Life Sciences 184 (2017) 58–70

Contents lists available at ScienceDirect

Life Sciences

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

Systematic prioritization of functional hotspot in RIG-1 domains using pattern based conventional molecular dynamic simulation

P. Raghuraman, R. Jesu Jaya Sudan, J. Lesitha Jeeva Kumari, C. Sudandiradoss *

Department of Biotechnology, School of Biosciences and Technology, VIT University, Vellore 632014, India

ARTICLE INFO

Article history:
Received 7 April 2017
Received in revised form 29 June 2017
Accepted 9 July 2017
Available online 11 July 2017

Keywords:
RIG-1
Motif
Distant ortholog
Mutagenesis
Hotspot
Essential dynamics

ABSTRACT

Background: Retinoic acid inducible gene 1 (RIG-1), multi-domain protein has a role-play in detecting viral nucleic acids and stimulates the antiviral response. Dysfunction of this protein due to mutations makes the route vulnerable to viral diseases.
Aim: Identification of functional hotspots that maintains conformational stability in RIG-1 domains.
Methods: In this study, we employed a systematic in silico strategy on RIG-1 protein to understand the mechanism of structural changes upon mutation. We computationally investigated the protein sequence signature for all the three domains of RIG-1 protein that encloses the mutation within the motif. Further, we carried out a structural comparison between RIG-1 domains with their respective distant orthologs which revealed the minimal number of interactions required to maintain its structural fold. This intra-protein network paved the way to infer hotspot residues crucial for the maintenance of the structural architecture and folding pattern.
Key findings: Our analysis revealed about 40 hotspot residues that determine the folding pattern of the RIG-1 domains. Also, conventional molecular dynamic simulation coupled with essential dynamics provides conformational transitions of hot spot residues among native and mutant structures. Structural variations owing to hotspot residues in mutants again confirm the significance of these residues in structural characterization of RIG-1 domains. We believe our results will help the researchers to better comprehend towards regulatory regions and target-binding sites for therapeutic design within the pattern recognition receptor proteins.
Significance: Our protocol employed in this work describes a novel approach in identifying signature residues that would provide structural insights in protein folding.

1. Introduction

The retinoic acid inducible gene 1 (RIG-1) like receptor (RLRs) which belongs to the IFN-stimulated gene (ISG) family [1] is a cytosolic receptor that sense pathogen-associated molecular patterns (PAMP) on viral RNA and thereby triggers an antiviral response by activating interferon [2,3]. RIG-1 receptor possesses three functional domains namely (i) N–terminal region consisting of (i) two tandem Caspase Activation and Recruitment Domain (CARD), (ii) a centrally located superfamily 2 (SF2) RNA helicase/ATPase domain (HEL) and (iii) a C-terminal Repressor Domain or Regulatory Domain (RD). While the CARD domain transduces signal upon viral stimulation, the central helicase domain is involved in recognition of viral nucleic acid substrates, following which it undergoes ATP hydrolysis to bind and unwind viral RNA. The C-terminal RD is involved in viral RNA recognition and is responsible for binding the 5′triphosphorylated RNAs [4–9]. RIG-1 signalling cascade involves series of steps which initiates with activation of HEL domain upon viral recognition at the RD domain. The activation of HEL domain is an ATP-dependent process which further allows the conformational change of the domain. The conformational change exposes the CARD domain which allows it to complex with the adaptor mitochondrial antiviral signalling (MAVS) protein [10–15]. This signalling complex further activates the transcription factors NF-κB and interferon regulatory factor (IRF)-3 to up-regulate the expression of pro-inflammatory and antiviral mediators [16]. However defective RIG-1 protein has been reported to increase susceptibility to viral disease and has shown unequivocal evidence on genetic mutation, especially single nucleotide polymorphism (SNP) which potentially alter the function of RIG-1. Several studies have listed out point mutations like E135A at CARD domain, K270A and D372A on ATP binding site region in HEL, and K888E in RD to be the most significant mutations that affect the function of RIG-1 [17–23].

Our keen interest was to identify and distinguish the impact of vital residues at the mutational site based on its contribution towards its structure and thereby its function. Since proteins exhibit sequence-

Abbreviations: RIG-1, Retinoic acid inducible gene 1; CARD, Caspase activation and recruitment domain; HEL, Helicase domain; RD, Repressor domain; DO, Distant ortholog; RMSD, Root mean square deviation; RMSF, Root mean square fluctuation.

* Corresponding author.
E-mail address: csudandiradoss@vit.ac.in (C. Sudandiradoss).

http://dx.doi.org/10.1016/j.lfs.2017.07.011
0024-3205/

structure-function relationship, sequence mutations will in most cases cause an aberrant loss of function through the loss or altered structural architecture of the proteins. In order to identify such structural changes upon mutation in RIG-1 domains we performed a comparative analysis using a novel approach that initiated with the prediction of RIG 1 domain signatures. A signature or a pattern represents a short continuous stretch of residues that is conserved among the protein of interest across different genera. Signatures are usually associated with the function or structure of the proteins therefore signature prediction is a crucial step to identify the structurally and functionally essential residues of the protein [24]. Signatures are obtained from the conserved region of the proteins, and are usually represented as X - [XYZ…]-Y-{X…}, where X, Y, Z without brackets are any amino acid that is highly conserved at that position, residues within [X] are possible substitutions that may occur at that site and within {X} are residues that has to be excluded at that site. Signatures are obtained by comparing two or more sequences of a specific protein from different genera through multiple sequence alignment. In this way, the signatures predicted from CARD, HEL and RD domains were eventually used to identify a distant ortholog (DO) of human RIG-1 protein. A distant ortholog in the present sense denotes RIG-1 domains identified from a cross-genera having a least similarity to the human RIG-1 protein. Previously, we have observed in NOD2 published elsewhere [25] with many other proteins resulted in such a way that despite having a wide variation between the sequences, the structure of the native domain and that of its DO display same structural fold [26–28]. This minimal sequence similarity should therefore provide substantial information to form the structural fold of the protein. As a result, we preferred distant ortholog over a homolog to identify the minimal number of residues required by the protein to maintain its structural architecture. Further, we rapidly generated the in silico protein models for DO's by utilizing homology modelling approach in all the context from the sequence identity to the alignment that falls in the safe zone (amino acid match >40%). Conventional molecular dynamics simulation of native, DO and mutant RIG-1 domains was also carried out to stabilize the structures [29–31]. It enhances our understanding on the atomistic behaviour of structural changes upon mutations [32], through analysis of bonded and non-bonded interactions, root mean square deviation (RMSD), root mean square fluctuation (RMSF), molecular motions and flexibility analysis. Structural comparison between the native RIG-1 domain and its ortholog, inter and intra protein interactions of RIG-1 domains and domains motions through principle component analysis was carried out to reveal the critical hotspots in RIG-1 domains that play an essential role in structure and function of the protein. From the results, we elucidated the sequence-structure-function relationship of RIG-1 and reported in terms of functional hotspots which serve as candidate sites for drug development.

2. Materials and methods

2.1. Datasets

The amino acid sequence of RIG-1 domains (CARD, HEL, and RD) was retrieved from the Uniprot/Swissprot database with the sequence ID: 095786 [33]. The entire protein sequence length was of 925 aa which comprise of CARD (1–180), HEL (251–430) and RD (735–925). The nonsynonymous single nucleotide polymorphisms (nsSNPs) of RIG-1 listed in UniProtKB were E135A (CARD), K270A (HEL) and K888E (RD) [34]. These mutations were cross-verified with Human Genome Mutation Database (HGMD) [35] a repository for disease-oriented and clinically validated mutations. All the three position-specific variants were reported significantly in experimental studies which correspond to diseases or decrease protein activity and thus were considered for our analysis.

2.2. Multiple sequence alignment and signature prediction

The sequence patterns that are highly conserved among a group of protein sequences are considered as protein signature which defines protein family, domains, active site or any other functional sites [24–25,36]. The conserved sequence motifs with high resemblance tend to share similar structural architecture which is considered to be more potential in folding. Therefore, we intended to predict the mutation based signatures for each RIG-1 domains from conserved regions of the sequences identified through multiple sequence alignment using ClustalW [37]. For this purpose, Gonnet matrix slow alignment method was used with gap penalties of 10 and 0.1 to open/extension for pairwise sequence alignment. For multiple sequence alignment the matrix was kept the same and gap penalties were set to 10 and 0.2. The minimum gap position was set to 15, the sequences having an E value >1.0 were excluded, and no iteration was specified which ensures relatively good score range for alignment [38–40]. For each domain, about, 10 RIG-1 protein sequences from different species were selected and subjected to multiple sequence alignment. From the conserved regions of each domain, signatures were identified which are derived from the conserved regions of the multiple sequence alignment. The signatures were written in the regular PROSITE format and further validated using scanprosite at PROSITE [39]. Subsequently, each of the predicted signatures was scanned across UniProtKB/SwissProt [33] and TrEMBL database [40]. The total number of hits obtained for each query was recorded, and the motifs are demonstrating the greatest number of hits with a similar protein, irrespective of varying taxonomical classes was considered the best. To ensure the uniqueness of the signature, patterns having an identity to any other protein were dismissed. The validated signatures were subjected to comparative analysis to identify sequence-structure relationships and to decipher the hotspots.

2.3. Pattern based search for identifying distant ortholog structures (DO)

Comparative structural analysis helps in identifying structurally similar or distinct patterns in an orthologous or paralogous group of proteins. In order to predict the sequence-structure-function relationship of the RIG-1 domains, we performed a signature based search using Pattern Hit Iterative blast (PHI-Blast) [41,42] which lists all possible proteins that belong to the same family. From the hits obtained we identified the least similar protein (distant ortholog structure-DO) for the comparative analysis. The aim of this approach was to identify the most common and conserved residues that are required to maintain the secondary structure and folded conformation of the protein [26,27,43]. This concept arises from the theory that protein structural similarity leads to functional similarity and vice-versa. The DO structure reported here, represents a distant ortholog of RIG-1 which share similar structure architecture despite a very low identity with the target sequence. The reason for choosing a DO structure was to predict the minimal number of residues that are required for folding into the specific functional architecture. The same approach was applied to all three RIG-1 domains individually, and their respective DO's were identified.

2.4. Structure prediction and validation

The 3D structures of the RIG-1 domains were obtained from Protein Data Bank (PDB) [44] using the PDB ID 4P4H for CARD, 4ON9 for HEL and 3TMI for RD. Similar search was made for DO domains identified through PHI-Blast. Since the structures of DO protein were unavailable, we employed homology modelling method for structure prediction using SWISS-MODEL server [45]. Here, we used the method of choice in which the query was searched on the basis of global sequence or structural similarity for known template [30,45–47]. The obtained structures of DO for each of the domains (CARD, HEL and RD) were further validated using Structure Analysis and Verification Server (SAVES), MolProbity [48], Swiss-PDB Viewer [49] 3D refine [50] and QMEAN