Why FRET over genomics?

DINO A. DE ANGELIS
Cellular Biochemistry and Biophysics Program, Memorial Sloan-Kettering Cancer Center, New York, New York 10021

De Angelis, Dino A. Why FRET over genomics? Physiol. Genomics 1: 93–99, 1999.—Genetic information is being uncovered quickly and in vast amounts through the largely automated sequencing of genomes from all kinds of organisms. As this information becomes available, enormous challenges are emerging on three levels: first, functions will have to be assigned to individual gene products; second, factors that influence the expression level of these gene products will have to be identified; and third, allelic variants that act alone or in combination to give rise to complex traits will have to be characterized. Because of the sheer size of genomes, methods that can streamline or automate these processes are highly desirable. Fluorescence is an attractive readout for such high-throughput tasks because of the availability of equipment designed to detect light-emitting compounds with great speed and high capacity. The following is an overview of the achievements and potential of fluorescence resonance energy transfer (FRET) as applied in three areas of genomics: the identification of single-nucleotide polymorphisms, the detection of protein-protein interactions, and the genomewide analysis of regulatory sequences.

green fluorescent protein; proximity imaging; high-throughput screening; single-nucleotide polymorphism; fluorescence-activated cell sorting; digital imaging spectroscopy

FLUORESCENCE RESONANCE ENERGY TRANSFER

Fluorescence resonance energy transfer (FRET) is a quantum mechanical phenomenon that occurs between a fluorescence donor (D) and a fluorescence acceptor (A) in close proximity (usually <100 Å of separation) if the emission spectrum of D overlaps with the excitation spectrum of A (Fig. 1A) (15). Under optimal FRET conditions, illumination at the excitation wavelength of D results in transfer of energy, not photons, to A. If A is nonfluorescent, it can be referred to as a quencher (Q); in this case, FRET between D and Q results in a net decrease in photon emission from D (Fig. 1C). If A is also fluorescent, the decrease in emission from D is concomitant with an increase in fluorescence at the emission wavelength of A (Fig. 1B). The efficiency of energy transfer (E) decreases very rapidly with increasing distance (R) between the donor and acceptor, according to the relationship $E \propto \left[1 + \left(R/R_0\right)^6\right]^{-1}$, where $R_0$ is the distance at which $E$ is 50%. To illustrate this point, simply making $R = 2 \times R_0$ decreases FRET efficiency from 50% to ~1%.

FRET-induced changes in fluorescence intensity (when A is a nonfluorescent quencher, Q) or in the ratio of emission intensities of the donor and acceptor (when A is fluorescent) have been widely exploited in biology to monitor proximity relationships between appropriately labeled macromolecules (31, 33). Absolute distances between donor and acceptor pairs can be measured with FRET; it has been coined a “spectroscopic ruler” (33). However, absolute distance measurements are not usually attempted because FRET efficiency is also influenced by the angle between the dipole moments of D and A, a parameter that is often difficult to quantify. The most common application of the technique is to dynamically monitor the relative distance between D and A after changes induced in the system under study.

ANALYSIS OF SINGLE-NUCLEOTIDE POLYMORPHISMS WITH FRET

One of the greatest challenges facing geneticists is the analysis of complex traits and diseases. Complexity arises because there is no simple correspondence between genotype and phenotype, for instance, when mutations in any one of several genes result in the same phenotype or when mutations in several genes are required to give rise to a particular trait (20). One way of zeroing in on candidate genes is through genetic
association studies that compare the prevalence of a marker (or a set of markers) between affected and nonaffected individuals. Single-nucleotide polymorphisms (SNPs) are genetic markers that are densely distributed throughout the genome, at an average frequency of 1 per 1,000 base pairs (8). Organizations such as the National Human Genome Research Institute (http://www.nhgri.nih.gov/) and the recently created SNP Consortium (23) are creating public databases cataloguing several hundreds of thousands of these markers. These databases, coupled with efficient methodologies to detect SNPs, will facilitate the identification of genes responsible for complex disorders such as diabetes, hypertension, or schizophrenia.

Although SNPs can be successfully analyzed by traditional methods based on gel separation (Sanger or Maxam-Gilbert sequencing, restriction fragment length polymorphism), several more efficient/automatable alternatives have been devised (19). Most methods are based on differences in hybridization to a given oligonucleotide probe between a perfectly matched and a mismatched sequence. For instance, DNA chips consist of thousands of oligonucleotide probes arrayed at high density on a solid support in a predetermined order (37). DNA samples are amplified by PCR using fluorescent nucleotide analogs; these labeled PCR products can only hybridize to perfectly matching oligonucleotide sequences immobilized on the chip. Images of the hybridized fluorescent probes are obtained by confocal microscopy of the chip, allowing direct comparison between patient and control hybridization patterns. The main advantage of this approach is that it enables the simultaneous detection of thousands of SNPs from a given genomic DNA sample; however, the fact that the probes are fixed to a chip that is difficult to manufacture hinders the detection of newly catalogued SNPs.

More flexibility is provided by simple homogeneous assays that allow the detection of SNPs in solution; four of these are based on FRET and have been reviewed in detail elsewhere (19). The TaqMan assay (21) is a protocol that utilizes short oligonucleotide probes that undergo FRET in their intact state because they are labeled at each end with a fluorescence donor and acceptor pair (Fig. 2A). Even in probes longer than 30 nucleotides, significant FRET can occur because of hydrophobic interactions between D and A (5). The probe is designed to hybridize to a sequence being amplified in a PCR reaction using Taq polymerase (downstream of one of the primers). If there is a perfect match between the probe and the target, the 5’-exonuclease activity of Taq polymerase can digest the hybridized TaqMan probe during the elongation cycle, separating D from A and resulting in a decrease in FRET.

The “molecular beacons” of Tyagi and Kramer (36) represent a second approach. These oligonucleotide probes are chemically modified with a fluorescence donor (EDANS) at their 5’ end and a nonfluorescent quenching acceptor (DABCYL) at their 3’ end. In the
absence of a perfectly matched target, they assume a stem-and-loop structure in solution: the loop is a DNA strand complementary to the target, and the stem is formed by intramolecular base-pairing of short complementary sequences at each end of the loop (Fig. 2B). This hairpin conformation positions D and Q in extremely close proximity, much closer than a pair of fluorophores at opposite ends of a randomly coiled oligonucleotide; this effectively quenches donor fluorescence. In the presence of a perfectly matched sequence, the oligonucleotide undergoes a conformational change that allows the hairpin loop to hybridize to the target, separating D from Q and resulting in a fluorescence increase (up to 900-fold; Fig. 2B). The ability of molecular beacons to form hairpin structures significantly enhances their specificity compared with standard oligonucleotide probes of the same size, allowing them to readily distinguish between a perfect match and a single base mismatch (3). Recently, molecular beacons with seven additional donors ranging from blue to red fluorescence emission have been developed, permitting the simultaneous detection of several targets within the same tube (35).

Dye-labeled oligonucleotide ligation (DOL) (7) is a third FRET-based technique that relies on the extreme sensitivity of DNA ligase to mismatches close to the ligation site. The method utilizes two oligonucleotides, one labeled with a fluorescence donor (D) and the other with an acceptor (A); this effectively quenches donor fluorescence. In the presence of a perfectly matched sequence, the oligonucleotides are designed to hybridize adjacent to a polymorphic nucleotide, allowing their ligation in the presence of a donor-acceptor pair (Fig. 2C). A mismatch located on the 3’ end of the primer will result in failure of the ligation, thus preventing FRET.

Finally, in template-directed dye-terminator incorporation (TDI) (6), a donor-labeled oligonucleotide primer designed with a donor-labeled nucleotide terminator analog corresponding to SNP of interest is added to a DNA target. In a mismatch situation, the analog is not incorporated at the 3’ end of the primer and FRET does not occur. In a perfect match, the analog is incorporated and FRET results.

Fig. 2. Four FRET-based approaches for analysis of single-nucleotide polymorphisms (SNPs) in solution. A: TaqMan assay. A, left: TaqMan probe does not match target DNA (gray horizontal line; mismatch represented by “X”) and undergoes FRET in the intact state upon illumination of D; Taq polymerase (Taq) can extend a PCR primer (arrow) located upstream of mismatched target sequence. A, right: hybridized TaqMan probe is digested by 5’-exonuclease activity of Taq polymerase during extension of PCR primer, resulting in separation of D from A and disruption of FRET signal. B: molecular beacons. B, left: in presence of a mismatched target, the probe adopts a stem-and-loop structure; because of proximity of 5’ and 3’ ends of probe, donor (D) fluorescence is directly coupled to nonfluorescent quenching acceptor (Q), resulting in very low levels of fluorescence emission from D. B, right: in presence of correct target, molecular beacon hybridizes perfectly to the target; D and Q are separated, giving rise to a disruption of FRET and an increase in emission from D. C: dye-labeled oligonucleotide ligation (DOL). C, left: a mismatch at junction between 2 contiguous oligonucleotides [one labeled with a fluorescence donor (D) and the other with an acceptor (A)] prevents DNA ligase (Lig) from joining the 2 fragments together, preventing energy transfer between D and A. C, right: a perfect match allows ligase to link the oligonucleotides, resulting in FRET. D: template-directed dye-terminator incorporation (TDI). DNA polymerase (Pol) and an acceptor-labeled nucleotide terminator analog corresponding to SNP of interest are added to a fluorescence donor-labeled oligonucleotide hybridized to a DNA target. D, left: in a mismatch situation, the analog is not incorporated at the 3’ end of the primer and FRET does not occur. D, right: with a perfect match, the analog is incorporated and FRET results.
DNA polymerase and an acceptor-labeled nucleotide terminator analog corresponding to the SNP of interest are added; FRET results only if the analog is incorporated at the 3’ end of the primer, indicating a perfect match. Although TDI is extremely easy to optimize, it necessitates addition of specific reagents twice after the initial reaction setup; the other three FRET-based methods are “walk-away” assays that only require proper thermal cycling of the initial reaction mixture.

Although these four methods do not approach the degree of parallelism or multiplexing offered by high-density DNA arrays, where simultaneous analysis of thousands of SNPs per reaction is customary, they are much more flexible: new SNPs can be detected without having to remanufacture a custom DNA chip because the specific probes and reagents required are easy to obtain or synthesize. Moreover, these FRET-based assays are fast and simple: with the exception of TDI (see above), all necessary reagents are initially combined and changes in FRET can be monitored in real time during the course of the assays without having to process the samples further and without the help of expensive imaging equipment.

ANALYSIS OF CODING SEQUENCES: PROTEIN-PROTEIN INTERACTIONS ANALYZED WITH FRET AND PRIM

Because most cellular processes are carried via protein assemblies, important clues to the function of a novel gene product can be obtained by identifying the set of proteins with which it interacts. On a genomewide basis, such information would constitute what has been termed “protein linkage maps” (13). Methods able to streamline the identification of protein-protein interactions are necessary to cope with the large amount of genomic data being generated (4). In the yeast two-hybrid system (14), the DNA binding domain of a transcription factor is fused to a sequence of interest in a vector, and a DNA library is fused to the transcriptional activation domain in a second vector. Protein-protein interactions are detected in cotransformed yeast colonies on the basis of reconstituted transcriptional activity, which is coupled to a functional effect (usually survival on selective media). Inserting a library in each of the vectors allows the mapping of protein-protein interactions on a genomewide basis, as was demonstrated with the bacteriophage T7 genome (2); larger genomes can be handled with multiple-round two-hybrid screens (16). In phage-display technologies, cDNA libraries are expressed on the surface of phage particles directly through fusion to a coat protein (18) or indirectly via strong noncovalent interaction with a modified coat protein (10). Typically, the phage libraries are presented to a target immobilized on a solid support; nonbinding phages are washed away and specific binding phages are enriched and analyzed (27). A combination of phage and bacterial display has recently been developed that can potentially detect interactions not only between a target protein and a library of coding sequences but also between two libraries (4, 22).

Since the initial cloning of green fluorescent protein from the jellyfish Aequorea victoria and its expression in different cells and organisms (5, 30), the coding sequence has been mutagenized extensively to improve or alter some of the spectral and biophysical properties of the original. The various mutants have recently been grouped under seven classes (34), including variants emitting light of different wavelengths: green, blue, cyan, and yellow fluorescent proteins (GFP, BFP, CFP, and YFP, respectively). The advent of these spectral variants has opened up another platform for the detection of protein-protein interactions: appropriately selected mutants can serve as D-A pairs in a FRET scheme. In proof-of-concept experiments (17, 25), BFP (D) and GFP (A) were linked together via spacers of 20 or 25 amino acids containing a protease cleavage site. As a result of forced proximity, substantial intramolecular energy transfer occurred from BFP to GFP in the uncleaved species; addition of the specific protease led to the disappearance of the FRET signal. GFP donors and acceptors have since been used in numerous FRET-based applications (29). Currently, the optimal D-A pair consists of optimized CFP and YFP variants, respectively (26). The drawbacks of using a 28-kDa protein as a fluorescent tag (large size, possible disruption of protein function) are outweighed by the advantages of specifically and homogeneously labeling a given protein and targeting it to any subcellular location for which there is a known consensus sequence.

The ease with which given coding sequences can be fused to GFP suggests that high-throughput methods may be devised to detect protein-protein interactions by FRET on a genomewide basis (1) (Fig. 3A). In such a scheme, an appropriate host (yeast or bacteria) would be transformed with a vector containing a given coding sequence (or a library) fused with CFP and a second vector with a library fused to YFP and then grown on selective media. Only a small fraction of the surviving clones are expected to have an altered fluorescence emission ratio, indicative of a protein-protein interaction; the major task would then be to identify and isolate these positive clones for further analysis. Fortunately, existing technologies can be adapted for such a task: isolation of positive clones from liquid cultures can be performed by iterative cycles of fluorescence-activated cell sorting (FACS), as was previously done for the identification of strongly fluorescent variants of GFP (9). Solid-phase screening directly from agar plates containing the transformants is also possible: digital imaging spectroscopy (DIS) can record spectral information from any two-dimensional surface in a massively parallel fashion (39). Thus DIS could be configured to report the CFP-to-YFP emission ratio of every pixel in a given field, allowing simultaneous imaging of up to 25 agar plates in certain setups. Using either FACS or DIS, positive clones could be expanded, and the protein-protein interactions giving rise to changes in emission ratio could be characterized.

Whereas FRET is ideally suited to detect heterotypic protein interactions, proximity imaging (PRIM) is a novel technique that can optically report homotypic interactions of GFP-labeled proteins (12). PRIM employs the original wild-type GFP with two amino acid

http://physiolgenomics.physiology.org
substitutions that help the protein fold at 37°C (32) [henceforth referred to as ttGFP (thermotolerant GFP)]. Unlike the GFP variants with simplified excitation spectra used in FRET, ttGFP possesses two excitation peaks (at 395 and 475 nm) and one emission peak (at 510 nm). The ratio of intensities of the excitation peaks, i.e., the excitation ratio, is the same for any ttGFP-labeled protein that is monomeric. However, on dimerization, deviations from this excitation ratio occur (Fig. 3B), detectable in vitro and in vivo (12, 24). Unlike FRET, which is a quantum mechanical interaction between any D−A pair, PRIM is GFP specific and results from orientation-dependent structural interactions between two ttGFP modules that affect their excitation ratio. High-throughput identification of proteins or protein domains capable of self-association could be achieved by adapting FACS or DIS for excitation ratio imaging. Exhaustive screening for homotypic interactions by PRIM would be unidimensional, requiring the expression of a single library of coding sequences fused to ttGFP. By comparison, exhaustive screening for pairwise heterotypic interactions (with any of the methods mentioned above) is a two-dimensional combinatorial process in which each coding sequence has to be tested against every other coding sequence, necessitating the expression of two libraries. A genome of 50,000 coding sequences thus would minimally require the screening of 50,000 clones for homotypic interactions and 0.5 × (50,000)² or 1,250,000,000 clones for heterotypic interactions.

Each of the above-mentioned techniques suffers from shortcomings in certain respects. Phage display techniques are limited to the study of relatively small- to medium-size proteins lacking eukaryotic posttranslational modifications (4). The yeast two-hybrid system is known to generate a high number of false positives due to nonspecifically interacting proteins and transcriptional autoactivation by the DNA-binding fusion protein (2). Moreover, interactions have to take place within the yeast nucleus. In contrast, GFP-based FRET or PRIM measurements can be performed in bacteria, yeast, or other cell types; moreover, interactions can theoretically be detected within any subcellular organelle. However, GFP-based measurements are hampered by several problems. 1) The GFP mutants used as FRET donors typically photobleach at much faster rates than acceptors, complicating the analysis somewhat (26); with PRIM care must be taken not to illuminate samples too intensely in the UV range because this can cause excitation ratio changes through a photoisomerization effect (11). 2) A large number of labeled molecules have to be expressed to be detectable, preferably at similar levels in the case of FRET with CFP- and YFP-tagged proteins; the detection threshold

![Fig. 3. Analysis of protein-protein interactions with green fluorescent protein (GFP)-based probes. A: detection of heterotypic protein interaction with FRET between cyan fluorescent protein (CFP) fusion to protein X and yellow fluorescent protein (YFP) to protein Y. Under optimal proximity and angular conditions, interaction between X and Y (right) causes a decrease in intensity of CFP fluorescence concomitant with an increase in intensity of YFP fluorescence. B: detection of homotypic protein interaction of protein X using proximity imaging (PRIM) with thermotolerant GFP (ttGFP). Under optimal proximity and angular conditions, self-association of X will produce a change in excitation ratio of ttGFP (depicted by change in shape of shading within ttGFP rectangle).](http://physiolgenomics.physiology.org)

http://physiolgenomics.physiology.org

![Fig. 4. Analysis of regulatory regions using β-lactamase cleavage of CCF2. A: CCF2 consists of blue-emitting coumarin (donor D) linked to green-emitting fluorescein (acceptor A) via a β-lactam bond. In cells lacking β-lactamase, excitation of D results in fluorescence emission predominantly from A due to FRET in the intact probe. B: in presence of β-lactamase (scissors), CCF2 is cleaved, disrupting FRET; excitation of D results in emission predominantly from D.](http://physiolgenomics.physiology.org)
is 200 nM or ~2,000 copies/cell for yeast and 200 copies/cell for E. coli (28). This can pose certain problems, for instance in cases where a protein is toxic to the host above a certain threshold concentration. 3) Stringent distance and angular requirements for FRET and PRIM signals might contribute to a large number of false negatives, i.e., proteins that actually interact but do not produce a change in the emission ratio (FRET) or excitation ratio (PRIM) because of unfavorable geometry of the GFP modules within the protein complex. Despite these shortcomings, screening protein-protein interactions using GFP-based techniques can be a valid alternative, especially when screening for homotypic protein interactions or when eukaryotic posttranslational modifications or specific subcellular localization of the target proteins is required.

**ANALYSIS OF REGULATORY SEQUENCES WITH FRET**

Functional genomics would be incomplete without an understanding of the way in which regulatory sequences dynamically control the level of gene expression in time and space and in response to specific signals. A powerful FRET-based method was devised with the potential to identify regulatory sequences that respond positively or negatively to any given stimulus in vivo (38). In this system, cell populations are stably transected with a linearized gene-trap vector consisting of two parts: a promoterless reporter gene encoding β-lactamase and a constitutively expressed neomycin-selectable marker. A splice acceptor sequence at the 5’ end of the vector ensures preferential insertion within coding regions, thus placing the promoterless β-lactamase under control of endogenous promoters throughout the genome. Stably transected cell populations are then incubated with a cell-permeable version of CCF2, a small molecule consisting of a blue-emitting fluorescence donor (coumarin) and a green-emitting acceptor (fluorescein) chemically linked together via a β-lactam ring. Because of the close proximity of the fluorophores within CCF2, intramolecular energy transfer occurs, yielding predominantly green fluorescence emission upon donor excitation (Fig. 4A). β-Lactamase cleaves CCF2 and results in predominantly blue fluorescence emission by disruption of FRET (Fig. 4B). After cell loading with CCF2, β-lactamase-positive (blue) and -negative (green) clones can be sorted by flow cytometry. Because the signal is based on enzyme-substrate kinetics, the method is much more sensitive than a GFP-based assay: J urkat T cells containing 1,000 copies of β-lactamase can accurately be scored positive; by increasing incubation time, as few as 100 copies can be detected, in contrast to the 30,000 copies of GFP required in a similar cell (38, 40).

In the report by Whitney and colleagues (38), a library of 1.5 million independent, stably transected J urkat T-cell clones was established; cells were loaded with CCF2 and sorted according to two schemes. In the induced sorting mode, β-lactamase-negative cell pools (green) were treated with various pharmacological agents, either alone or in combination, and resorted; clones that upregulated β-lactamase (blue) were isolated and analyzed. In the repressed sorting mode, the converse was performed: β-lactamase-positive clones (blue) were drug treated and resorted; clones that downregulated β-lactamase in response to pharmacological intervention (green) were selected. After isolation and expansion of the responsive clones, 14 genes were identified that either upregulated or downregulated the reporter, 10 of which were completely novel. In principle, this method can be extended to any propagating cell line and allows the identification of genes whose expression levels can be externally modulated.

**DISCUSSION**

Because the detection of fluorescent compounds can be automated, several aspects of genomics can benefit from techniques based on FRET. The identification of SNPs in solution, for which four different FRET-based assays have been developed, remains the most successful application (19). Analysis of regulatory regions throughout entire genomes is still in its infancy, but the work performed with β-lactamase and CCF2 in J urkat T cells shows enormous promise (38). Finally, genome-wide detection of protein-protein interactions is possible through the fusion of cDNA libraries to spectral variants of GFP capable of detecting heterotypic or homotypic interactions. Several alternatives to FRET-based assays for high-throughput analysis in each of these areas are available, each with its particular strengths and weaknesses; a comprehensive picture of functional genomics will likely require contributions from all existing techniques and novel, yet-to-be-devised approaches.

I thank James McNew, Boris Zemelman, Christine Hughes, and Matt Tector for critical reading of the manuscript.

D. A. DeAngelis was supported by a postdoctoral fellowship from the Fonds de la Recherche en Santé du Québec.

Address for reprint requests and other correspondence: D. A. DeAngelis, Cellular Biochemistry and Biophysics Program, Memorial Sloan-Kettering Cancer Center, 1275 York Ave. Box 251, New York, New York 10021 (E-mail: d-deangelis@ski.mskcc.org).

**REFERENCES**


