

Molecular Dynamics of the disordered γ -tubulin carboxyl terminus

Carlos G. Oliver

Master of Science

Department of Biology

McGill University

Montreal, Quebec

December 12, 2016

A thesis submitted to McGill University in partial fulfillment of the requirements of
the degree of Master of Science

©Carlos G. Oliver, 2016

DEDICATION

To my grandparents; Hilda Fernandez and Porfirio Oliver. *Gracias por su apoyo, amor, y por ser mi inspiración en todo. Los quiero mucho.*

ACKNOWLEDGEMENTS

I give my sincere thanks to my supervisor Dr. Jackie Vogel for her support and mentorship throughout my Master's, and for sharing with me her passion for understanding the complexity of life. I thank my colleagues in the Vogel lab for their friendship, guidance and useful discussions. I would also like to acknowledge the generous financial support from the Cellular Dynamics of Molecular Complexes fellowship and Dr. Vogel's grants from the Canadian Institutes for Health Research and the National Institutes of Health. I thank the Guillimin support team for their invaluable technical help in working with the Calcul Québec compute clusters. I would like to thank Roman Sarrazin Gendron for editing the French abstract. I also thank Vladimir Reinhartz for his kind help with computational problems, and for all the enriching discussions.

I would like to express my deepest gratitude to all those whom I am so fortunate to call my friends and loved ones. To my brother, whose friendship means everything. To my father for his care and support. And to my mother; for being there for me every day, in every way; for being my role model, and my best friend.

CONTRIBUTION OF AUTHORS

All data collection and results obtained during my M.Sc. are presented here as a standard format thesis. This work includes results and figures obtained by collaborators in the Department of Chemistry Dr. Anthony Mittermaier and Jason Harris who conducted NMR experiments and generated all NMR based figures. All computer simulation work and MD figures were done by myself. Collection of tubulin primary sequences was done by Roman Sarrazin Gendron. All text and literature review in this thesis was done by myself with feedback from my supervisor Jackie Vogel and collaborator Anthony Mittermaier.

ABSTRACT

With recent advances in experimental and computational methods in structural biology, it is becoming increasingly clear that protein function is not only dependent on stable architectures, but equally so on the absence of well-ordered domains. These elements, also known as intrinsically disordered proteins (IDPs) or regions (IDRs) are protein chains that do not adopt organized three dimensional structures, but are highly functional nonetheless. Due to their flexible backbones, IDPs/IDRs are able to explore a large ensemble of conformations and functions. The cell harnesses this flexibility by targeting IDRs for post-translational modifications (PTMs) such as phosphorylation. Altering the conformational sampling of IDRs through PTMs, serves as a useful tool for functional control and signal integration. This paradigm is crucial for ensuring the proper execution of complex cellular processes that involve a large number of multi-tasking proteins that need to act in a coordinated manner. However, the physical mechanisms linking IDRs to functional output remain largely unknown. In this work, we study the role of phosphorylation in the assembly of the mitotic spindle, which effects chromosome segregation during cell division. Phosphorylation at Tyrosine 11 (Y11) in the intrinsically disordered C-terminus of γ -Tubulin (γ -CT) has been shown to control key aspects of the assembly of microtubules in the mitotic spindle. *in vivo*, the phosphomimicking mutation of Tyr to Asp (Y11D) leads to a temperature sensitive growth phenotype and an overbuilt spindle in budding yeast. Through Molecular Dynamics computer simulations (MD) and Nuclear Magnetic Resonance spectroscopy (NMR) we show that non-phosphomimicking and

phospho-mimicking states of the γ -CT primarily sample disordered and collapsed conformations. However, the phosphomimicking (Y11D) mutant undergoes switch-like collective motions to an extended but also disordered state. We propose that this transition serves to control the binding of proteins involved in shaping microtubule dynamics. This is the first observation of switch-like behaviour in IDRs/IDPs where both states are disordered, making this a novel physical mechanism of control with the potential to regulate cellular function.

ABRÉGÉ

De récents progrès dans les méthodologies expérimentales et informatiques en biologie structurelle démontrent que la fonction des protéines ne dépend pas uniquement de leur architecture tridimensionnelle stable, mais aussi de l'absence d'une telle architecture. Les éléments dépourvus de structure, appelés protéines ou régions intrinsèquement désordonnées (PID, RID), peuvent explorer une vaste gamme de conformations et de fonctions. La cellule exploite cette flexibilité en ciblant les RID avec des modifications post-traductionnelles (MPT) comme la phosphorylation. L'altération de l'échantillonage conformationnel des RID par les MPT est un outil pour le contrôle fonctionnel et l'intégration des signaux cellulaires. Ce paradigme est essentiel pour assurer l'exécution correcte des processus cellulaires qui impliquent un grand nombre de protéines multitches devant agir de manière précise et coordonnée. Cependant, les mécanismes physiques qui établissent un lien entre les RID et les résultats fonctionnels sont en grande partie inconnus. Cette thèse présente une étude sur le rôle de la phosphorylation dans l'assemblage et l'organisation du fuseau mitotique, la machine moléculaire qui effectue la ségrégation des chromosomes durant la division cellulaire. La phosphorylation de l'acide aminé Tyrosine 11 (Y11) à son extrémité carboxy-terminale (γ -CT) intrinsèquement désordonnée contrôle des aspects essentiels de l'assemblage du fuseau mitotique. La mutation de cette tyrosine à un acide aspartique phosphomimétique *in vivo* (Y11D) provoque une sensibilité à la température et des défauts dans l'organisation des microtubules du fuseau mitotique. En utilisant des simulations informatiques de dynamique moléculaire et la résonance

magnétique nucléaire, nous montrons que les formes non-phosphomimétiques et phosphomimétiques du γ -CT balayent toutes deux principalement des conformations compactes et désordonnées, mais que la mutation Y11D entreprend des mouvements collectifs occasionnels vers une conformation mineure étendue, elle aussi désordonnée. Nous suggérons que cette transition peut servir à contrôler la liaison avec les protéines qui affectent la dynamique des microtubules. Il s'agit de la première observation d'une transition organisée dans un ensemble de structures désordonnées, un nouveau mécanisme physique de contrôle ayant le potentiel de réguler le fonctionnement cellulaire.

TABLE OF CONTENTS

DEDICATION	ii
ACKNOWLEDGEMENTS	iii
CONTRIBUTION OF AUTHORS	iv
ABSTRACT	v
ABRÉGÉ	vii
LIST OF FIGURES	xi
KEY TO ABBREVIATIONS	xiii
1 Introduction	1
1.1 Disorder in proteins	3
1.2 Physical mechanisms of IDR function in cellular machines	6
1.3 IDP function in the mitotic spindle	10
1.3.1 Microtubules	10
1.3.2 γ -Tubulin & the γ -CT	11
1.4 Experimental question	13
1.5 Approach	13
2 Theory & Methods	14
2.1 Molecular Dynamics Simulations	15
2.1.1 Computing atomic trajectories	15
2.1.2 Force Field	16
2.1.3 Preparing the system for a simulation	19
2.2 Trajectory Analysis	21
2.2.1 Root Mean Square Deviation	21
2.2.2 Radius of gyration	21
2.2.3 Diffusion coefficient and Hydrodynamic Radius	22

2.2.4	Covariance Analysis	23
3	Conformational analysis of the γ -Tubulin carboxyl terminus	26
3.1	NMR	27
3.2	MD simulation set-up	31
3.3	Conformational sampling of γ -CT	33
3.3.1	Collective motions correspond to transitions between extended and collapsed conformations	43
3.3.2	Whole protein simulations and conserved properties of γ -CT	49
3.4	Discussion	51
4	Conclusions	59
4.0.1	Summary of Findings	59
4.0.2	Future Aims	59
	References	61

LIST OF FIGURES

<u>Figure</u>		<u>page</u>
1–1 γ -Tubulin 3D Structure		4
2–1 Polyproline Simulation		19
3–1 NMR Diffusion Measurements		29
3–2 NMR Dispersion Measurements		30
3–3 NMR Secondary Structure Assignments		31
3–4 NMR Chemical Shifts		32
3–5 γ -CT Secondary Structure Assignments		35
3–6 RMSD		36
3–7 γ -CT predicted conformational ensemble		38
3–8 Distribution of diffusion coefficients		40
3–9 Time series of conformational sampling of γ -CT under Molecular Dynamics Simulations.		41
3–10 Contact Maps of Extended and Collapsed Ensembles		44
3–11 Eigenvalues of γ -CT PCA		46
3–12 Principal Component Analysis		47
3–13 Extreme projection along opening-closing axis.		48
3–14 YD Globular Contact Map		50
3–15 γ -Tubulin whole protein simulation		50
3–16 Multiple Sequence Alignment of γ -CT across eukaryotes		52

3–16 Full γ and β Tubulin CT Multiple Sequence Alignments	54
3–17 IDP Conformational Landscapes	56
3–18 γ -TuRC	58

KEY TO ABBREVIATIONS

D_t : Translational Diffusion Coefficient

R_g : Radius of Gyration

γ -CT: γ -Tubulin C-Terminus

γ -TuRC: γ -Tubulin Ring Complex

μ s: Microsecond

CBP: CREB Binding Protein

CREB: Cyclic AMP Response Element Binding Protein

DNA: Deoxyribonucleic Acid

GRIPS: γ -Tubulin Ring Proteins

GROMACS: GROnigen Machine for Chemical Simulatoins

IDP: Intrinsically Disordered Protein

IDR: Intrinsically Disordered Region

KID: Kinase Inducible Domain

MD: Molecular Dynamics

NMR: Nuclear Magnetic Resonance

NPT: Isothermal-isobaric Ensemble

ns: Nanosecond

NVE: Microcanonical Ensemble

NVT: Canonical Ensemble

OPLS-AA: Optimized Potential for Liquid Simulations - All Atom

PTM: Post-Translational Modification

RMSD: Root Mean Square Displacement

SPCE: Extended Single Point Charge

VMD: Visual Molecular Dynamics

WT: Wild-Type

YD: Y445D (Y11D)mutant

CHAPTER 1

Introduction

Everything existing in the universe
is the fruit of chance and necessity.

Democritus

Molecular machines are assemblies of proteins and associated molecules that work together in a coordinated manner to solve a biological problem. These biological problems, such as coordinating chromosome segregation during cell division, regulation of cell cycle timing, execution of protein synthesis, etc. encompass many processes essential to life. Given the necessity for survival in the face of ever changing environments, evolution has produced a large diversity of molecular mechanisms for solving these problems in a flexible yet robust manner. A biological machine that only functions properly under a narrow range of conditions is less likely to support the life of a single individual or a population. Therefore, at the core of each of those processes are highly complex networks of proteins that are able to assemble, communicate, coordinate, self-regulate and self-correct in order to accomplish the necessary task reliably. For example, the vital process of DNA replication is executed by a large multitude of proteins that each contribute to the process of copying the genome [4]. The replication machinery must first read environmental cues for initiating replication at the correct time. Meanwhile, complex combinatoric signaling networks ensure that DNA replication unfolds in a processive manner, enzymatic

components perform physical work to unwind DNA strand for copying, and others communicate with the DNA repair machinery to correct copying errors and avoid harmful mutations. It is clear that solving the biological problem of DNA replication requires the ability of participating proteins to interact with many different partners and mediate many different processes.

The structure and function of each protein is encoded in its unique sequence, or chain, of amino acids. Physical interactions between amino acids give rise to a specific 3D arrangement of the protein chain, also known as structure. The structure of each protein allows for specific interactions between proteins to assemble molecular machines, recruit necessary factors and mediate chemical reactions. See **Fig. 1–1** for a visualization of protein structure. Since the 1950s when the first X-ray crystallography protein structure was solved [33], we have learned a great deal about how 3D architecture and conformational sampling of the chains give rise to protein function. X-ray crystallography accesses atomic-scale conformations of folded protein domains, allowing us to infer that coordinated motions between structural conformations is the main element of control in protein function [19]. For example, X-ray crystallography experiments have shown that the activity of Calmodulin, an important signalling protein, is modulated by conformational changes of its α -helix domains brought about by binding of Ca^{2+} ions [43]. These conformational rearrangements initiate a clamping motion of the helical domains which allow Calmodulin to bind to its downstream targets. However, it is important to note that X-ray crystallography only offers static pictures of protein structure, and provides information mostly on the spatial arrangement of relatively large and stable domains. It therefore became

a long standing dogma that the stable 3D folds of a protein chain dictate a protein's function, also known as the "one structure - one function" paradigm.

However, as we saw with DNA replication, a single protein is often required to fulfill many functions, and interact with various different partners. It is therefore unlikely that such large scale and consequently slow structural motions can account for all of the precise and rapid control we observe in biological systems. A static description of proteins is not sufficient to explain the degree of functional flexibility and control that we observe. This leaves us with several questions. If one structure means one function, how can the same protein fulfill multiple functions and engage in many different interactions? How can molecular machines offer such precise control of functionality while counting only on a static architectures? The broad aim of this thesis is therefore to improve our understanding of the physical mechanisms underlying the functional complexity of molecular machines.

1.1 Disorder in proteins

In recent years, it has been recognized that functional plasticity can be found in regions of proteins that do not adopt stable architectures, also known as intrinsically disordered regions (IDRs). While IDRs are highly flexible and largely unstructured, functional studies have shown that they are necessary for many cellular processes [69].¹ Instead of relying on a structure for function, IDR functionality lies the

¹ Some works make the distinction between IDP and IDR where an IDR is an intrinsically disordered region and IDP is a fully disordered protein.

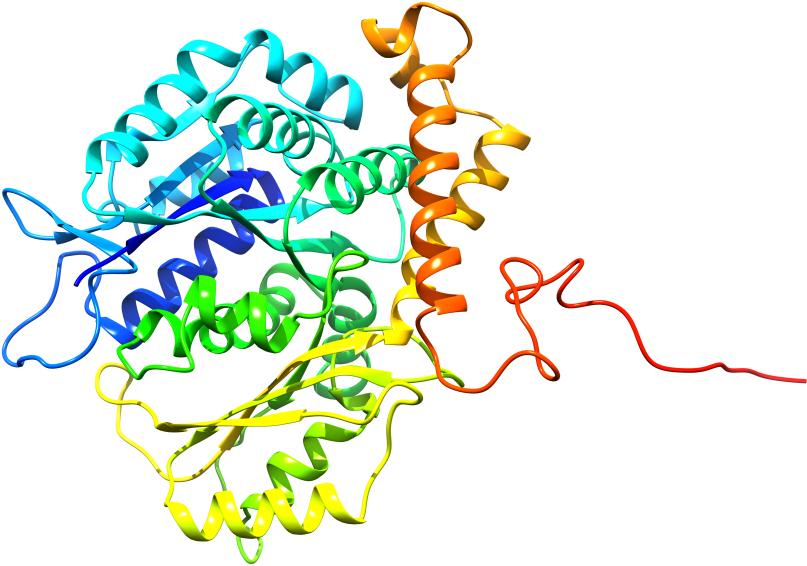


Figure 1–1: γ -Tubulin 3D Structure

Visual representation of the yeast γ -Tubulin protein based on the human γ -Tubulin structure derived from X-ray crystallography [1]. The protein is composed of a highly ordered globular domain stabilized by various α -helices and β -sheet domains and measures 2.20 nm in radius of gyration. In red, we see the disordered γ -CT region which was absent in the crystal structure and was modeled by RaptorX [28].

absence of structure. This flexibility offers the protein rapid access to a vast pool of conformations with which to fine-tune and diversify its function.

The study of IDRs is relatively new to structural biology. This is largely due to the fact that the main tool being used for structural biology in the past decades, X-ray crystallography, fails to detect patterns in unfolded chains, making it difficult to study highly dynamic elements in protein such as IDRs. Unstable protein domains such as IDRs that sample many conformations produce averaged out electron scatter patterns that cannot be interpreted [51]. Techniques that do produce information on dynamics, such as Nuclear Magnetic Resonance (NMR) only developed for proteins

until 1984 [70], 25 years after the first structure was solved by X-ray crystallography in 1958 [33]. Another approach for studying protein dynamics with atomic resolution is through computational simulation. Physical models of proteins whose motions are computed *in silico* have been shown to provide important information regarding detailed dynamics of biomolecules [30]. However, until recently, these techniques, such as Molecular Dynamics (MD), were greatly limited by shortcomings in computer power. However, with large advances in experimental and computational techniques in recent years, we have been able to study the dynamic properties of IDPs in great detail, and have found that they play key roles in the control of molecular machines.

Over 15,000 proteins in the Protein Data Bank have been predicted to contain long disordered regions [54]; it is therefore not surprising that IDPs have also been implicated in a multitude of cellular processes and disease states [61]. Interestingly, it has been shown that viral proteins use IDPs in their proteins to hijack cellular proteins and use the flexibility of IDPs to mimic host proteins and recruit host cellular machinery in order to propagate [10]. This would suggest that viral proteins use IDPs to make efficient use of their smaller genomes and obtain a greater range of function from the limited number of proteins in their genomes. It is now clear that IDPs, through their lack of structure, are an important adaptive feature that drive the functional complexity and robustness we observe in molecular machines.

1.2 Physical mechanisms of IDR function in cellular machines

In this section we will give a brief account of some of the physical mechanisms of IDR function that have been described in the literature.

Phosphorylation

A key aspect of dynamic control is the ability to modulate function in a precise and reversible manner. The cell needs to be able to induce and inhibit interactions in a time and space dependent manner. To solve this problem, the cell harnesses the structural malleability of IDRs/IDPs by coupling these elements with post translational modifications, most commonly, phosphorylation. Phosphorylation is the reversible covalent addition of a phosphate group effected by a protein kinase, which carries a negative charge to a tyrosine, serine or threonine amino acid. The reverse reaction is catalyzed by enzymes called phosphatases which remove the phosphate group. The addition of a phosphate group introduces the potential inter and intra-molecular hydrogen bonding which alters the electrostatic environment of the IDP/IDR [62]. This change can in turn bias the stochastic conformational sampling of the IDP/IDR in a particular direction. Because phosphorylation is reversible, it acts as a means for driving structural switching which can then be used to modulate a large range of interactions [34, 53, 49]. Not surprisingly, it has been seen in many studies that IDPs are prime targets for phosphorylation [24]. The use of phosphorylation as an information carrier has been described in various cellular systems. For example, in cell cycle control, there is a strong need for a specific temporal sequence of interactions to be enforced [71]. In such a case, the sequential

phosphorylation of a single target modifies the affinity for the same target to the next target in the pathway, ensuring that interactions take place in an ordered manner [69]. Phosphorylation can also be used to enforce thresholds, where in order to avoid the negative impact of accidental interactions, certain interactions will be blocked until an IDP/IDR has achieved a certain number of phosphorylations [46]. Having accumulated enough phosphorylations, a structural rearrangement favours the interaction.

Disorder-Order transitions

The best explored physical mechanism of IDP function is the fold-on binding paradigm [68]. In this case, IDPs in the free form are unstructured, and when they encounter their binding target, they undergo a folding transition (disorder to order) to form a stable complex. The lack of structure in the unbound state allows the IDP/IDR the necessary flexibility to recognize multiple targets, and it allows binding to be inducible instead of constitutive. A well studied example of this kind of mechanism is the binding of the transcriptional activator protein CREB and its co-activator CBP [15]. An IDR in CREB known as KID mediates binding to CBP where upon binding, the IDP folds into a pair of helices. However, this binding process is not

favoured spontaneously due to a high entropic barrier ². However, when the KID is phosphorylated, the phosphoryl group interacts with CBP by forming hydrogen bonds which result in a negative enthalpic change that compensates for the loss of entropy and thus makes the folding reaction favourable. Because of the inducible nature of this interaction, CBP is able to also interact with other co-factors, which has been reported in the literature [52]. This is an example of how even though the association state of the IDP is ordered, it is the intrinsic disorder and entropy of the unbound state which is the key for controlling the interaction.

² Protein folding implies an ordering of the protein backbone which typically results in a loss of conformational entropy to the system and is therefore energetically disfavoured. This is known as the entropic barrier; $\Delta G = \Delta H - T\Delta S$, where ΔH is the change in enthalpy, or heat energy available to the system, and ΔS is the entropy, or degree of disorder/conformational freedom available to the system. However, if the folding process releases sufficient heat (ΔH), the loss in entropy is overcome and the folding reaction is energetically favoured; $\Delta G < 0$. The release in heat by the folding of the protein typically increases the entropy of the surrounding water molecules thus preserving the second law of thermodynamics. It should also be noted that not all binding interactions are driven only by conformational entropy. Phenomena such as the hydrophobic effect can drive interactions through the entropic component of the solvent surrounding macromolecules.

Fuzzy interactions

Disorder to order transitions are not necessary for IDPs to confer functionality. There are a growing number of examples where IDPs/IDRs are involved in functional interactions while remaining in a disordered state [59]. Such interactions have been labeled with the term ‘fuzzy’ as they maintain a heterogenous conformational ensemble throughout their lifetimes. There exist various physical mechanisms by which fuzziness, or disorder, in binding interactions confers advantages to protein function. For example, binding interactions between an IDP/IDR and a target protein where the IDP/IDR is able to form alternate contacts with its binding target can help reduce the entropic cost of binding, as well as control the accessibility of different sites on the protein for modulating interactions with different targets [17, 14]. IDPs/IDRs can also play a role in interactions without making direct contacts with the binding partner by acting as flexible linkers for folded domains [5], or as ‘antennae’ for [56] recruiting further interactiors and stabilizing the binding of folded domains through long range interactions [74, 72]

It is becoming increasingly evident that nature has harnessed disorder as an adaptive mechanism for control in protein-function. High flexibility in protein conformational state allows switch-like control over interactions and activity, fine-tuning the kinetics of interactions, precise signal integration, controlled multiple partner binding, etc. It is due to these biophysical properties that the cell is able to carry out its complex tasks with such robustness and precision. Advances in this field have caused us to reconsider the ‘one structure, one function’ paradigm that has prevailed in structural biology for decades. However, this remains a relatively novel area

of structural biology, and there still remain many unsolved physical mechanisms in IDPs/IDRs.

1.3 IDP function in the mitotic spindle

In this section we will address the role of IDRs in controlling the function of the mitotic spindle. The mitotic spindle is a complex molecular machine composed of microtubules, force generators, chromosomes, and numerous effector molecules which act in a coordinated manner to accomplish the process of chromosome segregation during cell division [31]. This process ensures that genetic material is transferred from the mother to the daughter cell in a timely manner, and without errors which would in most cases result in fatality. Because cell division is a fundamental task in every cell’s lifetime, mitotic spindles share common design features throughout eukaryotes [38]. The task of properly arranging and segregating chromosomes is effected ultimately by hollow 25 nm filaments composed of polymerized tubulin, known as microtubules, which attach to chromosomes and exert forces to move the DNA into mother and daughter cells [35]. In order to study the underlying mechanisms at play, we work with the mitotic spindle of the budding yeast *Saccharomyces cerevisiae* due to its minimal, yet highly conserved construction [38].

1.3.1 Microtubules

Microtubules are constructed of alternating pairs, or heterodimers, of the globular proteins α and β tubulin. A cylindrical arrangement of tubulin dimers results in the formation of a microtubule, whose rigidity can be adapted in a length dependent manner. By adding or losing subunits, microtubules can increase and decrease in length, and push or pull directly on targets. Microtubules also bind proteins that link

them to chromosomes, to vesicles, and even to each other when forming microtubule bundles. [12]. Microtubules have a number of other functions, such as serving as roadways along which transport molecules can carry cargo, and form an adaptable cytoskeleton that contributes to cell shape and movement. The spontaneous assembly of tubulin dimers in solution into a microtubule is heavily disfavoured. However, when free floating tubulins encounter pre-formed nucleus or microtubule seed, the growth of a microtubule is greatly facilitated [26]. In cells, this template is known as the γ -Tubulin Ring Complex (γ -TuRC) which is a ring-like assembly of γ -Tubulin molecules held together by various other proteins known as γ -tubulin ring proteins (GTRPs) [37]. γ tubulin shares a similar structure to α and β tubulin and have been shown to act as nucleation templates for microtubules when assembled in a ring complex [37].

1.3.2 γ -Tubulin & the γ -CT

Microtubule nucleation was thought to be the sole function of γ tubulin for many years. However, recent evidence from budding yeast suggests that γ -tubulin might have additional functions in controlling microtubule properties [65, 64, 9, 47]. γ -tubulin bound to the spindle poles is phosphorylated in vivo at 8 sites [64, 32]. Several functional studies following up on the finding that γ -tubulin is regulated show that mutations altering phosphorylation sites, all of which lie in IDRs, have consequences on the organization and stability of microtubules but no effect on their nucleation. This opens a relatively unexplored field of functional coupling whereby the γ -TuRC acts not only as a microtubule nucleator, but also as a signal integration hub for regulating downstream events in microtubule organization.

One phosphorylation site in γ -Tubulin that has an important role in spindle function is the highly conserved tyrosine (Y) 445 which lies in the disordered carboxyl terminal tail of γ -Tubulin (γ -CT). The γ -CT is defined as the final 35 residues in the C-terminal portion of γ -tubulin, which lies outside of the folded globular domain. The γ -CT is essential for survival in budding yeast [65]. Substitution of an aspartic or glutamic acid (D/E) residue in the place of Y445 results in slow growing cells with unstable and misaligned mitotic spindles [64]. The Y445D/E substitutions are used as a means for constitutively mimicking the electrostatic environment of a phosphate group by introducing a negative charge. Defects in spindle function observed in these mutants suggest that phosphorylation is limited to a specific stage of the cell cycle and/or subset of molecules. However, very little is known about the interactions and physical mechanisms that phosphorylation of Y445 may control. The coupling of post-translational modifications (PTMs) to the C-terminal tails of tubulins is well described in the γ -tubulin orthologues α and β tubulin. Specific combinations of PTMs on the α and β tubulin tails act as a ‘tubulin code’ for selectively recruiting motor proteins and microtubule associated proteins to the microtubule lattice. A similar code has not yet been described for γ -tubulin despite evidence that it is regulated *in vivo*. While the functional importance of phosphorylation and IDRs in γ -tubulin is becoming increasingly clear, the physical mechanisms by which local modifications at IDRs can have global impacts on the large molecular machine remain unstudied.

1.4 Experimental question

We hypothesize that phosphorylation of the γ -CT IDR is a key event in the regulation of microtubule dynamics which allows cells precise control over the building of the mitotic spindle. Moreover we propose that such control is achieved via the local addition of negative charge modulates the global dynamics and conformational sampling of the γ -CT.

1.5 Approach

In order to study changes in conformational sampling of the γ -CT, we use a powerful computational technique known as Molecular Dynamics (MD) simulations. We will be simulating the dynamics of two forms of the γ -CT: WT, and Y11D (Y445 in the full protein). We will perform simulations on the γ -CT in isolation as well as in the presence of the entire γ -Tubulin protein. Analysis of MD simulations will be guided and validated by NMR experiments on the same system performed by collaborators.

CHAPTER 2

Theory & Methods

Life can only be understood
backwards; but it must be lived
forwards.

Søren Kierkegaard

The main technique I will use to study the behaviour of IDPs/IDRs is the computational technique of Molecular Dynamics (MD) simulations [18]. Protein dynamics are shaped by various types of physical interactions, acting on many conformational degrees of freedom, and on timescales spanning femtoseconds to milliseconds. This level of complexity makes it very challenging to predict the dynamics, or compute ensemble quantities of IDPs *ab initio*. MD is a brute-force approach which addresses this problem by iteratively solving the equations of motion for every interaction in a system of atoms in 3D space. What results is a trajectory and a velocity for every atom in the system in time, which we can use to visualize the conformational sampling of our IDP of interest and compute thermodynamic quantities. While this can be a computationally demanding task, it is currently the most reliable way of studying protein conformational sampling *in silico* and has been successfully applied to many biomolecular systems [30]. We will use this approach to study the conformational dynamics of two isoforms of the γ -CT: WT, and Y11D.

2.1 Molecular Dynamics Simulations

We represent our system as a set of N atoms represented as vectors $R = \{\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N\}$ in three dimensional space. We then use classical Newtonian mechanics to obtain the changes in position of the atoms as a function of time. For a peptide in solution, this system would consist of the atoms in the peptide, ions, water atoms, and the forces arising from interactions between them.

2.1.1 Computing atomic trajectories

MD is centered on the principle that the potential energy arising from interacting particles is a function of their positions in space. Given a function describing the potentials arising from interactions between the different atoms, which we call a force field, we can iterate through every atom in the system and calculate resulting forces as a function of potential energy. The potential energy given by the force field can be written as $V(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)$ and is a function of the positions of each atom. Using the classical definition of force as $\mathbf{F} = m\mathbf{a}$, we can combine the positions of each atom with the force field to compute the force acting on each atom as follows.

$$\mathbf{F}_i = -\frac{\partial V(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_i, \dots, \mathbf{r}_N)}{\partial \mathbf{r}_i} \quad (2.1)$$

Given that the force on an atom is the result of interactions with all other atoms in the system, we obtain the force on a particular atom as the sum of the force of the interactions with all other atoms j in the system, $\mathbf{F}_i = \sum_j \mathbf{F}_{ij}$. Given the total force on an atom, we can compute its trajectory in space by numerically integrating Newton's equations of motion. This process is repeated and trajectories are stored and updated for the desired number of steps in the simulation.

$$\frac{\partial^2 \mathbf{r}_i}{\partial t^2} = \frac{\mathbf{F}_i}{m_i} \quad (2.2)$$

2.1.2 Force Field

The functions for potential energy of every type of interaction in the system are defined in what we call a force field. The energy between two interacting atoms can be broken down into two broad types of interactions: bonded and non-bonded interactions.

$$E_{total} = E_{bonded} + E_{nonbonded} \quad (2.3)$$

The bonded energy term can be written as the sum of energies arising from the bond itself (E_{bond}) which is a function of the bond length, the potential arising from the angle formed by the bond (E_{angle}), as well as the torsional/dihedral angle ($E_{dihedral}$) arising from the rotation of three bonds about two intersecting planes.

Non-bonded interactions can have two contributing factors; electrostatic force, and van der Waals force. The electrostatic potential ($E_{electrostatic}$) arises from the interaction of the charges of particles, while the van der Waals potential ($E_{vanderWaals}$) arises from the attraction or repulsion between uncharged groups. Combining all of these terms, we can write the full description of forces in the system as:

$$E(\mathbf{r}_1, \dots, \mathbf{r}_N) = E_{bond} + E_{angle} + E_{dihedral} + E_{vanderWaals} + E_{electrostatic} \quad (2.4)$$

The MD algorithm evaluates $E(r_N)$ at every time step to obtain the force on each atom, and therefore the trajectory at each time step. Given this low-level description of the system, more complex phenomena such as the hydrophobic effect and hydrogen bonding which are known to be essential to protein dynamics do not need to be coded explicitly in the models. Instead, they arise naturally from the definition of the system.

Another key component to the force field is the definition of parameters for the different types of interactions and particles in the system. Key parameters include values for charge, mass, bond length, etc. and are obtained from experimental measurements. The force field must naturally also contain a set of definitions for the various types of atoms and functional groups it can model. Therefore, the choice of force field can have important consequences on the outcome of the simulations and must be chosen with care.

In this work, we will be using the Optimized Potentials for Liquid Simulations - All Atom (OPLS-AA) force field [25] that can be represented as:

$$E_{bond} = \sum_{bonds} K_r (\mathbf{r} - \mathbf{r}_0)^2 \quad (2.5)$$

$$E_{angle} = \sum_{angles} k_\theta (\theta - \theta_0) \quad (2.6)$$

$$E_{dihedrals} = \sum_{\text{dihedrals}} \left(\frac{V_1}{2} [1 + \cos(\phi - \phi_1)] + \frac{V_2}{2} [1 - \cos(2\phi - \phi_2)] + \frac{V_3}{2} [1 + \cos(3\phi - \phi_3)] + \frac{V_4}{2} [1 - \cos(4\phi - \phi_4)] \right) \quad (2.7)$$

Where ϕ is the dihedral angle and V_i are the coefficients in the Fourier series.

Non-bonded energies are computed as follows:

$$E_{nonbonded} = \sum_{i>j} f_{ij} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{C_{ij}}{r_{ij}^6} + \frac{q_i q_j e^2}{4\pi\epsilon_0 r_{ij}} \right) \quad (2.8)$$

Where A and C represent combining rules which allow us to obtain the interaction energy of dissimilar non-bonded atoms. OPLS uses standard combining rules where $A_{ij} = \sqrt{A_{ii}A_{jj}}$ and $C_{ij} = \sqrt{C_{ii}C_{jj}}$ [16].

We believe the OPLS-AA force field to be the best fit for modeling charge-dependent motions of the γ -CT. Whereas other popular force fields such as CHARMM were parametrized on X-ray crystallographic experimental values of folded globular proteins, the OPLS force field was optimized with quantum chemical calculations of charged short peptides [39]. This approach is likely to more accurately model the dynamics of unfolded peptides where other force fields may tend to favour the formation of collapsed stable secondary structure motifs [20, 60]. Furthermore, it has been reported that the small and localized treatment of charged groups in OPLS-AA is well suited for systems where local charge interactions drive global dynamics [63]. We show that a simulation under OPLS-AA is able to accurately produce an extended conformation for the ‘molecular ruler’ polyproline [55] **Fig. 2–1**. We

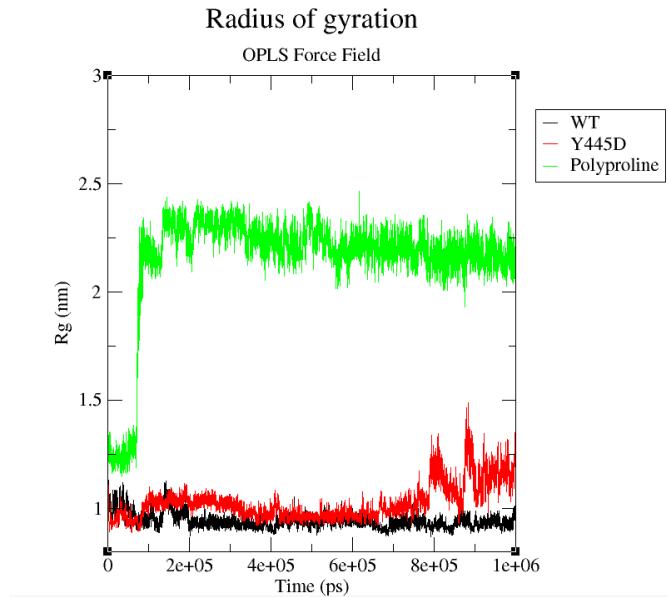


Figure 2–1: Polyproline Simulation

Simulation of 39-mer polyproline peptide known to be fully extended [55] behaves as expected and accesses high radius of gyration conformations in a $1\mu s$ simulation using the OPLS-AA force field.

also control the appropriateness of the force field by comparing MD values to NMR measurements on the same system.

2.1.3 Preparing the system for a simulation

The starting point of an MD simulation is a force field and a set of initial atomic coordinates for the system of interest. Before a simulation can be successfully run, there are several pre-processing steps that must be executed.

Hydrogen bonding and the hydrophobic effect play very important roles in shaping the dynamics of polypeptides therefore, the model must include water molecules. We place the peptide atoms in a simulated box under periodic boundary conditions where water molecules are introduced to fill the remaining space. All subsequent

force calculations in MD will consider solvent-solvent and solvent-peptide atomic interactions. We use an explicit water regime, meaning that all solvent atoms are modeled as discrete units in the system. While faster alternatives to this paradigm which represent the solvent with mean field behaviour, known as implicit solvent models are available, it is well documented that an explicit treatment currently provides results with the highest accuracy [48, 3, 73].

Once the peptide is solvated, any initial steric clashes between atoms must be allowed to relax. Typically this involves executing an energy minimization algorithm which searches for atomic coordinates that minimize the forces between atoms to move the system towards an energy minimum. No minimization algorithm guarantees convergence to a global minimum in finite time on a realistic system. However, convergence to a local minimum is often sufficient to eliminate significant clashes. In this work, we use the steepest descent energy minimization algorithm implemented in GROMACS [22].

At this point, we could begin an MD simulation and obtain trajectories in the NVE ensemble (constant number of particles, volume, and energy). However, we are often interested in comparing results from MD to experimental measurements such as those from NMR where the system is under constant temperature and pressure. It is therefore necessary to ensure that the forces in the system do not produce large fluctuations in the pressure and temperature of the ensemble. In order to keep the temperature constant and achieve an NVT sampling (constant number of particles, volume, and temperature) we use a thermostat. Since the temperature of a system is a function of the kinetic energy, a thermostat re-scales the velocities of

the atoms in the system to achieve a given temperature. Likewise, for maintaining constant pressure, a barostat adjusts the size of the simulation box to counteract fluctuations in pressure and thus achieving an NPT ensemble (constant number of particles, pressure, and temperature). During the equilibration step, we first let the system adjust to the desired temperature by executing a short simulation in NVT. Then under NPT we allow the system to adjust to the desired pressure. Once both equilibration simulations are complete, the system is ready for a full simulation.

2.2 Trajectory Analysis

The MD simulation generates a set of coordinates for every atom in the system as a function of time, $\mathbf{r}(t)$. From these trajectories we can compute several quantities to study conformational changes in the peptide over time.

2.2.1 Root Mean Square Deviation

We measure the square displacement between the coordinates of atom i at time t weighted by the mass of the atom m_i . We iterate this process for every atom in the peptide to obtain a measure of the degree of change between two conformations in time weighted by atomic masses in matrix M .

$$\text{RMSD}(t_1, t_2) = \left[M^{-1} \sum_{i=1}^N m_i \|\mathbf{r}_i(t_1) - \mathbf{r}_i(t_2)\|^2 \right]^{\frac{1}{2}} \quad (2.9)$$

2.2.2 Radius of gyration

The radius of gyration is a measure of a structure's compactness. To obtain the radius of gyration, we compute the mean squared distance from the position vector \mathbf{r}_i to the molecule's centre of mass r_{mean} .

$$R_g(\mathbf{r}) = \sqrt{N^{-1} \sum_{k=1}^N (\mathbf{r}_k - \mathbf{r}_{\text{mean}})^2} \quad (2.10)$$

2.2.3 Diffusion coefficient and Hydrodynamic Radius

Like radius of gyration, diffusion coefficient (D_t) and hydrodynamic radius (R_h) is a proxy for the compactness of a macromolecule. However, D_t and R_h are quantities that describe the size of the molecule in the context of their solvent. Because biomolecules perform all of their function in solution, and are shaped by their interactions with the solvent, these quantities are often more informative than R_g for our purposes.

The translational diffusion coefficient of a macromolecule is defined as the rate at which its center of mass is able to diffuse through a solvent of a given viscosity under a certain hydrodynamic model. Conformations with high diffusion rates experience rapid displacement of their center of mass, while conformations with greater viscous force with the solvent experience reduced diffusion coefficients.

The process of computing the translational diffusion coefficient for a particular conformation is done by the software package `hydroNMR` [11]. The method of calculating the diffusion coefficient will not be discussed here in detail as it is beyond the scope of this work. The main concept is that the software models each atom in the system, in our case a trajectory obtained by MD, as a spherical bead. The resulting chain of beads is packed into a hexagonal lattice and internal beads are removed to extract a topology of residues exposed to the solvent. From this topology, the software calculates the frictional force that a given conformation would exert on

the solvent, this quantity is contained in the translational friction tensor Ξ . The following expression gives us the translational diffusion tensor \mathbf{D}_t .

$$\mathbf{D}_t = k_B T \Xi_t^{-1} \quad (2.11)$$

Where k_B is the Boltzmann constant ($1.380 \times 10^{-23} \text{ m}^2 \text{ kg s}^{-2} \text{ K}^{-1}$) and T is the temperature in Kelvin. The trace of the translational diffusoin tensor is invariant regardless of the molecule's orientation and is thus used to define the translational diffusion coefficient D_t .

$$D_t = \frac{1}{3} \text{tr}(\mathbf{D}_t) \quad (2.12)$$

Knowing D_t , we can use the Stokes-Einstein equation to obtain an expression for the effective hydrodynamic radius which measures the diffusion of spherical particles in solvent.

$$R_h = \frac{k_B T}{6\pi\eta D_t} \quad (2.13)$$

Where R_h is the hydrodynamic radius and η is the viscosity of the solvent.

2.2.4 Covariance Analysis

When analyzing MD trajectories, we are often interested in observing coordinated, or correlated motions. This is because global motions are likely to be involved in some functional mechanism. However, molecular trajectories typically feature complex motions along many axes and time scales which can often make it difficult

to detect coordinated motions. Local rearrangements, vibrations, rotations, and random diffusion are examples of non-coordinated motions that likely do not contribute to a functional mechanism. The goal in MD trajectory covariance analysis is to obtain the axes of motion where atoms in the peptide of interest show a high degree of correlation which could be indicative of a global coordinated motion.

Covariance analysis, or principal component analysis is a mathematical tool which isolates principal axes, or components of correlated motion by computing the covariance between atoms at every time point in the simulation. We compute the covariance for pairs $r_i, r + j$ of N atoms in 3 dimensions resulting in a covariance matrix of size $3N$. M is again a matrix of atomic weights.

$$C_{ij} = \left\langle M_{ii}^{\frac{1}{2}}(\mathbf{r}_i(t) - \langle \mathbf{r}_i(t) \rangle) M_{jj}^{\frac{1}{2}}(\mathbf{r}_j(t) - \langle \mathbf{r}_j(t) \rangle) \right\rangle \quad (2.14)$$

The eigenvectors of the covariance matrix, C define the set of orthogonal axes along which maximize variance. Note that $\langle \cdot \rangle$ denotes a time average. Due to the constraints imposed by the backbone, only a limited number of eigenvectors are expected to contribute most to global movements.

$$R^T C R = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_{3N}) \quad \text{where} \quad \lambda_1 \geq \lambda_2 \geq \lambda_{3N} \quad (2.15)$$

Where R is the transformation matrix whose columns contain an eigenvector. Using this matrix to diagonalize C , we get a diagonalized C containing the set of eigenvalues λ_i for every eigenvector in R along its main diagonal. The magnitude of the eigenvalue tells us the amount of variance captured by its corresponding eigenvector and can thus be used to guide our projection toward the major axes of motion.

If we wish to visualize motions along a particular axis and filter out motions along other axes, we can project the coordinates of each atom along a specific eigenvector. We can use following transformation to obtain the new set of coordinates $\mathbf{p}(t)$.

$$\mathbf{p}(t) = R^T M^{\frac{1}{2}} \mathbf{r}(t) \quad (2.16)$$

The resulting trajectory lets us visualize motions along any component and is a useful tool for detecting coordinated structural changes.

CHAPTER 3

Conformational analysis of the γ -Tubulin carboxyl terminus

Only in disorder are we conceivable.

Roberto Bolaño

In this chapter we discuss the impact of phosphorylation on the global dynamics of the γ -Tubulin C-terminus (γ -CT). We begin with the finding that the γ -CT has been shown to be phosphorylated *in vivo* [32], and mimicking constitutive phosphorylation of a conserved tyrosine residue in the γ -CT perturbs the function of the mitotic spindle [64]. We hypothesize that transient phosphorylation at the γ -CT IDR is acting to regulate an as of yet unknown aspect of the function of the spindle via a structural mechanism. More specifically, we will be studying the highly conserved tyrosine 11 (Y11) of the γ -CT where we study the IDR which has been identified as a key phospho-site in this system [64]. This is a first step towards describing a molecular mechanism by which regulation of the γ -CT promotes proper spindle function and the identification of key residues involved in binding molecular partners.

In order to understand the effect of phosphorylation at Y11, we took two approaches. First, NMR experiments which provide critical structural information; including translational diffusion coefficients, fast motions at nsec timescales that correspond to sidechain dynamics as well as slow motions of the backbone that

correspond to dispersion, and the timescale of exchange between states. Collectively, these NMR experiments are used to probe the conformational sampling of non-phosphorylated and phosphomimetic forms of the γ -CT. These experiments revealed slow motions and exchange between two states that are only observed in the Y11D mutant. Thus molecular dynamics simulations were employed to gain deeper insight regarding the observed conformational sampling. Because phosphorylation is a challenge to implement controllably *in vivo* and *in vitro*, we approximate the electrostatic effects of a phosphorylation by introducing an acidic residue, Aspartic Acid (D), at the phosphorylation site through site directed mutagenesis. We therefore study the wild-type sequence, WT: LLRGAAEQDSYLDLVDDENMVGELEEDLDADGDHKLV, and the phospho-mutant Y11D: LLRGAAEQDSLDDVLVDDENMVGELEEDLDADGDHKLV using MD simulations. By comparing results from our simulations to experimental measurements previously performed with NMR spectroscopy on the γ -CT, we are able to propose that local changes in charge at the 11 residue in the polypeptide modulate the conformational sampling of the γ -CT. The changes in conformational sampling we observe point to a physical mechanism for regulating the availability of binding surfaces on γ -Tubulin. Furthermore, we describe a novel mode of IDP control whereby functionality arises from switch-like transitions that lie entirely within disordered states.

3.1 NMR

NMR experiments were done by collaborators in the Department of Chemistry at McGill on the WT and YD forms of the γ -CT. Through protein NMR spectroscopy we are able to obtain accurate *in vitro* measurements of the structural and dynamic

properties of polypeptides. Among these are secondary structure state, dynamic conformational changes, and structural properties such as diffusion rates. Given that the γ -CT is intrinsically disordered and therefore highly dynamic, this makes NMR a particularly well suited to this problem. Diffusion measurements show that the major conformation of both forms corresponds to that of a collapsed polypeptide **Fig. 3–1**. By computing the effective hydrodynamic radius of the polypeptides using the Stokes-Einstein equation 2.13 we can compare the compactness of the γ -CT to other known structures. For WT we obtain $D_t = 1.25 \times 10^{-6} \pm 1 \times 10^{-8} \text{ cm}^2 \text{s}^{-1}$, $R_h = 14.2 \text{ \AA} \pm 0.2$ and $D_t = 1.224 \times 10^{-6} \pm 3.503 \times 10^{-8} \text{ cm}^2 \text{s}^{-1}$, $R_h = 15.6 \text{ \AA} \pm 0.2$ for the YD mutant. Both are close to the hydrodynamic radius of the folded fibronectin binding protein D3 ($r = 14.9 \text{ \AA}$) [67] and well under the predicted Stokes radius of an extended 39 amino acid chain **Fig. 3–1**. This is a surprising finding given the high content of negatively charged residues (15 of 39 residues) in both forms would be expected to promote open conformations. However, because the γ -CT also has a relatively high number of hydrophobic residues (14/39), it is likely that there are other forces, such as the hydrophobic effect, acting on the packing of the polypeptide. We also note that the YD has a slightly higher Stokes radius than the WT suggesting some shifts in conformational sampling that are absent in the WT. Chemical shift secondary structure predictions show that both forms occupy a disordered random coil state and do not show evidence of adopting any secondary structure domains **Fig. 3–3**. We do not observe any disorder-order transitions that are typical of functionally-coupled phosphorylated IDPs **3–4**. However, relaxation-dispersion experiments, which detect changes in the chemical environment of residues

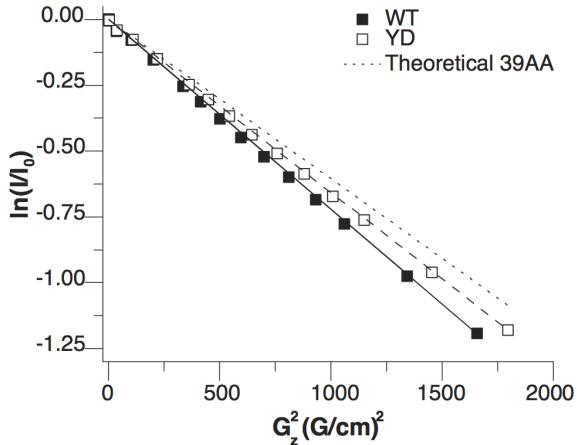


Figure 3–1: NMR Diffusion Measurements

NMR gradient diffusion data. The natural logarithms of normalized ${}^1\text{H}$ methyl peak intensities ($\ln \frac{I}{I_0}$) are plotted as a function of the square of the strength of the magnetic field gradient (G_z^2) applied during a diffusion delay for WT (filled squares) and Y11D (hollow squares) γ -CT polypeptides. The slopes (solid lines) are proportional to the translational diffusion constants. Experimental uncertainties are smaller than the symbols used. The dashed line corresponds to the expected $\ln \frac{I}{I_0}$ versus G_z^2 relationship for a typical random coil polypeptide of 39 amino acids [7].

on the micro-millisecond timescale, show that the Y11D mutant undergoes large coordinated transitions between two states, while the WT dispersion profile is flat

Fig. 3–2. The two states explored by the mutant are calculated to exchange between a major state populated to 97.5 % and a minor state populated to 2.5% with an exchange rate of 2252 sec^{-1} . Combining the dispersions brought about by the YD mutation with the shift in global D_t to more open conformations, we hypothesize that the YD mutation induces collective motions between a collapsed major state and an extended minor state.

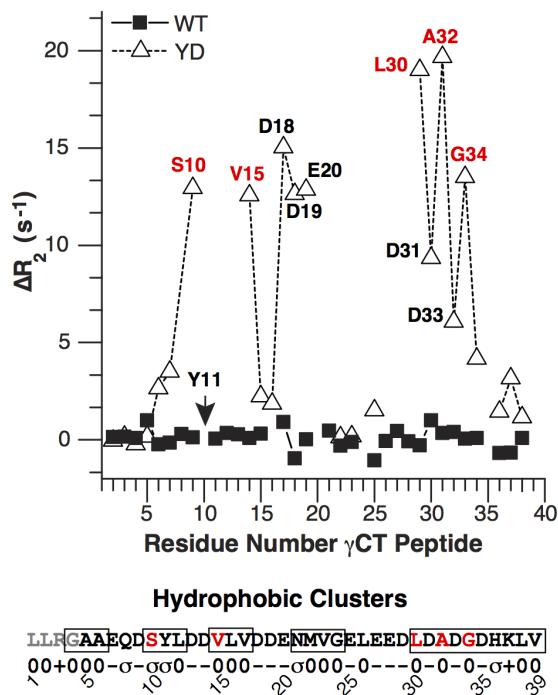


Figure 3–2: NMR Dispersion Measurements
 Magnitudes of ^{15}N relaxation dispersion profiles,
 $\Delta R_2 = R_2(\nu_{CPMG} = 1000\text{Hz}) - R_2(\nu_{CPMG} = 100\text{Hz})$, plotted as a function of residue number for WT (squares) and Y11D (triangles) γ -CT polypeptides. Gaps in the data are due to overlap in $^1\text{H}-^{15}\text{N}$ correlation NMR spectra. Large ΔR_2 values indicate micro-millisecond timescale changes in chemical environment per residue.

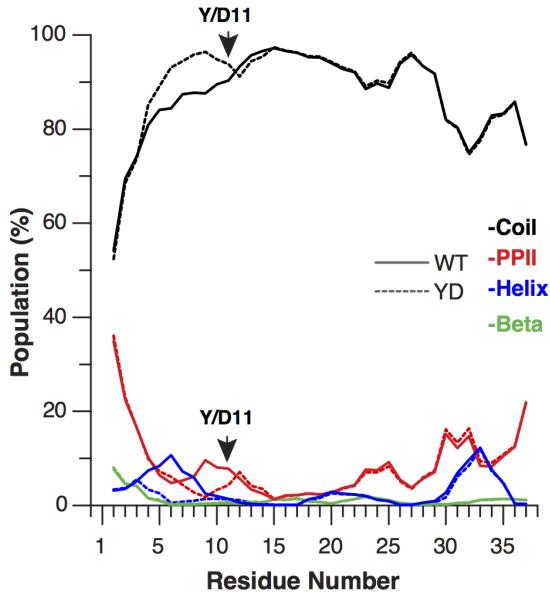


Figure 3–3: NMR Secondary Structure Assignments
 Secondary structure content determined from backbone NMR chemical shifts plotted as a function of residue number for WT (solid line) and Y11D γ -CT (dashed line) polypeptides with black, blue, green, and red indicating the percentage of random coil, α -helical, β -sheet, and poly-proline II helical conformations, respectively.

In this work we use MD simulations of both γ -CT forms to gain further insight into the transition hypothesized by NMR.

3.2 MD simulation set-up

Molecular Dynamics simulations (MDS) on WT and Y11D γ -CT were carried out using MPI-enabled GROMACS 4.6.6 software[22] and a CentOS 5 high performance computational cluster. Calculations were distributed over 64 Dual Sandy Bridge 8-core, 2.6 GHz computing nodes and run under periodic boundary conditions

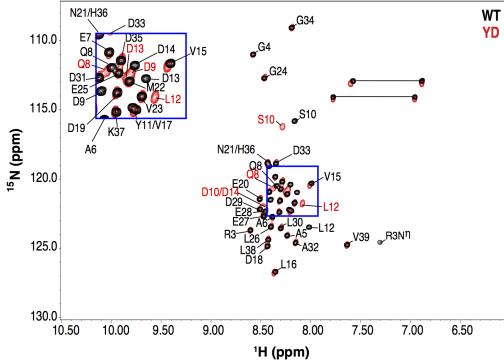


Figure 3–4: NMR Chemical Shifts

1^H - ^{15}N correlation spectra of WT (black) and Y11D (red) γ -CT polypeptides. The blue inset shows an enlargement of the central region. Residue-specific assignments are indicated for each peak. Black labels correspond to the WT and overlaying Y11D peaks. Assignments of perturbed Y11D peaks are indicated in red.

with the OPLS-AA (Optimized Potential for Liquid Simulations All Atom) force field [29]. The starting γ -CT polypeptide configurations were obtained from secondary and tertiary structure predictions by RaptorX [28] and solvated using the SPCE (extended single point charge) water model in a dodecahedral box while enforcing a minimum distance between the edge of the box and solute of 1 nanometer. The total charge of the system was neutralized by adding sodium ions to the solution. Energy minimization was carried out using a steepest descent algorithm for a maximum of 50,000 steps until a maximum force of 100 kJ/mol between atoms was achieved. A 1 nm cut-off was used for non-bonded interactions, and long-range electrostatics were calculated using a Particle Mesh Edwald Sum algorithm. The systems equilibrated under the constant NVT and NPT ensembles (288K and 1 atm) for 5 ns before the production 2 μ s simulations. Post-processing of all trajectories was done

using the `trjconv` module of GROMACS. Theoretical random-coil structural ensembles (10,000 conformers) were calculated based on the γ -CT primary amino acid sequence using Flexible Meccano software [50]. Translational diffusion coefficients were calculated for each structure using hydroNMR software [11].

3.3 Conformational sampling of γ -CT

We computed atomic trajectories for two separate peptides, the WT and YD forms of the γ -CT in all-atom MD simulations with durations of 2 μ s. We choose this simulation length because timescale of transitions observed in NMR are on the micro-millisecond order. While our nano-microsecond MD simulations do not reach such timescales due to computational limitations, long timescale motions measured in NMR have been reproduced in shorter timescale MD simulations but with faster exchange rates [42, 57, 21]. From the resulting trajectories we are able to capture dynamics and conformational sampling that are remarkably consistent with those measured in NMR. It should be noted that a major concern in MD simulations is the problem of sufficient sampling and convergence. While we are fortunate to have direct experimental validation available in the NMR data, it is left to future work to perform further repetitions of the simulation under different initial states to obtain deeper sampling. This is especially important in the case of disordered systems where the range of conformations sampled can be very high and a large amount of sampling is required to obtain appropriate ensemble averages.

γ -CT does not adopt any stable secondary structure

The first question we addressed was whether the structures in our MD trajectories feature the same lack of global structure that was observed in NMR. We used the `dssp` algorithm [27] in the VMD software package [23] to compute secondary structure assignments on each residue in the chain. Assignments by `dssp` are based on computations of hydrogen bonding energies between all atoms. Since hydrogen bonds are the primary stabilizing interaction that gives rise to backbone secondary structure, `dssp` is able to classify geometries arising from potential hydrogen bonds to a category of secondary structure motif found in a large database of annotated structures. We apply this algorithm to every frame in the trajectories at 1 ns intervals to assess secondary structure motifs at every residue as a function of simulation time **Fig. 3–5**. Apart from some local helicity in the middle residues, secondary structure assignment plots for both trajectories point to a consistent absence of global secondary structure motif. The major classes of secondary structure present are turn and coil geometries which correspond to a largely unstructured ensemble of conformations.

Although we do not observe ordered secondary structure rearrangements by looking at `dssp`, we compute RMSD to ask whether the simulations produce any conformational changes in the disordered ensemble. RMSD quantifies the distance between superimposed structures and is therefore a useful tool for detecting the presence of conformational changes in a trajectory. We therefore computed backbone RMSD values for every frame in the simulation with respect to the starting structure. Since the starting conformation is not derived from experimental data and is not expected to correspond to a native state, we also compute RMSD with a 10 ns sliding

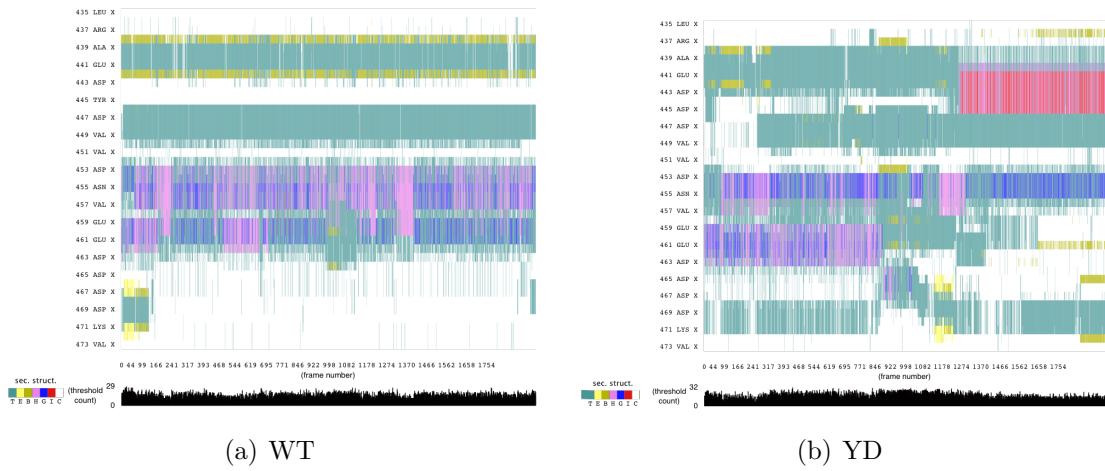


Figure 3–5: γ -CT Secondary Structure Assignments

Per-residue secondary structure assignments based on 3D coordinates are computed for every frame of the simulation. WT and YD trajectories lack global and persistent secondary structure motifs throughout the duration of the simulation.

Apart from turn motifs and small local α -helices, the γ -CT samples largely disordered conformations and does not undergo any disorder-order transitions. T: turn, E: extended, B: isolated bridge, H: α -helix, G: 3/10 Helix, I: π -helix, C: coil.

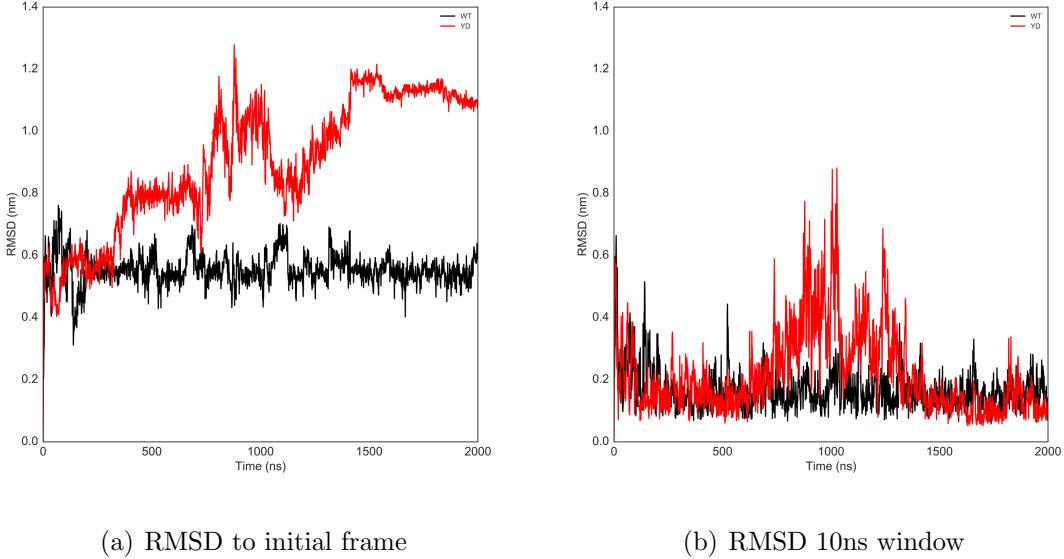


Figure 3-6: RMSD

RMSD plot over simulation time comparing each frame to the starting frame. **Fig. 3-6(a)** shows YD mutation induces conformational changes absent in WT. 10ns sliding window RMSD plot shows rapid conformational switching in YD simulation in the middle portion of the simulation. **Fig. 3-6(b)**

window where every frame is compared with the one 10ns before. In the middle portion of the YD simulation, both methods of computing RMSD contain sharp peaks which indicate the presence of large scale backbone rearrangements **Fig. 3-6**. In contrast, WT RMSD values remain stable throughout the simulation. This suggests that the YD mutation can modulate conformational exploration and the stability of the γ -CT.

γ -CT is largely collapsed

Given that NMR reports transitions between extended and collapsed conformations in the YD mutant, we hypothesize that a similar motion is driving the displacement observed in the RMSD computations. In order to obtain values of compactness that can be compared directly to NMR results, we compute the translational diffusion coefficient (D_t) of conformers in our simulations. As a reference point for interpreting the diffusion values of the trajectories, we use the software **flexiblemeccano** to compute an ensemble of disordered peptides of the YD polypeptide. **flexiblemeccano** takes a primary sequence as input and generates an ensemble of 3D conformations based on amino acid specific conformational potentials and volume exclusion. We then use **hydroNMR** to compute D_t values for each conformer and obtain a distribution for the D_t of the γ -CT. From this distribution **Fig. 3–7** we obtain a large range of conformations; from highly collapsed, to extended chains against which we can compare MD-derived values.

We computed global averages for the radius of gyration, and translational diffusion coefficient over the $2\text{ }\mu\text{s}$ simulations. Both simulations appear to occupy largely collapsed conformations which agrees with experimental findings **Fig. 3–8**. The WT polypeptide D_t mean is $D_t = 1.237 \times 10^{-6} \pm 1.5816 \times 10^{-8}\text{ cm}^2\text{ s}^{-1}$ while the value obtained through NMR is $D_t = 1.25 \times 10^{-6} \pm 1 \times 10^{-8}\text{ cm}^2\text{ s}^{-1}$. Similarly to what was seen by NMR, we find that the mean D_t of the Y11D γ -CT polypeptide is slightly lower than that of the WT γ -CT ($D_t = 1.224 \times 10^{-6} \pm 3.503 \times 10^{-8}\text{ cm}^2\text{ s}^{-1}$). These results confirm that the γ -CT, while disordered, is more compact than a fully denatured polypeptide chain, and that the YD γ -CT is more extended on average.

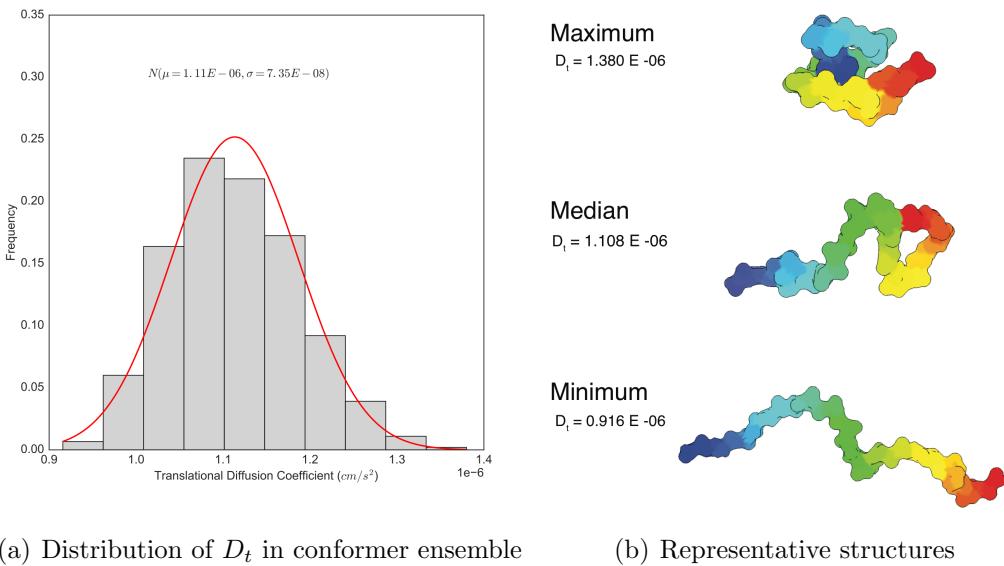


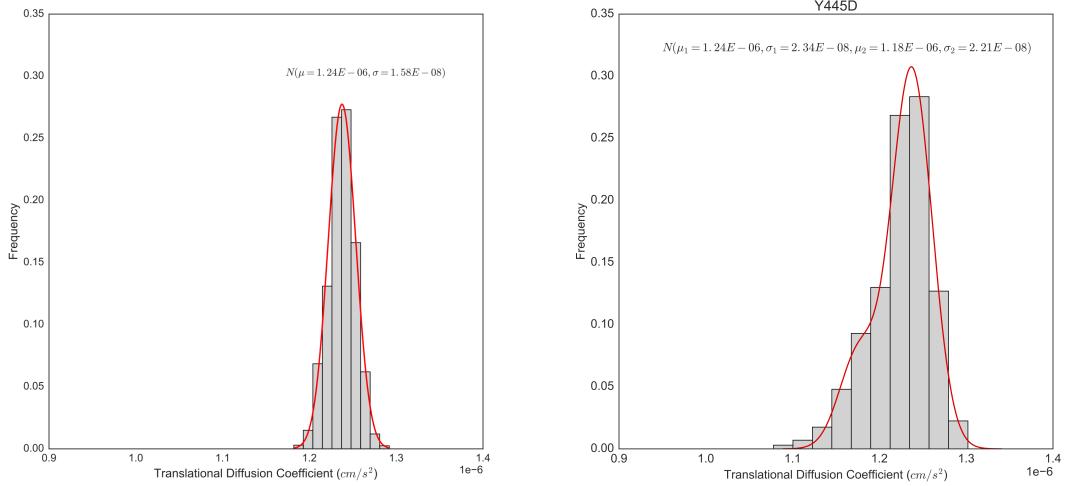
Figure 3-7: γ -CT predicted conformational ensemble

We use `flexiblemeccano` to obtain a conformer ensemble based on the primary sequence of the γ -CT and plot D_t for every conformer in the set of 10,000 sampled structures **Fig. 3-7(a)**. Structures corresponding to the maximum, minimum, and median are visualized D_t . **Fig. 3-7(b)**

Translational diffusion coefficient measurements in NMR report that both γ -CT forms primarily occupy collapsed conformations. The global experimental average for the WT polypeptide obtained through NMR is $D_t = 1.25 \times 10^{-6} \pm 1 \times 10^{-8} \text{ cm}^2 \text{ s}^{-1}$ which agrees well with the NMR-derived value ($D_t = 1.25 \times 10^{-6} \pm 1 \times 10^{-8} \text{ cm}^2 \text{ s}^{-1}$). Similarly to what was seen by NMR, we find that the mean D_t of the Y11D γ -CT polypeptide is slightly lower than that of the WT γ -CT ($D_t = 1.224 \times 10^{-6} \pm 3.503 \times 10^{-8} \text{ cm}^2 \text{ s}^{-1}$). These results confirm that the γ -CT, while disordered, is more compact than a fully denatured polypeptide chain **Fig. 3–8(c)**.

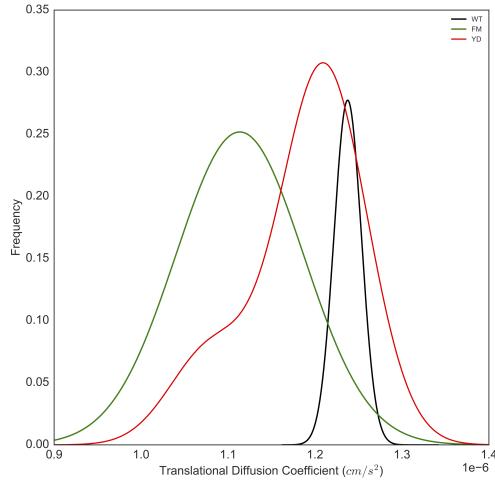
Although both forms of the γ -CT primarily occupy collapsed and disordered conformations, we do observe that, as in NMR, the YD γ -CT has a slightly lower mean D_t than WT. From NMR we hypothesize that this is caused by transitions between compacted and extended driven by enhanced dynamics in the YD mutant. Lending support to this hypothesis, we show that we are able to explain the distribution of diffusion coefficients in the YD mutant as the sum of two gaussian distributions with parameters $\mu_1 = 1.24 \times 10^{-6} \text{ cm}^2 \text{ s}^{-1}, \sigma_2 = 2.34 \times 10^{-8}, \mu_2 = 1.18 \times 10^{-6} \text{ cm}^2 \text{ s}^{-1}, \sigma_2 = 2.21 \times 10^{-8}$ **Fig. 3–8(b)**. This suggests that the YD dynamics likely give rise to a two-state system where a minor state occupies extended conformations **Fig. 3–8(c)**. Meanwhile, the WT D_t distribution is best explained by a single normal distribution which suggests that the peptide occupies a single stable state which corresponds to the compacted portion of conformation space **Fig. 3–8(a)**.

Looking at D_t as a function of simulation time, we found that the diffusion coefficient (D_t) of the WT γ -CT remains relatively constant throughout the simulation, ($D_t = 1.237 \times 10^{-6} \pm 1.5816 \times 10^{-8} \text{ cm}^2 \text{ s}^{-1}$ and agrees well with the NMR-derived



(a) WT

(b) YD



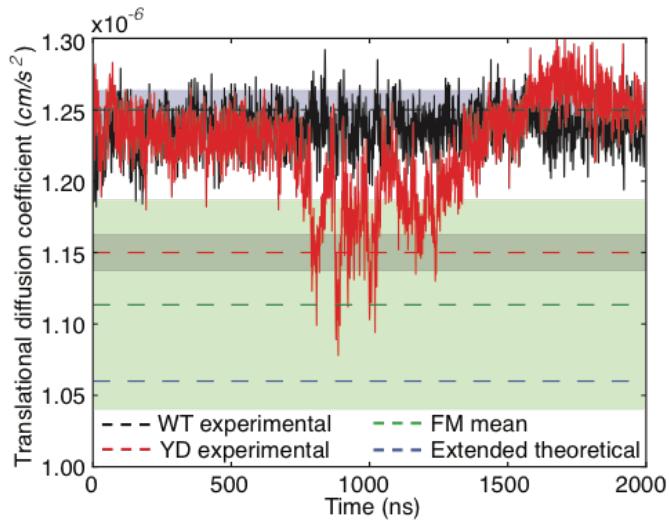
(c) Distribution Overlay

Figure 3–8: Distribution of diffusion coefficients

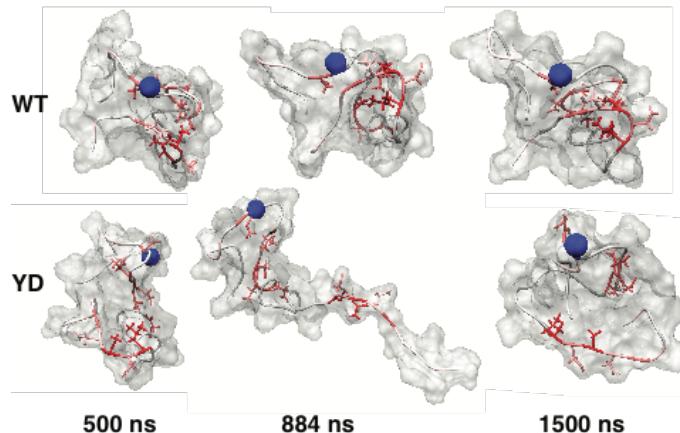
Frequency histograms of translational diffusion coefficient for both γ -CT forms. WT γ -CT **Fig. 3–8(a)** is closely centered around a compact conformation while the

YD mutant also occupies a major collapsed state but also explored a minor extended state **Fig. 3–8(b)**. The resulting distribution of YD can be expressed as the sum of two gaussian distributions, reflecting the two-state nature of the system.

Fig. 3–8(c) depicts an overlay of both γ -CT distributions with the predicted conformational ensemble computed by **flexiblemeccano** and shows that WT γ -CT remains in the collapsed region of the ensemble while YD explores more extended regions.



(a) D_t over simulation time



(b) Representative structures

Figure 3–9: Time series of conformational sampling of γ -CT under Molecular Dynamics Simulations.

Fig. 3–9(a) Translational diffusion coefficients of conformations of the WT (black) and Y11D (red) γ -CT polypeptides sampled at 1 ns intervals over a $2\mu\text{s}$ MD simulation (2000 conformations per condition). Experimentally determined D_t are shown as a line with the associated error indicated by a shaded region on either side. The theoretical D_t for a denatured 39 residue protein [7] is indicated by the dashed blue line. The shaded green color shows 1 standard deviation of a randomly generated ensemble of structures that an unfolded WT or Y11D γ -CT polypeptide can have based on the amino acid-specific conformational potentials and volume exclusion. This region corresponds to partially extended conformations as seen in

Fig. 3–7(a) **Fig. 3–9(b)** Structures taken from the simulation at 500, 887 (minimum D_t value) and 1500 ns for WT and Y11D γ -CT polypeptides.

value ($D_t = 1.25 \times 10^{-6} \pm 1 \times 10^{-8} \text{ cm}^2 \text{s}^{-1}$). Similarly to what was seen by NMR, we find that the mean D_t of the Y11D γ -CT polypeptide is slightly lower than that of the WT γ -CT ($D_t = 1.224 \times 10^{-6} \pm 3.503 \times 10^{-8} \text{ cm}^2 \text{s}^{-1}$). These results confirm that the γ -CT, while disordered, is more compact than a fully denatured polypeptide chain. Interestingly, between 762 to 1255 ns in the MDS, the Y11D γ -CT underwent transient excursions to less compact conformations with significantly lower diffusion coefficients (mean $D_t = 1.152 \times 10^{-6} \pm 2.0325 \times 10^{-8} \text{ cm}^2 \text{s}^{-1}$). This sub-population is more extended (i.e. diffuses more slowly) than any conformation sampled by the WT γ -CT throughout the entire MDS. While the Y11D γ -CT extended states do not overlap with the conformational ensemble of the WT γ -CT polypeptides, they do, however, lie close to the extended conformational space for a typical random-coil poly-peptide, as modeled by **flexiblemeccano** **Fig. 3–9**.

In order to characterize the two distinct states, we use the D_t distributions in **Fig. 3–8** to select the conformations within the top and bottom 1% (20 structures each) of the WT γ -CT and Y11D γ -CT. In the case of the WT, we do not expect the upper and lower D_t subsets to differ substantially, as the WT γ -CT conformations exhibit fairly homogeneous compactness overall. For Y11D γ -CT, we expect the upper D_t subset to resemble that of the WT γ -CT, while the lower D_t subset is expected to reflect the transient opening process. We plotted the mean distance between alpha carbons of all pairs of residues as contact maps for the set of collapsed and extended conformations of the WT γ -CT polypeptide **Fig. 3–10(a)** and the Y11D γ -CT polypeptide **Fig. 3–10(b)**. As expected, the upper and lower D_t subsets of the WT γ -CT and the upper D_t subset of the YD γ -CT polypeptides

show similar patterns of pair-wise contacts. In contrast, the C-terminal residues in the lower D_t subset of the Y11D γ -CT lose the majority of contacts with N terminal residues, as a consequence of the conformational expansion. Next, we isolated the three conformations from the upper and lower D_t subsets of Y11D γ -CT polypeptides with the lowest all-to-all RMS, also known as centroid structures, shown in **Fig. 3–10(c)**. This analysis shows that the extended conformations consist of a compact N-terminus with residues located in the C-terminal region of the γ -CT, (including dynamically-broadened residues L30, A32 and G34) isolated from the N-terminus and solvent-accessible. Through MD we are able to re-produce the anomalously rapid diffusion (i.e. high compactness) of the WT and Y11D ground-state γ -CT polypeptides. Moreover, we saw that the YD substitution caused relatively slow collective motions of the entire polypeptide chain, as observed by NMR. This serves as a possible explanation for how residues throughout a disordered polypeptide can experience concerted, two-state, dynamical behaviour in the presence of the Y11D mutation. More specifically, it is the separation of a cluster of residues located in N and C termini of the γ -CT polypeptide that drives a transition to extended conformations with a concomitant reduction of the translational diffusion coefficient.

3.3.1 Collective motions correspond to transitions between extended and collapsed conformations

Until this point, we have identified the presence of an extended sub-population of the YD γ -CT that is absent in the WT. This lends support to the hypothesis that the shift in D_t measured by NMR is due to transient expansions of the YD backbone into an extended state. Complementary to this finding, is the fact that residues in the YD polypeptide show evidence of collective motions detected as shifts

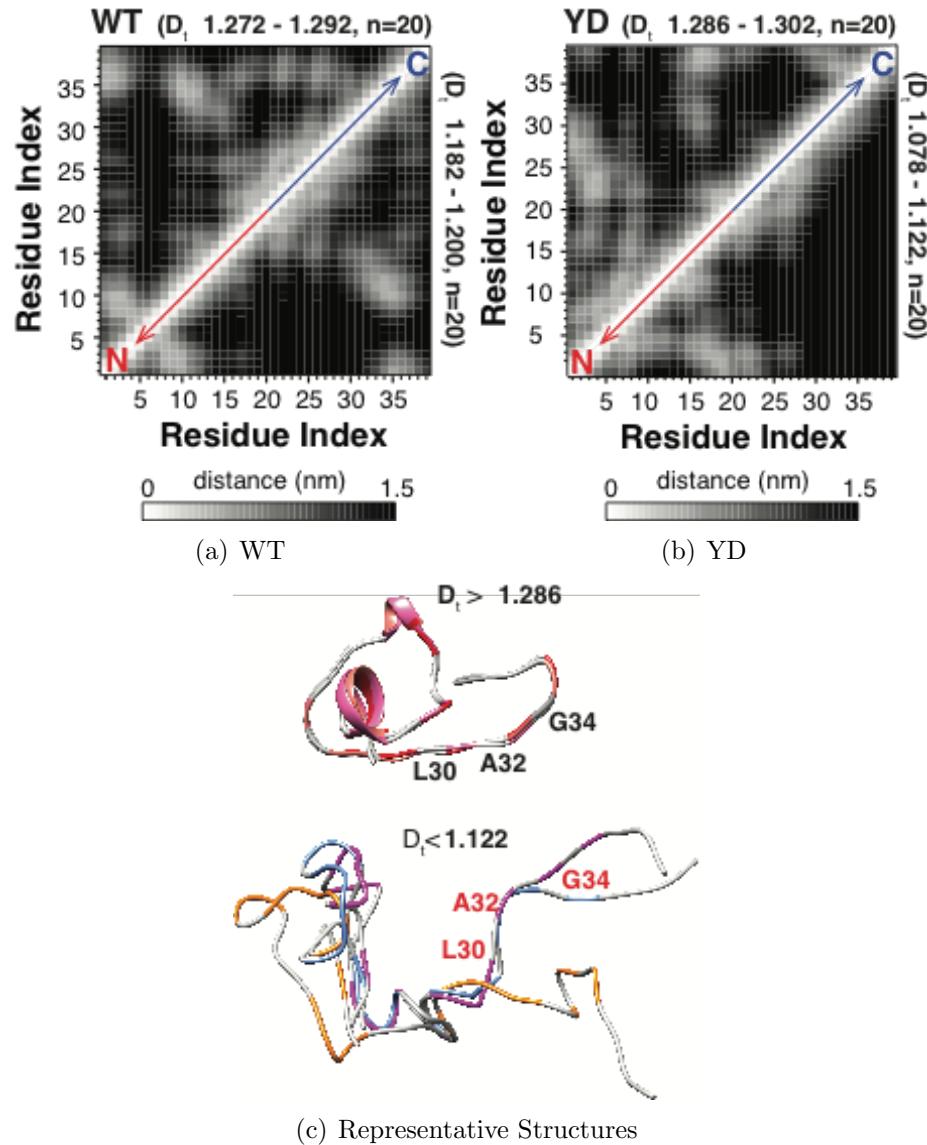


Figure 3–10: Contact Maps of Extended and Collapsed Ensembles
 All-to-all residue carbon- distances in structures with lowest 1% D_t (upper triangle) and highest 1% D_t (lower triangle). Comparison is shown for WT **Fig. 3–10(a)** and Y11D **Fig. 3–10(b)** simulations. Representative structures obtained identifying structures with lowest all-to-all RMSD values for the Y11D low diffusion coefficient group. High dispersion residues with $\Delta R2$ values greater than $5 s^{-1}$ are labeled in red.

in chemical environment through NMR. This suggests that the transition between states occurs in a coordinated manner. We therefore seek to test whether correlated motions are also present in the simulation, and if they are, whether they can explain the transitions between collapsed and extended states.

We performed covariance analysis, also known as Principal Component Analysis on γ -CT trajectories to identify major modes of correlated motion in our trajectories. We use `covar` and `anaeig` from the GROMACS package to build a backbone atom covariance matrix, extract principal modes, and perform dimensionality reductions through eigenvector projections. In order to eliminate rotational and diffusive translations which do not correspond to conformational motions, we align all frames in the trajectory to the average structure as computed by `covar` using RMSD based clustering. As is typical with molecular simulations which operate on a limited number of backbone degrees of freedom, the first few eigenvectors in both trajectories account for nearly all of the variation in the trajectories **Fig. 3–11**.

We therefore focus our attention on the two first major modes of motion. **Fig. 3–12** shows a 2 dimensional projection onto the first two eigenvalues of WT and YD trajectories. Each point represents a 3D conformation in the 2 μ s simulation projected along the first two eigenvectors. The WT projection shows a conformational space that is closely clustered, indicative of constrained motions consistent with the low dispersions found in NMR, and the single state behaviour suggested by RMSD and D_t analysis. However, the YD appears to be exploring multiple conformation clusters which is in agreement with the presence of high dispersion groups found in NMR. Furthermore, by coloring each conformation with a normalized D_t value we

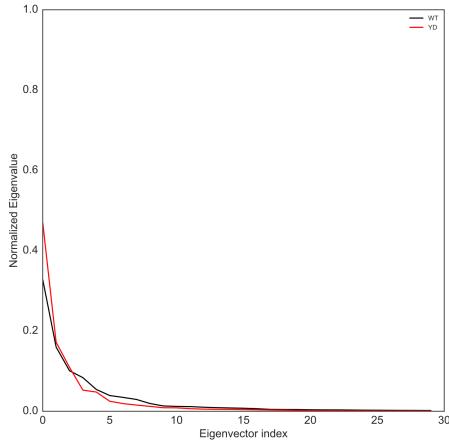


Figure 3–11: Eigenvalues of γ -CT PCA

Normalized eigenvalues of first 50 eigenvectors resulting from covariance matrix diagonalization of WT and YD trajectories.

are able to show that correlated motions along the major modes correspond with transitions between collapsed and extended states.

In order to visualize the transitions, we generate a porcupine plot depicting the direction of motion between conformations on two extremes of the second principal component projection **Fig. 3–13**. We show that the transition between collapsed and extended is indeed driven by a separation of the N and C terminus in a correlated fashion. With this analysis we are able to propose a physical mechanism to explain the concerted global dynamics and diffusion coefficient shift observed in NMR as the action of correlated motions brought about by a local change in electrostatic environment.

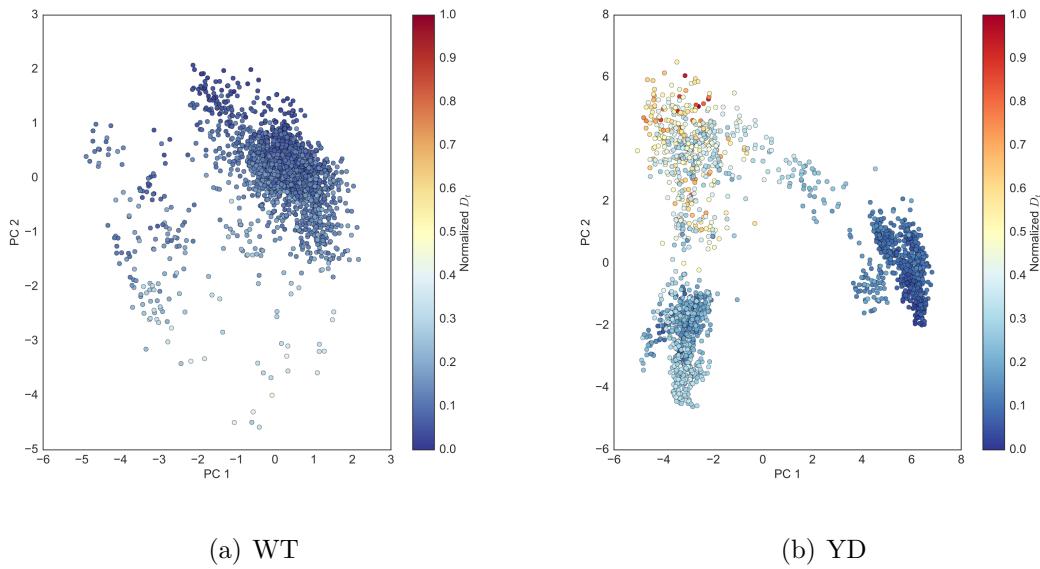
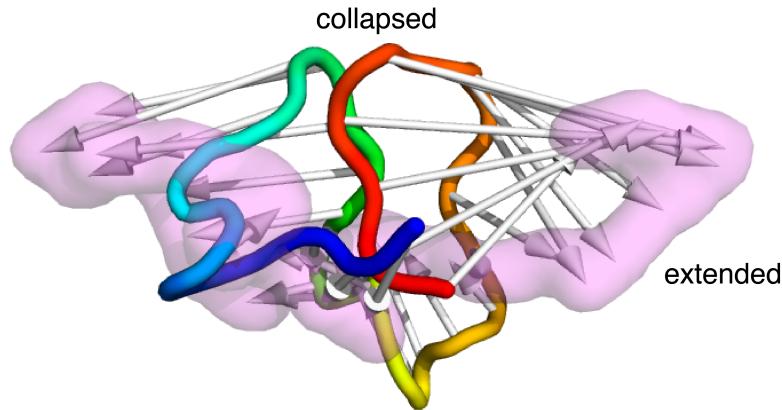
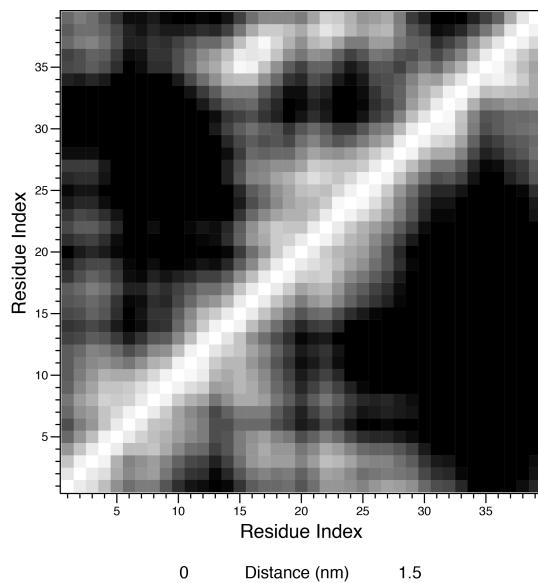


Figure 3–12: Principal Component Analysis
 2 dimensional projections of the WT **Fig. 3–12(a)** and YD trajectories
Fig. 3–12(b). Each point represents a 3D conformation projected along the first two principal axes. Conformations are colored according to a normalized D_t value where 1 is the lowest D_t (most extended) and 0 is the highest (most compact).



(a) Porcupine Plot



(b) YD Contact Maps

Figure 3–13: Extreme projection along opening-closing axis.

Fig. 3–13(a) Arrows indicate direction of motion along opening/closing axis from YD PCA where the purple cloud represents the most extreme projection along the axis. **Fig. 3–13(b)** Contact maps of the minimal (top triangle) and maximal (bottom triangle) projections along open/close axis show separation of N and C terminus.

3.3.2 Whole protein simulations and conserved properties of γ -CT

Our analysis of γ -CT structural properties using NMR and corresponding MDS are based on the WT and Y11D γ -CT polypeptides in isolation. In order to determine whether the conformations and dynamics we observed for the isolated γ -CT are physically consistent when attached to the full-length γ -Tubulin protein, we docked the minimum D_t γ -CT model **Fig. 3–15** onto the globular domain of an S.c. γ -Tubulin homology model as an initial structure for a whole protein simulation. Due to the substantial increase in system size, simulation times were reduced to 200 ns. Despite the shorter simulation time, with the Y11D γ -CT polypeptide, the γ -CT in the whole protein simulation underwent exchange between extended and compact conformations **Fig. 3–15**, suggesting both states are still accessible in the presence of the globular domain. We found no contacts between residues in the globular domain with the 39 residues of the γ -CT throughout the 200 ns simulation (minimal distance between any pair of residues is greater than 0.7 nm) **Fig. 3–14**. Structures for the full protein with the γ -CT at minimum radius of gyration (1.073 nm) and maximum radius of gyration (1.582 nm) are shown in **Fig. 3–15**. This suggests that the γ -CT opening-closing mechanism can act independently of the rest of the protein.

Finally, we obtain evolutionary evidence that the physical mechanism of the γ -CT described here is likely present in many other organisms, and that it is not akin to other tubulin tails. We used the MUSCLE alignment tool [13] to generate multiple sequence alignments from sets of 71 β and 85 γ -CTs primary sequences spanning various eukaryotic kingdoms from the UniProt database [2] **Fig. ??**. We see from the resulting alignments that both CTs are enriched in acidic residues (Asp,

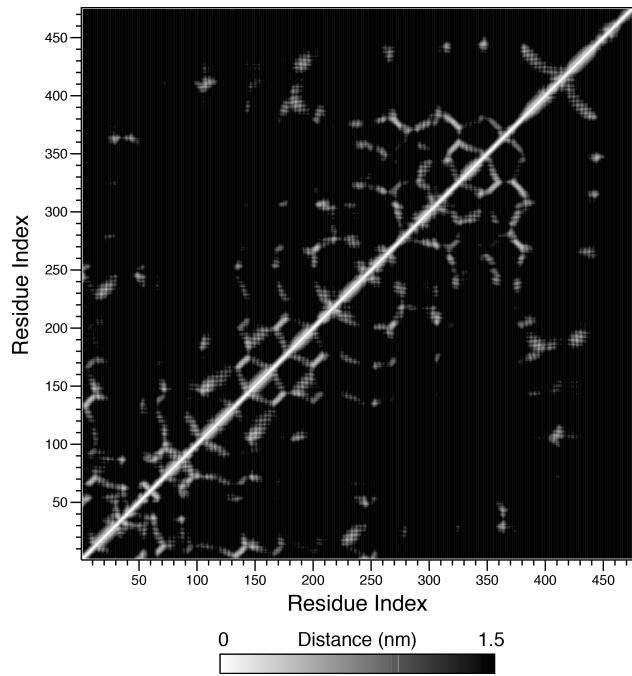


Figure 3–14: YD Globular Contact Map

Average contact map for full protein simulation with an extended YD γ -CT.

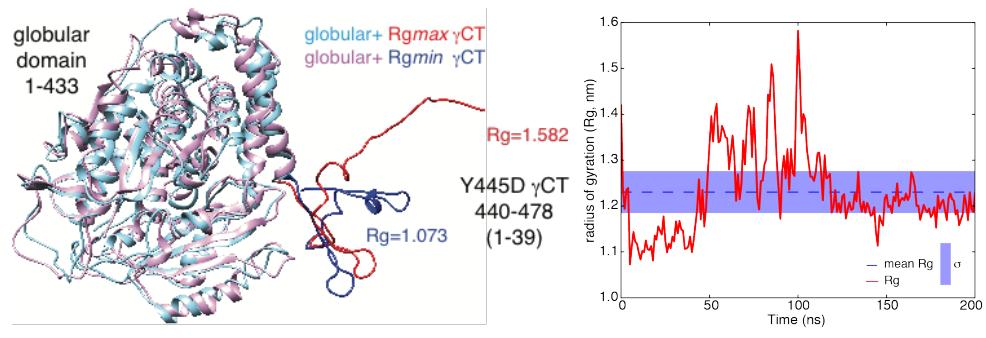


Figure 3–15: γ -Tubulin whole protein simulation

Fig. 3–15(a) Alignment of whole protein simulation structures with mean, maximum, and minimum γ -CT radius of gyration. **Fig. 3–15(b)** radius of gyration of γ -CT from whole protein simulation. The shaded blue region denotes 1 standard deviation around the mean radius of gyration value.

Glu). Whereas γ -CT across eukaryotes additionally contain clusters of hydrophobic or polar residues which are not found in α - or β -CTs. Interestingly, the residues most broadened in Y11D NMR spectra, i.e. those most affected by the compact-to-extended transition (V15, D19, E20, A32, G34), are all found in positions conserved either on a sequence level or on a physical property level (polarity/charge) in the consensus γ -CT sequence. We therefore hypothesize that clusters of hydrophobic residues, including those that contribute to transitions between compact and extended conformations in the S.c. Y11 γ -CT, are a conserved feature of an otherwise diverse set of γ -CTs across many eukaryotic organisms. This feature appears to be absent in the well characterized β -CT ‘e-hook’IDRs which suggests that the γ -CT mechanism does not behave like α and β . MD simulations of the β -tubulin CTs suggest an extended ground state, unlike the collapsed ground state observed in γ -CT [40]. Furthermore, given that the Y11 position is so pervasively conserved, it is likely that a similar mechanism of action to the one described in this work functions many other eukaryotes.

3.4 Discussion

Through NMR measurements and MD computer simulations, we demonstrate the first example of an IDP acting as a disorder-to-disorder regulatory switch, and propose a physical mechanism to explain the regulation of an essential biological machine.

Both NMR and MD are in strong agreement that the conformational sampling of the γ -CT lies entirely within the diordered ensemble, as no signal of secondary structure was detected by either method. Diffusion measurements show that the

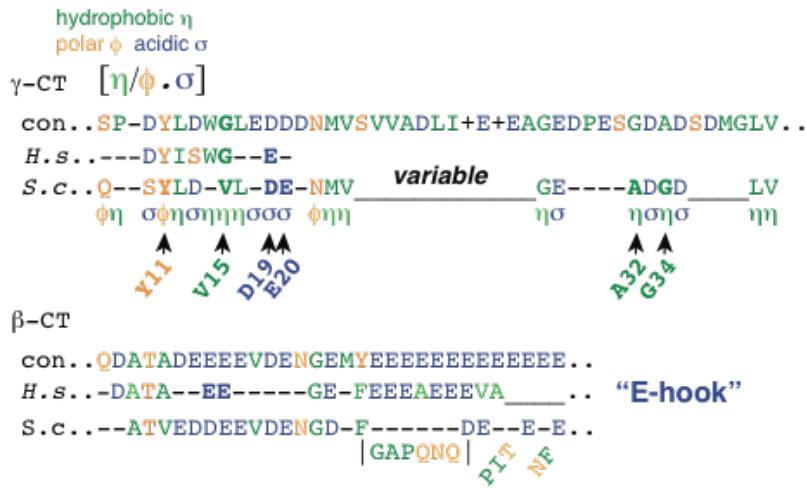
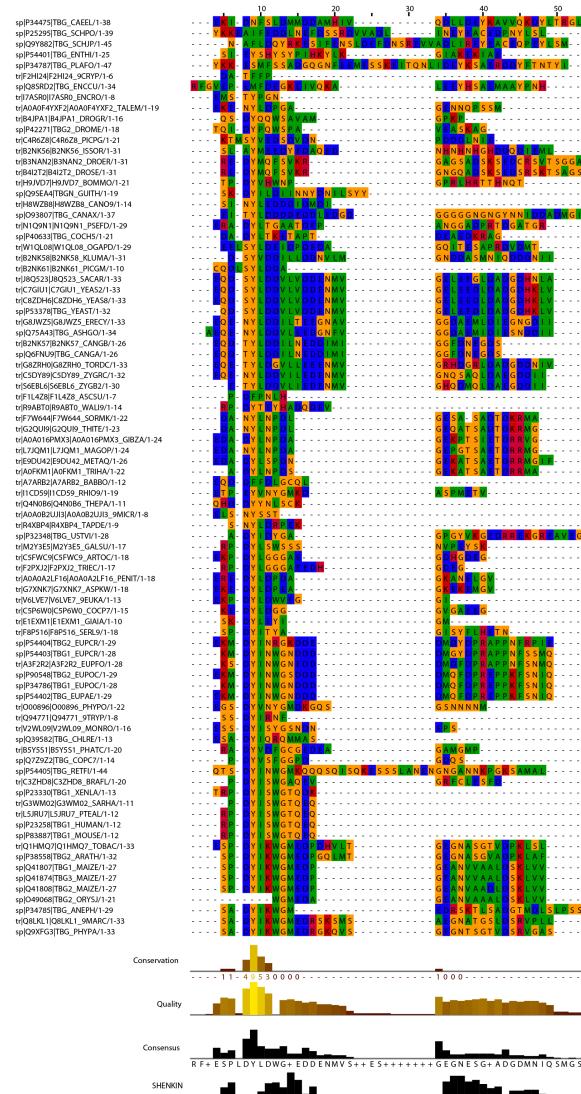


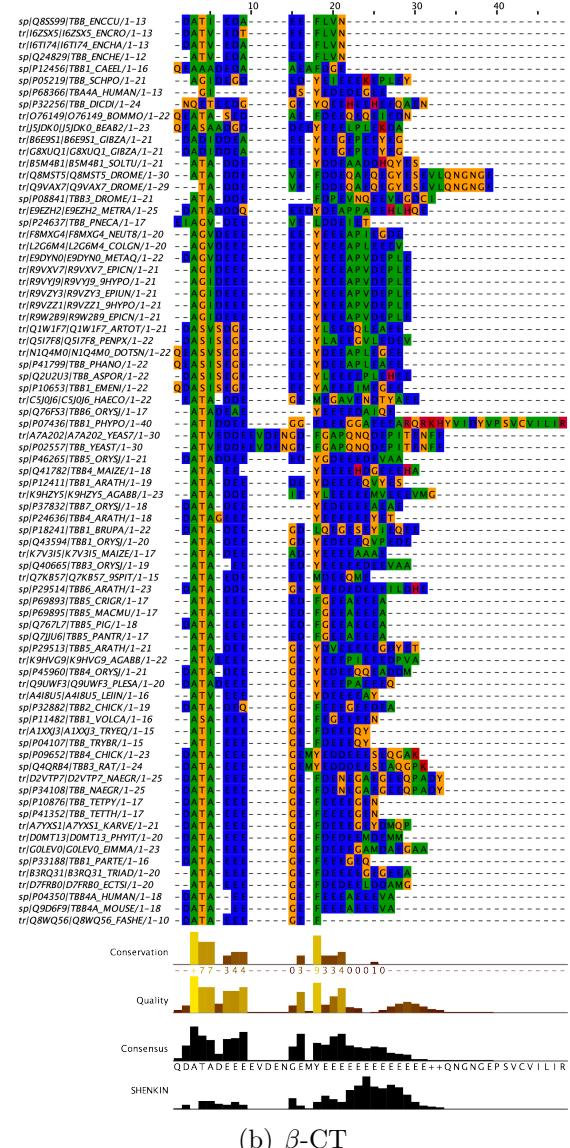
Figure 3–16: Multiple Sequence Alignment of γ -CT across eukaryotes

Primary sequence consensus from a multiple protein-sequence alignment of γ -CT across various eukaryotes. Conserved physical properties are shown below primary sequence consensus (hydrophobic η , polar, ϕ acidic σ). An equivalent analysis is shown below for β -CTs, also known as E-hooks.

major conformational state of the γ -CT in the WT and in the YD is collapsed, with the YD having a slightly larger hydrodynamic radius on average. However, NMR and MD simulations both provide evidence that a single point mutation to a negatively charged residue YD produces concerted motions along the entire polypeptide that are not present in the WT γ -CT. Analysis of the collective motions detected in MDS by PCA correspond closely to transitions between collapsed and extended states which leads us to hypothesize that the changes in chemical environment detected in MDS and difference in hydrodynamic radius is a product of an correlated opening and closing motion brought about by the YD mutation. This coordinated opening is likely brought about by the effects of electrostatic repulsion as the Asp substitution introduces further negative charge in an already acidic region. It has



(a) γ -CT



(b) β -CT

Figure 3–16: Full γ and β Tubulin CT Multiple Sequence Alignments
 Primary sequence alignments of visualized by software tool jalview [66] and colored by amino acid physical properties. Blue: acidic, Red: basic, Orange: on-polar, Green: polar.

been previously shown that charge has a strong influence on the conformational ensemble and diffusion rates in IDPs [41]. Once the γ -CT enters the open state, several hydrophobic residues in the middle of the polypeptide(L464, A466, G468) become accessible for protein-protein interactions.

Because both WT and YD appear to have the same collapsed native state and differ only when the YD undergoes transient excursion to an extended state, we can model such behaviour as a two well potential system **Fig. 3–17(d)**. The equilibrium between extended and collapsed conformations is shifted by phosphorylation to render extensions more accessible in the YD. Previously, such dynamics were mostly observed in the well characterized order-to disorder transitions, or the folding-on-binding paradigm [58]. This model stands in contrast to the view of disordered ensembles as uniform **Fig. 3–17(c)** spaces of random coils [44]. And instead, we observe that the disordered landscape has some structure that IDPs can selectively explore through PTMs to modulate functionality by giving rise to switch-like behaviour while still remaining disordered. A similar case has been described in [45] whereby multisite phosphorylation was observed by NMR to cause transient structural changes in a predominantly compact IDR that regulates receptor binding. These changes were not enough to induce global ordering of the structure and were shown to remain within a disordered ensemble. The system observed in this work lends additional evidence to the hypothesis that order is not required to modulate interactions, and presents a novel mechanism of disordered control that is based in switch-like transitions in structural compactness. We claim that this transition is

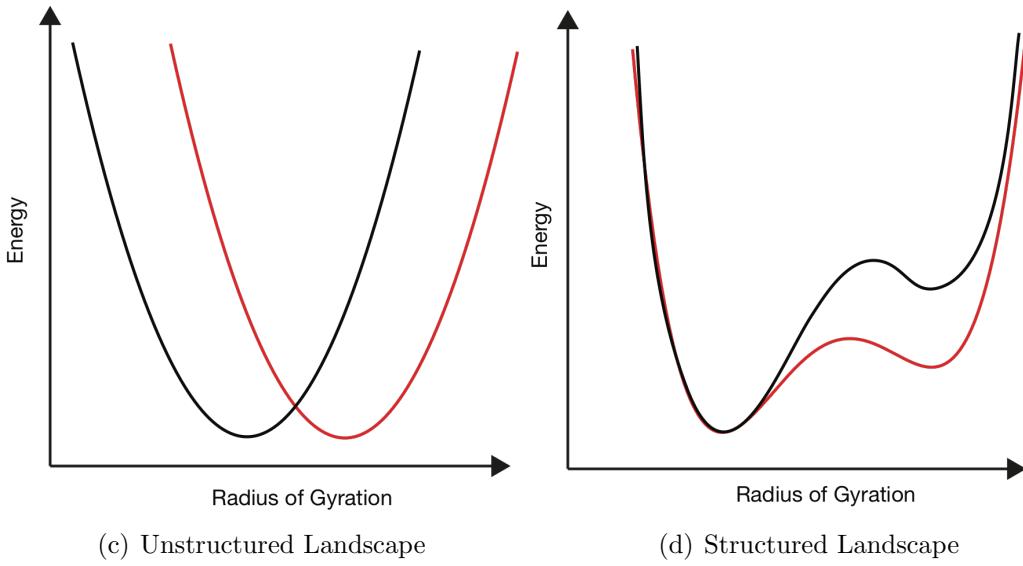


Figure 3–17: IDP Conformational Landscapes

Fig. 3–17(c) shows an energy landscape as a function of radius of gyration (black to red) where a modification in the IDP would lead to a uniform conformational shift towards uniformly less compact structures centered around a single energy minimum. **Fig. 3–17(d)** illustrates the behaviour observed in the γ -CT, where the energy landscape appears to have a two-well structure. This landscape, when combined with phosphorylation, renders a minor state accessible allowing for switch-like behaviour in the disordered ensemble.

dependent on a phosphorylation event, which we model as a negatively charged mutation. By modelling phosphorylation with a negatively charged mutation, we propose that changes in electrostatic environment are a major driver shifting conformational sampling in this system. It is interesting to note that several past works have shown that IDRs in various other systems have demonstrated the role of electrostatics in controlling IDP interactions [6, 8, 41].

From these findings we propose that phosphorylation at Y11 of the γ -CT acts as a regulatory switch to modulate protein-protein interactions. In order to better visualize this mechanism, we dock 3D structures obtained from full γ -Tubulin simulations into cryo electron microscopy models of the γ -TuRC [36] **Fig. 3–18** which previously lacked any information on the γ -CT due to its disordered nature. Based on this visualization we can propose a mechanism to explain the phenotype of hyper-stable microtubules in the Y11D mutants *in vivo*. The extensions that project outward from the complex brought about by the YD substitution, or phosphorylation, allow the γ -TuRC to selectively recruit effector proteins to the minus end of microtubules, making them available to the entire complex to subsequently act in regulating microtubule dynamics. The constitutive addition of negative charge in the YD mutant therefore shifts the equilibrium between the collapsed and extended states, leading to a misregulation of the recruitment of microtubule associated proteins and defects in microtubule dynamics. The identity of the protein(s) being recruited by the γ -CT remains an open question but there are several candidates known to affect microtubule stability and localize to the spindle poles [9] are currently being verified through various experimental screens. However, the characterization of γ -CT conformational sampling is an essential first step and a useful tool in identifying these potential interactors and understanding the complex mechanisms by which IDRs regulate large molecular machines.

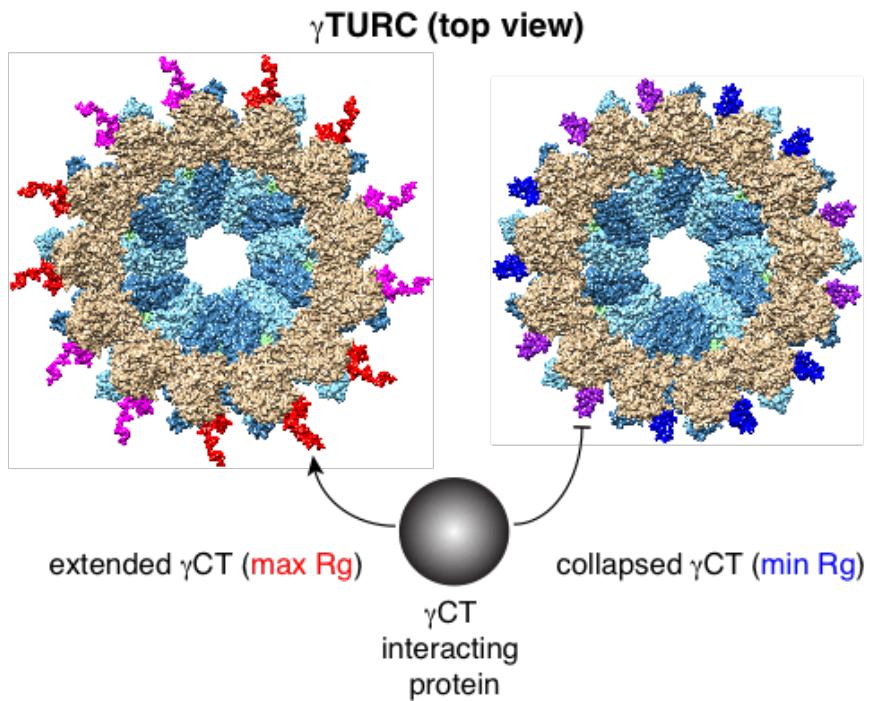


Figure 3–18: γ -TuRC

A working physical model for CT selective interaction with accessory proteins through sampling between ‘extended’ and ‘collapsed’ conformations. Docking of simulated γ -Tubulin models to γ TuRC models shows a possible mechanism for regulating the binding of γ -CT interacting proteins to the ring complex. Extended (high radius of gyration) γ -CT complexes are accessible to interactions, while collapsed γ -CT (low radius of gyration) inhibit interactions.

CHAPTER 4

Conclusions

A totally blind process can by definition lead to anything; it can even lead to vision itself.

Jacques Monod

4.0.1 Summary of Findings

In this thesis I presented computational models that shed light on a previously unreported type of IDP dynamics. The collective motion between extended and collapsed states we observed in the phosphorylated γ -CT constitute evidence for an organized disordered ensemble where shifting the equilibrium between sub-regions can give rise to switch-like behaviour without transitions to folded states. Furthermore, we present a detailed physical mechanism to explain the open question of the role of phospho-regulation of the mitotic spindle.

4.0.2 Future Aims

Due to the extremely high computational costs associated with long timescale MD simulations, we are still continuing work on achieving a more complete sampling of the γ -CT conformational space. To this end, we are running replicate simulations to further confirm our findings. However, with the experimental findings from NMR to guide simulations, we are confident that the MD results are sound.

Next, we would like to exploit the remarkably close agreement between NMR and MD to establish a pipeline for further characterizing the physical mechanisms at play in this system. That is, we would like to harness MD to inform further NMR experiments by making predictions *in silico* of mutants to the γ -CT that would affect the physical mechanism described here in some desired manner. We would then use these predictions to motivate NMR studies which can then go to further validate the mechanism and serve as experimental tools for probing the mechanism *in vivo*. Apart from allowing us to further our understanding of the physical mechanisms of IDRs, functional mutants can serve as a tool for identifying unknown interacting proteins and intra-molecular interactions. We propose to build a high throughput tool for searching the mutational space of IDPs and evaluating dynamic behaviour using a combination of genetic algorithms and MD. Due to the large number of computations this requires, we will have to employ ‘enhanced’MD sampling techniques such as implicit solvation and replica exchange MD to increase sampling rates. This software is already under development and will serve as a testable hypothesis generation tool that can be applied to understanding the structural mechanisms of any intrinsically disordered system.

References

- [1] Hector Aldaz, Luke M Rice, Tim Stearns, and David A Agard. Insights into microtubule nucleation from the crystal structure of human γ -tubulin. *Nature*, 435(7041):523–527, 2005.
- [2] Rolf Apweiler, Amos Bairoch, Cathy H Wu, Winona C Barker, Brigitte Boeckmann, Serenella Ferro, Elisabeth Gasteiger, Hongzhan Huang, Rodrigo Lopez, Michele Magrane, et al. Uniprot: the universal protein knowledgebase. *Nucleic acids research*, 32(suppl 1):D115–D119, 2004.
- [3] Gregory E Arnold and Rick L Ornstein. An evaluation of implicit and explicit solvent model systems for the molecular dynamics simulation of bacteriophage t4 lysozyme. *Proteins: Structure, Function, and Bioinformatics*, 18(1):19–33, 1994.
- [4] Stephen P Bell and Anindya Dutta. Dna replication in eukaryotic cells. *Annual review of biochemistry*, 71(1):333–374, 2002.
- [5] Roby P Bhattacharyya, Attila Reményi, Matthew C Good, Caleb J Bashor, Arnold M Falick, and Wendell A Lim. The ste5 scaffold allosterically modulates signaling output of the yeast mating pathway. *Science*, 311(5762):822–826, 2006.
- [6] Mikael Borg, Tanja Mittag, Tony Pawson, Mike Tyers, Julie D Forman-Kay, and Hue Sun Chan. Polyelectrostatic interactions of disordered ligands suggest a physical basis for ultrasensitivity. *Proceedings of the National Academy of Sciences*, 104(23):9650–9655, 2007.
- [7] Alexander H Boschitsch, Marcia O Fenley, and Huan-Xiang Zhou. Fast boundary element method for the linear poisson-boltzmann equation. *The Journal of Physical Chemistry B*, 106(10):2741–2754, 2002.
- [8] Tao Chen, Jianhui Song, and Hue Sun Chan. Theoretical perspectives on non-native interactions and intrinsic disorder in protein folding and binding. *Current opinion in structural biology*, 30:32–42, 2015.

- [9] Lara Cuschieri, Rita Miller, and Jackie Vogel. γ -tubulin is required for proper recruitment and assembly of kar9–bim1 complexes in budding yeast. *Molecular biology of the cell*, 17(10):4420–4434, 2006.
- [10] Norman E Davey, Gilles Travé, and Toby J Gibson. How viruses hijack cell regulation. *Trends in biochemical sciences*, 36(3):159–169, 2011.
- [11] J Garcia De la Torre, ML Huertas, and B Carrasco. Hydronmr: prediction of nmr relaxation of globular proteins from atomic-level structures and hydrodynamic calculations. *Journal of Magnetic Resonance*, 147(1):138–146, 2000.
- [12] Marileen Dogterom, Jacob WJ Kerssemakers, Guillaume Romet-Lemonne, and Marcel E Janson. Force generation by dynamic microtubules. *Current opinion in cell biology*, 17(1):67–74, 2005.
- [13] Robert C Edgar. Muscle: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research*, 32(5):1792–1797, 2004.
- [14] Marcos RM Fontes, Trazel Teh, and Bostjan Kobe. Structural basis of recognition of monopartite and bipartite nuclear localization sequences by mammalian importin- α . *Journal of molecular biology*, 297(5):1183–1194, 2000.
- [15] Stefano Gianni, Angela Morrone, Rajanish Giri, and Maurizio Brunori. A folding-after-binding mechanism describes the recognition between the trans-activation domain of c-myb and the kix domain of the creb-binding protein. *Biochemical and biophysical research communications*, 428(2):205–209, 2012.
- [16] Robert J Good and Christopher J Hope. New combining rule for intermolecular distances in intermolecular potential functions. *The Journal of Chemical Physics*, 53(2):540–543, 1970.
- [17] Thomas A Graham, Denise M Ferkey, Feng Mao, David Kimelman, and Wenqing Xu. Tcf4 can specifically recognize β -catenin using alternative conformations. *Nature Structural & Molecular Biology*, 8(12):1048–1052, 2001.
- [18] JM Haile. *Molecular dynamics simulation*, volume 18. Wiley, New York, 1992.
- [19] Hedi Hegyi and Mark Gerstein. The relationship between protein structure and function: a comprehensive survey with application to the yeast genome. *Journal of molecular biology*, 288(1):147–164, 1999.

- [20] Joao Henriques, Carolina Cagnell, and Marie Skepo. Molecular dynamics simulations of intrinsically disordered proteins: Force field evaluation and comparison with experiment. *Journal of chemical theory and computation*, 11(7):3420–3431, 2015.
- [21] Katherine Henzler-Wildman and Dorothee Kern. Dynamic personalities of proteins. *Nature*, 450(7172):964–972, 2007.
- [22] Berk Hess, Carsten Kutzner, David Van Der Spoel, and Erik Lindahl. Gromacs 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of chemical theory and computation*, 4(3):435–447, 2008.
- [23] William Humphrey, Andrew Dalke, and Klaus Schulten. Vmd: visual molecular dynamics. *Journal of molecular graphics*, 14(1):33–38, 1996.
- [24] Lilia M Iakoucheva, Predrag Radivojac, Celeste J Brown, Timothy R OConnor, Jason G Sikes, Zoran Obradovic, and A Keith Dunker. The importance of intrinsic disorder for protein phosphorylation. *Nucleic acids research*, 32(3):1037–1049, 2004.
- [25] William L Jorgensen and Julian Tirado-Rives. The opls [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *Journal of the American Chemical Society*, 110(6):1657–1666, 1988.
- [26] Harish C Joshi, Monica J Palacios, Leemore McNamara, and Don W Cleveland. γ -tubulin is a centrosomal protein required for cell cycle-dependent microtubule nucleation. 1992.
- [27] Wolfgang Kabsch and Christian Sander. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 22(12):2577–2637, 1983.
- [28] Morten Källberg, Haipeng Wang, Sheng Wang, Jian Peng, Zhiyong Wang, Hui Lu, and Jimbo Xu. Template-based protein structure modeling using the raptord web server. *Nature protocols*, 7(8):1511–1522, 2012.
- [29] George A Kaminski, Richard A Friesner, Julian Tirado-Rives, and William L Jorgensen. Evaluation and reparametrization of the opls-aa force field for proteins via comparison with accurate quantum chemical calculations on peptides. *The Journal of Physical Chemistry B*, 105(28):6474–6487, 2001.

- [30] Martin Karplus and J Andrew McCammon. Molecular dynamics simulations of biomolecules. *Nature Structural & Molecular Biology*, 9(9):646–652, 2002.
- [31] E Karsenti and I Vernos. The mitotic spindle: a self-made machine. *Science*, 294(5542):543–547, 2001.
- [32] Jamie M Keck, Michele H Jones, Catherine CL Wong, Jonathan Binkley, Daici Chen, Sue L Jaspersen, Eric P Holinger, Tao Xu, Mario Niepel, Michael P Rout, et al. A cell cycle phosphoproteome of the yeast centrosome. *Science*, 332(6037):1557–1561, 2011.
- [33] John C Kendrew, G Bodo, Howard M Dintzis, RG Parrish, Harold Wyckoff, and David C Phillips. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature*, 181(4610):662–666, 1958.
- [34] Dorothee Kern, Brian F Volkman, Peter Luginbühl, Michael J Nohale, Sydney Kustu, and David E Wemmer. Structure of a transiently phosphorylated switch in bacterial signal transduction. *Nature*, 402(6764):894–898, 1999.
- [35] Susan L Kline-Smith and Claire E Walczak. Mitotic spindle assembly and chromosome segregation: refocusing on microtubule dynamics. *Molecular cell*, 15(3):317–327, 2004.
- [36] Justin M Kollman, Charles H Greenberg, Sam Li, Michelle Moritz, Alex Zelter, Kimberly K Fong, Jose-Jesus Fernandez, Andrej Sali, John Kilmartin, Trisha N Davis, et al. Ring closure activates yeast γ -tubulin for species-specific microtubule nucleation. *Nature structural & molecular biology*, 22(2):132–137, 2015.
- [37] Justin M Kollman, Andreas Merdes, Lionel Mourey, and David A Agard. Microtubule nucleation by γ -tubulin complexes. *Nature reviews Molecular cell biology*, 12(11):709–721, 2011.
- [38] Donna F Kubai. The evolution of the mitotic spindle. *International review of cytology*, 43:167–227, 1976.
- [39] Andreas Kukol et al. *Molecular modeling of proteins*. Springer, 2008.
- [40] Tyler Luchko, J Torin Huzil, Maria Stepanova, and Jack Tuszyński. Conformational analysis of the carboxy-terminal tails of human β -tubulin isotypes. *Biophysical journal*, 94(6):1971–1982, 2008.

- [41] Albert H Mao, Scott L Crick, Andreas Vitalis, Caitlin L Chicoine, and Rohit V Pappu. Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. *Proceedings of the National Academy of Sciences*, 107(18):8183–8188, 2010.
- [42] Francesca Massi, Eric Johnson, Chunyu Wang, Mark Rance, and Arthur G Palmer. Nmr r 1ρ rotating-frame relaxation with weak radio frequency fields. *Journal of the American Chemical Society*, 126(7):2247–2256, 2004.
- [43] William E Meador, Anthony R Means, and Florante A Quiocho. Target enzyme recognition by calmodulin: 2.4 a structure of a calmodulin-peptide complex. *Science*, 257(5074):1251–1255, 1992.
- [44] Tanja Mittag and Julie D Forman-Kay. Atomic-level characterization of disordered protein ensembles. *Current opinion in structural biology*, 17(1):3–14, 2007.
- [45] Tanja Mittag, Stephen Orlicky, Wing-Yiu Choy, Xiaojing Tang, Hong Lin, Frank Sicheri, Lewis E Kay, Mike Tyers, and Julie D Forman-Kay. Dynamic equilibrium engagement of a polyvalent ligand with a single-site receptor. *Proceedings of the National Academy of Sciences*, 105(46):17772–17777, 2008.
- [46] Piers Nash, Xiaojing Tang, Stephen Orlicky, Qinghua Chen, Frank B Gertler, Michael D Mendenhall, Frank Sicheri, Tony Pawson, and Mike Tyers. Multisite phosphorylation of a cdk inhibitor sets a threshold for the onset of dna replication. *Nature*, 414(6863):514–521, 2001.
- [47] E Nazarova, P Ear, C Hall, E OToole, S Kaitna, M Booth, D Abd-Rabbo, J Moreno, D Chen, P Francois, et al. Cdk1 phosphorylation of gamma-tubulin controls microtubule pairing and overlap. In *MOLECULAR BIOLOGY OF THE CELL*, volume 24. AMER SOC CELL BIOLOGY 8120 WOODMONT AVE, STE 750, BETHESDA, MD 20814-2755 USA, 2013.
- [48] Alexey Onufriev. Implicit solvent models in molecular dynamics simulations: A brief overview. *Annual Reports in Computational Chemistry*, 4:125–137, 2008.
- [49] Camilla L Oxley, Nicholas J Anthis, Edward D Lowe, Ioannis Vakonakis, Iain D Campbell, and L Wegener. An integrin phosphorylation switch the effect of $\beta 3$ integrin tail phosphorylation on dok1 and talin binding. *Journal of Biological Chemistry*, 283(9):5420–5426, 2008.

- [50] Valéry Ozanne, Frédéric Bauer, Loïc Salmon, Jie-rong Huang, Malene Ringkjøbing Jensen, Stéphane Segard, Pau Bernadó, Céline Charavay, and Martin Blackledge. Flexible-meccano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables. *Bioinformatics*, 28(11):1463–1470, 2012.
- [51] Christopher D Putnam, Michal Hammel, Greg L Hura, and John A Tainer. X-ray solution scattering (saxs) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Quarterly reviews of biophysics*, 40(03):191–285, 2007.
- [52] Ishwar Radhakrishnan, Gabriela C Pérez-Alvarado, David Parker, H Jane Dyson, Marc R Montminy, and Peter E Wright. Solution structure of the kix domain of cbp bound to the transactivation domain of creb: a model for activator: coactivator interactions. *Cell*, 91(6):741–752, 1997.
- [53] Theresa A Ramelot and Linda K Nicholson. Phosphorylation-induced structural changes in the amyloid precursor protein cytoplasmic tail detected by nmr. *Journal of molecular biology*, 307(3):871–884, 2001.
- [54] Pedro Romero, ZORAN Obradovic, Charles R Kissinger, J Ernest Villafranca, ETHAN Garner, STEPHEN Guillot, and A Keith Dunker. Thousands of proteins likely to have long disordered regions. In *Pac Symp Biocomput*, volume 3, pages 437–448, 1998.
- [55] Benjamin Schuler, Everett A Lipman, Peter J Steinbach, Michael Kumke, and William A Eaton. Polyproline and the spectroscopic ruler revisited with single-molecule fluorescence. *Proceedings of the National Academy of Sciences of the United States of America*, 102(8):2754–2759, 2005.
- [56] Alexander Sigalov, Dikran Aivazian, and Lawrence Stern. Homooligomerization of the cytoplasmic domain of the t cell receptor ζ chain and of other proteins containing the immunoreceptor tyrosine-based activation motif. *Biochemistry*, 43(7):2049–2061, 2004.
- [57] Kate A Stafford, Nikola Trbovic, Joel A Butterwick, Robert Abel, Richard A Friesner, and Arthur G Palmer. Conformational preferences underlying reduced activity of a thermophilic ribonuclease h. *Journal of molecular biology*, 427(4):853–866, 2015.

- [58] Peter Tompa. Intrinsically disordered proteins: a 10-year recap. *Trends in biochemical sciences*, 37(12):509–516, 2012.
- [59] Peter Tompa and Monika Fuxreiter. Fuzzy complexes: polymorphism and structural disorder in protein–protein interactions. *Trends in biochemical sciences*, 33(1):2–8, 2008.
- [60] Hoang T Tran, Albert Mao, and Rohit V Pappu. Role of backbone- solvent interactions in determining conformational equilibria of intrinsically disordered proteins. *Journal of the American Chemical Society*, 130(23):7380–7392, 2008.
- [61] Vladimir N Uversky, Christopher J Oldfield, and A Keith Dunker. Intrinsically disordered proteins in human diseases: introducing the d2 concept. *Annu. Rev. Biophys.*, 37:215–246, 2008.
- [62] Alard A Van Dijk, Liesbeth CM De Lange, William W Bachovchin, and George T Robillard. Effect of phosphorylation on hydrogen-bonding interactions of the active site histidine of the phosphocarrier protein hpr of the phosphoenolpyruvate-dependent phosphotransferase system determined by nitrogen-15 nmr spectroscopy. *Biochemistry*, 29(35):8164–8171, 1990.
- [63] Andreas Vitalis and Rohit V Pappu. Absinth: A new continuum solvation model for simulations of polypeptides in aqueous solutions. *Journal of computational chemistry*, 30(5):673–699, 2009.
- [64] Jacalyn Vogel, Ben Drapkin, Jamina Oomen, Dale Beach, Kerry Bloom, and Michael Snyder. Phosphorylation of γ -tubulin regulates microtubule organization in budding yeast. *Developmental cell*, 1(5):621–631, 2001.
- [65] Jacalyn Vogel and Michael Snyder. The carboxy terminus of tub4p is required for gamma-tubulin function in budding yeast. *Journal of cell Science*, 113(21):3871–3882, 2000.
- [66] Andrew M Waterhouse, James B Procter, David MA Martin, Michèle Clamp, and Geoffrey J Barton. Jalview version 2a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, 25(9):1189–1191, 2009.
- [67] Deborah K Wilkins, Shaun B Grimshaw, Véronique Receveur, Christopher M Dobson, Jonathan A Jones, and Lorna J Smith. Hydrodynamic radii of native and denatured proteins measured by pulse field gradient nmr techniques. *Biochemistry*, 38(50):16424–16431, 1999.

- [68] Peter E Wright and H Jane Dyson. Linking folding and binding. *Current opinion in structural biology*, 19(1):31–38, 2009.
- [69] Peter E Wright and H Jane Dyson. Intrinsically disordered proteins in cellular signalling and regulation. *Nature Reviews Molecular Cell Biology*, 16(1):18–29, 2015.
- [70] Kurt Wüthrich. The way to nmr structures of proteins. *Nat Struct Biol*, 8(11):923–5, 2001.
- [71] Mi-Kyung Yoon, Diana M Mitrea, Li Ou, and Richard W Kriwacki. Cell cycle regulation by the intrinsically disordered proteins p21 and p27. *Biochemical Society Transactions*, 40(5):981–988, 2012.
- [72] Hongtao Yu, James K Chen, Sibo Feng, David C Dalgarno, Andrew W Brauer, and Stuart L Schreiber. Structural basis for the binding of proline-rich peptides to sh3 domains. *Cell*, 76(5):933–945, 1994.
- [73] Ruhong Zhou. Free energy landscape of protein folding in water: explicit vs. implicit solvent. *Proteins: Structure, Function, and Bioinformatics*, 53(2):148–161, 2003.
- [74] Tsaffrir Zor, Bernhard M Mayr, H Jane Dyson, Marc R Montminy, and Peter E Wright. Roles of phosphorylation and helix propensity in the binding of the kix domain of creb-binding protein by constitutive (c-myb) and inducible (creb) activators. *Journal of Biological Chemistry*, 277(44):42241–42248, 2002.