A peer-reviewed version of this preprint was published in PeerJ on 8 January 2015.

View the peer-reviewed version (peerj.com/articles/724), which is the preferred citable publication unless you specifically need to cite this preprint.

Singh GP. 2015. Association between intrinsic disorder and serine/threonine phosphorylation in *Mycobacterium tuberculosis*. PeerJ 3:e724 https://doi.org/10.7717/peerj.724

Association between intrinsic disorder and serine/threonine phosphorylation in *Mycobacterium tuberculosis*

Gajinder Pal Singh

School of Biotechnology,

KIIT University, Campus-XI, Patia, Bhubaneswar 751024,

Odisha, India

Tel : +91 674 2725732

Email: gajinder.pal.singh@gmail.com

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.564v1 | CC-BY 4.0 Open Access | rec: 29 Oct 2014, publ: 29 Oct 2014

Abstract

Serine/threonine phosphorylation is an important mechanism to regulate protein function. In eukaryotes phosphorylation occurs predominantly in intrinsically disordered regions of proteins. While serine/threonine phosphorylation and protein disorder are much less prevalent in prokaryotes, *M. tuberculosis* has both high serine/threonine phosphorylation and disorder. Here I show that, similar to eukaryotes, serine/threonine phosphorylation sites in *M. tuberculosis* are highly enriched in intrinsically disordered regions, indicating similarity in substrate recognition mechanism of eukaryotic and *M. tuberculosis* kinases. Serine/threonine phosphorylation has been linked to the pathogenicity and survival of *M. tuberculosis*, thus better understanding of how its kinases recognize their substrates could have important implications in understanding and controlling the biology of this deadly pathogen.

Introduction

Reversible phosphorylation of serine and threonine residues is a widespread post-translational modification in eukaryotes, with more than a third of proteins phosphorylated during their lifetime [1] . Phosphorylation can modify protein interactions, enzyme function, localization or degradation. While traditionally, regulation and signal transduction in bacteria was thought to be mediated by histidine and aspartate phosphorylation in two-component systems, recently the occurrence and importance of phosphorylation of serine/threonine (S/T) residues has gained much attention [2–4]. Large scale mass-spectrometry based analysis has revealed S/T phosphorylation in a number of bacteria [2]. *M. tuberculosis* shows the highest rates of phosphorylation among studied bacteria with 8 % of the proteins identified as phosphorylated [5].

One of the most interesting findings about serine/threonine phosphorylation in eukaryotes is its association with intrinsically disordered regions[6–11]. Intrinsically disordered regions lack a well-defined three-dimensional structure and are characterized by low content of hydrophobic amino acids and high net charge [12]. These characteristic physiochemical properties allow accurate prediction of disordered regions across proteomes [13]. Disordered regions are often associated with binding to multiple partners in a transient manner [14–19]. These regions may undergo disorder to order transition upon binding, with decrease in conformational entropy. This uncouples binding affinity and specificity, allowing highly specific interactions to be reversible [20,21]. It has been proposed that due to high surface accessibility and transient mode of interactions, disordered regions are ideally suited for regulation by reversible phosphorylation

[6,7,20,21]. Disordered regions are highly abundant in eukaryotes with about a third of the proteins predicted to have at least one long (> 30 residues) disordered region and about 19 % of the residues in disordered state [22]. In contrast, most bacteria have much less disorder in their proteome, with about 4% proteins with long disordered regions and about 6% disordered residues [22]. The association between disorder and serine/threonine phosphorylation has not been investigated in prokaryotes.

Among bacteria, *M. tuberculosis* has one of the highest disorder content, with about 10 % disordered residues [22]. Thus, M. tuberculosis has both high level of S/T phosphorylation and disorder content. This prompted me to ask whether the association seen between S/T phosphorylation and disorder in eukaryotes might be present in *M. tuberculosis*, which indeed was found to be the case.

Materials and Methods

Data on *M. tuberculosis* phosphoproteins and phosphosites was obtained from Prisic *et al.* [5]. This study identified 301 phosphoproteins and 500 S/T phosphorylation events. For 215 of these sites, the specific residue that was modified could be identified with high confidence. *M. tuberculosis* proteome was obtained from Tuberculist database [23]. For disorder prediction I utilized IUPred program [24]. IUPred method is based on the observation that disordered regions do not form sufficient favorable interactions to fold, and thus have high estimated energy content [25]. I also utilized ESpritz program [26], which is conceptually different from IUPred. This method is machine learning based predictor which was trained on experimentally characterized disordered regions (missing regions in X-ray structures in PDB). I also used MFDp2 disorder

prediction tool, which is an ensemble disorder prediction tool [27,28]. Secondary structure prediction was done at Network Protein Sequence Analysis (NPSA) server [29] using consensus approach [30]. To calculate statistical significance of enrichment of pS/T sites in disordered and loop regions, I also used randomization test. For example, 84 pS/T sites are observed as disordered by IUPred program out of total 215 pS/T. For testing enrichment, I randomly selected equal number of S/T sites from each phosphoprotein as observed (1000 randomizations), and counted number of randomizations in which 84 or more sites were observed in disordered regions. This was not observed in any of 1000 randomization; hence *p*-value is less than 1e-3. To analyze conservation of S/T sites, seven mycobacterium genus were chosen which are least similar to *M. tuberculosis* with respect to BLAST sequence similarity score (*Mycobacterium sp. MCS*, *M. vanbaalenii*, *M. smegmatis*, *M. chubuense*, *Mycobacterium sp. JDM601*, *M. gilvum* and *M. abscessus*). Orthologs of *M. tuberculosis* in mycobacteria were identified using reciprocal best blast approach [31], and aligned using Clustal Omega [32]. Alignment positions with gaps were excluded from the analyses. Out of 215 sites, 135 sites were present in proteins with orthologs in all 7 species. Of these 135, 103 sites were without gaps using Clustal Omega. Positions with replacement of serine with threonine and vice-versa were considered as conserved. Phosphosite data for other bacteria was obtained from respective publications [33–40].

Results

Mass-spectrometry based analysis had previously revealed 301 phosphoproteins in *M. tuberculosis* with 500 S/T phosphorylation sites [5]. First, I tested whether phosphoproteins in *M. tuberculosis* are more likely to be disordered (i.e. have long (>=30 residues) disordered

regions). I utilized IUPred program to predict disordered regions at the proteome wide level [25]. Phosphoproteins are about twice as likely to be disordered compared to non-phosphoproteins (29.6 % vs. 13.4% respectively, Fisher *p* 4e-12). Since longer proteins are also more likely to have long disordered regions, I tested whether phosphoproteins have higher percentage of disordered residues. Phosphoproteins also have higher percentage of disordered residues compared to non-phosphoproteins (16.7% vs. 12.0% respectively, two tailed t-test *p* 3e-5).

Out of 500 phosphorylation events detected in *M. tuberculosis*, the phosphoresidue could be identified for 215 sites with high confidence [5]. For these sites, I tested whether phosphorylated S/T (pS/T) residues are more likely to be disordered compared to non-phosphorylated S/T (npS/T) residues from the same set of proteins. Overall 39.1% of pS/T sites are disordered compared to 22.4% of npS/T sites (Fisher test *p* 6e-8 and randomization test $p < 1e-3$, Figure 1). The results are very similar when another disorder prediction method, Espiritz[26] was used (52.6% pS/T disordered compared to 27.8% npS/T sites, Fisher test *p* 8e-14, randomization test $p < 1e-3$, Figure 1). A more recent disorder predictor MFDp2 [27,28] also gives similar results (43.7% pS/T disordered compared to 19.9 % npS/T sites, Fisher test *p* 6e-15, randomization test $p < 1e-3$, Figure 1). Disordered regions are also characterized by high irregular secondary structure (coil regions). Thus, I tested whether pS/T residues are enriched in coil regions of proteins. pS/T residues are more likely to occur in predicted coil regions compared to npS/T residues (70.2% pS/T sites in coil compared to 55.7% npS/T sites in coil, Fisher test *p* 2e-5 and randomization test $p < 1e-3$, Figure 2). There is also depletion of pS/T residues in beta-sheet regions (4.2% pS/T sites in sheet compared to 11.3% npS/T sites in sheet, Fisher test *p* 4e-4 and randomization test $p < 1e-3$, Figure 2), while no significant difference was found for helix regions

(22.8% pS/T sites in helix compared to 27.1 % npS/T sites in helix, Fisher test *p* 0.2 and randomization test *p* 0.12, Figure 2).

Disordered regions evolve faster than ordered regions [41], but pS/T residues evolve slower than npS/T in eukaryotes [42,43], so I tested whether *M. tuberculosis* pS/T and npS/T differ in conservation among seven mycobacterial proteomes (see methods). Conservation of pS/T is not significantly different from npS/T residues among these mycobacterial proteins (5.0 species vs. 5.2 species respectively using, Wilcox test *p* 0.92, number pS/T sites 103). Separately analyzing disordered and ordered sites does not reveal significant difference in conservation between pS/T and npS/T sites (Wilcox p 0.85 and 0.87 respectively). Thus pS/T are not more conserved than npS/T among mycobacteria.

Prisic *et al.*[5] conducted *in vitro* phosphorylation of 13-mer synthetic peptides corresponding *to in vivo* phosphorylation sites with different purified kinases. They could find phosphorylation of about half of these peptides. I tested the whether different kinases have differential preference for predicted disordered phosphoacceptors. I find that PknA has slightly higher preference for disordered phosphoacceptors than other kinases (Figure 3); however the uncorrected Chi-square test p-value is 0.04. Other kinases show uncorrected p-values > 0.05.

Discussion

Here I show the enrichment of localized serine/threonine phosphosites in disordered regions of proteins in *M. tuberculosis.* This preference is similar to that observed in eukaryotes. Association

between protein disorder and phosphorylation may offer similar advantages as those proposed in eukaryotes, including binding to multiple partners and transient mode of interaction [6,14–21], which are a prerequisite for regulatory interactions. Thus *M. tuberculosis* and eukaryotic serine/threonine phosphorylation dependent regulation may be more similar than generally appreciated. While most bacteria have low amount of protein disorder, *M. tuberculosis* has high disorder content. I suggest that high disorder content might be the reason for association between disorder and phosphorylation in *M. tuberculosis.* This suggestion is supported by the observation that other bacteria with high disorder content (*Streptomyces coelicolor* and *Halobacterium salinarum*) also show association between disorder and S/T phosphorylation, but bacteria with low disorder content do not (Supplemental Table 1) except for *Thermus thermophilus*, where enrichment of phosphorylation in disordered regions with marginal statistical significance was observed (Supplemental Table 1). Previously, a large fraction of phosphosites in *Thermus thermophilus* were observed in loop regions [44], albeit the statistical test for enrichment was not performed.

I find that PknA has a slightly higher preference for disordered phosphoacceptor sites on synthetic 13-mer substrate peptides under *in vitro* conditions than other kinases (Figure 3). However these results should be taken with the caveat that the structure of a peptide in *in vitro* conditions might be very different from the *in vivo* structure in the context of the full protein. Thus differential preferences of different kinases towards disordered substrates under *in vivo* conditions remain an open question.

I did not find pS/T sites to be more conserved than npS/T sites among mycobacteria, this suggests that phosphorylation network in mycobacteria evolves fast and may contribute towards niche specific adaptation. Important questions for future include, whether disordered and ordered S/T phosphosites are functionally different, and whether different kinases differ in their preference for disorder in their substrates under physiological conditions. Incorporation of disorder information might also be useful to predict novel S/T phosphosites [45], as has been shown in eukaryotes [6,46]. S/T kinases and their substrates have been linked to the survival, pathogenesis and virulence of *M. tuberculosis* [3,47,48], thus these finding may facilitate understanding the basic biology this deadly pathogen.

Acknowledgements

I would like to thank Dr. Shampa Ghosh from Bionivid Technology, Dr. Rahul Modak and Dr. Avinash Sonawane from KIIT University for critical reading of the manuscript.

References

1. Albuquerque CP, Smolka MB, Payne SH, Bafna V, Eng J, Zhou H (2008) A multidimensional chromatography technology for in-depth phosphoproteome analysis. Mol Cell Proteomics 7: 1389-1396.

2. Mijakovic I, Macek B (2012) Impact of phosphoproteomics on studies of bacterial physiology. FEMS Microbiol Rev 36: 877-892.

3. Cousin C, Derouiche A, Shi L, Pagot Y, Poncet S, Mijakovic I (2013) Protein-serine/threonine/tyrosine kinases in bacterial signaling and regulation. FEMS Microbiol Lett 346: 11-19.

4. Kobir A, Shi L, Boskovic A, Grangeasse C, Franjevic D, Mijakovic I (2011) Protein phosphorylation in bacterial signal transduction. Biochim Biophys Acta 1810: 989-994.

5. Prisic S, Dankwa S, Schwartz D, Chou MF, Locasale JW, Kang CM, Bemis G, Church GM, Steen H, Husson RN (2010) Extensive phosphorylation with overlapping specificity by Mycobacterium tuberculosis serine/threonine protein kinases. Proc Natl Acad Sci U S A %20;107: 7521-7526.

6. Iakoucheva LM, Radivojac P, Brown CJ, O'Connor TR, Sikes JG, Obradovic Z, Dunker AK (2004) The importance of intrinsic disorder for protein phosphorylation. Nucleic Acids Res 32: 1037-1049.

7. Collins MO, Yu L, Campuzano I, Grant SG, Choudhary JS (2008) Phosphoproteomic analysis of the mouse brain cytosol reveals a predominance of protein phosphorylation in regions of intrinsic sequence disorder. Mol Cell Proteomics 7: 1331-1348.

8. Gnad F, de Godoy LM, Cox J, Neuhauser N, Ren S, Olsen JV, Mann M (2009) High-accuracy identification and bioinformatic analysis of in vivo protein phosphorylation sites in yeast. Proteomics 9: 4642-4652.

9. Landry CR, Levy ED, Michnick SW (2009) Weak functional constraints on phosphoproteomes. Trends Genet 25: 193-197.

10. Marchini FK, de Godoy LM, Rampazzo RC, Pavoni DP, Probst CM, Gnad F, Mann M, Krieger MA (2011) Profiling the Trypanosoma cruzi phosphoproteome. PLoS One 6: e25381.

11. Amoutzias GD, He Y, Lilley KS, Van de PY, Oliver SG (2012) Evaluation and properties of the budding yeast phosphoproteome. Mol Cell Proteomics 11: M111.

12. Uversky VN, Gillespie JR, Fink AL (2000) Why are "natively unfolded" proteins unstructured under physiologic conditions? Proteins 41: 415-427.

13. Monastyrskyy B, Kryshtafovych A, Moult J, Tramontano A, Fidelis K (2014) Assessment of protein disorder region predictions in CASP10. Proteins 82 Suppl 2:127-37. doi: 10.1002/prot.24391. Epub;%2013 Nov 22.: 127-137.

14. Singh GP, Ganapathi M, Dash D (2007) Role of intrinsic disorder in transient interactions of hub proteins. Proteins 66: 761-765.

15. Tompa P, Szasz C, Buday L (2005) Structural disorder throws new light on moonlighting. Trends Biochem Sci 30: 484-489.

16. Patil A, Nakamura H (2006) Disordered domains and high surface charge confer hubs with the ability to interact with multiple proteins in interaction networks. FEBS Lett 580: 2041-2045.

17. Dunker AK, Cortese MS, Romero P, Iakoucheva LM, Uversky VN (2005) Flexible nets. The roles of intrinsic disorder in protein interaction networks. FEBS J 272: 5129-5148.

18. Wright PE, Dyson HJ (1999) Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. J Mol Biol 293: 321-331.

19. Fink AL (2005) Natively unfolded proteins. Curr Opin Struct Biol 15: 35-41.

20. Dyson HJ, Wright PE (2005) Intrinsically unstructured proteins and their functions. Nat Rev Mol Cell Biol 6: 197-208.

21. Tompa P (2002) Intrinsically unstructured proteins. Trends Biochem Sci 27: 527-533.

22. Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT (2004) Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. J Mol Biol 337: 635-645.

23. Lew JM, Mao C, Shukla M, Warren A, Will R, Kuznetsov D, Xenarios I, Robertson BD, Gordon SV, Schnappinger D, Cole ST, Sobral B (2013) Database resources for the tuberculosis community. Tuberculosis (Edinb) 93: 12-17.

24. Dosztanyi Z, Csizmok V, Tompa P, Simon I (2005) IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. Bioinformatics 21: 3433-3434.

25. Dosztanyi Z, Csizmok V, Tompa P, Simon I (2005) The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. J Mol Biol 347: 827-839.

26. Walsh I, Martin AJ, Di DT, Tosatto SC (2012) ESpritz: accurate and fast prediction of protein disorder. Bioinformatics 28: 503-509.

27. Mizianty MJ, Peng ZL, Kurgan L (2013) MFDp2 – accurate predictor of disorder in proteins by fusion of disorder probabilities, content and profiles. Intrinsically Disordered Proteins 1:e24428.

28. Mizianty MJ, Uversky V, Kurgan L (2014) Prediction of intrinsic disorder in proteins using MFDp2. Methods Mol Biol 1137:147-62. doi: 10.1007/978-1-4939-0366-5_11.: 147-162.

29. Combet C, Blanchet C, Geourjon C, Deleage G (2000) NPS@: network protein sequence analysis. Trends Biochem Sci 25: 147-150.

30. Deleage G, Blanchet C, Geourjon C (1997) Protein structure prediction. Implications for the biologist. Biochimie 79: 681-686.

31. Wolf YI, Koonin EV (2012) A tight link between orthologs and bidirectional best hits in bacterial and archaeal genomes. Genome Biol Evol 4: 1286-1294.

32. Sievers F, Higgins DG (2014) Clustal Omega, accurate alignment of very large numbers of sequences. Methods Mol Biol 1079:105-16. doi: 10.1007/978-1-62703-646-7_6.: 105-116.

33. Yang MK, Qiao ZX, Zhang WY, Xiong Q, Zhang J, Li T, Ge F, Zhao JD (2013) Global phosphoproteomic analysis reveals diverse functions of serine/threonine/tyrosine phosphorylation in the model cyanobacterium Synechococcus sp. strain PCC 7002. J Proteome Res 12: 1909-1923.

34. Manteca A, Ye J, Sanchez J, Jensen ON (2011) Phosphoproteome analysis of Streptomyces development reveals extensive protein phosphorylation accompanying bacterial differentiation. J Proteome Res 10: 5481-5492.

35. Misra SK, Milohanic E, Ake F, Mijakovic I, Deutscher J, Monnet V, Henry C (2011) Analysis of the serine/threonine/tyrosine phosphoproteome of the pathogenic bacterium Listeria monocytogenes reveals phosphorylated proteins related to virulence. Proteomics 11: 4155-4165.

36. Parker JL, Jones AM, Serazetdinova L, Saalbach G, Bibb MJ, Naldrett MJ (2010) Analysis of the phosphoproteome of the multicellular bacterium Streptomyces coelicolor A3(2) by protein/peptide fractionation, phosphopeptide enrichment and high-accuracy mass spectrometry. Proteomics 10: 2486-2497.

37. Lin MH, Hsu TL, Lin SY, Pan YJ, Jan JT, Wang JT, Khoo KH, Wu SH (2009) Phosphoproteomics of Klebsiella pneumoniae NTUH-K2044 reveals a tight link between tyrosine phosphorylation and virulence. Mol Cell Proteomics 8: 2613-2623.

38. Aivaliotis M, Macek B, Gnad F, Reichelt P, Mann M, Oesterhelt D (2009) Ser/Thr/Tyr protein phosphorylation in the archaeon Halobacterium salinarum--a representative of the third domain of life. PLoS One 4: e4777.

39. Soufi B, Gnad F, Jensen PR, Petranovic D, Mann M, Mijakovic I, Macek B (2008) The Ser/Thr/Tyr phosphoproteome of Lactococcus lactis IL1403 reveals multiply phosphorylated proteins. Proteomics 8: 3486-3493.

40. Macek B, Mijakovic I, Olsen JV, Gnad F, Kumar C, Jensen PR, Mann M (2007) The serine/threonine/tyrosine phosphoproteome of the model bacterium Bacillus subtilis. Mol Cell Proteomics 6: 697-707.

41. Brown CJ, Takayama S, Campen AM, Vise P, Marshall TW, Oldfield CJ, Williams CJ, Dunker AK (2002) Evolutionary rate heterogeneity in proteins with long disordered regions. J Mol Evol 55: 104-110.

42. Macek B, Gnad F, Soufi B, Kumar C, Olsen JV, Mijakovic I, Mann M (2008) Phosphoproteome analysis of E. coli reveals evolutionary conservation of bacterial Ser/Thr/Tyr phosphorylation. Mol Cell Proteomics 7: 299-307.

43. Gray VE, Kumar S (2011) Rampant purifying selection conserves positions with posttranslational modifications in human proteins. Mol Biol Evol 28: 1565-1568.

44. Takahata Y, Inoue M, Kim K, Iio Y, Miyamoto M, Masui R, Ishihama Y, Kuramitsu S (2012) Close proximity of phosphorylation sites to ligand in the phosphoproteome of the extreme thermophile Thermus thermophilus HB8. Proteomics 12: 1414-1430.

45. Miller ML, Soufi B, Jers C, Blom N, Macek B, Mijakovic I (2009) NetPhosBac - a predictor for Ser/Thr phosphorylation sites in bacterial proteins. Proteomics 9: 116-125.

46. Neduva V, Linding R, Su-Angrand I, Stark A, de MF, Gibson TJ, Lewis J, Serrano L, Russell RB (2005) Systematic discovery of new recognition peptides mediating protein interaction networks. PLoS Biol 3: e405.

47. Pereira SF, Goss L, Dworkin J (2011) Eukaryote-like serine/threonine kinases and phosphatases in bacteria. Microbiol Mol Biol Rev 75: 192-212.

48. Av-Gay Y, Everett M (2000) The eukaryotic-like Ser/Thr protein kinases of Mycobacterium tuberculosis. Trends Microbiol 8: 238-244.

Figure Legends

Figure 1. Phosphorylated serine/threonine sites in *M. tuberculosis* are more likely to be disordered. Both phosphorylated and non-phosphorylated serine/threonine sites are from the same set of proteins. The disorder was predicted using IUPred, Espiritz and MFDp2 methods. The Fisher test *p* values are 6e-8, 8e-14 and 6e-15 respectively. pS/T - phosphorylated serine/threonine, npS/T- non-phosphorylated serine/threonine.

Figure 2. Association between predicted secondary structure and phosphorylation of serine/threonine sites in *M. tuberculosis*. Phosphorylated serine/threonine sites occur preferentially in coil regions, while being significantly depleted in sheet regions and show no significant difference in helix regions. The Fisher test *p* values are 2e-5, 4e-4 and 0.2 respectively. pS/T -phosphorylated serine/threonine, npS/T- non-phosphorylated serine/threonine.

Figure 3. Percentage of disordered phosphoacceptors are shown for different kinases. The phosphorylation was carried out by purified kinases under in vitro conditions on synthetic 13-mer peptides corresponding to in vivo phosphorylation sites (Prisic et al.). Uncorrected Chi-square p values were significant (<0.05) only for PknA.

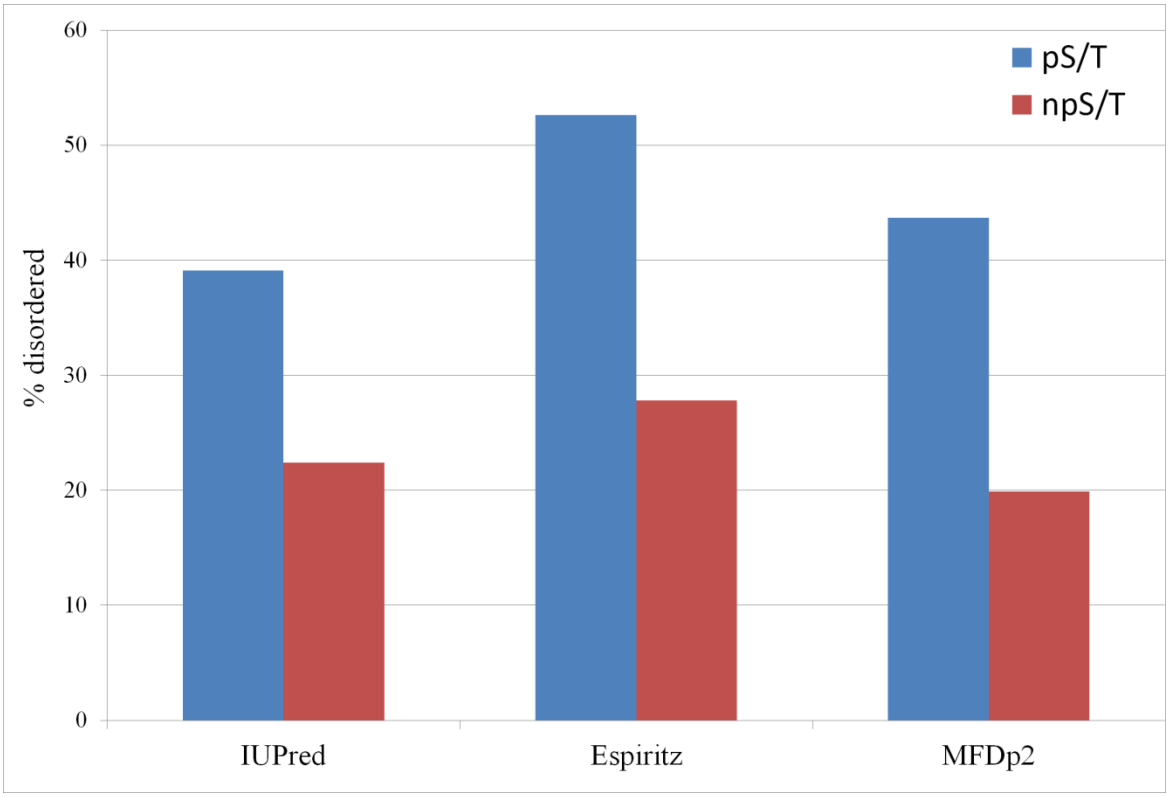

Figure 1

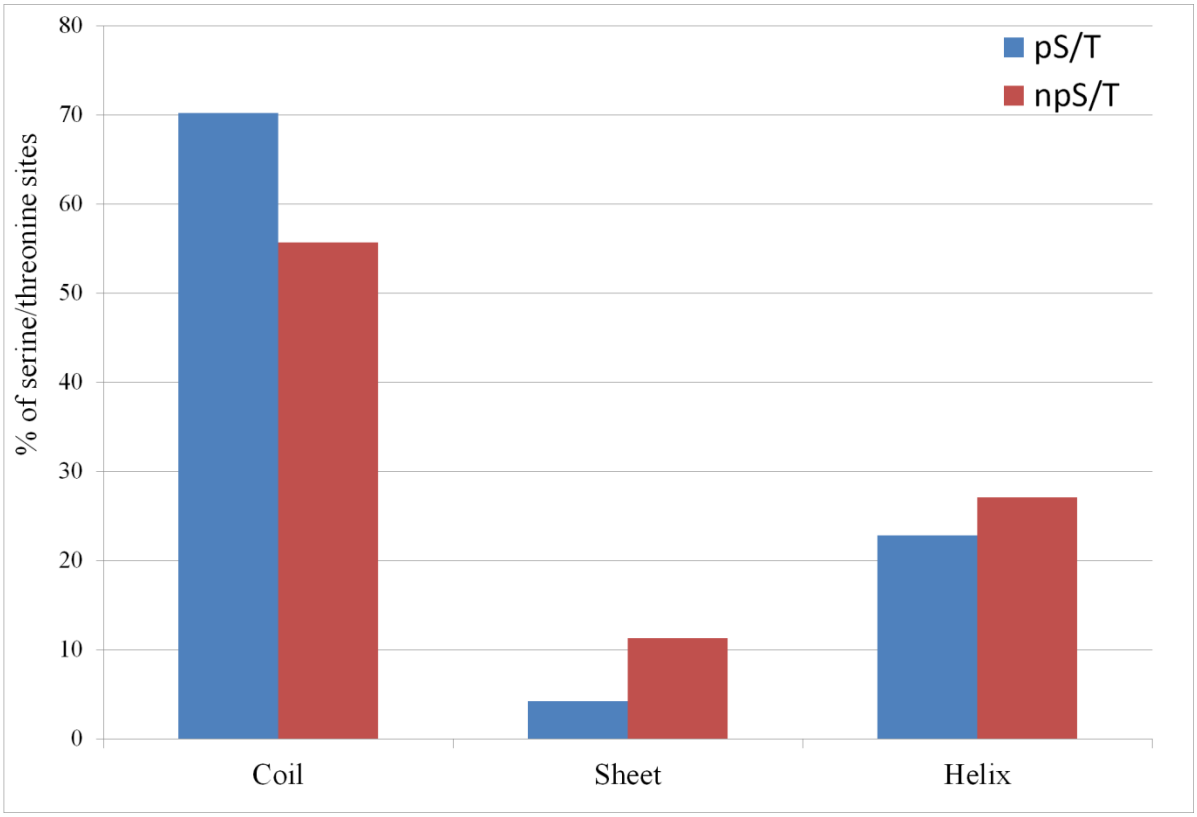

Figure 2

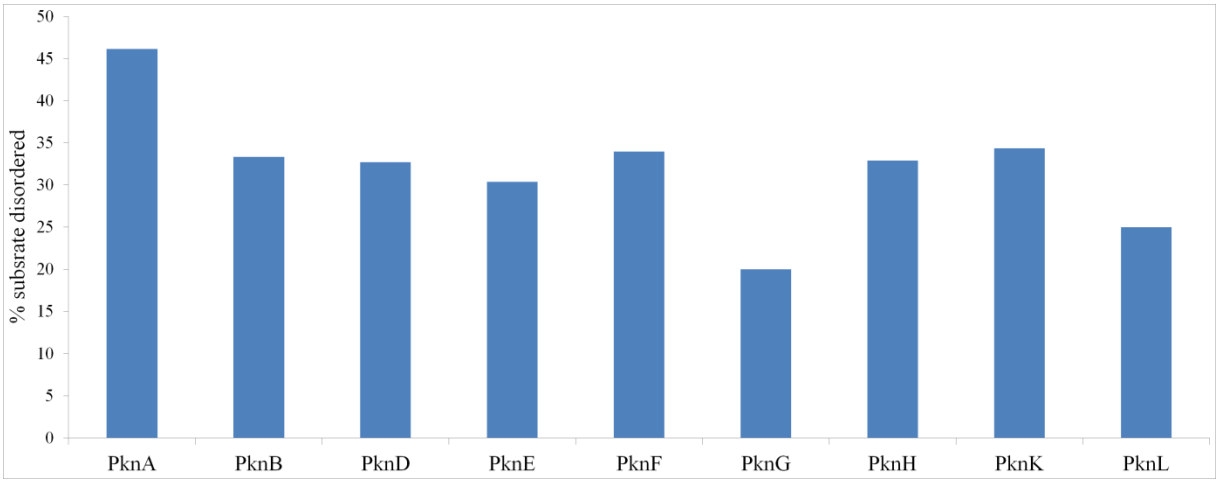

Figure 3