Journal of Theoretical Biology ■ (■■■■) ■■■–■■■

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

Role of structural flexibility in the evolution of emerin

Q1 Jia Yuan, Bin Xue *

Department of Cell Biology, Microbiology and Molecular Biology, School of Natural Sciences and Mathematics, College of Arts and Sciences, University of South Florida, 4202 E. Fowler Ave, ISA 2015, Tampa, FL 33620, USA

HIGHLIGHTS

- The middle disordered region of emerin contains multiple functional motifs.
- Q2 The substitution rates of majority of the functional motifs correlates positively with the predicted disorder score of these motifs.
- Human emerin sequence has evolved hydrophobic motifs in the middle regions, which may indicate the acquisition of novel functions during evolution.

ARTICLE INFO

Article history:
Received 8 April 2015
Received in revised form
7 August 2015
Accepted 17 August 2015

Keywords:
EMD
Phylogenetic analysis
Intrinsic disorder
Substitution
Mutation

ABSTRACT

Emerin is a short inner nuclear membrane protein with an LEM-domain at the N-terminal end and a transmembrane domain at the C-terminal end. The middle region of human emerin contains multiple binding motifs. Since emerin is often found in evolutionarily newer species, the functional conservation of emerin becomes an interesting topic. In this study, we have demonstrated that most of the functional motifs of emerin are intrinsically disordered or highly flexible. Many post-translational modification sites and mutation sites are associated with these disordered regions. The averaged substitution rates of most functional motifs between species correlate positively with the averaged disorder scores of those functional motifs. Human emerin sequence may have acquired new functions on protein–protein interaction through the formation of hydrophobic motifs in the middle region, which is resulted from accumulated mutations during the evolution process.

1. Introduction

Emerin is a major nuclear envelope protein that interacts with lamins and other nuclear lamin-associated proteins to form lamina, which is one of the most basic structural components of eukaryotic nuclear envelope and functions to maintain the stability of nucleus, regulate nuclear pore complexes, and coordinate the synergy between chromatin and transcription factors (Wilson and Dawson, 2011; Berk et al., 2013). Emerin has ~250 residues and belongs to the type-II inner membrane protein family. A group of members in this family, such as LAP2, Emerin, and MAN1, share a common domain, which is named as LEM after the initials of these three proteins (Tsuchiya et al., 1999; Lin et al., 2000; Lee et al., 2001; Lee and Wilson 2004). The LEM domain is a ~40 residue globular structural unit. NMR experiments revealed that the LEM domain is composed of an N-terminal helical turn and two large parallel alpha helices, with a set of conserved hydrophobic amino acids on the interaction interfaces (Cai et al., 2001; Laguri et al., 2001). In addition to the LEM domain, emerin at its C-terminal end has Q3 another hydrophobic transmembrane domain (TMD) of about 30 residues. These two domains account for about one third of the full length sequence of emerin. The rest two third of emerin sequence has no experimentally validated structural characteristics. Previous studies on the interaction between the intermediate region of emerin and other proteins found that the interaction follows the "touch and go" model in which two molecules interact with each other frequently and transiently (Shimi et al., 2004; Wheeler et al., 2007). This "touch and go" model of interaction is often found in Intrinsically Disordered Proteins (IDPs) (Dyson and Wright 2005; Yi et al., 2007; Dunker et al., 2008; Csizmok et al., 2008; Mittag et al., 2010; Zhuo et al., 2010; Nagulapalli et al., 2012; Khan et al., 2013; Mileo et al., 2013). Therefore, these observations may indicate that emerin is an intrinsically disordered protein or contains Intrinsically Disordered Regions (IDRs). The IDPs or IDRs are highly flexible proteins or regions that do not have enough hydrophobic interactions to hold different segments or amino acids together under physiological conditions (Arnone et al., 1971; Wright and Dyson 1999; Romero et al., 1999; Uversky et al., 2000; Tompa 2002). IDRs/IDPs do not have rigid structures but do play important biological functions especially in signaling and regulation through interaction

* Corresponding author. Tel.: +1 813 974 6007.
E-mail address: binxue@usf.edu (B. Xue).

http://dx.doi.org/10.1016/j.jtbi.2015.08.009
0022-5193/

with partners (Dyson and Wright 2005; Dunker et al., 2002; Minezaki et al., 2006; Iakoucheva et al., 2002; Uversky et al., 2005; Ward et al., 2004; Pawson and Scott 1997).

Emerin interacts with multiple partners. Emerin binds to lamins and therefore becomes one of the major structural components of lamina (Sakaki et al., 2001; Clements et al., 2000). The LEM domain of emerin interacts with BAF (barrier-to-autointegration factor) and BAF-DNA complexes. BAF dimers bind to double-stranded DNA non-specifically and thereby bridge DNA molecules to form a large, discrete nucleoprotein complex (Lin et al., 2000; Cai et al., 2001). The segment AA168–186 of emerin is consensus to a segment of adenomatous polyposis coli (APC) tumor repressor protein that binds to transcription activator β-catenin (Wheeler et al., 2007; Cartegni et al., 1997). Interestingly, the APC segment was demonstrated to be intrinsically disordered in our previous studies (Xue et al., 2012d, 2013b). High affinity between emerin and nuclear envelope protein nesprin was also reported (Mislow et al., 2002). In addition, emerin was shown to interact with splicing-associated factor YT521-B (Wilkinson et al., 2003), actin (Lattanzi et al., 2003), death-promoting transcriptional repressor Btf (Haraguchi et al., 2004), LIM domain containing protein Lmo7 (Holaska and Wilson, 2006), spermatogenesis associated protein GCL (Holaska et al., 2003), and many other proteins (Berk et al., 2013).

Due to its specific cellular location and central role in interacting with many other important proteins, malfunction of emerin is directly associated with several serious human diseases. Defects in the emerin gene are a cause of Emery-Dreifuss muscular dystrophy, an X-linked disorder characterized by early contractures, muscle wasting, weakness and cardiomyopathy (Bione et al., 1994, 1995; Yamada and Kobayashi, 1996). Another new aspect on the function of emerin is the recent controversial studies on the role of emerin in reducing the infectivity of HIV-1 on mammal cells through interaction between emerin and viral proteins (Jacque and Stevenson, 2006; Shun et al., 2007; Mulky et al., 2008). It seems that the roles of emerin in these studies depend on cell types and stages. In another study, HIV-1 tat protein interacts with human emerin, very likely through the interaction with lamin (Gautier et al., 2009). It was further reported that the emerin phosphorylation is a critical step in nuclear translocation of the viral preintegration complex (Bukong et al., 2010). Moreover, it was found that Herpes Simplex Viral proteins interact with human emerin and disrupt the association between emerin and lamin (Morris et al., 2007; Leach et al., 2007; Leach and Roller, 2010).

Although performing important functions, the emerin gene is not ubiquitous, especially in evolutionarily early species. This indicates that the emergence of emerin gene is evolutionarily late. This issue become more interesting when taking into consideration that emerin may be an intrinsically disordered protein, since the evolution of intrinsically disordered proteins has many unique features (Chen et al., 2006; Brown et al., 2010; Wrabl et al., 2011; Schlessinger et al., 2011; Brown et al., 2011; Jeong and Kim, 2012; Chemes et al., 2012; Xue et al., 2013a; Brunquell et al., in preparation). For this reason, we analyzed in this manuscript the predicted intrinsic disorder of emerin proteins from multiple species and explored the correlations among structural flexibility, functional motifs, and evolution. We proposed a quantitative measurement between averaged disorder score and averaged substitution rate of functional motifs. This study will greatly enhance our understanding on the role of structural flexibility in the process of evolution.

2. Methods

The major objectives of this project include correlating protein intrinsic disorder with evolution, and characterizing the change of functional motifs during evolution. Therefore, in addition to building a dataset of emerin sequences for evolutionary analysis, we also carried out disorder prediction using PONDR-FIT (Xue et al., 2010a) and PONDR©VLXT (Romero et al., 2001) predictors to measure disorder of single amino acids, using CH–CDF plot (Oldfield et al., 2005b; Xue et al., 2009) to show the disordered status of proteins at the level of entire protein, and completed prediction of binding motifs using MoRF predictor (Oldfield et al., 2005a).

2.1. Dataset

Human Emerin protein sequence (UniProt ID: P50402) was aligned against UniProt database using BLASTP (Altschul et al., 1990; Camacho et al., 2009) to search for possible homologs. Preliminary search with *E*-value of 10 resulted in 163 sequence hits. By removing fragments and restricting the *E*-value to 0.005, 72 sequences were selected for further analysis. The minimal sequence identity and alignment score of these sequences are 24% and 110%, respectively (the alignment score for human emerin against itself is 1335), which are comparable to the characteristics of final datasets in our previous studies on protein evolution (Xue et al., 2013b; Brunquell et al., in preparation). The sequences in the final dataset come from mammals, amphibian, reptile, and fish. The sequence length is normally around 250 residues, with several shorter sequences of ~100 residues in fish and several longer sequences of ~450 residues in non-emerin proteins. The sequences in the final dataset were also manually inspected in the UniProt database on their names, species, domain structures, and functions.

2.2. Disorder prediction

PONDR-FIT (Xue et al., 2010a) and PONDR©VLXT (Romero et al., 2001) were used to predict the per-residue disorder score for all sequences in the dataset. PONDR-FIT is one of the most accurate disorder predictors, reaching an overall accuracy of more than 80% (Yan et al., 2013; Zhang and Obradovic, 2011). PONDR©VLXT is a delicate tool to analyze the influence of local hydrophobicity on structural flexibility (Romero et al., 2001). The disorder scores were applied to infer the structural flexibility of amino acids or regions (Xu et al., 2012; Xue and Uversky, 2013). The similar structural flexibility often reflects similar structural dynamics and similar functional roles (Brunquell et al., in preparation; Glembo et al., 2012). The combination of disorder scores from PONDR©VLXT, PONDR-FIT, and other predictors was applied in many studies to explore a broad range of biological questions, such as methionine oxidation (Xu et al., 2012), phosphorylation (Xue and Uversky, 2013), P53 evolution (Brunquell et al., in preparation), binding motifs (Xue et al., 2010b), iPS transcription factors (Xue et al., 2012b), PTEN interactome (Malaney et al., 2013), Splicesome (Ribeiro et al., 2013), structural flexibility of viral proteins (Xue et al., 2012c), evolution across species (Xue et al., 2010c, 2012). Predictors IUPred (Dosztanyi et al., 2005) and PONDR-VSL2 (Peng et al., 2006) were also used in the disorder analysis as references.

2.3. CH–CDF plot

CH–CDF plot was employed to measure the flexibility of entire protein sequence (Oldfield et al., 2005b; Xue et al., 2009). CH–CDF plot was made from the outputs of another two individual predictors: Charge-Hydropathy (CH) plot (Uversky et al., 2000) and Cumulative Distribution Function (CDF) plot (Dunker et al., 2000). In each of these two individual plots, structured proteins and disordered proteins stay in different regions and are therefore

Please cite this article as: Yuan, J., Xue, B., Role of structural flexibility in the evolution of emerin. J. Theor. Biol. (2015), http://dx.doi.org/10.1016/j.jtbi.2015.08.009