Charles University

Faculty of Science

Study programme: Chemistry

Branch of study: Modelling of Chemical Properties of Nano- and Biostructures

Ing. Veronika Jurásková

Influence of phosphorylation on the conformation of peptides and proteins

Vliv fosforylace na konformaci peptidů a proteinů

Diploma thesis

Supervisor: RNDr. Jiří Vondrášek, CSc.

Prague 2017

Prohlašuji, že jsem tuto diplomovou práci vypracoval(a) samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona v platném znění, zejména skutečnost, že Univerzita Karlova má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle §60 odst. 1 autorského zákona.

V ..... dne ..... Podpis autora

Acknowledgement

At this place I would like to express my gratitude to the supervisor of this thesis RNDr. Jiří Vondrášek, CSc. for his advice and help during my work. I would further like to thank Mgr. Jiří Vymětal, Ph.D. for practical guidance and for providing the scripts for metadynamics. I really appreciate his thorough reading and commenting on the manuscript of this thesis as well as his answers on the number of my, sometimes naive, questions. My thanks belong also to all members of the Bioinformatics group for creating a nice working environment. Last but not least, I would like to thank my boyfriend Daniel for all his patience and support during my studies.

Access to computing and storage facilities owned by parties and projects contributing to the National Grid Infrastructure MetaCentrum provided under the programme "Projects of Large Research, Development, and Innovations Infrastructures" (CESNET LM2015042), is greatly appreciated.

Title: Influence of phosphorylation on the conformation of peptides and proteins

Author: Ing. Veronika Jurásková

Department: Department of Physical and Macromolecular Chemistry

Supervisor: RNDr. Jiří Vondrášek, CSc.

Consultant: Mgr. Jiří Vymětal, PhD.

Abstract: The aim of the thesis was to study the effect of phosphorylation on the conformation of peptides and proteins. I focused on three amino acids which are phosphorylated the most – serine, threonine and tyrosine. I studied the conformational changes upon phosphorylation in dipeptides and pentapeptides by the metadynamics technique. I found out that the phosphorylation of amino acid residues led to the conformational changes characteristic for each amino acid. Whereas the phosphorylation of serine increased the preference of right-handed alpha helix conformation, the phosphorylation of threonine led to the extended structure and the conformation of tyrosine was not influenced by phosphorylation at all. Using classical molecular dynamics, I also studied the conformational changes in longer peptides derived from the phosphorylation sites of disordered proteins. Moreover, I simulated phosphorylated and unphosphorylated variants of three proteins with a known structure from the RCSB PDB database. I compared the calculated results with available experimental and computational studies.

Keywords: phosphorylation, dipeptides, pentapeptides, metadyamics, molecular dynamics, post-translational modifications

Název: Vliv fosforylace na konformaci peptidů a proteinů

Autor: Ing. Veronika Jurásková

Katedra: Katedra fyzikální a makromolekulární chemie

Vedoucí práce: RNDr. Jiří Vondrášek, CSc.

Konzultant: Mgr. Jiří Vymětal, PhD.

Abstrakt: V předkládané práci jsem se zabývala studiem vlivu fosforylace na konformaci peptidů a proteinů. Zaměřila jsem se na tři nejčastěji fosforylované aminokyseliny - serin, threonin a tyrosin. Pomocí metadynamiky jsem studovala změnu konformace vlivem fosforylace u jednoduchých dipeptidů a také složitějších pentapeptidů. Zjistila jsem, že fosforylace jednotlivých aminokyselin vede ke strukturním změnám, které jsou charakteristické pro každou aminokyselinu. Zatímco fosforylace serinu zvyšuje preferenci pravotočivého alfa helixu, fosforylace threoninu vede k extendované struktuře a konformace tyrosinu není fosforylací ovlivněna. Dále jsem studovala změnu konformace serinu v delších peptidech získaných z fosforylačních míst nestrukturovaných proteinů pomocí klasické dynamiky. Pokusila jsem simulovat také fosforylované a nefosforylované varianty proteinů se známou strukturou publikovanou v RCSB PDB databázi. Získané výsledky jsem poté porovnávala s dostupnými výpočetními a experimentálními daty.

Klíčová slova: fosforylace, dipeptidy, pentapeptidy, metadynamika, molekulová dynamika, posttranslační modifikace

Contents

List of Abbreviations

α_L	Left-handed alpha helix
α_R, $\alpha_{R'}$	Right-handed alpha helix
Ace	Acetyl
Arg	Arginine
ATP	Adenosine triphosphate
CD	Circular Dichroism
CV, CVs	Collective Variable
Ext	Extended structure
FES	Free Energy Surface
IR	Infrared
LINCS	Linear Constraint Solver
MD	Molecular Dynamics
Nme	N-methyl
NMR	Nuclear magnetic resonance
NpT	Isothermal-isobaric ensemble
NVE	Microcanonic ensemble
PPII	Polyproline II
Pro	Proline
Ptr	Phosphotyrosine
RMSD	Root-Mean-Square Deviation
RMSF	Fluctuations of RMSD
Ser	Serine
Sep	Phosphoserine
Thr	Threonine
TIP3P	Transferable Intermolecular Potential with 3-points
Tpo	Phosphothreonine
Tyr	Tyrosine
VCD	Vibrational Circular Dichroism

Chapter 1

Introduction

1.1 Protein phosphorylation

Phosphorylation of proteins belongs to the most important post-translational modifications in both prokaryotes and eukaryotes. During phosphorylation reaction, the phosphoryl group $(PO_3)^{2-}$ is attached to an amino acid residue. This reaction is generally catalyzed by kinase enzymes, which belong to transferases enzyme class. Kinases transfer the phosphoryl group from phosphoryl donating molecules (e.g. adenosine triphosphate, ATP) to the substrates (see Fig. 1.1). The reverse reaction when phosphoryl group is removed from the amino acid is called dephosphorylation and it is catalyzed by phosphatases from hydrolases enzyme class.[21]

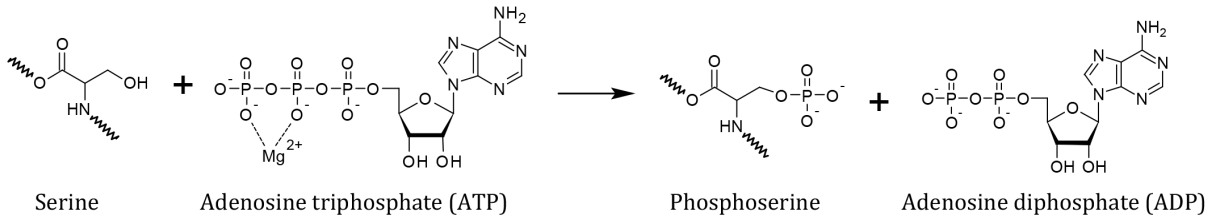

Fig. 1.1: Phosphorylation of serine: Kinase catalyses reaction of serine with ATP. Product of the reaction is phosphoserine and adenosine diphosphate (ADP).

Phosphorylation and dephosphorylation regulates enzyme activity, marks proteins for destruction and influences the stability and protein-protein interactions. It thus plays important role in many different metabolic pathways.[21] About 30 % of cellular proteins contain a phosphate group. The abnormal protein phosphorylation was even found to play role in some diseases such as Alzheimer's disease or cancer.[20]

The usual amino acids which undergo the phosphorylation are serine, threonine and tyrosine. These amino acids contain hydroxy group which forms stable phosphomo-

noester during the phosphorylation. Nevertheless, phosphorylation occurs also on the arginine, lysine or histidine forming phosphoramidates by reaction of phosphoryl group with an amino group. The other residues which can be the substrate of phosphorylation are cysteine (creating thiophosphate) and aspartate (acylphosphate).[19]

The phosphate group is negatively charged at physiological pH. It strongly interacts with surrounding water molecules, positively charged amino acids and other charged molecules. Therefore, phosphoryl group can induce conformational changes in peptides and proteins. Generally, there are two possible mechanisms of protein structure modification. In the first case, phosphorylation changes the intrinsic propensity at the level of secondary structure and influences the backbone conformation. The other mechanisms involves the conformational changes in the tertiary structure caused by the charge of the phosphoryl group.[27]

In the following sections, experimental and theoretical studies on the influence of the phosphorylation on the conformation of small peptides are presented. The last section of this chapter defines the aim of the thesis.

1.2 Theoretical studies of peptide phosphorylation

Computational studies on the effect of the phosphorylation are often focused on the changes in the local backbone conformation. These changes are caused by the interaction with the solvent molecules, side-chain groups and neighboring backbone amino acids.

Different conformations of the polypeptide backbone are described mainly by two backbone angles - ϕ a ψ. These angles define the respective orientation of the adjacent rigid peptide bond units connected through an alpha carbon (see Fig. 1.2).

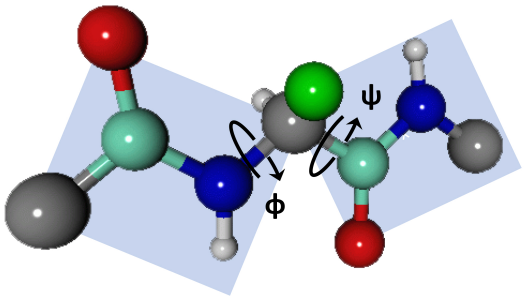

Fig. 1.2: Definition of dihedral angles ϕ and ψ: grey - alpha carbon, turquoise - carbon, blue - nitrogen, red - oxygen, white - hydrogen, green - side-chain residue.

Considering these two dihedral angles and the hybridization state of the involved

atoms, it is possible to deduce nine minima on the potential energy surface of a dipeptide, which is the minimal model of a peptide. These minima correspond to the nine different backbone conformations which can be present in dipeptides and in longer peptides. Terminology for the minima definition differs according to authors. In the Tab. 1.1, the terminology based on the works of Scarsdale et al.[47], Jalkanen et al.[30] and Wang et al.[58] is presented.

Tab. 1.1: Different conformations in the ϕ/ψ space.[57,30,58] C5 corresponds to the extended structure, $C7_{eq}$ and $C7_{ax}$ are cyclic hydrogen bonded structures, α conformers define different α helices and β corresponds to the left-handed polyproline helix (PPII).

		Φ		
		60°	180°	300°
Ψ	300°	$C7_{ax}$	α'	α_R
	180°	α_D	C5	β
	60°	α_L	β_2	$C7_{eq}$

Presented conformational terminology is often used in the computational and experimental studies of peptides. Simple model systems to study intrinsic propensities of different amino acids are e.g. small capped dipeptides in a form Ace-X-Nme, where Ace denotes the acetyl group, Nme is N-methyl and X states for studied amino acid (see Fig. 1.3). Note, that the name dipeptide refers to the number of peptide bonds in the molecule, not the number of amino acid residues.

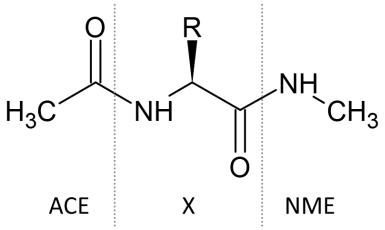

Fig. 1.3: Model capped dipeptide for conformation studies: ACE - acetyl, X - studied amino acid, NME - N-methyl.

The other frequently used model systems are host-guest peptides GGXGG and AAXAA, where G states for glycine and A for alanine residue. The serine and threonine host-guest peptides GGSGG and GGTGG as well as their phosphorylated forms were studied by He et al.[27]. They performed replica exchange molecular dynamics (REMD) simulations with 32 replicas using temperatures in range 278.0 - 478.0 K. According to their calculations, the phosphorylation of serine and threonine directly

influenced the intrinsic conformational propensities of the backbone. The phosphorylation of serine led to the increase in the propensity for the polyproline II (PPII) conformation. For the threonine residue, the situation was opposite. After the phosphorylation, the propensity towards PPII decreased. Phosphorylation of both amino acids also decreased the propensity of β-strand formation and increased the probability of the alpha helix conformation.

Shen et al.[50] reported the study of conformational preference of tetrapeptides GSSS and Ace-GSSS-NH_2 and their phosphorylated forms with the middle serine phosphorylated (single and double deprotonated states). The all-atom Brownian dynamics with implicit solvent was applied. They found out that the central serine phosphorylation shifted the conformational preferences to the alpha helix.

Later, Wong et al.[59] studied the tetrapeptide GSSS again. They tested more sophisticate water models including implicit and explicit solvent. They reported that both solvent models provide good agreement with NMR experiment. Unfortunately, they did not follow conformational changes during phosphorylation.

Another study[36] concerning tetrapeptide GSSX (X - Ser and its phosphorylated form) compared experimental data with simulations using different force fields. They demonstrated a strong dependency of conformation of studied peptides on the force field. According to the results of simulations, the most stable conformation of both peptides was right-handed alpha helix. The probability of the helical conformation also increased after the phosphorylation. On the other hand, the most stable state of both peptides measured by circular dichroism and infrared spectroscopy was found to be in extended conformation (ppII + β). They thus concluded that used force field were not sufficient for description of conformational changes.

Serine and threonine in the phosphorylated sites are often followed by the proline residue. Therefore, computational studies investigating the conformation of Ser/Thr - Pro peptides were performed.[18,26,54]

Byun et al.[18] studied the conformation of phosphorylated Ser-Pro and Thr-Pro in different solvents by density functional methods combined with implicit solvent method. They found significant differences in stable conformations in different solvents, however, they did not compare the phosphorylated and unphosphorylated structures.

Hamelberg et al.[26] decided to study a small peptide Thr-Ser-Pro-Ile in unphosphorylated and phosphorylated form by accelerated molecular dynamics. They focused on cis/trans isomerization of this peptide and also on the conformational changes after phosphorylation. According to their simulations, the serine phosphorylation changed the preferences in conformation – it led to increase of propensity to alpha helix conformation in serine and proline residues.

Many theoretical works were accompanied by experiments. In the next section, the most common experimental techniques for the elucidation of peptide conformation are presented and the experimental findings on the effect of phosphorylation are discussed.

1.3 Experimental studies of peptide phosphorylation

The influence of phosphorylation of peptide conformation can be studied by several experimental techniques such as circular dichroism (CD)[9,24,37,33,36], vibrational circular dichroism (VCD)[37,36], nuclear magnetic resonance (NMR)[52,33,37] or infrared vibrational spectroscopy (IR)[37,36,60].

The phosphorylation effect on the conformation of simple capped dipeptides was studied by Lee et al.[37] Using the CD technique, they found out that all studied dipeptides (Ser, Thr and their phosphorylated forms) occupied mainly PPII and extended β-strand conformations. The conformation of serine was almost unaffected by the phosphorylation, whereas the conformation of threonine changed from PPII to β-strand. They also studied the formation of hydrogen bonds in these dipeptides. According to their data, the dianionic phosphate group in threonine could form hydrogen bond with both amide protons (see Fig. 1.3). Unlike threonine, phosphorylated serine formed hydrogen bond only with the amide proton at the acetylated end of the dipeptide.

Kim et al.[33] also studied the effect of phosphorylation on the dipeptide conformation by CD and NMR. Their results were slightly different from those published in the previous study. They found out that phosphorylation of serine led to the increase in the PPII propensity, whereas the phosphorylation of threonine had the opposite effect. The difference in behaviour of both peptides was probably caused by the longer side chain in threonine residue which influenced the formation of intramolecular hydrogen bonds.

Analogous to small dipeptides, conformational properties of tetrapeptides GSXS were also investigated. Lee et al.[36] studied the conformation of GSSS and its phosphorylated form GSSepS by CD, VCD and IR spectroscopy and compared their results with molecular dynamics simulations. As already discussed in the previous section, they concluded that the studied peptides occurred in a mixture of PPII and β-strand structures. Their results also did not correspond to the simulations which predicted preferred alpha helix conformations.

The NMR study on tetrapeptides was done by Tholey et al.[52] They performed 1D and 2D ^{1}H NMR experiments with tetrapeptides in a form GSXS, where X marked Ser,

Tyr, Thr and their phosphorylated forms. They showed that conformation of peptides containing serine and threonine is affected by phosphorylation, whereas phosphorylation of tyrosine did not influenced the backbone conformation preferences. However, they did not infer the nature of structural changes, instead they presented the measured coupling constants.

Phosphorylation of longer peptides was also examined, interested reader can look e.g. at Ref. 9, 24. In the both publications, the effect of serine phosphorylation on the stability of alpha helix is described. According to the position of serine in the chain, the phosphorylation can stabilize or destabilize the alpha helix.

1.4 The aim of this thesis

The main goal of this thesis is to study the effect of phosphorylation on the conformation of different peptide/protein models and real systems by methods of computational chemistry. The methods used for calculations are the classical molecular dynamics and metadynamics.

The first step was to investigate the changes in conformational intrinsic propensities of phosphorylated amino acids in small dipeptides - serine, threonine, tyrosine and their phosphorylated forms (mono- and dianionic variants). Then the focus was changed on the conformation properties of host-guest pentapeptides of these amino acids in context of glycine and alanine peptides (GGXGG and AAXAA forms). All of these systems were studied by metadynamics.

The next step was the study of conformational behaviour of larger real systems including peptides made of fifteen amino acids extracted from intrinsically disordered proteins and whole proteins which structures are available in PDB in both phosphorylated and unphosphorylated forms. Because of the size of these systems, the classical molecular dynamics was used. Although the bias exchange metadynamics would provide better sampling of the conformational space, it is too computationally demanding. Also the choice of collective variables of such large systems is difficult.

Chapter 2

Computational methods

This chapter introduces the computational methods utilized in this work. The first section is focused on the description of classical molecular dynamics and general simulation protocol. This topic is covered in many textbooks[32,41,25] and manuals of simulation packages[8]. In the second section the basic concepts of metadynamics are presented. Metadynamics was applied in the conformational study of dipeptides and pentapeptides. Then the tools used for the analysis of the calculated trajectories are presented. In the last part the simulation protocols used for different studied systems are briefly described.

2.1 Classical molecular dynamics

If the quantum nature of the atomic nuclei are neglected, the movement of atoms and molecules can be described be the solution of Newton equations of motions. These equations are given as

$$\frac{\mathrm{d}\mathbf{r}_i}{\mathrm{d}t} = \frac{\mathbf{p}_i}{m_i}, \tag{2.1}$$

$$\frac{\mathrm{d}\mathbf{p}_i}{\mathrm{d}t} = \mathbf{f}_i = -\frac{V}{\mathrm{d}\mathbf{r}_i}, \tag{2.2}$$

where $\mathbf{r}_i$ are Cartesian coordinates of position, $\mathbf{p}_i$ are momenta, $\mathbf{f}_i$ are forces and m_i are masses of different atoms. V stands for interaction potential which is defined by used force field.

Newton equations of motion represent the system of ordinary differential equations. If the initial conditions (the initial positions and velocities of all the atoms)are known, this equations can be solved by different numerical methods. The positions of atoms are usually given by the chosen structure. The initial velocities can be then assigned randomly from the Maxwell-Boltzmann distribution.

The most frequently used method for the integration of equations of motion is the Verlet algorithm. This method can be understood as a Taylor expansion of the atom position as a function of the time t. The final equation for the positions is then given as a sum of the forward and backward Taylor expansions and it has the form

$$\mathbf{r}_i(t+\Delta t) = 2\mathbf{r}_i(t) - \mathbf{r}_i(t-\Delta t) + \frac{\mathbf{f}_i(t)}{m_i}\Delta t^2 + O(\Delta t^4). \tag{2.3}$$

The estimated new position contains an error of order Δt^4.

The velocity of the particles is not used explicitly in computation of their positions in the standard Verlet algorithm. However, it can be calculated straightforwardly as

$$\mathbf{v}_i(t) = \frac{\mathbf{r}_i(t+\Delta t) - \mathbf{r}_i(t-\Delta t)}{2\Delta t}. \tag{2.4}$$

In this thesis, the Newton equation of motion were solved using leap-frog algorithm, which is implemented in the program Gromacs[7]. This algorithm is equivalent to the Verlet algorithm and provides the identical trajectory. In the leap-frog algorithm, the velocities are used for the calculation of the updated positions, but the velocities themselves are related to the time shifted by the half of the time interval Δt. The final equations for the atomic positions and velocities are

$$\mathbf{r}_i(t+\Delta t) = \mathbf{r}_i(t) + \Delta t\mathbf{v}_i(t+\frac{1}{2}\Delta t) \tag{2.5}$$

$$\mathbf{v}_i(t+\frac{1}{2}\Delta t) = \mathbf{v}_i(t-\frac{1}{2}\Delta t) + \frac{\Delta t}{m}\mathbf{f}_i(t). \tag{2.6}$$

2.1.1 General simulation protocol

Although the different systems vary in a size and properties, the simulation protocol used to study them is often the same. It can be written as follows

Generate the initial structures → Define simulation box and environment → Minimization → Equilibration → Production run → Analysis

However, the detailed scheme depends on the character of studied problem. The scheme mentioned above is suitable for equilibrium processes when one trajectory is sufficient for calculation of time averages.

If one wants to simulate non-equilibrium processes like interactions with external field, it is necessary to run more than one molecular dynamics (MD) trajectory. Average properties of the system are then calculated over whole set of trajectories.

The first information one needs is the structure of the system. It can be obtained directly from databases containing X-ray or NMR data (e.g. RCSB Protein Data

Bank[4]) or it can be build in one of many available programs. In this thesis, I built extended peptide structures in the Pymol program.[48]

Once the molecular structure is available, one can start to prepare the simulation. At the beginning of the simulation, it is necessary to thoroughly choose a force field which describes the interaction potential of the molecule.

The studied system is then placed into a box. The box represents a space in which the system is simulated and it keeps the required conditions. The choice of a box is quite important. The box needs to be large enough to enable satisfactory movement of the molecule. At the same time, it should not be too large, because it would increase the computational cost of the simulation especially when the processes in solutions are studied. There is also a need to have a shape which can fully fill the space. Such shapes are e.g. cube, rectangular and triclinic parallelepipeds, rhombic dodecahedron or truncated octahedron.

After the system is placed in a box, the remaining empty space is filled with a solvent (if used). In order to avoid serious interface and boundary artifacts, the periodic boundary conditions are applied on the movement and interaction of the atoms and molecules in the box. Virtually, the whole box with the system is surrounded by a infinity set of its replicas. During the simulation, the molecules in a box can travel between different replicas as all replicas of molecules move simultaneously – when one molecule leaves the box, its replica enters into the box from the opposite side.

When the periodic boundary conditions are used, it is important to avoid calculating artificial interactions of a molecule with its periodic replica. Therefore, the minimum-image convention is usually followed for evaluation of interactions of a molecule with the rest of the simulated system. It means that interactions are calculated only within the few closest non-equivalent neighbouring molecules. This approach also leads to the significant simplification in forces calculations.

Another possible simplification of these calculations is the so called interaction cut-off. In this approach, the interactions with atoms which are further than the cut-off are neglected or described only in an approximative way. This is done by the shifting the potential to zero (using some constant value) or by using some shifting function which smooths the potential. The cut off is used mainly for the short-range interactions (e.g. Lennard-Jones) which decrease quickly with distance. The long-range interactions such as Coulombic interactions (ion-ion, dipole-dipole etc.) cannot be easily truncated without creating simulation artifacts. However, they can be efficiently evaluated under periodic boundary condition by Ewald-summation-based methods such as Particle-Mesh-Ewald (PME) summation[22].

The other thing one needs to think about before any simulation is the length of

the time step. If the chosen time step is too short, the simulation is not effective and its computational cost is high. On the other hand, a too long time step leads to large integration errors in the simulations. The length of the time step generally depends on the fastest motion in the simulation. The higher frequencies should be described, the shorter time step should be used. For example in biomolecules, the fastest motions are the vibrations of bonds involving hydrogen atoms. Therefore the appropriate time step should be about 1 fs. This is quite impractical if long simulations are required e.g. for description of the proteins and other biomolecules - a simulation with total length 1 μs would require 10^9 steps!

Hydrogen atoms are problematic also for their quantum character. Therefore their description by classical mechanics is not correct. For all these reasons, the chemical bonds of hydrogen atoms are "frozen" during the simulation. This can be done by constraining algorithms like SHAKE[46] or LINCS[28](LINear Constraint Solver). In this thesis, the LINCS algorithm was used. It is a simple, non-iterative method which resets bonds after unconstrained update to their correct lengths.

After completing steps described above the MD simulation run is almost prepared. However, the structure should not contain any inappropriate geometry or steric clashes. This could lead to large initial forces and to the problems with integrator or constraining algorithm. Therefore, the geometry of the initial structure is optimized to obtain lower initial energy and realistic geometry. In this work, the steepest descent algorithm was used. It is a robust method which uses only the first derivatives to find the direction to the local minimum.

The solution of Newton equations of motion leads to the energy conservation. At the same time, the volume of the system is constant because of the used box. The simulation thus describes microcanonical ensemble (NVE). However, typically one is interested in different kinds of ensembles, such as canonical ensemble NVT at constant temperature or NpT (isothermal-isobaric) ensemble at constant temperature and pressure. These kinds of ensembles can be generated from modified equations of motion.

Constant temperature during the simulation can be maintained by so called thermostats. The thermostats mimic the coupling to some external heat bath with given temperature. The simplest used thermostat (so called Berendsen thermostat[14]) rescales the velocity of particles after each step of the simulation to correct the temperature difference between the system and the heat bath. Although the Berendsen thermostat is quite simple, it does not provide real canonical ensemble. In this thesis, the v-rescale thermostat[16] was applied. It is similar to the Berendsen thermostat, but it adds additional stochastic force ensuring the correct distribution for kinetic energy. This thermostat thus simulates correct canonical ensemble keeping the advantages of

the velocity-rescale approach.

Temperature of the simulation should not be changed abruptly. Before the simulation starts, the system is shortly equilibrated as NVT ensemble. It means that the system is slowly heated to the simulation temperature. When the simulation of the system at constant pressure is required, the next short equilibration period is applied to establish the chosen pressure.

The constant pressure can be achieved by changing the volume of the box during the simulation. This change is maintained by a barostat. The most simple Berendsen barostat[14] rescales the vectors describing the box according to the reference pressure. In this thesis, Berendsen thermostat was adjusted at NpT equilibration. However, Berendsen barostat (similarly as Berendsen thermostat) does not provide correct NpT ensemble. Therefore, after the NpT equilibration, the Parinello-Rahman barostat[42], which gives the true NpT ensemble, was used.

After the equilibration, the production simulation is prepared to run. The chosen simulation time depends on the studied system and on the time scale of the observed processes. The typical affordable simulation time for proteins and other biomolecules is nowadays from 50 ns to several μs.

The main outputs from simulations are trajectories of the atoms and information about the energy, velocities, temperature and pressure. They can be used for further analysis.

2.2 Metadynamics

In this thesis, the effect of phosphorylation on peptide conformation was primarily studied by metadynamics. Metadynamics is an efficient method for sampling of the free energy surface (FES) as a function of few chosen collective variables (CVs). It uses external history-dependent bias potential, which is added to the Hamiltonian of the system. The first algorithm of metadynamics was published in 2002 by A. Laio and M. Parrinello[35]. Since that time, metadynamics has been rigorously derived and applied in different fields e.g. study of chemical reactions[23], protein folding[17] and material science[40]. In the following sections, I introduce the concept of this technique and briefly describe its extensions used in this work.

2.2.1 Principle of method

Classical molecular dynamics is a powerful method for studying behaviour of different systems. However, in biomolecules and similar complicated systems, the relevant con-

figurations often represent deep minima on the FES separated by high energy barriers. In such cases, the molecule is entrapped in these potential wells and can not overcome them easily and frequently by thermal motion. Therefore, sampling of the FES for such systems by molecular dynamics is quite difficult and computationally demanding.

Metadynamics[11,34] is one of the techniques which simplify sampling of the FES by the application of the external bias potential. In metadynamics, the bias potential is history-dependent. Application of this external potential is often described by a model of man trapped in a deep hole.[34] If we assume that the man has an infinity large amount of sand, he can slowly deposit this sand onto the place he visited. After a while, the amount of sand behind his legs is sufficient to help him to escape from the hole. Metadynamics uses the same principle. As the molecule moves, the external potential is slowly deposited along the FES and fills the minima. Molecule is then pushed from the well and can explore the other regions on the FES. Metadynamics thus provides information on the configurations which are difficult to sample by classical molecular dynamics.

The external potential used in metadynamics can be written as a sum of Gaussians or as an integral for so-called continuous direct metadynamics in the form[11]

$$V_G(S,t) = \int_0^t \omega \quad \exp\left(-\sum_{i=1}^{d} \frac{(S_i - Si(R(t')))^2}{2\sigma_i^2}\right), \tag{2.7}$$

where $S(R) = (S_1(R), ..S_n(R))$ is set of n CVs characteristic for the system, ω is the energy rate and σ_i is the width of Gaussian. The energy rate can be further written as

$$\omega = \frac{W}{\tau_G}, \tag{2.8}$$

where W is a Gaussian height and τ_G represents frequency of adding the Gaussians. These parameters directly influence the accuracy and efficiency of the sampling.

The height and width of a Gaussian define the volume which is occupied by adding one Gaussian to the potential. If the volume of the Gaussian is too large, the FES is sampled too quickly with low accuracy. On the other hand, if the Gaussians are small, the accuracy of reconstructed FES is higher, but the simulation is longer and less effective. The similar trend is valid for the τ_G. The deposition time should not be too short, because the system needs some time for a relaxation. However, the longer times lead to slower sampling of the space and then to higher computational cost.[11]

In metadynamcs, it is assumed that after sufficiently long time the bias external potential V_G provides unbiased estimation of the free energy according to

$$V_G(S, t \to \infty) = -F(S) + C, \tag{2.9}$$

where C is an additive constant and $F(S)$ is the free energy defined as a function of the CVs

$$F(S) = -k_B T \ln \left(\int dR \delta(S - S(R)) \mathrm{e}^{-\beta U(R)} \right), \tag{2.10}$$

where $U(R)$ is the potential energy.[34]

The error ε in the calculated FES was shown to be proportional to Gaussian parameters as[11]

$$\varepsilon \propto \sqrt{\frac{W k_B T}{\tau_G D}}, \tag{2.11}$$

where k_B is Boltzmann constant, T is thermodynamic temperature and D is the diffusion coefficient in the CVs space.

According to the assumption in Eq. 2.9, metadynamics can reconstruct the free energy surface without previous knowledge on the landscape required. It can be thus used to accelerate sampling far from the local minima. Due to these properties, metadynamics enables studying of new reaction pathways.[11,34]

In the Eq. 2.7, the bias potential is defined using so called collective variables which properly describe the system. In real systems, these variables are not generally known. Nevertheless, they need to be chosen before the beginning of the simulation. This choice is often difficult because a poorly chosen set of CVs often leads to an inaccurate reconstruction of the FES (see Eq. 2.10).

In the next section, I discuss the selection of appropriate CVs and show a few commonly used variables.

2.2.2 Collective variables

Wisely chosen collective variables are vital for correct calculations of the FES. The CVs should fulfill few requirements mentioned in ref. 11. They should:

- be able to describe all the relevant intermediates.
- include all slow modes of the system.
- be limited in a number.

Typically, small set of two or three CVs is chosen. The reason is that the large number of CVs leads to multidimensional problem which is difficult and computationally demanding to solve.

The choice of CVs always depends on the nature of the studied problem. Especially for the biosystems, the selection of efficient set of CVs is quite challenging, because of

large number of different configuration and reaction pathways. For study of peptide conformation, two dihedral angles ϕ and ψ are commonly used.[55]

Except for these geometrical parameters, the frequently used CVs are also coordination numbers, potential energy, hydrogen bonds or radius of gyration.[11]

2.2.3 Well-tempered metadynamics

The quality of the FES calculated by metadynamics depends on few factors. Besides the force field used and Gaussians parameters discussed in previous sections, the another problematic question is when to stop the simulation. In an ideal case, the bias potential should converge to the negative value of the free energy. Nevertheless, the real bias potential fluctuates around the correct value and metadynamics thus overfills the FES.

The general rule can be formulated as follows. If the metadyamics is used e.g. to find the lowest saddle point of studied minimum, it should be stopped when the minimum is filled and molecule escapes. On the other side, if the metadynamics tries to reconstruct the FES, it should be terminated when the molecule starts to move diffusively in the studied region.[11,34]

More sophisticated solution of the termination problem is provided by the well-tempered metadynamics.[12] In this method, the additional bias potential can be expressed in the form

$$V(S,t) = k_B \Delta T \ln\left(1 + \frac{\omega N(S,t)}{k_B \Delta T}\right), \tag{2.12}$$

where ΔT is an input parameter with the dimension of temperature, ω has the same meaning as in previous case and $N(S,t)$ represents histogram of collective variables during the biased simulation.

Time derivative of this potential is written as

$$\dot{V}(S,t) = \frac{\omega k_B \Delta T \delta_{S,S(t)}}{k_B \Delta T + \omega N(S,t)} = \omega e^{[-V(S,t)/k_B \Delta T]} \delta_{S,S(t)}. \tag{2.13}$$

Eq. 2.13 defines the rate of the changes of the potential. If the term $\delta_{S,S(t)}$ in the equation is replaced with a finite width Gaussian, the resulting equation is equivalent to the time derivative of Eq. 2.7.

This method is implemented into the metadynamics code by rescaling the height of the Gaussian. The new height is calculated as

$$W = \omega \tau_G \mathrm{e}^{\frac{V_G(S,t)}{k_B \Delta T}}, \tag{2.14}$$

where ω stands for initial bias deposition rate. The bias deposition rate decreases to zero as $1/t$. This means that for long times, the simulated system gets closer to the

equilibrium. However, the bias potential does not converge according to the Eq. 2.15. The convergence limit is given as

$$V_G(S, t \to \infty) = -\frac{\Delta T}{T + \Delta T} F(S) + C. \tag{2.15}$$

The CVs probability distribution is then given as

$$P(S) \propto \mathrm{e}^{-\frac{F(S)}{k_B(T+\Delta T)}}. \tag{2.16}$$

The choice of the ΔT parameter directly affects the character of the simulation. The limit case $\Delta T = 0$ leads to classical MD with no bias potential. On the other hand, $\Delta T \to \infty$ corresponds to the standard metadynamics.

Moreover, the well-tempered metadynamics allows to reconstruct correctly the canonical distribution of the unbiased degrees of freedom, e.g by reweighting technique developed by M. Bonomi et al.[15]

2.2.4 Extensions: Bias exchange

In metadynamics, the optimal sampling is reached when two or three CVs are used. In the most systems, the number of relevant slow degrees of freedom is higher, though. That is the reason why some extensions to metadynamics were added. They are generally based on the running of several replicas helping to improve the statistics of the FES estimation.

In this thesis, I focus on so called Bias exchange metadynamics.[43,34] In this approach, n replicas at the same temperature are running. In each replica, one or two collective variables are biased by time-dependent external potential V_G. After chosen time interval, bias potentials between two randomly selected replicas are exchanged. The probability of the acceptance of the exchange is given as

$$P_{ab} = \min\left(1, \exp\left\{\frac{1}{k_B T}\left[V_G^a(x^a, t) + V_G^b(x^b, t) - V_G^a(x^b, t) - V_G^b(x^a, t)\right]\right\}\right), \tag{2.17}$$

where $V_G^i(x^i, t)$ is bias potential of replica i in a point x^i and $V_G^i(x^j, t)$ is the same potential after the change of the coordinates between replicas.

Except running of the replicas biasing the CVs, it is also useful tu run one so called neutral replica with zero biasing potential. This replica exchanges with the other replicas according the Eq. 2.17. The difference is that this replica samples the approximative canonical distribution with unbiased potential. If $\omega/\tau_G = 0$, the canonical distribution is exact.[43]

2.3 Trajectory analysis

The following section describes the tools used for the trajectory analysis. These are especially the Ramachandran plot, radius of gyration and root-mean-square deviation.

2.3.1 Ramachandran plot

Peptide and protein secondary structure can be described by two backbone dihedral angles ϕ and ψ (see Fig. 1.2). Distribution of these angles is often depicted by so called Ramachandran plot.[44] In the Ramachandran plot, the x axis represents the values of angle ϕ and the y axis displays the angle ψ. Both axis are usually plotted in the interval (-180° - 180°).

Angles ψ and ϕ can occupy few allowed regions providing information about the secondary structure of peptides and proteins. Lowell et al.[39] collected the data about approximately 100 000 residues and refined the original allowed and disallowed regions. In the Fig. 2.1 the Ramachandran plots for a general residue, glycine, proline and pre-proline residues are shown.

Glycine and proline play special role because they differ in Ramachandran diagram from other amino acid residues. Glycine does not contain any side chain and thus can occupy many different conformations without any steric clashes. On the other hand, proline is the only imine among the amino acids. Due to its structure, the rotation around ϕ angle is restricted and can achieve only narrow interval of values. This restriction is so strong that it can even influence the residue situated in front of the proline (so called pre-Pro residues). These residues thus have special Ramachandran diagram showing their conformational preferences.

Every occupied area in the Ramachandran diagram represents different backbone conformation. These conformations are presented in Fig. 2.2. In the left figure, there are conformations typical for proteins as originally presented by Ramachandran et al.[45]. The right figure[56] depicts the distribution of conformations which is characteristic for dipeptides and which was used in this thesis.

2.3.2 Root-Mean-Square Deviation

Root mean square deviation (RMSD) describes the changes in protein structure during the simulation. In the Gromacs program, RMSD is calculated with respect to the reference molecule which is often the initial structure gained e.g. from the PDB database. In this thesis, the backbone fitting is used. The studied structure is then least-square

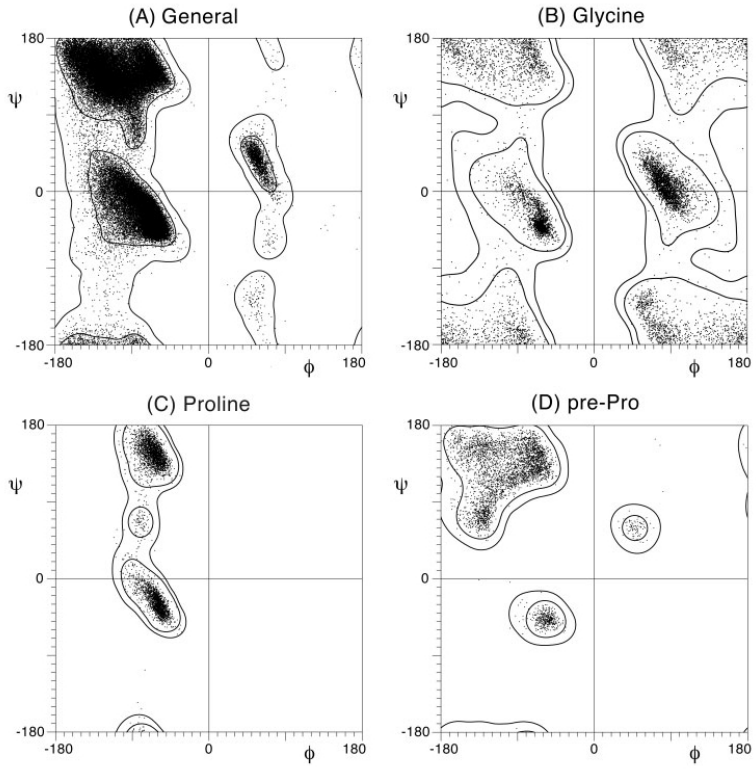

Fig. 2.1: Ramachandran diagrams as presented by Lovell et al.[39].

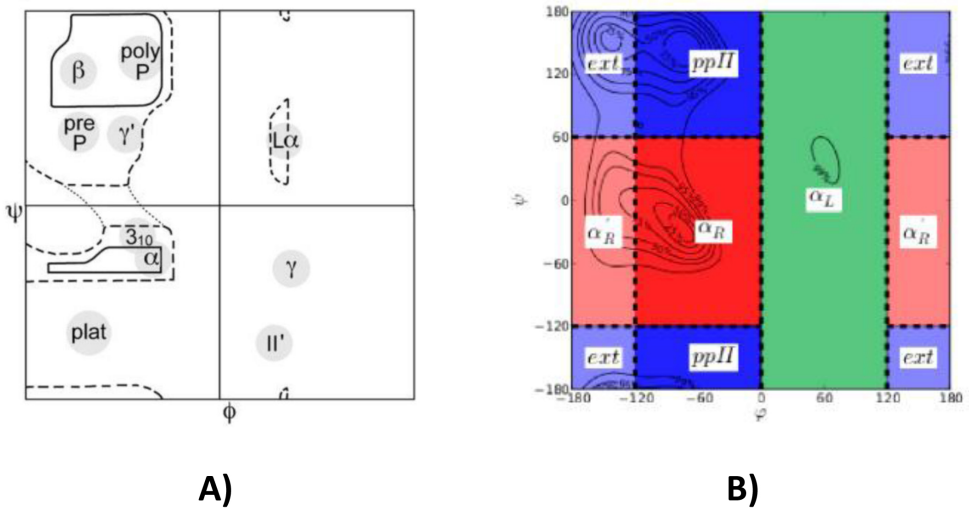

Fig. 2.2: Areas in the Ramachandran diagram represent different secondary structures. A) General Ramachandran diagram for proteins defined by hard sphere atomic overlaps[45], B) Ramachandran diagram for general dipeptides[56].

fitted to the chosen reference structure. RMSD can be calculated as

$$\mathrm{RMSD}(t_1, t_2) = \left[\frac{1}{M}\sum_{i=1}^{N} m_i ||\mathbf{r}_i(t_1) - \mathbf{r}_i(t_2)||^2\right]^{\frac{1}{2}}, \tag{2.18}$$

where m_i is a mass of atom i, M is the sum of mass of all atoms in molecule and r_i is atom position in given time.

Except the total RMSD, it is also possible to calculate Root-Mean-Square fluctuations (RMSF) of individual atoms or whole residues.

2.3.3 Radius of gyration

The radius of gyration describes the compactness of a studied structure. It is given as

$$R_g = \left(\frac{\sum_i ||r_i||^2 m_i}{\sum_i m_i}\right)^{\frac{1}{2}}, \tag{2.19}$$

where m_i is the mass of atom i and r_i is the position of atom with respect to the centre of mass of the molecule.

The calculation of radius of gyration is implemented directly in Gromacs. In this thesis, I used radius of gyration to compare the shape and compactness of unphosphorylated and phosphorylated proteins. Proteins with smaller radius of gyration are generally more compact.

2.4 Simulation protocols

In the following sections, simulation protocols used for different studied systems are described. All protocols were based on the general protocol presented in section 2.1.1 with slight differences. All molecular dynamics simulations were performed in the Gromacs 5.1 program.[7] To run the metadynamics, a variant of Gromacs extended by the Plumed 2.3 plugin[53] was used. All calculations were performed at the facilities provided by the National Grid Infrastructure MetaCentrum[1] and on resources provided by UNINETT Sigma2 - the National Infrastructure for High Performance Computing and Data Storage in Norway[5].

2.4.1 Simulation of dipeptides

For the simulations of dipeptides, the metadynamics enabling the extended sampling of conformational space was applied. This was necessary for reliable quantitative assessment of populations of individual conformers.

The capped structures of serine, threonine, tyrosine and their phosphorylated forms (monoanionic and dianionic) were studied. All these structures were built in the Pymol program in the extended conformation and they are depicted in the Fig. 2.3.

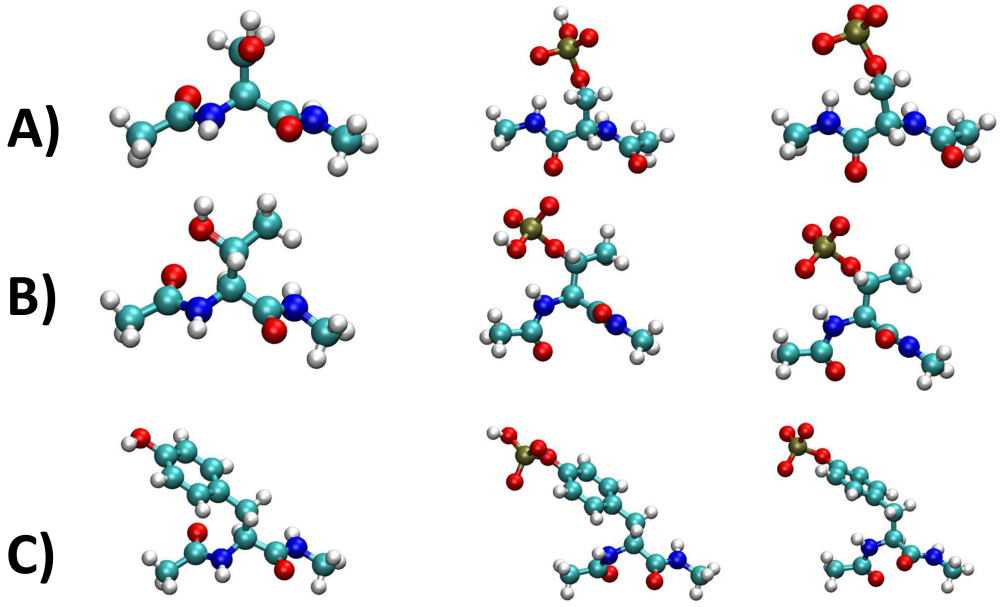

Fig. 2.3: The capped dipeptides in dephosphorylated, monoanionic and danionic phosphorylated forms: A) serine, B) threonine and C) tyrosine.

For the simulation, the Amber99SB-ILDN[38] force field extended by parameters for phosphorylated amino acids[29] was used. The structures were placed in a cubic box with the edge size about 3.6 nm. Then, the studied systems were solvated by explicit water molecules using the TIP3P water model[31]. The number of water molecules around each structure was about 1500.

The effect of the ionic strength on the conformation was tested. Two sets of structures were prepared – the first set did not contain any ions whereas the second set included about 30 Na^+ and Cl^- ions corresponding to the concentration about 1 mol/L.

Then, the structures were minimized using steepest descent gradient method. Afterwards, the whole system was equilibrated in two steps. In the first step, the NVT ensemble was simulated using the v-rescale thermostat. The final temperature was set to 298.15 K and the simulation time was 100 ps.

In the second part of the equilibration, the Berendsen barostat was employed to adjust the pressure to 1.0 bar. This system correspond to the NpT ensemble. The NpT equilibration took also 100 ps and the overall equilibration time was thus equal to 200 ps.

All dipeptides were simulated using well-tempered bias exchange metadynamics with five replicas. Each replica biased one collective variable. I used different dihedral angles - ϕ, ψ, χ_1 and χ_2. The fifth replica was always neutral (without any bias).

The definition of two backbone dihedral angles ϕ and ψ were mentioned in the section 1.2. The other two dihedral angles — χ_1 and χ_2 — represent side-chain dihedral angles. The χ_1 angle describes rotation around the bond between the alpha and beta carbons in peptide. Similarly, the side chain dihedral angle χ_2 is defined by the rotation around the bond between beta and gamma carbon. In the case of serine and threonine which do not contain gamma carbons, the angle χ_2 was defined by rotation around the bond between beta carbon and oxygen of the hydroxy or phosphate group.

The bias exchange potential was added each 500 steps (i.e. 1 ps) during the whole simulation, therefore 100 000 Gaussian were deposited along each FES. The width of Gaussian was equal to 0.05 rad and the initial Gaussian height was set to 1.0 kJ/mol.

The conditions in all replicas were the same. The systems were simulated at a constant temperature of 298.15 K maintained by the v-rescale thermostat. The pressure in simulation was 1.0 bar as in the equilibration; however, the Parinello-Rahman barostat was applied. In all cases, the isothermal compressibility of the system was set to the value of the pure water ($4.5 \cdot 10^{-5}$ bar^{-1}). All bonds in the simulations were constrained using the LINCS algorithm. The total length of simulation was 100 ns with the time step of 2 fs.

2.4.2 Simulation of pentapeptides

Apart from the dipeptides, I also studied the host-guest pentapeptides of the sequence Ace-GGXGG-Nme and Ace-AAXAA-Nme, where X stands for serine, threonine and their phosphorylated forms (both monoanionic and dianionic). The examples of structures are depicted in the Fig. 2.4

All capped pentapeptides were built in the Pymol program in extended conformations. The structures were simulated with the force field Amber99Sb-ILDN supplemented with parameters for phosphorylated amino acids[29]. All simulations were performed using a cubic box with the initial edge length of 5.8 nm. The structures were simulated in explicit solvent (about 6 500 water molecules in a box) using the TIP3P water model. The solution did not contain any ions.

The minimization, equilibration and metadynamics parameters were the same as in the simulation of dipeptides using the same settings for thermostats and barostats. The simulated time was 100 ns.

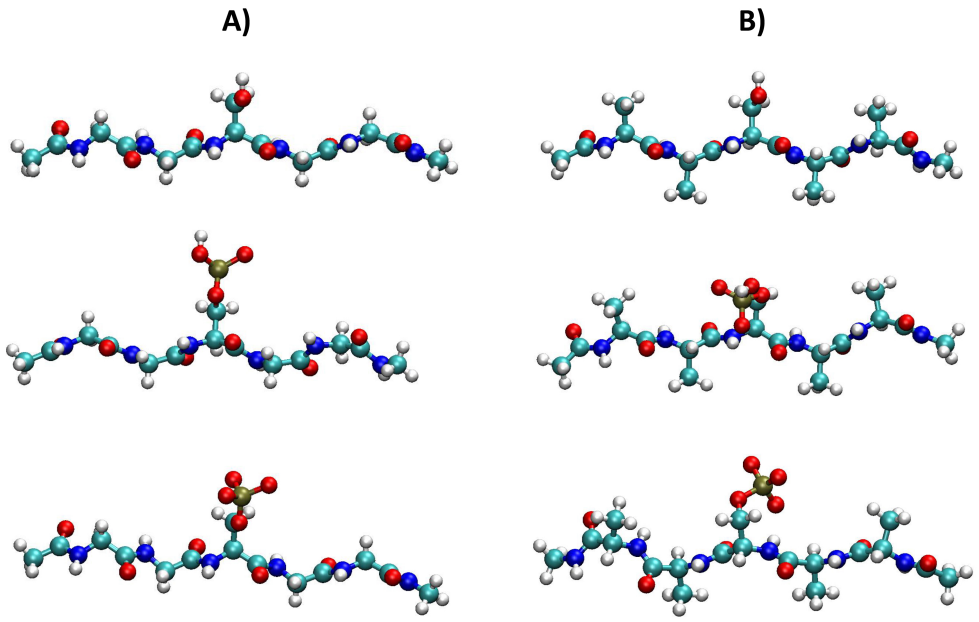

Fig. 2.4: The structures of pentapeptides containing serine and monoanionic and dianionic phosphoserine: A) Pentapeptides of the GGSGG sequence, B) pentapeptides of the AASAA sequence.

2.4.3 Simulation of peptides from disordered proteins

Additionally for the model dipeptides and pentapeptides, fifteen longer capped peptides made of fifteen amino acids were studied. These peptides were selected from the database of experimentally proven phosphorylation sites of different intrinsically disordered proteins. All peptides contained serine residue at the eighth position of the chain. At the same time, other phosphorylation sites were not present in the peptide.

All chosen peptides also originated from three disordered proteins with different properties. The information about the protein secondary structure were obtained from the UniProt database[6] and proven by Jpred4 – A protein Secondary Structure Prediction Server[?] .

The chosen disordered proteins were RAF proto-oncogene serine/threonine-protein kinase (P04049), Adenomatous polyposis coli protein (P25054) and Breast cancer type 1 susceptibility protein (P38398).

Extended structure of all peptides were built in the Pymol program. Both ends of the peptides were capped by acetyl and N-methyl groups to avoid interactions with charged residues.

All structure were placed in a rhombic dodecahedron box and simulated using force field Amber99SB-ILDN with parameters for phosphorylated amino acids[29]. Water

Tab. 2.1: Sequences of fifteen selected peptides from three different disordered proteins.

P04049	P25054	P38398
FDGSSCISPTIVQQF	QLNSGRQSPSQNERW	KAEFCNKSKQPGLAR
FGYQRRASDDGKLTD	EFSSGAKSPSKSGAQ	DPKDLMLSGERVLQT
VSSQHRYSTPHAFTF	MPKKKKPSRLKGDNE	QKGELSRSPSPFTHT
PKINRSASEPSLHRA	IPRSESASKGLNQMN	EDCSGLSSQSDILTT
	TPFNYNPSPRKSSAD	KSSEYPISQNPEGLS
		LQNRNYPSQEELIKV

environment was modeled by explicit solvent molecules (about 9 000 water molecules in each box) described by the TIP3P solvent model. All boxes also contained Na^+ and Cl^- ions in concentration 150 mmol/L which corresponds to the physiological conditions in a cell.

The studied structures were then minimized by the steepest descent algorithm and equilibrated as NVT and NpT ensemble. For equilibration, the v-rescale thermostat and Berendsen barostat were used. The peptides and solvent were thermostated independently. The final temperature of the system were 298.15 K and the pressure was set to 1.0 bar. Both equilibrations took 100 ps (i.e. 200 ps for equilibration in total).

The disordered peptides were simulated using classical molecular dynamics. V-rescale thermostat and Parinello-Rahman barostat were applied to maintain the conditions in the box. The constraints were applied only on the bonds involving hydrogen atoms using the algorithm LINCS. The total simulation time was 100 ns with time step 2 fs.

2.4.4 Simulation of proteins

Last studied systems were whole proteins containing phosphorylation sites. The protein structures for simulations were selected from the RCSB PDB database. The criteria for selection of proteins were the availability of phosphorylated and unphosphorylated structures and the size of the protein.

According to these requirements, three proteins were chosen – the Mengovirus Leader Protein, Anti-sigma F factor antagonist and Oxoglutarate dehydrogenase inhibitor.

The Mengovirus Leader Protein was simulated in three different forms. The first form did not contain any phosphorylated residue (structure 2MMH), the second structure was phosphorylated on threonine Thr51 (structure 2MML) and the third structure

included two phosphorylated residues – Thr51 and Tyr45 (structure 2MMK). All structures were determined by solution NMR.[10] The representative structures are depicted in Fig. 2.5.

The Anti-sigma F factor antagonist was studied in its two forms - one form was unphosphorylated and the second form contained phoshorylation on the Ser57 (see Fig. 2.6). The structures were obtained by X-ray diffraction with resolution 1.61 Å(unphosphorylated form, structure 1H4Y) and 1.16 Å(phosphorylated form, structure 1H4X).[49]

The last protein – Oxoglutarate dehydrogenase inhibitor was also simulated in its two forms (see Fig. 2.7) – the unphosphorylated structure (2KB3) and structure phosphorylated on Thr15 (2KB4). Both structures were gained from solution NMR.[13]

The simulation protocol of proteins were similar to disordered peptides. All proteins were placed in a rhombic dodecahedron box and solvated with explicit water molecules (about 30 000 water molecules in each box). The solution also contained Na^+ and Cl^- ions in concentration 150 mmol/L. The force field Amber99SB-ILDN with parameters for phosphorylated amino acids[29] was used for simulation.

All structures were minimized by the steepest descent algorithm and then equilibrated as NVT and NpT ensembles. Temperature 298.15 K was maintained by v-rescale thermostat. In the NpT equilibration, Berendsen barostat was adjusted (final pressure was set to 1.0 bar). Both equilibration parts took 100 ps.

The molecular dynamics was performed at constant temperature 298.15 K and 1.0 bar. The thermostat and barostat were the same as in peptide simulation. All H-bonds were constrained by the LINCS algorithm. The total simulation time was 100 ns with time step 2 fs.

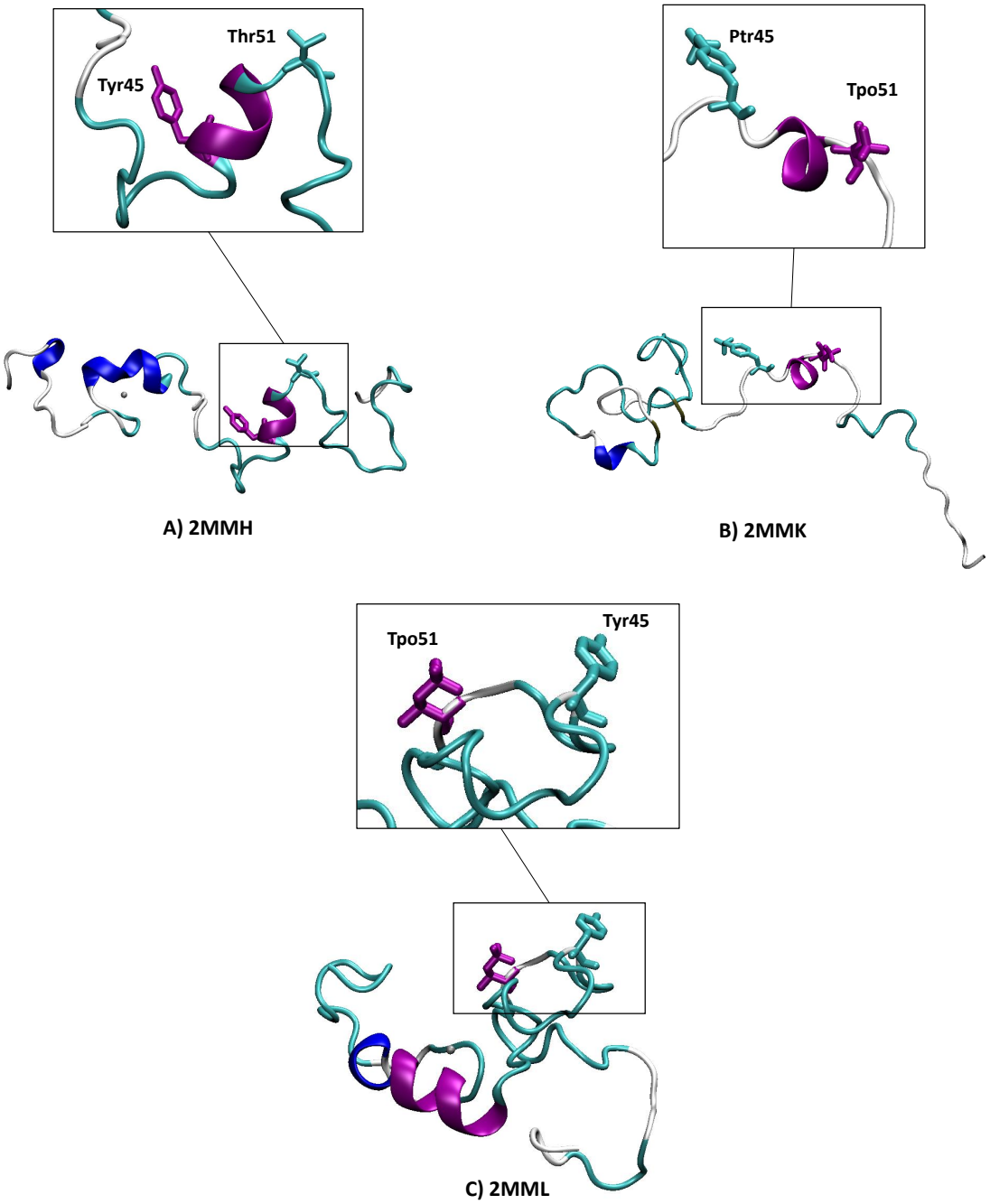

Fig. 2.5: The Mengovirus Leader Protein from *Cardiovirus a* – solution NMR structures[10]: A) 2MMH - unphosphorylated protein (phosphorylation sites Tyr45 and Thr51 are marked), B) 2MMK - phosphorylation on Thr51 (residue Tpo) and Tyr45 (residue Ptr), C) 2MML - phopsphorylation on Thr51 (residue Tpo).

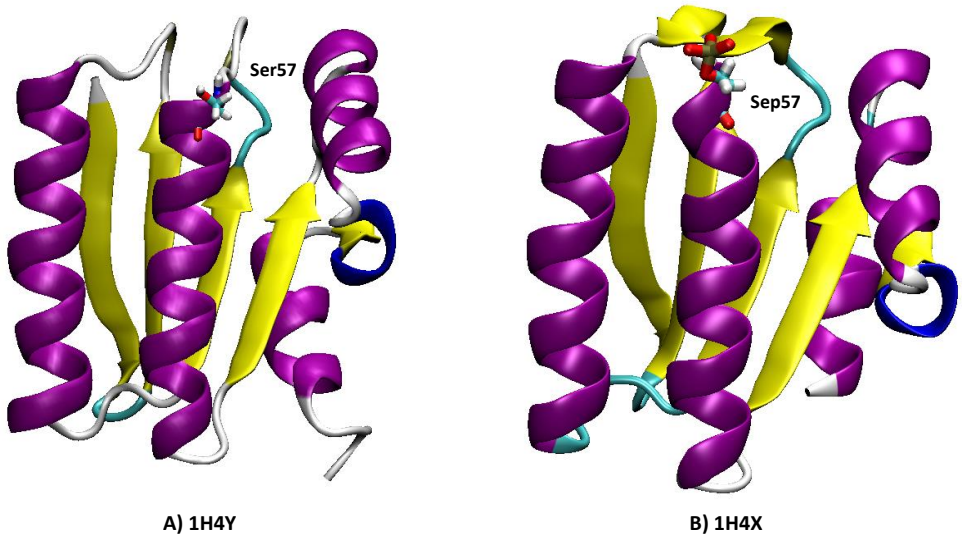

Fig. 2.6: The Anti-sigma F factor antagonist from *Lysinibacillus sphaericus*, X-ray diffraction structure[49]: A) 1H4Y - unphosphorylated structure (phosphorylation site Ser57 marked), B) 1H4X - phosphorylation on Ser57 (residue Sep).

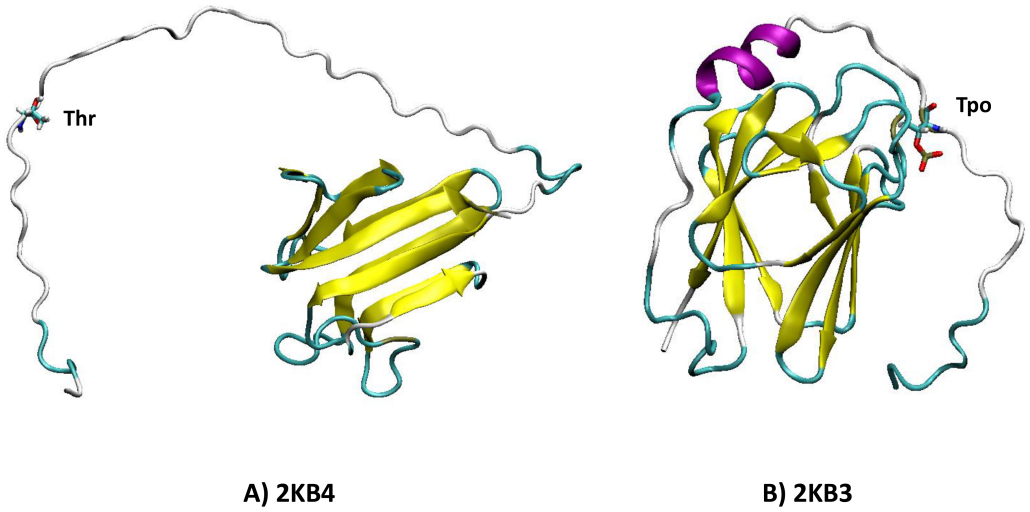

Fig. 2.7: Oxoglutarate dehydrogenase inhibitor from *Corynebacterium glutamicum*, structures from solution NMR[13]: A) 2KB4 - unphosphorylated form (phosphorylation site Thr15 marked), B) 2KB3 - phosphorylation on Thr15 (residue Tpo).

Chapter 3

Results

The following chapter presents the results of this thesis. The first section deals with the effects of phosphorylation on the conformational intrinsic propensities of capped dipeptides. The next section is focused on the conformational preferences in host-guest pentapeptides. Both model systems – dipeptides and pentapeptides – were investigated by the metadynamics. The results collected from classical simulations of the peptides from disordered proteins are also discussed. The last section describes the effect of the phosphorylation on the conformation of real proteins.

3.1 Dipeptides

The effect of phosphorylation on the conformation of simple dipeptides was addressed in this part. Serine, threonine, tyrosine and their phosphorylated forms were studied. All these systems were simulated in an aqueous solution at constant temperature 298.15 K and pressure 1.0 bar using the Amber99sb-ILDN force field. The bias exchange well-tempered metadynamics was used for the investigation of probability of different conformations. The possible conformations were identified according to the values of the dihedral angles ϕ and ψ. Throughout the text the used nomenclature corresponds to the areas of the Ramachandran diagram depicted in Fig. 2.2.

3.1.1 Serine

In the Tab. 3.1, the distribution of different conformations for serine and its two phosphorylated forms is shown. The conformations taken into account were extended form (Ext) and those typical for polyproline II (PPII) helix, right-handed alpha helix (α_R and $\alpha_{R'}$) and left-handed alpha helix (α_L). The reported values were calculated as an arithmetic average of values obtained from all five replicas. The statistical weight

of individual conformers in the biased replicas were evaluated by reweighting scheme of Bonomi et al.[15] The values (populations) obtained from different replicas differed at most by 10 %, but the typical difference was about 1.5 %.

Tab. 3.1: Comparison of probability of individual conformers[%] for different serine forms.

Structure	**Ext**	**PPII**	**Ext + PPII**	α_R	$\alpha_{R'}$	$\alpha_R + \alpha_{R'}$	α_L
Serine	20	37	57	27	9	36	7
Phosphoserine (1-)	18	25	43	43	12	55	2
Phosphoserine (2-)	17	17	34	50	15	65	1

It is obvious that the phosphorylation of serine led to changes in local intrinsic propensities. The most probable conformation of unphosphorylated serine was PPII structure (37 %), the probability for whole extended region and PPII structure was 57 %. The probability of PPII and Ext conformations decreased upon phosphorylation. The most probable conformation of phosphorylated serine thus became α_R conformation corresponding to the right handed alpha helix. This effect was stronger in the dianionic phosphoserine.

The Fig. 3.1 shows the most probable conformations of serine and phosphoserine forms.

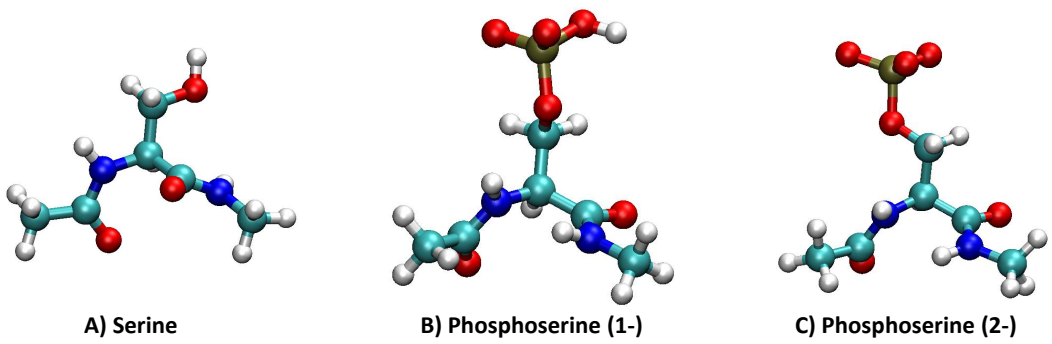

Fig. 3.1: Most probable conformations of serine and phosphoserine: A) Serine - PPII, B) Phosphoserine (1-) - α_R, C) Phosphoserine (2-) - α_R.

The Ramachandran diagrams for studied serine versions are depicted at Fig. 3.2. These Ramachandran were calculated from the neutral replica of metadynamics simulations.

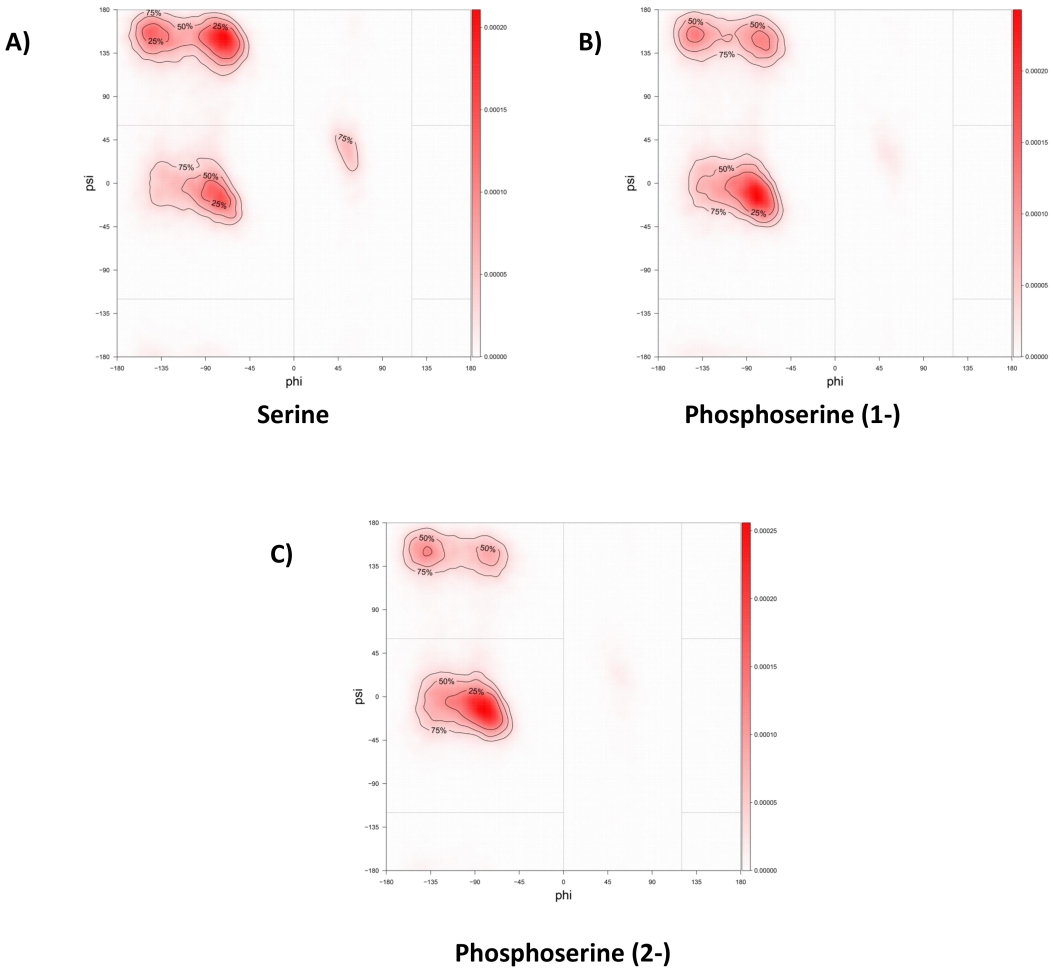

Fig. 3.2: Ramachandran diagrams of serine (panel A) and phosphorylated serine (monoanionic (panel B) and dianionic (panel C) forms) calculated from the neutral replica of metadynamics.

3.1.2 Threonine

Tab. 3.2 contains values for probability distributions of different forms of threonine. The most probable conformation of unphosphorylated threonine was PPII conformation (34 %) as in the serine case. However, the phosphorylation of threonine caused an increase of the probability for extended conformation to almost 50 %. Phosphorylated threonine thus reached 69 % probability for extended and PPII conformation. It means that the phosphorylation of threonine led to the increase of the intrinsic propensity for extended structure, whereas the phosphorylation of serine led to increase in right-handed alpha helix conformation. The prevailing conformations of threonine and phosphothreonine dipeptides are shown in Fig. 3.2.

The Ramachandran diagram of all threonine forms are shown in Fig. 3.4.

Tab. 3.2: Comparison of probability of individual conformers [%] for different threonine forms.

Structure	Ext	PPII	Ext + PPII	α_R	α'_R	$\alpha_R + \alpha_{R'}$	α_L
Threonine	19	34	53	34	12	46	1
Phosphothreonine (1-)	49	20	69	23	8	31	0
Phosphothreonine (2-)	55	18	73	21	6	27	0

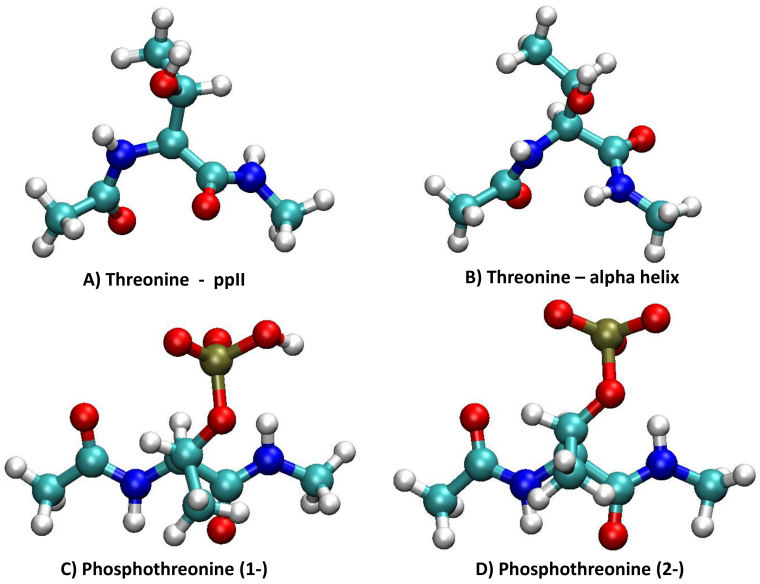

Fig. 3.3: Most probable conformations of threonine and phosphothreonine: A) Threonine - ppII, B) Threonine - α_R, C) Phosphothreonine (1-) - Ext, D) Phosphothreonine (2-) - Ext.

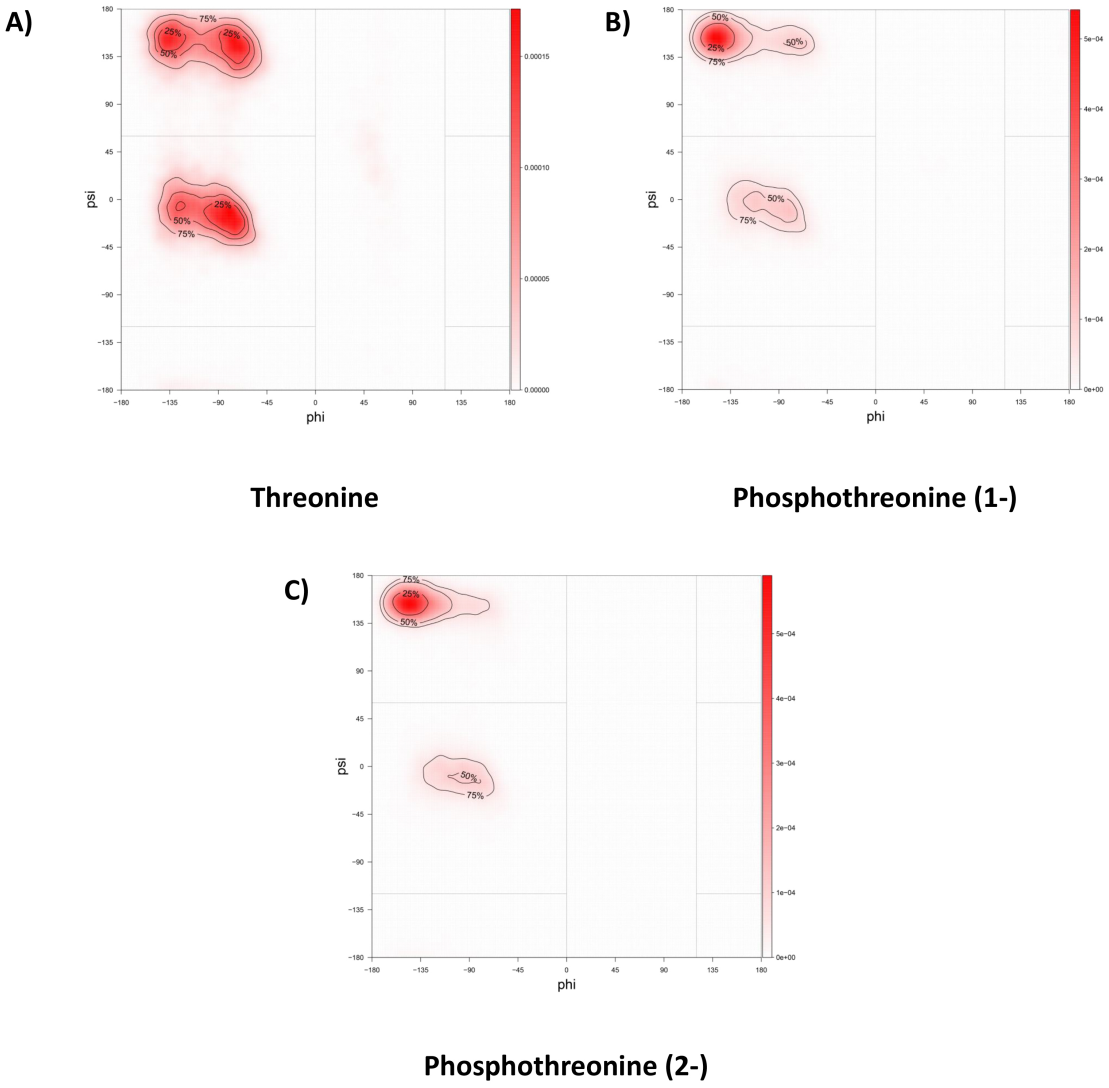

Fig. 3.4: Ramachandran diagrams of threonine (panel A) and phosphorylated threonine (panels B and C, monoanionic and dianionic forms) calculated from the neutral replica of metadynamics.

3.1.3 Tyrosine

Tab. 3.3 summarizes the conformation probabilities for different tyrosine forms. Whereas the phosphorylation of serine and threonine changed the intrinsic propensities of dipeptide backbone, the phosphorylation of tyrosine seems to have almost no effect on the dipeptide conformation. This behaviour might be caused by the location of the phosphorylation site at the tyrosine residue, which is more distant from the backbone than in case of serine or threonine.

The conformations of tyrosine and phosphotyrosine dipeptides are presented in Fig. 3.5. The most occupied conformation is ppII for all cases.

The Ramachandran diagrams of tyrosine are shown in Fig. 3.6.

Tab. 3.3: Comparison of probability of individual conformers[%] for different tyrosine forms.

Structure	Ext	PPII	Ext + PPII	α_R	$\alpha_{R'}$	$\alpha_R + \alpha_{R'}$	α_L
Tyrosine	32	43	75	15	8	23	2
Phosphotyrosine (1-)	24	54	78	16	5	21	1
Phosphotyrosine (2-)	25	49	74	19	6	25	1

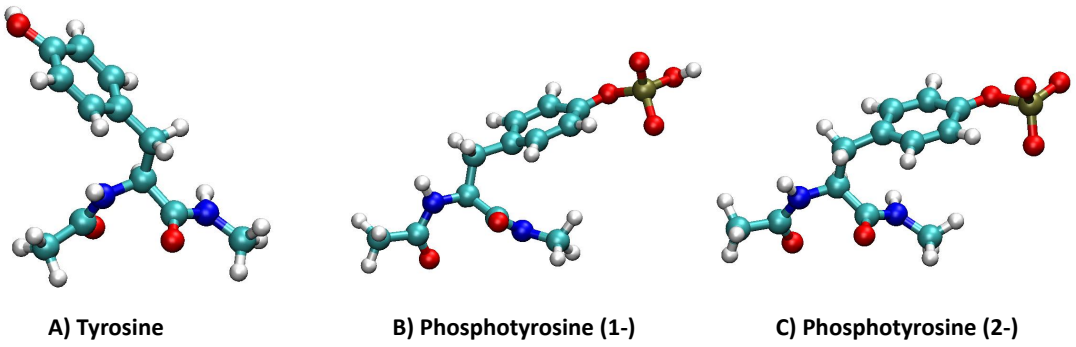

Fig. 3.5: Most probable conformations of tyrosine and phosphotyrosine: A) Tyrosine - ppII, B) Phosphotyrosine (1-) - ppII, C) Phosphotyrosine (2-) - ppII.

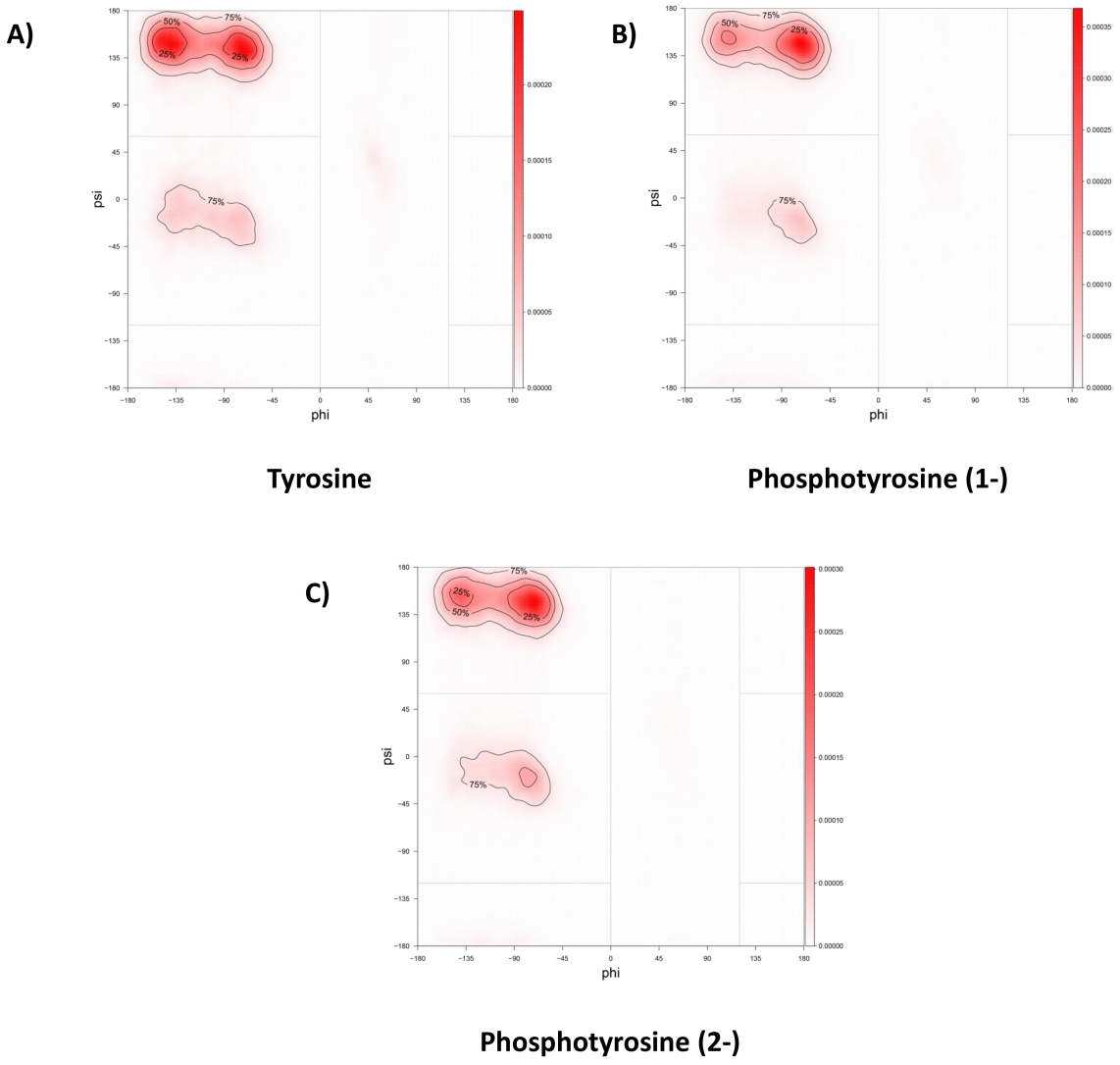

Fig. 3.6: Ramachandran diagrams of tyrosine (panel A) and phosphorylated tyrosine (monoanionic and dianionic forms, panels B and C) calculated from the neutral replica of metadynamics.

3.1.4 Effect of the ionic strength

Previously presented data for the dipeptide conformation preferences were calculated for dipeptides in aqueous solution with no added ions. Further we studied the effect of Na^+ and Cl^- ions on the conformation of serine, threonine and tyrosine and their dianionic phosphorylated forms. All dipeptides were placed in a solution containing 1 mol/L NaCl. The normal physiological concentration in a cell is about 150 mmol/L NaCl.

Tab. 3.4 contains summary of probabilities for conformers in all three amino acids. The preferences are quite similar to those with no added ions. Presence of the salt thus did not influence the effect of phosphorylation on the conformation of backbone.

Tab. 3.4: Comparison of probability of individual conformers [%] for different amino acids - the preferences calculated with no ions are shown in the parentheses.

Structure	**Ext**	**PPII**	**Ext + PPII**	α_R	$\alpha_{R'}$	$\alpha_R + \alpha_{R'}$	α_L
Ser	21 (20)	36 (37)	57 (57)	26 (27)	9 (9)	35 (36)	8 (7)
Sep (2-)	18 (17)	17 (17)	35 (34)	47 (50)	16 (15)	63 (65)	2 (1)
Thr	20 (19)	35 (34)	55 (53)	31 (34)	12 (12)	43 (46)	2 (1)
Pto (2-)	49 (55)	16 (21)	65 (73)	26 (21)	8 (6)	34 (27)	1 (0)
Tyr	32 (32)	43 (43)	75 (75)	15 (15)	8 (8)	23 (23)	2 (2)
Ptr (2-)	24 (25)	49 (49)	73 (74)	20 (19)	6 (6)	26 (25)	1 (1)

3.1.5 Hydrogen bond formation

Dipeptides containing hydroxy or phosphate group can form intramolecular hydrogen bonds with the amide backbone protons on both sides of a dipeptide. These bonds are illustrated in Fig. 3.7. In panel A, the hydrogen bond is formed between the serine hydroxy group and the both amide protons. In the panel B, the phosphate group oxygen creates bond with the both amide protons.

A simple distance criterion was used to detect hydrogen bond. The minimal distance between the particular amide atom and oxygen atom on the hydroxy group or the phosphate group in the side chain of the amino acid residue was at first evaluated. The distances were plotted as a histogram and compared with the values for unphosphorylated and dianionic phosphorylated dipeptides (see Fig. 3.8). The hydrogen bond length was set to have 3 Å maximum.[51]

Histograms show that serine dipeptide formed hydrogen bonds with the proton on the acetyl end of molecule whereas phosphoserine did not form any internal hydrogen

bonds.

In threonine, hydrogen bond is created between the hydroxy group and proton on the acetyl end of dipeptide. Phosphothreonine can create hydrogen bond between phosphate group and proton on the N-methyl end of molecule.

In tyrosine and phosphotyrosine dipeptides no hydrogen bonds to amide protons were observed for hydroxy or phosphate group at the tip of the side chain.

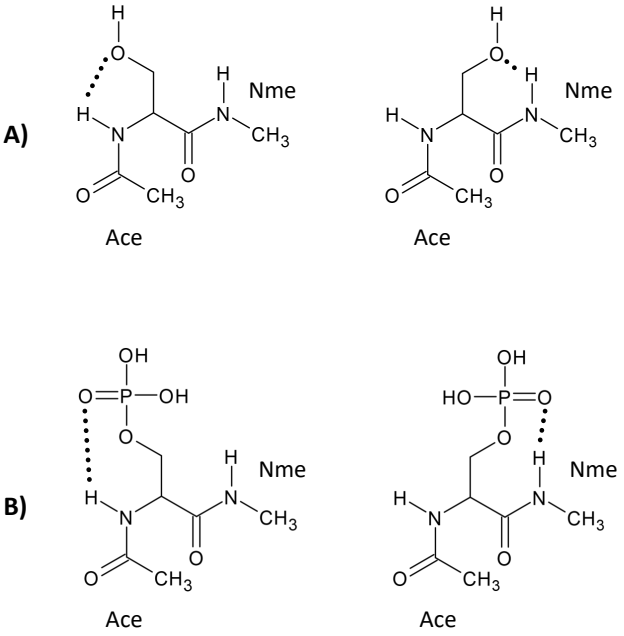

Fig. 3.7: Hydrogen bonds between hydroxy (phosphate) group and amide protons: Panel A - theoretical hydrogen bonds in serine, panel B - theoretical hydrogen bonds in phosphoserine.

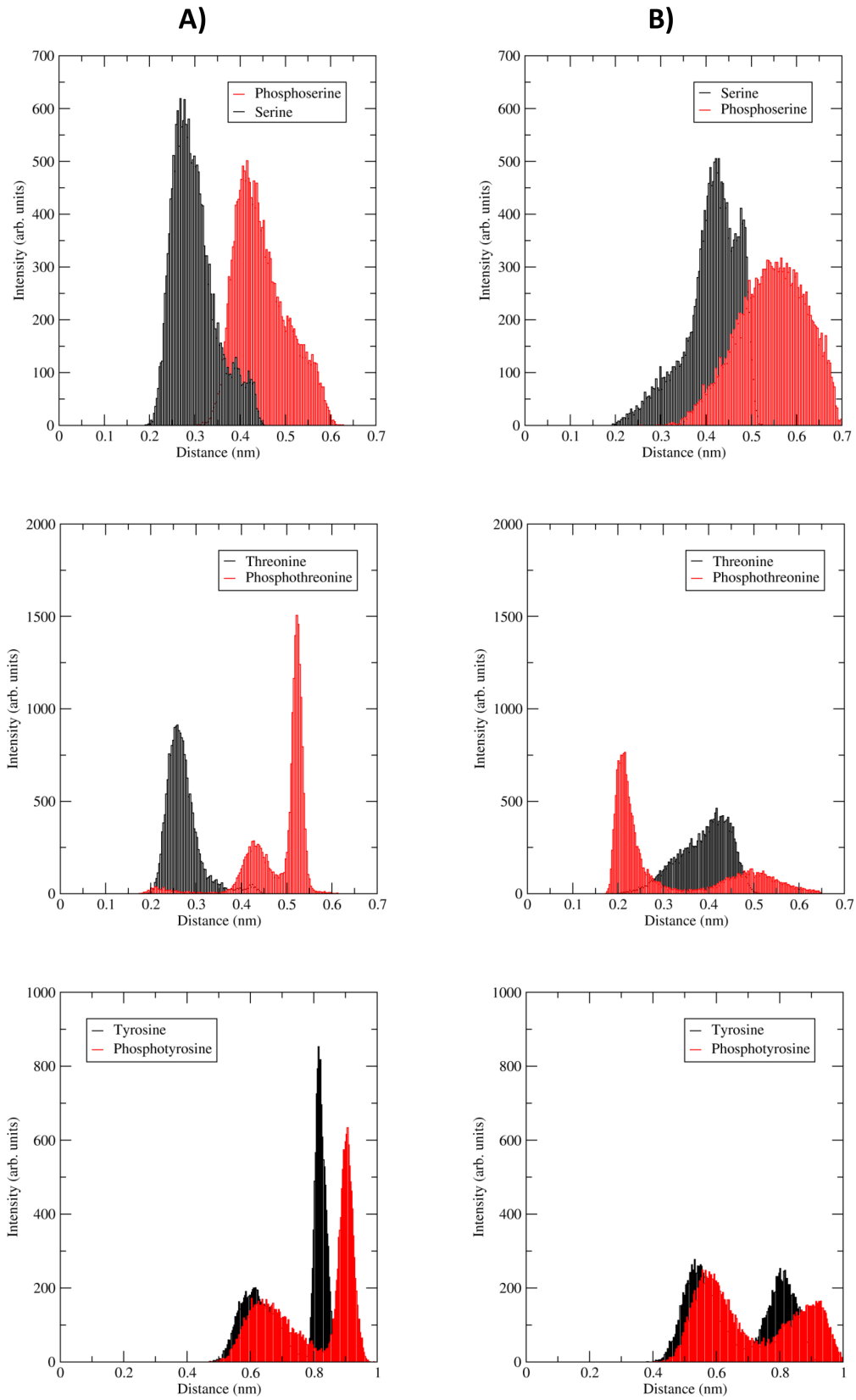

Fig. 3.8: Distribution of distances between side-chain oxygens and backbone amide protons: A) Amide proton on the acetyl end of dipeptide, B) amide proton on N-methyl group.

3.2 Pentapeptides

This section presents data calculated for capped glycine and alanine host-guest pentapeptides of the sequences GGXGG and AAXAA. The symbol X represents serine, threonine and their phosphorylated forms. These two amino acids were chosen for further study because of the observed effect of the phosphorylation on the backbone intrinsic propensity in dipeptides. In this thesis, the sequences of amino acids in pentapeptides are written by combination of one and three-letter amino acids abbreviations. The three-letter abbreviations are used for highlighting of the phosphorylated and unphosphorylated forms of amino acids (e.g. GGSerGG vs. GGSepGG).

Short peptides made of glycine and alanine do not contain any charged or polar groups and thus do not interact with the charged phosphate group. At the same time, they represent a good model systems to study longer peptides.

All pentapeptides were simulated in water solution with the Amber99sb-ILDN force field at constant temperature of 298.15 K and constant pressure of 1.0 bar. The bias exchange metadynamics of the same setup as in the dipeptide case were applied in simulation.

3.2.1 Serine

In the Tab. 3.5, the calculated probabilities of different conformations of pentapeptides are summarized.

Tab. 3.5: Comparison of probability of individual conformers for GGXGG and AAXAA host-guest peptides containing serine and its phosphorylated forms.

Structure	**Ext**	**PPII**	**Ext + PPII**	α_R	$\alpha_{R'}$	$\alpha_R + \alpha_{R'}$	α_L
GGSerGG	15	28	43	37	10	47	10
GGSepGG (1-)	13	19	32	46	12	58	10
GGSepGG (2-)	9	13	22	60	12	72	6
AASerAA	17	40	57	25	6	31	12
AASepAA (1-)	15	24	39	47	8	55	6
AASepAA (2-)	10	13	23	60	10	70	7

The results show that serine in the GGSerGG pentapeptide had similar probability of being in Ext or PPII conformation (43 %) as in α_R or $\alpha_{R'}$ conformation (47 %). The highest probability was observed for α_R area corresponding to the right-handed alpha helix. In the contrary, AASerAA pentapeptide occupied mainly the PPII region

(40 %). Also the probability of being in Ext or PPII conformation was notably higher than the probability of both right-handed alpha helix conformations ($\alpha_R + \alpha_{R'}$).

However, serine phosphorylation influenced the intrinsic propensities in both pentapeptides the same way. After the phosphorylation, the probability of PPII conformation decreased according to the same behaviour pattern observed in dipeptides. In all cases, the probability of α_R conformation significantly increased.

3.2.2 Threonine

Tab. 3.6 presents data on the capped host-guest pentapeptides GGXGG and AAXAA containing threonine and its phosphorylated monoanionic and dianionic forms.

Tab. 3.6: Comparison of conformation probabilities for GGXGG and AAXAA host-guest peptides, where X states for threonine and its phosphorylated forms.

Structure	**Ext**	**PPII**	**Ext + PPII**	α_R	$\alpha_{R'}$	$\alpha_R + \alpha_{R'}$	α_L
GGThrGG	14	28	42	43	12	55	3
GGTpoGG (1-)	41	15	56	31	12	43	1
GGTpoGG (2-)	68	20	88	10	2	12	0
AAThrAA	18	35	53	34	10	44	3
AATpoAA (1-)	57	21	78	18	4	22	0
AATpoAA (2-)	70	19	89	9	2	11	0

The table shows that pentapeptide GGThrGG occurred mainly in right-handed alpha helix conformation α_R (43 %), whereas pentapeptide AAThrAA had almost equal probabilities of being in conformation PPII and α_R. The preferences for Ext or PPII conformation was higher than for alpha helix conformation.

After the phosphorylation, the probability of Ext structure significantly increased in both studied phosphorylated pentapeptides. At the same time, the preferences to the other conformations decreased. The effect of phosphorylation was observed stronger in dianionic peptides. This trend is the same as was observed for threonine dipeptides.

Fig. 3.10 depicts the Ramachandran plots for threonine and its phosphorylated forms calculated from the neutral replica.

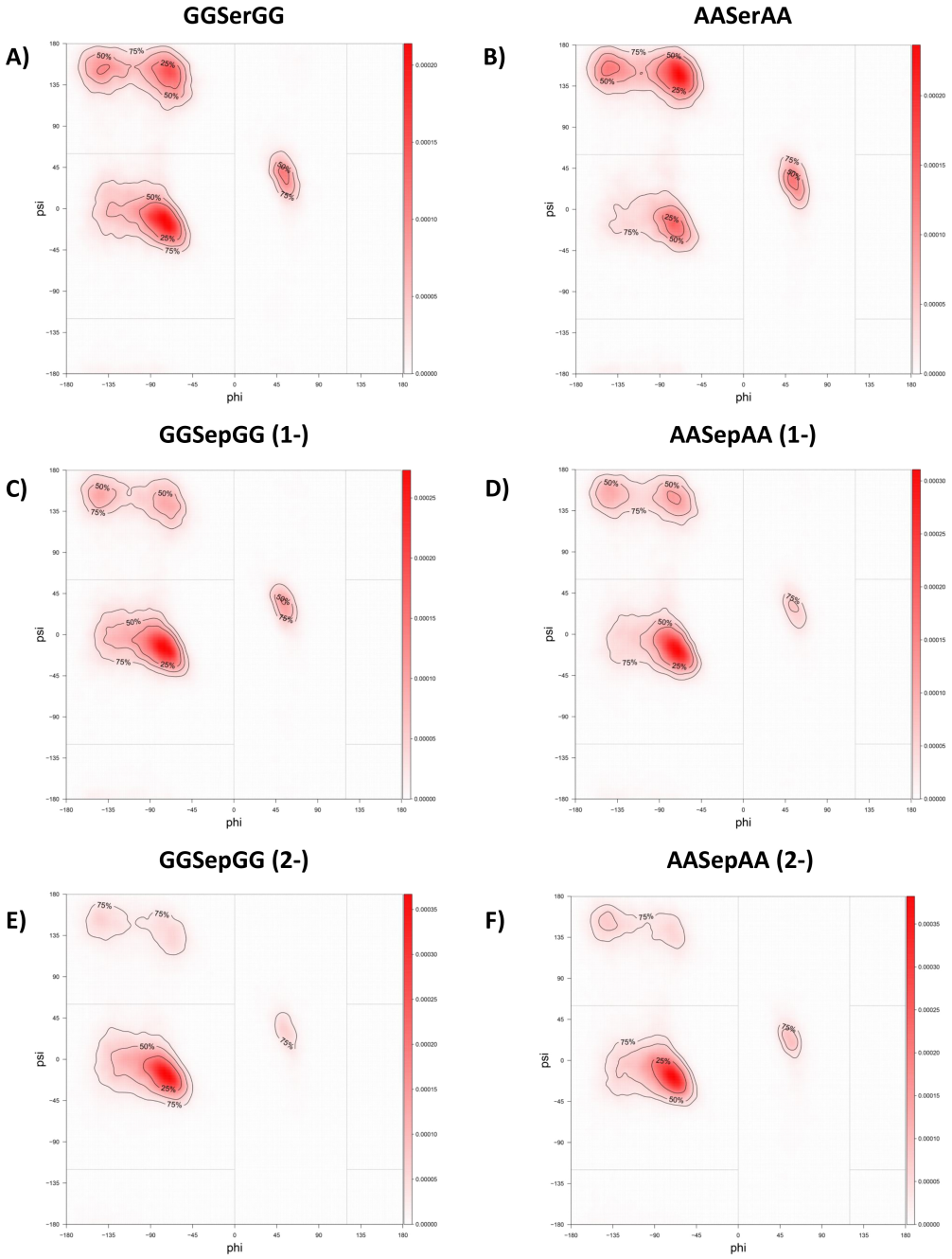

Fig. 3.9: Ramachandran diagrams of serine and its phosphorylated forms in host-guest pentapeptides calculated from the neutral replica of metadynamics.

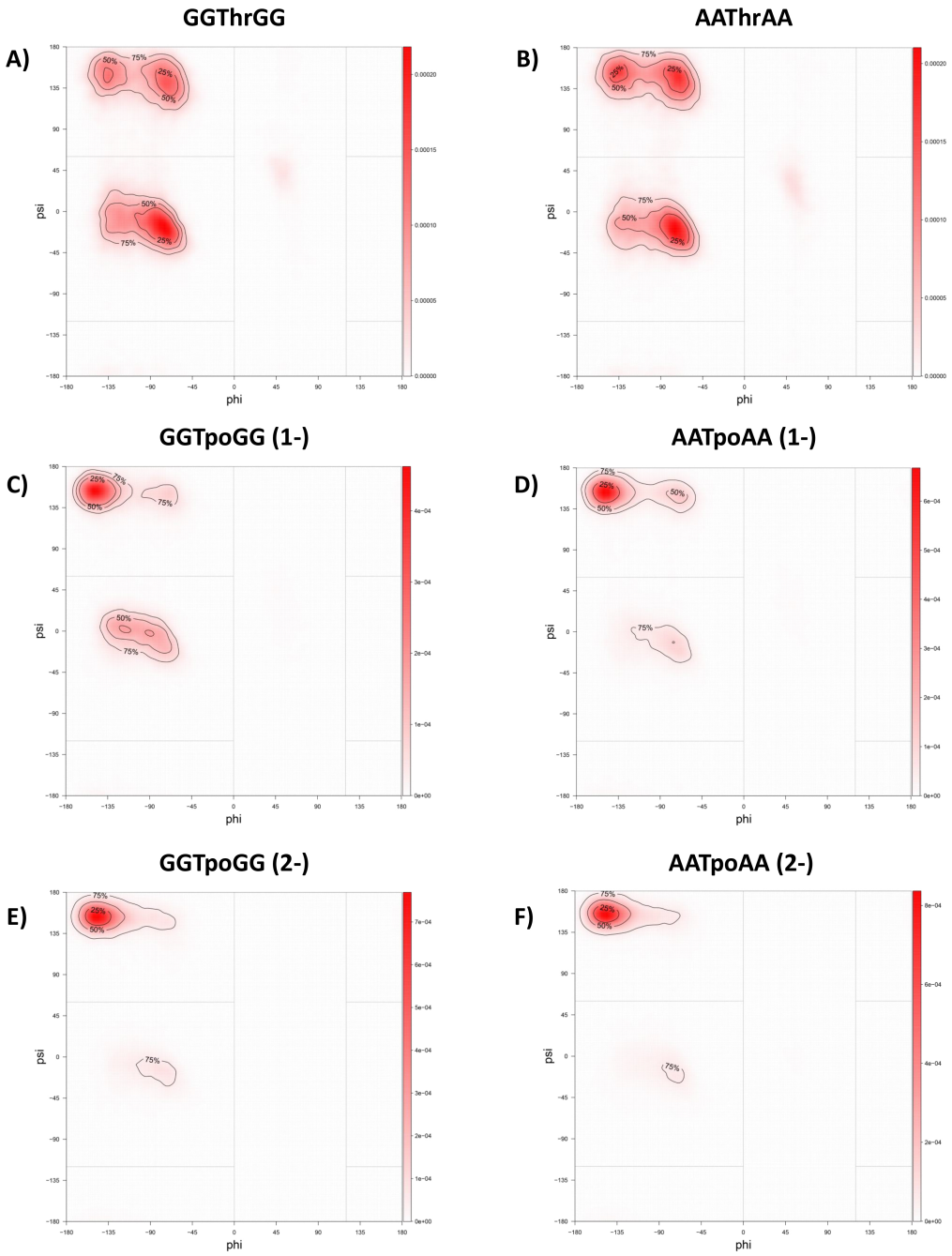

Fig. 3.10: Ramachandran diagrams of threonine and its phosphorylated forms in host-guest pentapeptides calculated from the neutral replica of metadynamics.

3.3 Peptides from disordered proteins

This section summarizes results calculated for the peptides derived from three different disordered proteins. The chosen proteins were all human proteins – i) RAF proto-oncogene serine/threonine-protein kinase, ii) Adenomatous polyposis coli protein and iii) Breast cancer type 1 susceptibility protein. Fifteen peptides containing the serine phosphorylation site according to the databases of these sites – PHOSIDA[2] and PhosphoSitePlus[3] were chosen for the purpose of this study. All peptides were fifteen amino acids long with serine residue placed at the middle of the chain and peptides were capped on both sides. Modelled structures of peptides are shown in Appendix A.

In the figures and diagrams presented in this section, the serine phoshorylation site is labelled by three-letter abbreviation Ser/Sep whereas the other amino acids are labelled by one-letter abbreviations. This marking was introduced to highlight the phosphorylation sites and different forms of amino acid residues.

All studied peptides can be sorted into two groups. In the first group the serine residue was followed by any of the amino acids except proline. The second group contains peptides in which serine residue is followed by proline residue.

All peptides were simulated in unphosphorylated and phosphorylated form. All structures were built in the Pymol program and the initial conformation was fully extended. Peptides were simulated in the water solution containing 150 mmol/L NaCl.

The nomenclature used for different conformation regions corresponds to the Ramachandran diagram for general proteins shown in Fig. 2.2.

3.3.1 Serine phosphorylation sites in disordered proteins

The peptides containing the serine phosphorylations site followed by amino acids different than proline are as follows:

DPKDLML**S**GERVLQT
EDCSGLS**S**QSDILTT
FGYQRRA**S**DDGKLTD
IPRSESA**S**KGLNQMN
KAEFCNK**S**KQPGLAR
KSSEYPI**S**QNPEGLS
LQNRNYP**S**QEELIKV
MPKKKKP**S**RLKGDNE
PKINRSA**S**EPSLHRA
VSSQHRY**S**TPHAFTF

Fig. 3.11 shows the Ramachandran diagrams for central serine residues of two peptides. Ramachandran diagrams for the other peptides are part of the Appendix B. The diagrams show that the intrinsic propensities in unphosphorylated peptides differ significantly due to character of their sequences. However, the phosphorylation changed all these propensities in favour to the right-handed alpha helix conformation region. This trend is similar for all studied peptides.

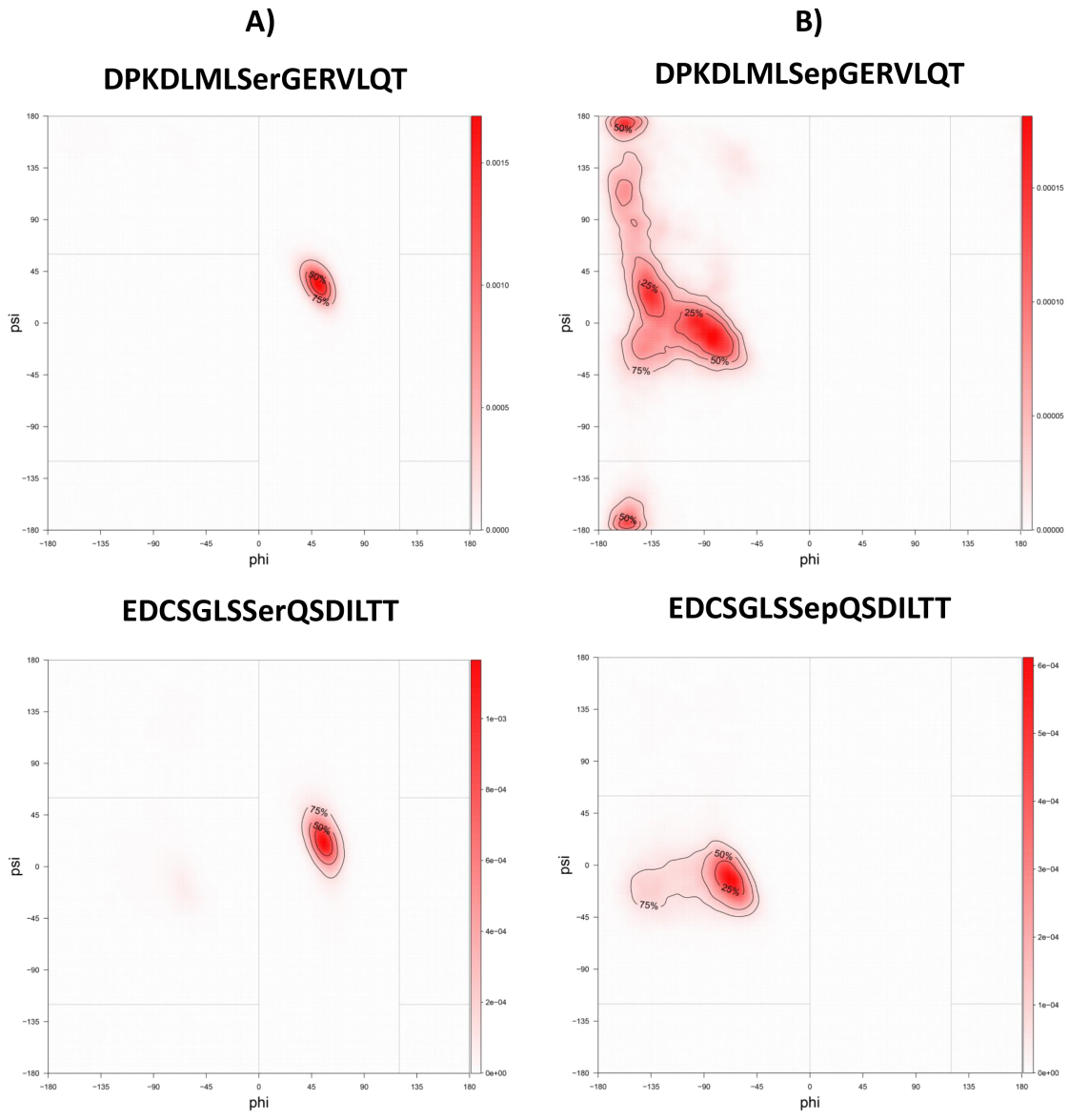

Fig. 3.11: Ramachandran diagrams calculated for the central serine residue: The column A - unphosphorylated forms, the column B - peptides phosphorylated on the central serine.

Fig. 3.12 depicts the radii of gyration calculated for two representative peptides.

The diagrams for the other studied peptides are presented in the Appendix B. Because of the disordered origin of studied peptides, as you can see no general trends were found.

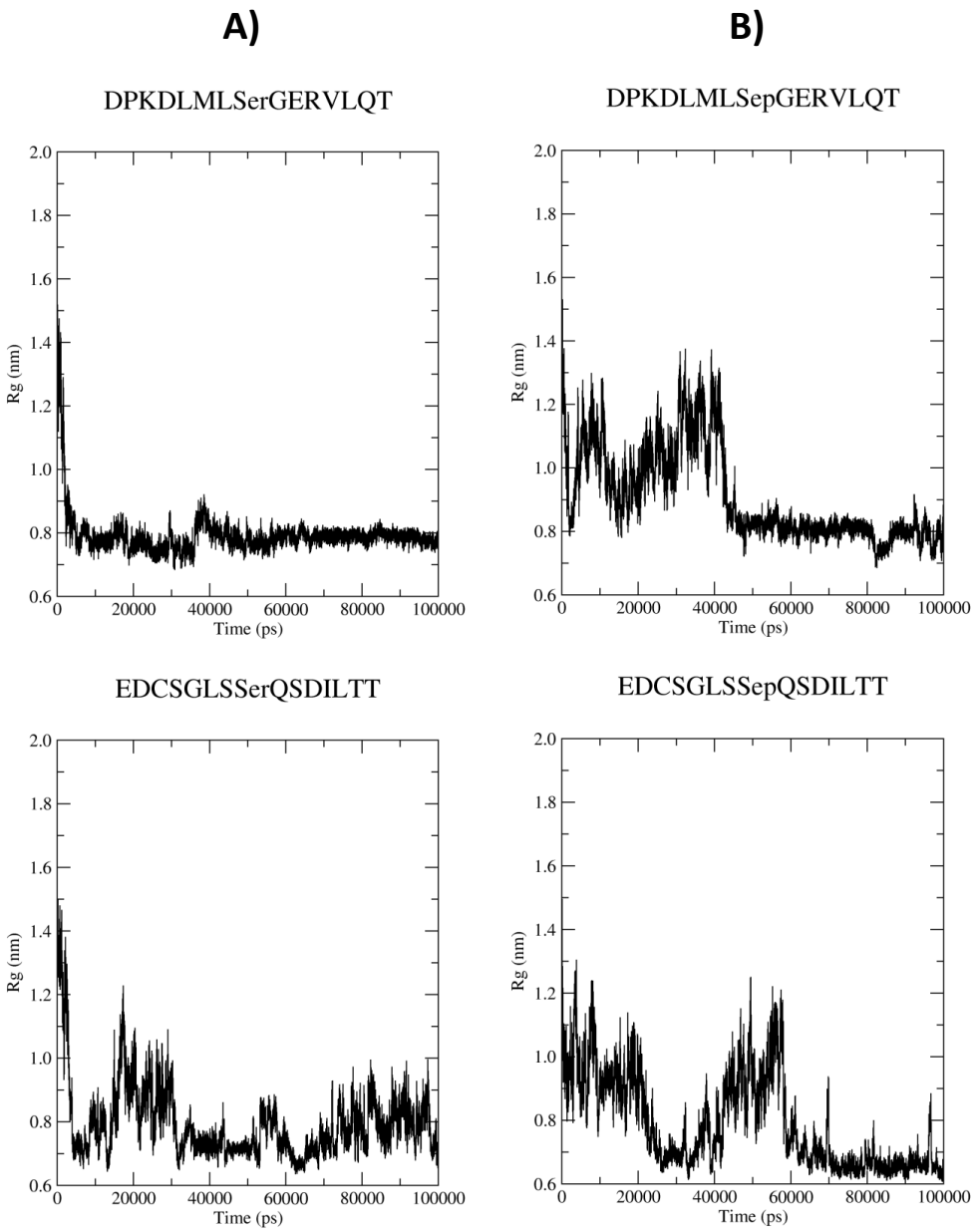

Fig. 3.12: Radii of gyration for two disordered peptides. Column A - unphosphorylated forms, column B - peptides phosphorylated on the central serine.

3.3.2 Ser-Pro phosphorylation sites in disordered proteins

In the second group of studied peptides, the central serine residue was followed by proline. The sequences of these peptides are as follows.

EFSSGAKSPSKSGAQ
FDGSSCISPTIVQQF
QKGELSRSPSPFTHT
QLNSGRQSPSQNERW
TPFNYNPSPRKSSAD

The Fig. 3.13 presents the Ramachandran diagrams for the central serine residue in two disordered peptides. Ramachandran diagrams for other peptides are included in the Appendix B.

According to these diagrams, the conformation of serine residue in unphosphorylated peptides is mostly Ext or PPII. After the phosphorylation, the intrinsic propensity of this serine almost did not change. In the peptides with phosphorylation sites containing Ser-Pro motives thus phosphorylation did not change the serine conformation toward right-handed alpha helix. This is consistent with the Ramachandran diagrams for pre-Pro residues (see Fig. 2.1) presented in Lovell et al.[39] paper.

The Fig. 3.14 shows the radii of gyration for two disordered peptides containing Ser-Pro motif. The radii of gyration graphs show that there is again no general effect of the phosphorylation on the compactness of the peptides. The figures of the rest of studied peptides are enclosed in the Appendix B.

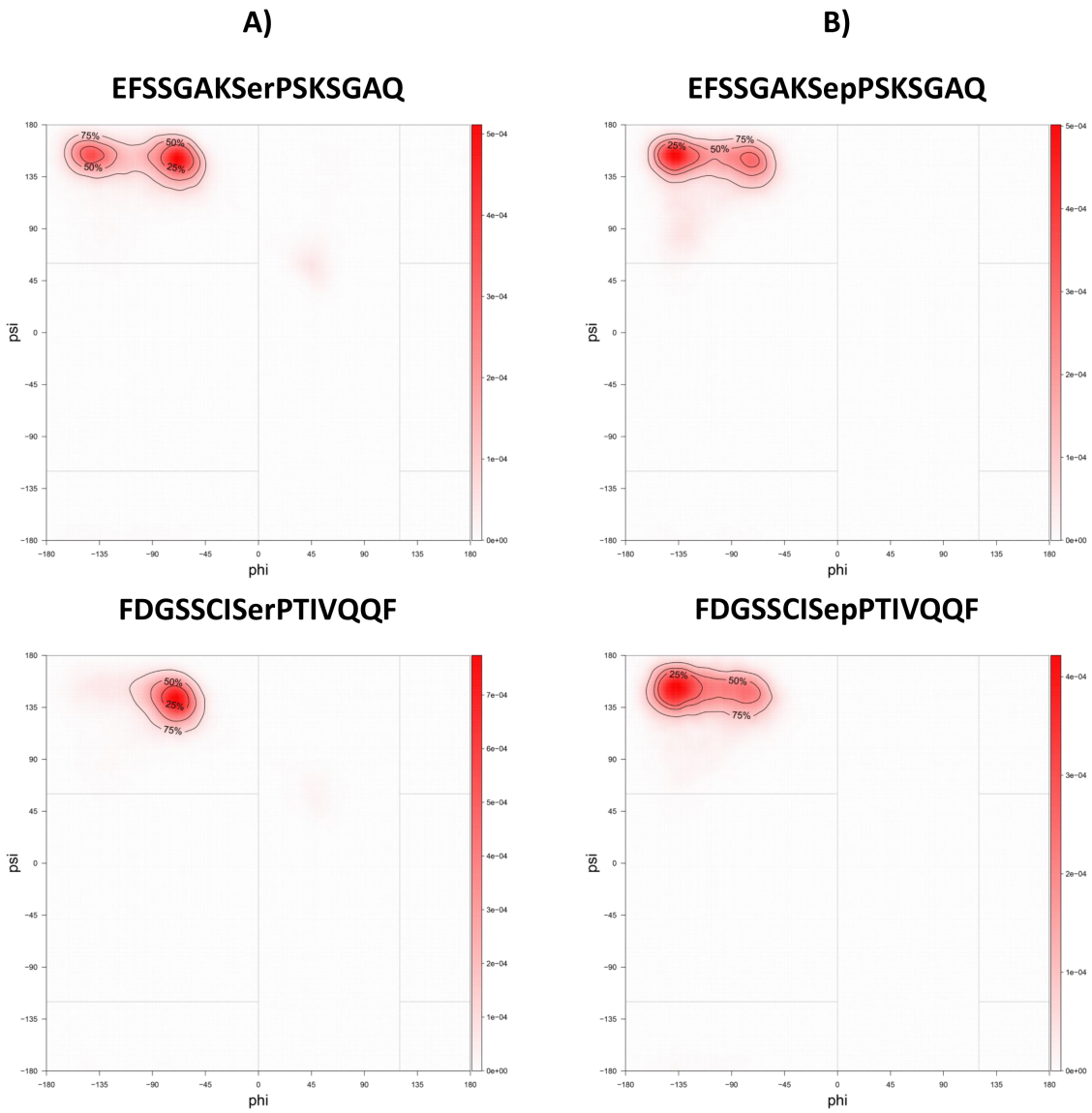

Fig. 3.13: Ramachandran diagrams calculated for the central serine residue of disordered peptides: Left - the unphosphorylated forms, right - peptides containing the phosphorylated serine.

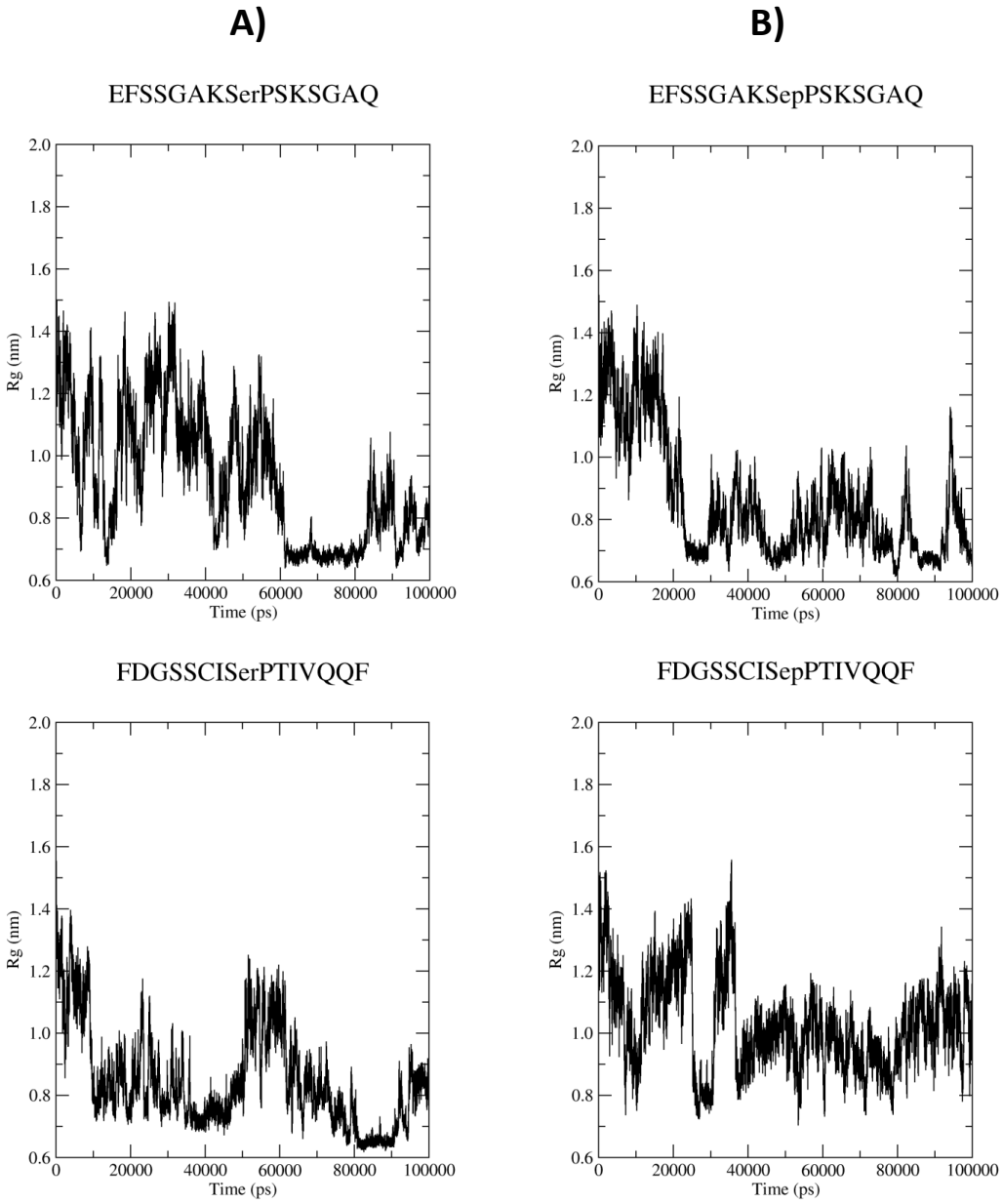

Fig. 3.14: Radii of gyration for two disordered peptides. Left - unphosphorylated forms, right - peptides phosphorylated on the central serine.

3.4 Proteins

The last part of this diploma thesis focuses on the simulations of proteins containing phosphorylated amino acids. Seven protein structures from RCSB PDB protein data bank were identified for this thesis. These structures represent three different proteins i) Mengovirus Leader Protein (from *Cardiovirus a*), ii) Anti-sigma F factor antagonist (from *Lysinibacillus sphaericus*) and iii) Oxoglutarate dehydrogenase inhibitor (from *Corynebacterium glutamicum*) in unphosphorylated and phosphorylated forms.

Two structure forms of protein i) were simulated. Phosphorylated structure 1HX4 comprised 111 amino acid residues (see Fig. 2.6), where Ser57 was phosphorylated. Structure 1H4Y represents unphosphorylated version of the protein containing 115 amino acid residues. The four extra amino acid residues are located at the C-terminal of the protein.

Protein ii) was simulated in its two forms 2KB3 and 2KB4 (see Fig. 2.7). Both proteins contained 143 amino acid residues. The structure 2KB3 was phosphorylated on Thr15.

The last protein iii) was studied in three different forms containing 71 amino acid residues and Zn^{2+} atom (see Fig. 2.5). Structure 2MMH is unphosphorylated version of the protein, whereas structure 2MMK was phosphorylated on Thr51 and Tyr45 and structure 2MML only on Thr51.

Both structure versions of protein i) were obtained by X-ray diffraction, whereas all the protein forms ii) and iii) were gained from solution NMR experiments. They were also more flexible and disordered than the structures from X-ray diffraction.

All proteins were simulated in aqueous solution containing 150 mmol/L NaCl at constant temperature of 298.15 K and constant pressure of 1.0 bar. Force field Amber99SB-ILDN was used. Total length of each simulation was 100 ns.

3.4.1 Anti-sigma F factor antagonist

Fig. 3.15 shows comparison of radius of gyration for proteins 1H4Y and 1H4X during the simulation. The plots show that radius of gyration of phosphorylated proteins stay the same during the simulation. The unphosphorylated structure was less compact at the beginning of the simulation. However, during the simulation the radius of gyration converged to the similar value as in phosphorylated structure.

On the other hand, the RMSD values for both structures differed a lot (see Fig. 3.16). Whereas the RMSD of phosphorylated structure 1H4X remained almost constant during the whole simulation, the RMSD of structure 1H4Y increased and fluctuated more. However it converged after the 70 ns of simulation, it also became constant.

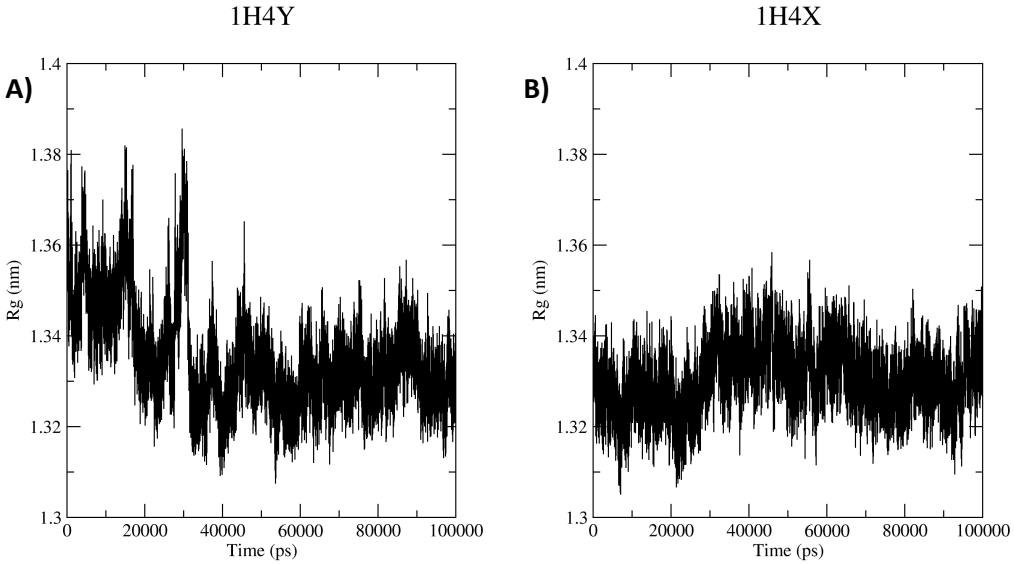

Fig. 3.15: Comparison of radius of gyration for 1H4Y and 1H4X in course of simulation. 1H4X contains phosphorylated Ser57.

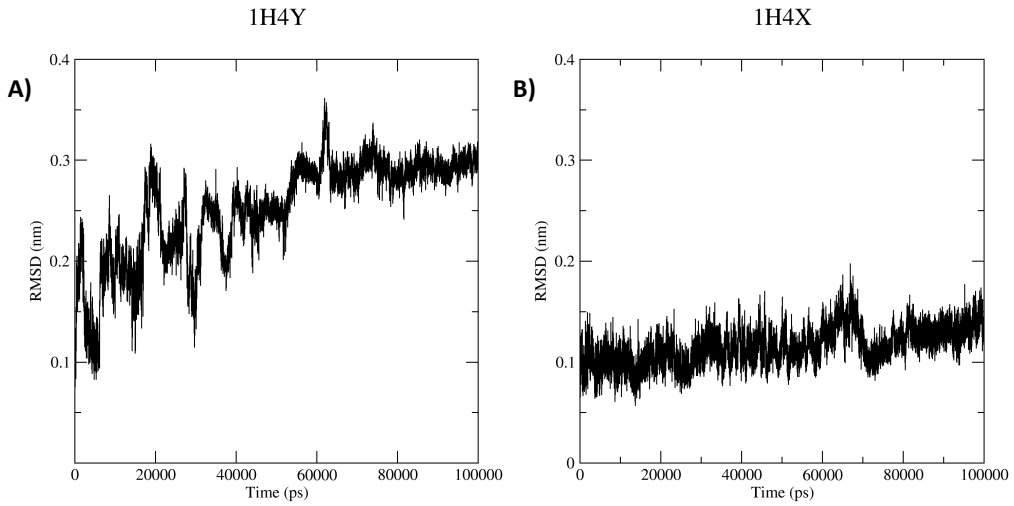

Fig. 3.16: Comparison of RMSD for 1H4Y and 1H4X in course of simulations. The protein backbone was used for superposition. Structure 1H4Y does not contain any phosphorylated residues, structure 1H4X is phosphorylated on Ser57.

The Fig. 3.17 compares the RMSD fluctuations in both structures. The fluctuations are averaged over all simulation time at each residue. Both structures differ in a phosphorylated residue Ser57. Although both structures had quite different RMSD during simulation, the fluctuations on all residues are similar.

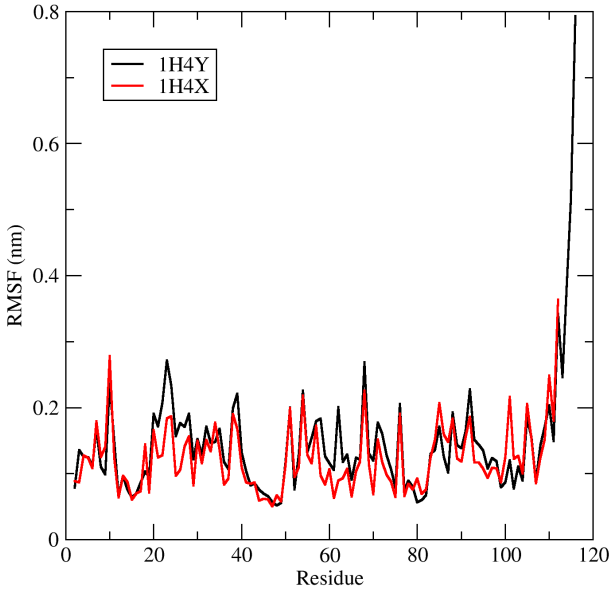

Fig. 3.17: Fluctuations of RMSD averaged over residues: Comparison of 1HX4 and 1H4Y structures. 1H4X is phosphorylated on Ser57.

3.4.2 Oxoglutarate dehydrogenase inhibitor

Fig. 3.18 shows the radius of gyration of structures 2KB3 and 2KB4. Whereas the radius of gyration of phosphorylated structure 2KB3 remained almost constant during the simulation, the structure 2KB4 had quite large radius of gyration which then quickly decreased to value similar to the phosphorylated form. This large radius of gyration of structure 2KB4 was likely caused by the long disordered part of the protein (see Fig. 2.7) which had more flexibility at the beginning of the simulation. However, during the simulation protein formed compact structure similar to the phosphorylated one.

In the protein structure 2KB3, the salt bridge the salt bridge was created between the phosphorylated threonine and arginine residue in position 87. This salt bridge was formed at the beginning of the simulation and remained stable during the whole simulation time. It thus probably contributes to the larger compactness of the structure. Fig. 3.19 shows the histogram of distances between residues Tpo15 and Arg87. I also presented detailed structure of the formed bridge.

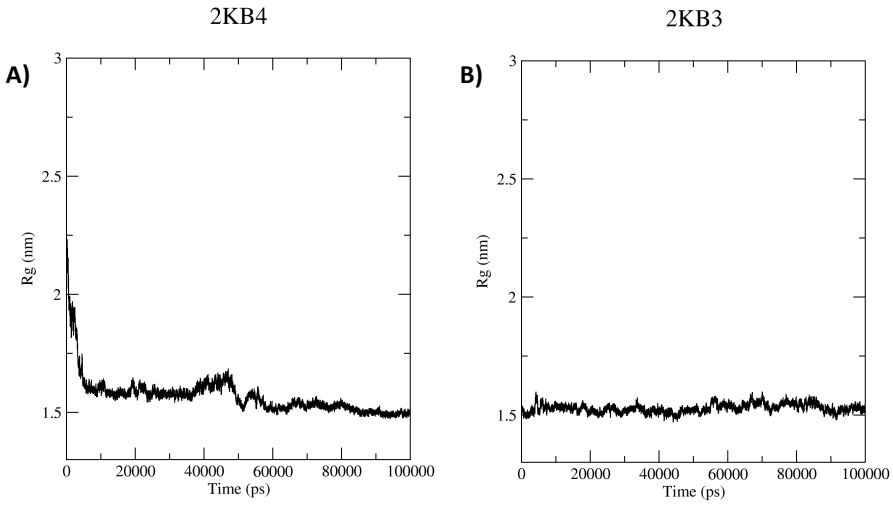

Fig. 3.18: Comparison of radius of gyration of structure 2KB4 and 2KB3. 2KB3 contains phosphorylated threonine - Tpo15.

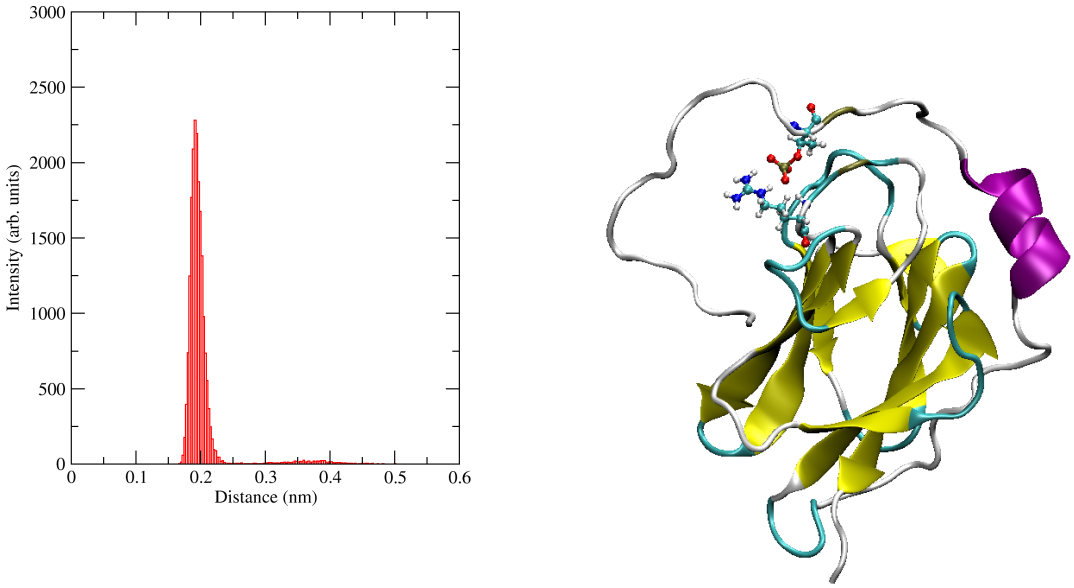

Fig. 3.19: Analysis of the salt bridge formed between residues Tpo15 and Arg87: Left - histogram of distances between residues Tpo15 and Arg87, right - structure of protein form 2KB3, residue Tpo15 and Arg87 are highlighted.

The RMSD of both structures also differs (see Fig. 3.20). The RMSD of phosphorylated structure is significantly lower than the RMSD of unphosphorylated form.

At the same time, the RMSD values of both structure are higher than in previous case. It is caused by flexible disordered part near the N-terminus of protein. Therefore, I calculated RMSD values only for the folded part of proteins (see Fig. 3.21). This RMSD values fluctuate around 3 and 4 Å. The folded part of protein is more stable during the simulation then the disordered part which contributes to the RMSD most.

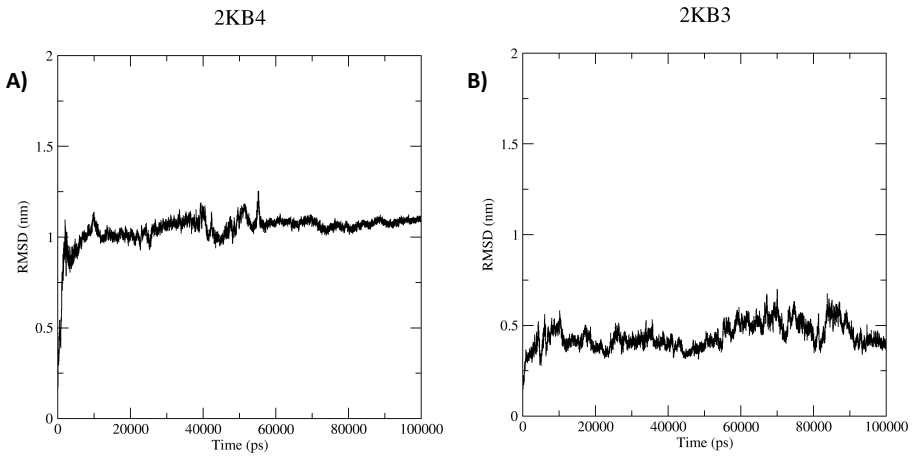

Fig. 3.20: Comparison of RMSD of 2KB4 and 2KB3 in course of simulation. The backbone of the proteins was used for superposition. Structure 2KB3 is phosphorylated on Thr15.

Fig. 3.22 depicts the RMSF for both structures averaged over each residue. In unphosphorylated structure 2KB4, the RMSF were high in all residues whereas in the phosphorylated structure 2KB3 the RMSF was significantly lower. The higher fluctuations were observed only in about twenty first residues corresponding to the more flexible disordered part of the protein.

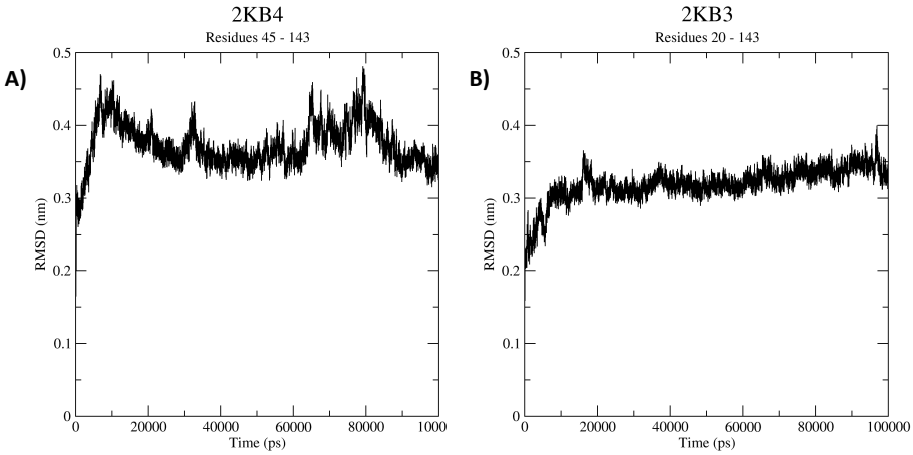

Fig. 3.21: The RMSD values calculated for the folded core part of proteins: Panel A - Structure 2KB4 without the first 44 residues, panel B - Structure 2KB3 without the first 19 residues.

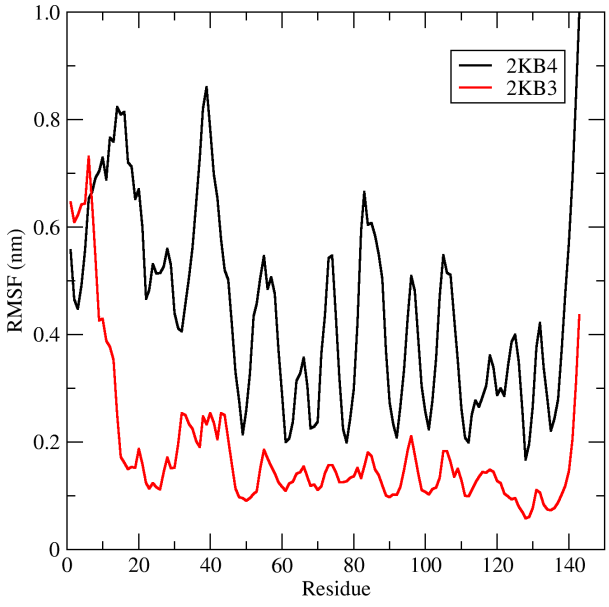

Fig. 3.22: Fluctuation of RMSD values in structures 2KB3 and 2KB4. Structure 2KB3 is phosphorylated on residue Thr15.

3.4.3 Mengovirus Leader Protein

The last studied protein was the Mengovirus Leader Protein in its three forms. Fig. 3.23 shows the radius of gyration of all this forms during the simulation. The values show that the unphosphorylated structure form 2MMH started from less compact structure, but it then reached similar radius of gyration as 2MML structure. The structure 2MML containing phosphorylation on Thr51 was stable during the whole simulation. The structure 2MMK containing two phosphorylated residue Thr51 and Tyr45 was not stable during the simulation and its radius of gyration highly fluctuated. The simulation length of 100 ns proper sampling was probably not sufficient.

The RMSD values of all three protein structures are high during the whole simulation. The initial structures used for simulations had only small stability during the simulation. It is consequence of the NMR origin of these structures.

When we compare the RMSF averaged over the residues (see Fig. 3.25), we see that the structure 2MML with one phosphorylated residue had the lowest fluctuations, whereas the fluctuations in the both phosphorylated structures were significantly higher.

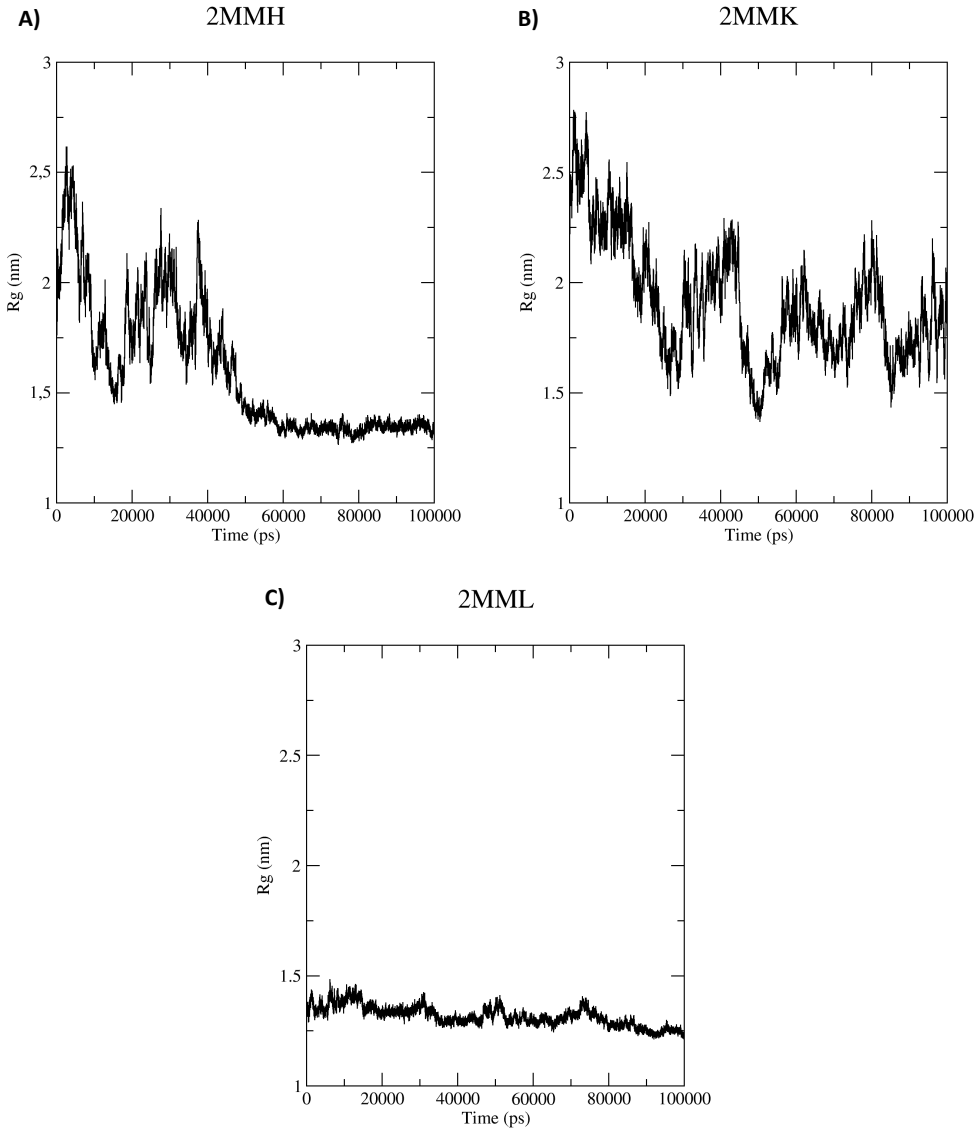

Fig. 3.23: Radius of gyration of 2MMH, 2MMK and 2MML protein structures. Structure 2MMH did not contain any phosphorylated residue, 2MML was phosphorylated on Thr51 and 2MMK had two phosphorylated residues - Thr51 and Tyr45.

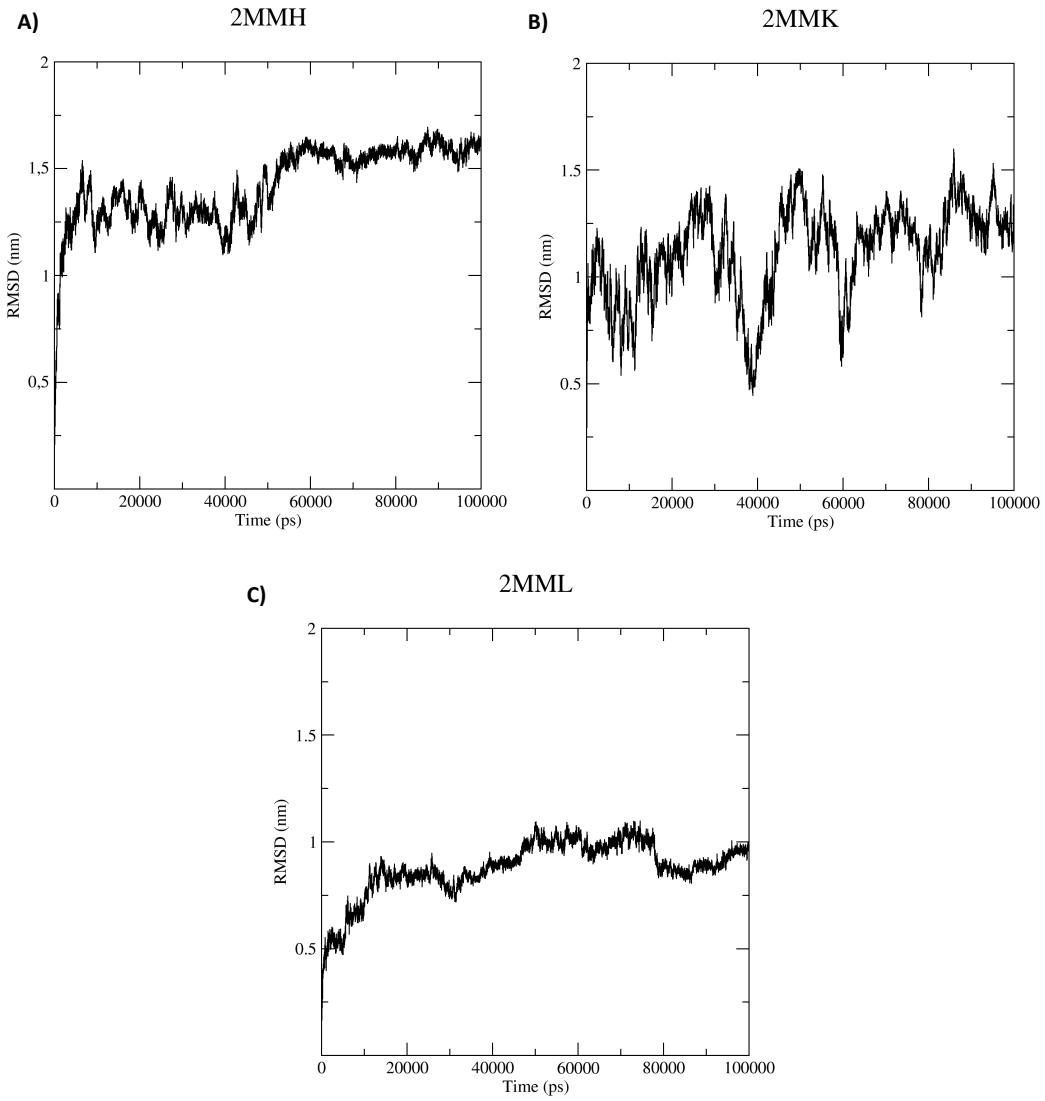

Fig. 3.24: Comparison of RMSD of 2MMH, 2MMK and 2MML.

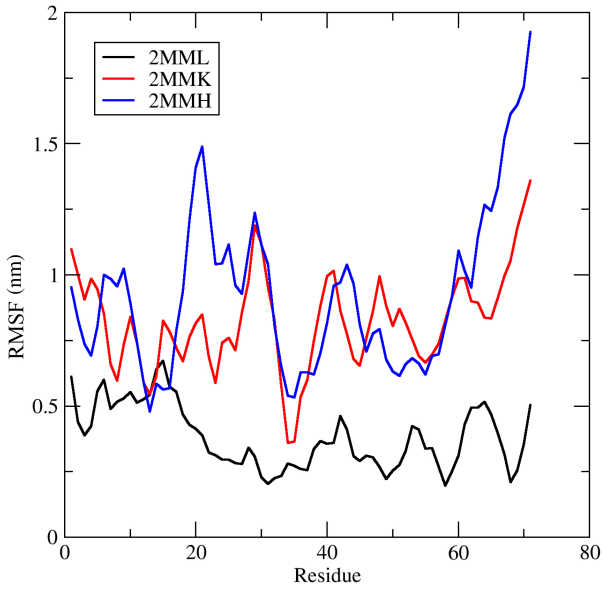

Fig. 3.25: RMSD fluctuations of 2MMH, 2MMK and 2MML. Structure 2MMH did not contain any phosphorylated residue. Structure 2MML was phosphorylated on Thr51 and structure 2MMK comprised phosphorylation on residues Thr51 and Tyr45.

Chapter 4

Conclusions and Outlook

Focus of this thesis was to map the effect of serine, threonine and tyrosine phosphorylation on the conformational intrinsic propensities in different model as well as in real systems. The phosphorylation was studied on systems starting from simple capped dipeptides, via the longer host-guest pentapeptides, peptides from disordered proteins and finally proteins with known structures.

First task was to calculate probability distribution for different conformations in simple dipeptides of serine, threonine, tyrosine and their monoanionic and dianionic phosphorylated forms. The changes in intrinsic propensities after phosphorylation were studied by bias exchange metadynamics.

In the serine dipeptide, the phosphorylation induced the changes in conformation preferences from PPII to α_R. However, our results are not fully compatible with the experimental data. According to the data published by Lee et al.[37], unphosphorylated serine dipeptide preferred mainly the PPII conformation and the phosphorylation did not lead to any significant changes. In the contrary, Kim et al.[33] showed that phosphorylation of serine led to the increase in propensity toward PPII conformation.

The larger stability and preferences for alpha helix conformation in our study can be probably assigned to the artificial effect of used force field. On the other hand, if this would be the case, this effect would influence also the conformation of unphosphorylated dipeptide towards helical conformation (as in work by Lee et al. 36) what we did not observe.

In the threonine dipeptides, the phosphorylation led to the changes of conformational preferences from the PPII and α_R conformation to the extended one. According to the experimental data presented by Lee et al.[37], threonine phosphorylation changed the conformation from PPII to β-strand. However, the β-strand region used in the work seems to partially overlap with Ext region defined in this thesis. The observed result thus agreed in both studies. Kim et al.[33] also presented changes of the intrinsic

conformational propensities of threonine from PPII to β-strand region.

The conformation of last studied tyrosine dipeptides were not influenced by phosphorylation notably. The most probable conformation of both tyrosine forms was PPII. The similar result were obtained by 2D ^{1}H NMR study of tetrapeptides GSXS by Tholey et al.[52].

Any influence of salt concentration on the phosphorylation effect was found. The changes of conformation preferences were the same in water with no ions as well as in solution with NaCl.

Formation of intramolecular hydrogen bonds with protons on the acetyl and N-methyl end of dipeptide was also studied. According to our results, serine dipeptides form the intramolecular hydrogen bond with the proton on the acetyl end of dipeptide whereas the dianionic phosphorylated serine did not form any hydrogen bonds. Monoanionic form of phosphorylated serine was not studied. This observation does not confirm experimental data obtained by Lee et al.[37] In their study, the phosphoserine created hydrogen bond with serine on the acetyl end of dipeptide.

In the threonine dipeptide, hydroxy group formed hydrogen bond with the proton on acetyl end of dipeptide. In the phosphorylated threonine, hydrogen bond between the phosphate group and the proton on N-methyl end of dipeptide was observed. Lee et al.[37] found phosphate group in phosphothreonine formed hydrogen bonds with protons on the both ends of dipeptide.

Tyrosine and phosphotyrosine dipeptide did not form any intramolecular hydrogen bonds.

The study on the capped dipeptides made of serine, threonine and tyrosine showed that the phosphorylation influences the conformation of amino acids at different extent. Whereas the conformation of serine and threonine was directly influenced by phosphorylation, tyrosine was almost untouched.

Second part of the thesis was focused on larger systems - the host-guest pentapeptides containing serine and threonine. I simulated pentapeptides in forms GGXGG and AAXAA by metadynamics method using the same simulation protocol as in previous case.

In the host-guest pentapeptides containing serine residue, the most probable conformations were α_R (for GGSerGG) and PPII (for AASerAA). After the phosphorylation, the intrinsic propensity systematically increased in favour of α_R region in all pentapeptides.

In the pentapeptides containing threonine, the preferred conformations were divided between α_R (for both GGThrGG and AAThrAA) and also PPII (AAThrAA). The phosphorylation of threonine then led to the increase in probability of extended

structure in all pentapeptides.

Therefore, we can conclude that the general trends observed in the pentapeptides were similar as in dipeptide study upon phosphorylation event.

Theoretical study on pentapeptides GGXGG done by He et al.[27], provided just opposite trend. Whereas the phosphorylation of serine in this study increased the probability for right-handed alpha helix region and decreased the preferences of PPII conformation, the results from He et al. showed that serine phosphorylation increases the probability of PPII and also right-handed alpha helix. The effect on the alpha helix conformation was however small. He reported that the phosphorylation of threonine led to significant increase of the preferences for alpha-helix conformation and decrease in PPII and β-strand (corresponding to the extended conformation). We found that the phosphorylation of threonine led to the increase of Ext conformation propensities. These both studies differed in the used method and force fields. In this thesis, metadynamics method with force field Amber99Sb-ILDN was applied whereas He et al. used replica exchange molecular dynamics with force fields Amberff10 and CHARMM36.

The next logical step was to study longer peptides extracted from real disordered proteins. Two parameters influencing the effect of phosphorylation on the peptide conformation was found in the chosen peptides containing serine phosphorylation site. If the serine phosphorylation site in a peptide was followed by amino acid different from proline, the phosphorylation increased the preferences for right-handed alpha helix conformation as observed in smaller systems. However, if the serine residue was followed by proline, the phosphorylation of serine did not influenced the conformational preferences in peptides or increased the probability of extended structure (β-strand). There is no agreement between our study and computational study on TSPI peptides presented by Hamelberg et al.[26] where the serine phosphorylation increased the alpha-helix conformation.

The last systems studied in this thesis were whole proteins in unphosphorylated and phosphorylated forms. These systems are more complex than short peptides and they also involves more interactions influencing the tertiary structure. In the studied proteins, we found that the phosphorylation increased the compactness of protein structures. However, the stable salt bridge (as one possible element of rigidity) between the phosphate group and arginine residue was observed only in 2KB3 structure. The stabilization in other proteins could be probably assigned to other factors.

The outlook of this thesis could be explaining the discrepancies between experimental studies and calculated data. It is possible to run simulations again with different force field. For example, the force field CHARMM36 contains parameters for phosphorylated amino acids and is also available for use in program Gromacs.

It is the fact that the available experimental data are not uniform in the determination of phosphorylation effect on the amino acid conformations. The major complication could be the fact that ϕ and ψ values used for definition of different conformations are also not described well and comparison of calculated data with experiment is thus difficult.

Another further work could also comprise the investigation of conformation temperature dependence. It would be also useful to repeat simulations of disordered peptides to obtain longer trajectories in more than one replica to get better statistics on the peptide conformations. The longer simulation would be suitable in the protein study as well.

The next step could be also the simulation of artificially phosphorylated proteins to get better comparison of the behaviour after the phosphorylation.

Last but not least, conformational study on the rest of amino acids which can be the object of phosphorylation e.g. histidine would be also interesting. However, the simulation of these amino acids could be probably more complicated because of the force field parameters required.

The important summary of the thesis is following. We found out significant effect of the phosphorylation on the conformational preferences on the studied amino acids. The phosphorylation influenced serine, threonine and tyrosine residues differently a this effect was transferable between dipeptides and pentapeptides. However this transferability is limited only for the small peptides. In the study of large systems is need of sampling comparable for all systems.

Bibliography

1. Metacentrum - virtual organization. https://metavo.metacentrum.cz/en/index.html.

2. Phosida: Posttranslational modification database. http://141.61.102.18/phosida/index.aspx.

3. Phosphositeplus. https://www.phosphosite.org/homeAction.action.

4. Rcsb pdb protein data bank. http://www.rcsb.org.

5. Uninett sigma2 as. https://www.sigma2.no/.

6. Uniprot. http://www.uniprot.org/.

7. Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., and Lindah, E. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX 1-2* (2015), 19–25.

8. Abraham, M. J., van der Spoel, D., Lindahl, E., Hess, B., and the GROMACS development team. Gromacs user manual version 5.0.7. http://www.gromacs.org, 2015.

9. Andrew, C. D., Warwicker, J., Jones, G. R., and Doig, A. J. Effect of phosphorylation on alpha-helix stability as a function of position. *Biochemistry 41*, 6 (2002), 1897–1905.

10. Bacot-Davis, V. R., Ciomperlik, J. J., Basta, H. A., Cornilescu, C. C., and Palmenberg, A. C. Solution structures of Mengovirus Leader protein , its phosphorylated derivatives , and in complex with nuclear transport regulatory protein , RanGTPase. *PNAS 111*, 44 (2014), 15792–15979.

11. Barducci, A., Bonomi, M., and Parrinello, M. Metadynamics. *WIREs Comput. Mol. Sci. 1*, 5 (2011), 826–843.

12. Barducci, A., Bussi, G., and Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett. 100*, 2 (2008), 020603.

13. Barthe, P., Roumestand, C., Canova, M. J., Kremer, L., Hurard, C., and Molle, V. Article Dynamic and Structural Characterization of a Bacterial FHA Protein Reveals a New Autoinhibition Mechanism. *Struct. Des. 17*, 4 (2009), 568–578.

14. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., and Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys. 81* (1984), 3684–3690.

15. Bonomi, M., Barducci, A., and Parrinello, M. Reconstructing the Equilibrium Boltzmann Distribution from Well-Tempered Metadynamics. *J. Comput. Chem. 30*, 11 (2009), 1615–1621.

16. Bussi, G., Donadio, D., and Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys. 126*, 1 (2007).

17. Bussi, G., Gervasio, F. L., Laio, A., and Parrinello, M. Free-Energy Landscape for Hairpin Folding from Combined Parallel Tempering and Metadynamics. *J. Am. Chem. Soc.*, 4 (2006), 13435–13441.

18. Byun, B. J., and Kang, Y. K. Conformational Preferences and Prolyl Cis-Trans Isomerization of Phosphorylated Ser/Thr-Pro Motifs. *Biopolymers 93*, 4 (2009), 330–339.

19. Cieśla, J., Fraczyk, T., and Rode, W. Phosphorylation of basic amino acid residues in proteins: Important but easily missed. *Acta Biochim. Pol. 58*, 2 (2011), 137–148.

20. Cohen, P. The regulation of protein function by multisite phosphorylation - A 25 year update. *Trends Biochem. Sci. 25*, 12 (2000), 596–601.

21. Cohen, P. The origins of protein phosphorylation. *Nat. Cell Biol. 4*, 5 (2002), E127–E130.

22. Darden, T., York, D., and Pedersen, L. Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems. *J. Chem. Phys. 98* (1993), 10089.

23. Ensing, B., De Vivo, M., Liu, Z., Moore, P., and Klein, M. L. Metadynamics as a tool for exploring free energy landscapes of chemical reactions. *Acc. Chem. Res. 39*, 2 (2006), 73–81.

24. Errington, N., and Doig, A. J. A phosphoserine-lysine salt bridge within an alpha-helical peptide, the strongest alpa-helix side-chain interaction measured to date. *Biochemistry 44*, 20 (2005), 7553–7558.

25. Frenkel, D., and Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications*. Academic Press, 2002.

26. Hamelberg, D., Shen, T., and McCammon, J. A. Phosphorylation effects on cis/trans isomerization and the backbone conformation of serine-proline motifs: Accelerated molecular dynamics analysis. *J. Am. Chem. Soc. 127*, 6 (2005), 1969–1974.

27. He, E., Yan, G., Zhang, J., Wang, J., and Li, W. Effects of phosphorylation on the intrinsic propensity of backbone conformations of serine/threonine. *J. Biol. Phys. 42*, 2 (2016), 247–258.

28. Hess, B., Bekker, H., Berendsen, H. J. C., and Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem. 18*, 12 (1997), 1463–1472.

29. Homeyer, N., Horn, A. H. C., Lanig, H., and Sticht, H. AMBER force-field parameters for phosphorylated amino acids in different protonation states: Phosphoserine, phosphothreonine, phosphotyrosine, and phosphohistidine. *J. Mol. Model. 12*, 3 (2006), 281–289.

30. Jalkanen, K. J., and Suhai, S. N-acetyl-L-alanine N'-methylamide: A density functional analysis of the vibrational absorption and vibrational circular dichroism spectra. *Chem. Phys. 208*, 1 (1996), 81–116.

31. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys. 79*, 2 (1983), 926.

32. Jungwirth, P. Klasická a kvantová molekulová dynamika. `http://marge.uochb.cas.cz/~jungwirt/`.

33. Kim, S. Y., Jung, Y., Hwang, G. S., Han, H., and Cho, M. Phosphorylation alters backbone conformational preferences of serine and threonine peptides. *Proteins 79*, 11 (2011), 3155–3165.

34. Laio, A., and Gervasio, F. L. Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Rep. Prog. Phys. 71*, 12 (2008), 126601.

35. Laio, A., and Parrinello, M. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci. USA 99* (2002), 12562.

36. Lee, K. K., Joo, C., Yang, S., Han, H., and Cho, M. Phosphorylation effect on the GSSS peptide conformation in water: Infrared, vibrational circular dichroism, and circular dichroism experiments and comparisons with molecular dynamics simulations. *J. Chem. Phys. 126*, 23 (2007), 1–15.

37. Lee, K. K., Kim, E., Joo, C., Song, J., Han, H., and Cho, M. Site-selective intramolecular hydrogen-bonding interactions in phosphorylated serine and threonine dipeptides. *J. Phys. Chem. B 112*, 51 (2008), 16782–16787.

38. Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R. O., and Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins 78*, 8 (2010), 1950–1958.

39. Lovell, S. C., Davis, I. W., Adrendall, W. B., de Bakker, P. I. W., Word, J. M., Prisant, M. G., Richardson, J. S., and Richardson, D. C. Structure validation by C alpha geometry: phi,psi and C beta deviation. *Proteins 50*, August 2002 (2003), 437–450.

40. Martoňák, R., Donadio, D., Oganov, A. R., and Parrinello, M. Crystal structure transformations in SiO2 from classical and ab initio metadynamics. *Nat. Mater. 5*, 8 (2006), 623–626.

41. Nezbeda, I., Kolafa, J., and Kotrla, M. *Úvod do počítačových simulací. Metody Monte Carlo a molekulární dynamiky.* Karolinum, Praha, 2002.

42. Parrinello, M., and Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys. 52*, 12 (1981), 7182–7190.

43. Piana, S., and Laio, A. A Bias-Exchange Approach to Protein Folding. *J. Phys. Chem. B 111* (2007), 4553–4559.

44. Ramachandran, G., Ramakrishnan, C., and Sasisekharan, V. Stereochemistry of polypeptide chain configurations. *J. Mol. Biol. 7*, 1 (1963), 95–99.

45. Ramachandran, G. N., and Sasisekharan, V. Conformation of Polypeptides and Proteins. *Adv. Protein Chem. 23* (1968), 283–437.

46. Ryckaert, J. P., Ciccotti, G., and Berendsen, H. J. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys. 23*, 3 (1977), 327–341.

47. Scarsdale, J. N., Van Alsenoy, C., Klimkowski, V. J., Schaefer, L., and Momany, F. A. Ab initio studies of molecular geometries. 27. Optimized molecular structures and conformational analysis of N.alpha.-acetyl-N-methylalaninamide and comparison with peptide crystal data and empirical calculations. *J. Am. Chem. Soc. 105*, 11 (1983), 3438–3445.

48. Schrödinger, LLC. The PyMOL molecular graphics system, version 1.8. November 2015.

49. Seavers, P. R., Lewis, R. J., Brannigan, J. A., Verschueren, K. H. G., Murshudov, G. N., and Wilkinson, A. J. Structure of the Bacillus Cell Fate Determinant SpoIIAA in Phosphorylated and Unphosphorylated Forms. *Structure 9*, 01 (2001), 605–614.

50. Shen, T., Wong, C. F., and McCammon, J. A. Atomistic Brownian dynamics simulation of peptide phosphorylation. *J. Am. Chem. Soc. 123*, 37 (2001), 9107–9111.

51. Steiner, T. The hydrogen bond in the solid state. *Angew. Chem. Int. Ed. 41*, 1 (2002), 49–76.

52. Tholey, A., Lindemann, A., Kinzel, V., and Reed, J. Direct Effects of Phosphorylation on the Preferred Backbone Conformation of Peptides: A Nuclear Magnetic Resonance Study. *Biophys. J. 76*, 1 (1999), 76–87.

53. Tribello, G. A., Bonomi, M., Branduardi, D., Camilloni, C., and Bussi, G. PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun. 185*, 2 (2014), 604–613.

54. Velazquez, H. A., and Hamelberg, D. Conformational selection in the recognition of phosphorylated substrates by the catalytic domain of human pin1. *Biochemistry 50*, 44 (2011), 9605–9615.

55. Vymětal, J., and Vondrášek, J. Metadynamics as a tool for mapping the conformational and free-energy space of peptides - The alanine dipeptide case study. *J. Phys. Chem. B 114*, 16 (2010), 5632–5642.

56. Vyměтal, J., and Vondrášek, J. A Critical Assessment of Current Force Fields. Short Peptide Test Case. *J. Chem. Theory Comput. 9* (2013), 441–451.

57. Vyměтal, J. Theoretical Study of the Conformational Behavior and Structural Preferences of Amino Acids in Peptides by Metadynamics. Master's thesis, Charles University in Prague, 2008.

58. Wang, Z. X., and Duan, Y. Solvation effects on alanine dipeptide: A MP2/cc-pVTZ//MP2/6-31G** study of (Phi, Psi) energy maps and conformers in the gas phase, ether, and water. *J. Comput. Chem. 25*, 14 (2004), 1699–1716.

59. Wong, S. E., Bernacki, K., and Jacobson, M. Competition between Intramolecular Hydrogen Bonds and Solvation in Phosphorylated Peptides: Simulation with Explicit and Implicit Solvent. *J. Phys. Chem. B 109* (2005), 5249–5258.

60. Xiang, S., Gapsys, V., Kim, H. Y., Bessonov, S., Hsiao, H. H., Möhlmann, S., Klaukien, V., Ficner, R., Becker, S., Urlaub, H., Lührmann, R., De Groot, B., and Zweckstetter, M. Phosphorylation drives a dynamic switch in serine/arginine-rich proteins. *Structure 21*, 12 (2013), 2162–2174.

Appendix A

Structures of peptides from disordered proteins

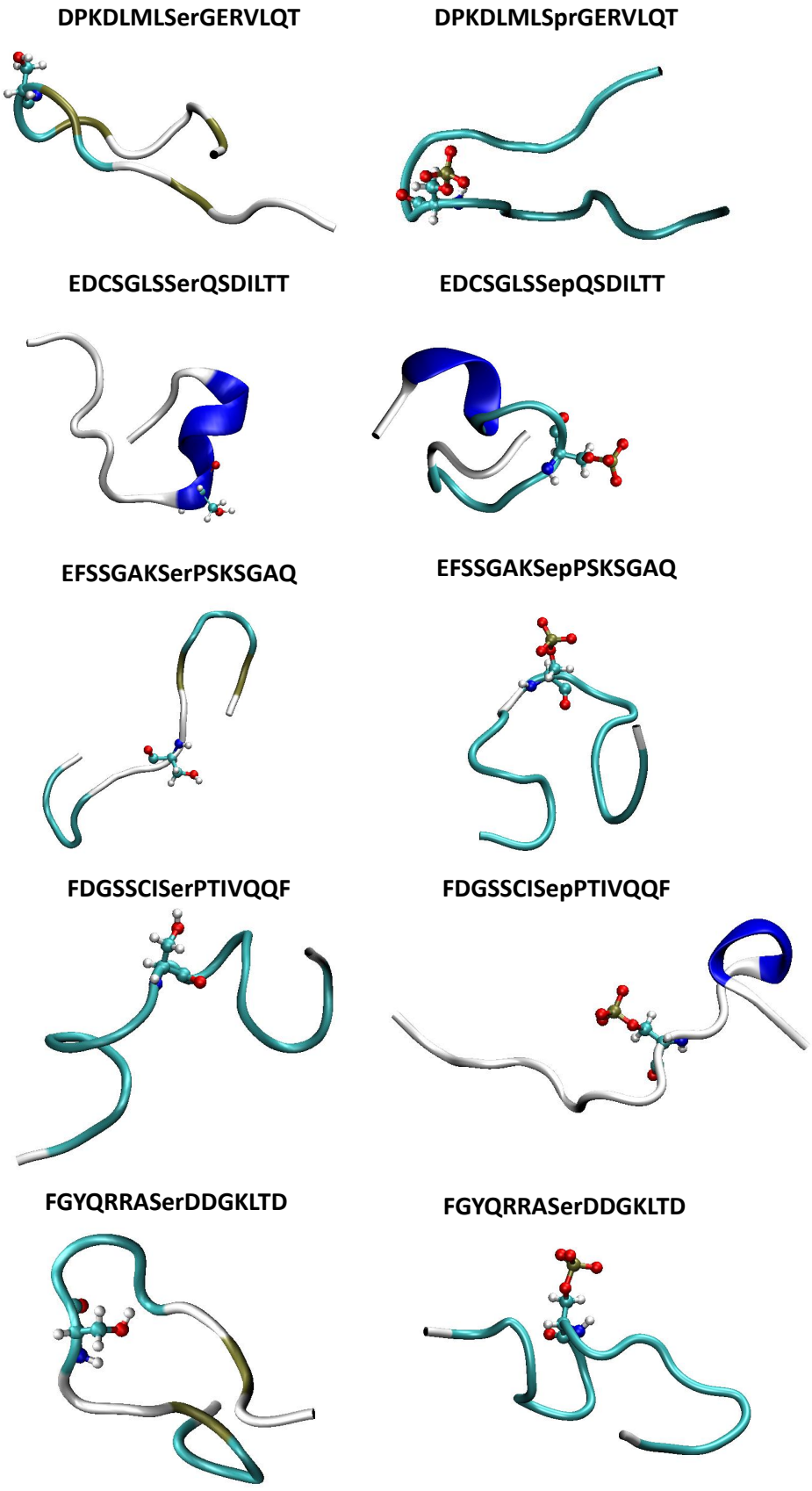

Fig. A.1: Structures of disordered peptides containing serine phosphorylation site in the middle of the chain.

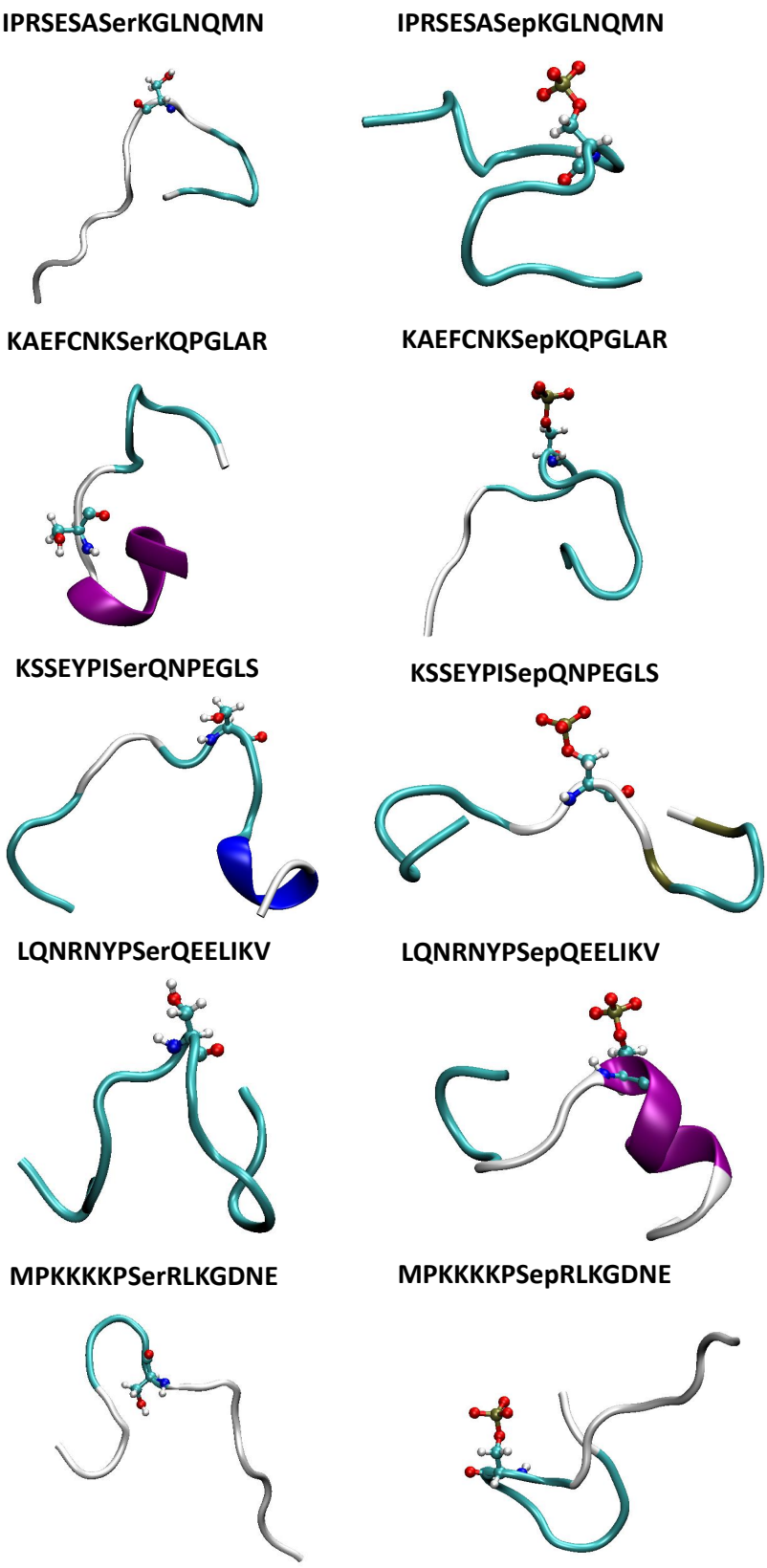

Fig. A.2: Structures of disordered peptides containing serine phosphorylation site in the middle of the chain.

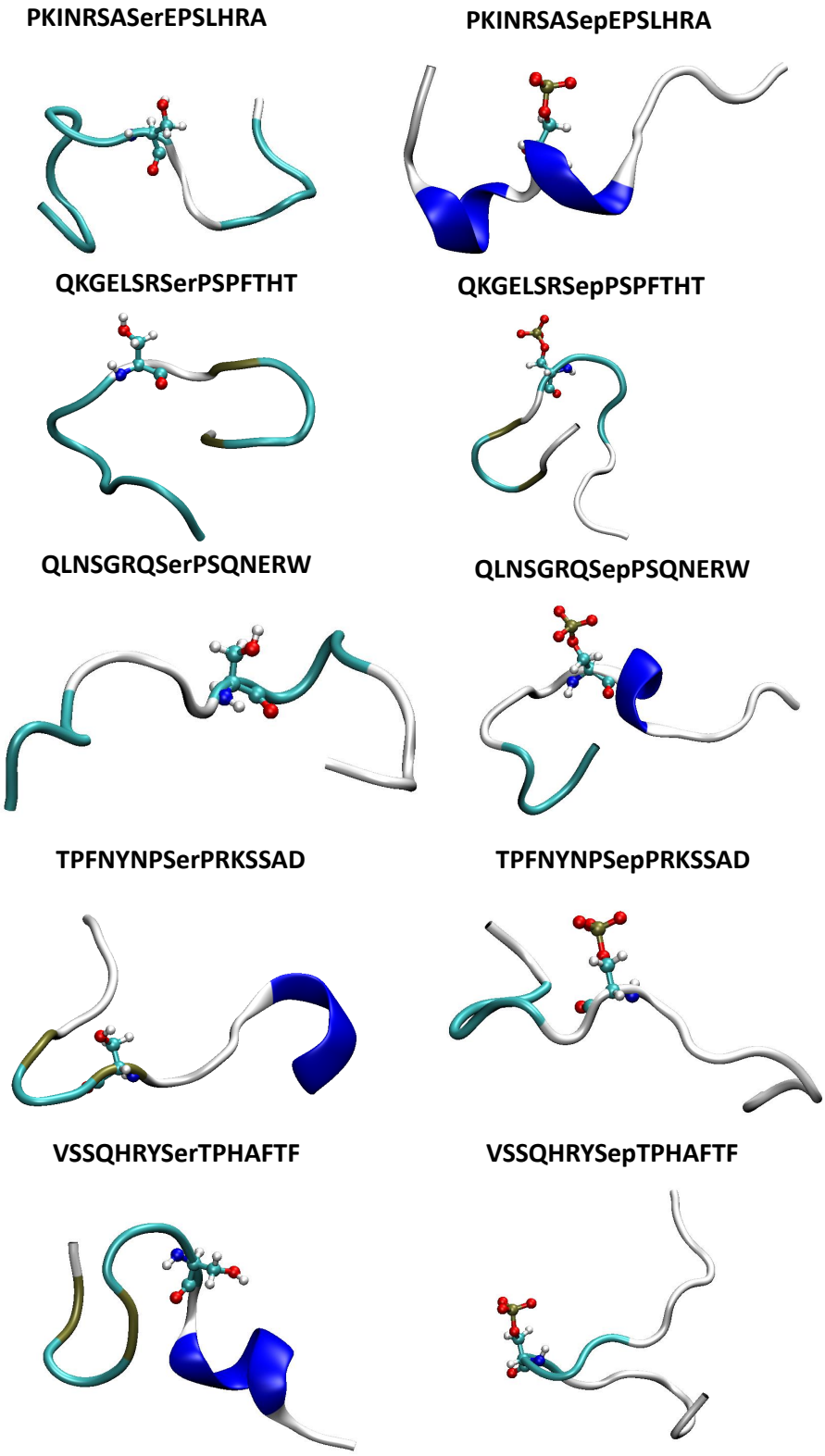

Fig. A.3: Structures of disordered peptides containing serine phosphorylation site in the middle of the chain.

Appendix B

Ramachandran diagrams and radii of gyration

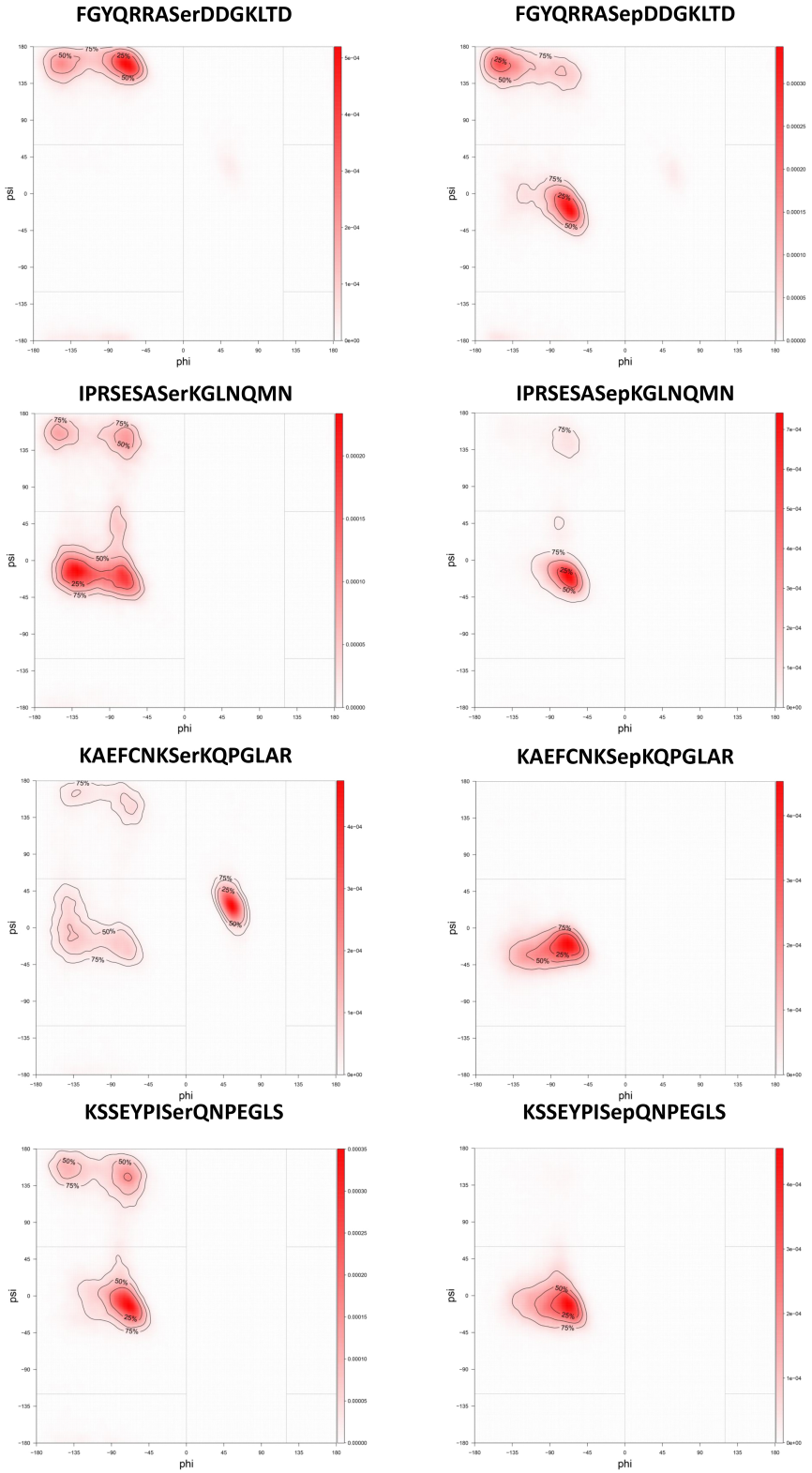

Fig. B.1: Ramachandran diagrams of disordered peptides: Left column - unphoshorylated forms, right column - peptides containing phosphorylated serine.

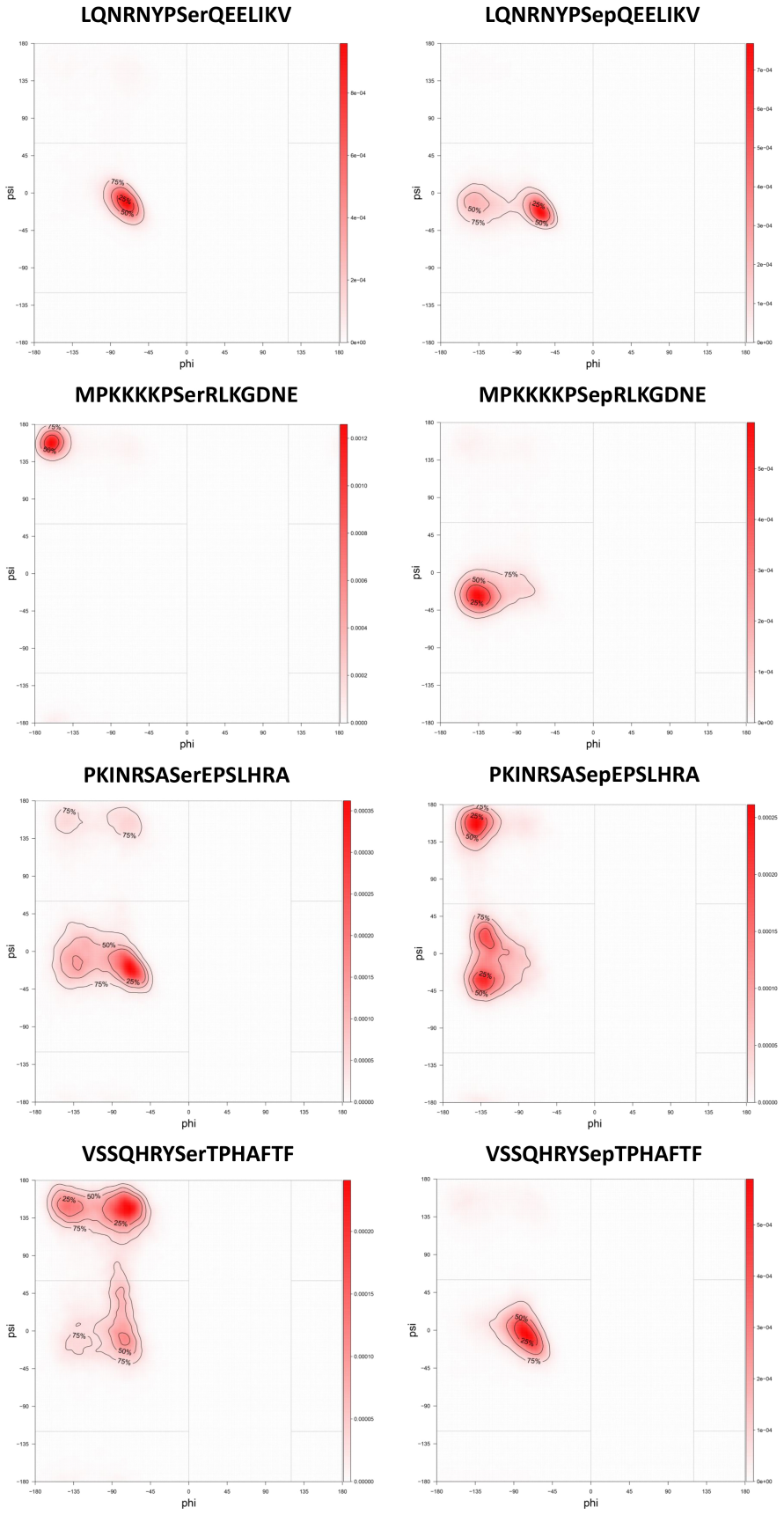

Fig. B.2: Ramachandran diagrams of disordered peptides: Left column - unphoshorylated forms, right column - peptides containing phosphorylated serine.

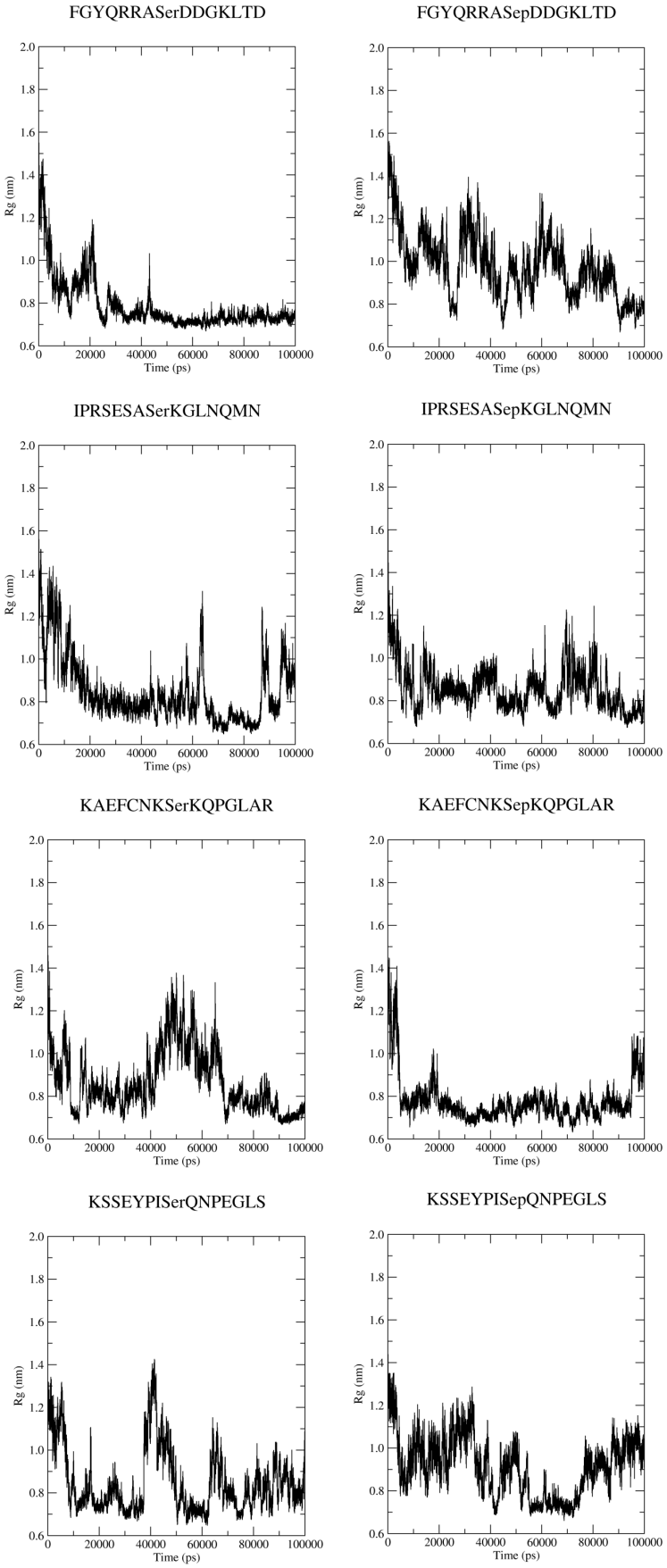

Fig. B.3: Comparison of radii of gyration calculated for disordered peptides: Left column - unphosphorylated peptides, right column - peptides containing phosphorylated serine.

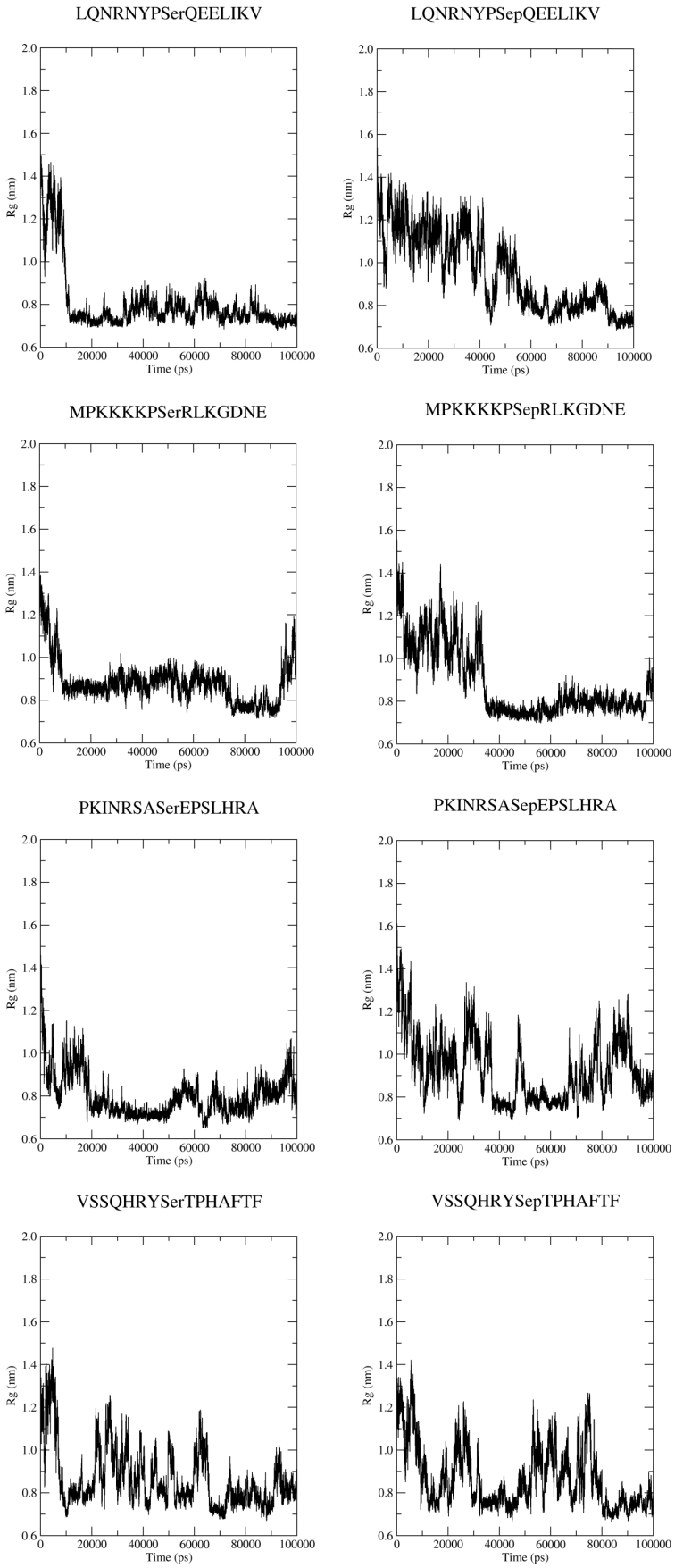

Fig. B.4: Comparison of radii of gyration calculated for disordered peptides: Left column - unphosphorylated peptides, right column - peptides containing phosphorylated serine.

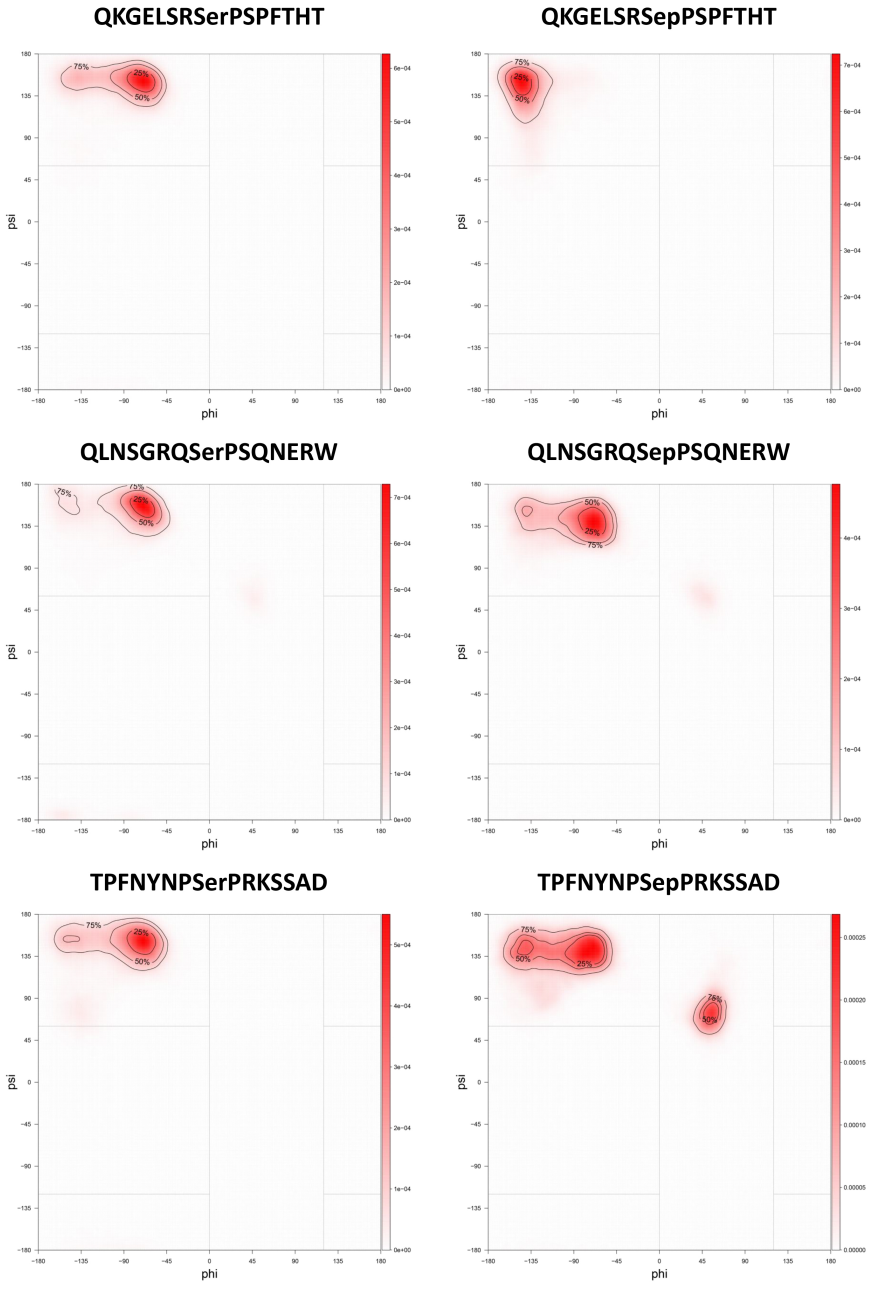

Fig. B.5: Ramachandran diagrams of disordered peptides with phosphorylation site containing Ser-Pro: Left column - unphoshorylated forms, right column - phosphorylated peptides.

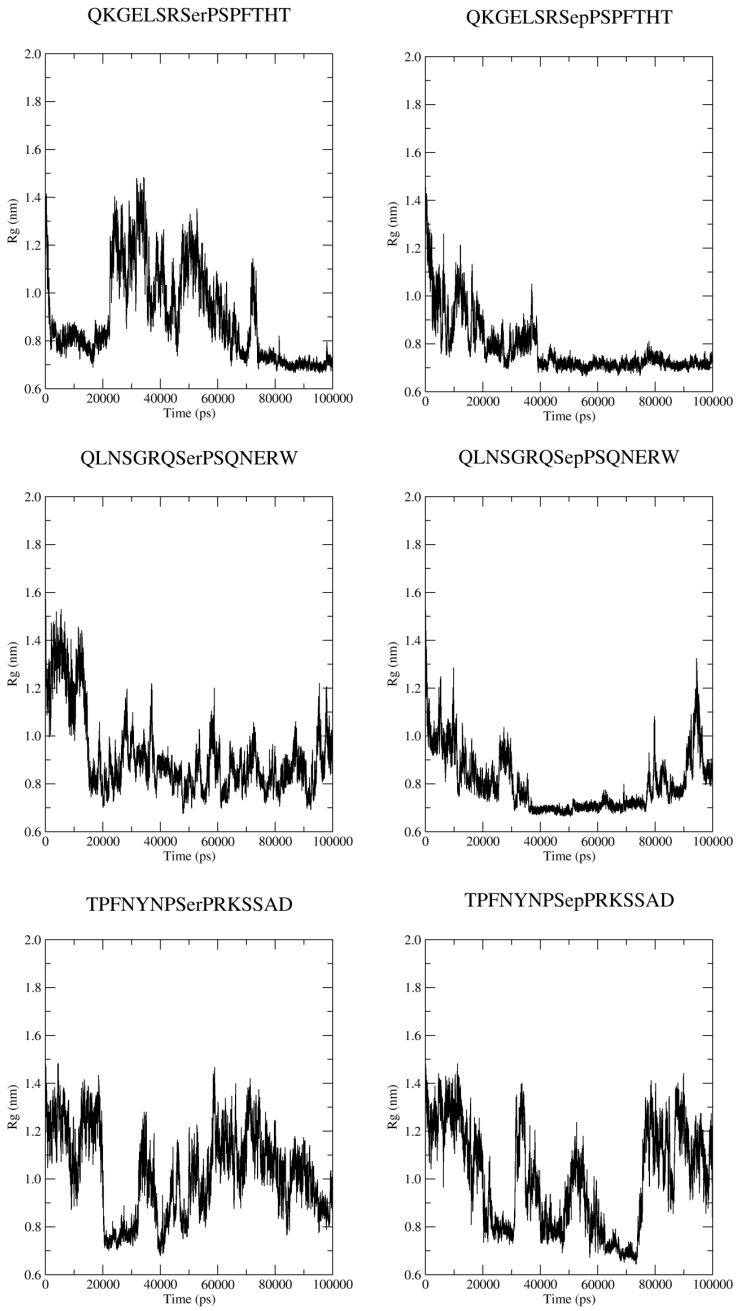

Fig. B.6: Comparison of radii of gyration calculated for disordered peptides woth phosphorylation site containing Ser-Pro: Left column - unphosphorylated peptides, right column - phosphorylated peptides.