Role of highly central residues of P-loop and its flanking region in preserving the archetypal conformation of Walker A motif of diverse P-loop NTPases

Ekta Pathak¹, Neelam Atri² & Rajeev Mishra†*

¹Department of Bioinformatics (MMV), Banaras Hindu University, India; ²Department of Botany (MMV), Banaras Hindu University, Varanasi - 221005, India; Rajeev Mishra – Email: mishrarajeev@gmail.com; Phone +91 9935338891; *Corresponding author

Received December 03, 2012; Accepted December 05, 2012; Published January 09, 2013

Abstract:
P-loop NTPases represent a large and highly diverse protein family that is involved in variety of cellular functions. Walker A motif forms a typical arched conformation, necessary to accommodate the phosphate moiety of the nucleoside tri (or di-) phosphate in P-loop NTPases. The feature that maintains the ancient architecture of P-loop is unidentified and uncharacterized. Here, using a well established global network parameter, closeness centrality, we identify that Walker A and its flanking regions (N- and C-terminal) have high density of globally connected residue positions. We find that closeness centrality of these residue positions are conserved across common structural core of diverse domains of P-loop NTPase fold. Our results suggest the potential role of globally connected residues in maintaining the local conformation of P-loop.

Background:
P-loop NTPases represent a large protein family that are involved in variety of cellular functions, for example, in signal transduction, translation, protein transport and localization, signal-sequence recognition, chromosome partitioning, and membrane transport [1-3]. Walker A also known as phosphate binding loop (P-loop) is a common feature of P-loop NTPase fold that bind nucleotide. The consensus sequence of Walker A (GXXXXG[K/S/T], where X is any residue) is often used as a motif for identifying new members of this group [4-6]. Walker A sequences are also present in many proteins that do not form P-loop, for example, peroxidases, and enzymes like α-amylase, glutamate dehydrogenase, Taq polymerase, carbonic anhydrase, binding proteins (lectin, trypsin inhibitor), proteases, and others [7]. Here, we investigated the features that maintain the P-loop architecture by employing a well established global network parameter closeness centrality. Protein structures can be represented as a residue-residue interaction network where the residues are nodes and interactions between them constitute edges. This approach has been useful in various studies like predicting functional residues in enzyme families [8], protein structure flexibility [9], protein folding [10], and side-chain clusters [11]. Closeness centrality is a global network parameter that correlates more accurately with critical residues than any other centrality measurement tested [12]. High closeness residues interact directly or by a few intermediates with all other residues of the protein [13]. By definition, closeness-centrality is calculated by mean distance of a node (residue) to all other nodes (residue) in the network. Amitai et al., [8] have shown that important residue positions like those involved in substrate and co-factor binding, catalysis, and mutation intolerant residues show high closeness centrality in networks. Del sol et al. [13] have shown that centrality residues integrate and propagate the information to all other residues in protein. Here, we show that Walker A and its flanking regions (N- and C-terminal) have high density of high closeness centrality.
residue positions in P-loop NTPases. We report that closeness centrality of these residue positions are conserved across common structural core of Ras superfamily and diverse domains of P-loop NTPase fold. No such high densities of high centrality residue positions are observed in the proteins containing Walker A sequence that do not form P-loop. The presented data clearly indicate the role of globally connected residues in conservation of the local conformation of an ancient motif such as Walker A.

**Figure 1:** A) Ribbon diagram of typical architecture of P-loop (Red) with bound nucleotide molecule (stick) of Ras superfamily proteins (Ras (green), Rab (cyan), Rho (blue), Ran (yellow), and Arf (magenta)); B) Ribbon diagrams of typical architecture of P-loop (red) in representatives of diverse P-loop containing NTPases. 4 letter words are the PDBID.

**Methodology:**

**Selection of structures of P-loop containing NTPases**

High resolution X-ray crystallographic structures of diverse domain of P-loop containing NTPases were used in the study. Initially, ScopTree search of protein databank (http://www.rcsb.org/pdb) was used to retrieve a set of 1203 structures of P-loop containing nucleoside triphosphate hydrolase. The search was then refined to 227 distinctly related protein structures by using ScopTree homologue removal tool at 30% sequence identity cutoff. This was primarily done to avoid redundancy and utilize the diversity present in the P-loop NTPases. Complete structures (i.e., without chain breaks or missing residues) with resolution ≤ 2.4 were chosen. Finally, we selected 23 structures of P-loop NTPases Table 1 (see supplementary material). We retrieved 22 PDB files for protein structures containing Walker A sequence (GXXXXGKS/T) that do not form the P-loop Table2 (see supplementary material) [7].

**Computation of closeness centrality**

Protein structures can be represented as a residue-residue interaction graphs in which amino acid residues serve as the nodes and their interatomic contacts are the edges. Closeness centrality correlates more accurately with critical residues than any other centrality measurement tested [12]. Therefore, we used SARIG server which efficiently calculates the closeness centrality (please see supplementary material for calculation and explanation).

Beginning with the atomic coordinates of a protein structure, server calculates the interaction between each pair of atoms by using the CSU program [14]. Closeness values were calculated for each residue and standardized by calculating their standard deviation from the mean value. The z-score of the closeness centrality was calculated by z-score = (C(x) − μ) / σ, where μ is the mean value of closeness and σ is the standard deviation. The residues with z-score ≥ 1.0 were considered significant (for detailed descriptions, please refer to Amitai et al [8]). Protein structure analysis was performed using Chimera (http://plato.cgl.ucsf.edu/chimera).

**Results and Discussion:**

Walker A motif forms a typical architecture in P-loop fold NTPase (Figure1A & 1B). A distortion in the P-loop conformation makes it incompatible with the binding of nucleotides [15]. The features that contribute in preserving the architecture of this ancient motif remain unidentified and uncharacterized. Therefore, an important and open question is how P-loop forms a typical architecture in structurally and functionally diverse P-loop NTPases. Here, we used a well established closeness centrality network parameter to study the global impact of residues on the typical local conformation of P-loop. Residues with high closeness value are central in network and interact with other residues directly or by a few intermediates [8].

**High closeness residue positions around P-loop and its flanking regions in Ras Super family members**

In order to understand the P-loop architecture, we first analyzed the residue-residue interaction network of Ras superfamily (Ras: 5P21; Rab: 3RA; B; Ran: 1IBR; Rho: 1M7B and Arf: 1R4A) experimental structures in GTP bound form. Interestingly, Walker A and its flanking regions showed high density of high closeness residue positions (Table 1). Here, the
High density of highly conserved residue positions in P-loop and its flanking regions in diverse set of P-loop NTPases

Since the Ras superfamily belongs to P-loop NTPase fold, we then extended the centrality analysis on high resolution X-ray crystallographic structures of P-loop NTPases (Table 1). The structural overlay of highly diverse P-loop NTPase fold showed that the typical P-loop architecture is maintained (Figure 1B). In order to avoid redundancy and utilize the diversity present in the P-loop NTPases, we selected a set of 23 NTPase structures at 30% sequence identity cutoff (see methodology). We wanted to look at the impact of sequence diversity on the closeness value of the residues of P-loop and its flanking region. Intriguingly, the highly diverse P-loop NTPase exhibited a similar pattern of high density of conserved high closeness centrality residue positions around Walker A motif, as seen in Ras super family. Here, the conserved high closeness centrality positions are defined as those positions with statistically significant closeness values (z-score ≥1.0) in at least 60% of the structures of P-loop NTPase fold (Figure 2 & Table 1). 11 such residue positions around Walker A and its flanking regions showed high closeness value. Four contiguous residue positions (N2-N5) of the N-terminal, two residue positions of C-terminal (C2-C3) and five residue positions of Walker A (W1, W2, W5, W6, W7) were showing high closeness centrality. The residue positions N4 (100%), W7 (96%) and C2 (100 %) were highly conserved in their centrality across the diverse structures. The invariant residue positions (G, K, S/T) and variant residue positions (W2) of Walker A showed high closeness centrality (Table 1). Walker A sequence has wider distribution and observed in many proteins that do not bind nucleotides [7]. The structural analysis revealed that these proteins do not form the conspicuous P-loop architecture [7]. To test our prediction, we calculated the closeness value in Walker A sequences that do not form P-loop (Table 2). We did not observe high density of high closeness centrality pattern.

Our results indicate the high density of conserved high closeness residue positions in P-loop and its flanking regions in P-loop fold NTPase and underscore its role in supporting the architecture of P-loop. The study presented is in concord with the observation that highly central residue positions correlate well with active site residues or their neighbors that provide supportive scaffold [13]. However, high closeness value of invariant (G, K, and S/T) residues of Walker A indicates its role in catalysis. P-loop lysine interacts and forms hydrogen-bond with oxygen of γ-phosphate of bound nucleotide and serine/threonine binds with Mg2+ [16, 17]. Recently Grüber et al. [15] demonstrated the role of conserved glycine residues of Walker A motif in guarding the active site region for nucleotide entrance in archaea-type ATP synthases. The altered conformation of the P-loop resulted in the active-site region being closed to nucleotide entry [15].

Conclusion:

In the context of network, protein structural scaffold and sequence diversity can be visualized as a dramatic change in the type of node, and also the connections between the nodes. Regardless of such diversity, depicted in Ras superfamily and diverse domains of P-loop fold NTPase, the closeness centrality of residue positions in P-loop and its flanking regions are remarkably maintained to be high. Thus, our finding supports the observation that centrality of a residue is maintained evolutionarily to assure the proper functioning of protein [8, 13]. We did not find such high centrality residue positions in proteins containing Walker A motif that do not form P-loop. This strengthens the evidence that required geometry of
archetypal P-loop is achieved by high density of residue positions which are globally connected in short steps.

Acknowledgment: EP acknowledges CSIR, India for the JRF and SRF. We thank Department of Bioinformatics, MMV, Banaras Hindu University, for providing computational facility.

References:
[1] Koonin EV et al. Trends Biochem Sci. 2000 25: 223 [PMID: 10782090]
[2] Saraste M et al. Trends Biochem Sci. 1990 15: 430 [PMID: 2126155]
[3] Vetter IR et al. Q Rev Biophys. 1999 32: 1 [PMID: 10800520]
[4] Walker JE et al. EMBO J. 1982 1: 945 [PMID: 6329717]
[5] Wolf YI et al. Genome Res. 1999 9: 17 [PMID: 9927481]
[6] Ma BG et al. Biochem Biophys Res Commun. 2008 366: 607 [PMID: 18073136]
[7] Ramakrishnan C et al. Protein Eng. 2002 15: 783 [PMID: 12468712]
[8] Amitai G et al. J Mol Biol. 2004 344: 1135 [PMID: 15544817]
[9] Jacobs DJ et al. Proteins. 2001 44: 150 [PMID: 11391777]
[10] Vendruscolo M et al. Phys Rev E Stat Nonlin Soft Matter Phys. 2002 65: 0619101 [PMID: 12188762]
[11] Kannan N & Vishveshwara S. J Mol Biol. 1999 292: 441 [PMID: 10493887]
[12] Thibert B et al. BMC Bioinformatics. 2005 6: 213 [PMID: 16124876]
[13] del Sol A et al. Protein Sci. 2006 15: 2120 [PMID: 16882992]
[14] Sobolev V et al. Bioinformatics. 1999 15: 327 [PMID: 10320401]
[15] Kumar A et al. J Mol Biol. 2010 396: 301 [PMID: 1994110]
[16] Crampton DJ et al. Biochemistry. 2001 40: 3710 [PMID: 11297439]
[17] Runquist JA et al. Biochemistry. 2001 40: 14530 [PMID: 11724566]

License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited

Edited by P Kanguane
Citation: Pathak et al. Bioinformation 9(1): 023-028 (2013)
Methodology: Computation of closeness centrality

Protein structures can be represented as a residue-residue interaction graphs in which amino acid residues serve as the nodes and their interatomic contacts are the edges. Closeness centrality correlates more accurately with critical residues than other centrality measurement tested [12]. Therefore, we used SARIG server which efficiently calculates the closeness centrality (http://bioinfo2.weizmann.ac.il/~pietro/SARIG/V3/index.html). Closeness centrality of node \( x \) is calculated as follows:

\[
C(x) = \frac{1}{\sum d(x, y)} \quad \forall x, y
\]

Where \( d(x, y) \) is the shortest-path between node \( x \) and any node \( y \). \( U \) is the set of all nodes and \( N \) is the number of nodes in the network.

### Table 1: Closeness centrality (z-scores) of residues of Walker A and flanking region sequence in P-loop NTPases H (%) is the percentage of structures with z-score ≥ 1.0. HCR is the number of High closeness residues with z-score ≥ 1.0.

| PDB ID | N1 | N2 | N3 | N4 | N5 | W1 | W2 | W3 | W4 | W5 | W6 | W7 | W8 | W9 | C3 | C4 | C5 | HCR |
|--------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|-----|
| 1FMJ_A | ASP | VAL | PHE | VAL | ALA | SER | TYR | GLN | ARG | SER | GLY | THR | THR | THR | THR | THR | GLN | GLU | LEU | 12 |
| 1D2N_SA | LEU | LEU | VAL | LEU | GLY | GLY | PRO | PRO | HIS | GLY | LYS | THR | THR | ALA | LEU | ALA | LYS | LYS | LYS | 15 |
| 1J1F_A | ALA | TYR | LEU | PHE | SER | GLY | THR | ARG | VAL | GLY | VAL | LYS | THR | SER | ILE | ALA | ARG | LEU | 13 |
| 1M4V_SA | TRP | TRP | ILE | LYS | PRO | ILE | ASP | SER | GLY | LYS | THR | THR | LEU | ALA | ALA | LYS | LYS | LYS | 9 |
| 1R3F_SA | LEU | ILE | VAL | MET | VAL | VAL | VAL | VAL | VAL | VAL | VAL | VAL | VAL | VAL | VAL | VAL | VAL | VAL | 10 |
| 1C2P_SA | GLN | GLY | ILE | TYR | GLY | LYS | LYS | LYS | LYS | LYS | SER | LYS | LYS | LYS | LYS | LYS | LYS | LYS | 15 |
| 1G3Q_A | ILE | SER | ILE | GLY | GLY | LYS | LYS | GLY | THR | THR | LYS | THR | THR | THR | VAL | THR | VAL | ALA | ASN | 14 |
| 1N6P_B | LEU | LEU | PHE | VAL | VAL | PHE | VAL | PHE | GLU | THR | LYS | THR | THR | LEU | LYS | LYS | LYS | LYS | 9 |
| 1Y1B_A | TRP | VAL | ILE | ALA | GLN | SER | GLN | SER | GLY | THR | LYS | THR | THR | LEU | LYS | LYS | LYS | LYS | 12 |
| 2HYC_A | ASP | ILE | VAL | ILE | GLN | SER | GLN | SER | GLY | THR | LYS | THR | THR | ALA | PHF | SER | ILE | LYS | 9 |
| 2J0S_A | ASP | ILE | LEU | ILE | GLY | SER | LYS | GLN | SER | LYS | LYS | LYS | TH | THR | LEU | LYS | LYS | LYS | 9 |
| 1L1Y_A | Ile | LEU | TRH | ILE | GLY | CYS | ILE | ARG | GLY | SER | LYS | LYS | SER | THR | TRP | ALA | ARG | GLU | 11 |
| 1H65_A | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | 15 |
| 1CHT_A | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | 15 |
| 1KQf_A | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | 15 |
| 1NOA_A | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | 15 |
| 1U1J_A | ASP | GLY | VAL | GLY | SER | GLY | GLY | THR | THR | THR | THR | THR | THR | THR | THR | THR | THR | THR | THR | 15 |
| 1BZ5_A | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | 15 |
| 1KQN_A | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | 15 |
| 1UVI_A | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | ASP | 15 |

ISSN 0973-2063 (online) 0973-8894 (print)
Biotechnology Information 91:123-208 (2013) 27
© 2013 Biomedical Informatics
## Table 2: Closeness centrality (z-score) of residues of proteins containing Walker A sequence, which do not form P-loop. H (%) is the percentage of structures with z-score ≥ 1.0

| PDB ID | N1 | N2 | N3 | N4 | W1 | W2 | W3 | W4 | W5 | W6 | W7 | W8 | C1 | C2 | C3 | C4 | C5 |
|--------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| GMTH-A | 1.051 | 0.901 | 0.155 | 0.561 | 0.546 | 0.494 | 1.454 | 0.519 | 0.342 | 0.036 |
| 1QPA-A | 0.349 | 0.462 | 0.135 | 0.789 | 0.908 | 0.696 | 1.675 | 1.518 | 0.975 | 0.314 |
| 2CYP-A | 0.889 | 0.293 | 1.175 | 1.736 | 1.388 | 1.254 | 1.874 | 1.072 | 0.776 | 1.834 | 0.642 | - | - | - | - | - |
| 1DS4-A | 0.799 | 0.185 | 1.136 | 1.779 | 1.368 | 0.972 | 1.91 | 0.402 | 0.694 | 1.951 | 0.782 | 0.585 | - | - | - | - |
| 1MTD-D | 0.101 | 0.079 | 0.912 | -0.04 | 0.186 | - | - | - | - | - | - | - | - | - | - | - |
| 1IMQ-D | 0.236 | 0.654 | 0.639 | 0.115 | - | - | - | - | - | - | - | - | - | - | - | - |
| 1STE-A | 0.244 | 0.697 | 0.667 | 0.023 | 0.135 | 0.059 | 0.805 | 0.871 | 0.384 | 0.044 | - | - | - | - | - | - |
| 1CQ5-A | 1.242 | 1.386 | 1.361 | 2.039 | 2.054 | 0.989 | 1.998 | 1.533 | 0.926 | 1.102 | 0.155 | - | - | - | - | - |
| 1QGW-A | 0.132 | 0.721 | 0.934 | 0.531 | 0.449 | 0.439 | 0.225 | 0.9 | 1.925 | 1.751 | 1.477 | 1.184 | 2.247 | 1.626 | - | - |
| 1CVW-A | 1.315 | 1.311 | 1.045 | 1.065 | - | - | - | - | - | - | - | - | - | - | - | - |
| 1FS7-A | 1.130 | 1.366 | 1.014 | 1.59 | 1.191 | 0.851 | 0.106 | 0.409 | - | - | - | - | - | - | - | - |
| 1CGH-A | 0.141 | 0.451 | 0.65 | 1.253 | 0.764 | 1.688 | 2.126 | 2.004 | 1.437 | 0.605 | - | - | - | - | - | - |
| 1CN_A | 1.124 | 0.244 | 0.297 | - | 0.946 | 0.965 | 0.298 | 0.366 | 0.388 | 0.229 | 0.885 | 0.121 | 0.433 | 0.388 | 1.247 | 0.889 | 0.552 |
| 1HLY-A | 0.358 | 0.868 | 0.611 | 0.557 | 0.473 | 0.051 | 1.191 | 0.692 | 1.044 | 0.311 | - | - | - | - | - | - |
| 1SA-A | 0.055 | 0.014 | 0.018 | 0.018 | - | - | - | - | - | - | 0.769 | 0.034 | -1.04 | - | - | - |
| 1JAE-A | 1.063 | 1.615 | 1.107 | 1.048 | 1.915 | 2.186 | 2.067 | 2.333 | 1.314 | 2.135 | 1.455 | 1.428 | 0.691 | - | - | - |
| 1DDZ-A | 0.524 | 0.566 | 0.872 | 1.187 | 2.068 | 1.835 | 1.582 | 0.925 | 0.865 | 0.954 | 0.912 | 1.488 | 1.039 | 1.874 | 1.196 | 2.024 | 1.857 |
| 1DYW-A | 0.856 | 1.327 | 0.637 | 0.417 | 0.471 | 0.293 | 0.543 | 0.097 | 0.829 | 0.707 | 0.176 | 0.246 | 0.655 | 0.057 | - | - | - |
| 1QHB-A | 0.036 | 0.579 | 0.903 | 0.938 | 0.061 | 0.838 | 0.485 | 0.855 | 0.801 | - | - | - | - | - | - | - | - |
| 1CA-A | 1.476 | 0.553 | 1.191 | 1.426 | 0.836 | 0.261 | 0.973 | 0.671 | - | - | - | - | - | - | - | - | - |
| 1FO7-A | 0.51 | 0.153 | 0.229 | 0.951 | 1.062 | 1.374 | 1.634 | 1.748 | 1.454 | 1.428 | 1.032 | 2.096 | 0.941 | - | - | - | - |
| 4PGA-A | 0.237 | 0.752 | 0.74 | 0.681 | 0.889 | - | - | - | - | - | - | - | - | - | - | - | - |

**H (%)** is the percentage of structures with z-score ≥ 1.0.