Green Fluorescent Protein

Synopsis:

- Introduction
- Obelin
- GFP sequence
- Vector and Extremophile selection
- Codon Bias
- Primer Designing
- Cloning and Expression
- Commercial Applications

Introduction:

Certain marine organisms produce calcium-activated photoproteins that allow them to emit light for a variety of purposes, such as defense, feeding, breeding, etc. Even though there are many bioluminescent organisms in nature, only a few photoproteins have been isolated and characterized. The mechanism of emission of light in the blue region is the result of an internal chemical reaction. Since there is no need for excitation from external irradiation for the

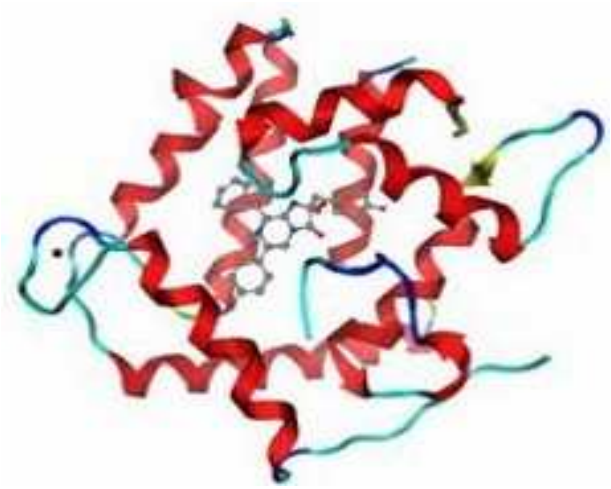
emission of bioluminescence, the signal produced has virtually no background. This allows for detection of the proteins at extremely low levels, making these photoproteins attractive labels for analytical applications. In that regard, the use of certain photoproteins, namely, aequorin, obelin, and the green fluorescent protein (GFP) as labels has increased rapidly in recent years.

Photoproteins are scientifically interesting for two reasons.

1) They are a fascinating vehicle in which to observe, and manipulate, the energy harnessed in matter changing its form from potential kinetic energy to radiative energy. Unlike fluorescence, the photons of a bioluminescent reaction are produced from molecules relaxing from a chemically, and not photonically, excited electronic state. The chemical reaction alone produces the emission of light, without the aid of an excitatory light source.

2) Photoproteins serve as a highly sensitive label in a variety of scientific applications. The light emitted from photoproteins allows the researcher to ‘see’ where the protein is, and how much is present, following calcium addition. For that reason, photoproteins are excellent labels and reporter proteins.

Structure Of Obelin:



The structure of the obelin molecule is highly compact and globular with a radius of 25 Å. The molecularstructure is formed by two sets of four helices designated A ~16–29, B ~39–54,C ~58–74, and D ~85–105, in the N-terminal domain, and E ~110–122, F ~132–142, G ~148–157, and H ~168–180, in the C-terminal domain. Both the N- and C-terminal domains can be thought of as a cup whose insides are lined with hydrophobic residues.The coelenterazine-oxygen substrate resides in an internal cavity, which is surrounded by hydrophobic residues from the eight helices.

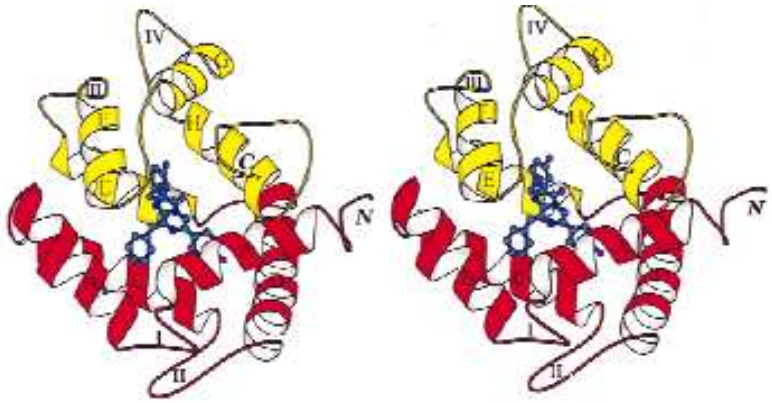
The obelin molecule consists of four sets of helix-turn-helix (HTH) structural motifs, which are characteristic of EF-hand calcium-binding domains. The A- and B-helices are joined by a 9-residue loop,which together with three residues of the B-helix forms the first expected calcium-binding site. The amino acid composition of this site is similar to the calcium-binding sites of other calcium binding proteins .The C- andD-helices are linked by a 10-residue loop .The amino acid residues forming this loop are known to be unable to coordinate calcium and, consequently, this

HTH motif II is not functional for calciumbinding. The E- and F-helices are connected througha 9-residue loop, which together with three residues of the F-helixforms HTH motif III ~the second expected calcium-binding site.The G- and H-helices are linked via a 10-residue loop, nine of which together with three residues of the H helix forms HTH motif IV ~the third calcium-binding site. The amino acid composition and structural organization of calcium-binding sites III and IV are typical for canonic calcium-binding sites and consequently, both HTH motifs of the C-terminal can be expected to bind calcium ions.

The electrostatic potentials of the surfaces of the three potential calcium-binding sites ~HTH motifs I III, and IV are negative. These provide likely regionsfor binding of calcium ions. On the other hand, the electrostatic potential at the corresponding position of the loop of HTH motif IIis neutral. This explains that while I, III, and IV HTH motifs canbind calcium ions, HTH motif II cannot. In addition, the surface electrostatic potential of the loops at the HTH motif III and IV is more negative than that of HTH motif I. This indicates that the calcium-binding constants for III and IV calcium-binding sites are probably higher than for site I.

The N-terminal cup is formed by the interactions between HTH motifs I and II while the C-terminal cup is formed by the interactions between HTH motifs III and IV. Although no hydrogen bonding was identified between the eight helices in the obelin structure, several close interhelical contacts were observed including a salt bridge between Arg112 of helix E and Asp169 of helix H .

Structure of the coelenterazine-oxygen binding pocket:



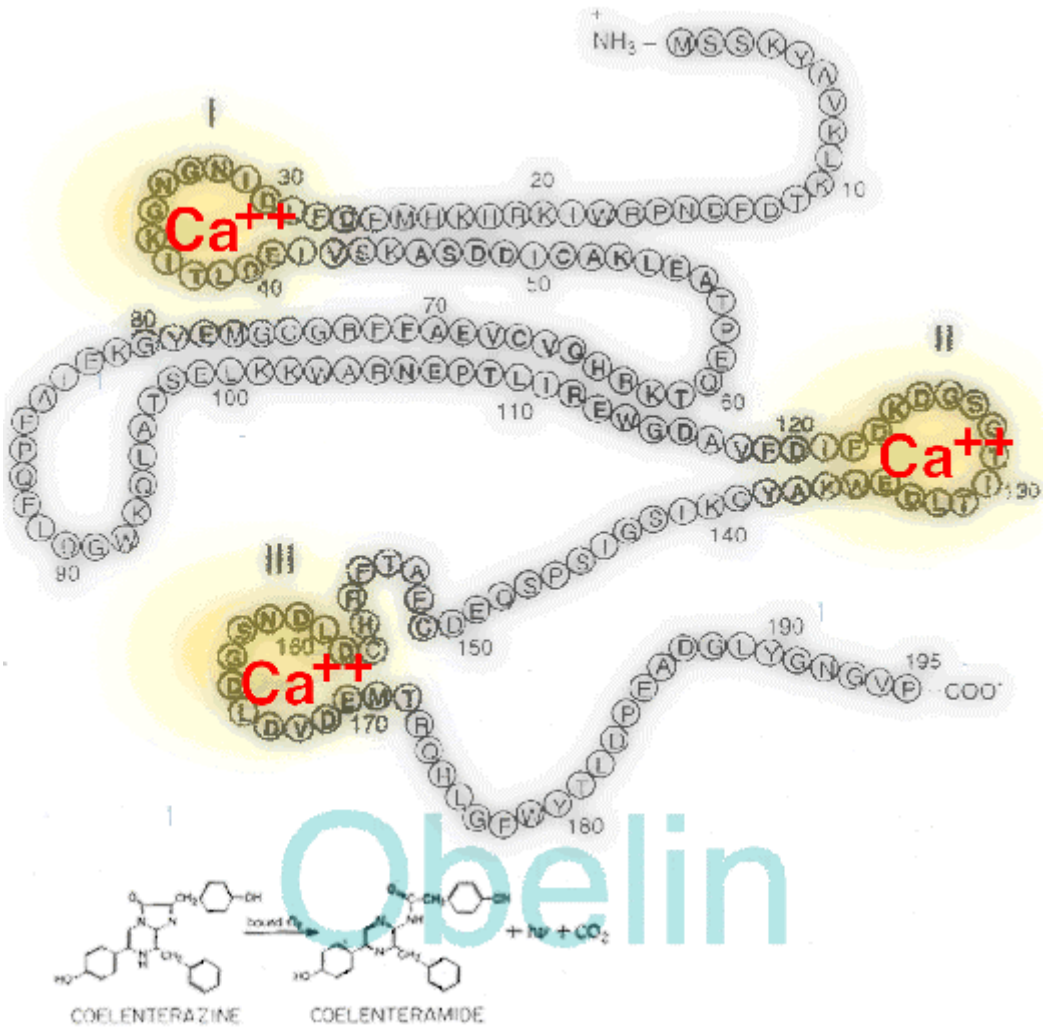
A stereoview of the overall crystal structure of obelin. The N- and C-terminal calcium-binding domains of the molecule are colored red and yellow, respectively. The coelenterazine-oxygen molecule is colored blue.

The coelenterazine-oxygen substrate and surrounding residues ~within 4 Å. The binding pocket is highly hydrophobic and is formed by residues originating from helix A ~Met25 and Leu29, helix B ~Ile42, Ala46, and Ile50, helix C ~Phe72, helix D ~Phe88 and Trp92, helix E ~Ile111, rp114, Val118, and Phe119, helix F ~Trp135, Ile144 from the loop linking helices F and G, and helix H ~Met171 and Trp179. In addition, several hydrophilic side chains are directed into the pocket. These are His22, Tyr138, and His175 localized in helices A, F, and H,respectively, and Tyr190 located near the C-terminus of the protein.

The most protein-coelenterazine contacts are observed between residues of helix E in HTH motif III including a short ~3.45 Å. contact to the main chain of Gly115. Almost all the residues forming the coelenterazine binding

pocket are conserved among all calcium-regulated photoproteins and consequently, the structure of the coelenterazine-oxygen binding sites should be conserved for all calcium-regulated photoproteins. Site-directed mutagenesis and direct chemical modifications of photoproteins have shown certain amino acid residues to be important for photoprotein bioluminescence. These are Trp, His, Cys, and the C-terminal Pro. Mutants of aequorin with replacement of Trp residues for Phe exhibited various luminescent activity and spectra including a W86F mutant that gave a bimodal emission spectrum with maxima at 455 and 400 nm ~Ohmiya et al., 1992!. It was suggested that Trp86 in aequorin may be involved in the generation of the product excited state during photoprotein luminescence. The obelin structure supports this suggestion since Trp92, Trp114, Trp135, and Trp179 are among the residues that make close contact with the coelenterazine oxygen substrate. For example, the side chains of Trp92 and Trp179 “sandwich” the p-hydroxyphenyl ring of coelenterazine. The planes of the phenyl ring and the Trp92 indole are almost parallel. The side chains of Trp114 and Trp135 are localized near the hydroxybenzyl group of coelenterazine.

Calcium-regulated photoprotein:



Green Fluorescent Protein (GFP) Sequence:

The obelin sequence is about 781 nucleotides.The sequence is shown below.

1 gatctcaaac gatcaaaaat cagcttctgc agctttcacc aaaaaaagaa gaaaaatggc

61 ttccaaatac gcagtcaaac ttcaaactga ctttgacaat ccaaaatgga tcaaaagaca

121 taaatttatg tttgattatc ttgacatcaa cggaaatggt caaatcacac ttgacgaaat

181 cgtatccaaa gcatctgatg acatttgtaa aaatcttgga gccacaccag cacaaactca

241 acgtcatcaa gattgcgttg aagctttctt cagaggttgc ggtttggaat atggcaaaga

301 aaccaaattc ccagaatttc ttgaaggatg gaagaacttg gcaaatgcag atctggcaaa

361 atgggcaaga aacgaaccga cacttattcg tgagtgggga gacgcagtat ttgacatatt

421 cgacaaggat ggcagtggta caatcacttt ggacgaatgg aaagcttatg gaagaatctc

481 tggtatctct ccatcagaag aagattgtga aaagaccttt caacattgtg atttggataa

541 cagtggtgag cttgatgttg atgagatgac aagacaacat ttgggattct ggtacacctt

601 ggatccagaa gctgatggtc tttacggaaa tggagtcccc taaatattat tttattttag

661 cagatctttg tacctctctc actaaaatat gtcttgacta acttttact aacttttaat

721 tttcttttgt cagctctttt aattaaagaa atacaaccaa agaaactaca ttccaaaaaa

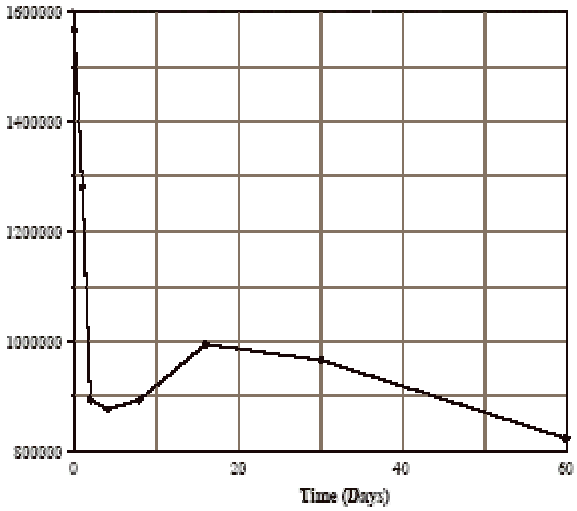
781 aaaaaaaaaa aaaaaaaaaa aa

Obelin :Protein sequence:

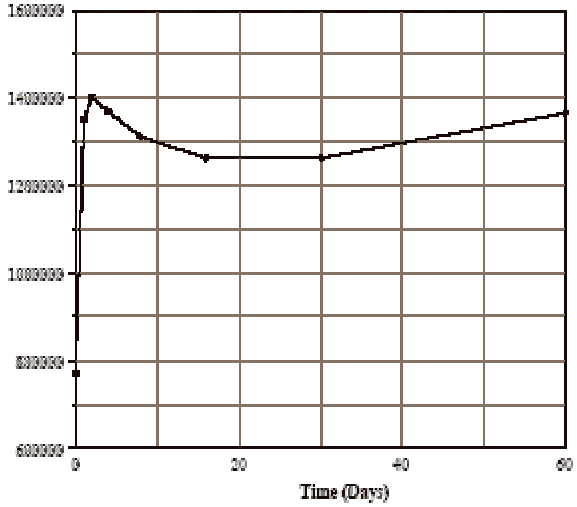
"MASKYAVKLQTDFDNPKWIKRHKFMFDYLDINGNGQITLDEIVS
KASDDICKNLGATPAQTQRHQDCVEAFFRGCGLEYGKETKFPEFLEGWKNLANADLAK
WARNEPTLIREWGDAVFDIFDKDGSGTITLDEWKAYGRISGISPSEEDCEKTFQHCDL
DNSGELDVDEMTRQHLGFWYTLDPEADGLYGNGVP"

Stability of Obelin with Aequorin:

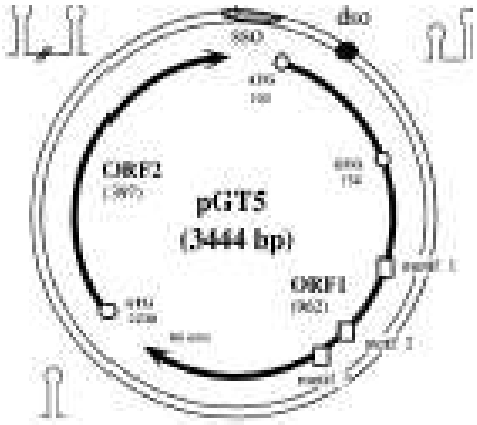
Stability of aequorin at 4 C



Stability of Obelin at 4 C



Vector Diagram:



Vector Sequence:

1 attatccttt agaaacgcgt tgggtattgg gggagccccc cttagtgggg agctcccct
61 aaacaccccc aagacaatag gaaaccatag gaggcccctg caatggtgat atacacttca
121 aagtttaaaa actctttgct tgatgggttg ggggttgggc accttagcta tgatgaccag
181 cctattctat gtaatgaggt tcatccaacc cttaccctcg acactttcat ctcgggtggg
241 tcatctggtt ctcgaccccg cccccgttgg gtttatcttg atatatccac aaccaatgaa
301 ggaatttcag aagagttctc ttcaaatact gagagtaaaa gccctctatt gaaggtgtgt
361 ggtgtcagtt ctaagtcatc tgagggtgat ggcagtagtt ttatctggtt tgataagtat
421 cgttcagtca tttcaaggct tgagcactct ggttttgtag aagttgaaag aactgtcctt
481 aagctcagga aaatctccaa agaactccgc tctataaaca agcagttatc gaggcttttc
541 cttgatgata gggagagagc aaagcttctc tctcgcaagc gtaagtatct tgattttgct
601 cgtgctctca taggttctat ttccaagtct cttactcttt atgcggatag gttctttgtt
661 gaggttcccc aagattacgc aaagttaatc caaaagctcg ggttctcaag tgatacgctt

721 cttcttcacc tattcgtgaa tagtggtgtt cttgaggtgt ttcttgatga taatagttct
781 cataagttcc gtattgctta tatttcaaag gttcacgctg gtaagtatca tcctgtgaag
841 ggtatctcca agggttctca ggaggctaag agggttctta gggatttgct tgtgcttagt
901 gaacttcttg agggttctct tgttagctac aggagtgggg gggttgaaac tattcatcac
961 cttattcctg tccgtcattt tgtccttacc gctcctaagg acgttagttt cagtatttgg
1021 gcttctctca agaaggggga tagttcattg tttagggctt tcaaggatgc tggtgctaaa
1081 gccatcaagg agtttttatc ataccttgct tccaaggaac atatctctgg taatctccta
1141 tttggtttta ctatcaatgt tcatgttact ggtgataaga acccctttga gcctcacttc
1201 catattgatg ctatcgttac tttcatttgc tatgataagt cttcaaccaa atggtttagg
1261 cttaatcctc ttctcagcga gagtgatcta aagaaattgc gtgatatttg gaagaacgtg
1321 ttgctgtcct atttcggtga gcttctctct gaggatacta agtcaaagga ttttgatgtt
1381 tgggctggtg ataattatta ctctttgcct cttgatgttc ctcaggtctt cttcgagctc
1441 aagtatgcgt cgaggaagct tttcgtcaat tttgttaatt actttgagca gtctaatttt
1501 gatgagagct ccgtgtccga ttgggatttt gtccgtttcg tcttcgagta tagtaatagg
1561 actgaacgtt atggtttcct cactaacatt aaacgttatc tttctatgtc ttgctctcac
1621 ctcgtagaga agcgtgttca ggagcttgaa gagttcatta gccgtattga gtttgatttg
1681 tctgtgaatg gtaataagat gagtgattca cttaagaggg ctctccttga gcgtcttgag
1741 tatttgaagg atgagctttc agaacttaag gagcgtggtt tcgagtatct ctttgagcgt
1801 gcccttgaga aggcggagga gttgttatct aacgataatc tcactcttga gcgtgttatt
1861 catattttgg aaactttgtt caccgctctt ggtaagtcta ttgtcaatta caatttctat
1921 gttgagcttg aggatgtctc cttcagggag tttgtcgatt atctttatga taatcatctc
1981 tctgatgtcc ttgtctttag tgataggcac cgttctatta ccattattcg gcttattcct
2041 cctcctgatg ggggtgtgcc tgtatgacac cttccaactt tattgtaagt ttgggtgtca
2101 ttttggcacc gcaccccaca ctctctcctc tccccctcg cctcgccccc cgctggtgcc
2161 tccttggtgc cactacccta cgtttccatc tttttataac tctcgtgcac ttagtaggtg
2221 ggatggtcca tgaggttgca cgtccatatt aatgacaagt tgggtgaact gttcaagaag
2281 cgtgtccacc ttgtgtatgg ggataggcgt ggtgctttgt ctcacgccat tgaggaagct
2341 attatcctgt ggttgctcca tactgactat attgtaaagc ttcagaaggt gaacgctatg
2401 agccttcaaa ctcaaaggat tgacagcgac tcctctcatt atttgcctat ctctcggttg
2461 cctcagacta ttgatgagtt tagggctttc tataatagat ttggtgctaa ggctactttg
2521 cttgtttacc ttcttgttag gggtaagcct ctccgtttgt atgacctttt taggtttcat
2581 caggctattg ctggtgtgtc ttctgataat tctactcgta aggttctcaa gaggttgttg
2641 cgggatggtt taatctatag aaaaggtaag ctctattatc cttcccctga tgctcttttg
2701 cttatgtggg attctattga ttggtcgagg gtgaggacta agcgttatac tggtaatggt
2761 gtggtgtcta agccttctaa acgtagggtt cgtggtgttg gttctcaggt taggcgggtt
2821 gttcgtattg cgaggagttt ggctcggcgt ggggatactt tgagggctgt tagtttgctt
2881 acgcatactc ttttgggtgt taggcagact ggtgtgctcc tttatagaaa aggtggcttg

2941 tttattgtgt atgagcataa ggacaggtct gttagggttc ttttctctga gcggattggt
3001 aagttgtttg atgatttggg gcttcctgag gggtctcttt atatgcatat gtattatcgt
3061 gctggtagga ttattcacaa tgtgtttggg gggtatcggg ttgctcggcg catgttctat
3121 tttctccgtg agcatggttt cttctatgtt cccgcttctg aggttgatag ggttggtgaa
3181 tatcgctatg accctcatag cgggaagctc attgttcgta ttggtaattg gacgtatgag
3241 gttgatgctg atcttgatgc cattgatagt tacgttgaaa atcttaattc gaggaggact
3301 cgtaagtcta agaagaagtc ttcgtctctc gttagttatc ctcgtgagag gtttggtatt
3361 gtgcgtgttt caaagcatac gaaggtggag aacgatgttt actacctcaa ttatactccg
3421 taggagttat aaatccttga gaat

It is a naturally occurring plasmid in Pyrococcus abyssi GE5 strain. Plasmid pGT5 (3,444 bp) from the hyperthermophilic archaeon Pyrococcus abyssi GE5 has been completely sequenced. Two major open reading frames with a good coding probability are located on the same strand and cover 85% of the total sequence. The larger open reading frame encodes a putative polypeptide, which exhibits sequence similarity with Rep proteins of plasmids using the rolling-circle mechanism for replication.

A putative single-stranded origin exhibits similarities both to bacterial primosome-dependent single-stranded initiation sites and to bacterial primase (dnaG) start sites.

A single-stranded form of pGT5 corresponding to the plus strand was detected in cells of P. abyssi. These data indicate that pGT5 replicates via the rolling-circle mechanism and suggest

that members of the domain Archaea contain homologs of several bacterial proteins involved in chromosomal DNA replication.

Cloning in Plasmid pGT5:

The vector is opened with the restriction enzyme NspBII at position 2225.The codon bias was done with codon usage table for Pyrococcus abyssi GE5 strain.There is a numerous variations in the codon usage. So codon bias is done. If there is a change in one or two amino acids, we can approach site directed mutagenesis. But here it has got a various changes.So codon bias is done. The GFP nucleotide sequence is ligated in this region flanking the this NspBII enzyme. So that ligation would be proper. For this particular sequence forward primer and reverse primer is designed.

Primer designing:

Forward primer sequence:

5’GGTCCAGATCTCAACGC 3’

Reverse primer sequence:

5’GGTCCACCTTACAGCAGC3’

	forward primer	reverse primer
Length	**18**	**18**
Tm: (in C)	**52**	**54.94**
GC% content:	**55.56**	**61**
Dimer:	**1**	**Nil**
Cross dimer:	**0**	**Nil**
Hairpin:	**Nil**	**Nil**
Repeat and Runs	**1**	**Nil**

Applications of Obelin:

1) Ca^{++}-regulated photoprotein assays are highly sensitive and non hazardous with rapid and simple light-producing reactions.

2) Photoproteins are therefore attractive for applications already utilising reporter enzymes like alkaline phosphatase or Horseradish peroxidase.

3) The feasibility of using photoproteins such as obelin as non-radioactive tags is already established in a variety of *in vitro* applications for the sub-nanogram detection of various substances.

4) PhotoLight™ Obelin-labeled molecules are for the use in High Throughput Screening (HTS) detection systems.

5) The PhotoLight™ systems are ideally suited for applications requiring high sensitivity, such as highly sensitive DNA probe assays and

immunoassays for thyroid peptide hormones , free hormone fractions, steroids and drugs.

REFERENCES:

1) Sequence of Plasmid pGT5 from the Archaeon *Pyrococcus abyssi*: Evidence for Rolling-Circle Replication in a Hyperthermophile. JOURNAL OF BACTERIOLOGY, June 1996, p. 3232–3237.

2) Obelin from the Bioluminescent Marine Hydroid *Obelia geniculata*: Cloning, Expression, and Comparison of Some Properties with Those of OtherCa2+-Regulated Photoproteins *Biochemistry* **2002,** *41*, 2227-2236.

3) Structural Chemistry of a Green Fluorescent Protein Zn Biosensor

4) Bioorg Khim. 2001 Sep-Oct;27(5):364-71.

5) FEMS Microbiol Rev. 1996 May;18(2-3):93-104

6) Crystal Structure of a Ca^{2+}-discharged Photoprotein