Engineering fluorescent proteins

Bob Keenan
Department of Biochemistry & Molecular Biology
University of Chicago

Principles of Fluorescence Techniques
April 10, 2009

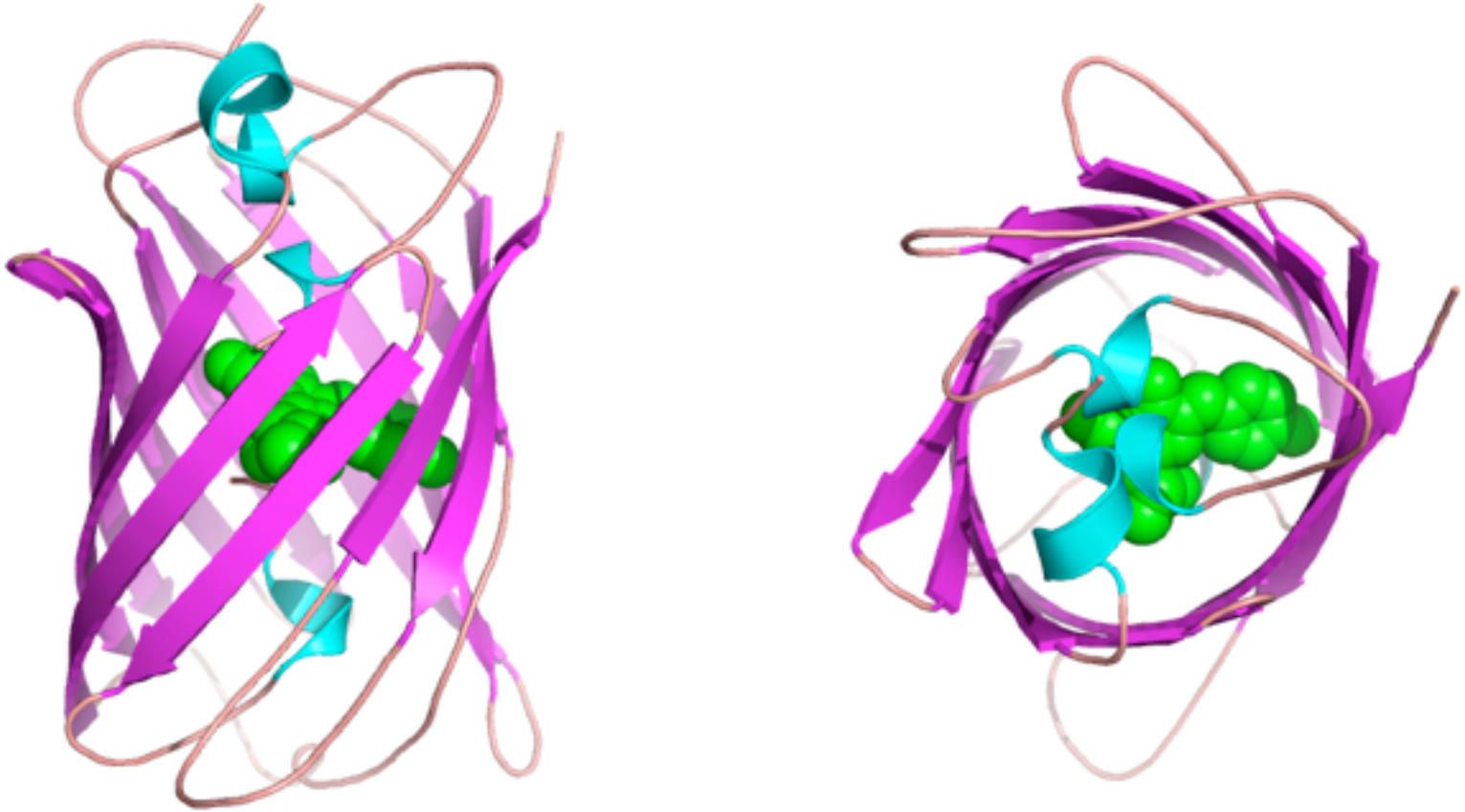
GFP History

- Gene discovered by Osamu Shimomura in 1962
- Cloned by Douglas Prasher in 1992
- Heterologous expression of recombinant GFP in *E. coli* and *C. elegans* by Martin Chalfie in 1994
  - Demonstration that GFP forms spontaneously, not requiring additional enzymes or co-factors
  - Powerful tool: genetically encoded fluorescent marker

Aequoria victoria

GFP structure

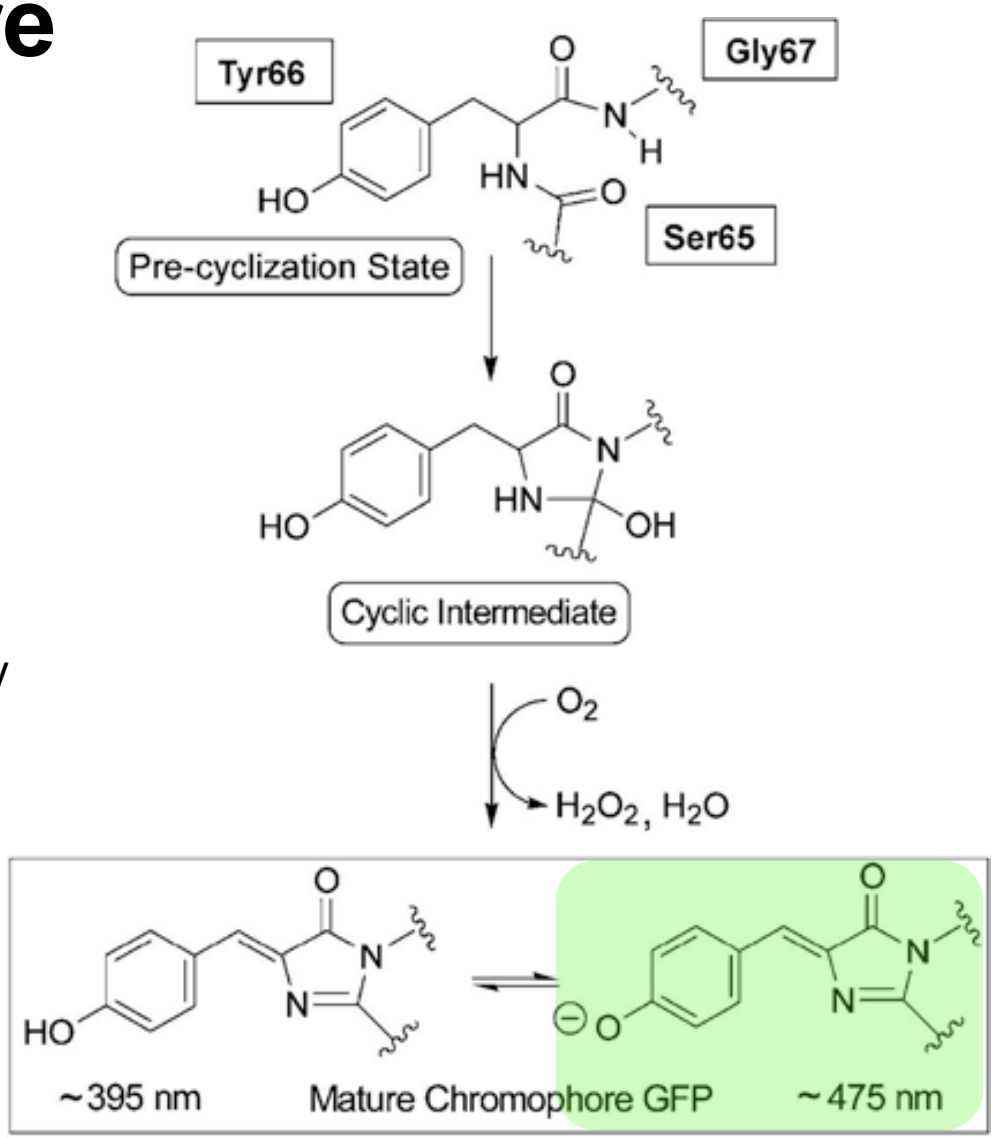
- 27 kD protein; 238 residues
- 11-stranded beta-barrel wrapped around central helix
- Tripeptide-derived chromophore is shielded from bulk solvent in well-packed protein interior



Ormo *et al*., **Science**, (1996)

GFP chromophore

- Tripeptide sequence
  - Ser-Tyr-Gly
  - Only Gly is essential for chromophore formation
- Mechanism
  - main-chain cyclization, followed by oxidation-dehydration steps
  - autocatalytic: NO external cofactors (except O_2)
- Maturation
  - $t_{0.5}$ approx. 20-80 minutes



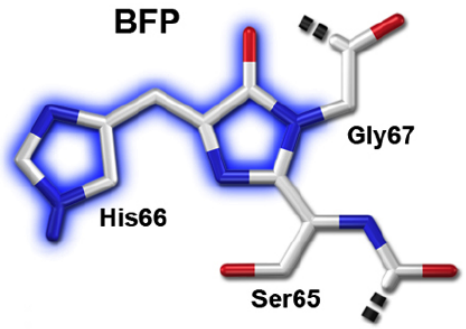
from Nienhaus and Wiedenmann, **ChemPhysChem** (2009)

Early GFP derivatives

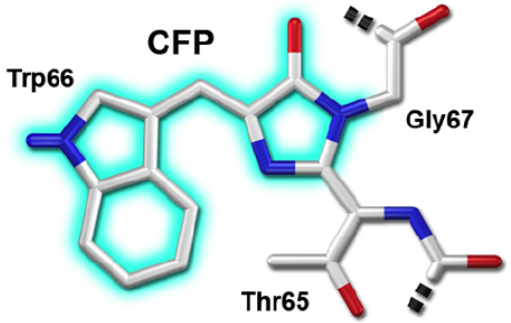
Pioneering work by Roger Tsien in mid-1990's:

- chromophore mutations: large shifts in absorbance/ emission spectra
- mutations in surrounding region: more modest shifts

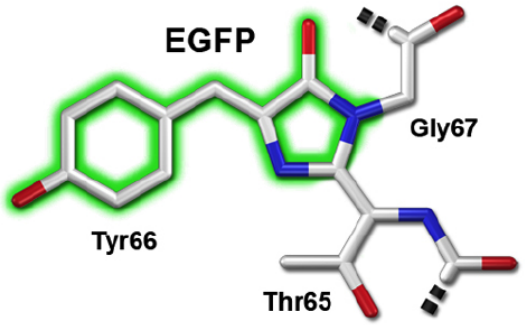
Despite these engineering successes, no ORANGE or RED variants of GFP!



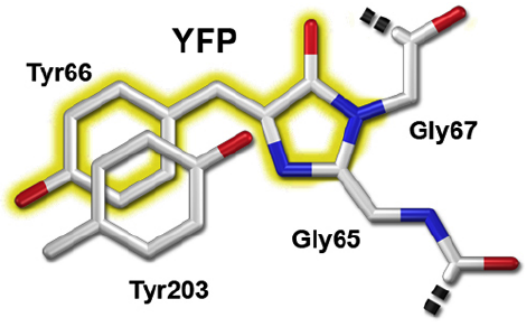
383/448 nm



433/475 nm



488/507 nm



514/527 nm

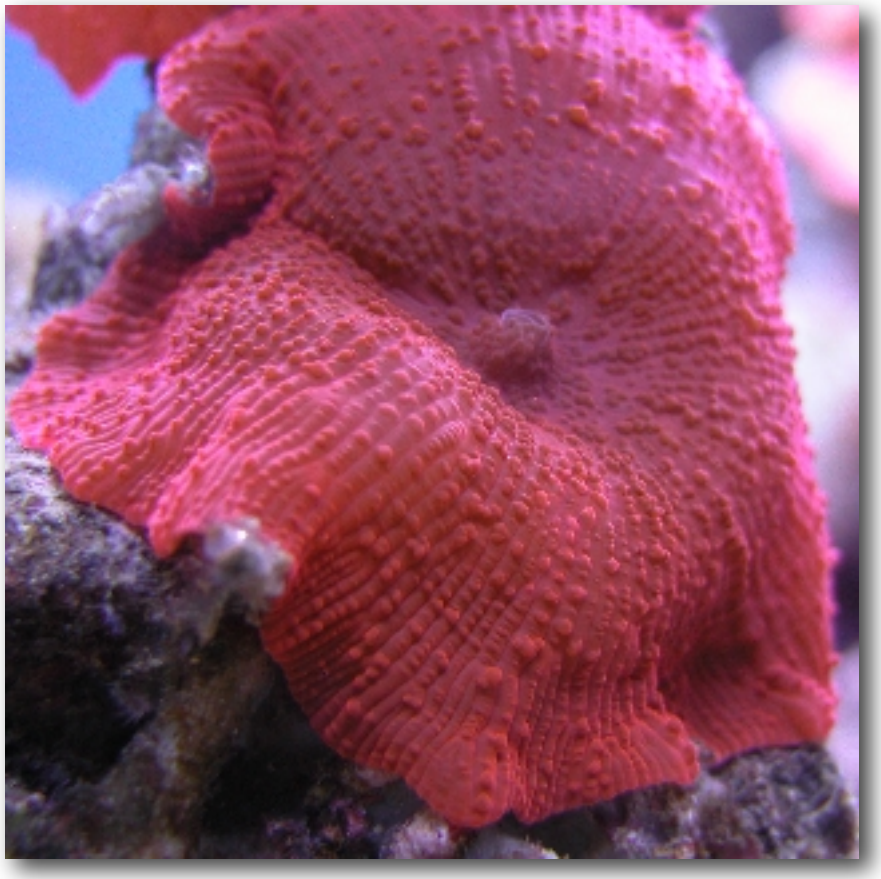
adapted from Shaner *et al.*, **J. Cell Sci.** (2007)

Anthozoan FPs

- First discovery of FPs from non-bioluminescent organisms
  - Matz *et al.,* **Nature Biotechnol.** (1999)
  - series of FPs with novel properties, including green, yellow, orange and red emission

- **DsRed** (ex/em = 559/584 nm). Long-wavelength provides advantages:
  - light is scattered less and causes less toxicity in tissues and cells than wavelengths used to excite GFP
  - improved signal:noise due to less cellular autofluorescence
  - dual-color experiments with GFP

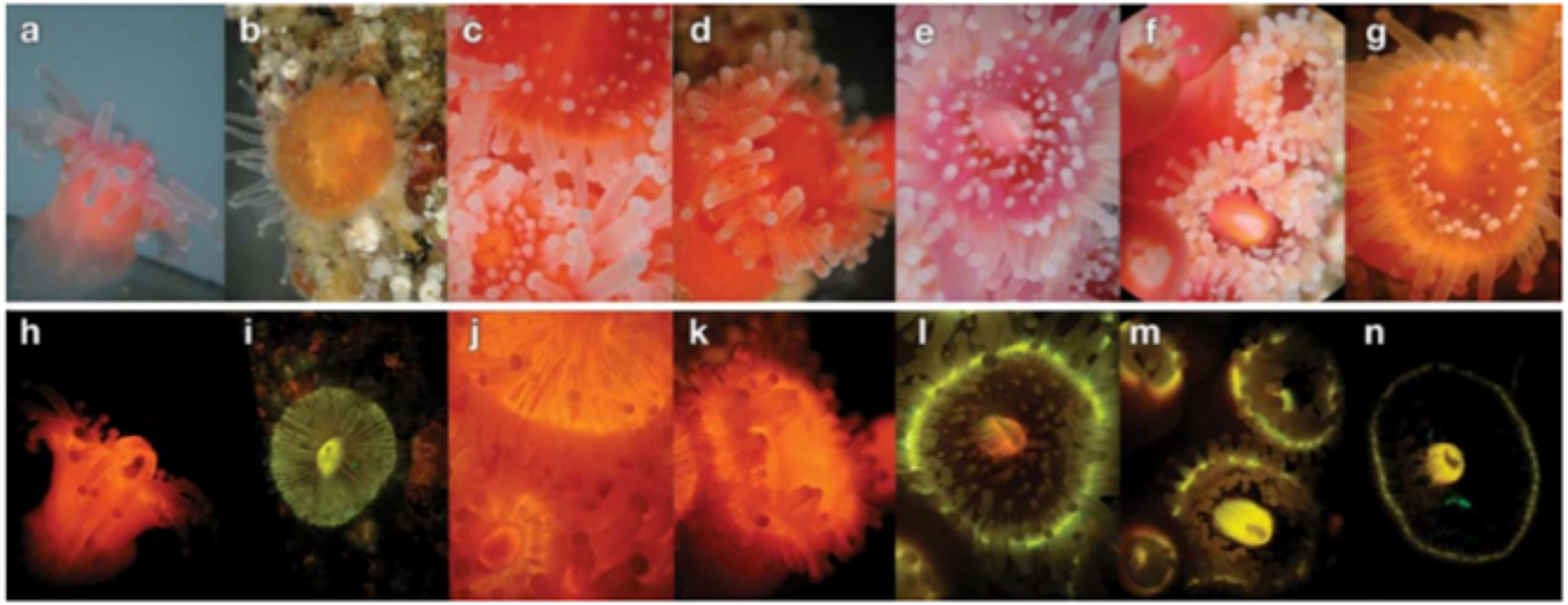
- Explosion of FPs discovered over the last decade provide a rich source of starting material for engineering...



Discosoma sp.

Natural FP diversity

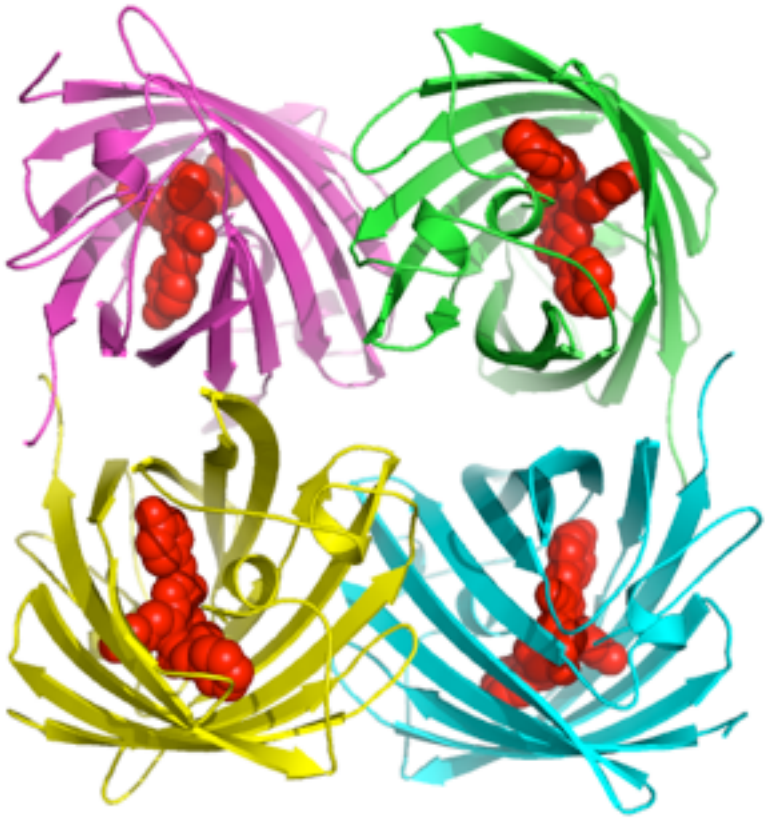
Corynactis californica
(in collaboration with Steve Haddock, MBARI)

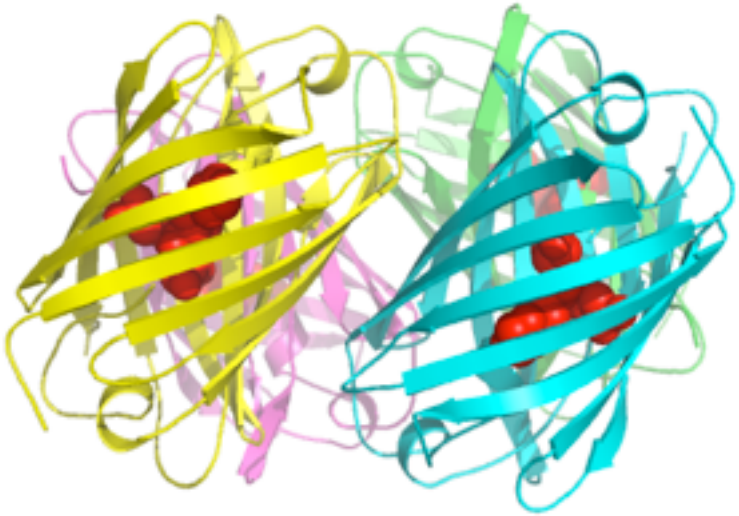


Six FPs: Green, yellow, orange and red emission

Schnitzler et al., Marine Biotech., (2008)

DsRed structure



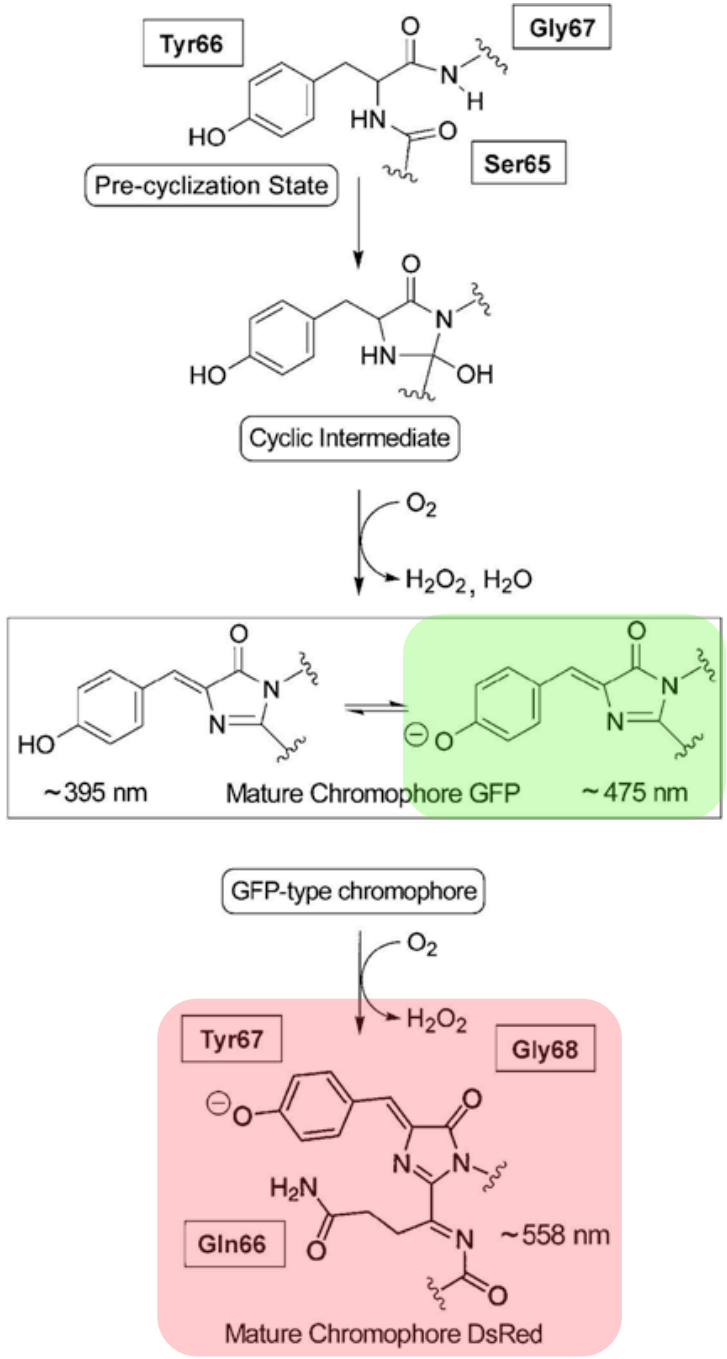


- 225 amino-acids
- 23% identical to *A. victoria* GFP
- Similar fold
- Tetrameric structure

Wall *et al.*, **Nature Structural Biology** (2000)

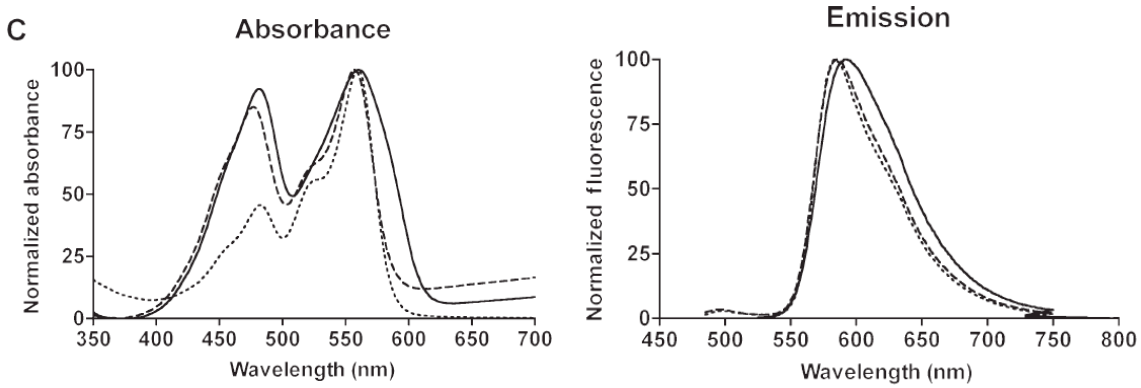
DsRed chromophore

- Tripeptide sequence
  - Gln-Tyr-Gly
- Mechanism
  - derived from GFP-like chromophore
  - additional oxidation resulting in formation of acylimine at position 66
  - extension of pi-orbital conjugation results in red-shifted emission
- Maturation
  - $t_{0.5}$ approx. 12 hours!



Incomplete maturation

- Mixture of green (trans/sp3) and red (cis/sp2) chromophore
- Observed pure RED emission is due to intra-tetramer FRET



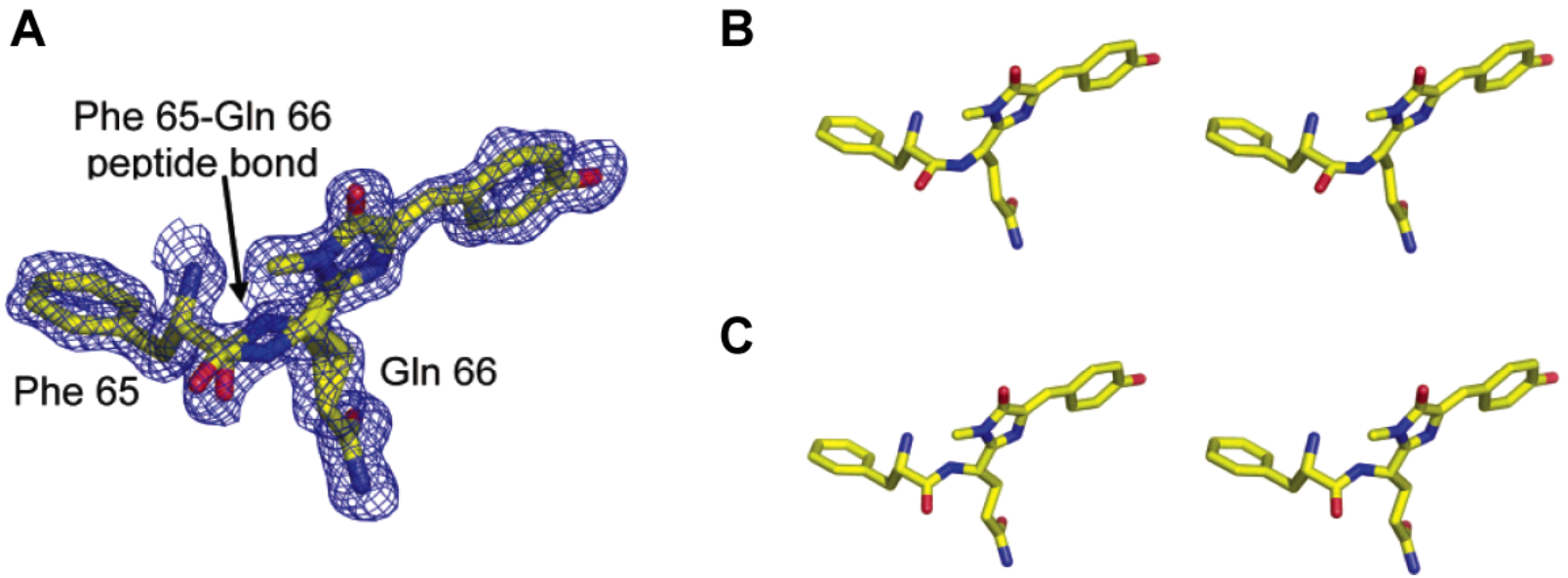


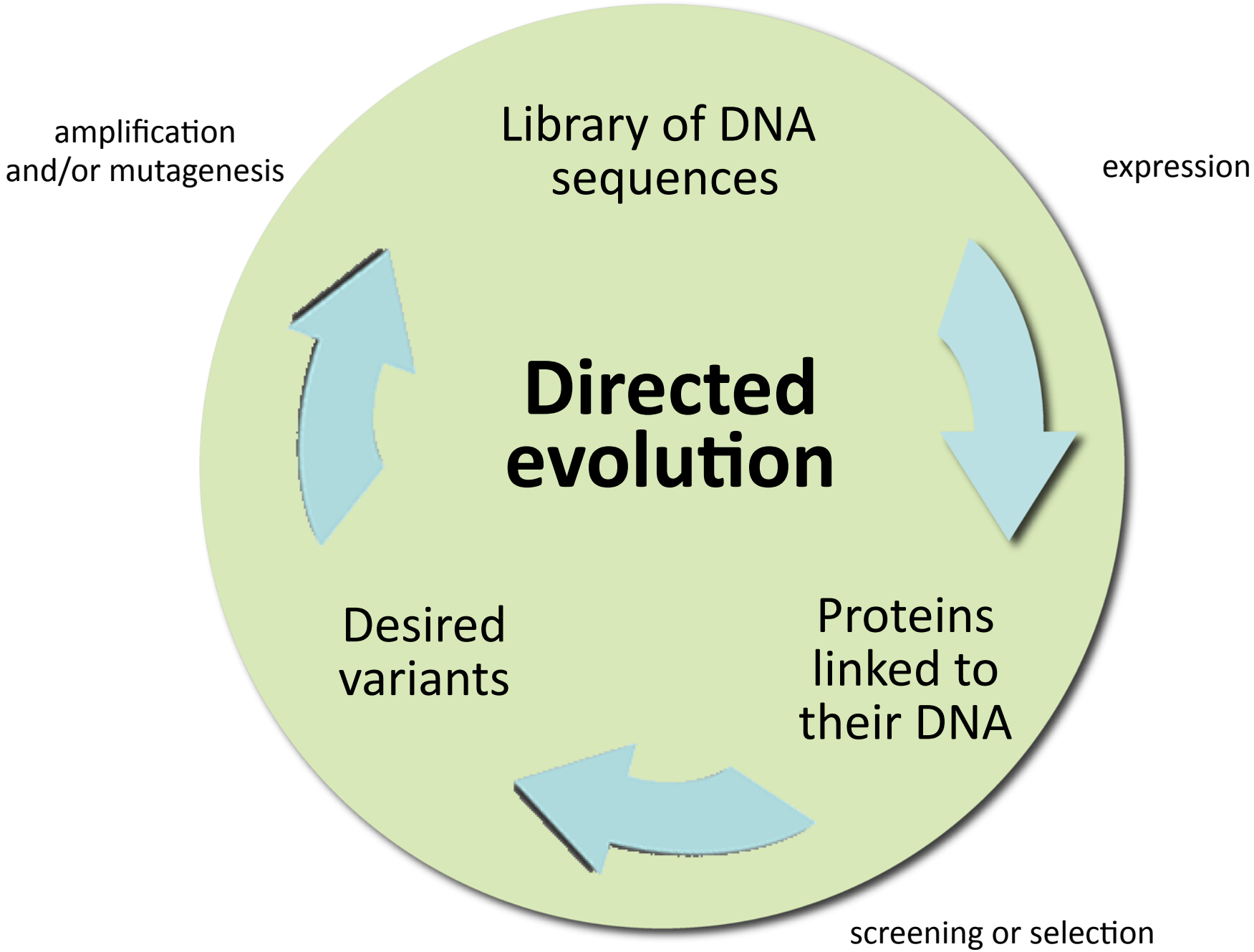
FIGURE 2: Refined high-resolution DsRed structure revealing *trans* and *cis* chromophore conformations. DsRed wild-type chromophore with $2F_o - F_c$ electron density omit map contoured at 1σ (A). Electron density supports both *trans* and *cis* Phe 65−Gln 66 peptide bonds. The Gly 68 carbonyl carbon and oxygen atoms are not shown. Stereoview of the mature red (B) and immature green (C) chromophores. Gln 66 Cα (center) is sp^2-hybridized with planar geometry in (B) and sp^3-hybridized with tetrahedral geometry in (C). Oxygen atoms are colored red and nitrogen atoms blue.

Tubbs *et al.*, **Biochemistry** (2005)

Limitations

- Anthozoan FPs (esp. reds) generally suffer from a number of serious limitations:
  - *slow maturation rate*
  - *incomplete maturation (green & red chromophores)*
  - *tetramerization*
  - *tendency to aggregate at high concentrations*
  - *cytotoxicity*

- Focus on recent advances in protein engineering that have led to new and improved FP variants (emphasis on reds)...


amplification
and/or mutagenesis
Library of DNA
sequences
expression
Directed
evolution
Desired
variants
Proteins
linked to
their DNA
screening or selection

Sequence and function space

- Sequence space is vast:
  - 20^{225} sequence variants in a 225-residue protein!

- But *functional* sequence space is tiny:

  "…the overall prevalence of sequences performing a specific function by any domain-sized fold may be as low as 1 in 10^{77}, adding to the body of evidence that functional folds require highly extraordinary sequences." –D. Axe, J. Mol. Biol. (2004)

- State-of-the-art assay formats allow functional characterization of ~10^{12} variants.

- GOAL is to focus the search of sequence space in regions of functional interest...

Library construction methods

1. Random mutagenesis
2. Targeted mutagenesis
3. Recombination-based mutagenesis

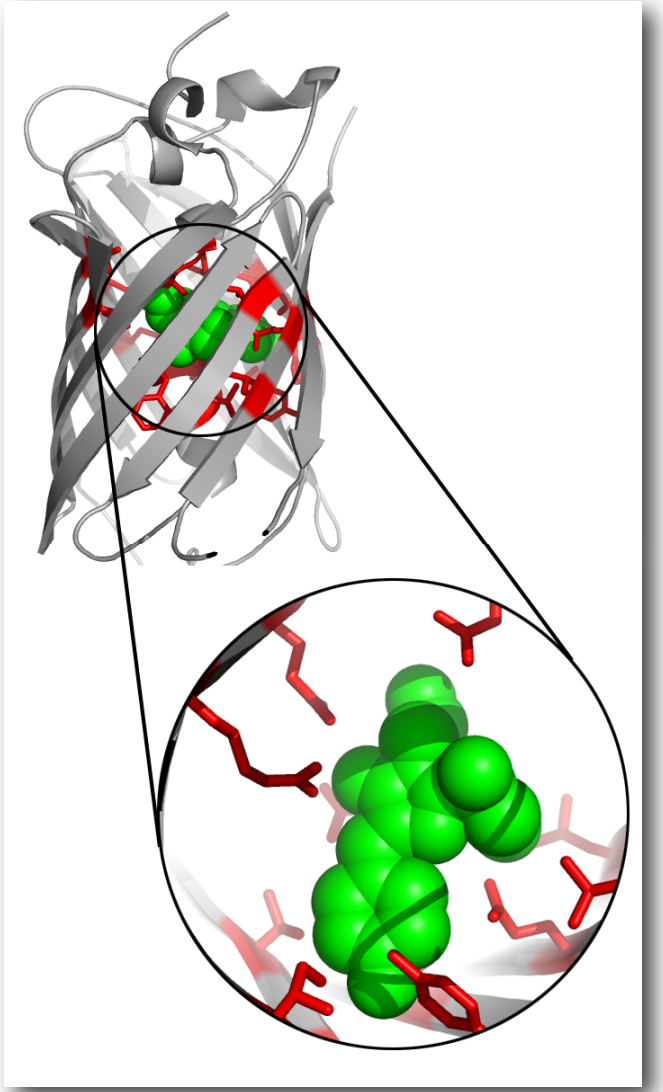
*These approaches differ mainly in the **type**, **location** and **number** of changes that can be introduced into a library*

Random mutagenesis

- Error-prone PCR to introduce a few mutations per gene (on average)
- Requires no prior structural or mechanistic information
- **Number** of mutations (load) must be kept low
  - *Most random mutations are neutral or deleterious to protein function*
  - *Beneficial combinations of mutations are difficult to uncover*
- **Types** of mutations are restricted
  - *Single nucleotide changes to a codon typically access ~6/19 possible amino-acid substitutions*
- Relatively unbiased **location** of mutations

Targeted mutagenesis

- Structural and/or biochemical information used to select a subset of positions for mutation
  - *e.g.*, focus sequence search on residues surrounding chromophore
- No restriction on the **types** of mutations (all 20 amino-acids may be encoded at any position)



Considerations

- **Number** of variable positions evaluated is restricted:

  <u>Simultaneous</u> mutation of 4 residues = 20^4-1 = library of 10^5 variants
  <u>Independent</u> mutation of 4 residues = 4 x 19 = library of 76 variants.

- **Location** of mutations are user-biased:

  -*The effect of mutations distant from a functional site are even more difficult to predict* a priori*, but they are often important!*
  -*e.g., Kazaluskas and coworkers (**Chem. Biol.** 2005) compared random vs. targeted mutagenesis to enhance esterase enantioselectivity. Beneficial mutations found in and distant from the active site. Caveat emptor...*

Recombination-based library construction

- Sequence diversity obtained by recombination of homologous genes
- Requires no prior structural or mechanistic information
- Example: **DNA shuffling**
  - Stemmer, **PNAS** (1994)
  - *Many* variations exist

DNA shuffling

2 homologous genes (~80% here)

5′-…GTCATCGATCGAGCTAGCTAGGATAGGCTAACGCGAA…-3′
3′-…CAGTAGCTAGCTCGATCGATCCTATCCGATTGCGCTT…-5′

5′-…TTCGTCCATTGAGCTAGCTAGGATATCCTGACGCGAG…-3′
3′-…AAGCAGGTAACTCGATCGATCCTATAGGACTGCGCTC…-5′

Digest into ~50-200 bp fragments

5′-…GTCATCGATCGAGCTAGCTAGGAT-3′
3′-TCGATCGATCCTATAGGACTGCGCTC…-5′

“Primerless” extension at crossover points

5′-…GTCATCGATCGAGCTAGCTAGGATATCCTGACGCGAG…-3′
3′-…CAGTAGCTAGCTCGATCGATCCTATAGGACTGCGCTC…-5′

Full-length chimeric genes

Considerations

- **Type** of diversity is "functional"
  - pre-selected by natural evolutionary pressures
- **Number** of substitutions tolerated is quite high relative to other methods
  - substitutions are compatible with protein structure and function
  - Exploration of beneficial combinations of substitutions
- **Location** of substitutions: distributed throughout protein structure
  - difficult-to-predict beneficial mutations may be discovered (e.g., long-range, non-active site substitutions)

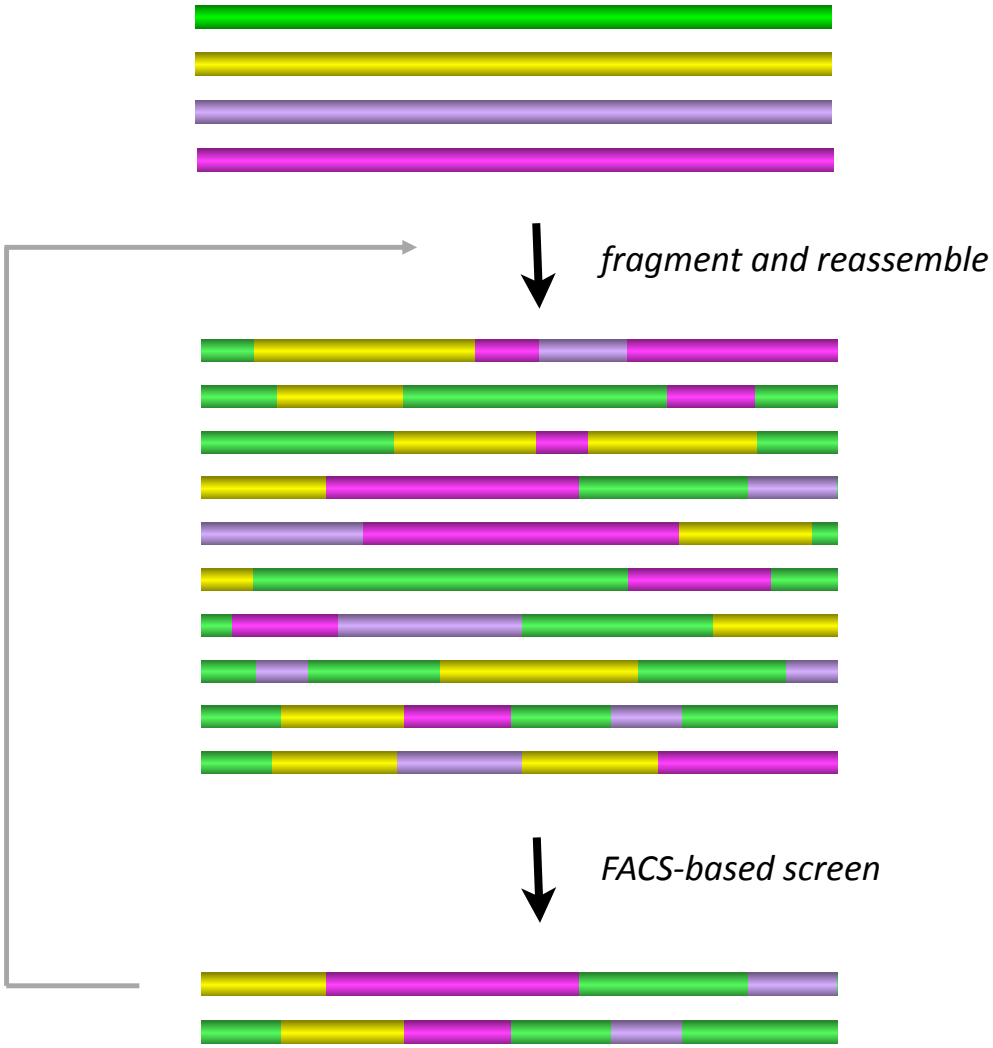
General design considerations: FPs

- Internal vs. external mutations:
  - to a first approximation, we find it useful to distinguish between external mutations which affect surface properties (solubility, oligomerization state etc.) and internal mutations which modulate chromophore properties

- Chromophore mutations: potential for large spectral shifts.
  - e.g., Ser-**His**-Gly in BFP vs. Ser-**Tyr**-Gly in EGFP (60 nm shift in emission)

- Surrounding mutations can lead to large shifts:
  - Typically up to 40 nm shifts in emission
  - BUT sometimes quite large: e.g., **Met**-Tyr-Gly sequence is able to form chromophore spanning 175 nm emission range!
  - AmCyan1 (Clonetech; 486 nm) vs. AQ14 (Shkrob *et al.*, 2005; 663 nm)

Multi-gene DNA shuffling & directed evolution



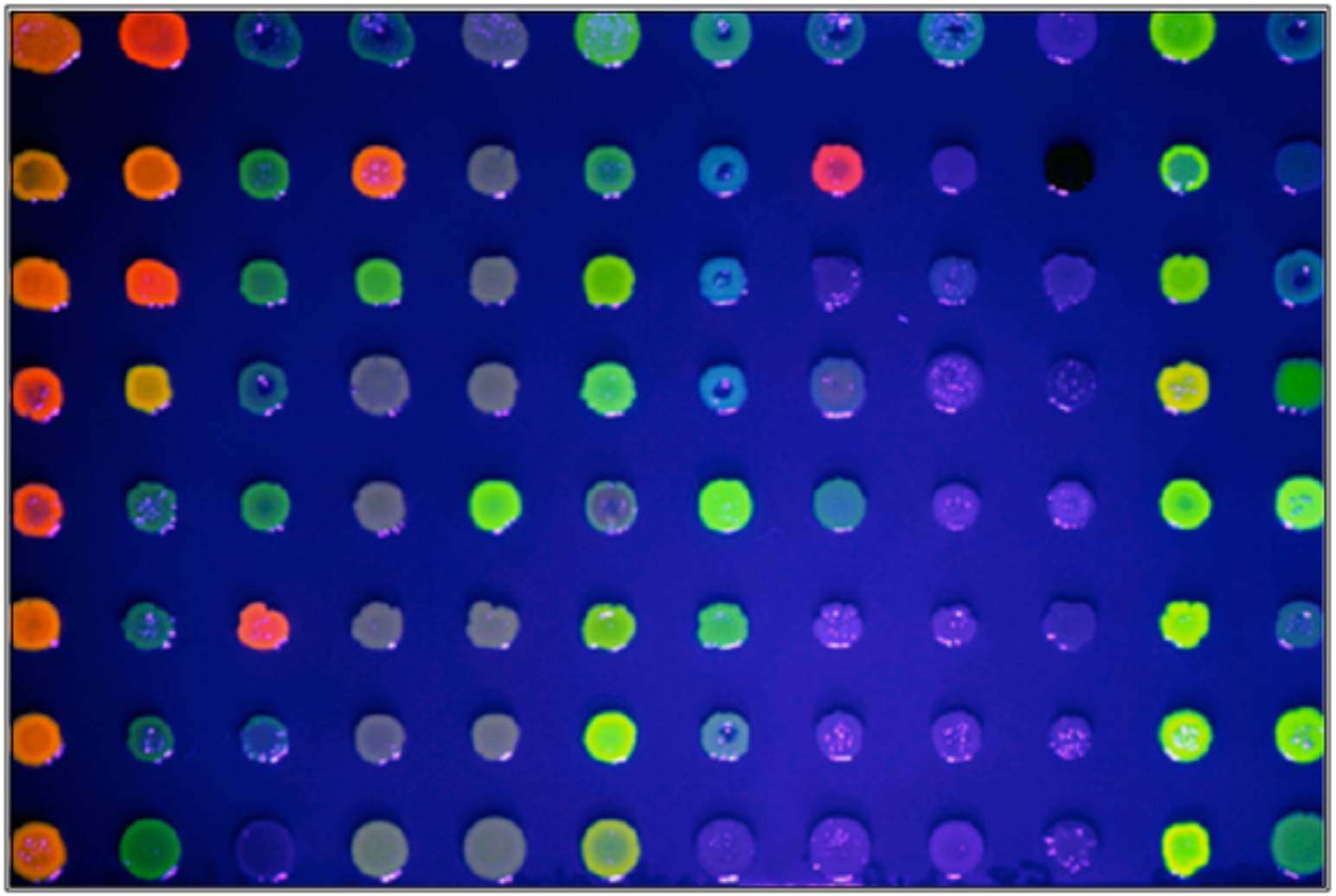
family of related
FP genes

library of chimeric
FP genes

identify interesting
spectral variants

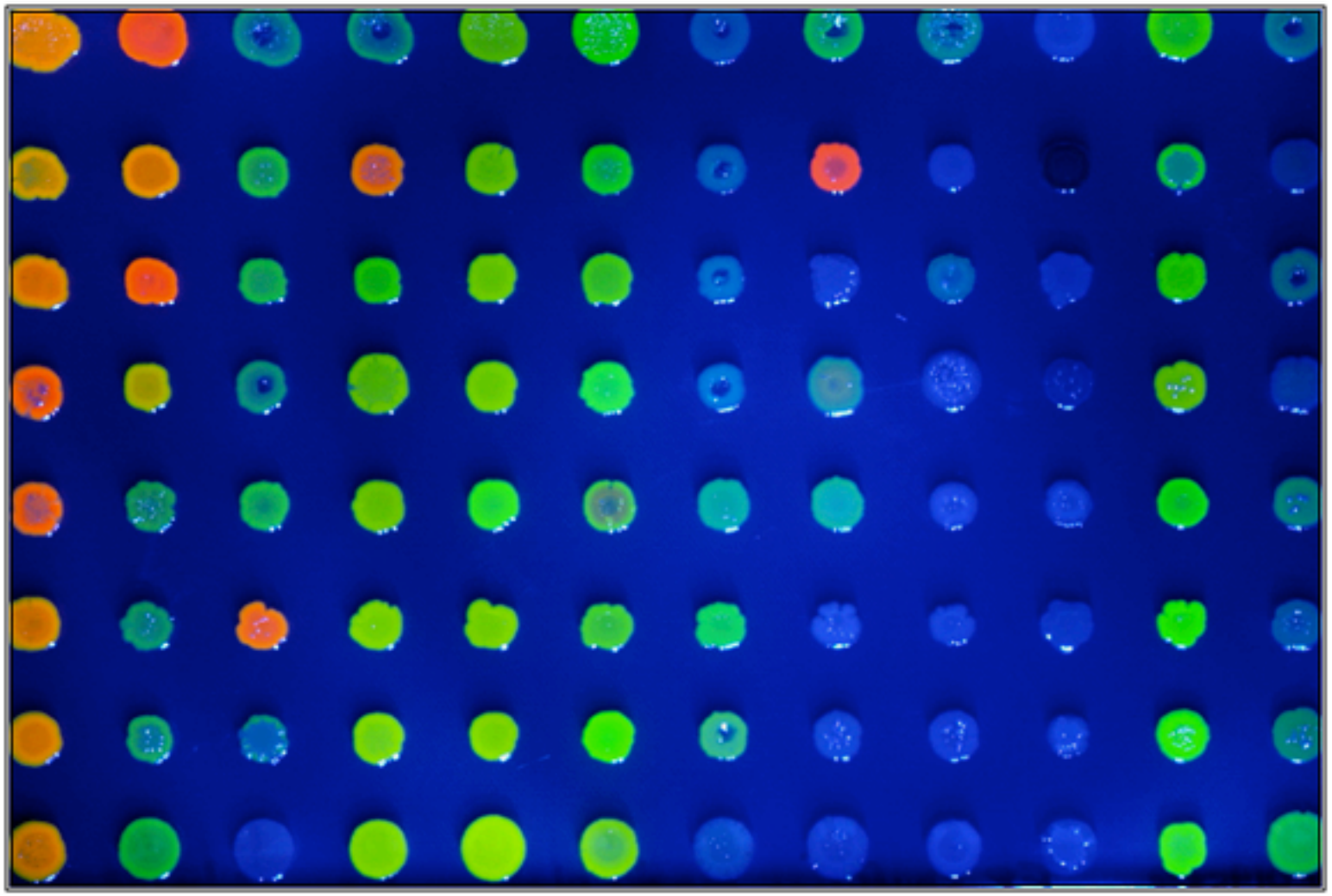
Shuffled variants with distinct phenotypes

PURPLE (400 nm) excitation



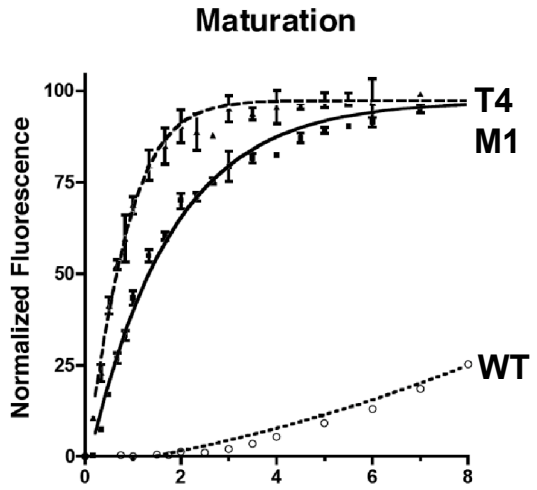
Shuffled variants with distinct phenotypes

BLUE (488 nm) excitation



Faster maturation

- Directed evolution strategy:
  - 6 rounds of error-prone PCR; *E.coli* expression
  - visual screen for rapid acquisition of red emission (4,000-100,000 variants screened per round)
- Result: DsRed.T4 ("**DsRed-Express**") 15-fold faster maturation
- Key mutation: N42Q, adjacent to chromophore



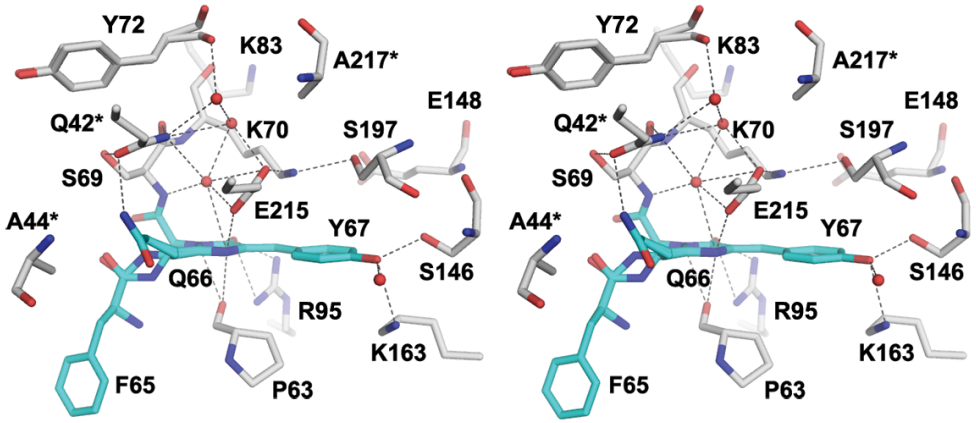


Table 1. Properties of the mature DsRed variants[a]

DsRed variant	Excitation maximum (nm)	Emission maximum (nm)	Maximal extinction coefficient (M^{-1} cm^{-1})	Quantum yield	Relative brightness[b]	Maturation half-time (h)[c]
DsRed1	558	583	52,000	0.68	(1.00)	11
DsRed2	561	587	43,800	0.55	0.68	6.5
DsRed.T1	554	586	30,100	0.42	0.36	0.70
DsRed.T3	560	587	49,500	0.59	0.83	1.3
DsRed.T4	555	586	30,300	0.44	0.38	0.71

Bevis & Glick, **Nat. Biotechnol.** (2002)

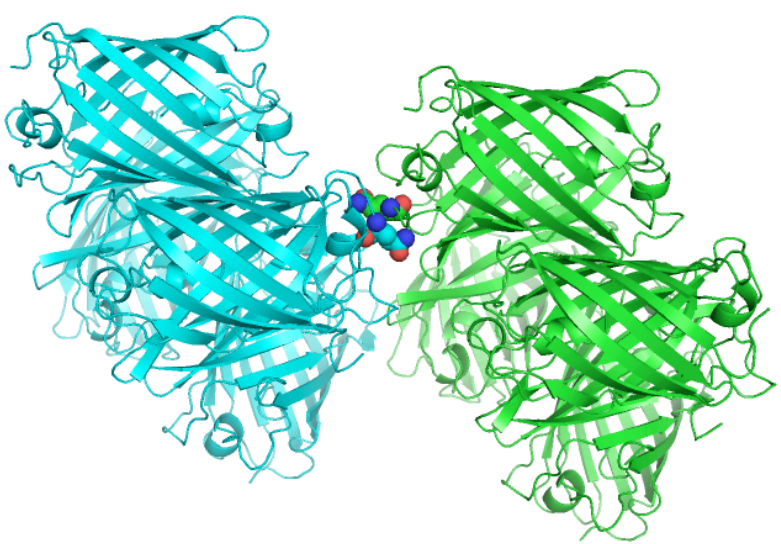
Reducing cytotoxicity in DsRed for whole cell labeling

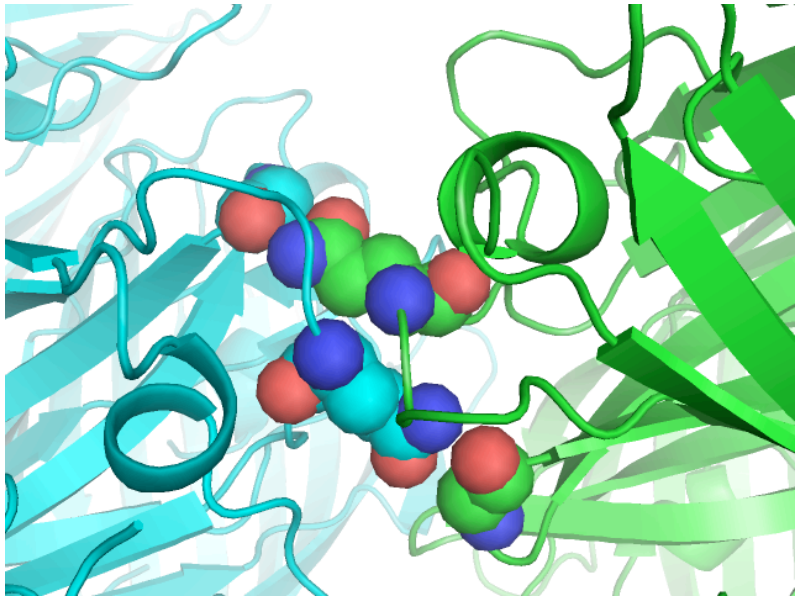
EGFP Is a Nearly Ideal Long-term Expression Tracer for Hematopoietic Stem Cells while DsRed-Express Fluorescent Protein Is Not

– Tao et al., *Stem Cells*, 2006

- Hypothesis: DsRed aggregation leads to cytotoxicity
- GOAL: eliminate "sticky" patches on surface to decrease higher order aggregation and minimize cytotoxicity
- Start with DsRed-Express:
  - bright
  - very photostable
  - fast maturation
  - least cytotoxic of red FPs tested

- Mutagenesis:
  - targeted: analysis of tetrameric xtal structures to identify patches that are consistently involved in crystal packing interactions
  - random





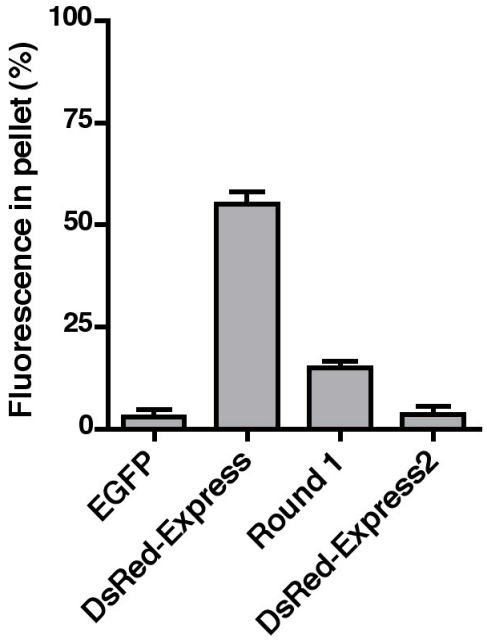
Decreased aggregation

- Two-tiered assay:
  1. visual screen for fluorescent colonies
  2. bacterial lysis assay

- Screened >30,000 variants

- DsRed-Express2 contains 16 mutations relative to DsRed-Express

- Minimal cytotoxicity in bacterial and mammalian systems



Reduced cytotoxicity in murine HSC

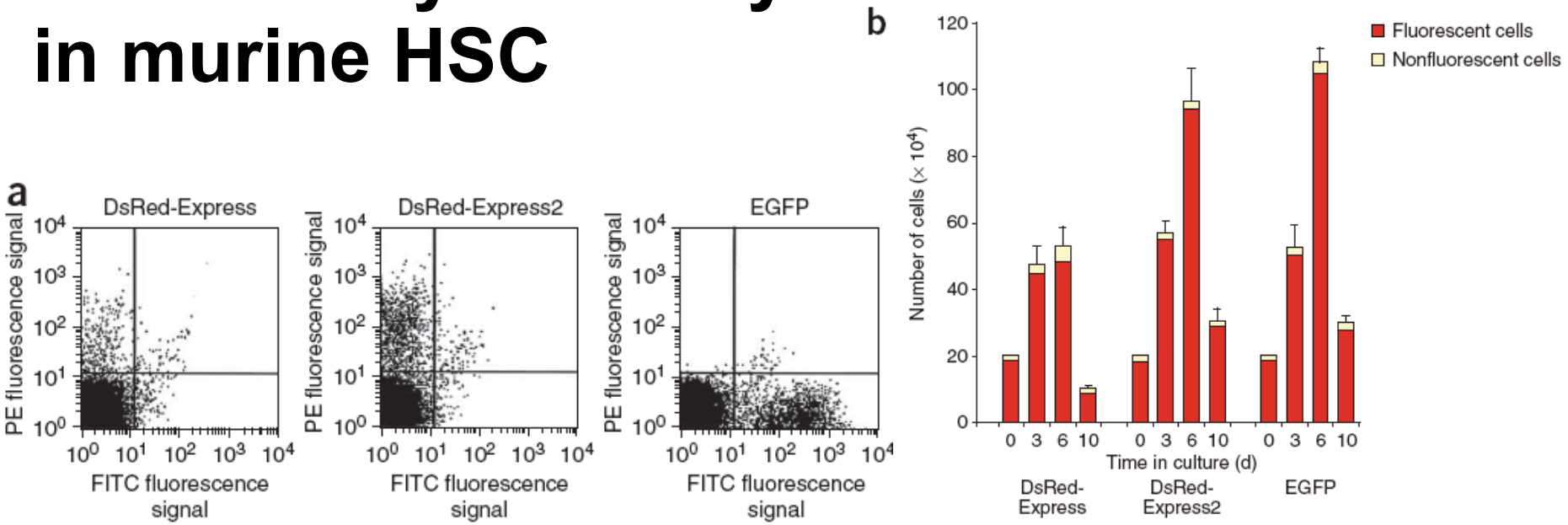
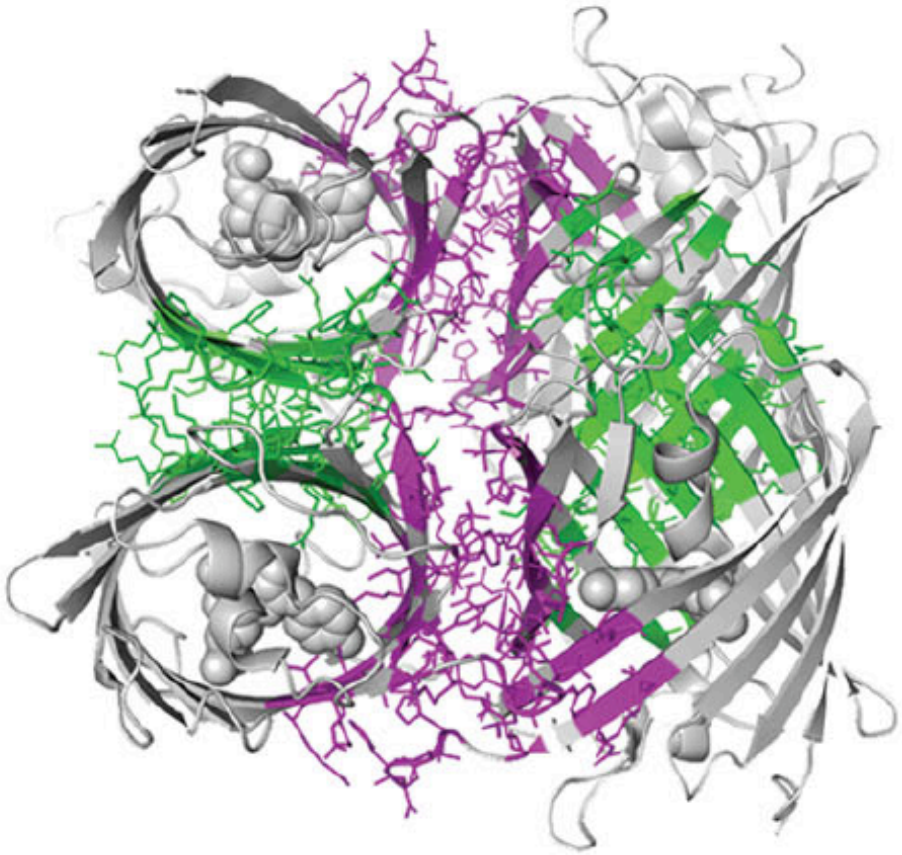


Figure 3 | Robust growth of mouse bone marrow hematopoietic stem and progenitor cells expressing DsRed-Express2. (**a**) Mononuclear bone marrow cells were transduced with retroviral vectors encoding DsRed-Express, DsRed-Express2 or EGFP, and fluorescent cells were sorted after 87 h. Red and green fluorescence signals were detected using PE and FITC filter sets, respectively. The lines represent gates defined by analyzing untransduced cells. (**b**) Sorted cells from **a** were cultured under conditions favoring preservation and growth of hematopoietic stem cells, and the cultures were analyzed by flow cytometry after 3, 6 and 10 d. Three wells were analyzed for each data point. Error bars represent s.d. The decrease in total cell number at day 10 is not fluorescent protein–related but reflects senescence that is routinely observed under *in vitro* culture conditions[11].

Engineering monomeric DsRed

- Tetrameric FPs are useful as whole-cell labels, but NOT as fusions with target proteins
- GOAL: generate true monomers that retain favorable properties of DsRed
- WT DsRed possesses two large tetrameric interfaces

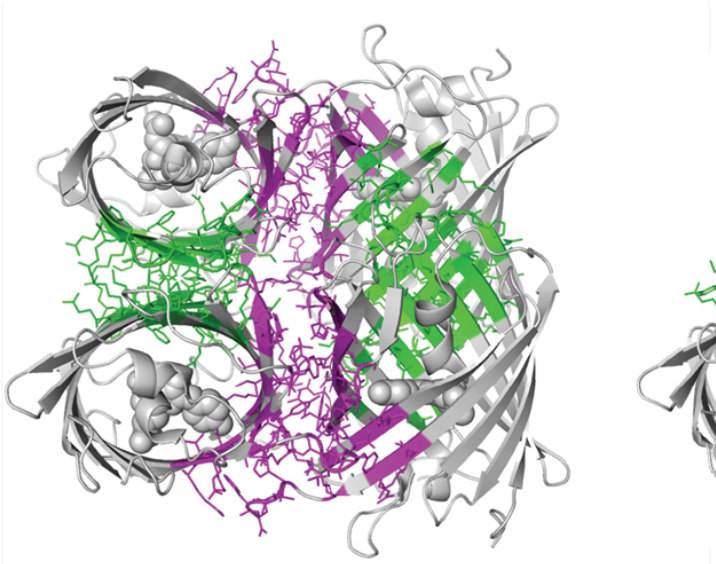
"polar", 1,300 Å^2 BSA
"hydrophobic", 1,000 Å^2 BSA



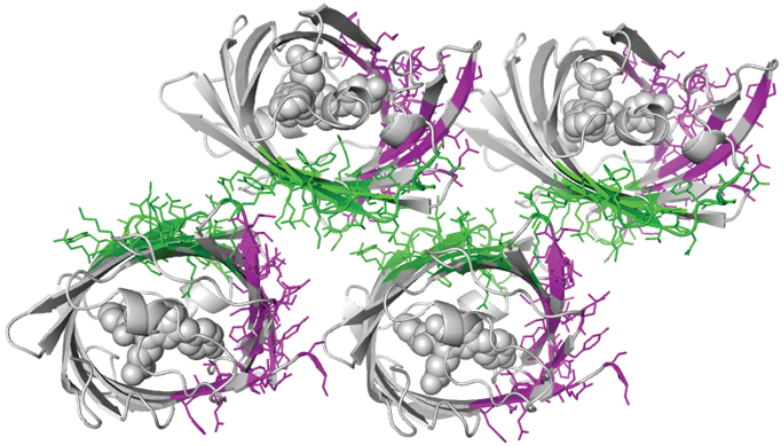
- Two-stage strategy:
  - starting point: DsRed-Express
  - targeted mutagenesis based on x-ray structure
  - random mutagenesis to restore brightness
- Screening strategy: *E. coli* expression; visual screening; pseudo-native SDS/PAGE

- RESULT: monomeric by analytical ultra and x-ray crystallography
- First accomplished by Campbell *et al.*, **PNAS** (2002) with mRFP1
  - 33 mutations
- Followed by DsRed-monomer (Clonetech)
  - 45 mutations

Crystal packing comparison:



WT DsRed

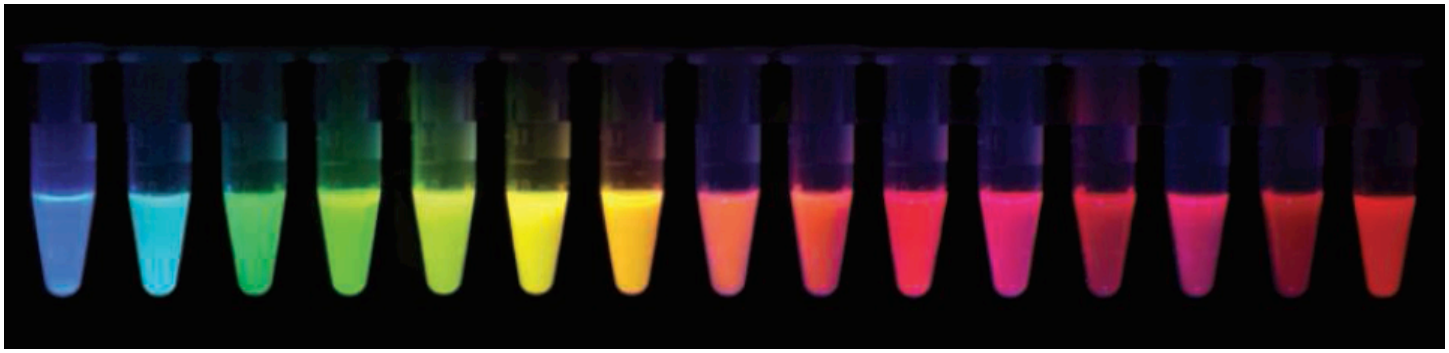


DsRed-monomer (45 mutations)

Strongin *et al.*, **Prot. Eng. Des. & Sel.**, (2007)

A palette of monomeric FPs

- GOAL: generate series of monomeric FPs with range of emission maxima, improved brightness and photostability.
- Derived from mRFP1
- Mutagenesis: targeted (residues within and surrounding chromophore) and random
- Screening: *E. coli* expression; visual and FACS-based assays
- RESULT:

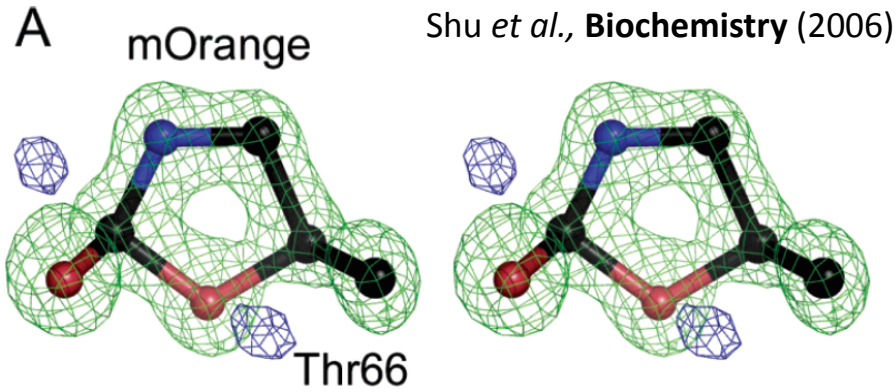


- Modest improvements in brightness and photostability for RED proteins
- Rare orange FP -- mOrange (562 nm emission)

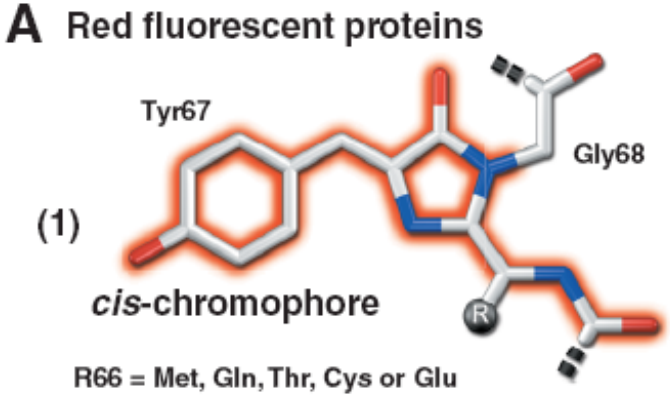
Shaner *et al.,* **Nat. Biotech.** (2004)

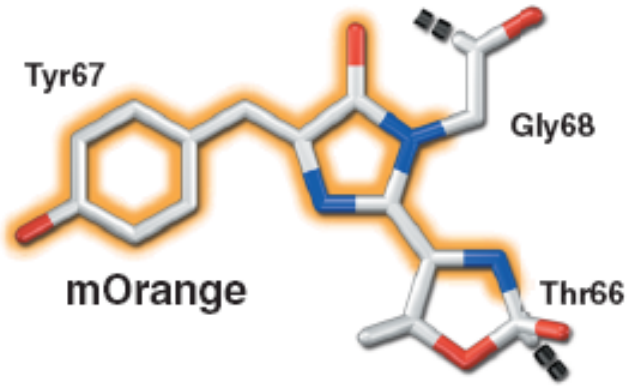
Orange chromophores

- Novel five-membered oxazole ring
- first observed in mOrange



Shu *et al.*, **Biochemistry** (2006)





blue-shift due to elimination of conjugation between mainchain carbonyl of residue 65 and the chromophore

mOrange: **Thr**-Tyr-Gly
mKO: **Cys**-Tyr-Gly
ccaOFP1: **Thr**-Tyr-Gly

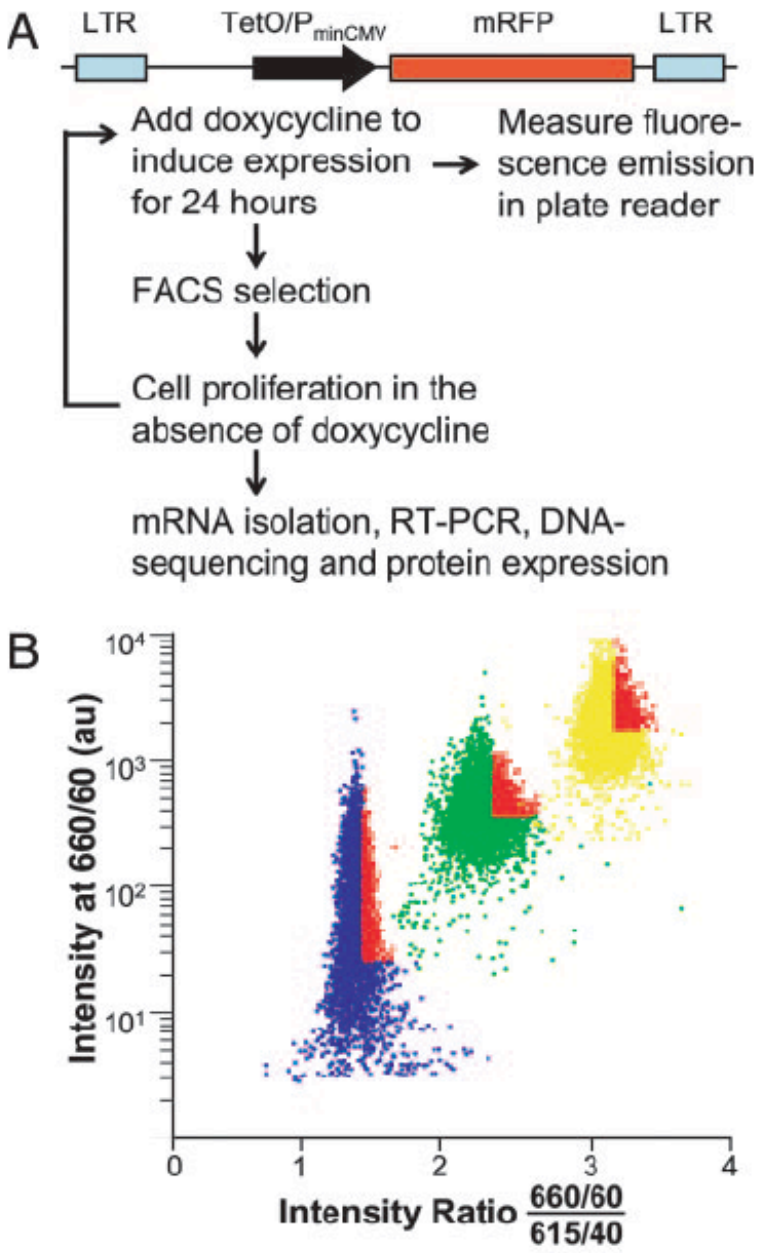
iterative somatic hypermutation

- GOAL: develop system for *in situ* directed evolution of target genes.
  - faster than *in vitro* methods: minimal human intervention during each iteration
  - exploration of larger sequence spaces
- Novel strategy:
  - directed evolution by somatic hypermutation (SHM)
  - Human B lymphocytes can specifically mutate Ig-genes at a rate of ~1 x 10^{-3} mutations/bp/generation (~10^6x higher than rest of genome)
  - integrate single-copy of target gene into Burkitt lymphoma Ramos B cells (a constitutively hypermutating line) --> screen for improved variants...
  - Note: integration at Ig-locus results in higher mutation rates, but SHM can apparently occur at other loci within the genome

Wang *et al.*, **PNAS** (2004)

A far-red FP

- Far-red FPs useful for optical imaging in thick tissues and animals
- Starting gene: mRFP (em = 612 nm)
- After 23 rounds of SHM and FACS (for long-wavelength emission), **mPlum:**
  - 7 substitutions, em = 649 nm
  - Control experiments using traditional strategy in *E coli* gave best em = 632 nm, suggesting that SHM samples larger amount of sequence space



Current limitations of monomeric FPs

- Lower extinction coeff & QY, and faster photobleaching than tetrameric RFPs
  - *Seems to be generally true for monomerized FPs*

- Residual aggregation at high concentrations means that monomeric RFPs still not as robust as GFP as fusion partners *in vivo*...

Optical highlighters

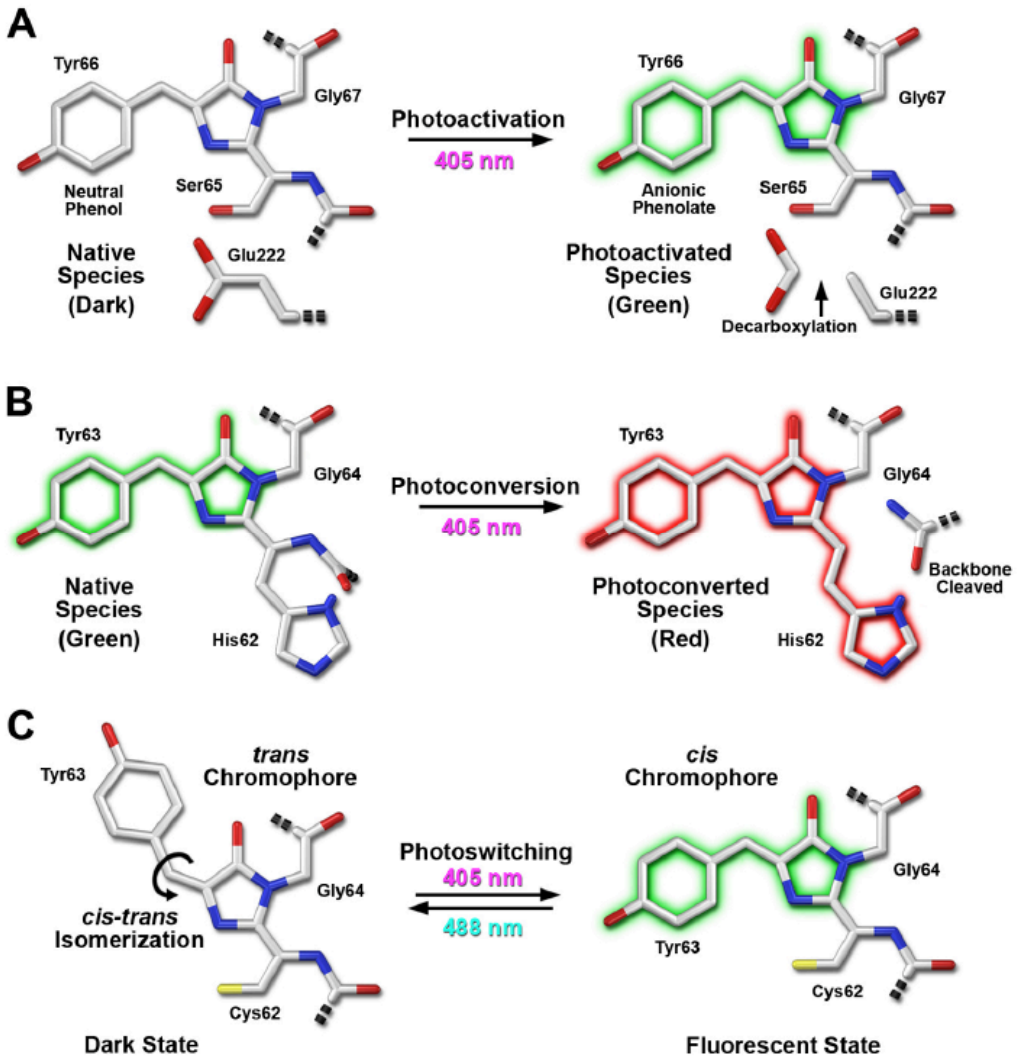


Fig. 4. Photoactivation, photoconversion and photoswitching mechanisms for optical highlighter FPs. (A) Photoactivation of PA-GFP (illustrated) and PS-CFP2 is believed to occur due to decarboxylation of Glu222 followed by conversion of the chromophore from a neutral to anionic state. (B) Green-to-red photoconversion for Kaede, KikGR, Dendra2 and Eos, all of which contain the HYG chromophore, occurs when the FP is illuminated with ultraviolet or violet radiation to induce cleavage between the amide nitrogen and α-carbon atoms in the His62 residue leading to subsequent formation of a conjugated dual imidazole ring system. (C) Photoswitching of Dronpa involves cis-trans photoisomerization induced by alternating radiation between 405 nm and 488 nm. A similar isomerization mechanism is suggested to operate in mTFP0.7 and KFP1.

from Shaner *et al.*, **J. Cell Sci.** (2007)

Outlook

- Great strides have been made over the past decade, both in terms of discovery and engineering. Dozens of FPs now available to researchers...but there is still plenty of room for improvement:
  - monomeric, bright, non-aggregating, photostable, far-red emission, non-toxic

- Other desirable properties:
  - near-IR fluorescent proteins
  - "Switchable" FPs: reversible photoactivation, photoconversion etc.

- Evolution of FPs that possess combinations of these desirable properties will require new HT screening strategies that allow multiple parameters to be evaluated in parallel...