



Analysis of Recombinase Polymerase Amplification Products by Lateral Flow Detection

Abstract

RPA read on a lateral flow strip gives a 20–30 minute, instrument-free nucleic acid test that runs at 37–42 °C and is read by eye. The amplicon is built to carry two haptens, one per end: gold conjugate in the strip binds the first, an immobilised capture reagent at the test line binds the second, so only a double-labelled product makes a line. What reaches the strip matters as much as what the reaction makes — gp32 stays bound to the amplicon and fouls the membrane, so product is diluted or, better, purified first.

What follows covers the chemistry of both halves, primer and probe design that keeps artefacts off the test line, a bench protocol, clean-up, interpretation and performance limits, and the two failure modes that decide whether the assay survives outside a laboratory: primer-dimers and amplicon carryover.

Why RPA with a lateral flow readout

The bottleneck in decentralised testing is readout, not amplification. PCR needs a thermal cycler and, for closed-tube detection, an optical channel — both costly and power-hungry relative to the test itself.

RPA removes the cycler: isothermal at 37–42 °C, which body heat or a chemical heater supplies, with detectable product in 5–20 minutes — faster than LAMP, NASBA or helicase-dependent amplification. Detection is what it leaves behind. Gels are lab-bound and smear in the reaction's protein and crowding-agent load; real-time fluorescence needs a reader.

A lateral flow strip closes that gap. Diluted product is applied and capillary flow carries it through a conjugate pad and a nitrocellulose membrane striped with a test and a control line. Within 5 minutes there is a coloured line or there is not, with no instrument.

The two fit chemically as well. RPA tolerates 5' haptens, internal abasic sites and 3' blocking groups, so the amplicon is labelled during amplification rather than after, and it tolerates crude samples — blood, plant lysates, swab eluates — better than Taq, which keeps extraction as simple as the readout.

The trade-offs occupy most of this article: RPA amplifies primer artefacts readily, strips are binary and near-non-quantitative, and the tube is opened after amplification.

How RPA generates the amplicon

RPA replaces thermal melting with a recombinase system borrowed from T4 bacteriophage DNA repair. Three protein activities do the work.



Component	Typical identity	Role
Recombinase	T4 UvsX, with UvsY loading factor	Forms a nucleoprotein filament on each primer and searches duplex DNA for the homologous site, then invades it
Single-strand binding protein	T4 gp32	Stabilises the displaced strand of the D-loop so it cannot re-anneal
Strand-displacing polymerase	Bacillus subtilis Pol I large fragment (Bsu)	Extends the invading primer from its 3' end, displacing the non-template strand
Crowding agent	PEG 35000, ~5%	Concentrates the macromolecules and drives filament assembly and product formation
Energy and cofactors	ATP, phosphocreatine, creatine kinase, dNTPs, Mg(OAc) ₂	Powers recombinase cycling and synthesis

The cycle is continuous, not stepwise: a primer–recombinase filament invades, the polymerase extends, the recombinase disassembles as ATP is hydrolysed, and both strands become available again. With no denaturation step gating it, growth is exponential from the start and usually complete in 5–20 minutes.

Five practical consequences follow from this mechanism.

- **Magnesium starts the clock.** Magnesium acetate is added last, normally to the tube lid, and the reaction begins on mixing. Reactions assembled on the bench will creep before incubation.
- **The reaction is viscous and protein-rich.** gp32 remains bound to the amplicon when amplification ends, and PEG, recombinase and polymerase travel with it, so the product cannot go onto a strip as it stands. Dilution or clean-up is a required step, not an optional one.
- **Mixing mid-reaction raises yield.** A brief vortex and spin at 3–5 minutes measurably improves sensitivity in most published assays, because filament assembly is diffusion-limited in a crowded mixture.
- **Short amplicons win.** Strand displacement over long templates is slow; 100–200 bp is the practical sweet spot and performance degrades beyond roughly 500 bp.
- **RNA needs a reverse transcriptase.** RT-RPA simply adds a thermostable RT to the same tube; no separate cDNA step is required.

The mechanism also explains RPA's main weakness. The recombinase loads on any primer and invades any partially complementary sequence, including the other primer, so artefacts amplify as efficiently as target. A strip cannot tell them apart unless the labelling scheme is designed to.

How the strip detects the amplicon

A nucleic acid strip is an immunoassay run on a DNA analyte: it detects the haptens, never the sequence. Four overlapping pads sit on a backing card — sample pad, conjugate pad holding dried gold conjugated to an antibody against the first hapten, nitrocellulose membrane striped with test and control lines, and an absorbent wick that pulls flow.



The test line carries a reagent against the second hapten, usually streptavidin against biotin. A single-labelled molecule picks up gold but is not retained; an unlabelled one does neither. Only a doubly labelled product bridges gold to the test line. The control line captures the conjugate itself and confirms only that flow occurred and the gold was active.

Three states are worth distinguishing. A dual-labelled product binds gold through its FAM label and is held at the test line by streptavidin, giving a visible line. A biotin-only artefact, such as a primer-dimer under the probe scheme, is held at the test line but carries no gold, so it leaves no colour. A FAM-only fragment carries gold but is not captured, and flows on to the wick.

Three labelling schemes are used, and the choice determines how trustworthy the test line is.

Two labelled primers. One primer carries 5' biotin, the other 5' FAM or digoxigenin, so every amplicon is dual-labelled by construction — and so is every primer-dimer joining the two, which gives a full-strength test line. Use it only where dimer formation has been excluded empirically.

Probe-based (info) detection. The standard for commercial kits and the scheme to default to. A 46–52 nt probe carries a 5' label (FAM or FITC), an internal tetrahydrofuran (THF) abasic site about 30 nt in, and a 3' block such as a C3 spacer, so it cannot be extended as supplied. On hybridising within the amplicon, endonuclease IV (Nfo) cleaves the THF — it cuts only in duplex — removing the block and turning the probe into an extendable FAM-labelled primer. The reverse primer carries 5' biotin. A dual-labelled product therefore requires an internal sequence match, which dimers cannot provide.

Hybridisation after amplification. The amplicon is made with one biotinylated primer, then denatured and hybridised to a labelled probe before running. This adds a sequence check and allows several targets from one reaction, but it adds steps and opens the tube twice. On-strip hybridisation kits (AU2061) and barcode strips (AU2065) package the format ready-made.

Two details matter at the bench. The product cannot be applied neat — it must be diluted 1:10 to 1:50 or purified, for the reasons below. And the gold is finite: too much dual-labelled product saturates it, giving a weak line at high target.

Dilution or purification before the strip

This step, not the amplification, decides whether replicate strips agree. The raw reaction carries ~5% PEG, milligrams per millilitre of protein, and free labelled primer and probe. The crowding agent distorts the flow front; gp32, the main source of artefacts, stays bound to amplicon and free oligonucleotide, masks the haptens and aggregates gold at the application point, giving a smear at the origin and weak or absent test lines; and unincorporated label competes for both the conjugate and the test line. Each varies between tubes, so inconsistently diluted product gives intensities that do not track concentration.

Two routes fix it, and the choice is a trade between reproducibility and field practicality.

Approach	Typical handling	Reproducibility	Cost
Dilution in running buffer	1:10 to 1:50, fixed and pipetted accurately	Moderate; gp32 bound to the amplicon is diluted, not removed	None; keeps the assay instrument-free
Column or magnetic-bead clean-up	Silica column or SPRI beads, elute in 20–50 µL	Best; removes PEG, SSB and free labelled oligonucleotide	Adds 10–15 min, consumables, and a second tube opening

Purify wherever the workflow allows. Dilution lowers the SSB concentration but does not strip protein off the amplicon, so gp32 artefacts scale down rather than disappear; a column or bead clean-up removes protein and free label together. Expect a sharper line, a clean origin, and a visual limit of detection better by a factor of a few. For development, and for any comparison of intensities, purification is the default, not the fallback.

For deployed tests, dilute — but treat the factor as a fixed assay parameter: validate one, keep one buffer lot, pipette it the same way every time. A 1:10 that suits one primer set may blush for another with a heavier dimer load. Where dilution is marginal, 1:50 with a two-minute stand before dipping usually restores consistency.

Designing primers and the nfo probe

RPA primer design does not follow PCR rules. Melting temperature is nearly irrelevant, because no annealing step is thermodynamically gated; length, 3' behaviour and the absence of self-complementarity carry the weight.

Parameter	RPA-LFA target	Why
Primer length	30–35 nt	Shorter primers form unstable recombinase filaments; longer ones add dimer risk
Amplicon length	100–200 bp, ≤ 500 bp	Strand displacement slows over distance
GC content	40–60%	Extremes destabilise invasion or promote secondary structure
3' terminus	Avoid long G/C runs; no self-complementarity	Dictates extension and dimer initiation
5' terminus	G avoided; hapten attached here	Protects label accessibility
Probe length (nfo)	46–52 nt	Needed for stable duplex formation around the THF
THF position	≥ 30 nt from the 5' end, ≥ 15 nt from the 3' end	Leaves a stable primer after cleavage and a stable block before it
Probe 3' end	C3 spacer, biotin-TEG or ddC	Prevents extension of uncleaved probe

Design 4–6 candidates per region and screen them combinatorially rather than trusting prediction. An eight-combination screen against a dilution series and a no-template control is the normal first bench day, and one or two usable pairs is a normal yield.

The no-template control is the criterion that matters most. A pair giving a clean line with template and a faint one without has failed, not nearly passed: dimer lines do not improve with optimisation of time or temperature, because the artefact amplifies by the same mechanism as the target.

Three mitigations are worth trying before abandoning a dimer-prone pair: shorten the primers to 30 nt, which often removes the complementary 3' stretch; shift one primer 5–10 nt along the target; or move to the nfo scheme, where a dimer carries only biotin and so is captured but invisible. The probe scheme does not stop dimers forming — it stops them being read.

For multiplexing, keep the biotinylated reverse primer common where biology allows and give each target its own probe label, with a matching line on a multi-line strip. Beyond two or three targets



the conjugate chemistry and the strip real estate run out, and a reader-based or CRISPR readout is the better route.

A representative protocol

The following is a typical 50 μ L nfo-probe reaction read on a generic FAM/biotin strip. Volumes follow freeze-dried pellet formats; adjust to the kit in use.

1. **Reconstitute.** Prepare a master mix of rehydration buffer (29.5 μ L per reaction), forward primer and biotinylated reverse primer (2.1 μ L each of 10 μ M), FAM-labelled nfo probe (0.6 μ L of 10 μ M), and nuclease-free water to 47.5 μ L less the sample volume. Add to the lyophilised enzyme pellet and mix until dissolved.
2. **Add sample.** Add 1–5 μ L of extracted nucleic acid or crude lysate. For RNA targets, include reverse transcriptase in the master mix.
3. **Start with magnesium.** Pipette 2.5 μ L of 280 mM magnesium acetate into each tube lid, close, invert and spin briefly. The reaction starts here; move all tubes to the incubator together.
4. **Incubate.** 39 °C for 20 minutes. Remove, vortex and spin at the 4-minute mark, then return.
5. **Prepare the product.** Purify on a silica column or with SPRI beads and elute in 30 μ L; this removes gp32, PEG and free labelled oligonucleotide, and gives the most reproducible strips. Where a clean-up step is not practical, transfer 2 μ L of product into 98 μ L of strip running buffer instead (a 1:50 dilution), mix by pipetting without vortexing, and hold that factor constant across every run.
6. **Run the strip.** Dip the sample pad into the diluted product, keeping the liquid below the conjugate pad. Read at 5 minutes.
7. **Record.** Photograph the strip under consistent lighting against a neutral background at the stated read time, before the lines dry and shift in intensity.

Controls to run on every batch, not just during development:

- **No-template control** — water in place of sample. Any test line here invalidates the run and usually indicates dimers or contamination.
- **Positive control** — a plasmid or synthetic fragment at roughly 10 $\times$ the limit of detection, confirming enzyme and strip activity. A pre-labelled control amplicon (such as the FAM/biotin 23S fragment, NA2032) applied directly to a strip additionally separates strip failure from reaction failure.
- **Internal amplification control** — a second target (a host housekeeping gene, or a spiked synthetic template) on its own test line, which distinguishes a true negative from an inhibited reaction.
- **Extraction control** — a sample processed from the start, when the extraction method is itself new.

Reading the strip and bounding performance

Four-line patterns are possible, and the one to read carefully is a test line with no control line.



Control line	Test line	Interpretation
Present	Present	Positive — dual-labelled product captured
Present	Absent	Negative, valid only if the internal amplification control passed
Absent	Present	Positive, if the test line is strong: the conjugate was consumed at the test line. Confirm on a further dilution. A weak test line here means flow or conjugate failure
Absent	Absent	Invalid — strip did not run; repeat

A missing control line is not automatically an invalid run. The control captures gold that was not retained upstream, so a very strong test line can strip the conjugate out of the flow before it gets there. With a strong test line, read it as positive and confirm on a 10-fold dilution, where both lines should appear. Excess free labelled oligonucleotide does the same from the other direction, competing for gold and driving the hook effect — one more reason to clean the product up first.

Read at the stated time, not later. Lines continue to darken for several minutes, and a strip judged at 20 minutes will show faint bands that were absent at 5 and that correlate poorly with target presence.

Published assays typically reach 1–100 copies per reaction — qPCR-comparable for well-designed targets, within one to two logs for most — in 25–35 minutes end to end. Specificity comes from the primer and probe design, not the strip, which reads haptens only.

Quantification is the weak point. Intensity tracks concentration over one order of magnitude at best, compressed at both ends: the conjugate saturates high, and exponential kinetics erase input differences by plateau. Treat a strip result as qualitative unless the assay has been calibrated. A graded series of labelled controls — the purpose of a stoichiometry QC kit such as AU2063 — maps the working range against the gold load, the cheapest way to find the saturation point before real samples do.

For semi-quantitative output there are three options, in increasing order of effort: shorten the incubation so the reaction is read before plateau, which widens the dynamic range; use a handheld reflectometric reader, optionally with a fluorescent label — quantum dot, europium or R-phycoerythrin — in place of gold; or partition the reaction and count positives, which gives true quantification but costs the instrument-free property that motivated the format.

One artefact to recognise on sight: a diffuse blush across the membrane rather than a discrete line indicates insufficient dilution of the reaction; increase the running-buffer dilution to 1:100 and re-run. The other, the hook effect, is worth setting out in full.

The hook effect

The hook effect — high-dose hook, or prozone — is a false negative or weak positive caused by too much analyte. Signal rises with concentration to a maximum, then falls, so the curve hooks down. A sample a hundred-fold above the limit of detection can read fainter than one near it; far above, it can read negative.

The cause is fixed binding capacity against unbounded amplification output. Three saturations contribute, usually all at once.



- **The gold conjugate is consumed.** Every anti-FAM paratope on the dried gold is occupied before the sample has finished rehydrating it. Dual-labelled amplicon arriving afterwards travels naked: captured at the test line, but colourless.
- **The test line is blocked.** Streptavidin at the test line is a finite number of sites. Free biotinylated primer and biotin-bearing fragments vastly outnumber full-length amplicon and occupy those sites.
- **Labelled species compete in solution.** Unincorporated FAM probe and primer bind gold as well as the amplicon does, and at primer concentrations — 240–480 nM against sub-nanomolar amplicon — so most of the conjugate leaves the pad bound to oligonucleotide that will never be captured.

The signs: a test line that weakens as target increases, a weak or absent control line beside a strong test line, and a non-monotonic dilution series with its strongest line in the middle of the range.

Diagnosis takes one extra strip: re-run at a further 10- to 100-fold dilution. If the line strengthens, the original was in the hook.

Four things keep assays out of it: clean the product up, which removes most of the competing free label; keep primer and probe at the low end of what still amplifies; fix a dilution validated against a dilution series rather than one that happened to work on a weak sample; and map the strip's working range once per lot against a graded series of labelled controls.

Where sample concentrations range widely, running two dilutions in parallel — say 1:10 and 1:1000 — costs one strip and removes the ambiguity.

Troubleshooting

Observation	Likely cause	What to do
Test line in the no-template control	Primer-dimer amplified and dual-labelled	Redesign or shorten primers; move to the nfo probe scheme so dimers carry only one label
Test line in every well, including water, appearing suddenly in a working assay	Amplicon carryover	Physically separate pre- and post-amplification areas; decontaminate with 10% bleach; discard reagent aliquots
No test line with known positive template	Reaction failed, or strip chemistry failed	Check the positive control and the control line; if the control line is also absent, the strip or the dilution buffer is at fault
No test line, control line fine, positive control fine	Inhibition by the sample matrix	Dilute the sample 1:10 before amplification, or improve extraction
Weak line at high target, strong at lower dilutions	Conjugate saturation (hook effect)	Dilute the product further and re-read
Smear or blush instead of a line	gp32 and PEG carried onto the membrane	Purify the product, or increase the running-buffer dilution to 1:100
Lines fade or shift after reading	Read outside the time window	Fix the read time and photograph immediately



Observation	Likely cause	What to do
Intermittent failures across a batch	Magnesium added unevenly, or reactions started at different times	Add magnesium to lids, close all tubes, spin together, incubate together
Line intensity varies between replicate reactions	Variable carryover of gp32, PEG and free labelled oligonucleotide into the strip	Purify every sample the same way; if diluting instead, fix one factor and one buffer lot and pipette identically

Carryover deserves separate emphasis. Each reaction produces roughly 10^{12} amplicon copies, and the strip format requires opening the tube and pipetting that product on an open bench. One aerosol event contaminates reagents, pipettes and surfaces, and the false positives that follow are indistinguishable from real results. Three defences work: physically separate pre- and post-amplification workspaces, with their own pipettes and coats; closed cartridges in which the strip never leaves the reaction chamber; and single-use reagent aliquots, so contamination costs one run rather than a batch. dUTP/UNG decontamination is compatible in principle and worth considering at high throughput.

Sourcing the components

Most groups building an RPA-LFA assay buy the strip side rather than make it. The catalogue below is [Attogene's](#), mapped to the steps described above.

Workflow step	Product	What it provides
Readout, labelled-primer or nfo-probe format	AU2034 Universal Lateral Flow Assay Kit (anti-FAM/FITC gold)	Dipsticks with anti-FAM gold conjugate for biotin + FAM amplicons; 2-, 3- and 4-line versions for multiplexing, with running buffer, wells and dual-labelled controls
Readout, streptavidin-gold format	AU2034 Universal Lateral Flow Assay Kit (Strep-Gold)	The inverted architecture: gold carries streptavidin and the test line captures the second hapten
Post-amplification hybridisation	AU2061 On-Strip Hybridisation NALF Starter Kit	Immobilises a biotinylated nucleic acid on nitrocellulose and runs a FAM-labelled probe against it, for adding a sequence check to an otherwise unlabelled amplicon
Multiplexed or barcoded detection	AU2065 NALF Barcode Test Strips	Capture-probe strips with a dedicated hybridisation running buffer, for several targets on one strip
Sample preparation	NA2012-01 Plant and Algae DNA and NA2014-01 RNA Isolation Kits , NA2018 magnetic racks	Magnetic-bead extraction for the inhibitor-heavy matrices RPA is usually chosen for
Strip positive control	NA2032 FAM and biotin end-labelled 23S amplicon	A known dual-labelled DNA that separates strip failure from reaction failure
Label	AU2063 FITC/FAM Nucleic Acid	A graded series of FAM-labelled



Workflow step	Product	What it provides
stoichiometry QC	Stoichiometry Lateral Flow QC Kit	controls that maps the working range of the strip against the gold, which is how the hook effect is bounded before it appears in real samples
Buffer optimisation	AU2035 RNase-Free LFA Running Buffer Screening Pack	Panel of running buffers for finding the dilution and formulation that migrate cleanly
Building a custom strip	AU2060 anti-FITC/FAM colloidal gold , AU2017 streptavidin gold , AU2054 blank strips , AU2059 Lateral Flow Starter Kit , AU2064 BYO-particle kit	Membranes, pads, conjugates and cassettes for a strip built around a specific hapten pair
Reader-based or fluorescent readout	AU2037 QDot , AU2049 europium and AU2053 R-phycoerythrin universal kits	Fluorescent labels in place of gold, where a reader is available and a wider quantitative range is wanted
CRISPR-coupled detection	PN2056 , PN2057 , PN2058 and PN2103 ssDNA Cas12a reporters	Dual-labelled collateral-cleavage reporters for RPA–Cas12a lateral flow

Two caveats when buying rather than building. Universal strips are hapten-agnostic — they report the labels, so specificity still has to come from the probe design. And the working range belongs to the strip, not the reaction, which is why a stoichiometry check per strip lot beats inferring the range from clinical samples.

Applications, alternatives and where the field is going

RPA–LFA is used wherever a reader is impractical: point-of-care detection of SARS-CoV-2, HIV, malaria and tuberculosis; field diagnosis of plant pathogens, where tolerance of crude leaf lysates is decisive; food and water testing for *Salmonella*, *Listeria* and norovirus; and veterinary and biosecurity screening on farms and at borders.

Set against other readouts of the same amplification:

Readout	Instrument	Quantitative	Tube opened	Main limitation
Lateral flow	None	No	Yes	Carryover risk; binary result
Real-time fluorescence (exo probe)	Fluorometer	Semi	No	Needs a reader and power
Agarose gel	Tank, imager	No	Yes	Lab-bound; smeared by reaction matrix
Turbidity / dye	Eye or simple reader	No	No	Poor specificity; no sequence check

The most consequential recent direction is coupling RPA to CRISPR effectors — Cas12a for DNA, Cas13a for RNA. A guide-matched amplicon activates the effector, which then cleaves reporter oligonucleotides collaterally; because a dual-labelled reporter loses signal on cleavage, these assays



are often read as inverted strips. The gain is a second sequence-recognition step after amplification, which largely solves the dimer problem and allows single-nucleotide discrimination. The costs are an extra incubation, another reagent to stabilise (the ssDNA reporter, such as the PN2056–PN2103 Cas12a series), and a readout logic that confuses operators trained on conventional strips.

Two other directions are worth watching: sealed cartridges that run amplification and strip in sequential chambers, removing the carryover failure mode without giving up instrument independence; and lyophilised all-in-one reagents, which remove the cold chain — often the real barrier to deployment rather than the chemistry.

Conclusion

RPA with a lateral flow readout is the most practical instrument-free nucleic acid test available, reaching single- to double-digit copy sensitivity in under 35 minutes at body temperature. Its ceiling is set by three design decisions rather than by the amplification: whether the labelling scheme makes primer artefacts invisible, whether the reaction proteins are removed before the product meets the membrane, and whether the workflow keeps amplicon away from the next reaction.

An assay that uses an nfo probe, purifies or consistently dilutes the product, carries an internal amplification control, is read at a fixed time, and runs in separated pre- and post-amplification spaces will behave reliably. One that skips any of these produces lines that cannot be trusted — and the failure looks like a working test.