Pergamon

Tetrahedron Letters 43 (2002) 1999–2003

TETRAHEDRON LETTERS

Unnatural amino acid derived FRET cassettes, terminators and their DNA sequencing potential

Satyam Nampalli,[a,*] Weihong Zhang,[a] T. Sudhakar Rao,[a] Haiguang Xiao,[a] Lakshmi P. Kotra[b] and Shiv Kumar[a]

[a]*Amersham Biosciences, 800 Centennial Avenue, Piscataway, NJ 08855, USA*
[b]*Faculty of Pharmacy, University of Toronto, 19 Russell Street, Toronto, ONT M5S 2S2, Canada*

Received 21 December 2001; revised 22 January 2002; accepted 28 January 2002

Abstract—An unnatural, tri-functional amino acid, *t*-Boc-L-*para*-amino-phenylalanine (**1**), was utilized in the design and syntheses of highly efficient FRET cassettes (**4**–**7**) and converted them into 2′,3′-dideoxynucleotide terminators (**8**–**11**) for investigating DNA sequencing potential coupled with a thermostable DNA polymerase, Thermo Sequenase™ II.

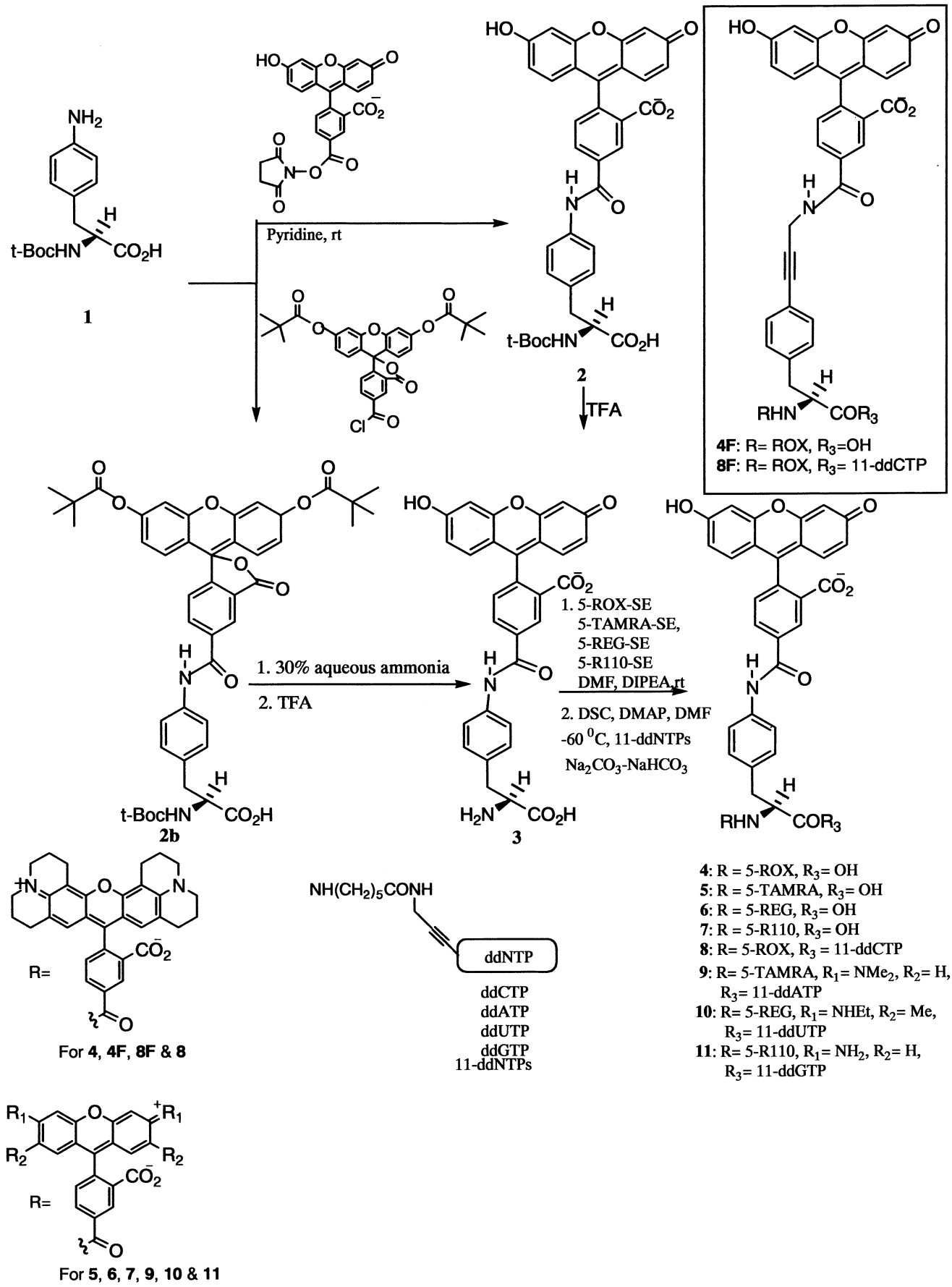
Recently, there has been a substantial interest in developing FRET-labeled primers[1] and dideoxynucleotide terminators[2] for high throughput DNA sequencing. The initial whole human genome mapping efforts hinged on the employment of FRET-labeled terminators on the capillary gel based, automated sequencers. Researchers in the genomics world are directing the sequencing efforts not only toward finishing the whole genome sequence and determining polymorphisms, but also mapping genomes of other organisms for comparison. For analysing the order of any piece of DNA using thermal cycle sequencing,[2,3] one needs a target template, primer, natural dNTPs, thermally stable DNA polymerase,[4] and four dye labeled-ddNTPs (terminators). FRET labeled-terminators are superior to single-dye labeled-terminators because they generate highly sensitized, enhanced fluorescence signals for the four kinds of DNA sequencing fragments. Each of the four dye-pair labeled-terminators comprises a common fluorescein donor dye, different rhodamine acceptor dye, and ddNTP covalently attached to a core molecular cassette. A variety of molecular cassettes[2] have been used for the construction of FRET terminators and the energy transfer efficiency of these depends upon the type of molecular cassette that serves as a bridge between the donor and acceptor dyes. FRET-based terminators coupled with Thermo Sequenase™ DNA polymerase significantly improve the quality and sensitivity of automated DNA sequencing. In our continued efforts to improve the quality and sensitivity, we have synthesized and evaluated yet another set of novel FRET terminators in a relatively fewer number of steps starting with a commercially available unnatural amino acid.

Utilizing *t*-Boc-*para*-amino-phenylalanine (**1**, Scheme 1), an unnatural amino acid, a common, single-dye labeled-cassette construct (**3**) was envisaged to be synthesized and converted to the four different dye-pair cassettes and their corresponding FRET terminators.

Employing **1** in a conjugation reaction with NHS ester of 5-carboxyfluorescein in pyridine produced **2**,[5] which upon TFA induced removal of *t*-Boc protection from the α-amino group provided the desired common single dye cassette **3** in low yield. Obviously, low nucleophilicity of the *para*-amino (sp^2 in nature and participates its lone-pair of electrons in resonance with the π-electrons) functional group in **1** towards NHS ester of the fluorescein resulted in poor yields of **2**. In an attempt to improve the yield of the single dye cassette, a freshly prepared acid chloride of the di-piv-carboxyfluorescein was reacted with **1** in CH_2Cl_2/pyridine/DMF to obtain **2b** in significantly improved 56% yield. Ammonolysis of compound **2b** with 28% aqueous ammonia followed by treatment with TFA afforded the same single dye labeled-cassette (**3**). In order to synthesize the FRET cassettes, the common single dye-labeled cassette (**3**) was individually reacted with isomerically pure NHS esters of ROX, TAMRA, REG, and R110 dyes in DMF in the presence of DIPEA to provide the FRET

* Corresponding author. Fax: (732) 457-8353; e-mail: satyam.nampalli@am.amershambiosciences.com

0040-4039/02/$ - see front matter
PII: S0040-4039(02)00196-X



Scheme 1. Synthesis of FRET cassettes (**4**–**7**) and terminators (**8**–**11**).

dye-pair cassettes (**4**–**7**)[6] in yields ranging from 70 to 85%. The overall fluorescence quantum yield[7] (QE%) of these FRET cassettes (**4**–**7**)—27, 30, 37, and 45%, respectively—were measured against the PA-phenylalanine derived ones—6, 13, 14, and 29%—to display significantly improved percentages for all the four constructs. To help explain the enhanced brightness of the FRET cassettes and terminators of this study, a molecular modelling analysis was undertaken for ascertaining relative spatial orientation, distance between the dye centroids, and intramolecular functional group interactions.

Thus, using Sybyl molecular modelling program,[8] analysis was particularly carried out on the FRET constructs (**4** and **4F**), in which the absorption spectra of the acceptor dyes overlapped with the fluorescence emission spectra of the donor dye to the least extent. The lowest energy conformations for **4** and **4F** showed electrostatic interaction between the quaternary nitrogen of the acceptor, ROX and the 3-carboxyl of the donor, fluorescien dye moieties as a common feature (shown as white arrows in Figs. 1 and 2). Additionally, **4** showed hydrogen-bonding interaction between the 3-carboxyl of the acceptor, ROX with *para*-amide N–H of the cassette-core (shown as yellow broken line in Fig. 1). While the distances between the dye centroids in case of **4** with two energy minimized conformations were measured to be 8.9 (in cyan) and 9.6 (in red) Å, **4F** measured an increased distance of 10.2 Å. In both cases the dipole-induced dipole orientation was found to be parallel.

Clearly, shorter distance (Foster distance R_0)[9] coupled with the presence of H-bonding in case of **4** compared to **4F** corroborates the enhanced energy transfer or several-fold brightness demonstrated by the FRET cassettes (**4**–**7**) and their corresponding terminators (**8**–**11**).

For achieving the syntheses of the 2′,3′-dideoxynucleotide terminators (**8**–**11**, Scheme 1), the FRET cassettes (**4**–**7**) were individually converted to their corresponding NHS esters employing the previously optimized DSC/DMAP/DMF/−60°C reaction conditions,[2a] and in

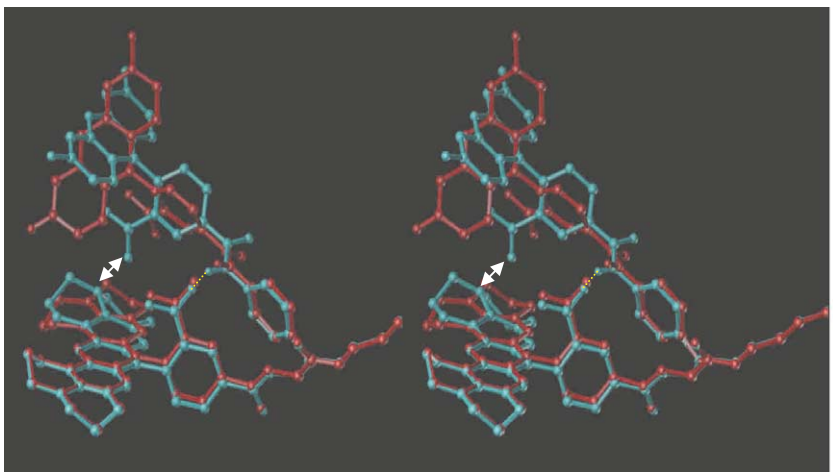


Figure 1. Overlapped stereo views of the two lowest energy conformations of the *para*-amino-phenylalanine derived ET cassette (**4** in cyan and red)–overlapping was carried out to see the difference between the two in stereo viewing and all four stereo views belong to **4**.

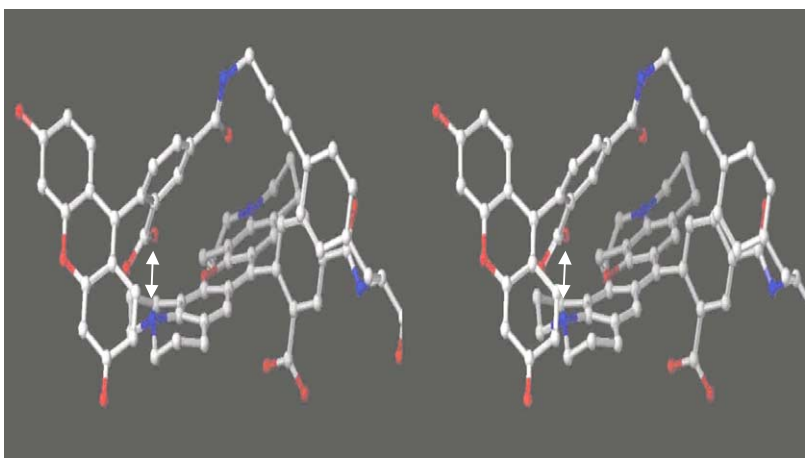


Figure 2. Stereo view of the lowest energy conformation of the *para*-propargylamino-phenylalanine derived ET cassette (**4F**)—both the stereo views shown belong to **4F**.

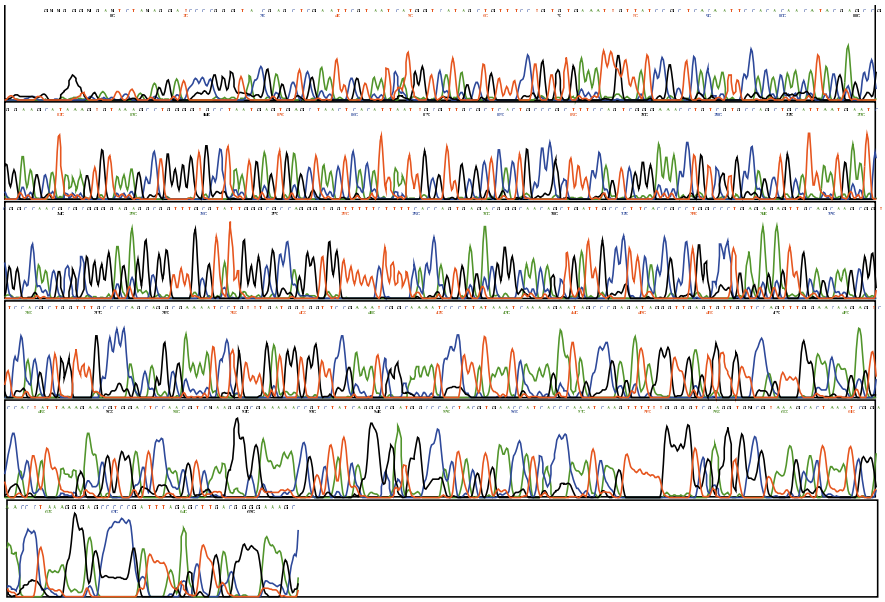


Figure 3. Electropherogram generated by FRET terminators (**8**–**11**) in combination with Thermo Sequenase™ II for sequencing a plasmid template on the MegaBACE 1000 sequencer. Read-length shown in excess of 650 bases.

situ conjugated with 11-ddNTPs (11-ddCTP, 11-ddATP, 11-ddUTP, and 11-ddGTP)[10] at −30°C to provide the desired FRET dye-pair labeled-terminators (**8**–**11**)[11] in yields ranging from 15 to 20%. The fluorescence emission enhancement rates of these ET terminators compared to the corresponding single rhodamine dye labeled-terminators were found to be 34 (ROX), 8.9 (TAMRA), 8 (REG), and 3 (R110). The QE% of these FRET terminators (**8**–**11**)—36, 22, 43, and 15%, respectively—were measured to be superior to PA-phenylalanine derived terminators—19, 16, 26, and 8%.

In order to investigate a four-color DNA sequencing potential of the synthesized FRET terminators (**8**–**11**), they were tested in sequencing reactions using Thermo Sequenase™ II DNA polymerase on the MegaBACE™ 1000 sequencer. Careful examination of the electropherogram (Fig. 3) revealed that the FRET terminators (**8**–**11**) derived from *para*-aminophenylalanine indeed generated high quality data with good peak uniformity, accuracy and read-lengths in excess of 650 bases (eyeball estimation). Based on the comparative gel images (Fig. 4) created by the **8** and **8F** labeled-amplicons from a slab gel DNA sequencer, the fluorescent signatures of the former were displayed to be two-fold brighter than the latter.

In conclusion, an efficient synthetic route has been developed for a four-color, FRET dye-pair labeled terminators from a commercially available starting material, and in combination with a suitable DNA polymerase, demonstrated their DNA sequencing potential by generating highly sensitized fluorescence signals representing the sequential order of the bases involved. Investigations into the photophysical properties, molecular modelling, and DNA sequencing ability of these new FRET terminators, not only offers another DNA sequencing paradigm, but also shows useful pointers for further advancements.

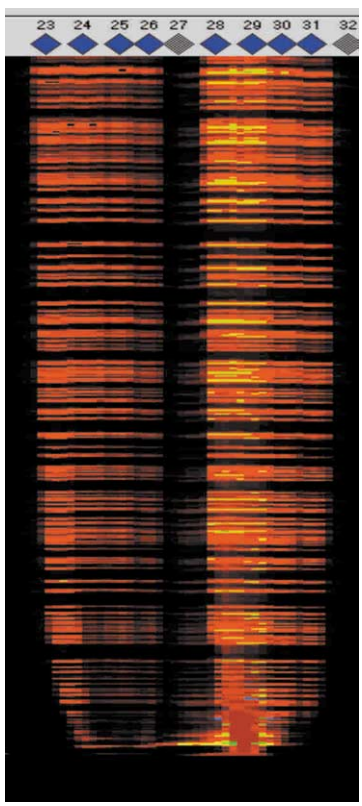


Figure 4. Gel image obtained from a slab gel DNA Sequencer. Ladder 1 generated by **8F** and ladder 2 generated by **8** showing two-fold brighter signals than **8F**.

Acknowledgements

The authors wish to thank Dr. C. Y. Chen for fluorescence spectral analysis, Matthew Bull for mass spectral analysis and Dr. Carl Fuller for many helpful discussions.

References

1. (a) Ju, J.; Kheterpal, I.; Scherer, J. R.; Ruan, C.; Fuller, C. W.; Glazer, A. N.; Mathies, R. A. *Anal. Biochem.* **1995**, *231*, 131–140; (b) Metzker, M. L.; Lu, J.; Gibbs, R. A. *Science* **1996**, *271*, 1420–1422; (c) Ju, J.; Glazer, A. N.; Mathies, R. A. *Nucleic Acids Res.* **1996**, *24*, 1144–1148; (d) Lee, L. G.; Spurgeon, S. L.; Heiner, C. R.; Benson, S. C.; Rosenblum, B. B.; Menchen, S. M.; Graham, R. J.; Constantinescu, A.; Upadhya, K. G.; Cassel, J. M. *Nucleic Acids Res.* **1997**, *25*, 2816–2822.
2. (a) Nampalli, S.; Khot, M.; Kumar, S. *Tetrahedron Lett.* **2000**, *41*, 8867–8871; (b) Rao, T. S.; Nampalli, S.; Lavrenov, K.; Zhang, W.; Xiao, H.; Nelson, J.; Kumar, S. *Nucleosides, Nucleotides Nucleic Acids* **2001**, *20*, 673–676; (c) Rosenblum, B. B.; Lee, L. G.; Spurgeon, S. L.; Khan, S. H.; Menchen, S. M.; Heiner, C. R.; Chen, S. M. *Nucleic Acids Res.* **1997**, *25*, 4500–4504.
3. Sanger, F.; Nicklen, S.; Coulson, A. R. *Proc. Natl. Acad. Sci. USA* **1977**, *74*, 5463–5467.
4. (a) Tabor, S.; Richardson, C. C. *Proc. Natl. Acad. Sci. USA* **1995**, *61*, 6339–6343; (b) Tabor, S.; Richardson, C. C. *J. Biol. Chem.* **1990**, *265*, 8322–8328.
5. ^{1}H NMR of the single dye cassette **2** (DMSO-d_6): 1.39 (9H, s, *t*-Bu), δ 2.70 (2H, m, benzylic), 4.20 (1H, s, chiral), 6.50 (4H, m, aromatic), 7.04 (2H, d, $J=10$ Hz, aromatic), 7.25 (2H, d, $J=9.0$ Hz, aromatic), 7.39 (2H, d, $J=9.0$ Hz, aromatic), 7.39 (1H, d, $J=6$ Hz, aromatic), 7.64 (2H, d, $J=6.0$ Hz, aromatic), 8.15 (1H, d, $J=6.0$ Hz, aromatic), 8.59 (1H, s, aromatic). TOF MS m/z ES for $C_{35}H_{45}N_2O_{10}$, 637.86 (exact mass 638).
6. ^{1}H NMR of **4** (DMSO-d_6): δ 1.82 (8H, br s, homo-benzylic), 1.98 (8H, br s, benzylic), 2.63 (8H, br s, ROX-methylenes adjacent N), 3.05 (2H, m, benzylic), 4.44 (1H, m, chiral), 6.50 (4H, m, aromatic), 6.62 (2H, s, aromatic), 6.95 (1H, d, $J=9.0$ Hz, aromatic), 7.28 (5H, m, aromatic), 7.96 (1H, d, $J=6.0$ Hz, aromatic), 8.00 (1H, d, $J=6.0$, aromatic), 8.34 (1H, s, aromatic), 8.46 (1H, s, aromatic). Compounds **5**: TOF MS m/z ES+ for $C_{55}H_{43}N_4O_{12}$, 951.67 (exact mass 951.29); compound **6**: TOF MS m/z ES+ for $C_{57}H_{45}N_4O_{12}$, 978.9 (exact mass 979.00) and **7**: $C_{51}H_{34}N_4O_{12}$, 894.36 (exact mass 894.22). Compounds **5**, **6**, and **7** showed satisfactory ^{1}H NMR data.
7. Using fluorescein as the quantum efficiency (QE%) standard, the quantum efficiencies of these FRET cassettes (**4**–**7**) in 1X TBE+8 M urea were measured from their acceptor dye emissions when excited at the common donor dye wavelength (488 nm) of an Argon laser. A corrected, photon counting fluorescence spectrometer was used with highly dilute samples to avoid complexation and inner-filter effects. The QE% values reported are the number of photons emitted per 100 photons absorbed.
8. Sybyl version 6.6 Tripos force field was applied for 5000 iterations. The atomic charges on the molecules were calculated by Pullman method as implemented in Sybyl. Of the total 2052 conformations and 2602 distances between the dyes, the lowest energy conformations for **4** and **4F** were taken into account.
9. The transfer of energy is inversely proportional to the sixth power of the distance between the donor and acceptor dyes. Lakowicz, J. R. *Principles of Fluorescence Spectroscopy*, 2nd Ed.; Kluwer Academic/Plenum publishers: New York, 1999; pp. 367–390.
10. Hobbs, F. W., Jr.; Cocuzza, A. J. US Patent 5 047 519, 1991.
11. UV–vis in 1.0 M TBE+8.0 M urea for **8**: λ_{max} 499, 586 nm, ε_{max} 609 nm; **9**: λ_{max} 497, 556 nm, ε_{max} 582 nm; **10**: λ_{max} 500, 530 nm ε_{max} 558 nm; **11**: λ_{max} 500 nm ε_{max} 526 nm.

Glossary of terms: FRET (fluorescence resonance energy transfer), DIPEA (*N*,*N*-di-isopropylethylamine), DMAP (4-*N*,*N*-dimethylaminopyridine), DSC (*N*,*N*-disuccinimidylcarbonate), DMF (*N*,*N*-dimethylformamide), PCR (polymerase chain reaction), ddNTPs (1′,2′-dideoxynucleotide triphosphate), 11-ddNTPs (number 11 denotes the number of atoms in the linear spacer arm at 5-position of pyrimidine and 7-position of purine bases), PA-phenylalanine (*para*-propargylamino phenylalanine), QE% (quantum efficiency or energy transfer efficiency), DYEnamic and MegaBACE are the trademarks of Amersham Biosciences, Inc.