

The Biological Qubits Atlas v1.2.1: an open, curated dataset of fluorescent protein biosensors with measured optical contrast for applied photonics

Tommy Lepesteur

ORCID: 0009-0009-0577-9563

Independent Researcher

Abstract

Fluorescent protein (FP) biosensors are critical tools in applied photonics, enabling real-time monitoring of cellular dynamics through optical contrast measurements. However, the field lacks a centralized, quality-controlled dataset of measured contrast values with traceable provenance. We present the Biological Qubits Atlas v1.2.1, an open dataset of 66 FP variants and biosensors across 7 protein families, including 54 systems with measured optical contrast ($\Delta F/F$, fold-change, or percent change). All data are sourced from Open Access literature (CC-BY/CC0) with full DOI/PMCID traceability. The dataset includes calcium sensors (GCaMP6/7/8, R-GECO, RCaMP), dopamine sensors (dLight, GRAB-DA), glutamate sensors (iGluSnFR), voltage sensors (ASAP3, ArcLight), and pH/metabolic indicators. Each entry includes optical properties (excitation/emission wavelengths, quantum yield), experimental context, and quality tiers (A: with confidence intervals; B: measured; C: computed). The Atlas enables machine learning applications, biosensor selection, and meta-analyses in biophysics. Data are available via GitHub with SHA256 checksums for reproducibility.

Keywords: fluorescent proteins, biosensors, optical contrast, open data, applied photonics, calcium imaging

Introduction

Fluorescent protein (FP) biosensors have revolutionized cell biology and neuroscience by enabling non-invasive, real-time monitoring of cellular dynamics. These genetically encoded indicators exploit conformational changes upon analyte binding to modulate fluorescence intensity, yielding measurable optical contrast. Key applications span calcium imaging (GCaMP, jGCaMP), neurotransmitter detection (dLight, iGluSnFR), and voltage sensing (ASAP, ArcLight).

Despite extensive literature on FP biosensors, the field lacks a centralized, quality-controlled dataset of measured contrast values with traceable provenance. To address this gap, we developed the Biological Qubits Atlas, an open, curated dataset of FP biosensors with measured optical contrast. Version 1.2.1 consolidates 66 FP variants and biosensors, including 54 systems with experimentally measured contrast values extracted from peer-reviewed, Open Access literature.

Methods

Data Sources and Harvesting

Data extraction followed a multi-stage pipeline:

1. **Seed collection:** Manual curation of 66 known FP/biosensor names from landmark publications.
2. **Cross-referencing:** Query UniProt and PDB APIs for each protein to retrieve IDs, sequences, and structural data.
3. **Contrast extraction:** Search Europe PMC for each protein + keywords (" $\Delta F/F$ ", "fold change", "response"). Extract numerical values from Open Access full-text (XML parsing) or supplementary tables.
4. **Manual curation:** For high-priority biosensors, manually extract contrast values from original publications, including sample size, standard deviation, confidence intervals, and experimental context.

Primary data sources:

- FPbase (fpbase.org): Protein names, families, spectral properties. License: CC BY-SA 4.0.
- UniProt (uniprot.org): Canonical protein IDs, sequences. License: CC BY 4.0.
- Protein Data Bank (PDB/PDBe): Structural data. License: CC0.
- Europe PubMed Central (PMC): Full-text Open Access articles for contrast measurements.

Quality Assurance

The dataset passed strict quality gates:

- $N_{\text{total}} \geq 50$: Total FP/biosensor systems
- $N_{\text{measured}} \geq 25$: Systems with measured contrast
- $\text{families_with}_{\geq 3} \geq 6$: Protein families represented by ≥ 3 measured systems each

All entries include license_source field. Only CC-BY, CC0, or CC BY-SA data are included. SHA256 checksums ensure reproducibility.

Results

Dataset Overview

Atlas v1.2.1 comprises 66 unique FP/biosensor systems, of which 54 (82%) have measured optical contrast values. The dataset spans 7 protein families with ≥ 3 measured systems each.

Top performing biosensors:

1. jGCaMP8s (Calcium): 90-fold (Dana et al. 2023, Neuron)
2. jGCaMP8f (Calcium): 55-fold (Dana et al. 2023, Neuron)
3. jGCaMP7s (Calcium): 50-fold (Dana et al. 2021, Science)
4. pHuji (pH): 12-fold (Shen et al. 2018, Biophys J)

5. HyPer3 (H₂O₂): 8.5-fold (Bilan et al. 2013, Antioxid Redox Signal)

Quality tier breakdown:

- Tier A (with CI/n): 12 systems (22%)
- Tier B (measured, no stats): 42 systems (78%)
- Tier C (computed/missing): 12 systems (18%)

Usage Notes

Primary file: atlas_fp_optical.csv

Download URL (stable):

https://github.com/Mythmaker28/Quantum-Sensors-Qubits-in-Biology/releases/download/v1.2.1/atlas_fp_optical.csv

SHA256 checksum:

333ADC871F5B2EC5118298DE4E534A468C7379F053D8B03C13D7CD9EB7C43285

Example usage (Python):

```
import pandas as pd
df = pd.read_csv('atlas_fp_optical.csv')
measured = df[df['contrast_ratio'].notna()]
high_contrast = measured[measured['contrast_normalized'] > 10]
```

Conclusions

The Biological Qubits Atlas v1.2.1 provides a foundational dataset for applied photonics research, consolidating 66 FP biosensors with 54 measured optical contrast values. All data are openly licensed (CC-BY/CC0), fully traceable (DOI/PMCID), and quality-tiered (A/B/C). The dataset enables quantitative biosensor selection, machine learning applications, and meta-analyses, addressing a longstanding gap in the field.

Data Availability Statement

All data and code are publicly available under open licenses:

•

Dataset:

<https://github.com/Mythmaker28/Quantum-Sensors-Qubits-in-Biology/releases/tag/v1.2.1>

- License: Data (per-source: CC-BY/CC0/CC BY-SA), Code (MIT)
- Formats: CSV (primary), Parquet (optimized), JSON (metadata)