Lanthanide-tagged proteins – An illuminating partnership

Running title: Lanthanide-tagged proteins

Karen N. Allen\textsuperscript{1*} and Barbara Imperiali\textsuperscript{2*}

\textsuperscript{1}Department of Chemistry, Boston University,
590 Commonwealth Avenue,
Boston, MA 02215-2521, USA,
e-mail drkallen@bu.edu

\textsuperscript{2}Department of Chemistry and Department of Biology,
Massachusetts Institute of Technology,
77 Massachusetts Avenue,
Cambridge, MA 02139, USA,
e-mail imper@mit.edu

\*Corresponding authors
Summary

Lanthanide-tagged proteins are valuable for exploiting the unique properties of Ln ions for investigating protein structure, function, and dynamics. Introduction of the Ln into the target is accomplished via chemical modification with synthetic lanthanide-chelating prosthetic groups or by coexpression with peptide-based binding tags. Complexed Ln-tags offer a heavy-atom site for solving the phase problem in X-ray crystallography. In NMR, paramagnetic lanthanide ions induce residual dipolar couplings and pseudo-contact shifts that yield valuable distance constraints for structural analysis. Lanthanide luminescence-based techniques and Ln-tagged proteins are valuable for investigating the functions and dynamics of large proteins and protein complexes and have been applied in vivo. Overall, the reach of Ln-tagged proteins will increase our ability to understand cellular functions on the molecular level.

Introduction

The unique photophysical and electronic properties of lanthanide ions (Ln), coupled with the absence of these rare earths in living systems, has stimulated the development of methods that exploit these properties to provide new insight into protein structure, function, and dynamics [1]. For selected proteins, the similarity between trivalent lanthanides (Ln\(^{3+}\)) and divalent calcium (Ca\(^{2+}\)) in terms of ionic radius and oxophilicity may enable direct substitution into calcium-binding proteins, providing a valuable spectroscopic handle for structural and dynamic studies [2-4].

In order to expand the number of targets and exploit the potential of lanthanides in the study of complex systems, the lanthanide ion must be site-specifically complexed to the target protein. In this context, two approaches have been developed (Figure 1). Synthetic lanthanide-chelating prosthetic groups may be incorporated into proteins either by integration into the side chain of a non-natural amino acid, or by chemical modification of uniquely reactive amino acids such as cysteine. The synthetic chelates may additionally incorporate organic fluorophores as sensitizers. For example, a diethylenetriaminepentaacetate (DTPA) chelate that is modified via the pendant carboxylate arms with a chromophore, carbostyril 124, and a thiol-reactive moiety, maleimide, has been used extensively for the generation of luminescent Ln-tagged proteins [5]. The attachment of lanthanide-ion binding tags through cysteine thiol modification is advantageous since it affords a rational means of generating Ln-tagged proteins with desired orientations.
between the lanthanide ion and the target protein for NMR applications [6] and allows strategic placement of probes for luminescence studies [7].

Alternatively, peptide-based lanthanide-ion binding tags (LBTs) may be incorporated into proteins by standard molecular biology techniques, thereby avoiding steps that may be necessary for optimizing site-selective chemical modification [8]. In this case, native protein sequences such as the Ca$^{2+}$-binding EF hand motifs, which show intrinsic binding to lanthanides, have been used as the starting point for the development of encoded peptides with lanthanide-binding properties. In an important early study Szabo established that the 14-residue peptide corresponding to an EF-hand motif from calmodulin could form a luminescent Tb$^{3+}$-chelate when the fluorescent tryptophan residue was incorporated at position 7 of the sequence [9]. Studies of this LBT prototype were consistent with a Dexter-type electron exchange model of energy transfer from the indole to the Tb$^{3+}$ center. This peptide later formed the foundation for split-and-pool based screens for the identification of improved LBTs [10], which maintained Ln binding when appended to the C- or N-termini of proteins and when integrated into intrinsic protein-loop structures. The LBT peptides have been subjected to extensive analyses revealing $K_{d}$s in the low nM range [11] and a highly ordered chelate structure including only peptide-based ligands without water in the inner complexation sphere [12] (Figure 2A), which is critical for minimizing Ln luminescence quenching [1]. Peptide-based LBTs have also been conjugated to proteins via cysteine modification [13].

Over the past six years Ln-tagging via the integration of synthetic or peptide-based chelates has been applied in studies to elucidate protein structure, conformational dynamics, protein-protein interactions and protein-ligand interactions. The inherent structural properties of the tags have allowed their use in membrane-bound proteins as well as in the cellular milieu. Herein we review selected applications of both synthetic and peptide-based lanthanide-ion binding tags and highlight their promise for future studies.

**Expanding the tool set for protein-structure determination**

All of the lanthanides provide excellent X-ray scattering power, and therefore complexed lanthanide-ion binding tags offer a programmable heavy-atom binding site for solving the phase problem in X-ray crystallography. This is especially useful in the absence of native metal-binding sites or when incorporation of selenomethionine is unfeasible due to limitations imposed by protein expression and stability. In phase determination using the anomalous signal from heavy
atoms, the metal must be well ordered in relation to the target protein. The concept that increased steric bulk would decrease mobility and promote formation of crystal contacts inspired the construction of a tag utilizing tandem LBT sequences in a double LBT (dLBT) as a macromolecular phasing tool [14] (Figure 2B). The bound Tb\(^{3+}\) in the dLBT was used to solve the phase problem for the structure determination of a construct encoding the dLBT tag as an N-terminal fusion to ubiquitin by single-wavelength anomalous diffraction. The structure of ubiquitin was unaffected by fusion with the dLBT. Moreover, as anticipated, the presence of the dLBT led to formation of a new crystal form of ubiquitin with the dLBT contributing to the crystal contacts. Future uses of lanthanide-ion binding tags may facilitate the crystallization of intractable proteins by modulating the protein surface, as has been previously done using site-directed mutagenesis or chemical modification.

In NMR, paramagnetic lanthanide ions that are fixed in position relative to the attached protein induce paramagnetic effects such as residual dipolar couplings (RDC) and pseudo-contact shifts (PCS). These paramagnetic effects have been utilized to provide long-range distance and angular information for proteins, thus yielding valuable constraints for NMR structural analysis of proteins. Information over large distances is especially useful in applications to multidomain proteins and multiprotein complexes, wherein PCS can provide distance and angular information up to 40\(\text{Å}\) from the nuclei under observation. Moreover, the employment of different lanthanide ions allows modification of the magnetic susceptibility anisotropy tensors and alignment tensors (Figure 3A) [13,15] and facilitates peak assignments by pairing peak positions from spectra of complexes with paramagnetic versus diamagnetic samples [16]. These biophysical effects have been exploited in NMR using intrinsic sites in metalloproteins [3], but the available effects and the target set have been greatly expanded through the use of lanthanide-ion binding tags [15-17].

The mobility of the tag, and hence the bound lanthanide ion reduces the anisotropic effect in structure determination via NMR. Although attachment at the protein terminus has provided useful RDC information [15-17], modifications that diminish mobility lead to increased signal. For instance, the use of a double lanthanide-binding tag to increase molecular mass (also used in crystallography, see above) led to a 3-fold enhancement in the RDCs versus the single LBT sequence [17]. Increased rigidity and signal enhancement has also been achieved via the symmetrical design of synthetic chelators to provide two attachment points through adjacent cysteines on the protein target [18], however, chiral purity and availability of these chelates has
limited development. This problem has been addressed by creation of a dipicolinic acid tag (that forms non-chiral metal complexes) bearing a single thiol group for cysteine attachment and leaves free coordination sites on bound lanthanides for coordination with nearby protein carboxylates [19]. Recently, the use of a lanthanide-binding peptide tag anchored to the target protein at two points via a disulfide and the N-terminus of an immunoglobulin binding domain produced 2-3 fold stronger anisotropic paramagnetic effects compared to a single point of attachment through the disulfide alone [20]. The improved rigidity of the tag was demonstrated from the size of the magnetic susceptibility tensor and alignment tensors obtained from PCS and RDC analysis. It is anticipated that further modifications to stabilize the point(s) of attachment and limit conformational freedom of ligating residues will additionally enhance the observed paramagnetic effects.

The use of RDCs and PCS in NMR via lanthanide-ion binding tags allows the acquisition of data where it would otherwise be thorny. For example, when ligands are carbohydrates, or otherwise make extensive hydrogen-bonding networks, NOEs between ligand and protein are diminished or eliminated. This problem was circumvented in the determination of the structure of a complex between Galectin-3 and lactose by using paramagnetism-based constraints introduced via a C-terminally fused peptidic lanthanide-ion binding tag loaded with Dy$^{3+}$ [16]. Additionally, improvements in the precision of NMR structures is obvious (Figure 3B) not only from the inclusion of RDCs and PCS but also from the addition of data from multiple lanthanide-ion tags of differing structure even when attached to a single site on the protein target [21].

Additional information can be obtained about the relative orientations of mobile protein domains through NMR by insertion of a lanthanide ion into one domain and utilizing RDCs to provide a solution structure in conjunction with PCS to independently obtain the global orientation tensor [4]. This was accomplished recently for calmodulin-peptide complexes through binding of the lanthanide to a single domain of the protein after introduction of a mutation to shift binding from Ca$^{2+}$ to Ln$^{3+}$ (N60D) [2]. In this study, the use of lanthanides as orienting devices in NMR structure determination demonstrated significant rearrangements between the solution and solid-state structures of the calmodulin-peptide complexes.

Lanthanide binding tag utility has also been extended to the study of interprotein dynamics via RDCs, which are sensitive to motions in the ps - ms time regime. Metal ion-induced alignment affords a fixed reference alignment for one component, allowing detection of motions of the
other. For example, the introduction of a caged lanthanide probe with two-point attachment to cytochrome C was utilized to measure the interactions with adrenodoxin in an electron-transfer complex between the two proteins, showing high relative mobility of the adrenodoxin component [18]. Lanthanide binding tag utility has thus been extended to the structural study of conformations of multi-domain proteins and protein complexes. These approaches studying intra- and inter-protein conformational dynamics via NMR can be complemented by those using luminescence through the use of the same LBT toolset.

**Protein trafficking and localization**

Luminescent lanthanide chelates are distinguished from the more common organic fluorophores in several respects that render lanthanide-tagged proteins valuable for biological studies [7]. In this context, Tb$^{3+}$ and Eu$^{3+}$ are the most commonly applied since these lanthanides emit light in the visible range and are more intense than others in the series. In addition, Tb$^{3+}$ and Eu$^{3+}$ exhibit long, ms, excited-state lifetimes and emission quantum yields are high. Importantly, while many lanthanide ions exhibit advantageous luminescence emission spectra, due to specific $f$–$f$ electronic transitions, these transitions are forbidden, and therefore it is more facile to excite these ions by sensitization with appropriate organic fluorophores [1]. For Tb$^{3+}$ the availability of tryptophan as a convenient sensitizing fluorophore means that this encoded amino acid can be used to sensitize Tb$^{3+}$. In luminescence experiments, the ms lifetimes of Tb$^{3+}$ and Eu$^{3+}$ provide greatly increased sensitivity and elimination of background fluorescence from typical short-lived organic fluorophores since time-gated data acquisition can be applied for the selective detection of Ln-tagged species.

Luminescent LBTs are useful handles for direct protein expression profiling of LBT-tagged proteins from crude cell extracts since the avid Tb$^{3+}$ binding and robust luminescence enables reliable quantification even in the presence of detergents and denaturants [22]. Recently, Hiroaki and coworkers have demonstrated that exogenously expressed, dual LBT/Protein Transduction Domain (PTD)-tagged proteins can be delivered into cells and visualized following fixation by observation of the Tb$^{3+}$ luminescence, when the LBT is loaded with Tb$^{3+}$ prior to cellular delivery [23]. LBTs have also been employed in protein constructs designed to assess the PTD activity of the heparin-binding domains of the human insulin-like growth binding proteins IGFBP-3 and IGFBP-5, thus revealing a new application of Ln-tagged proteins in the study of protein trafficking from outside to the inside of cells [24]. Notably, the “Tb$^{3+}$-loaded” LBT-tagged
proteins do not appear to cause toxicity to live mammalian cells and the tagged protein complexes appear to retain bound Ln during the course of the imaging experiments. In this context, recent advances also address enhancements of the LBTs to include approaches for sensitizing the lanthanide ions at longer wavelengths. Strategies for the introduction of Ln-tags into proteins that combine the binding properties of LBT peptides with the advantageous photophysical properties of synthetic sensitizers have been introduced [25]. For example, Ln-tagged proteins can be prepared using native chemical ligation to conjugate a synthetic C-terminal thioester LBT peptides that include acridone or carbostryril sensitizers, to expressed proteins thereby providing access to systems with either sensitized Eu\(^{3+}\) or Tb\(^{3+}\) emission. This strategy is useful since it provides for luminescence sensitization at longer wavelengths (340 nm for carbostryril and 390 nm for acridone vs 280 nm for the tryptophan indole) as well as the long wavelength sensitized Eu\(^{3+}\) emission (615 nm). Multiphoton excitation may also be employed for sensitizing Tb\(^{3+}\) and Eu\(^{3+}\) complexes and the two-photon excited luminescence of selected complexes have been reported [26].

**Protein interactions and dynamics**

Lanthanide luminescence-based techniques and Ln-tagged proteins are valuable tools for investigating the functions and dynamics of large proteins and protein complexes including ion channels [7,27], small molecule transporters [28], and the RNA polymerase complex [29,30]. In contrast to fluorescence resonance energy transfer (FRET), which has been a cornerstone of biological studies since the introduction of the “spectroscopic ruler” by Stryer and Haugland over 50 years ago [31], lanthanide-based or luminescence resonance energy transfer (LRET), was first introduced by Selvin in 1994 [32]. LRET, which relies upon the interaction of a luminescent lanthanide complex as donor and an organic fluorophore as acceptor, offers advantages over FRET for the measurement of distances in large biological complexes and represents a powerful tool for the study of multimeric integral membrane proteins. Distances up to 100 Å can be measured with considerably greater accuracy than by FRET since the Ln donor emission is unpolarized. Additionally, since measurements are based on the lifetime of the sensitized luminescent lanthanide, background emission from the acceptor fluorophore can be completely eliminated and complications from incomplete labeling are also circumvented since the measurements are concentration independent.
Integrated Ln-binding peptides for LRET have been applied in a study of lactose permease (LacY; *E. coli*), which catalyzes the co-transport of lactose and protons into cells [28]. In these studies, an engineered EF hand motif with a tryptophan sensitizer, was inserted into a predicted cytoplasmic loop of LacY (Figure 4A). The luminescent Tb\(^{3+}\) complex that was formed was then employed as an LRET donor with strategically placed acceptor fluorophore-labeled cysteines in the transmembrane helix (TMH) VI of the protein. LRET measurements in reconstituted proteoliposomes, using three different acceptor dyes at two locations on the TMH, are in excellent agreement and provide consistent information in the membrane-spanning 20-50-Å distance range. Genetically-encoded LBTs have also be applied in an LRET study of the *Shaker* K\(^+\) channel in *Xenopus* oocytes that exploits a transition-metal bound 6-His tag with Ni\(^{2+}\) or Cu\(^{2+}\) as the energy transfer acceptor [33] (Figure 4B).

The structural dynamics of the *Shaker* K\(^+\) channel, have also been subject to intensive study by Selvin and coworkers who have applied LRET using a synthetic lanthanide-chelating prosthetic group (Figure 1A) that is introduced via chemical modification with a maleimide derivative (Figure 5A). The channel is a homotetramer with each subunit comprising six transmembrane domains (S1-S6) (Figure 5B). One domain, S4, includes 7 positively charged residues and serves as the voltage sensor domain (VSD) together with the three equivalent domains from the other subunits. Several structures of the channel in the open state have been solved using X-ray crystallography [34] and there is considerable interest in developing an understanding of the dynamic states of the channel and how they relate to the channel function. LRET has been used to measure 19 different VSD to VSD (intersubunit) and VSD to pore-blocking charybdotoxin (subunit-ligand) distances, in channels expressed in *Xenopus* oocyte membranes [27] and controlled with a whole oocyte two-electrode voltage clamp (Figure 5C and D). The studies employed LRET between modified cysteines and BodipyFl or Atto465 fluorophores as acceptors, which were introduced either into another subunit or into the charybdotoxin (CTX) ligand. The studies provided Ångstrom-resolution measurements on the independent dynamics of S4 and S3 and suggested that these components of the subunits do not move as a rigid body as previously proposed. With the continued emphasis on high resolution X-ray structure determination and the further development of synergistic biophysical tools, such as LRET, there is considerable optimism that in the near future it will be possible to develop a unified picture of channel dynamics and how these relate to biological function.
Encoded LBT sequences also make LRET-based approaches readily applicable to the analysis of protein/ligand interactions. For example, the interactions of SH2 domains with partner phosphopeptide ligands can be evaluated by using LRET [35]. In one study the Src and Crk SH2 domains were expressed with C-terminal LBTs that could be loaded with Tb$^{3+}$, and LRET studies with cognate and non-cognate phosphotyrosine peptides tagged with BodipyFl and Bodipy-TMR were carried out to define affinities and provide distance information on the binding interactions. In addition to these studies of ligand interactions, the luminescence properties of LBTs have also been exploited for the development of protein kinase [36] and protease [37] assays as well as for the detection of insulin bound receptors for high-throughput screening [38].

**Conclusions**

As the need to understand the next dimension of cell function at the molecular level becomes critical, so does the necessity of addressing ever larger multi-protein and nucleic-acid complexes. The increasing size of structural problems tackled by NMR and X-ray crystallography produces a need that LBTs can fill. Moreover, the biosciences will benefit from lanthanide luminescent bioprobes, which can provide many advantages in sensing biomolecules and deciphering the signaling processes between them in living cells. Information on structure-function relationships of how protein complexes penetrate and locate in cells is lacking and can greatly benefit from the unique properties of Ln-tagged proteins.

**Acknowledgements**

We gratefully acknowledge Dr. Matthieu Sainlos and Kelly Daughtry for help in preparation of figures. We thank Dr. Harald Schwalbe for many helpful discussions concerning the applications of LBTs in NMR-based analyses. This work is supported by NSF MCB 0744415 to K.N.A. and B.I.
Figure Legends

Figure 1. Two methods for the site-specific incorporation of lanthanide-binding ion tags into target proteins
A) Attachment of synthetic lanthanide-chelating prosthetic groups (red) via chemical modification between a reactive group (star) and cysteines (shown here) or other amino-acids. LBT peptides may also be linked to the target protein by chemical modification of target cysteines.
B) Peptide-based lanthanide-binding ion tags (red) are encoded at the DNA level and fused to either the N or C-terminus of the target protein or into a loop.

Figure 2. Structure of the dLBT-ubiquitin fusion protein [14]
A) Detail of the high-resolution structure of the N-terminal Tb$^{3+}$-bound LBT (numbered from ligand 1)
B) The dLBT-ubiquitin structure (grey) was obtained with phases from the anomalous scattering of the bound Tb$^{3+}$. The structure highlights the compact, ordered nature of the dLBT and demonstrates it does not perturb the structure of the fusion partner (native ubiquitin, blue).

Figure 3. Uses of Lanthanides in NMR
A) Lanthanides span a wide range of magnetic anisotropies, such that refinement of NOE-based solutions using PCS is effective in concentric spheres of increasing distance from the lanthanide ion.
B) Improvements after refinement with PCS show that the Ln can be used to further constrain regions distal to the metal binding site (left panel, no metal, right panel Ce$^{3+}$) with improvements of rmsds from 0.74 and 1.10 Å for the backbone and all heavy atoms without PCS to 0.54 and 0.95 Å with PCS using Ce$^{3+}$ [39].

Figure 4. Use of Peptide-based LBTs for Investigating Intramolecular Distances in Membrane-bound Proteins via LRET
A) Transmembrane structure of LacY highlighting distances measured between an Ln-binding EF-hand motif inserted into a central cytoplasmic loop and fluorophore-labeled cysteine residues in transmembrane helix VI.
B) Use of Tb$^{3+}$-loaded LBT and Ni$^{2+}$- or Cu$^{2+}$-loaded 6-His tag for the measurement of intramolecular distances in *Xenopus* oocyte-expressed *Shaker* K$^+$ channel.

**Figure 5. Investigation of the Structural Dynamics of the *Shaker* K$^+$ Channel**

A) Maleimide-based cysteine modification agent (DTPA-cs124-EMPH) incorporating carbostyril-124 sensitizer and a DTPA chelate.

B) Topology diagram of one subunit of homotetrameric *Shaker* K$^+$ channel.

C) Top view of channel identifying location of donor and acceptor probes for VSD to VSD LRET-based measurements.

D) Top view of acceptor-modified charydotoxin (CTX) bound to *Shaker* K$^+$ channel identifying VSD to CTX LRET-based measurements.

Figures B, C, and D adopted from Ref. 27.
References and Recommended Reading

Papers of particular interest, published within the period of review, have been highlighted as:

• of special interest
  •• of outstanding interest

  •• This work uses X-ray data as a starting model and, with only a minimal set of NMR data, determines two previously undescribed calmodulin-peptide complexes. RDCs and PCS are implemented via lanthanides bound to a native Ca$^{2+}$ binding site.
  • Enhancements in NMR structure determination are obtained by attachment of lanthanide-binding peptides through different positions on the peptide to the same site on the protein target together with coordination to a series of lanthanide ions.
   • A double lanthanide-ion binding peptide is developed and utilized in X-ray crystallographic structure determination as a fusion to the N-terminus of ubiquitin. Notably, phases were determined solely from the bound lanthanide and were robust, allowing automated model building.
   • The problem of decrease or disappearance of NOEs in extensively hydrogen-bonded structures is circumvented in this study by the use of RDCs and PCSs incorporated through use of a lanthanide-ion binding peptide fused to Galectin at the C-terminus.
   • This paper reports the first demonstration of the uptake and subsequent visualization of an LBT-protein conjugate in living cells.


Figure 5