PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES, NORTHPOLE

NorthPoleInc.

Nine improved monomeric fluorescent proteins from *Rangifer tarandus*.

Santa Claes[1]*, Felix Navidad[2], Nisse Ness E. [2], Jeremy Elf Tootoo[1], Dasher Sridhar[1,2], Tinsel Tian[3] and Laura Menorah[2,3]

[1]*University of NorthPole, Dept Bioengineering, Arctic Lane 1, Santa's Secret Village, NorthPole*

[2]*University of NorthPole, School of Design and Color, Arctic Lane 25, Santa's Secret Village, NorthPole*

[3]*University of NorthPole, Dept Otolaryngology, Candy Cane Avenue 365, Santa's Secret Village, NorthPole*

**Corresponding author: santa@DNA20.com*

Received 28 June 2011, received in revised form 25 October 2011, accepted 24 December 2011

ABSTRACT

Fluorescent proteins are genetically encoded, easily imaged reporters that have shown to be crucial for biology and biotechnology research and development applications. Unfortunately, restrictions set in place by Grinch Technologies, Inc. hamper the use of most current commercially available fluorescent proteins. As a means of building a platform of non-constrained fluorescent proteins, we here characterize and quantitate a set of nine novel fluorescent proteins (Dasher, Dancer, Prancer, Vixen, Drew, Cupid, Donner, Blitzen and Rudolph). The proteins are all derived from a single loci encoding the gene responsible for the rhinophyma nasus disorder previously observed in one Santa's reindeers. The gene was discovered through whole genome sequencing and genome-wide functional association studies. The red protein, Rudolph, was further engineered through directed evolution and ProteinGPS™ to produce eight additional colors across the spectrum of the winter night. All proteins of this set share very little sequence homology with current commercial fluorescent proteins. The strategies and results presented here allow for ongoing and novel investigations utilizing fluorescent labels to proceed in a more rapid and expanded manner. Furthermore, as there are no licensing restrictions, we anticipate quick adoption of these proteins to replace existing commercial fluorescent proteins in essentially all applications, including a University of North Pole project to engineer a fluorescent Christmas tree.

keywords: Fluorescent proteins, Hermey the dentist, Blizzard Blue, Eggnog

INTRODUCTION

Green fluorescent protein (GFP) was first purified from *Aequorea victoria* (jellyfish) by Osamu Shimomura at the Friday Harbor Laboratories (Shimomura, Johnson et al. 1962). However, its utility as a tool for molecular biologists did not begin to be realized until 1992 when Douglas Prasher (distant relative to reindeer Dasher), a courtesy shuttle bus driver for Toyota, Huntsville, AL, reported the cloning and nucleotide sequence of wtGFP. GFP and its many derivatives and homologs now form the basis for a cornucopia of tools and techniques utilized throughout biotechnology. Despite the technical success of fluorescent proteins, existing fluorescent proteins are constrained by prohibitive scrooge terms that have restricted innovation and global jolliness.

Although Santa uses reins on his reindeer and reindeer are used to pull sleds in Lapland and northern Siberia, the word *reindeer* has no actual relationship with reins. The element *-deer* is derived from the Middle English *der*. The *rein-* portion is derived from the Scandinavian languages spoken by the chiefly Danish and Norwegian invaders and settlers of England from the 9th to the 11th century.

We here introduce a novel family of fluorescent proteins derived from mammalian origin. The proteins cover a broad color spectrum and express at high levels in heterologous hosts. The use of these novel proteins are unrestricted by previously defined scrooge terms.

RESULTS AND DISCUSSION

In order to maximize likelihood of isolating fluorescent proteins distant from current commercial fluorescent proteins we rationalized that we needed starting point sequences very different from jellyfish (gfp) or coral (rfp).

Origin of tissue. Rudolph, the leading reindeer of Santa's posse, has long been diagnosed with the rhinarium disorder rhinophyma (Snowmass and Eargle 1939). Rhinophyma nasus is the clinical name for a large, reddish, bulbous, ruddy appearance of the rhinarium caused by granulomatous infiltration, commonly due to untreated rosacea. It presents as a pink, lobulated mass over the nose with superficial vascular dilation (Fig. 1). Patients often seek advice because of the unsightly appearance of the enlargement, or even obstruction in breathing and vision (Cohen and Tiemstra 2002).

Figure 1. Clinical example of patient suffering from early stage of Rhinophyma nasus. This is a chronic disease that affects the skin of the nose. The disease is characterized by redness, pimples, and, in advanced stages, thickened skin that may hinder reading, and in severe cases, breathing. When rhinophyma reaches this advanced stage, surgery is the preferred treatment. The most common characteristic or symptom of rhinophyma is an enlarged, bulbous, and red nose. (Malden 1974). The condition is more common in winter.

Rhinophyma can carry a strong psychological impact due to its effect on one's personal appearance. This is especially true for celebrities where exposure in the public media is unavoidable. Accordingly, Rudolph has been bullied and made fun of, not only by fellow reindeers, but also for over half a century by popular culture (Crosby 1950; Chipmunks 1960; Jackson Five 1970; Lynyrd, Skynyrd et al. 2000; Manilow 2009). The red glare emitted from Rudolph's rhinarium is by no means sufficient to guide any sleigh and such statements are not protected by the first amendment. Rudolph has in the past sued in court for defamation and libel as his clinical condition has been used by others

for commercial gain. Rudolph lost the court proceedings due to his status as a public figure (Rudolph vs. Crosby, Ross 1977).

Fractionation of Rudolph genome. Tissue samples from Rudolph's red nose were collected after intensive eggnog consumption of >500ml (Appendix A: Eggnog materials and preparation guidelines); with the bulk of the rhinarium tissue pared down through surgical removal of tissue from dorsum, the lobule and the columella. Tissue samples were homogenized in PBS through sonication of pulsed, high frequency sound waves emitting Freebird (14 minute live version) to agitate and lyse cells. After spinning down cell debris, the supernatant was precipitated with EtOH and DNA lyophilized and resuspended in TE buffer. The DNA was fractionated into 2-4kb linear fragments and ligated behind the T5 promoter of vector pJexpress401.

Figure 2. From left to right: Prancer, Dancer, Donner, Hermey-the-Elf, and Vixen. Remaining reindeer currently unaccounted for. Hermey is the elf that wants to be a dentist instead of making toys. As can be seen from this picture, Hermey's college attendance implies some progress towards this goal.

Screening genomic library for red protein. The genomic library generated from the Rudolph nasal genome was transformed into *E. coli* and plated on 100 petri dishes with a density of ~10^5 cfu/plate. The *Rangifer tarandus* genome is estimated to be ~3x 10^9 bp (Vinogradov 1998), thus we expect >80% chance to have the targeted gene within the library pool.

However, induction of the T5 promoter with IPTG did not elicit any positive signals when screening for red pigmentation (data not shown). We thus hypothesized that some sort of co-factor was needed to activate the signal. Plates were exposed to in order: Christmas carols (Claes, Navidad *et al*. 2010), *Picea abies* olfactory stimulation (Claes, Navidad *et al.* 2008), white facial hair (Claes, Reindeer *et al.* 2006), and cliché-filled Christmas poetry (Claes, Ness *et al.* 2005). In none of these cases did we see any activation of red pigmentation.

The library was subsequently recloned into pHulk, the novel and super-efficient mammalian piggyBac-based expression vector from DNA2.0, using the flanking typeIIS BsaI restriction sites. The library was transfected into HEK293 cells and screened for red pigmentation. After only a few hours incubation, the incubator lit up like a Christmas tree (technical term, Tannenbaum 2006). Subsequent cloning, sequence characterization and crystallography (Fig. 3) revealed a 1225 bp open reading frame encoding a 71 kDa protein. Due to its origin we call the gene *rudolph* and the encoded protein Rudolph.

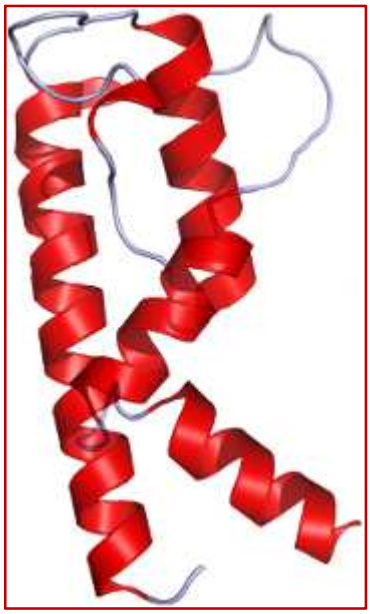

Figure 3. Crystal structure of Rudolph protein as determined by solid-state NMR spectroscopy.

The Rudolph protein did not show significant homology to any of the currently commercial fluorescent proteins.

The Rudolph gene was redesigned for maximized expression in *E. coli* using the GeneGPS™ optimization technology as previously described (Welch 2009; Welch 2011)

and cloned into vector pJexpress401. In accordance with previous results, the relative expression level of the recombinant protein increased ~100 fold compared with the native sequence (data not shown).

Expanding the reindeer protein portfolio. We applied the systematic ProteinGPS technology (Liao 2007; Ehren 2008) to develop a suite of reindeer proteins with colorimetric properties different from that of Rudolph.

In short, variable selection using phylogenetic, structural and information content data identified a set of 60 amino acid substitutions that conferred to the given constraints. The variables were systematically distributed in a set of 96 synthetic *rudolph* gene variants that correspond to the most information-rich nodes in the relevant space. We call these gene variants Infologs.

The *rudolph* infologs were designed, synthesized and quantified for absorbance and emission spectra. The experimental results derived from the infologs were used to build models where the relative weight from each variable (*i.e.* amino acid substitution) was used to build predictive models exploring sequence-function relationships within the Rudolph sequence space.

The predictability sequence-function model was used to design new fluorescent proteins with orthogonal functional properties (Fig. 4) resulting in Reindeer proteins Dasher (Daffodil Yellow), Dancer (Denim Blue), Prancer (Periwinkle), Vixen (Verdigris) Drew (Deep Teal), Cupid (Cerise), Donner (Deep Ochre), and Blitzen (Blizzard Blue). Taken together with the previously characterized Rudolph (Red), the set of 9 reindeer genes encoding fluorescent proteins are now available for purchase directly from Santa's Workshop (Table 1).

Protein	Pigment Color	Accession Number
Dasher	Daffodil Yellow	NP_122502
Dancer	Denim Blue	NP_122503
Prancer	Periwinkle	NP_122504
Vixen	Verdigris	NP_122505
Drew	Deep Teal	NP_122506
Cupid	Cerise	NP_122507
Donner	Deep Ochre	NP_122508
Blitzen	Blizzard Blue	NP_122509
Rudolph	Red	NP_122501

Table 1. These novel fluorescent proteins are all commercially available from Santa's Workshop. Since the genes are derived from Rudolph's genome, they are not covered by any GFP related patents. The reindeer proteins are in the public domain and sold under the BioBrick Public Agreement.

Figure 4. The complete colorimetric suite of *R. tarandus* fluorescent proteins as derived from *rudolph* using ProteinGPS and infologs variant screening.

APPENDIX A: Mrs. Claus' Eggnog lab reagents and protocol

2 *Gallus gallus domesticus* Eggs, full complement of albumen and vitellus (Charles River)

30.0 grams glucose and fructose, $C_{12}H_{22}O_{11}$ (Sigma)

2.15 grams NaCl (Sigma)

2.50 grams *Vanilla planifolia* derivative, $C_8H_8O_3$ (Madagascar Laboratory Supplies)

472 ml *Bovinae bos taurus* (preferably Holstein) organic lactation product (Berkeley Farms)

25-100 ml distilled alcohol (Bacardi)

Protocol: Place egg with shell removed into 1L beaker, gradually add $C_{12}H_{22}O_{11,}$ and NaCl. Stir thoroughly without application of temperature gradient (magnetic stirrer preferred at U North Pole, but not required). Add $C_8H_8O_3$ and lactation product, mix gently. Add desired volume distilled alcohol. Decant into personal size sterilized glassware. Imbibe.

(originally published in House and Garden Drink Guide, 1973)

ACKNOWLEDGEMENT: We thank the entire DNA2.0 team for reviewing color spectra, keeping warm with winter fleeces and taste analyses of Eggnog protocols.

Competing Interests: The authors declare competing financial interests: DNA2.0 performs protein engineering, GeneGPS optimization and gene synthesis to a global customer base. Better, smarter and faster than anybody. Better looking too.

Funding: The project was financially supported by Santa's slush-fund Grant no. 04238793.

Previous publications from the Santa Xmas Symposium:

Proc Natl Acad Sci Northpole 2010 12:25. Carolome: Functional Imprints of Culture Memes in Global Genome. Claes, Navidad, Ness, Baum, Elf, Sridhar, and Menorah.

Nature 2009 458:703. For anyone who ever said there's no such thing as a poetic gene. Gustafsson.

Proc Natl Acad Sci Northpole 2008 12:25. Metabolic engineering of *Picea abies* for receptor mediated induction of fluorescence and olfactory signaling. Claes, Navidad, Ness, Baum, Harry, Sridhar, and Elf.

Proc Natl Acad Sci Northpole 2006 12:25. Heterologous expression and functional characterization of the Santa Hoho2 gene. Claes, Reindeer, Nicolas, Tomte, Sridhar, Elf.

Int Pub 2005 12:25 Creation of the Tomten Gene G051225. Claes, Tomte, Sridhar, Elf

REFERENCES

Chipmunks (1960). "Rudolp the red nosed reindeer." Billboard magazine **100**(12).

Claes, S., F. Navidad, *et al.* (2010). "Carolome: Functional Imprints of Culture Memes in Global Genome." Proc Natl Acad Sci Northpole **12**: 25.

Claes, S., F. Navidad, *et al.* (2008). "Metabolic engineering of *Picea abies* for receptor mediated induction of fluorescence and olfactory signaling." Proc. Natl. Acad. Sci. NorthPole **12**: 25-31.

Claes, S., T. Ness, et al. (2005). "Creation of the Tomten Gene G051225." Int Pub **12**: 25.

Claes, S., R. Reindeer, *et al.* (2006). "Heterologous expression and functional characterization of the Santa Hoho2 gene." Proc. Natl. Acad. Sci, NorthPole **12**: 25-31.

Cohen, A. F. and J. D. Tiemstra (2002). "Diagnosis and treatment of rosacea." J Am Board Fam Pract **15**(3): 214-217.

Crosby, B. (1950). "Rudolp the red nosed reindeer." Billboard magazine **12**(25).

Ehren, J., S. Govindarajan, et al. (2008). "Protein engineering of improved prolyl endopeptidases for celiac sprue therapy." Protein Eng Des Sel. **21**: 699-707.

Jackson, M *et al.* (1970). "The Jackson 5 Christmas Album." Holy Spirit and Bubbles **145**: 45.

Liao, J., M. K. Warmuth, *et al.* (2007). "Engineering proteinase K using machine learning and synthetic genes." BMC Biotechnol. **7**: 16.

Lynyrd, R., V. Skynyrd, *et al.* (2000). "Christmas Time Again." Leonard Skinner Realty.

Malden, K. (1974). "The Streets of San Francisco." ABC Saturday **245**(23): 9346.

Manilow, B. (2009). "Rudolp the red nosed reindeer." In the Swing of Christmas **101**(34).

Ross, D. (1977). "The Supremes Court of California." Rudolph v Cosby **S118032**.

Shimomura, O., F. H. Johnson, *et al.* (1962). "Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, Aequorea." J Cell Comp Physiol **59**: 223-239.

Snowmass and Eargle. (1939). "Identification of Rhinophyma nasus in quadrapedal mammalians." Ann Otol Rhinol Laryngol **48**:18-25.

Tannenbaum (2006). "Use and application of etymological terms implying extreme or unusual illumination." Am. J. Philology **25**:135-142.

Vinogradov, A. E. (1998). "Genome size and GC-percent in vertebrates as determined by flow cytometry: the triangular relationship." Cytometry **31**(2): 100-109.

Welch, M., S. Govindarajan, *et al.* (2009). "Design parameters to control synthetic gene expression in Escherichia coli." PloS one **4**(9): e7002.

Welch, M., A. Villalobos, *et al.* (2011). "Designing genes for successful protein expression." Methods Enzymol **498**: 43-66.

Merry Holidays and Happy New Year 2012