Expression of Fusion Proteins: How to Get Started with the HaloTag® Technology

ABSTRACT Appending fusion proteins such as affinity tags and fluorescent proteins is important for the study of protein structure and function. The HaloTag® Technology gives the functionalities of different fusion partners with a single fusion construct. Here we discuss basic considerations and steps to creating a HaloTag® fusion construct.

Melissa McCornack, Trista Schagat, and Michael Slater, Promega Corporation

There are 3 steps to getting started with the HaloTag® Technology: planning (i.e., choosing a vector and designing strategy), cloning and confirming.

INTRODUCTION

Fusion proteins are used to understand protein function. They facilitate protein expression and purification, study of protein:protein and protein:nucleic acid interactions, and protein labeling in vivo for localization. Standard methods require use of different tags for different applications, necessitating the creation of multiple fusion constructs. The HaloTag® Technology[a–c] is a versatile fusion protein platform that allows all of these applications and more with a single fusion partner (Figure 1; 1). As with any fusion protein, there are three basic steps to getting started with the HaloTag® Technology: planning (i.e., choosing a vector and designing a strategy), cloning and confirming.

PLANNING

Prior to generating a fusion protein, there are a couple of considerations that will influence vector choice: expression system and structural design.

Expression Systems

Fusion proteins may be expressed in cell-based and cell-free systems. Downstream applications will dictate the choice, and the chosen systems will determine the regulatory elements that must be present in the expression vector. To study protein localization and trafficking, mammalian cell-based expression is typically used, which necessitates a Kozak sequence and mammalian promoter such as the CMV promoter. For large-scale protein expression and purification, bacterial expression is used, dictating either a bacterial or T7 promoter, depending on the bacterial strain, and an RBS (prokaryotic ribosome binding sequence). Cell-free systems (eukaryotic [e.g., TNT® systems] and prokaryotic [e.g., S30 systems]) are fast, amenable to high-throughput, and often used for pull-down experiments and enzymatic assays (2,3); these systems require a bacteriophage promoter such as T7 or SP6. Promoter and vector choice also may be influenced by the levels of protein expression required, particularly in mammalian cells where high-level expression can adversely affect normal cellular biology (4).

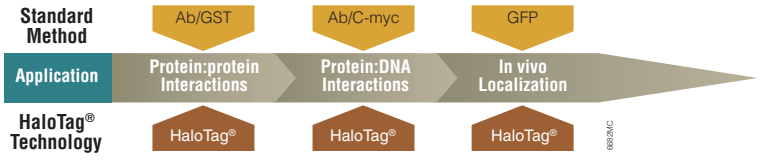

Figure 1. HaloTag® Technology offers a broad range of applications with a single fusion construct.

Structural Considerations

The position of the tag on the fusion protein can affect protein functionality and expression level. The optimal configuration must be determined empirically. For in vivo expression, there are two advantages of appending the tag on the C-terminus of the protein of interest: 1) N-terminal signal sequences of the protein of interest remain available for processing, and 2) subsequent detection of the tag provides assurance that the full-length translation product was produced. Appending the tag on the N-terminus of the protein of interest may improve expression and solubility, in some cases.

Another structural consideration is the peptide linker sequences. The flexibility, hydrophilicity and linker length are important for the production of functional fusion proteins (5). Typically at least 15 small uncharged amino acids are used (e.g., $[Ser\text{-}Gly_4]_3$). The linker also may be engineered to contain a protease cleavage site, which facilitates tag removal for protein purification.

CLONING

Creating an expression construct requires cloning the coding region of the protein of interest, the linker and the fusion tag into a vector that contains all the regulatory elements needed for expression. The cloning

Table 1. HaloTag® Flexi® Vectors. The expression system, expression strength, tag position, and antibiotic resistance marker determine which Flexi® Vector should be used. The orange and blue color scales indicate relative expression strength. Dark represents the strongest and each shade lighter represents approximately twofold less expression within each system for the orange color scale and tenfold less for the blue scale. In green are the best general purpose vectors. In the Flexi® naming convention, "pF" indicates the vector is a Flexi® Vector. The letter after "pF" indicates the position of the HaloTag® tags relative to the cloning site (e.g., "N" for an N-terminal HaloTag® protein; "C" for a C-terminal HaloTag® protein). The number associated with the vector is a unique identifier. A terminal "A" or "K" indicates the antibiotic resistance to ampicillin (Amp) or Kanamycin (Kan; not indicated here). Prok, prokaryotic; Euk, eukaryotic. NR, not recommended due to absence of critical regulatory sequence and/or low expression levels.

Expression Systems				Vector Names/ Catalog Numbers						
Cell-Free		Cell-Based		N-Terminal HaloTag® Protein			C-Terminal HaloTag® Protein			
Prokaryotic	Eukaryotic	*E. coli*	Mammalian	Vector	Amp	Kan	Vector	Amp	Kan	Comments
T7	NR	T7	NR	pFN18	G2751	G2681				Best for high-level *E. coli* expression in vivo and in vitro; lacks regulatory elements for eukaryotic expression.
T7	T7, SP6	T7	NR	pFN19	G1891	G1841	pFC20	G1681	G1691	Good for *E. coli* and in vitro expression, lacks regulatory elements for mammalian cell-based expression.
NR	T7	NR	CMV	pFN21	G2821	G2831	pFC14	G9651	G9661	Strictly for eukaryotic expression in vivo and in vitro; maximum yield in mammalian cells. Lacks regulatory elements for prokaryotic expression.
T7	T7, SP6	T7	CMV*d1*	pFN22	G2841	G2851	pFC15	G1611	G1601	Good expression in *E. coli* and mammalian in vivo and in vitro; promoter deletion series to optimize mammalian expression levels (4).
T7	T7, SP6	T7	CMV*d2*	pFN23	G2861	G2871	pFC16	G1591	G1571	
T7	T7, SP6	T7	CMV*d3*	pFN24	G2881	G2981	pFC17	G1551	G1321	

The best all-purpose vectors are pFN19 (for strong bacterial and cell-free expression) and pFN22 (for mammalian, bacterial and cell-free expression).

strategy used will depend on the vector system chosen. Commercially available cloning systems such as the Flexi® Vector System are designed to facilitate easy cloning and transfer between vectors to suit experimental needs.

Flexi® Vectors

The Flexi® Vector System[a,d,e] is a simple, yet powerful, directional cloning method for protein-coding sequences based on rare-cutting restriction enzymes (6). Flexi® advantages include rapid, efficient cloning, minimal appendage of extraneous amino acids, and high-fidelity transfer of protein-coding regions between vectors. Once a Flexi® construct is created, the insert may be transferred easily between Flexi® Vectors. We recommend starting with an N-terminal HaloTag® fusion, because the Flexi® cloning strategy makes these inserts more easily transferred to other Flexi® Vectors. C-terminal fusion constructs also can be transferred, but the transfer is not as convenient, and subsequent transfer to an N-terminal Flexi® Vector is not possible (7).

The HaloTag® coding region is available in a variety of Flexi® Vectors (Table 1). The expression system, expression level needed, and tag position will determine the initial Flexi® Vector choice. The best all-purpose vectors are pFN19 (for strong bacterial and cell-free expression) and pFN22 (for mammalian, bacterial and cell-free expression). If protein quantity is most important, plan for prokaryotic expression with pFN18. The pFN21 vector is ideal for high eukaryotic expression. The pFN22–24 vectors are all-purpose vectors with deletions of the CMV promoter to optimize cell-based mammalian expression (4).

Basic Steps

The basic steps for cloning the coding region of the protein of interest into a fusion protein expression vector are outlined in Table 2. Tools available for each step of the process are indicated. Use of the HaloTag® Flexi® Vector System simplifies the procedure, and a free web-based tool for primer design helps ensure that the open reading frame (ORF) of the protein of interest is cloned in-frame with the HaloTag® ORF. All HaloTag® Flexi® Vectors contain an optimized peptide linker sequence with a TEV protease cleavage site so that design and cloning of linker sequences is not needed. When using expression vectors other than Flexi® Vectors, the protein of interest ORF should first be cloned into the Flexi® Vector to generate the full fusion construct. Unique restriction enzyme sites then can be used to transfer the full fusion construct into an alternate expression vector (7).

These basic steps apply to the initial creation of a Flexi® construct. Once the ORF is cloned into a Flexi® Vector, it can be transferred to other Flexi® Vectors (i.e., to change tag orientation or for different expression or tag needs) by beginning the process at step 4 using rapid digestion and ligation protocols (8).

CONFIRMING

The last step in preparing fusion protein constructs is to confirm the insert sequence. A list of sequencing primers that can be used to sequence inserts in Flexi® Vectors is available (9). Depending on the error rate of the DNA polymerase used in PCR, multiple colonies may need to be screened to find a clone with the exact protein-coding sequence. Because transfer between Flexi® Vectors

Table 2. Steps and Tools for Cloning into a Fusion Protein Expression Vector.

Cloning Steps	Tools
1. Design PCR primers to append restriction sites to the ORF and ensure the ORF is in-frame with the HaloTag® ORF.	Flexi® Vector Primer Design Tool **www.promega.com/techserv/tools/flexivector/**
2. Amplify ORF.	Polymerase (e.g., *Pfu* (Cat.# M7741)*, GoTaq® DNA Polymerase** (Cat.# M3001))
3. Purify PCR product.	Wizard® SV Gel and PCR Clean-Up System
4. Digest vector and PCR product.	Restriction enzymes, fragment purification kit, ligase (e.g., Flexi® System, Entry/Transfer (Cat.# C8640)***)
5. Gel-purify digested DNA.	
6. Ligate vector and insert.	
7. Transform into bacteria.	Competent *E. coli* (e.g., Single Step (KRX) Competent Cells [Cat.# L3001] or JM109 [Cat.# L2001])
8. Pick colonies and grow cultures.	LB broth + antibiotic
9. Purify plasmid.	Plasmid purification kit (e.g., PureYield™ Plasmid Miniprep System (Cat.# A1221)

*_Pfu_ DNA Polymerase is not for sale within the United States.
**_Taq_ Polymerase may be used but the higher error rate of this polymerase will require sequencing multiple colonies to identify an error-free clone.
***Product indicated is for N-terminal entry into Flexi® Vectors only and includes the Wizard® SV Gel and PCR Clean-Up System, restriction enzymes, ligase and buffers. C-terminal entry into Flexi® Vectors will require purchase of the Carboxy Flexi® Enzyme Blend (Cat.# R1901).

is high-fidelity, sequencing is not required when creating additional Flexi® constructs with the same ORF.

After obtaining a clone with the correct sequence, the HaloTag® fusion protein can be tested for full-length expression, folding, and activity. The speed and ease of use of cell-free expression systems are well suited for this testing in a single afternoon by expressing, labeling the HaloTag® protein with a fluorescent ligand, and analyzing by fluorescence scan of SDS-PAGE (1). The presence of a fluorescently labeled band indicates functional HaloTag® protein was expressed, and the label allows visualization and estimation of protein size by gel.

SUMMARY

Use of fusion proteins requires three basic steps to get started: planning, cloning and confirming. The multifunctional nature of the HaloTag® Technology means one fusion tag can be used for a wide variety of applications. The Flexi® Vector System gives the convenience of rapid cloning and a breadth of vector options. Together these technologies simplify the generation of fusion constructs and offer greater options and flexibility in experimental design.

PROTOCOLS

- *Flexi® Vector System Technical Manual* #TM254, Promega Corporation. **www.promega.com/tbs/tm254/tm254.html**
- *HaloTag® Technology: Focus on Imaging Technical Manual* #TM260, Promega Corporation. **www.promega.com/tbs/tm260/tm260.html**

TOOLS

Flexi® Vector Primer Design Tool
www.promega.com/techserv/tools/flexivector/default.htm

REFERENCES

1. Los, G. *et al.* (2005) *Promega Notes* **89**, 2–6.
2. In Vitro Protein Expression Guide (www.promega.com/guides/ive_guide/)
3. Kobs, G. (2008) *Cell Notes* **21**, 6–9.
4. Slater, M. *et al.* (2008) *Promega Notes* **100**, 16–8.
5. Arai, R. *et al.* (2001) *Pro. Eng.* **14**, 529–32; and references therein.
6. Slater, M. (2006) *Promega Notes* **93**, 8–10.
7. Manuscript in preparation.
8. *Flexi® Vector Systems Technical Manual* #TM254, Promega Corporation. (www.promega.com/tbs/tm254/tm254.html)
9. Wheeler, S., Strauss, E. and Schagat, T. (2008) *Promega eNotes* online www.enotes.applications/ap0099.htm

ORDERING INFORMATION

Product	Size	Cat.#
Flexi® System, Entry/Transfer	5 entry/ 20 transfer reactions	C8640
Flexi® System, Transfer	100 reactions	C8820
Carboxy Flexi® System, Transfer	50 reactions	C9320
Carboxy Flexi® Enzyme Blend (SgfI & EcoICRI)	50 µl	R1901

[a]Patent Pending.

[b]For research use only. With respect to manufacture or sale of research products or services, or any diagnostic, therapeutic or prophylactic uses, please contact Promega for supply and licensing information.

[c]The CMV promoter and its use are covered under U.S. Pat. Nos. 5,168,062 and 5,385,839 owned by the University of Iowa Research Foundation, Iowa City, Iowa, and licensed FOR RESEARCH USE ONLY. Research Use includes contract research for which monetary or other consideration may be received. Other commercial users must obtain a license to these patents directly from the University of Iowa Research Foundation.

[d]For research use only. Persons wishing to use this product or its derivatives in other fields of use, including without limitation, commercial sale, diagnostics or therapeutics, should contact Promega Corporation for licensing information.

[e]Licensed under U.S. Pat. No. 5,945,288.