ISSN: 2277-9655
Impact Factor: 4.116

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

SERINE PHOSPHORYLATION EFFECT ON SECONDARY STRUCTURE PREDICTION OF INTRINSICALLY UNSTRUCTURED AND ORDERED STRUCTURED PENTAPEPTIDES BY RAMACHANDRAN ANALYSIS

Simon M. Longela*, Håvard J. Haugen, Jiri Vymetal and Jiri Vondrášek
*Biomaterials Department, IKO, University of Oslo, Norway *
Biomaterials Department, IKO, University of Oslo, Norway
Institute of Organic Chemistry and Biochemistry Academy of Science of the Czech Republic, Prague
Institute of Organic Chemistry and Biochemistry Academy of Science of the Czech Republic, Prague

DOI: 10.5281/zenodo.57943

ABSTRACT

Many proteins associated with cell signalling pathways are often targets of post-translational modifications such as phosphorylation, glycosylation, ubiquitination, nitrosylation, methylation, acetylation, lipidation and proteolysis. Such modifications can lead to the induction or disruption of secondary structural elements of the modified protein.
This paper describes the structural behaviour of a set of intrinsically disordered peptides (IDP) and ordered structured peptides (OP) modified by phosphorylation on a serine residue. These pentapeptides (IDP and OP) derived from fragmentation of proteins from different organisms.
Ramachandran analysis revealed that both, IDP and OP, have an overall propensity for α-helical structure that is greatest in IDP and non-existent in OP cluster conformations. Phosphorylation of OP caused a decrease in the helical propensity in the serine phosphorylated residues, whereas for IDP with high disorder tendency, the opposite was observed and phosphorylation engendered alpha left-handed helical propensity.

KEYWORDS: Phosphorylation, Disorder, α-Left-helix, Ramachandran plot, Simulation

INTRODUCTION

Phosphorylation, the addition of a phosphoryl group (HPO_3) to the side chain of an amino acid, is a posttranslational modification of proteins that has been intensively studied lately [1]. Except for being unusually common, about 33% of all proteins in the cell are phosphorylated at some point during their lifecycle [2], and as this is interesting from a biological point of view, it is also medically relevant. Several studies have shown that many untreatable diseases, like cancer [3], diabetes [4], and neurodegenerative disorders [5], can be associated with abnormal phosphorylation in the affected cells. A recent study has also presented that phosphorylation might play a key role in the viral infection mechanism [6]. Thus, a good analysis tool, one specially developed for phosphorylated protein, is crucial in order to both understand the origin of and find working treatments for several of the common severe chronic diseases in today's society.
Within a protein, phosphorylation can occur on several amino acids [7]. Phosphorylation on serine is the most common, followed by threonine [1]. Tyrosine phosphorylation is relatively rare but is at the origin of protein phosphorylation signalling pathways in most of the eukaryotes. Because of the central role of phosphorylation in the regulation of life, diligent efforts have been focussed on the development of methods for characterising protein phosphorylation [8].
Biological activities have shown to exert strong influence on the cellular behaviour of IDP from both structural and functional viewpoints [9]. Sites of protein phosphorylation, for example, can be found in regions of structural disorder, as well as within regions of well-ordered structure [10]. However, the similarity in sequence complexity, amino acid composition, and flexibility parameters between protein phosphorylation sites and disordered protein regions suggests that intrinsic disorder in and close to modification sites constitutes a common recognition feature

for many eukaryotic serine kinases [11]. With regard to the variety of structural consequences of protein phosphorylation, conformational responses have been observed by Ramachandran analysis.
The aim of this study was to understand the structural difference of peptides' residues upon phosphorylation, by scrutinising secondary structures derived from molecular dynamic predictions.
From molecular dynamic predictions executed with bioinformatics' tools and algorithms, different simulations were conducted mimicking the behaviour of peptides subjected to Serine phosphorylation with the purpose to understand and visualise phosphorylation response on different conformations adopted by peptides.

MODELS AND METHODS

This paper focuses on phosphorylation of serine comprises in the middle of short pentapeptides trimmed from longer IDP and OP taken from different organism' proteomes from the Uniprot database [12].
A set of 20 proteins were randomly selected from different organism; however, computing longer proteins' sequences would require much longer sampling times, we then trimmed those longer proteins' sequences into short five amino acid long peptides containing Serine in the middle, to form pentapeptides which were later used for molecular dynamic simulations. We randomly selected 6 representative pentapeptides (3 IDP and 3 OP) to proceed further phosphorylation and simulations.
The underlined Serines in **Table 1** have been phosphorylated for the purpose of this study. Extended forms of IDP and OP were graphically built with PyMOL (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC) [13], as well as the phosphorylated molecules which were inputs for molecular dynamic simulations.

Table 1: Selected pentapetides

Peptide type	Peptide ID	Phospho-rylated ID	Trimmed Sequence	Organism	RCSB PBD ID*
IDP	D2	D2p	VGSPP	Homo sapiens	A0JNW5-1
IDP	D3	D3p	SSSPT	Homo sapiens	A0MZ66-1
IDP	D5	D5p	FNSPS	Homo sapiens	A0MZ66-8
OP	O2	O2p	EGSGR	Homo sapiens	Q16555-2
OP	O4	O4p	VYSLD	Rattus Norvegicus	P61765-2
OP	O5	O5p	TRSLK	Pongo Abeii	Q5R6D2-1

*Protein Data Bank Identity [16]

Free energy profiles were shown in Ramachandran plots were visualized with Grace 5 (Grace Development Team, Portland, OR) [14].
The computing facility for the molecular dynamic simulations were performed on High Performance Cluster for Biological application based on Stallo [15], a Norwegian computer cluster with HP BL 460c Gen 8 blade servers with configuration shown in **Table 2**.

Table 2: High Performance Computer Hardware [15]

Name	System	Type	Number of nodes	Number of cores	CPU type	Theoretical total peak	Total memory	Total disk capacity
Stallo	HP BL460c Gen8	Cluster	518	9132	Intel E5-2670	101 Tflop	12800 GB	2.1 PB

Molecular Dynamic Simulation Setup

Molecular dynamic simulations were performed by GROMACS 4.5.5 package under modified AMBER force-field parameters for phosphorylated amino acids in different protonation states [16],with polarized water model TIP3P [17]. Each peptide was then placed in a neutral periodic cubical box filled with water molecules [17]. After the relaxation of the systems using energy minimisation, simulations were initiated by randomly distributing initial velocities of atoms according to a Maxwell-Boltzmann distribution.
The initial distribution of velocities was determined from a random distribution with the magnitudes conforming to

the required temperature and corrected so there is no overall momentum (Eq. 1), i.e.,

$$P = \sum_{i=1}^{N} m_i v_i = 0 \quad (1)$$

The velocities vi are often chosen randomly from a Maxwell-Boltzmann at at 299 K, which gives the probability (Eq. 2) that an atom i has a velocity vx in the x direction at a temperature T.

$$P(v_i) = \left(\frac{m_i}{2\pi k_B T}\right)^{1/2} \exp\left[-\frac{1}{2}\frac{m_i v_{ix}^2}{k_B T}\right] \quad (2)$$

The temperature (Eq. 3) was computed calculated from the velocities using the relation

$$T = \frac{1}{(3N)} \sum_{i=1}^{N} \frac{|P_i|}{2m_i} \quad (3)$$

where N is the number of atoms in the system.

A gradual positional restraint procedure was used to avoid distortion of peptides. Each system was subjected to two 100ps consecutive equilibration runs in which all heavy peptide atoms were restrained to their starting positions (using a force constant of 1,000 and 100 kJ mol^{-1} nm^{-2}, respectively) whereas water was left free to appropriately form solvation layers around the solute. Next, a 100ps simulation at 300K with weaker restraints (using a force constant of 50 kJ mol^{-1} nm^{-2}) was performed. Finally, production runs without positional restraints were performed with configurations stored every 100 ns for analysis. In total, six simulations were performed. The protonation states of all protonated groups were set as appropriate at neutral pH.

Ramachandran Diagrams

The two torsion angles of the polypeptide chain, also called Ramachandran angles describe the rotations of the polypeptide backbone around the bonds between N-Cα (called Phi, φ) and Cα-C (called Psi, ψ). The Ramachandran plot provides a simple manner or viewing the distribution of torsion angles of a protein structure [18-20]. Ramachandran plots graphed over time were useful in identifying mobile residues in the secondary structure. It was possible to look at the complete

Ramachandran plot for all residues at certain time points, and it was possible to plot the phi/psi angles for the phosphorylated Serine over time are shown on **Figures** 1 and 2. It further provided an overview of allowed and disallowed regions (Beta sheet, αRight handed-helical and αLeft handed-helical) of torsion angle values, serving as an important indicator of the quality of protein secondary structures [21].

Disorder prediction for IDP

An investigation on the disorder propensity in IDP revealed to be important to understand their intrinsic behaviours and conformation preferences. Protein disorder and conformational variations are intrinsically related to protein flexibility [22].

Figure 3 shows disorder prediction scoring of our selected disordered pentapetides. Recent studies have established the correlation between protein flexibility and its intrinsic disorder [23].

These results helped us to later discuss the influence of disorder on conformational preferences of phosphorylated peptides.

RESULTS AND DISCUSSION

After an initial visual inspection of the trajectory, additional thorough checks regarding quality of the simulations were performed. Energy minimisation assesses whether or not equilibrium was reached, involving tests for the convergence of thermodynamic parameters such as the temperature, the pressure, the potential and the kinetic energy. Structural analyses were performed by analysing Ramachandran plot calculations.

Various Ramachandran plots were retrieved from trajectories of molecular dynamic simulations at different times in order to follow the evolution of serine-phosphorylated: first at 100 ns, which marks the end of the simulation, a second time at 50 ns (half time of the simulation), and finally at 100 ns for all amino acid residues.

After plotting Ramachandran of four different stages of the simulation, we specifically analysed Rama SEP-4 (phosphorylated Serine plots for the entire simulation duration 100 ns) and Rama SEP-4 tail (phospho-serine conformation preferences at the final stage of the simulation) which show phosphorylated serine conformational preferences of each peptide.

Our simulations revealed large structural differences between the phosphorylated ordered and disordered peptides. We managed to plot the following Ramachandran maps for each peptide:

ISSN: 2277-9655
Impact Factor: 4.116

- **Rama half time** which represents the half time simulation Ramachandran of all residues (after 50 ns); obtained with the following command:

g_rama –f traj.xtc –s topol.trr –o Rama_halftime.xvg b 50000

- **Rama full time** which represents the Ramachandran plot for all residues after 100 ns, obtained with the following command:

g_rama –f traj.xtc –s topol.tpr –o Rama.xvg

- **Rama SEP-4** which represents phosphorylated Serine plots for the entire simulation duration (100 ns); obtained with the following command:

grep "\(ˆ@\|SEP-4\)" rama.xvg > Rama_SER4.xvg

- **Rama SEP-4 tail** which represents phosphorylated Serine conformation preferences at the final stage of the simulation; obtained with the following command:

grep"\(ˆ@\|SEP-4\)"rama.xvg|tail –n 50000 >> Rama_SEP4_tail.xvg

After obtaining all simulations steps, we analysed and realised that:

Case 1: **Rama halftime** (**Figures 1.A, 1.B, 1.C, 2.A, 2.B and 2.C**)

- *Beta sheets*: higher density on both IDP and ODs;
- *α-Right handed-helix*: high density on both IDP and ODs;
- *α-Left handed-helix*: high density on IDP but lower density or inexistent ODs.

Case 2: **Rama full time** (**Figures 1.D, 1.E, 1.F, 2.D, 2.E and 2.F**)

- *Beta sheets*: high density on both IDP and ODs;
- *α-Right handed helix*: high density on both IDP and ODs;
- *α-Left handed helix*: high density on IDP but lower density on ODs.

Case 3: **Rama SEP-4** (**Figures 1.G, 1.H, 1.I, 2.G, 2.H and 2.I**)

- *Beta sheets*: higher density on both IDP and ODs;
- *α-Right handed-helix*: low density on IDP but higher ODs;
- *α-Left handed-helix*: high density on IDP (D2 and D3) but inexistent ODs and one IDP (D5).

Case 4: **Rama SEP-4 tail** (**Figures 1.J, 1.K, 1.L, 2.J, 2.K and 2.L**)

- *Beta sheets*: higher density on both IDP and ODs;
- *α-Right handed-helix*: low density on IDP but higher ODs;
- *α-Left handed-helix*: high density on IDP (D2 and D3) but inexistent ODs and one IDP (D5)

In cases 1 and 2, peptides adopted similar conformations. IDP and OP strongly populated the Beta sheets region (upper left quadrant). Almost the same has been observed on αRight-handed helical regions for both IDP and OP; this region was relatively populated.

By isolating phosphor-Serine on cases 3 and 4, we realised a drastic preference change on α-Left handed helical region, where IDP in general would prefer to populate, while all ODs would not prefer (**Figure 4, A** and **B**).

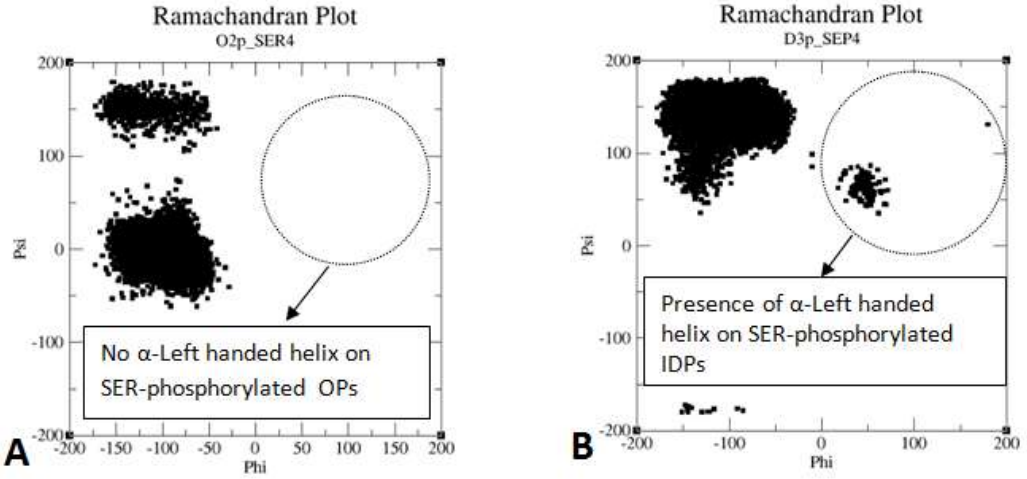

Figure 4. OP presents no occupancy of α-Left helix region (A). IDP D3 presents α-Left helix (B)

ISSN: 2277-9655
Impact Factor: 4.116

αLeft- handed helical region (around φ, ψ: 50°, 50°) was not populated by any individual phosphorylated Serine of OP but was populated by D2 and D4 IDP. However, for IDP, due to their intrinsic disorderliness, as they lack tertiary structure, they can adopt random multiple conformations as time evolves including the αLeft-helical region. Interestingly, this has been observed on IDP pentapeptides showing higher disorderliness on IUPred predictor while less disordered IDP tend to show little preference in α-Left helical regions. This is due to the fact that lower disordered pentapeptides appear to behave like a structured peptide (OP). IDP disorder tendency predicted by IUPRED [24] is shown in **Figure 3** where D5 presents a quite low disorder tendency (39.83%), affecting its phosphorylated serine Ramachandran plot as D5 behaved like an ordinary structured peptide.

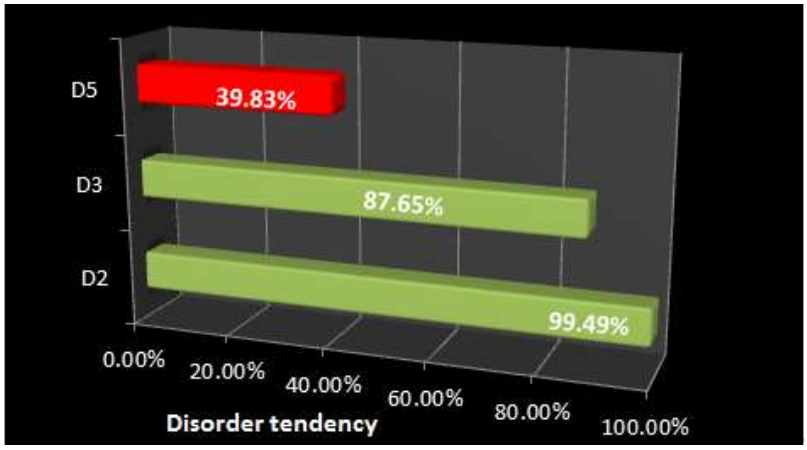

Figure 3. Disorder tendency of IDP

The key conclusion of this study is that our dynamic molecular method appears to be capable of reproducing conformational changes induced by post-translational phosphorylation, with near-atomic resolution in most cases considered here, which are limited to relatively modest conformational changes and not, e.g., more drastic order–disorder transitions. This work thus represents a significant step toward a broadly applicable method for predicting structural effects of phosphorylation. Through case studies 3 and 4, we also illustrated that computational methods can be used to provide new understanding of how phosphorylation drives conformational change on ODs while this process helps IDP with significant disorder to maintain flexibility of adopting multiple conformations. This behaviour has been observed by electrospray liquid chromatography (LC) ion trap (LCQ) system in a previous in vitro study [25]. And it has been explained that as the phosphate P–O bond is much weaker than the peptide bond. It has been hypothesized that the charged amino-acid residues interact with the phosphate group to stabilize it (perhaps in the form of a salt bridge [26]), resulting in a different chemical structure from the neutral peptide and the peptide ions of lower charge states. The long timescale in the ion trap may facilitate this process. As for phosphoserine-containing peptides, the phosphate is so unstable that the interaction with the charged group is insufficient to stabilize it. Another in vitro study with circular dichroism method, in agreement with our computational finding, showed that upon using of trifluoroethanol as an indicator of the alpha-helix, the stability of the alpha-helix conformation was enhanced by phosphorylation [27].

ISSN: 2277-9655
Impact Factor: 4.116

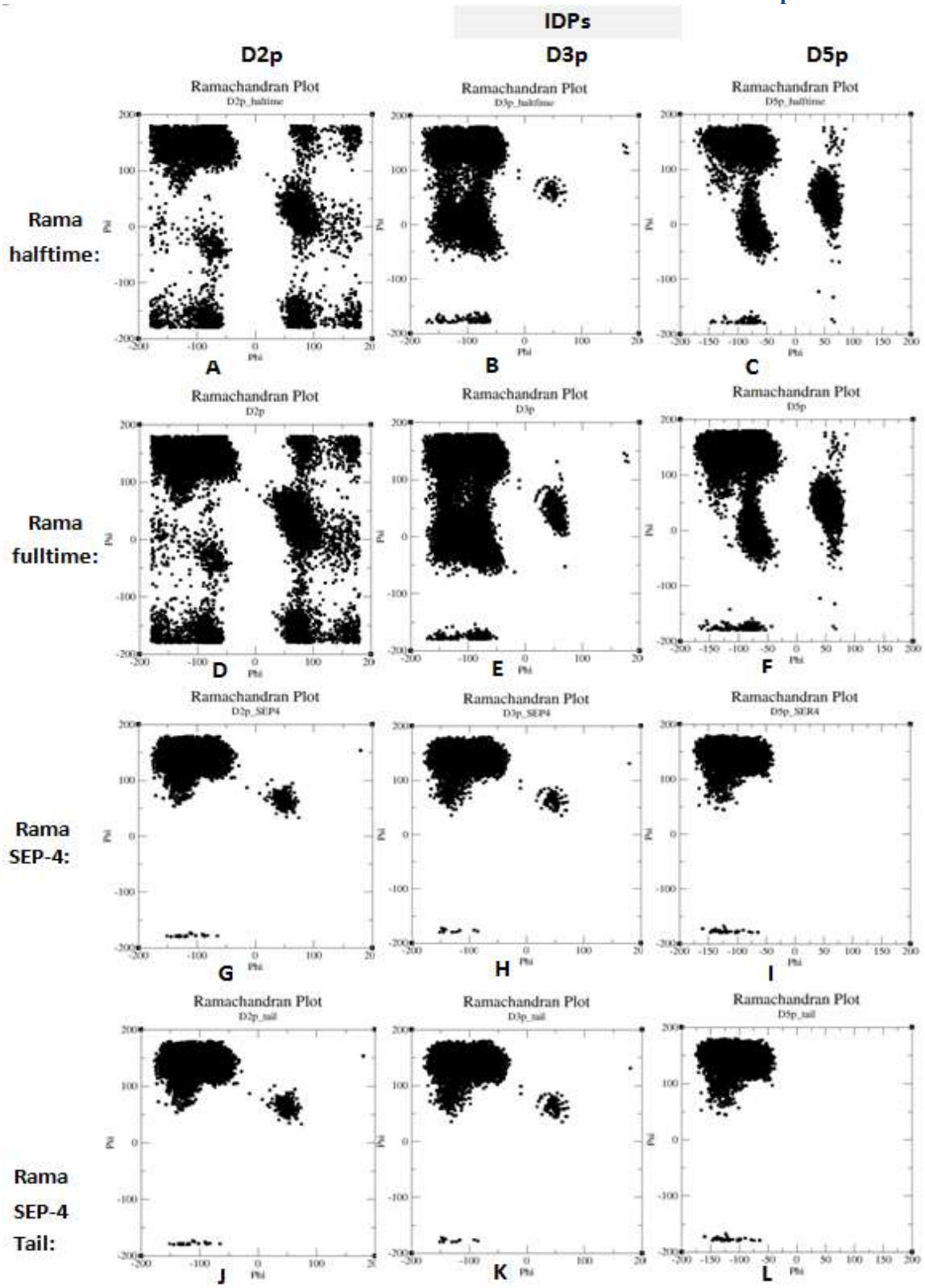

Figure 1. Ramachandran plots results for IDP

ISSN: 2277-9655
Impact Factor: 4.116

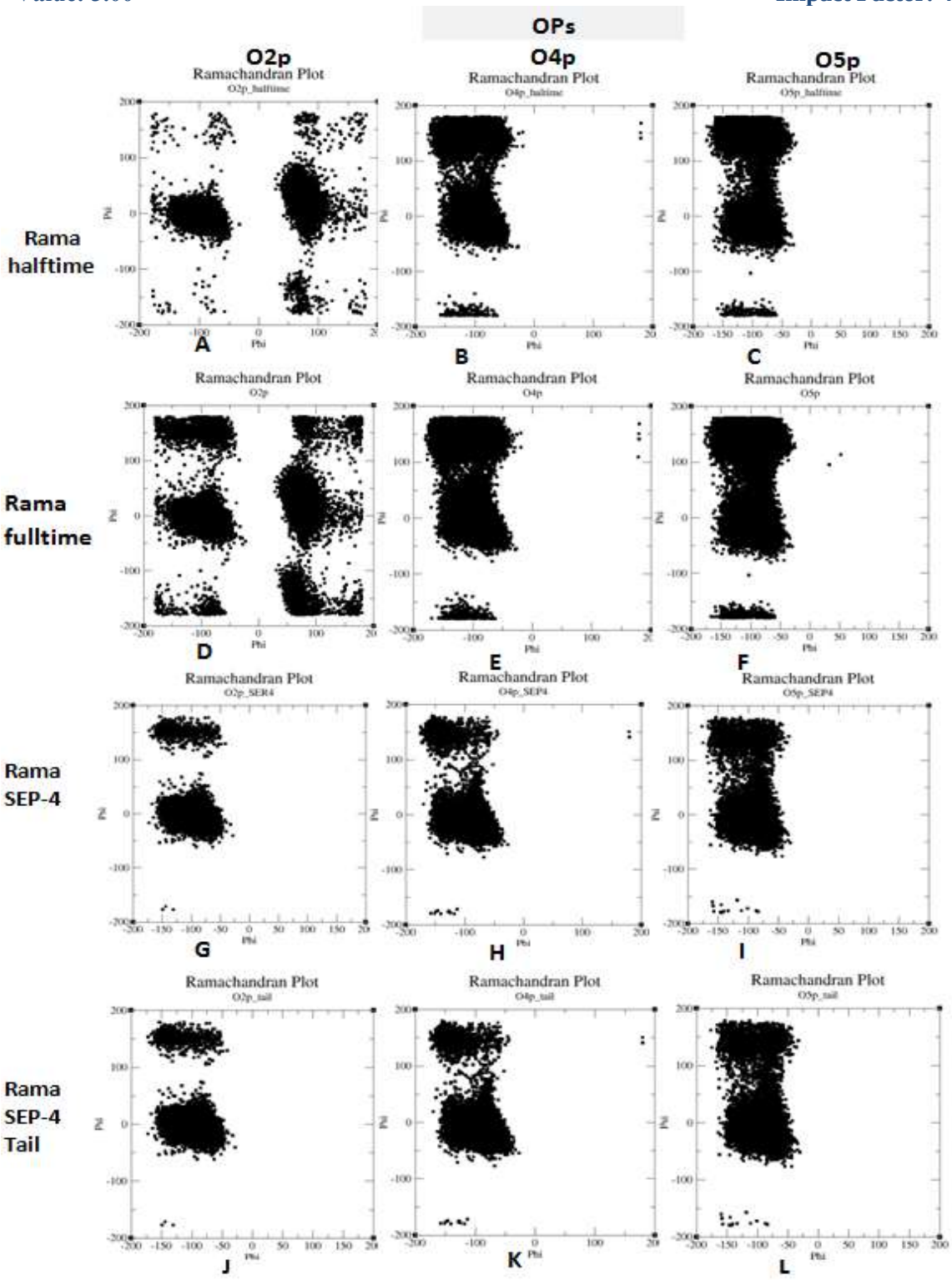

Figure 2. Ramachandran plots results for OP

CONCLUSION

Evolutionary understanding is indispensable to comprehend life as a biological system. Taken together, we have analysed the evolutionary aspects of eukaryotic interactome and gleaned some significant findings associating network features and peptide evolution.

We analysed Ramachandran free energy profiles of serine residues with and without phosphorylation. Computing longer proteins sequences would require much longer sampling times, however, by simulating short peptides, one can minimize the effects of tertiary interactions and focus on analysing secondary ones and we can then isolate a portion of a sequence to understand the specific conformation changes.

Our study indicates that for the phosphorylated IDP cases, phosphorylation stabilises α-Left handed helix conformations for peptides showing high disorder tendency.

We also observed that disorder tendency plays a critical role in IDPs behaviour. High disorderliness of IDP favours αL-helical conformation contrary to lower disordered IDPs which tend to not adapt αL-helical conformation behaving like OP.

Finally, our study confirmed that phosphorylated serine IDPs show the conformation flexibility of the backbone as they are able to populate all allowed regions of the Ramachandran diagram.

Recommendations

Future works can focus on similar molecular dynamic simulations investigations on longer phosphorylated proteins to access their conformational preferences on phosphorylated sites.

This paper opens a path to future works which may investigate on the same method used in this paper, but specially by checking the behaviour of OP and IDP by the addition of a co-solvent. 2-2-2 tri-fluoro-ethanol or hexafluoro-2-propanol (HFIP) as co-solvent can promote alpha helix conformation within proteins. Other helix promoters might be investigated to determine whether phosphorylation with co-solvents can lead to build future highly flexible multi cell functions proteins.

ACKNOWLEDGEMENTS

We acknowledge the support of the Norwegian Metacenter for Computational Science (NORTUR) which provided computational resources and services. This work has been funded by the Norwegian Research Council under grant code 232424.

REFERENCES

1. Eriksson AI. Enrichment and Separation of Phosphorylated Peptides on Titanium Dioxide Surfaces: Applied and Fundamental Studies. Acta Universitatis Upsaliensis Digital Comprehensive Summaries of Uppsala. 2013: 1059. 52 pp.
2. Delom F, Chevet E. Phosphoprotein analysis: from proteins to proteomes. Proteome Sci. 2006;4 (1):15.
3. Chong PK, Lee HY, Kong JWF, Loh MCS, Wong CH, Lim YP. Phosphoproteomics, oncogenic signaling and cancer research. Proteomics. 2008;8:4370-82.
4. Hancock JT. Cell signalling. 3rd ed. Oxford, UK: Oxford University Press; 2010.
5. Bahk YY, Ahsan-Ul-Bari M, Kim YJ. Biomedical application of phosphoproteomics in neurodegenerative diseases. Journal of microbiology and biotechnology. 2013;23:279-88.
6. Bergstrom Lind S, Artemenko KA, Elfineh L, Zhao Y, Bergquist J, Pettersson U. The phosphoproteome of the adenovirus type 2 virion. Virology. 2012;433:253-61.
7. Ubersax JA, Ferrell Jr JE. Mechanisms of specificity in protein phosphorylation. Nat Rev Mol Cell Biol. 2007;8:530-41.
8. Hanks SK, Hunter T. Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. FASEB J. 1995;9:576-96.
9. Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, et al. Intrinsically disordered protein. J Mol Graph Model. 2001;19:26-59.
10. Johnson LN, Lewis RJ. Structural basis for control by phosphorylation. Chem Rev. 2001;101:2209-42.
11. Iakoucheva LM, Radivojac P, Brown CJ, O'Connor TR, Sikes JG, Obradovic Z, et al. The importance of intrinsic disorder for protein phosphorylation. Nucleic acids research. 2004;32:1037-49.
12. UniProt C. Activities at the Universal Protein Resource (UniProt). Nucleic acids research. 2014;42:D191-8.

13. DeLano WL. Pymol: An open-source molecular graphics tool. CCP4 Newsletter On Protein Crystallography. 2002;40. San Carlos. CA. USA.
14. Turner P, Stambulchik E. Grace: A 2d plotting tool. In: Team GD, editor. Weizmann Institute of Science. Rehovot . Israel: Grace Development Team; 1998.
15. AS US. PORTALS. Fishing and the interplay. 2013; Notur II project – The Norwegian metacenter for computational science. Trondheim, Norway:26 p.
16. Homeyer N, Horn AHC, Lanig H, Sticht H. AMBER force-field parameters for phosphorylated amino acids in different protonation states: phosphoserine, phosphothreonine, phosphotyrosine, and phosphohistidine. Journal of molecular modeling. 2006;12:281-9.
17. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of Simple Potential Functions for Simulating Liquid Water. J Chem Phys. 1983;79:926-35.
18. Ramachandran GN, Ramakrishnan C, Sasisekharan V. Stereochemistry of polypeptide chain configurations. Journal of molecular biology. 1963;7:95-9.
19. Vymětal J, Vondrášek J. The DF-LCCSD(T0) correction of the φ/ψ force field dihedral parameters significantly influences the free energy profile of the alanine dipeptide. Chem Phys Lett. 2011;503:301-4.
20. Ting D, Wang GL, Shapovalov M, Mitra R, Jordan MI, Dunbrack RL. Neighbor-Dependent Ramachandran Probability Distributions of Amino Acids Developed from a Hierarchical Dirichlet Process Model. Plos Comput Biol. 2010;6.
21. Hovmoller S, Zhou T, Ohlson T. Conformations of amino acids in proteins. Acta Crystallogr D Biol Crystallogr. 2002;58:768-76.
22. Gu J, Hilser V. The significance and impacts of protein disorder and conformational variants. J Gu and PE Bourne, John Wiley & Sons, Inc, New Jersey. 2009:939-62.
23. Janin J, Sternberg MJ. Protein flexibility, not disorder, is intrinsic to molecular recognition. F1000 biology reports. 2013;5.
24. Dosztanyi Z, Csizmok V, Tompa P, Simon I. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. Bioinformatics. 2005;21:3433-4.
25. DeGnore J, Qin J. Fragmentation of phosphopeptides in an ion trap mass spectrometer. Journal of the American Society for Mass Spectrometry. 1998;9:1175-88.
26. Lee S-W, Kim HS, Beauchamp J. Salt bridge chemistry applied to gas-phase peptide sequencing: selective fragmentation of sodiated gas-phase peptide ions adjacent to aspartic acid residues. Journal of the American Chemical Society. 1998;120:3188-95.
27. Mendieta J, Fuertes MA, Kunjishapatham R, Santa-María I, Moreno FJ, Alonso C, et al. Phosphorylation modulates the alpha-helical structure and polymerization of a peptide from the third tau microtubule-binding repeat. Biochimica et Biophysica Acta (BBA)-General Subjects. 2005;1721:16-26.