



Construct Specification

Enzyme Sourcing and Expression

HUMAN FRONTIER LABS

Document 4 of 4 · October 2026

Construct Specification and Enzyme Sourcing

1. Purpose

Documents 2 and 3 specify the assay and the oligonucleotides. This document specifies the enzyme. It names the exact protein source, the coding sequence to synthesise, the expression route, and a low-cost pilot that should be run before any custom synthesis is ordered.

2. Step Zero: Commercial RT Pilot

Run the assay in Document 2 with an off-the-shelf reverse transcriptase before committing to gene synthesis. This costs reagents already present in most laboratories and returns an interpretable result either way.

- **Candidate enzymes:** MMLV-RT or AMV-RT, standard commercial preparations.
- **Templates:** consensus and iso-thermal control from Document 3, unchanged.
- **Conditions:** as Document 2, substituting the commercial enzyme for the cell-free expression mixture.

PILOT OUTCOME	INTERPRETATION	NEXT STEP
Arrest on consensus, read-through on control	Motif-dependent pausing occurs with a generic RT	Proceed to custom construct; the effect is not enzyme-specific
Arrest on both templates	Pausing is composition-driven, not sequence-specific	Re-examine the hypothesis before further spend
No arrest on either	Generic RT does not pause here	Custom construct still worth testing; the effect may be specific to the ART enzyme

WHY THIS COMES FIRST

Two of the three outcomes change what gets ordered next, and the pilot costs a fraction of a synthesis run. A negative result with a commercial enzyme does not falsify the hypothesis, because the proposed mechanism is specific to the ART tuning pocket; it does narrow what the custom construct must demonstrate.

3. Custom Construct: MarshHill Reverse Transcriptase

FIELD	VALUE
-------	-------

Source accession	MW248466.1
Organism	<i>Staphylococcus</i> phage MarsHill, complete genome, 266,637 bp
Locus tag	CPT_MarsHill_206
Protein ID	QQM14740.1
Product	RNA-directed DNA polymerase
CDS coordinates	215,089 – 216,594
CDS length	1,506 bp (502 codons, including stop)
Protein length	501 aa
Translation table	11 (bacterial)

COORDINATE PRECISION

215,089 is the **CDS** start. The **gene** feature for this locus begins at 215,072 and includes a 17 bp ribosome binding site upstream of the start codon. Ordering 215,072–216,594 yields 1,523 bp, which is not divisible by three and will not translate in frame. Synthesise the CDS coordinates above.

Protein sequence (QQM14740.1, 501 aa)

```

MFKFNNVERINLIGNRKMDYFELQESLYTLEPKFKKIVKNDKGNFLLIFSDGRKEYISNT
NAENTFRSKMMVNSIDDYNLFGKNKQEHSEINNLDYSLIHHLRSFIYEDYKFIIITKK
FRIYEEDKIFNLIKLKSTTENDKDIFKIEGILKIKKDLINIEELELIKDI LNDCRTSKER
QEYILNSLELLLSIDLKNNVNTFDTFYIPKSDGTKREINSPVEEIRPILKKQANILNA
ILNLRINKYNLGENIIAYRNGKSIKDNADYHKENKTIKFDISKFFDNCHFDYWKENLFF
SFKDEDYVMERFEKHMEGLLFRKNNKGLYMGSPSPAASNLMVPMRYINNILSKFKTD
IKVTYVADDITFSSNEINDEFNPKRLKSMVNHAFKVNLKFKLKEEKSVMKNHKRKITG
LTVNHNDPTISQKRYRQLRTIIHLLNKGVEFKDIKFKTPLELVSLINYYLDVDTTEKV
KRLMDKNERIFNELKKEAGVN

```

4. Synthesis and Expression

PARAMETER	SPECIFICATION
Synthesis route	Commercial gene synthesis (Twist, IDT or equivalent)
Codon usage	Optimise for <i>E. coli</i> ; PURExpress is reconstituted from <i>E. coli</i> components

Vector backbone	T7-driven pIVEX or pET-derived, PURExpress compatible (NEB #E6800L)
Cloning junctions	NdeI / BamHI, N-terminal 6×His tag for immunoblot verification
Template per reaction	250 ng linearised plasmid per 25 µL translation reaction
Expression conditions	30 °C, 2 hours, per Document 2 step 1

EXPRESSION CONSIDERATIONS

PURExpress is a prokaryotic reconstituted system. It supplies no eukaryotic chaperones and performs no post-translational modification, which is relevant if human ORF2p constructs are attempted later but not for this phage enzyme. A 501 aa product is within normal range for the system. Confirm synthesis by anti-His immunoblot before assuming the enzyme is present, since a failed translation and an inactive enzyme produce the same blank gel.

5. Scope and Evidence Class

- **This specification covers** enzyme sourcing, the expression template, and the pilot route.
- **It does not cover** in vivo assays or structural determination.
- **The assay in Documents 2 and 3 tests one thing:** whether the motif causes sequence-specific polymerase arrest. It does not test the GNAT-fold acetylation mechanism, and a positive result would not confirm it.
- **The GNAT pocket volumetric figures in Document 1** are comparative structural predictions. Testing them requires structural or mass-spectrometric work outside this pipeline.