the model's top pick failed, and one retrain moved the search to the right region. This is the honest version of the benchmark story. - Product roadmap items driven by real use: - automatic context management (F11); - write-gating for recipe edits (F12); - enforcing constraints inside the optimizer (F14); - deterministic plate and volume tools (F17).

4. Open questions and actions

1. **Before r4 runs (time-sensitive):** the r4 setup sheet introduces another new DNA prep (pET21a-mScarlet 10/21a, 451 nM stock, diluted to 60 nM). Two facts point to the DNA prep as a possible cause of the PAP-specific loss: - the r3 drop was PAP-specific and coincided with a new prep; - Fig. 4a showed plasmid identity matters only in PAP reactions.

Suggestion: add ctr and SY-V18-c20 wells made with the original 10/14 prep (or the 10/20 prep) to separate the effect of the prep from that of the protein titration. 2. **Workspace memory:** correct the protein-dispense convention, which still credits protein separation for the PAP recovery (V2, F19). 3. **Export versus workspace:** the 10/23 export shows r1021 with 37 active wells and no exclusions, while the log shows IGOR excluding 9 r1021 wells. Was the export taken before those exclusions? 4. **r4 results:** needed to test V18 (protein titration) and the claim that fresh protein restores r2-level yields.