

Structure, Stability and Dynamics of an Intrinsically Disordered Protein AtPP16-1 and its Mutants

A Thesis

Submitted for the Degree of

DOCTOR OF PHILOSOPHY

By

DIDLA SANTI SWARUPINI



SCHOOL OF CHEMISTRY

UNIVERSITY OF HYDERABAD

INSTITUTE OF EMINENCE

HYDERABAD 500 046

INDIA

February 2022

**DEDICATED TO
MY BELOVED PARENTS**



School of Chemistry

University of Hyderabad

Hyderabad – 500 046

STATEMENT

I hereby declare that the matter embodied in this thesis is the result of investigations carried out by me in the School of Chemistry, University of Hyderabad, Hyderabad, under the supervision of **Prof. Abani K Bhuyan**.

In keeping with the general practice of reporting scientific observations, due acknowledgments have been made whenever the work described is based on the finding of other investigators. Any omission which might have occurred by oversight or error is regretted.

Hyderabad

February 2022

D. S. Swarupini
Didla Santi Swarupini

(16CHPH05)



DECLARATION

I, **Didla Santi Swarupini** hereby declare that the thesis entitled “**Structure, stability, and dynamics of the intrinsically disordered protein AtPP16-1 and its mutants.**” submitted by me under the supervision of Prof. **Abani K. Bhuyan** is a bonafide research work which is free from plagiarism. I also declare that it has not been submitted previously in part or in full to this University or any other University or Institution for the award of any degree or diploma. I hereby agree that my thesis can be deposited in Shodhganga/INFLIBNET.

A report on plagiarism from the University Library is enclosed.

Prof. Abani K. Bhuyan

Thesis Supervisor

ABANI K. BHUYAN, Ph.D.
Professor
University of Hyderabad
School of Chemistry
Hyderabad - 500 046.

Didla Santi Swarupini
Reg. No.:16CHPH05

Prof. Abani K. Bhuyan
Professor
School of Chemistry
University of Hyderabad
Hyderabad-500 046, India



Phone: +91-40-2313 4810 (O)
Fax: +91-40-2301 2460
E-mail: akbsc@uohyd.ac.in

CERTIFICATE

This is to certify that the thesis entitled **“Structure, stability, and dynamics of an intrinsically disordered protein AtPP16-1 and its mutants”** submitted by **Ms. Didla Santi Swarupini** bearing registration number 16CHPH05 in partial fulfillment of the requirements for award of the Doctor of Philosophy (Ph. D) is a bonafide work carried out by her under my supervision and guidance in the School of Chemistry, University of Hyderabad. This thesis is free from plagiarism and has not been submitted previously in part or in full to this or any other University or Institution for award of any degree or diploma.

Further, the student has two publications before submission of the thesis for adjudication and has produced evidences for the same in the form of reprints.

Parts of this thesis have been published in the following two publications:

1. Santi Swarupini, D., & Bhuyan, A. K. (2018). Amyloid fibrillation of an intrinsically disordered plant phloem protein AtPP16-1 under acidic condition. *Biophysical chemistry*, 237, 1–8. (Chapter 2)
2. Negative Thermal Expansion on of an Intrinsically Disordered Protein, accepted for publication in *Chemical Physics*, 2022 (Chapter 3).

She has also made presentations in the following conferences:

1. Oral presentation at “CHEMFEST-2021” Annual in-house Symposium of School of Chemistry, University of Hyderabad.
2. Poster presentation at “CHEMFEST-2019” Annual in-house Symposium of School of Chemistry, University of Hyderabad.

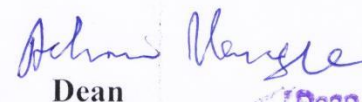
Further the student has passed the following courses towards fulfillment of the coursework requirement for Ph. D. degree:

Sl. No.	Course	Title	Credits	Grade/Status
1.	CY-801	Research Proposal	3	B/pass
2.	CY-802	Chemistry Pedagogy	3	B/pass
3.	CY-803	Instrumentation-A	3	B+/pass
4.	CY-455	Biological Chemistry	3	C/pass



Prof. Abani K. Bhuyan

Thesis Supervisor
ABANI K. BHUYAN, Ph.D.
Professor
University of Hyderabad
School of Chemistry
Hyderabad - 500 046.


Dean

SCHOOL OF CHEMISTRY
University of Hyderabad
Hyderabad-500 046
School of Chemistry
University of Hyderabad

ACKNOWLEDGEMENTS

I express my gratitude to my PhD supervisor Prof. Abani K. Bhuyan, school of chemistry, University of Hyderabad, for his guidance, freedom, and valuable suggestions given me throughout the PhD program.

I wish to thank Prof. Ashwini Nangia, Dean of School of Chemistry, as well as all former Deans of the School for providing the infrastructure for carrying out the research work. I extend my sincere thanks to all faculty members of the school for their cooperation on various occasions.

I express my deepest gratitude to the thesis Doctoral Committee members Prof. Lalitha Guruprasad, Prof. M.J. Swami, who have been supportive throughout the course. I extend my sincere gratitude to all of my teachers, who have guided me up to this stage.

I was extraordinarily fortunate to work with a very good company of people in the laboratory. Many thanks to Dr. P. Sashi, Girija, Noorul, Mujahid, Kirthi, and Ramesh for providing a lovely working environment for all these years, they helped to make my time spent in the laboratory a very memorable one indeed.

I thank all the research scholars of the School of Chemistry and especially those who helped me in need. I would also acknowledge all the non-teaching staff, the School of Chemistry and CIL for their help. Non-Net fellowship and financial support from university are gratefully acknowledged.

I thank all my school and college teachers and my college friends for their support. I wish to express my heartfelt gratitude to my dear sir, Dr. Srinivas for his constant support and encouragement.

I would like to thank all faculty members in the Department of Chemistry, Andhra University, for their encouragement and guidance up to this stage.

I am very pleased to thank all my friends and classmates namely, Ms. Nayomi, Ms. Sabitha, Ms. Prasannatha, Ms. Prachurita, Ms. Tejaswi, Ms. Shanthi Priya, Ms. Swathi Priya, Ms. Jyothi, Ms. Sahitya, Ms. Maneesha, Ms. Adhi Lakshmi, Ms. Jhansi, Ms. Kalyani, Ms. Kumari, Ms. Sonali, Ms. Tejaswini, Ms. Esther, Ms. Vijaya, Ms. Mahalakshmi, Ms. Jenny, Ms. Anjali, Mr. Raju Mr. Chuku, Mr. Kirankumar, Mr. Calvin

Samuel, Dr Yerranna, Dr Y Rajesh, Dr Sujai, Dr. Srujana, Dr Manikanta, Dr Lavanya, Ms,Shakunthala family for great fellowship and friendship I enjoyed every movement with you all, without you I wouldn't have been enjoyed here. I thank Mr. & Ms. Andrew & Arapana family for their prayers and encouragement, my beloved sisters Ms. Baby, Ms. Nalini, Ms. Sujana, Ms. Jaya, Ms. Gowri, Ms. Priya for their prayers and support.

This thesis would be incomplete without my family members who have stood beside me at all stages of this journey. My mother and father deserve special mention for their inseparable support and prayers. With all their love, blessings and sacrifice only this could be possible. I am very grateful to my beloved parents, D.V. Ramakrishna and D. Madhavi and my special thanks to my beloved brother Dr Syam for his care and love, constant support and motivation in the hard times.

Above all, I would like to thank Almighty who granted and bestowed blessings upon me.

D. SANTI SWARUPINI

ABBREVIATIONS

<i>AtPP16-1</i>	<i>Arabidopsis thaliana</i> phloem protein 16-1
IDP	Intrinsically Disordered Protein
LB	Luria Broth
IPTG	Isopropyl -D-1-thiogalactopyranoside
Ni-NTA	Nickel-Nitrilotriacetic acid
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
ANS	8-Anilinonaphthalene-1-sulfonate
CD	Circular dichroism
RBP	RNA-Binding Protein
FWHM	Full width at half maximum
DSS	4,4-dimethyl-4-silapentane-1-sulfonate
MRE	Mean Residue Ellipticity

CONTENTS

Statement	i
Declaration	ii
Certificate	iii-iv
Acknowledgement	v-vi
Abbreviations	vii

CHAPTER 1: Introduction

1.1 Intrinsically Disordered Proteins	1-5
1.1.1 Unique 3D structure versus ensemble of conformers	1
1.1.2 Evolution, mutation, and disorder	2
1.1.3 Amino acid composition and backbone conformation	3
1.1.4 Binding and functions	4-5
1.2 <i>AtPP16-1</i> , an IDP	5-8
1.3 References	8-12

CHAPTER 2: Amyloid fibrillation of *AtPP16-1* under acidic condition

2.1 Abstract	13
2.2 Introduction	14-16
2.3 Materials and methods	16-17
2.3.1 Protein preparation	16
2.3.2 pH titration and ANS binding	17
2.3.3 Kinetics of fibrillation	17
2.3.4 NMR measurement	17
2.3.5 AFM, FESEM, and TEM imaging	17
2.4 Result	18-25
2.4.1 Purified <i>AtPP16-1</i> and pH-dependent conformation	18
2.4.2 Kinetics of <i>AtPP16-1</i> fibrillation	18-22

2.4.3	Images and dimensions of AtPP16-1 fibrils	22-24
2.5.	Discussion	25-31
2.5.1	The protein AtPP16-1 forms amyloid fibrils, not amorphous aggregates	25
2.5.2	pH-dependence and the mechanism of precursor aggregate (PA) formation	26-30
2.5.3	Condensation of precursor aggregates to protofibrils	30-31
2.5.4	Fibrillation of IDP	31
2.6	References	32-38
CHAPTER 3: Negative Thermal Expansion of AtPP16-1		39-60
3.1	Abstract	39
3.2	Introduction	40-41
3.3	Methods	42
3.4	Results and Discussion	42-55
3.4.1	Temperature window to study thermal shift of resonances	42-44
3.4.2	Positive temperature coefficient for the chemical shift of backbone atoms	44-45
3.4.3	Contracted H $\cdots$ O distances in the >N-H $\cdots$ O=C< H-bonds.	45-48
3.4.4	Temperature dependence of backbone C $_{\alpha}$ and $^1\text{H}^{\text{C}_{\alpha}}$ chemical shifts	49
3.4.5	Thermal shortening of global correlation time, τ_c	49-50
3.4.6	Thermally demoted volume fluctuation.	50-51
3.4.7	Basis for Negative Thermal Expansion	52-53
3.5	References	53-60
CHAPTER 4: Electrostatic effect on the folding of AtPP16-1		61-72
4.1	Abstract	61
4.2	Introduction	61-62
4.3	Materials and methods	62
4.4	Results and Discussion	63-71
4.4.1	A revisit to the pH dependence of AtPP16-1 conformers	63

4.4.2	Folding and stability at pH 4.1	64-66
4.4.3	The excitement of charge screening	66-69
4.4.3.1	pH 6.3	66-68
4.4.3.2	pH 4.1	68-69
4.4.4	The NaCl stabilized/structured proteins are off-pathway	69-71
4.5	Conclusions	71
4.6	References	71-72
CHAPTER 5 Cysteine substitution mutagenesis of AtPP16-1 alters the average conformational ensemble throughout the backbone		73-98
5.1	Abstract	73
5.2	Introduction	74-76
5.3	Materials and methods	76-78
5.3.1	Site-directed mutagenesis and protein purification	76
5.3.2	pH titration	77
5.3.3	Equilibrium unfolding	77
5.3.4	NMR Spectroscopy	78
5.4	Results and Discussion	78-92
5.4.1	Secondary and tertiary structures of Ala→Cys mutants	78-80
5.4.2	Stability and folding	80-85
5.4.3	pH-dependent changes in the structure	85-88
5.4.4	¹ H, ¹⁵ N backbone assignment of the cysteine mutants	89-92
5.5	Conclusions	92
5.6	References	93-98
	List of publications	99
	Conferences/ Presentations	100
	Plagiarism Certificate	

CHAPTER 1

INTRODUCTION

1.1. Intrinsically Disordered Proteins

Divulging the principles of structure, stability, folding, and function of proteins is a century-old huge burden mainly because the protein system presents a many-body problem which is compounded by the aqueous environment in which it thrives and works. Progress has been made in deciphering some of the basic principles of protein phenomena even if newer problems continue to surface. Research of the past three decades has also shown that a sizable set of proteins work only with inherent structural disorder. This finding has greatly influenced the idea of the structure-function paradigm, which initially considered that a uniquely folded 3D structure of a protein is needed for its functional performance. The repertoire of structurally disordered proteins engaged in the cellular functioning insists that the structure-function paradigm is not unique. It has also raised numerous questions in the areas of amino acid sequence evolution, protein folding, and functional preference of proteins. This dissertation describes facets of structure, stability, and dynamics of a disordered protein by basic optical and NMR spectroscopic methods. Some principles and the identity of structurally disordered proteins are briefly reviewed below.

1.1.1. Unique 3D structure versus ensemble of conformers

With respect to structure and compactness, proteins are of two kinds – structurally ordered globular and loosely packed disordered. A globular protein has a well-defined 3D structure, which is the lowest energy structure the unfolded chain folds to, and which is needed to carry out its specific function. Electron-transfer proteins, native functional enzymes, and oxygen transporter and storage proteins all are ordered globular proteins, the 3D structures of many of which have been determined to high resolution by X-ray

crystallography and NMR spectroscopy. On the other hand, an amino acid sequence containing polypeptide segments which can collapse but do not fold to a defined compact structure produces an ensemble of conformers containing unstructured or disordered regions. Since these disordered regions are inherent or intrinsic, obviously due to unbalanced composition of amino acids, the proteins containing such structural disorder are called intrinsically disordered proteins (IDPs).

The flexibility of the disordered regions produces an ensemble of IDP conformers within which the protein molecule equilibrates. These conformers can be treated as ‘static disorder’ in the parlance of crystallography because the diffraction of X-rays is instantaneous so that the different conformers are not resolved, and instead one obtains a time-averaged diffraction pattern. In contrast, spectroscopic methods such as NMR and fluorescence provide information about dynamic disorder, meaning the equilibrating conformers are resolved during measurement. The prevalence of disorders in proteins was however suspected in the 1970s when crystallographers began to sense limited and widely separated conformers emanating from conformational flexibility of proteins.¹ The disorder in the solution structure of many proteins has been decisively established by NMR spectroscopy since the 1990s.^{2–15} At least a third of the cellular proteins are now thought to be intrinsically disordered.

1.1.2. Evolution, mutation, and disorder

IDPs are abundant across proteomes, from viruses to eukaryotes, and their complexity increases with evolutionarily higher organisms.^{16–19} Of the respective total proteins, disorder-containing proteins are in the range 7–33%, 9–37%, and 36–63% in bacterial, archaeobacterial, and eukaryotic proteomes, respectively.¹⁷ Since the 3D structure is encoded in the primary structure, the disorder of IDP structure must also be coded by some composition of amino acids in the sequence distinct from that coding for compact globular structures. The amino acid composition of IDPs is less complex compared to the complexity encountered in ordered proteins^{20–23} which is likely due to higher rate of evolution of disordered proteins.^{24,25} By corollary, the disordered regions within an IDP generally evolve more rapidly than the structured regions.²⁶ Since IDPs scarcely have

long-range intramolecular contacts, the sequence is not constrained with respect to evolutionary changes.²⁷

Evolution of IDPs by point mutations changes their sequences at a higher frequency, albeit in a less conservative manner in comparison to that of ordered proteins.^{28,29} It should be obvious that such changes must be of the missense type, because silent mutations will bring about no change in the sequence and the nonsense type will halt translation, abrogating the protein. There are also abundant addition and deletion mutations leading to expansion or collapse of the IDP dimension.³⁰ Some mutations in the disordered regions can lead to a loss of function and could be deleterious even. Mutations leading to a disorder-to-order transition or various degrees of deviation in the conformation of the disordered regions are expected to cause functional revocation because IDPs function only when the disordered regions continue to be disordered. Some mutations may result in aggregation of IDPs, and the plaques so formed can lead to neurodegenerative human conditions.^{31,32}

1.1.3. Amino acid composition and backbone conformation

The low complexity amino acid sequence of IDPs^{21,23} have evolved to fold to an ensemble of compact conformers, albeit the extent of compaction is much less compared to that of ordered proteins. Such sequences are generally low in hydrophobic amino acids and high in mean net charge.³³ The low hydrophobicity should mean the lack of a hydrophobic core, or a much loosely packed core if it exists. But the low hydrophobic content may also prevent a chain collapse, an event that initiates folding. It is possible that the IDP backbone, which is fairly naked due to a high content of the neutral residue glycine, interacts unfavorably with water – a poor solvent, and hence contracts.³⁴

IDPs carry a high load of amino acids with charged side chains (Asp, Glu, Arg, His, and Lys), accounting for ~30% of the total residues,³³ suggesting that Coulombic interactions may influence conformation and dimension of disordered proteins. The number and distribution of opposite charges along the polypeptide is thought to determine the chain compaction and conformations.^{35,36} Thus expansion and compaction of IDP conformers depend on the charge content and intramolecular Coulombic attraction

and repulsion. One senses that an imbalance of the interactions can lead to the loss of both conformation and function. The imbalance can be brought about intrinsically by substitution mutations, and extrinsically by screening the charges by solvent ions.

1.1.4. Binding and functions

IDPs bind to effective molecules for posttranslational modifications, are involved in binding to target molecules for signaling and molecular recognition and assembly, and bind to DNA and RNA molecules.^{32,37} When posttranslational modification of an IDP is required, a short disordered region (usually ~10 residues), often also called interaction peptide motif, interfaces with the partner molecule – an enzyme, for example. For instance, the incising motif of caspase-3 interfaces with the [DE]XXD[GSAN] motif of the disordered region of the substrate protein and cleaves the site.³⁸

In carrying out its signaling function, the IDP binding to the target protein leads to folding of extended regions of the former, even though disorder in small regions may still prevail. The binding energy of such interactions is low, meaning low affinity interaction, but the binding is highly specific. It has been proposed that lower binding affinity (characterized by the equilibrium association constant $K_a = k_a/k_d$, where k_a is the diffusion-controlled bimolecular binding rate, and k_d is the dissociation rate of the complex) raises the dissociation rate according to $k_a = K_a k_d$, resulting in a short life of the bimolecular complex and hence a rapid release of the target protein at the end of signaling.³⁹ For example, the cyclin dependent kinase (CDK) whose activity directs DNA replication and cell division exists as a CDK/cyclin complex, and the IDP p27^{kip1}, also called p27 simply, binds to the complex to inhibit the kinase activity of CDK at G1/S and G2/M check points of the cell cycle.⁴⁰

IDPs have been shown to function as a hub for protein networking for molecular assemblies⁴¹ that are often complex in size and structure. For example, AP2 is an adaptor (hub) protein whose different domains bind cargo proteins and accessory proteins to produce an assembly of proteins during endocytosis.⁴² The disordered flexible linkers essentially serve as hinges for moving the proteins bound to the adaptor molecule.

Kinases, transcription factors, and splicing factors bind to nucleic acids by using

their disordered regions. To elucidate, intrinsically disordered transcription factors, often also called homeodomain proteins because they contain a highly conserved DNA-binding homeodomain encoded by a homeobox (~180 base pair DNA sequence). The homeodomain generally has 60 amino acids which fold to an N-terminal disordered region followed by three helices. While the helices help in docking the homeodomain near the surface of DNA it is the N-terminal linker which establishes multiple contacts with the minor groove aided by electrostatic interactions. A well-studied example of modulation of gene expression by IDPs is the interaction of HOX transcription factors with DNA, the modulation of the expression of which regulate developmental patterns.⁴³ In some instances, two homeodomain proteins may collaborate to regulate gene expression.

1.2. AtPP16-1, an IDP

The *Arabidopsis thaliana* phloem protein 16-1 (AtPP16-1) is a RNA-binding protein (RBP) and a putative mRNA transporter. There are at least 200 different RBPs coded by the genome of *A. thaliana*,⁴⁴ and the homologs AtPP16-1 and AtPP16-2, both expressed highly in flower and stem but sparsely or not at all in leaf⁴⁵ are identical by 35% in terms of conserved amino acid positions. The sequence conservation for AtPP16-1 and AtPP16-2 appears to suggest that they are functionally similar. These speculations provided the initial rationale to inquire into the structure and function of AtPP16. Since not much is known about RBP structures and in vivo roles of RBP in phloem transport, the solution NMR structure of AtPP16-1 was solved in this laboratory (Sashi et al 2018).¹⁴ It is not a compact globular protein, the structure consists of several disordered regions, the longest one being the stretch of residues 10–55 that connects strands β I and β II (Figure 1.1).

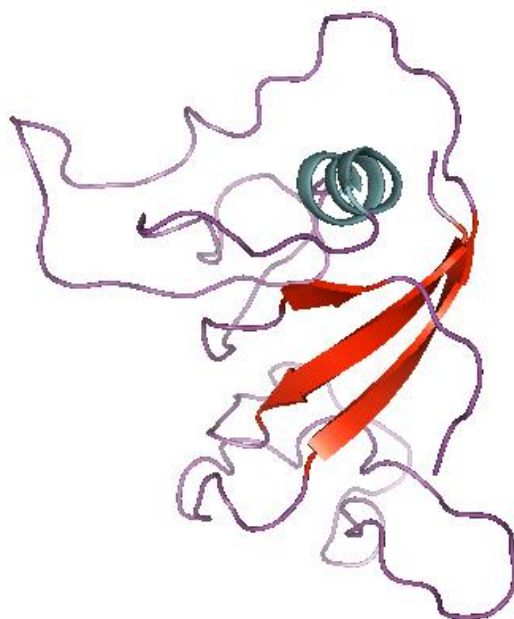


Figure 1.1. Ribbon diagram of AtPP16-1 showing many disordered regions.

The amino acid composition of AtPP16-1 listed in Table 1.1 is characteristic of the composition of IDP – low in hydrophobic but rich in polar and charged residues. There also are 18 Gly (G) residues, which are important because generic IDPs are rich in glycine.

Table 1.1. Amino acid composition of AtPP16-1*

Hydrophobic		Polar (neutral)		Acidic (charged)		Basic (charged)	
Ala (A)	6	Asn (N)	3	Asp (D)	14	Arg (R)	6
Ile (I)	8	Cys (C)	0	Glu (E)	11	His (H)	3
Leu (L)	14	Gln (Q)	5			Lys (K)	14
Met (M)	4	Ser (S)	13				
Val (V)	13	Thr (T)	6				

Phe (F) 6 Pro (P) 4

Trp (W) 3

Tyr (Y) 5

* There are 18 Gly (G) residues that are considered neutral.

AtPP16-1 binds nucleic acids. By combining NMR data and simulations, the structure of the complex of the protein with Dickerson-Drew dodecamer self-complementary sequence ds-(5' d-CGCGAATTCGCG 3') was derived (Sashi et al 2018). Residues K14, L16, Y28, R60, T92, and S123 are found to interact with the dodecamer DNA by H-bonding interactions (Figure 1.2), the hydrogen bond lengths being in the range 2.4–3.0 Å. Residues K14, L16, Y28, and T92 interact with DNA by respective main-chain NH, while the main-chain carbonyls of T92 and S123 form two H-bonds each. The side-chain OH of Y28, NH1 of R60, and OH of S123 also interact with DNA by H-bonding. The DNA-binding residues belong to disordered sequence segments, albeit a continuous stretch of interacting residues have not been identified, meaning several disordered regions participate in the interaction.

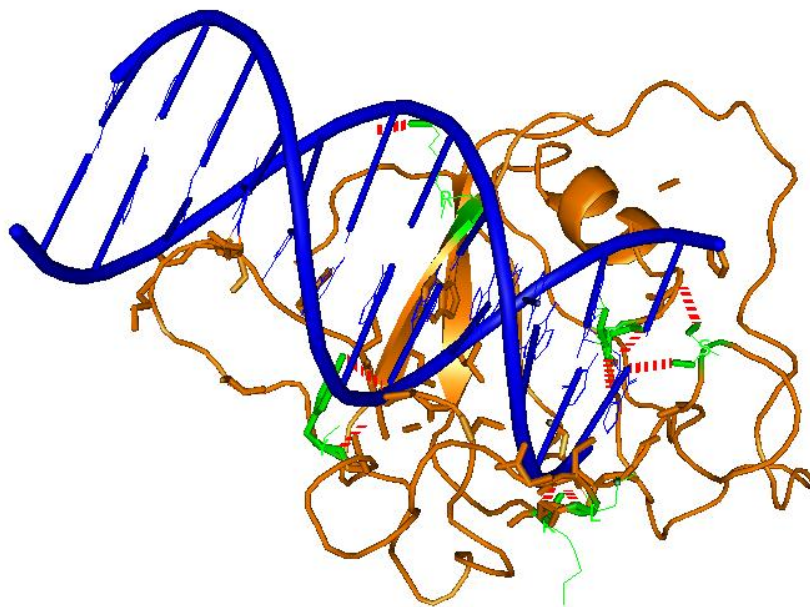


Figure 1.2. Interaction of AtPP16-1 with the dodecamer DNA. Hydrogen bonding residues and bond lengths (Å) are labeled.

1.3. REFERENCES

1. Huber, R. (1979). Conformational flexibility and its functional significance in some protein molecules. *Trends Biochem. Sci.* **4**, 271–276.
2. Riek, R., Hornemann, S., Wider, G., Billeter, M., Glockshuber, R., and Wuthrich, K. (1996). NMR structure of the mouse prion protein domain PrP(121–321). *Nature* **382**, 180–182.
3. Muchmore, S. W., Sattler, M., Liang, H., Meadows, R. P., Harlan, J. E., Yoon, H. S., Nettesheim, D., Chang, B. S., Thompson, C. B., Wong, S. L., Ng, S. L., and Fesik, S. W. (1996). X-ray and NMR structure of human Bcl-x_L, an inhibitor of programmed cell death. *Nature* **381**, 335–341.
4. Kriwacki, R. W., Hongst, L., Tennant, L., Reed, S. I., and Wright, P. E. (1996). Structural studies of p21^{Waf1/Cip1/Sdi1} in the free and Cdk2-bound state: conformational disorder mediates binding diversity. *Proc. Natl. Acad. Sci. USA* **93**, 11504–11509.

5. Fletcher, C. M., and Wagner, G. (1998). The interaction of eIF4E with 4E-BP1 is an induced fit to a completely disordered protein. *Protein Sci.* **7**, 1639–1642.
6. Wright, P. E., and Dyson, H. J. (1999). Intrinsically unstructured proteins: re-assessing the protein structure–function paradigm. *J. Mol. Biol.* **292**, 321–331.
7. Dyson, H. J., and Wright, P. E. (2004). Unfolded proteins and protein folding studied by NMR. *Chem. Rev.* **104**, 3607–3622.
8. Mittag, T., and Forman-Kay, J. D. (2007). Atomic-level characterization of disordered protein ensembles. *Curr. Opin. Struct. Biol.* **17**, 3–14.
9. Wright, P. E., and Dyson, H. J. (2009). Linking folding and binding. *Curr. Opin. Struct. Biol.* **19**, 31–38.
10. Eliezer, . Biophysical characterization of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* **19**, 23–30.
11. Mittag, T., Marsh, J., Grishaev, A., Orlicky, S., Lin, H., Sicheri, F., Tyers, M., and Forman-Kay, J. D. (2010). Structure/function implications in a dynamic complex of the intrinsically disordered Sic1 with the Cdc4 subunit of a SCF ubiquitin ligase. *Structure* **18**, 494–506.
12. Kosol, S., Contreras-Martos, S., Cedeno, C., and Tompa, P. (2013). Structural characterization of intrinsically disordered proteins by NMR spectroscopy. *Molecules* **18**, 10802–10828.
13. Felli, I., Connelli, L., and Pieratelli, R. (2014). In-cell ^{13}C NMR spectroscopy for the study of intrinsically disordered proteins. *Nat. Protocols* **9**, 2005–2016.
14. Sashi, P., Singarapu, K. K., and Bhuyan, A. K. (2018). Solution NMR structure and backbone dynamics of partially disordered Arabidopsis thaliana phloem protein 16-1, a putative RNA transporter. *Biochemistry* **57**, 912–924.
15. Adamski, A., Salvi, N., Maurin, D., Magnat, J., Milles, S., Jensen, M. R., Abyzov, A., Moreau, C. J., and Blackledge, M. (2019). A unified description of intrinsically disordered protein dynamics under physiological conditions using NMR spectroscopy. *J. Am. Chem. Soc.* **141**, 17817–17829.

16. Dunker, A. K., Obradovic, Z., Romero, P., Garner, E. C., and Brown, C. J. (2000). Intrinsic protein disorder in complete genomes. *Genome Inform. Ser. Workshop Genome Inform.* **11**, 161–171.
17. Dunker, A. K., Lawson, J. D., Brown, C. J., Williams, R. M., Romero, P., Oh, J. S., Oldfield, C. J., Campen, A. M., Ratcliff, C. M., Hipps, K. W., Ausio, J., Nissen, M. S., Reeves, R., Kang, C., Kissinger, C. R., Bailey, R. W., Griswold, M. D., Chiu, W., Garner, E. C., and Obradovic, Z. (2001). Intrinsically disordered protein. *J. Mol. Graph Model* **19**, 26–59.
18. Ward, J. J., Sodhi, J. S., McGuffin, L. J., Buxton, B. F., and Jones, D. T. (2004). Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J. Mol. Biol.* **337**, 635–645.
19. Oldfield, C. J., Cheng, Y., Cortese, M. S., Brown, C. J., Uversky, V. N., and Dunker, A. K. (2005). Comparing and combining predictors of mostly disordered proteins. *Biochemistry* **44**, 1989–2000.
20. Romero, P. R., Obradovic, Z., and Dunker, A. K. (1999). The complexity of disorder. *Biophys. J.* **76**, A1–A525.
21. Sim, K. L., and Creamer, T. P. (2002). Abundance and distributions of eukaryote protein simple sequences. *Mol. Cell. Proteomics* **1**, 983–995.
22. Radivojac, P., Iakoucheva, L. M., Oldfield, C. J., Obradovic, Z., Uversky, V. N., and Dunker, A. K. (2007). Intrinsic disorder and functional proteomics. *Biophys. J.* **92**, 1439–1456.
23. Weathers, E. A., Paulaitis, M. E., Woolf, T. B., and Hoh, J. H. (2007). Insights into protein structure and function from disorder-complexity space. *Proteins* **66**, 16–28.
24. Shafee, T., Bacic, A., and Johnson, K. (2020). Evolution of sequence-diverse disordered regions in a protein family: order within the chaos. *Mol. Biol. Evol.* **8**, 2155–2172.
25. Kastano, K., Erdős, G., Mier, P., Alanis-Lobato, G., Prompons, V. J., Dosztányi, Z., and Andrade-Navarro, M. A. (2020). Evolutionary study of disorder in protein sequences. *Biomolecules* **10**, 1413.

26. Brown, C. J., Takeyama, S., Campen, A. M., Vise, P., Marshall, T. W., Oldfield, C. J., Williams, C. J., and Dunker, A. K. (2002). Evolutionary rate heterogeneity in proteins with long disordered regions. *J. Mol. Evol.* **55**, 104–110.
27. Dunker, A. K., Olfield, C. J., Meng, J., Romero, P., Yang, J. Y., Chen, J. W., Vacic, V., Obradovic, Z., and Uversky, V. N. (2008). The unfoldomics decade: an update on intrinsically disordered proteins. *BMC Genomics* **9** (Suppl 2), S1.
28. Brown, C. J., Johnson, A. K., Daughdrill, G. W. (2010). Comparing models of evolution for ordered and disordered proteins. *Mol. Biol. Evol.* **27**, 609–621.
29. Brown, C. J., Johnson, A. K., Dunker, A. K., and Daughdrill, G. W. (2011). Evolution and disorder, *Curr. Opin. Struct. Biol.* **21**, 441–446.
30. Das, R. K., Ruff, K. M., and Pappu, R. V. (2015). Relating sequence encoded information to form and function of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* **52**, 102–112.
31. Vacic, V., Iakoucheva, L. M. (2012). Disease mutations in disordered regions – exception to the rule? *Mol. Biosyst.* **8**, 27–32.
32. Babu, M. M. (2016). The contribution of intrinsically disordered regions to protein function, cellular complexity, and human disease. *Biochem. Soc. Trans.* **44**, 1185–1200.
33. Uversky, V. N., Gillespie, J. R., and Fink, A. L. (2000). Why are “natively unfolded” proteins unstructured under physiologic conditions? *Proteins* **41**, 415–427.
34. Tran, H. T., Mao, A., and Pappu, R. V. (2008). Role of backbone-solvent interaction in determining conformational equilibria of intrinsically disordered proteins. *J. Am. Chem. Soc.* **130**, 7380–7392.
35. Mao, A. H., Crick, S. L., Vitalis, A., Chicoine, C. L., and Pappu, R. V. (2010). Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. *Proc. Natl. Acad. Sci. USA* **107**, 8183–8188.
36. Das, R. K., and Pappu, R. V. (2013). Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proc. Natl. Acad. Sci. USA* **110**, 13392–13397.

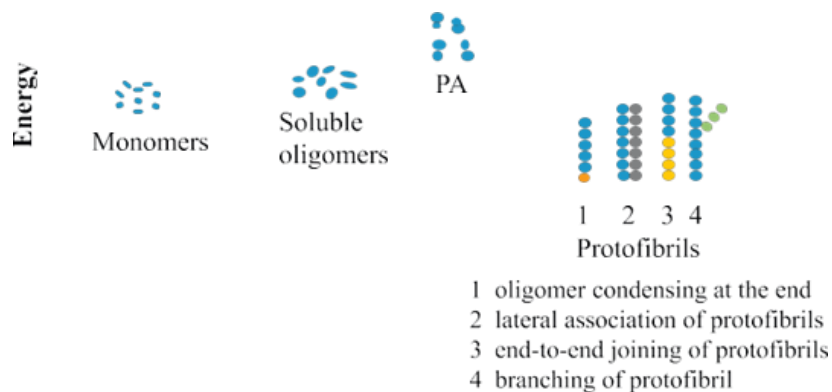
37. Radivojac, P., Vacic, V., Haynes, C., Cocklin, R. R., Mohan, A., Heyen, J. W., Goebel, M. G., and Iakoucheva, L. M. (2010). Identification, analysis, and prediction of protein ubiquitination sites. *Proteins: Struct. Funct. Bioinf.* **78**, 365–380.
38. Tompa, P., Davey, N. E., Gibson, T. J., and Babu, M. M. (2014). A million peptide motifs for the molecular biologist. *Mol. Cell.* **55**, 161–169.
39. Zhou, H. –X. (2012). Intrinsic disorder: signaling via highly specific but short – lived association. *Trends Biochem. Sci.* **37**, 43–48.
40. Yoon, M-K., Mitrea, D. M., Ou, L., and Kriwacki, R. W. (2012). Cell cycle regulation by the intrinsically disordered proteins, p21 and p27. *Biochem. Soc. Trans.* **40**, 981–988.
41. Dunker, A. K., Cortese, M. S., Romero, P., Iakoucheva, L., and Uversky, V. N. (2005). Flexible nets. The roles of intrinsic disorder in protein interaction networks. *FEBS J.* **272**, 5129–5148.
42. Schmid, E. M., and McMahon, H. T. (2007). Integrating molecular and network biology to decode endocytosis. *Nature* **448**, 883–888.
43. Dunker, A. K., Bondos, S. E., Huang, F., and Oldfield, C. J. (2015). Intrinsically disordered proteins and multicellular organisms. *Semin. Cell. Dev. Biol.* **37**, 44–55.
44. Lorković, Z. J. (2009). Role of plant mRNA-binding proteins in development, stress response and genome organization. *Trends Plant Sci.* **14**, 229–236.
45. Hsi, M. (2007) Genetic analysis of *AtPP16-1* and *AtPP16-2*. *Explorations* **10**, 89–96.

CHAPTER 2

Amyloid fibrillation of *AtPP16-1* under acidic condition

2.1. ABSTRACT

Arabidopsis thaliana Phloem Protein 16-1 (*AtPP16-1*) is a 156-residue intrinsically disordered nucleic acid binding protein which is putatively involved in long-distance systemic transport of RNA to budding regions in plants. Dimerization or oligomerization of the protein at pH higher than about 4.1 leaves no apolar surface exposed for interaction with the dye 8- Anilinonaphthalene-1-sulfonate (ANS). The most stable monomeric state is found near pH 4 where the structure of the protein is determined to have three short β -strands and a single α - helix. By surveying the pH-dependent propensity of fibrillation we find the protein enters the amyloidogenic state at pH 2, 60°C. The reaction product is not amorphous aggregate, but simple amyloid fibrils with sparse or no branching. The mean diameters of the fibril population scaled from AFM images are 13.2 and 21.2 nm for precursor aggregates (PA) and proto- or elongated fibrils, respectively. These values are somewhat larger than the fibril diameters generally cited, and the reason could be larger lateral association for both PA and protofibrils. The question emanating is: do phloem proteins fibrillate in vivo, and if so what implication fibrillation has for plant physiology?



2.2. INTRODUCTION

The association of un-branched elongated fibrils with a variety of amyloid diseases, including prion and Alzheimer^{1,2} has been known for a long time. But numerous globular proteins have been shown to form fibrils or forced to enter the amyloidogenic state under a variety of conditions³⁻⁶ that may not even be physiologic. In general, the proteins are partly denatured in acid or alkali, or by adding solvent additives so as to convert them to branched or un-branched fibrillar-states. Fibrillation observed for overwhelmingly large number of proteins establish that the fibrillar-state is the fourth state of proteins – the native, unfolded, and molten globule states are the other three. In fact it is the premolten globule-like state that is often found to enter the amyloidogenic state.⁷ Structurally, the fibrillar-state model of Astbury is characterized by a cross- β -sheet fibril.⁸ Recent high-resolution NMR has shown extensive rearrangement of the native-state β -structures to produce a β strand conformation in the fibril core.⁹ Since not only β -sheet proteins but also those having both α -helix and β -sheet structures fibrillate, it is thought that the α -helices may unfold and then refold to β -strands in the course of fibrillation.¹⁰ More recent amyloid fibrillation studies endeavor to understand the details of the mechanism of fibrillation, often at residue level.¹¹⁻¹⁴ Another important subject involves studies on plant protein fibrillation. Amyloidosis of plant proteins should never be a surprise because the fibrous state is a generic state of proteins as speculated by Astbury. However, reports of plant protein amyloid fibrillation have been relatively scarce, because of either sparse study or a general lack of propensity for fibrillation or both. Plant protein fibrillation most often reported are those of the well-studied monellin.¹⁵⁻¹⁷ Amyloid fibrillation and amorphous aggregation of concanavalin A that belongs to the legume lectins family have been studied in fair detail.^{14,18} Very recently concanavalin A has been induced to fibrillate by the use of cationic Gemini surfactant.¹⁹ Amyloid-like inclusions of maize Trans glutamines that share structural features with Alzheimer's and Parkinson's-associated aggregates have been located in the chloroplasts of tobacco trans plastomic plants.²⁰ Even though in vivo fibril-like aggregate formation in plant cells might be different from the well-studied amyloid formation in mammalian cells, designed protein aggregation has

been used for selective protein knockdowns in plants.¹³ Recently, we solved the solution NMR structure of AtPP16-1,²¹ a 17,316 Da protein that presumably serves as a RNA transporter in the phloem conduit that runs into shoot or bud regions. The protein is selective about the working pH that we set to 4.1 for structure determination. The secondary structure consists of three β -strands, β 1, β 2 (56–62), and β 3 (133–135), and an α -helix, and are formed of residue stretches 3–9, 56–62, 133–135, and 96–110, respectively (Fig. 2.1A).

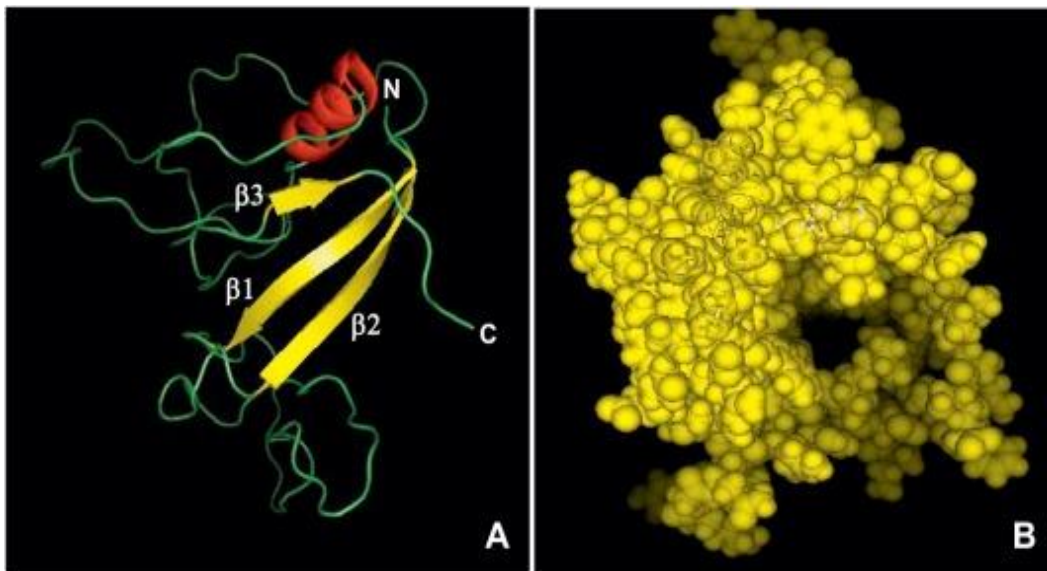


Figure 2.1. Solution structure of AtPP16-1 (PDB ID 5YQ3). (A) The secondary structure includes three antiparallel β -strands (β I, β II, and β III corresponding to residues 3–9, 56–62, and 133–135, respectively, and an α -helix (residues 96–105). The disordered tertiary fold is noticeable. The stretch of residues 10–55 that connects strands β I and β II form a small lobe. (B) The space-filled model showing surface overhangs, a sign of disordered structure. A deep cleft that is supported laterally by the β II strand runs through the molecule and is thought to be important for DNA and RNA binding.

The remainder of the chain segments, the N-terminal residues 10–55 in particular, are largely disordered with surface overhangs of side chains. As the space-filled structural model shows, the surface ledges arising out of some disordered stretches

produce irregular depressions with a conspicuous cleft (Fig. 2.1B). These structural features have already been incorporated in the description of AtPP16-1 as a prolate shaped intrinsically disordered protein (IDP) whose global correlation time is 8.4 ($\pm$ 0.7) ns.²¹

This identification makes the protein even more attractive for fibrillation studies, because IDPs are thought to undergo some refolding at the earliest stage of fibrillation. This work shows growth of unbranched or less branched amyloid fibrils under acid destabilized conditions where most of the secondary structure but little tertiary structure is retained.

2.3. MATERIALS AND METHODS

2.3.1. Protein preparation

Cloning and purification of Atpp16-1 has already been described.²¹ Briefly, the cDNA is initially cloned into pTZ57R/T vector, sub cloned into pET28a (+) vector, and transformed into BL21 (DE3) RIL cells for protein expression. Cells are grown in Luria Bertani broth containing kanamycin (50 μ g/mL) at 37 °C up to 0.6 optical density at 600 nm before allowing protein expression at 28 °C by the addition of IPTG (0.5 mM). After 5 h of growth cells are harvested and suspended in phosphate buffer saline. All purification steps are carried out at 4 °C. Cell lysis is achieved by a ‘20s on and 40s off’ sonication cycle for 5 min in 20 mM Tris-HCl, 50 mM NaCl, 5 mM imidazole, pH 8. The supernatant of the sonicated cell suspension is collected by a 20 min centrifugation at 23447g, passed through a Ni-NTA column equilibrated with lysis buffer, washed by flowing the wash buffer (20 mM Tris-HCl, pH 8, 50 mM NaCl, 40 mM Imidazole), and eluted with the elution buffer (20 mM Tris-HCl, 50 mM NaCl, 150 mM imidazole, pH 8). The eluted fractions are dialyzed in 50 mM sodium acetate at pH 4, the part of the protein precipitated during dialysis is removed by centrifugation, and the purity is checked by SDS-PAGE. For production of ¹⁵N-labeled protein, all steps are identical except that the cells are grown in M9 medium containing ¹⁵NH₄Cl.

2.3.2. pH titration and ANS binding

The buffer mixture consisted of 10 mM glycine, 4 mM each of sodium acetate, HEPES and PIPES, 8 mM Tris, and 2 mM CAPS. The protein concentration was $\sim 8 \mu\text{M}$. Fractions of the stock solution were set to different pH values in the range 1.6 – 10.75 using minimal volumes of HCl, acetic acid, and NaOH. Following equilibration at 25°C for 2 h 218-nm CD and fluorescence spectra (280-nm excitation) were measured. Each sample of 450 μL volume was then combined with 10 μL of a 1 mM ANS solution so as to obtain $\sim 23 \mu\text{M}$ in ANS. Fluorescence spectra were measured again by ANS excitation (380 nm). CD measurements were done in a AVIV SF420 spectrometer, and fluorescence spectra were taken using a Jasco FP-8300 instrument.

2.3.3. Kinetics of fibrillation

A 12 μM solution of Atp16-1 that contained $\sim 50 \mu\text{M}$ ThT (4-[3,6- dimethyl benzothiazol-2-yl]-N,N-dimethylaniline) was prepared in 50 mM Glycine-HCl buffer, pH 2. The solution was placed in a quartz cuvette and heated up to 60°C within ~ 4 min, and ThT fluorescence at 478 nm (excitation, 432 nm) was monitored as a function of time at the same temperature for 5 h.

2.3.4. NMR measurement

A 2D [1 H–15N] HSQC spectrum of Atp16-1 ($\sim 120 \mu\text{M}$) in D2O buffer of 20 mM sodium acetate, pH 4.1, 25 °C, was recorded as the control. The pH of the sample was then lowered to 2 by adding glycine, HCl, and spectra were recorded at 60 °C continuously for 6 h. The FID size was 2048 $\times$ 128 for F2 and F1, and the duration of each spectrum was ~ 19 min. A 500 MHz Bruker spectrometer was employed.

2.3.5. AFM, FESEM, and TEM imaging

Fibers were grown in 50 mM glycine-HCl, pH 2, 60°C up to different times in the 0–1800 min range. Samples for AFM were deposited on freshly cleaved mica sheets and dried under a stream of nitrogen for ~ 20 min. The film was then washed with deionized water and dried again under nitrogen. Semi contact-mode imaging was done using a 3 μM scanner in a NT-MDT Solver microscope. Images were processed and analyzed later

using the NOVA software. FESEM images of gold coated samples were taken in a ZEISS Ultra 55 instrument operating at 30 kV. For TEM imaging samples were deposited on copper grids, treated with 2 mM uranyl acetate for ~30 s, and washed and dried. Images were taken in a EDAX AMETEK instrument.

2.4. RESULTS

2.4.1. Purified AtPP16-1 and pH-dependent conformation

The N-terminal His6-tagged protein is easily purified to homogeneity (Fig. 2.2), although ~12 mg protein obtained per liter of culture growth is not considered a high yield. The reason for seemingly low yield is the precipitative loss of the protein that occurs while we change the solution condition of the purified protein from 20 mM Tris-HCl, pH 8 to 50 mM sodium acetate, pH 4.1 by dialysis. On the other hand, this change of the solution condition was necessary because the protein is more stable and monomeric at pH 4.1 as observed in our NMR structure.²¹ The precipitated protein however does not appear conformationally different from the one in solution, because the CD and fluorescence spectra of the former re-dissolved in the same buffer (50 mM sodium acetate, pH 4.1) are identical to the corresponding spectra for the latter (not shown). A detailed look at the pH-dependent conformation was however necessary in order to check for protein aggregation, and the propensity and rate of fibrillation so as to obtain a pH condition for growing the fibrils. The experiments involved fluorescence and CD-probed protein titration in the pH range 1.6–10.8.

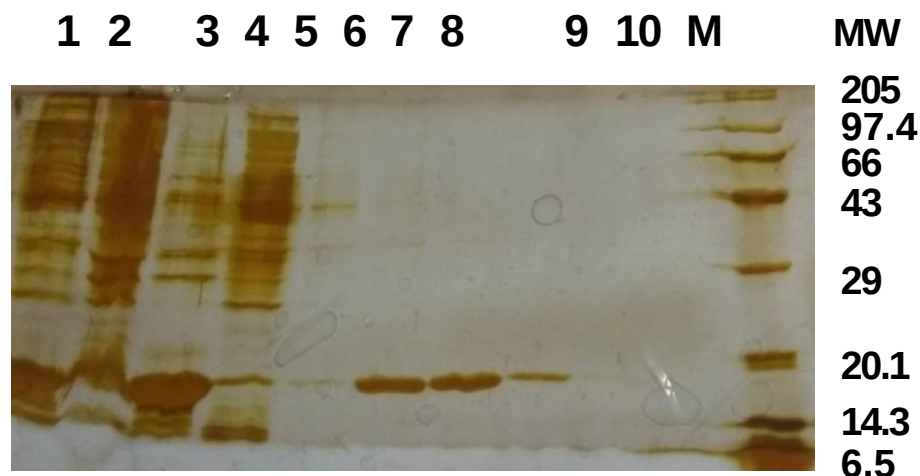


Figure 2.2 Silver-stained SDS-PAGE of samples in the purification of AtPP16-1. Lane labels are: 1, cell lysate before IPTG induction; 2, cell lysate after IPTG induced growth; 3, supernatant obtained by centrifuging the cell lysate; 4, pellet; 5, the flow-through fraction in Ni-NTA chromatography; 6–10, eluted fractions; M, marker proteins. The molecular mass of AtPP16-1 is 17.3 kDa.

The maximum of tryptophan fluorescence emission is Stokes-shifted from 331 to 347 nm as the pH increases across the scale (Fig. 2.3a, b). The shift up to pH 4 is related to acid-induced perturbation of the tertiary structure, and the larger red shift above pH 9 arises from alkali-induced large-scale conformational unfolding. The constant emission maximum of ~335 nm between in the neutral-pH region is likely due to structure loss by some mechanism the details of which we have not understood yet. It is also likely that AtPP16-1 dimerizes in this pH range, but there is little confirmatory evidence except that the pH-dependent emission intensity at the corresponding wavelength of maximum emission shows the presumed dimerization (Fig. 2.3c, open circles). Fig. 2.3c also plots the fluorescence intensity of the dye ANS that was added to the same set of pH-varying protein solutions that was used for the titration. The ANS binding shows that the protein perhaps dimerizes or even oligomerizes at pH values above 4, because ANS cannot bind and hence would not fluoresce if the hydrophobic surfaces of the individual monomers are concealed by self-association. The ANS fluorescence transition indicates that the self-

association is significantly stronger in the neutral pH region, and the alkali-denatured state does not bind ANS at all. The observation here that ANS binds to the protein at pH 4, the conditions used in our earlier NMR structural studies,²¹ is consistent with the fact that AtPP16-1 is an intrinsically disordered protein (IDP). The ANS fluorescence at pH < 4 suggests the presence of acid-induced exposure of hydrophobic patches on the monomeric protein surface that bind the dye. It is important to realize that the already disordered protein at pH 4 undergoes further structural disorder when the experimental pH is lowered. The fluorescence signal at pH < 4 in the presence and absence of ANS also needs to be closely scrutinized. In the absence of ANS the intrinsic aromatic fluorescence intensity of the protein reports on the loss of tertiary structure, and in the presence of the ANS dye the fluorescence intensity reports on the binding of the dye. They are thus complementary, because the dye would bind when the tertiary structure is perturbed that would result in exposure of hydrophobic surfaces leading to dye binding. Hence, the slight increase in the ANS binding below pH 4 means increase in hydrophobic patches at the surface of the protein already disordered at pH 4. In an attempt to understand the pH-dependent structural and conformational transitions further, we studied the pH titration by near- and far-UV CD. The near-UV CD spectrum of AtPP16-1 at pH 4.1 (Fig. 2.3d) shows markedly larger absorption in the 260–285 nm region peaking at ~273 nm, suggesting dominating contributions from side chains of Phe and Tyr that are 6 and 4 in number, respectively, but the contribution of the two Trp residues that absorb in the 280–290 nm region is minor.

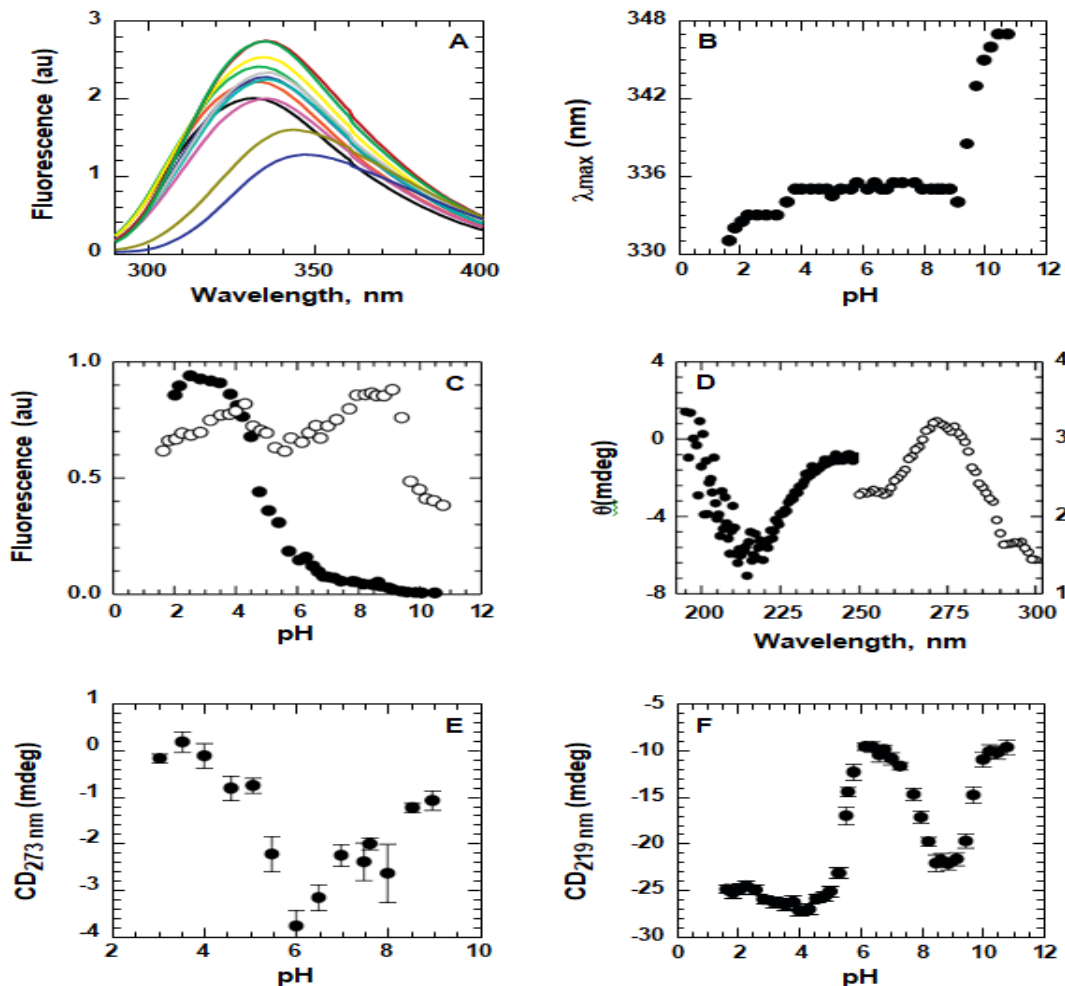


Figure 2.3. The pH-dependent conformations. (A) In tryptophan fluorescence spectra taken without dye addition the emission maximum is Stokes-shifted from 331 to 347 nm as the pH increases from 1.6 to 10.8. To avoid crowding, spectra are not labeled with respective pH values. (B) Starting from pH 1.6 the emission maximum red-shifts by 4 nm up to pH 4, remains constant in the pH 4–9 region, and then shifts by 12 nm up to pH 10.5 The small red shift in the acidic region is due to acid destabilization of tertiary structure, but large-scale conformational unfolding under alkaline conditions gives rise to relatively larger red shift. (C) The intensity at the corresponding emission maximum shows the acid and alkali effects on the tertiary structure (○). The changes between pH 4 and 8 are likely due to reversible dimerization or oligomerization that quenches the fluorescence. The binding of ANS monitored at the ANS emission maximum of 477 nm

(●) shows exposed hydrophobic surfaces below pH 5. Maximum binding occurs near pH 2.5. (D) The near and far-UV CD spectra of AtPP16-1 at pH 4.1, 25 °C. (E) The result of pH titration monitored by 273-nm CD is similar to that observed using tryptophan fluorescence shown in panel C. (F) The presence of substantial secondary structure (CD at 219 nm) in the acid pH region facilitates fibrillation. While the alkali transition arises from large-scale denaturation, the transitions between pH 5 and 8 are not fully understood.

The far-UV CD spectrum at pH 4.1, apparently noisy in Fig. 2.3d due to the acetate buffer, provides 219 nm as the peak wavelength. The pH titration monitored by 273-nm CD (Fig. 2.3e) shows loss of tertiary structure in the 4.2–8.5 region of pH, very similar to the result obtained with fluorescence (Fig. 2.3c). In the 218-nm CD probed titration (Fig. 2.3f), marginal loss of secondary structure below pH 4 and relatively large loss above pH 9 are related to acid destabilization and alkali unfolding, respectively. The origin of the two sharp transitions, one in the 4–6.2 range of pH and the other in the 6.8–9, is however difficult to understand. If the presumed dimerization of AtPP16-1 indeed occurs about pH 6, then the CD signal should have been more negative. Even if the dimerization assumption is not entertained, the CD signal of roughly -9 near pH 6 that implies a large loss of secondary structure compared to pH 4 or 9 is not easily perceived. This issue will be taken up with more experiments later, but on the basis of present results the fibrillation experiments at higher pH (> 4) was not contemplated. We rather chose to work with 50 mM Glycine-HCl, pH 2, for growing fibrils. The rationale for choosing pH 2 was to destabilize the tertiary and secondary structures to some extent (Figs. 2.3c, f). Under this condition the protein is monomeric with exposed hydrophobic surfaces that bind ANS, and has appreciable secondary and tertiary structures.

2.4.2. Kinetics of AtPP16-1 fibrillation

Fibrillation kinetics was followed by monitoring the fluorescence of the dye thioflavin T or ThT (4-(3,6-dimethylbenzothiazol-2-yl)-N, N-dimethylaniline), which is widely used to stain fibrils.^{22,23} A 12 μ M protein solution containing ~ 50 μ M ThT was set to pH 2 in

50 mM Glycine-HCl buffer, and incubated at 60 °C while monitoring time-base fluorescence of ThT fluorescence at 478 nm. After a lag time of ~30 min the fluorescence rises apparently in a single-exponential phase to reach the steady-state at ~300 min (Fig. 2.4a). This kinetics is specific to the pH and temperature used here, and is expected to vary with altered conditions. Fig. 2.4b shows the fibrillation kinetics up to an hour monitored by 218-nm CD, although the signal continues to be more negative at longer time (inset). The 218-nm CD monitored kinetics was found irreversible when the temperature is lowered to 25 °C after incubation for an hour, suggesting that fibrillation is accompanied by irreversible increase in the secondary structure content, because the.

The fibrillation kinetics was also checked with 2D [¹H-¹⁵N] HSQC spectra. Poor chemical shift dispersion for a large number of resonances in the spectrum of the native protein (~150 μM) at pH 4, 25 °C, is characteristic of intrinsic structural disorder (Fig. 2.5a). In the earliest spectrum at pH 2, 60 °C, the dispersion narrows down considerably, typical of a denatured protein (Fig. 2.5b, resonance contours plotted in black), and continues to collapse with time up to ~5 h (spectra plotted in red), which is better seen in the time stack plot produced by 1D processing of the first FID of each 2D HSQC spectrum (Fig. 2.5c). The result that elongated mature fibrils are formed in ~5 h appears consistent with the fluorescence-monitored fibrillation kinetics (Fig. 2.4a) even though the protein concentration used for NMR was ~12-fold more than that for the fluorescence experiment. Fibril growth is generally fast when higher protein concentration is used, and hence fluorescence should have reported with slower kinetics.

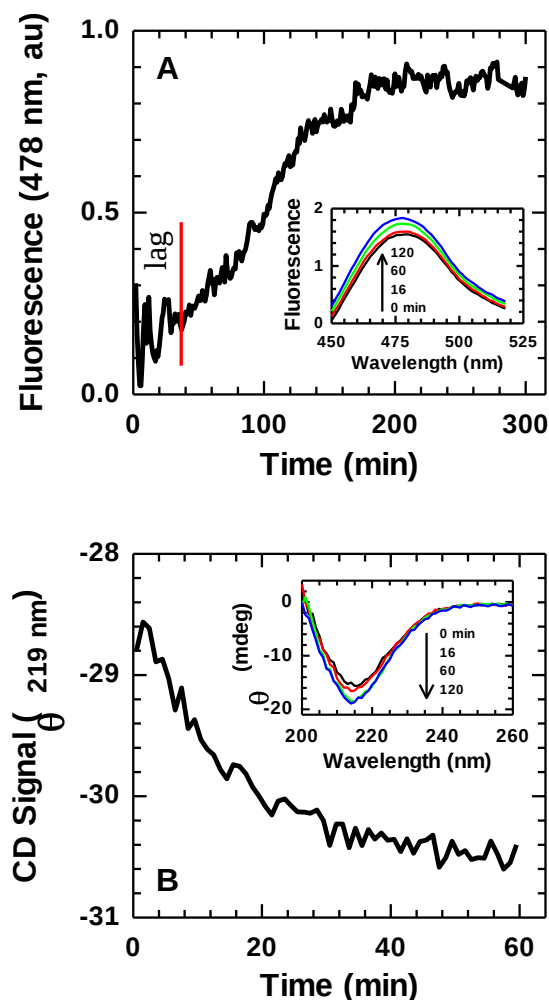


Figure 2.4. Fibrillation kinetics. (A) The reaction solution containing 12 μM Atpp16- 1, $\sim 50 \mu\text{M}$ ThT, and 50 mM Glycine-HCl, pH 2 was heated up to 60 $^{\circ}\text{C}$ within ~ 4 min, and time base ThT fluorescence at 478 nm (excitation, 432 nm) was measured thereafter at the same temperature. The initial lag phase, which is a signature of the process of amyloid fibrillation differentiated from amorphous aggregate, is indicated. The inset shows a few ThT fluorescence spectrum up to 2 h. (B) Time-base 219-nm CD signal of a 9 μM AtPP16-1 solution, 50 mM Tris-Glycine, pH 2, 60 $^{\circ}\text{C}$. The inset plots a few spectra at different times, 60 $^{\circ}\text{C}$.

2.4.3. Images and dimensions of AtPP16-1 fibrils

Figure 2.6 shows FESEM images of fibrillar structures raised up to ~ 2 h at pH 2, 60 $^{\circ}\text{C}$. They are simple and mostly linear; occasional sparse branching was seen in a fewer samples. Similar scanty branching is also observed in the TEM images taken after 1 and 2.5 h of heating at 60 $^{\circ}\text{C}$ (Figure 2.7). To resolve precursor aggregates and protofibrillar forms fibrillation was allowed with $\sim 12 \mu\text{M}$ protein, and were imaged by AFM at different stages of growth. Smaller precursor aggregates, their association to elongated forms and nascent protofibrillar structures are seen soon after the growth lag phase of ~ 30 min (Figure 2.8a), and moderately elongated protofibrils appear by 60 min (Figure 2.8b). The image taken 90 minutes after allowing the fibrillation reaction samples all of these

forms (Figure 2.8c). Prolonged growth appears to cause elongation and aggregation of different forms to produce an irregular mesh-like supra structure (Figure 2.8d).

For a dimensional distinction of the precursor aggregates and elongated fibrils we determined the distribution of the diameter of the respective form using a population size of 70 each, and fitted the diameter distribution as percent population (Figure 2.9) to the Gaussian formula

$$P(d) = P(d_0) \exp \left[-0.5 \left(\frac{d - d_0}{b} \right)^2 \right]$$

in which d is the measured diameter, $P(d_0)$ the amplitude of the mean diameter d_0 , and b the full width at half maximum (FWHM). The d_0 values for precursor aggregates and elongated fibrils are 13.25 and 21.15 nm; the higher value for the elongated fibrils presumably arises from larger content of β -sheet. The FWHM of 10.4 and 8.4 nm for precursor aggregates and elongated fibrils, respectively, suggest larger heterogeneity in the diameter of the former, and the heterogeneity implies content of different proportion of α -helix and β -sheet amongst the aggregates.

2.5. DISCUSSION

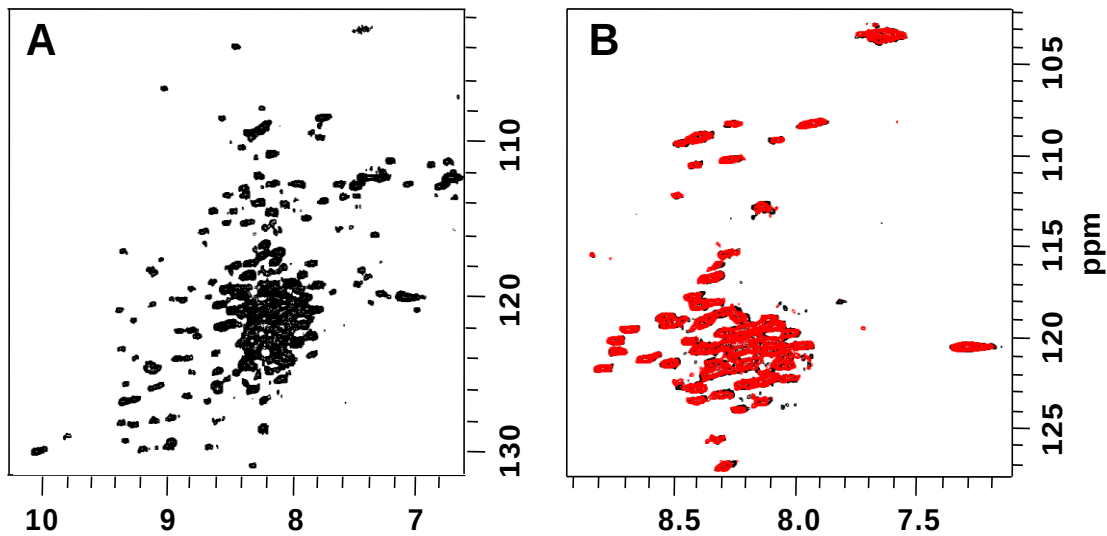
2.5.1. The protein AtPP16-1 forms amyloid fibrils, not amorphous aggregates

Protein chains can gather together to form glass-like amorphous or crystal-like amyloid aggregates. Although both are toxic to the cell, the growth mechanism and morphology of the two are distinct. The kinetics of amyloid fibrillation is characteristically preceded by a monomer concentration-dependent lag phase.^{24,25} This mode of fibrillation involves substantial changes of both secondary and tertiary structure,¹⁸ and the fibrils are distinctly viewed in microscopic images as rigid elongated structures. All of these requirements are resolved for AtPP16- 1 – the presence of a lag phase and changes in secondary structure followed by CD (Fig. 2.4B) and NMR resonances (Fig. 2.5C) leading to the appearance of long and rigid fibers (Figs. 2.6-2.8), indicating that the aggregation at pH 2 leads to amyloid growth. This distinction is important because concanavalin A, a lectin-family

plant protein, has been reported to undergo amorphous aggregation at $\text{pH} < 5$.¹⁸ More specifically, fibril-like aggregate formation in *Arabidopsis thaliana* and *Zeamays* plant tissues appear different from the typical amyloid formation in mammalian cells.^{13,20} It will be interesting to see if the deposition of *Arabidopsis* phloem proteins in vivo produces amorphous aggregates.

2.5.2. pH-dependence and the mechanism of precursor aggregate (PA) formation

Although the assembly and growth of amyloid structures is a complex process a thematic principle of initiation seems to arise from the results of numerous reports of pH-dependent propensity of fibrillation.^{26–34} Fibrillation is often achieved by destabilizing proteins under acidic or basic conditions that produce the generic A- and B-states, respectively. The acid-denatured state that serves as the precursor of the molten globule state is highly mobile and generally characterized by fluctuating secondary structure but little tertiary structure,³⁵ consistent with the observations made here for acidic AtPP16-1 (Fig. 2.4).



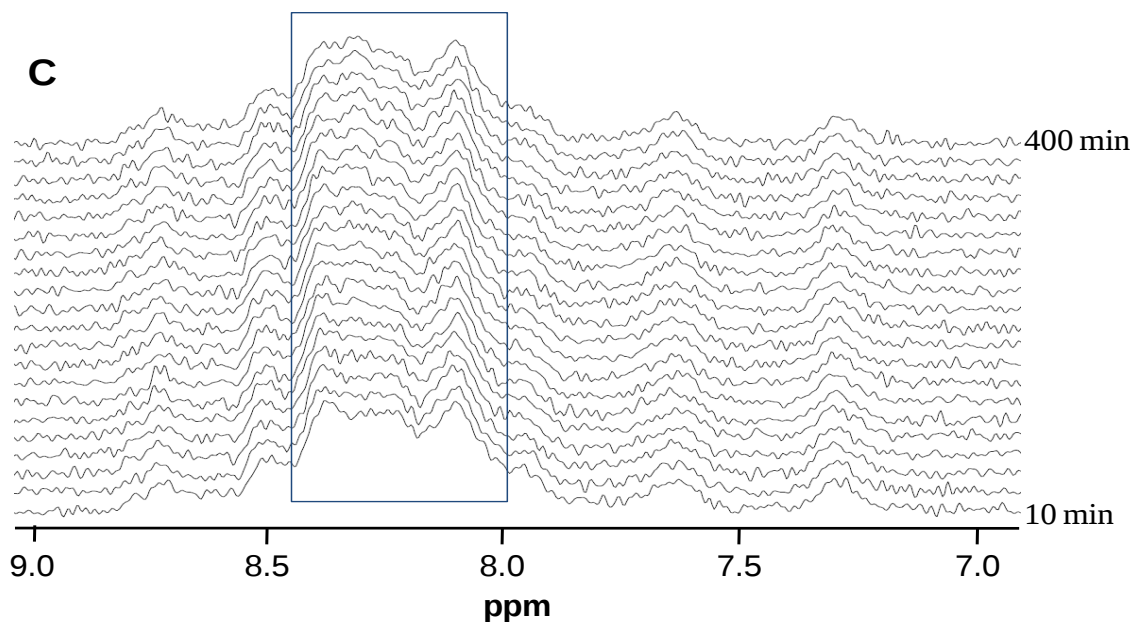


Figure 2.5. Fibrillation kinetics observed by 2D [^1H - ^{15}N] HSQC spectra. (A) The chemical shift dispersion in the spectrum of native AtPP16-1, pH 4, 25 °C, is still not impressive which is due to intrinsic disorder. (B) Spectra after 10 min (black resonances) and ~5 h (red resonances) of heating the protein up to 60°C, pH 2. The first spectrum already shows substantial collapse of dispersion. (C) Stack plot of the first 1D spectrum of each of the 2D spectra showing changes in chemical shifts as fibrillation progresses. The resonances boxed correspond to the most narrowed dispersion region in the 2D spectra.

These characteristics provide facile conditions for the initiation of fiber growth. First, the loss of tertiary structure exposes hydrophobic surfaces on the protein chain that can interface with the sequentially identical hydrophobic surfaces of another chain leading to the formation of a steric zipper.³⁶ Second, the conformational freedom of the protein facilitates collapse of the self-complementary hydrophobic surfaces from different chains that give rise to soluble oligomers called precursor aggregates (PA) or nuclei.²⁵ The formation of PA is not a favorable process, because this leads to not only a large decrease in segmental.

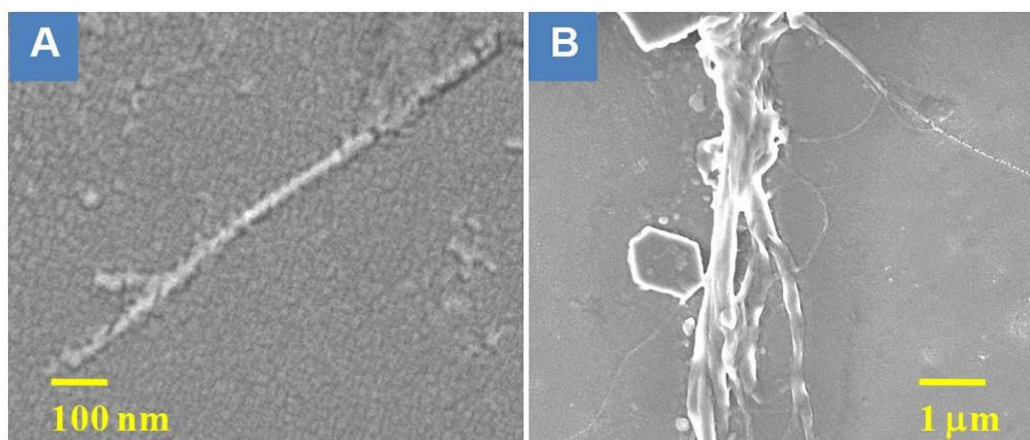


Figure 2.6. FESEM images of fibrils after 2 h (A) and 8 h (B) of incubation of AtPP161 at 60 °C, pH 2. The fibrils are linear and occasionally sparsely branched at shorter times of incubation.

Flexibility of the individual monomers but also sequestration of charges from the regions of chain contact and their burial in the interior. While reduced segmental flexibility is entropically unfavorable, charge burial in the low dielectric interior is energetically expensive. As such, the energy of soluble precursor aggregates must be higher than that of monomers from which they are formed. Even when fibrillation is initiated under conditions where the protein is not highly charged the factor of reduced entropy alone will raise the energy of the aggregate. A few studies under mild to highly alkaline conditions have also shown formation of soluble oligomers^{27,30} and elongated amyloid fibrils.^{14,18,32} Many alkali-destabilized pre-molten globule states of proteins including AtPP16-1 appear to be highly unstable so as to adopt near-random structure (see Fig. 2.3d). The lack of the required structural stability and perhaps the segmental flexibility might be the reason for their not entering the amyloidogenic state.

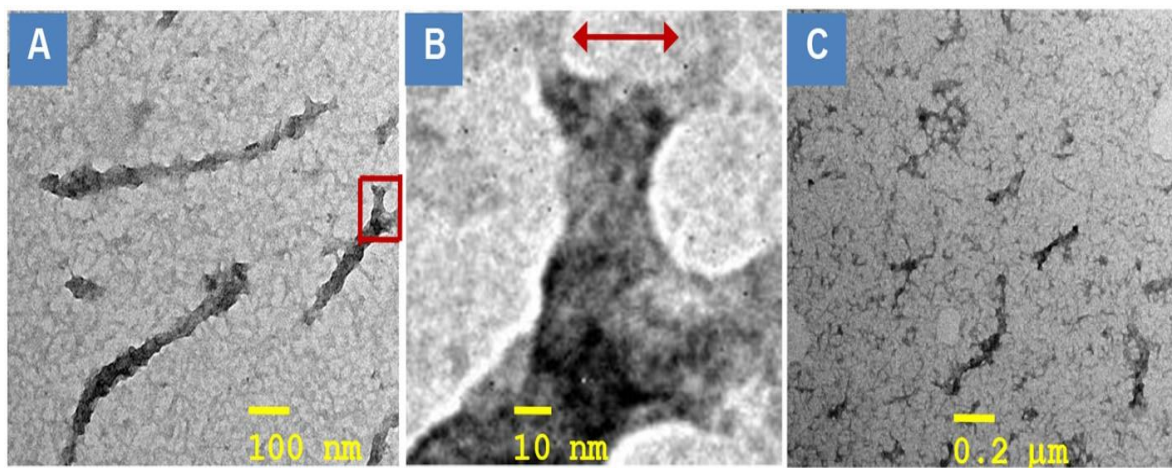


Figure 2.7. TEM images of fibrils. (A) Fibrillation after 1 h of initiating the reaction. (B) The region boxed in A is zoomed to show the onset of branching (arrowheads). (C) Fibrillation after 2.5 h of heating at 60 °C.

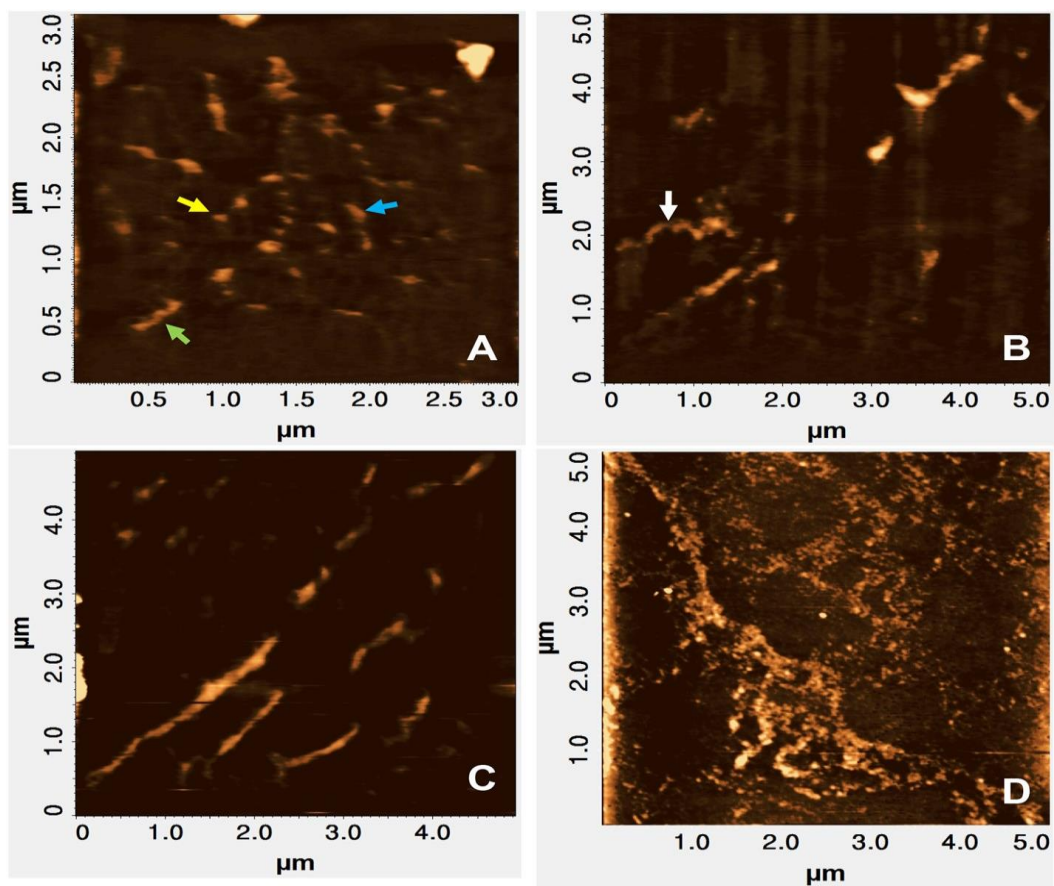


Figure 2.8. AFM images of AtPP16-1 fibrils. (A) Structures at 30 min after incubation at 60 °C. This time bin corresponds to the lag phase of fibrillation kinetics. The colored arrowheads point to precursor aggregates (yellow), elongated precursor aggregates (blue), and the initial stage of protofibril formation (green). (B) Elongation of protofibrils at 1 h (white arrowhead). (C) At 1.5 h protofibrils growth continues and larger ones appear. (D) After 6 h of growth irregular mesh-like supra structures of fibrils appear.

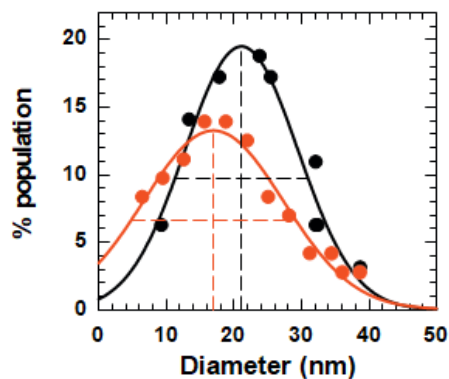
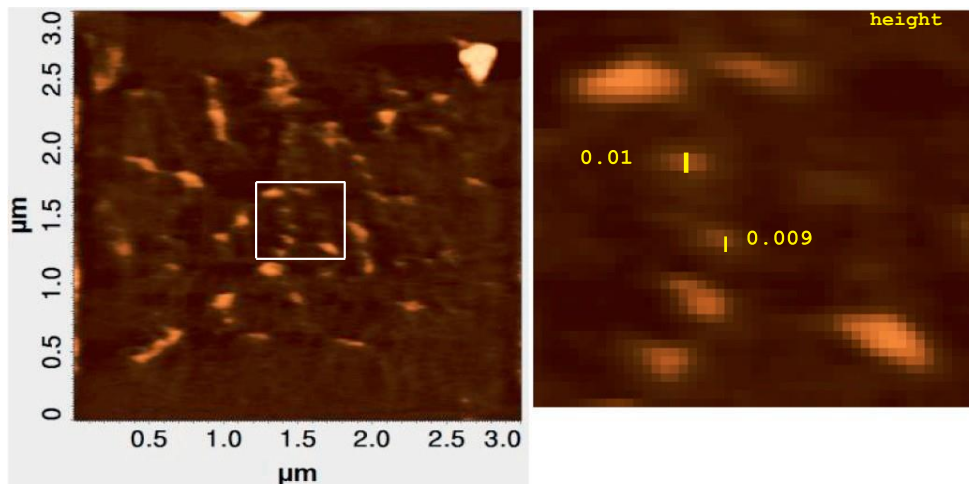


Figure 2.9. Gaussian distribution of population with diameters. Measurement of heights is exemplified using an expanded region of the image. Grouping the readings, and the determination of mean diameter of each group along with the corresponding percent population are described in the text. Solid lines in the percent population vs diameter graph are fits to data according to Eq. (1). Plotted in red are data for PA whose mean diameter is 13.25 nm, and data plotted in black pertain to elongated fibrils, the mean diameter for which is 21.15 nm. The FWHM of the distributions are 10.4 and 8.4 nm for precursor aggregates and elongated fibrils, respectively.

2.5.3. Condensation of precursor aggregates to protofibrils

Even though this work has not studied the PA $\rightarrow$ protofibril condensation the dimensional data scaled from the microscope images provide some indications regarding the elongation process. The mean diameter of PAs is 13.25 nm which increases to 21.15 nm for protofibrils (Fig. 2.8), suggesting that the growth of the latter occurs not only by condensation of the PAs to the end of the emerging protofibrils, but also by lateral association of the fibrils (Fig. 2.9). It must be the lateral collapse of two protofibrils that would account for their larger diameter. This type of condensation has been mentioned in the growth of amyloid- β (A β) protein fibrils, although the monomer and protofibril diameters recorded there are nearly one-third of that found here,^{37–39} which must be due to the smaller size A β (1–42).

2.5.4. Fibrillation of IDP

A conundrum arises from the conditions we used to induce fibrillation of AtPP16-1. It is thought that IDPs fibrillate by first refolding to some extent⁴⁰ whereas ordered proteins partly unfold first so as to enter the amyloidogenic state.^{4,41–45} Intrinsic disorder refers to a relative scale of structural disorder, and a large majority of proteins can be listed under the IDP class. The IDPs that are fairly structured like AtPP16-1 for which we used destabilizing acidic conditions may rather have to unfold to some extent in order to fibrillate. This is best seen from the chemical shift-collapsed HSQC NMR spectrum taken immediately after induction of fibrillation, where the loss of chemical shift dispersion indicates structural unfolding. Furthermore, the entire primary sequence is often not needed to initiate fibrillation, as demonstrated with fibrillation of a 21-residue peptide of the prion protein,⁴⁶ for example. If the involved stretch of amino acid sequence already fulfils the required level of structural content and segmental flexibility, fibrillation would appear favorable without undergoing further structural folding or unfolding. In essence, the assumed distinction of requirement of structural folding and unfolding between IDP and ordered proteins, respectively, should not be stressed too far. This work provided an opening report of amyloid-fibrillation of a plant phloem protein that is intrinsically disordered and presumably acts as a RNA transporter in the phloem conduit. There is a

large class of such RNA-binding proteins that are thought to take part in long-distance systemic transport of RNAs in plants,^{47,48} and it is quite likely that these RNA transporters have the propensity to fibrillate in test tube. But do they fibrillate or aggregate in vivo, and if so, it will be interesting to know the implications of such suprastructures for plant physiology.

2.6. REFERENCES

1. Terry, W. D., Page, D. L., Kimura, S., Isobe, T., Osserman, E. F., & Glenner, G. G. (1973). Structural identity of Bence Jones and amyloid fibril proteins in a patient with plasma cell dyscrasia and amyloidosis. *The Journal of clinical investigation*, 52(5), 1276–1281. <https://doi.org/10.1172/JCI107295>
2. Sunde, M., Serpell, L. C., Bartlam, M., Fraser, P. E., Pepys, M. B., & Blake, C. C. (1997). Common core structure of amyloid fibrils by synchrotron X-ray diffraction. *Journal of molecular biology*, 273(3), 729–739. <https://doi.org/10.1006/jmbi.1997.1348>
3. Sunde, M., & Blake, C. C. (1998). From the globular to the fibrous state: protein structure and structural conversion in amyloid formation. *Quarterly reviews of biophysics*, 31(1), 1–39. <https://doi.org/10.1017/s0033583598003400>
4. Fink A. L. (1998). Protein aggregation: folding aggregates, inclusion bodies and amyloid. *Folding & design*, 3(1), R9–R23. [https://doi.org/10.1016/S1359-0278\(98\)00002-9](https://doi.org/10.1016/S1359-0278(98)00002-9)
5. Pertinhez, T. A., Bouchard, M., Tomlinson, E. J., Wain, R., Ferguson, S. J., Dobson, C. M., & Smith, L. J. (2001). Amyloid fibril formation by a helical cytochrome. *FEBS letters*, 495(3), 184–186. [https://doi.org/10.1016/s0014-5793\(01\)02384-5](https://doi.org/10.1016/s0014-5793(01)02384-5)
6. Chiti, F., & Dobson, C. M. (2009). Amyloid formation by globular proteins under native conditions. *Nature chemical biology*, 5(1), 15–22. <https://doi.org/10.1038/nchembio.131>

7. Booth, D. R., Sunde, M., Bellotti, V., Robinson, C. V., Hutchinson, W. L., Fraser, P. E., Hawkins, P. N., Dobson, C. M., Radford, S. E., Blake, C. C., & Pepys, M. B. (1997). Instability, unfolding and aggregation of human lysozyme variants underlying amyloid fibrillogenesis. *Nature*, 385(6619), 787–793. <https://doi.org/10.1038/385787a0>
8. Astbury, W. T., Dickinson, S., & Bailey, K. (1935). The X-ray interpretation of denaturation and the structure of the seed globulins. *The Biochemical journal*, 29(10), 2351–2360.1. <https://doi.org/10.1042/bj0292351>
9. Su, Y., Sarell, C. J., Eddy, M. T., Debelouchina, G. T., Andreas, L. B., Pashley, C. L., Radford, S. E., & Griffin, R. G. (2014). Secondary structure in the core of amyloid fibrils formed from human β_2m and its truncated variant $\Delta N6$. *Journal of the American Chemical Society*, 136(17), 6313–6325. <https://doi.org/10.1021/ja4126092>
10. Walsh, D. M., Hartley, D. M., Kusumoto, Y., Fezoui, Y., Condron, M. M., Lomakin, A., Benedek, G. B., Selkoe, D. J., & Teplow, D. B. (1999). Amyloid beta-protein fibrillogenesis. Structure and biological activity of protofibrillar intermediates. *The Journal of biological chemistry*, 274(36), 25945–25952. <https://doi.org/10.1074/jbc.274.36.25945>
11. Varela, L., Morel, B., Azuaga, A. I., & Conejero-Lara, F. (2009). A single mutation in an SH3 domain increases amyloid aggregation by accelerating nucleation, but not by destabilizing thermodynamically the native state. *FEBS letters*, 583(4), 801–806. <https://doi.org/10.1016/j.febslet.2009.01.033>
12. Baskakov I. V. (2009). Switching in amyloid structure within individual fibrils: implication for strain adaptation, species barrier and strain classification. *FEBS letters*, 583(16), 2618–2622. <https://doi.org/10.1016/j.febslet.2009.05.044>
13. Betti, C., Vanhoutte, I., Coutuer, S., De Rycke, R., Mishev, K., Vuylsteke, M., Aesaert, S., Rombaut, D., Gallardo, R., De Smet, F., Xu, J., Van Lijsebettens, M., Van Breusegem, F., Inzé, D., Rousseau, F., Schymkowitz, J., & Russinova, E. (2016). Sequence-Specific Protein Aggregation Generates Defined Protein Knockdowns in Plants. *Plant physiology*, 171(2), 773–787. <https://doi.org/10.1104/pp.16.00335>

14. Carrotta, R., Vetri, V., Librizzi, F., Martorana, V., Militello, V., & Leone, M. (2011). Amyloid fibrils formation of concanavalin A at basic pH. *The journal of physical chemistry. B*, 115(12), 2691–2698. <https://doi.org/10.1021/jp1042409>
15. Konno, T., Murata, K., & Nagayama, K. (1999). Amyloid-like aggregates of a plant protein: a case of a sweet-tasting protein, monellin. *FEBS letters*, 454(1-2), 122–126. [https://doi.org/10.1016/s0014-5793\(99\)00789-9](https://doi.org/10.1016/s0014-5793(99)00789-9)
16. Konno T. (2001). Multistep nucleus formation and a separate subunit contribution of the amyloidogenesis of heat-denatured monellin. *Protein science : a publication of the Protein Society*, 10(10), 2093–2101. <https://doi.org/10.1110/ps.20201>
17. Assarsson, A., Hellstrand, E., Cabaleiro-Lago, C., & Linse, S. (2014). Charge dependent retardation of amyloid β aggregation by hydrophilic proteins. *ACS chemical neuroscience*, 5(4), 266–274. <https://doi.org/10.1021/cn400124r>
18. Vetri, V., Canale, C., Relini, A., Librizzi, F., Militello, V., Gliozzi, A., & Leone, M. (2007). Amyloid fibrils formation and amorphous aggregation in concanavalin A. *Biophysical chemistry*, 125(1), 184–190. <https://doi.org/10.1016/j.bpc.2006.07.012>
19. Ahmad, E., Kamranur Rahman, S., Masood Khan, J., Varshney, A., & Hasan Khan, R. (2009). Phytolacca americana lectin (Pa-2; pokeweed mitogen): an intrinsically unordered protein and its conversion into partial order at low pH. *Bioscience reports*, 30(2), 125–134. <https://doi.org/10.1042/BSR20090035>
20. Villar-Piqué, A., Sabaté, R., Lopera, O., Gibert, J., Torne, J. M., Santos, M., & Ventura, S. (2010). Amyloid-like protein inclusions in tobacco transgenic plants. *PloS one*, 5(10), e13625. <https://doi.org/10.1371/journal.pone.0013625>
21. Sashi, P., Singarapu, K. K., & Bhuyan, A. K. (2018). Solution NMR Structure and Backbone Dynamics of Partially Disordered Arabidopsis thaliana Phloem Protein 16-1, a Putative mRNA Transporter. *Biochemistry*, 57(6), 912–924. <https://doi.org/10.1021/acs.biochem.7b01071>
22. Harry LeVine (1995) Thioflavine T interaction with amyloid β -sheet structures, *Amyloid*, 2:1, 1-6, DOI: [10.3109/13506129509031881](https://doi.org/10.3109/13506129509031881)

23. Naiki, H., Higuchi, K., Hosokawa, M., & Takeda, T. (1989). Fluorometric determination of amyloid fibrils in vitro using the fluorescent dye, thioflavin T1. *Analytical biochemistry*, 177(2), 244–249. [https://doi.org/10.1016/0003-2697\(89\)90046-8](https://doi.org/10.1016/0003-2697(89)90046-8)
24. Yoshimura, Y., Lin, Y., Yagi, H., Lee, Y. H., Kitayama, H., Sakurai, K., So, M., Ogi, H., Naiki, H., & Goto, Y. (2012). Distinguishing crystal-like amyloid fibrils and glass-like amorphous aggregates from their kinetics of formation. *Proceedings of the National Academy of Sciences of the United States of America*, 109(36), 14446–14451. <https://doi.org/10.1073/pnas.1208228109>
25. Arosio, P., Knowles, T. P., & Linse, S. (2015). On the lag phase in amyloid fibril formation. *Physical chemistry chemical physics : PCCP*, 17(12), 7606–7618. <https://doi.org/10.1039/c4cp05563b>
26. Fraser, P. E., Nguyen, J. T., Surewicz, W. K., & Kirschner, D. A. (1991). pH-dependent structural transitions of Alzheimer amyloid peptides. *Biophysical journal*, 60(5), 1190–1201. [https://doi.org/10.1016/S0006-3495\(91\)82154-3](https://doi.org/10.1016/S0006-3495(91)82154-3)
27. Srinivasan, R., Jones, E. M., Liu, K., Ghiso, J., Marchant, R. E., & Zagorski, M. G. (2003). pH-dependent amyloid and protofibril formation by the ABri peptide of familial British dementia. *Journal of molecular biology*, 333(5), 1003–1023. <https://doi.org/10.1016/j.jmb.2003.09.001>
28. Guo, M., Gorman, P. M., Rico, M., Chakrabarty, A., & Laurents, D. V. (2005). Charge substitution shows that repulsive electrostatic interactions impede the oligomerization of Alzheimer amyloid peptides. *FEBS letters*, 579(17), 3574–3578. <https://doi.org/10.1016/j.febslet.2005.05.036>
29. Rao, P. N., Reddy, K. S., & Bhuyan, A. K. (2009). Amyloid fibrillation of human Apaf-1 CARD. *Biochemistry*, 48(32), 7656–7664. <https://doi.org/10.1021/bi900626u>
30. Ahmad, B., Winkelmann, J., Tiribilli, B., & Chiti, F. (2010). Searching for conditions to form stable protein oligomers with amyloid-like characteristics: The unexplored basic pH. *Biochimica et biophysica acta*, 1804(1), 223–234. <https://doi.org/10.1016/j.bbapap.2009.10.005>

31. Shammass, S. L., Knowles, T. P., Baldwin, A. J., Macphee, C. E., Welland, M. E., Dobson, C. M., & Devlin, G. L. (2011). Perturbation of the stability of amyloid fibrils through alteration of electrostatic interactions. *Biophysical journal*, 100(11), 2783–2791. <https://doi.org/10.1016/j.bpj.2011.04.039>
32. Ramakrishna, D., Prasad, M. D., & Bhuyan, A. K. (2012). Hydrophobic collapse overrides Coulombic repulsion in ferricytochrome c fibrillation under extremely alkaline condition. *Archives of biochemistry and biophysics*, 528(1), 67–71. <https://doi.org/10.1016/j.abb.2012.08.013>
33. Garrec, J., Tavernelli, I., & Rothlisberger, U. (2013). Two misfolding routes for the prion protein around pH 4.5. *PLoS computational biology*, 9(5), e1003057. <https://doi.org/10.1371/journal.pcbi.1003057>
34. Diaz-Espinoza, R., Nova, E., & Monasterio, O. (2017). Overcoming electrostatic repulsions during amyloid assembly: Effect of pH and interaction with divalent metals using model peptides. *Archives of biochemistry and biophysics*, 621, 46–53. <https://doi.org/10.1016/j.abb.2017.03.003>
35. Ptitsyn O. B. (1995). Molten globule and protein folding. *Advances in protein chemistry*, 47, 83–229. [https://doi.org/10.1016/s0065-3233\(08\)60546-x](https://doi.org/10.1016/s0065-3233(08)60546-x)
36. Goldschmidt, L., Teng, P. K., Riek, R., & Eisenberg, D. (2010). Identifying the amyloids, proteins capable of forming amyloid-like fibrils. *Proceedings of the National Academy of Sciences of the United States of America*, 107(8), 3487–3492. <https://doi.org/10.1073/pnas.0915166107>
37. Harper, J. D., Wong, S. S., Lieber, C. M., & Lansbury, P. T. (1997). Observation of metastable Abeta amyloid protofibrils by atomic force microscopy. *Chemistry & biology*, 4(2), 119–125. [https://doi.org/10.1016/s1074-5521\(97\)90255-6](https://doi.org/10.1016/s1074-5521(97)90255-6)
38. Nichols, M. R., Moss, M. A., Reed, D. K., Lin, W. L., Mukhopadhyay, R., Hoh, J. H., & Rosenberry, T. L. (2002). Growth of beta-amyloid(1-40) protofibrils by monomer elongation and lateral association. Characterization of distinct products by light scattering and atomic force microscopy. *Biochemistry*, 41(19), 6115–6127. <https://doi.org/10.1021/bi015985r>

39. Paranjape, G. S., Gouwens, L. K., Osborn, D. C., & Nichols, M. R. (2012). Isolated amyloid- β (1-42) protofibrils, but not isolated fibrils, are robust stimulators of microglia. *ACS chemical neuroscience*, 3(4), 302–311. <https://doi.org/10.1021/cn2001238>
40. Uversky V. N. (2010). Targeting intrinsically disordered proteins in neurodegenerative and protein dysfunction diseases: another illustration of the D(2) concept. *Expert review of proteomics*, 7(4), 543–564. <https://doi.org/10.1586/epr.10.36>
41. Kelly J. W. (1998). The alternative conformations of amyloidogenic proteins and their multi-step assembly pathways. *Current opinion in structural biology*, 8(1), 101–106. [https://doi.org/10.1016/s0959-440x\(98\)80016-x](https://doi.org/10.1016/s0959-440x(98)80016-x)
42. Lansbury P. T., Jr (1999). Evolution of amyloid: what normal protein folding may tell us about fibrillogenesis and disease. *Proceedings of the National Academy of Sciences of the United States of America*, 96(7), 3342–3344. <https://doi.org/10.1073/pnas.96.7.3342>
43. Bellotti, V., Mangione, P., & Stoppini, M. (1999). Biological activity and pathological implications of misfolded proteins. *Cellular and molecular life sciences : CMLS*, 55(6-7), 977–991. <https://doi.org/10.1007/s000180050348>
44. Rochet, J. C., & Lansbury, P. T., Jr (2000). Amyloid fibrillogenesis: themes and variations. *Current opinion in structural biology*, 10(1), 60–68. [https://doi.org/10.1016/s0959-440x\(99\)00049-4](https://doi.org/10.1016/s0959-440x(99)00049-4)
45. Zerovnik E. (2002). Amyloid-fibril formation. Proposed mechanisms and relevance to conformational disease. *European journal of biochemistry*, 269(14), 3362–3371. <https://doi.org/10.1046/j.1432-1033.2002.03024.x>
46. Ku, S. H., & Park, C. B. (2008). Highly accelerated self-assembly and fibrillation of prion peptides on solid surfaces. *Langmuir : the ACS journal of surfaces and colloids*, 24(24), 13822–13827. <https://doi.org/10.1021/la802931k>
47. Lucas, W. J., Yoo, B. C., & Kragler, F. (2001). RNA as a long-distance information macromolecule in plants. *Nature reviews. Molecular cell biology*, 2(11), 849–857. <https://doi.org/10.1038/35099096>

48. Thompson, G. A., & Schulz, A. (1999). Macromolecular trafficking in the phloem. *Trends in plant science*, 4(9), 354–360. [https://doi.org/10.1016/s1360-1385\(99\)01463-6](https://doi.org/10.1016/s1360-1385(99)01463-6).

CHAPTER 3

Negative Thermal Expansion of AtPP16-1

3.1. ABSTRACT

Despite immense interest in thermal expansion of solid materials, negative thermal expansion (NTE) in particular, expansivity of proteins near ambient temperature is minimally reported. One reason for the paucity of protein expansion studies could be the susceptibility of proteins to conformational changes often within sudenaturing temperatures, given that the expansivity phenomenon should include only vibrational fluctuations in atom positions in the absence of a conformational change. This study encounters a disordered plant protein called AtPP16-1 (*Arabidopsis thaliana* phloem protein type 16-1) which shows negative thermal expansion in the 292–309 K range of temperature at no expense of conformational transition. Temperature dependence of NMR chemical shifts of backbone atoms, global tumbling time derived from ^{15}N relaxations, and volume compressibility measurements yield a negative linear expansion coefficient $\langle\alpha\rangle \sim -5.3 \times 10^{-3} \text{ K}^{-1}$ averaged over a sizable set of hydrogen bonds, and a decrease in the root mean square volume fluctuation by $\sim 40\%$. A frequency spectrum of normal modes in which one or a fewer low frequency vibrational modes of large fluctuation amplitudes are excited at the lower end, and many high-frequency vibrational modes of lower fluctuation amplitudes are excited at the higher end of the temperature range can explain negative thermal expansion.

Keywords: negative thermal expansion of protein; thermal expansion coefficient; temperature coefficient of NMR chemical shift; negative volume expansion coefficient; frequency spectrum of vibrational modes

3.2. INTRODUCTION

Greater motion of atoms with higher thermal energy should increase the interatomic equilibrium distances in matter. It is not always so. In one of the earliest studies of materials, Hummel observed that thermal expansions of binary lithium aluminates are very large, one form of the ternary lithium-aluminium-silica compound is very low, and another form of the ternary compound is negative.¹ The finding of negative thermal expansion (NTE) of more materials in later years²⁻¹² led to envisage engineering devices built with material composites of desired thermal expansion. Some success in device building has been achieved even as the area of NTE of materials continues to grow rapidly.

Thermal expansion of proteins on the other hand has been studied less. The first report of Frauenfelder and colleagues¹³ that both the crystal lattice and structure of myoglobin in the 80–300 K range expand with the expansion coefficient $\alpha \sim 115 \times 10^{-6} \text{ K}^{-1}$ was followed up by studies with a few other proteins, including lysozyme,¹⁴⁻¹⁷ ovalbumin,¹⁴ RNase A,¹⁸ and bovine pancreatic trypsin inhibitor.¹⁹ It is also suggested that intramolecular dissipation of thermal energy, from the photo-excited heme to the globin matrix of myoglobin for example, can transiently expand the protein matrix.²⁰ It has been generally recognized that proteins having larger content of secondary structures, α -helix in particular, exhibit greater thermal expansion. The roles of surface hydration, molar volumes of water, and internal void spaces have also been discussed.¹⁹ A better understanding of thermal expansion of biopolymers is thought to find application in industry²¹ and pharmaceuticals.

Thermal expansion coefficient varies from one protein to another because of protein size and variation in the content of physical determinants and structural elements. Anomaly in the temperature dependence of thermal expansion may also be detected.¹⁴ No investigation has however described NTE protein in contrast with numerous reports of NTE materials and composites. One may encounter seeming instances of NTE proteins example, the amyloid peptide $A\beta_{42}$ in aqueous medium has been reported to exhibit an apparent thermal expansion coefficient $\alpha^{\text{app}} \sim -0.8 \times 10^{-3} \text{ K}^{-1}$.²² Although credited for the rigor of the work, $A\beta_{42}$ undergoes substantial structural and conformational changes in the temperature range described. As we discuss later in this paper, a description of thermal expansion or contraction must rely on atom motions, normal modes precisely, that do not lead to structural and

conformational changes.

This work finds negative linear and volume expansion of AtPP16-1 in a narrow temperature range in which no conformational change of the protein is detected. The protein AtPP16-1 is involved in a variety of functions, including calcium binding and nutrient transport. It is also likely to play a role in mRNA delivery, as suggested by studies with phloem protein homologs.²³ Recently, we determined the solution NMR structure of the protein and found it to be intrinsically disordered.²⁴ Briefly, only 32 of a total of 156 residues are found in secondary structural regions formed of residues 3–9 (β I), 56–62 (β II), 133–135 (β III), and 96–110 (α -helix); most of the chain otherwise is folded disorderly (Figure 1). The disordered N-terminal lobe (residues 10–55) adjoins β I to the rest of the molecule by a large irregular cleft which appears important for interactions with nucleic acids by hydrogen-bonding. The structure was solved at pH 4, because the protein unfolds in the region of physiological pH. We have now carried out chemical shift analyses of the backbone atoms $^1\text{H}^{\text{N}}$, $^1\text{H}^{\text{C}\alpha}$, and C_{α} , NMR relaxation, and adiabatic compressibility measurements in the 292–309($\pm$ 1) K window in which the protein conformation does not transform. We determine thermal expansion coefficient of the H-bonds and find negative linear expansion for all of them. It emerges that thermal fluctuations lead to contraction of the native state.

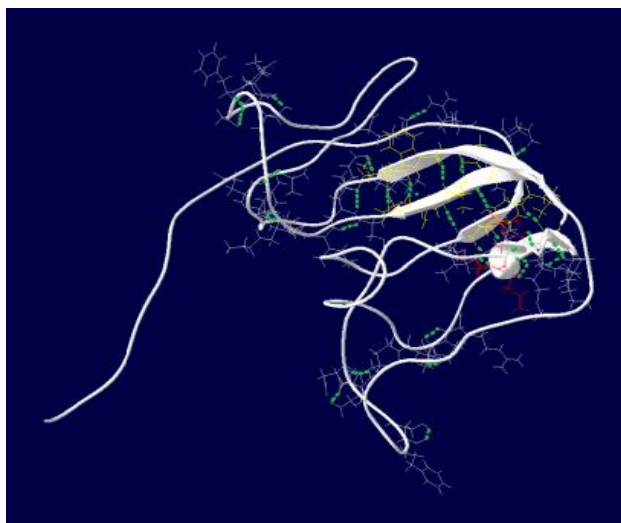


Figure 3.1. A cartoon of AtPP16-1 structure showing the intra- and intermolecular H-bonds.

3.3. METHODS

Cloning, expression, and purification of ^{15}N -labeled recombinant AtPP16-1 have been reported earlier.^{24,25} All experiments were done at pH 4.1 buffered with 7 mM acetate. Thermal denaturation of the protein was monitored by both far- and near-UV CD using a thermoelectric Peltier in an AVIV SF420 spectrometer. Equilibration time at each temperature was 2 minutes. Reversibility of thermal denaturation was checked by a repeat in which the sample heated from 293 to 363 K was cooled to the initial temperature of 293 K. Tryptophan fluorescence anisotropy was monitored by 280-nm excitation and 360-nm emission using the standard four settings of the polarizers $\text{HH}_{0,0}$, $\text{VH}_{90,0}$, $\text{HV}_{0,90}$, and $\text{VV}_{90,90}$. The anisotropy r as a function of time was calculated without applying the grating correction factor

$$G = I_{\text{HV}}/I_{\text{HH}},$$

$$r = \frac{I_{\text{VV}} - GI_{\text{VH}}}{I_{\text{VV}} + 2GI_{\text{VH}}}.$$

Two-dimensional [^1H - ^{15}N] HSQC spectra were recorded with a 9% D_2O sample of $\sim 150\ \mu\text{M}$ protein in the 296–308 K range at an interval of 2 K. Chemical shifts were referenced with respect to DSS (4,4-dimethyl-4-silapentane-1-sulfonate) whose chemical shift has minimal respectively, and the FID size for both F1 and F2 was 2048×128 . The ^{15}N spin-lattice and spin-spin relaxation times, T_1 and T_2 , were measured by using the inversion recovery and CPMG sequences executed with [^1H - ^{15}N] HSQC.²⁶ Spectra were recorded in a 500 MHz BRUKER spectrometer, and processed using Topspin 3.5 and Sparky software.

3.4. RESULTS AND DISCUSSION

3.4.1. Temperature window to study thermal shift of resonances.

The characterization of thermal expansion of materials should necessarily rely on fluctuations of the atom positions rather than thermally altered structure of molecules. This vital restriction on any drift in molecular structure mandates a careful evaluation of structure and conformation in response to a change in temperature. For a protein system in particular, the use of a set of temperature coefficients to characterize thermal expansion will be

legitimate only in the absence of any local or global conformational transitions. One therefore needs to choose a range of temperature well below that corresponding to a conformational change or unfolding. The most appropriate approach to ascertain the lack of conformational drift could have been the determination of crystal structure of the protein in a range of temperature to be chosen for thermal expansion. Generating X-ray maps in small temperature intervals however are fraught with practical limitations, especially in the case of a disordered protein that may not even crystallize.

With due consideration to these restrictions we examined the temperature dependence of conformational transitions of AtPP16-1 using basic spectroscopic measurements. Figure 3.2a, b show CD-monitored thermal transitions, which are fortunately reversible and reasonably sharp for an IDP. The slope of pre- and post-transition baselines in the far-UV CD melt (Figure 3.2a) is typical of thermal transitions of many proteins, and is probably not related to changes in secondary structure. The temperature dependence was then probed at 273 nm which reports on the asymmetric environment of Phe and Tyr side chains. The baselines of the melt are fairly flat here (Figure 3.2b), suggesting no detectable local conformational transition. Further, the temperature dependence of tryptophan fluorescence anisotropy (Figure 3.2c) does not indicate any considerable change in the fluorophore mobility that would have changed the conformation. These results suggest the availability of a window in the 292–308 K range of temperature to study coefficients of NMR chemical shifts and compressibility.

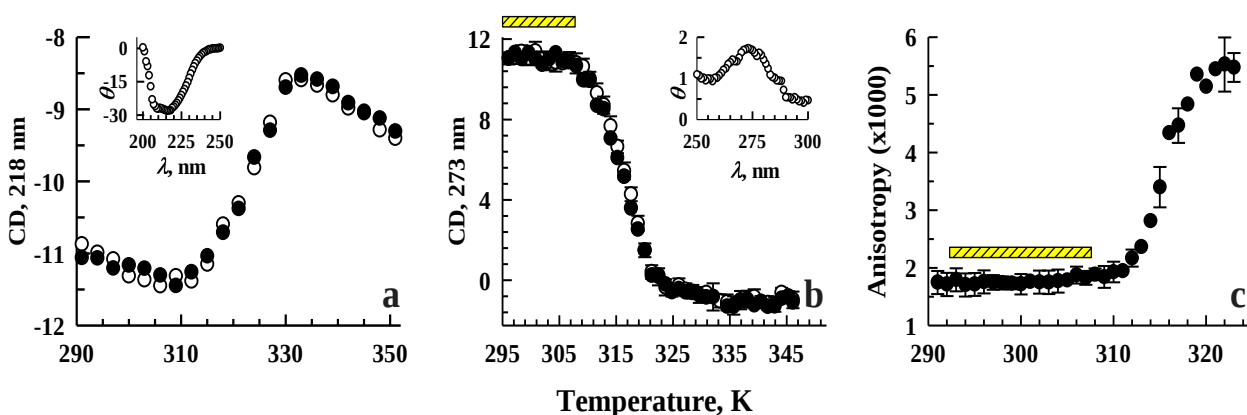


Figure 3.2. Thermal melts of AtPP16-1 at pH 4.1 monitored by CD at 218 nm (a), CD at 273 nm (b), and tryptophan fluorescence anisotropy (c). The CD spectrum in the far- and near-UV regions is shown in the *inset* of the first two panels. The hatched rectangles in yellow fill is spread over the temperature range of 292–309(±1) K employed for NMR and compressibility measurements.

3.4.2. Positive temperature coefficient for the chemical shift of backbone atoms.

Temperature coefficient ($\Delta\delta/\Delta T$) of protein backbone atoms was measured by NMR chemical shifts (δ) of $^1\text{H}^{\text{N}}$, $^1\text{H}^{\text{C}\alpha}$, and C_α atoms in the 293–310 K range of temperature referenced directly with respect to the chemical shift of DSS. Since AtPP16-1 is intrinsically disordered, the chemical shift dispersion is rather narrow; however, working with the resonances away from the crowded regions we could assign 33 residues in the [^1H – ^{15}N] HSQC and 47 residues in the [^1H – ^{13}C] HSQC spectra. As shown for a few analyzed peaks (Figure 3.3), the δ of $^1\text{H}^{\text{N}}$, $^1\text{H}^{\text{C}\alpha}$, and C_α atoms of all residues shift linearly downfield from DSS, indicating temperature-induced high-frequency shift or deshielding, and hence positive temperature coefficient ($\Delta\delta/\Delta T > 0$). The ranges of temperature coefficients observed are +4 to +10.5 ppb K^{-1} ($^1\text{H}^{\text{N}}$), +8.4 to +13.3 ppb K^{-1} ($^1\text{H}^{\text{C}\alpha}$), and +0.1 to +26.1 ppb K^{-1} (C_α).

This is somewhat surprising, because positive values of $\Delta\delta/\Delta T$ hints at linear thermal contraction, which is atypical of the commonly observed low-frequency shift or shielding of $^1\text{H}^{\text{N}}$ resonances with increasing temperature. A survey of data for $^1\text{H}^{\text{N}}$ in ~200 records published during the past 25 years reveal that the temperature coefficients of folded and unfolded states of proteins lie in the range from ~0 to ~ –12 ppb K^{-1} unlike the positive coefficients found from these results. Isolated instances of positive temperature coefficient have also been reported for a very few $^1\text{H}^{\text{N}}$ nuclei in ubiquitin,²⁷ GB3 protein,²⁸ and cardiac troponin C.²⁹ But the present finding of $\Delta\delta/\Delta T > 0$ for a large number of $^1\text{H}^{\text{N}}$, $^1\text{H}^{\text{C}\alpha}$ and C_α suggests the possibility that the protein undergoes thermal contraction.

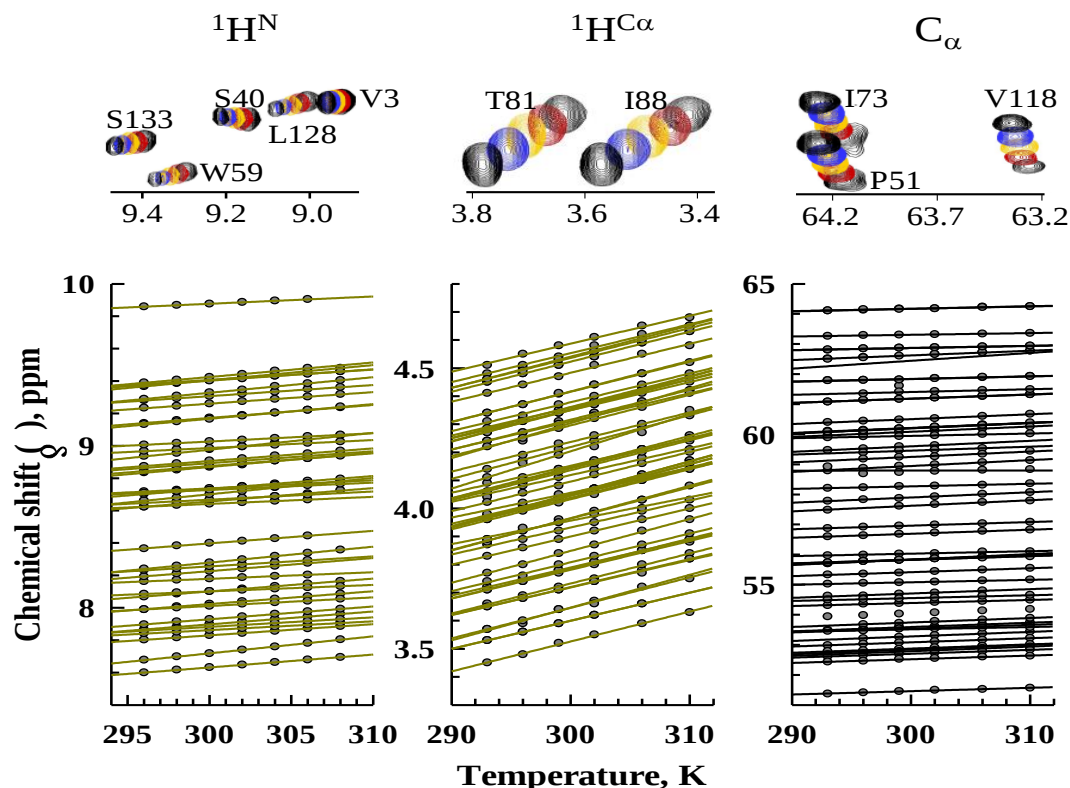


Figure 3.3. Temperature dependence of chemical shifts of $^1\text{H}^{\text{N}}$ (left), $^1\text{H}^{\text{C}\alpha}$ (middle), and C_α (right) resonances of AtPP16-1 in the 293 to 310 K range. The pattern of chemical shifts in HSQC spectra is illustrated by a few assigned peaks on top of each panel.

3.4.3. Contracted H...O distances in the $>\text{N}-\text{H}\cdots\text{O}=\text{C}<$ H-bonds.

Chemical shifts of backbone $^1\text{H}^{\text{N}}$ in proteins are particularly difficult to analyze by reference to the structural model, which is why their quantitative use in structural analysis remains extremely limited. The reason for this is their exquisite sensitivity to a multitude of factors including hydrogen bonding, local solvent composition at the protein surface, magnetic anisotropies of the surrounding groups, ring currents of the nearby aromatic systems, and local dihedral angles. The entire body of literature on the use of the amide proton temperature coefficients can be separated into two parts. The first one focuses on the idea that a hydrogen bond between an amide proton and a water molecule is on average more labile and thus more susceptible to temperature changes than an intramolecular $>\text{N}-\text{H}\cdots\text{O}=\text{C}<$ H-bond. Therefore,

experimental $^1\text{H}^{\text{N}}$ temperature coefficients are used to separate the amide groups in the interior of the protein and those facing the solvent. This separation is distant from perfect with a considerable overlap between the two distributions as observed from extensive surveys including multiple proteins. Another group of studies focuses on extraction of intramolecular hydrogen bonding parameters, mainly $\text{H}\cdots\text{O}$ distances in a $\text{N-H}\cdots\text{O}=\text{C}<\text{H}$ -bond, from the observed amide chemical shift temperature coefficients of native-state proteins.

Taking H-bond strength as the major factor affecting chemical shift of $^1\text{H}^{\text{N}}$ nuclei^{30–32} we look at their hydrogen bonding status. From the conformation-averaged solution structure of AtPP16-1 whose secondary structure contains only 20% of a total of 156 amino acids²⁴ we obtain only 33 intramolecular H-bonds. However, the fact that δ_{H} of all amide protons sizably shift with temperature should mean that all of them are H-bonded. This supposition is consistent with the argument that backbone polar groups not engaged in hydrogen-bonding are energetically expensive, and that all potential H-bonding donors and acceptors are believed to be H-bonded intramolecularly or with solvent water for a sufficient fraction of time.³³ In general, H-bonds contribute favorably to protein stability^{34–37} and are thus indispensable. We therefore assume that all amide hydrogens are H-bonded on the NMR timescale. Since H-bonds in our dataset are both intra and intermolecular we did not measure $^3\text{J}_{\text{NC}'}$ values, but relied on the following empirical relation between δ_{H} and r_{OH} due to Wagner et al³⁸

$$\Delta\delta_{\text{H}} \cong \frac{19.2}{r_{\text{OH}}^3} - 2.3 \quad (1)$$

for determination of the changes in H-bond distances with temperature. Here, r_{OH} is the H-bond distance, and $\Delta\delta_{\text{H}}$ is obtained by subtracting the corresponding random coil shift from the observed chemical shift.

To obtain $\Delta\delta_{\text{H}}$ values for each of the experimental temperatures, the Poulsen IDP/IUP random coil chemical shifts for the AtPP16-1 sequence were calculated at the appropriate temperatures using the NIH server (<https://spin.niddk.nih.gov/bax>). Figure 4a shows the length of H-bonds donated by NHs whose proton chemical shifts are measured. We note that the H-bond length given here is the $\text{H}\cdots\text{O}$ distance (r_{OH}), and not the donor-acceptor N–O distance (r_{NO}) as found in many X-ray as well as some NMR literature. Four H-bonds,

donated by NHs of K17, K114, L98 (not shown), and M101 appear relatively longer. Since L98 and M101 belong to the sole helix of AtPP16-1, the larger length for the two corresponding H-bonds of the $>\text{N-H}\cdots\text{O}=\text{C}<$ type might mean a bent in the helix such that the two residues lie on the outer side of the helix curvature. The helix indeed appears slightly curved in the model of the averaged NMR structure. The mean of r_{OH} for all other residues is 2.44 (SD 0.57) Å at 296 K but decreases linearly to 2.27 (SD 0.50) Å as the protein is heated up to 308 K (Figure 4a), suggesting thermal strengthening of the H-bonds.

The slope of the temperature dependence of r_{OH} in Figure 4a forms the metric for the linear thermal expansion coefficient (α) of the corresponding H-bond. In the temperature interval of T_1 and T_2 ,

$$\alpha(T_1, T_2) = \frac{1}{\langle r_{\text{OH}} \rangle} \frac{\partial r_{\text{OH}}}{\partial T} \quad (2)$$

where we take 2.3 Å for the mean length $\langle r_{\text{OH}} \rangle$. The variation of the α value in the temperature range 296 to 308 K for hydrogen bonds of donor residues is shown in Figure 4b. Clearly, the linear expansion coefficient is negative for all residues, reflecting thermal contraction of the H-bonds, the mean expansion coefficient $\langle \alpha \rangle \sim -5.6 \times 10^{-3} \text{ K}^{-1}$, the modulus of which is the average thermal contraction coefficient for all H-bonds measured.

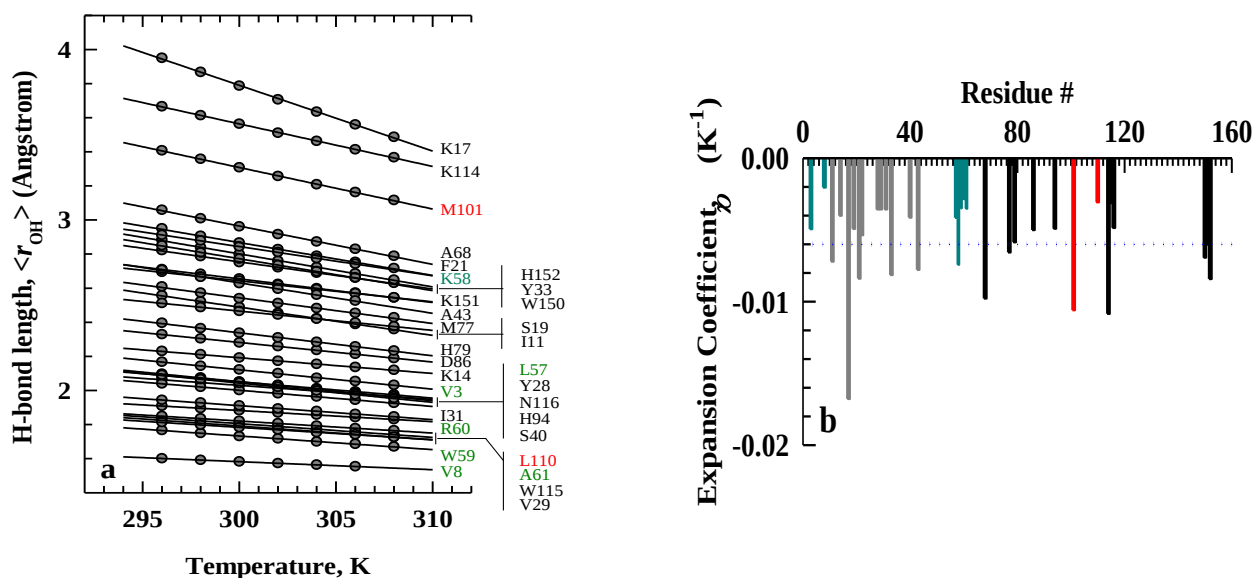


Figure 3.4. Length and linear expansion coefficient of hydrogen bonds. (a) The linear thermal dependence of hydrogen bond length of the indicated donor residues. Labels in red and cyan correspond to residues in the α -helix and β -strands, respectively. The decrease in bond length is a measure of increase in the bond strength. (b) The expansion coefficient of each of the measured H-bonds is negative. The bars for donor residues in the α -helix and β -strands are distinguished by red and cyan, respectively. The dotted horizontal line represents the average expansion coefficient.

The slope of the temperature dependence of r_{OH} in Figure 4a forms the metric for the linear thermal expansion coefficient (α) of the corresponding H-bond. In the temperature interval of T_1 and T_2 ,

$$\alpha(T_1, T_2) = \frac{1}{\langle r_{\text{OH}} \rangle} \frac{\partial r_{\text{OH}}}{\partial T} \quad (2)$$

where we take 2.3 Å for the mean length $\langle r_{\text{OH}} \rangle$. The variation of the α value in the temperature range 296 to 308 K for hydrogen bonds of donor residues is shown in Figure 4b. Clearly, the linear expansion coefficient is negative for all residues, reflecting thermal contraction of the H-bonds. The modulus of the mean expansion coefficient, $\langle \alpha \rangle \sim -5.6 \times 10^{-3} \text{ K}^{-1}$, is the average thermal contraction coefficient for all H-bonds measured.

Instances of thermal contraction of H-bonds in proteins are not abundant. Of the $>\text{N}-\text{H}\cdots\text{O}=\text{C}<$ type of H-bonds measured for ubiquitin²⁷ and GB3 protein,²⁸ only one in the former and three in the latter show negative thermal expansion. Cardiac troponin C, whose sequence size is comparable with that of AtPP16-1, has 15 H-bonds with negative thermal expansion.³⁹ Likewise, BPTI has about 15 H-bonds that register negative expansion coefficient.³⁸ The present result however appears unusual in that $\langle \alpha \rangle$ is negative for all the analyzed $\text{H}\cdots\text{O}$ distances. Since no other instance of a protein with negative expansion for all measured hydrogen bonds is known, we posit that the root-mean-square volume fluctuation of the protein should decrease with increasing temperature.

3.4.4. Temperature dependence of backbone C_α and $^1H^{C_\alpha}$ chemical shifts.

The difficulties associated with the interpretation of chemical shifts are substantially less with regard to C_α and $^1H^{C_\alpha}$ atoms, because their shifts are sensitive to a fewer factors including magnetic anisotropies of the surrounding groups and backbone torsion angles. Magnetic anisotropy arising from the bond magnetic susceptibility of $O=C<$ should produce a minor shielding (σ) because of large proximity of the C_α nucleus, although the shielding arising from the susceptibility of the $C_\alpha-C'$ bond is likely to have relatively larger effect. It has been discussed widely that the predominant factors affecting the chemical shift of C_α are the backbone torsion angles ϕ and ψ .^{40–42} But the chemical shift of $^1H^{C_\alpha}$ seems to depend little on the torsion angle ψ while both ϕ and the peptide magnetic anisotropy produce shielding at the $^1H^{C_\alpha}$ nucleus.⁴³ These considerations appear to suggest that the temperature coefficients of C_α ($\sim +0.1$ to $\sim +26.1$ ppb K^{-1}) and $^1H^{C_\alpha}$ ($\sim +8.4$ to $\sim +13.3$ ppb K^{-1}) derived from Figure 2 truly reflect deviations of the backbone atoms from their equilibrium positions due to thermal fluctuations.

3.4.5. Thermal shortening of global correlation time, τ_c .

To examine if the purported negative linear expansion is also reflected in hydrodynamics, we calculated the variation of τ_c of AtPP16-1 in the 298–304 K range of temperature by using the relaxation times T_1 and T_2 of the resolved ^{15}N resonances (Supporting Information). One assumes that fast internal motions influence T_1 and T_2 uniformly, and T_2 is not shortened by chemical exchange or faster pseudo-first-order processes. The T_1/T_2 ratio of all residues is not useful, but only those for which the ratio is within one standard deviation from the mean.⁴⁴ The global rotational correlation time is then calculated by

$$\tau_c = (4\pi\nu_N)^{-1} \sqrt{\frac{6T_1}{T_2} - 7} \quad (3)$$

where ν_N is Larmor frequency of ^{15}N (=50.68 MHz). The decrease of τ_c in such a narrow range of temperature (Figure 3.5) indicates that the protein has contracted by $\sim 40\%$. The contraction is not because of a conformational collapse, because the thermal melts probing secondary and tertiary structures and tryptophan side-chain anisotropy (Figure 3.2) did not indicate a conformational transition in this range of temperature. The contraction may rather

arise from decreasing volume fluctuation which is much faster than spin relaxation and hydrodynamics.

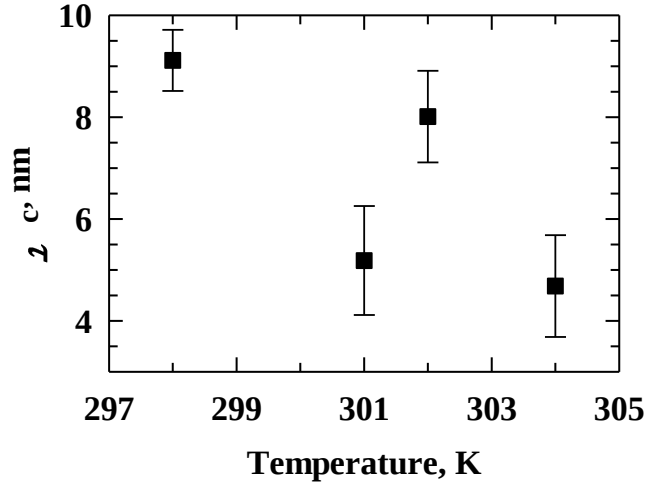


Figure 3.5. Global correlation time of AtPP16-1 with temperature.

3.4.6. Thermally demoted volume fluctuation.

To find out if atom fluctuations indeed decrease with temperature, we carried out protein compressibility measurements using a vibrating tube density meter. Measured sound velocity in the protein molecule (u) and the equilibrium density of the protein (ρ) were used to calculate adiabatic compressibility coefficient $\beta_S = 1/(\rho u^2)$ in the range 292–309 K (Figure 3.6a). For an isotropic protein system in aqueous medium the adiabatic and isothermal compressibility coefficients, β_S and β_T , respectively, are related as⁴⁵

$$\beta_T \equiv \beta_S + \frac{\kappa^2 T}{\rho C_p} \quad (4)$$

where C_p is the heat capacity of the protein and κ is the volume coefficient of thermal expansion. Since the magnitude of C_p is very large compared to that of κ , the second term in equation 4 is often neglected to project $\beta_S \sim \beta_T$ for aqueous solutions⁴⁵ so that β_S can be used to estimate volume fluctuation of the protein by the precise thermodynamic relation^{46,47}

$$\langle \delta V^2 \rangle = k_B T V_M \beta_T \quad (5)$$

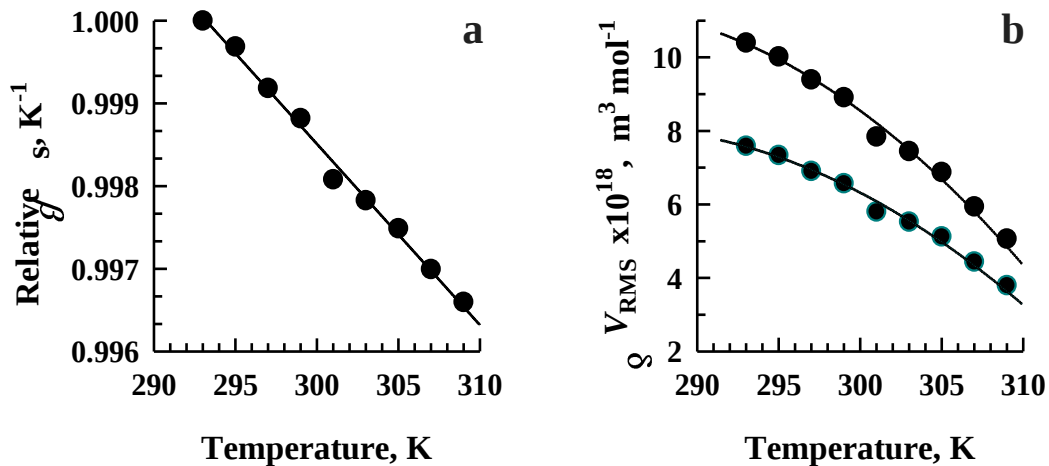


Figure 3.6. (a) Decrease of β_S with temperature. (b) Root-mean-square volume fluctuation of AtPP16-1, uncorrected (black) and corrected (cyan) for temperature dependence of the intrinsic volume V_M .

where $\langle \delta V^2 \rangle$ is the mean-square volume fluctuation. The intrinsic volume V_M , which is the sum of van der Waals volumes of the protein atoms and void spaces in the protein, can be taken as 1.27 M .^{48,49} The molecular mass of AtPP16-1 is 17.3 kDa, which yields $V_M \sim 2.19 \times 10^{-26} \text{ m}^3$ at ambient temperature. It is however difficult to obtain the temperature dependence of V_M . To this end, we used published data on amyloid peptide $A\beta_{42}$ ²² to approximate the temperature gradient near the ambient temperature so as to estimate the V_M values for AtPP16-1 in the temperature range 292–309 K. Figure 3.6b shows the root mean square volume fluctuation, δV_{RMS} , as a function of temperature using both temperature-corrected and uncorrected values of V_M . Clearly, δV_{RMS} decreases non-linearly with temperature irrespective of introducing the correction for V_M . The decrease of δV_{RMS} from 292 to 309 K (Figure 3.6c) is $\sim 37\%$, which is in excellent agreement with $\sim 40\%$ reduction in τ_c observed by NMR relaxation measurements (Figure 3.5), suggesting that the intrinsic thermal expansion of the native protein is negative.

3.4.7. Basis for Negative Thermal Expansion.

We conjecture a qualitative model without proof. Any model pertaining to thermal expansion must be based on concerted motions of vibrational modes resulting in linear or volume or both expansions. The frequency spectrum of $3N-6$ normal modes can be given in terms of both number density of normal modes and δV_{RMS} (Figure 3.7). As general it is, the number of low-frequency modes decreases to just one or two with the frequency approaching as low as $\sim 1 \text{ cm}^{-1}$. This should happen because concerted in-phase fluctuation of hundreds of atoms is a rare event. It may so happen that a fewer low-frequency modes cause in-phase fluctuation of all atoms, or nearly so, to bring about large δV_{RMS} , whose amplitude reduces with thermally promoted high-frequency modes.

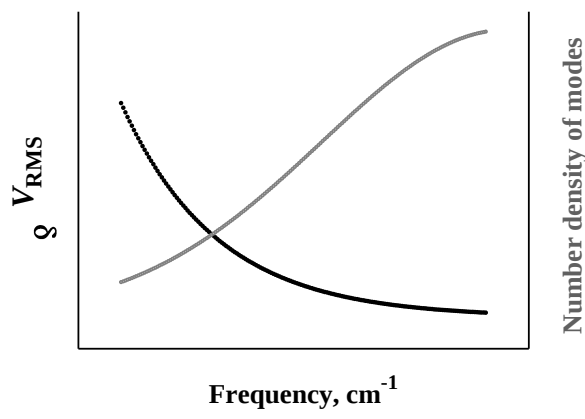


Figure 3.7. Normal mode densities (grey curve) and root mean square volume fluctuations (black curve) as a function of vibrational frequency. Large amplitude collective fluctuations corresponding to only one or a fewer modes of very low frequency take place at lower temperature.

Such frequency and amplitude spectra of normal modes of proteins are known from earlier work,^{50,51} but the question is why negative thermal expansion of proteins is a rare phenomenon? Although there is no definite answer as yet, one of the factors could be the

corrugations of IDP surfaces which influence the density of surface hydration. Since fluctuations of protein atoms are directly related to hydration dynamics, it is possible that thermal fluctuations in normally structured and intrinsically disordered proteins are very different. The rationale for this suggestion is consistent with several recent reports that temperature induces compaction of intrinsically disordered proteins,^{52–61} even though there is no clear mention if these proteins undergo conformational transition or they expand negatively with temperature. The contraction must not result in conformational changes or chain folding. The model surmised may not entirely reproduce experimental results, but further experiments and normal mode computations can provide additional clues.

3.5. REFERENCES

1. Hummel, F. A. Thermal expansion properties of some synthetic lithia minerals. *J. Am. Ceram. Soc.* **1951**, 34, 235–239.
2. Graham, J.; Wadsley, A. D.; Weymouth, J. H.; Williams, L. S. A new ternary oxide, ZrW_2O_8 . *J. Am. Ceram. Soc.* **1959**, 42, 570.
3. Martinek, C.; Hummel, F. A. Linear thermal expansion of three tungstates. *J. Am. Ceram. Soc.* **1968**, 51, 227–228.
4. Boilot, J. P.; Salanie, J. P. Phase transformation in $\text{Na}_{1+x}\text{Si}_x\text{Zr}_2\text{P}_{3-x}\text{O}_{12}$ compounds. *Mater. Res. Bull.* **1979**, 14, 1469–1477.
5. Lenain, G. E.; McKinsty, H. A.; Alamo, J.; Agarwal, D. K. Structural model for thermal expansion in $\text{MZr}_2\text{P}_3\text{O}_{12}$ (M=Li, Na, K, Rb, Cs). *J. Mater. Sci.* **1987**, 22, 17–22.
6. Roy, R.; Agarwal, D. K.; McKinsty, H. A. Very low thermal expansion coefficient materials. *Annu. Rev. Mater. Sci.* **1989**, 19, 59–81.
7. Limaye, S. Y.; Agarwal, D. K.; Roy, R.; Mehrotra, Y. Synthesis, sintering and thermal expansion of $\text{Ca}_{1-x}\text{Sr}_x\text{Zr}_4\text{P}_{16}\text{O}_{24}$ – an ultralow thermal-expansion ceramic system. *J. Mater. Sci.* **1991**, 26, 93–98.
8. Sleight, A. W. Thermal contraction. *Endeavour* **1995**, 19, 64–68.
9. Korthuis, V.; Khosrovani, N.; Sleight, A. W.; Roberts, N.; Dupree, R.; Warren, W.

- W. Negative thermal expansion and phase transitions in the $\text{ZrV}_{2-x}\text{P}_x\text{O}_7$ series. *Chem. Mater.* **1995**, 7, 412–417.
10. Mary, T. A.; Evans, J. S. O.; Vogt, T.; Sleight, A. W. Negative thermal expansion from 0.3 to 1050 Kelvin in ZrW_2O_8 . *Science* **1996**, 272, 90–92.
 11. Lind C. (2012). Two Decades of Negative Thermal Expansion Research: Where Do We Stand?. *Materials (Basel, Switzerland)*, 5(6), 1125–1154. <https://doi.org/10.3390/ma5061125>
 12. Gao, Q., Wang, J., Sanson, A., Sun, Q., Liang, E., Xing, X., & Chen, J. (2020). Discovering Large Isotropic Negative Thermal Expansion in Framework Compound $\text{AgB}(\text{CN})_4$ via the Concept of Average Atomic Volume. *Journal of the American Chemical Society*, 142(15), 6935–6939. <https://doi.org/10.1021/jacs.0c02188>
 13. Frauenfelder, H., Hartmann, H., Karplus, M., Kuntz, I. D., Jr, Kuriyan, J., Parak, F., Petsko, G. A., Ringe, D., Tilton, R. F., Jr, & Connolly, M. L. (1987). Thermal expansion of a protein. *Biochemistry*, 26(1), 254–261. <https://doi.org/10.1021/bi00375a035>
 14. Heibl, M.; Maksymiw, R. Anomalous temperature dependence of the thermal expansion of proteins. *Biopolymers* **1991**, 31, 161–167.
 15. Young, A. C., Tilton, R. F., & Dewan, J. C. (1994). Thermal expansion of hen egg-white lysozyme. Comparison of the 1.9 Å resolution structures of the tetragonal form of the enzyme at 100 K and 298 K. *Journal of molecular biology*, 235(1), 302–317. [https://doi.org/10.1016/s0022-2836\(05\)80034-8](https://doi.org/10.1016/s0022-2836(05)80034-8)
 16. Kurinov, I. V., & Harrison, R. W. (1995). The influence of temperature on lysozyme crystals. Structure and dynamics of protein and water. *Acta crystallographica. Section D, Biological crystallography*, 51(Pt 1), 98–109. <https://doi.org/10.1107/S0907444994009261>
 17. Sassi, P., Perticaroli, S., Comez, L., Giugliarelli, A., Paolantoni, M., Fioretto, D., & Morresi, A. (2013). Volume properties and spectroscopy: a terahertz Raman investigation of hen egg white lysozyme. *The Journal of chemical physics*, 139(22), 225101. <https://doi.org/10.1063/1.4838355>
 18. Sasisanker, P.; Oleinikova, A.; Weingärtner, H.; Ravindra, R.; Winter, R. Solvation

- properties and stability of ribonuclease A studied by dielectric relaxation and differential scanning /pressure perturbation calorimetry. *Phys. Chem. Chem. Phys.* **2004**, 6, 1899–1905.
19. Dellarole, M., Kobayashi, K., Rouget, J. B., Caro, J. A., Roche, J., Islam, M. M., Garcia-Moreno E, B., Kuroda, Y., & Royer, C. A. (2013). Probing the physical determinants of thermal expansion of folded proteins. *The journal of physical chemistry. B*, 117(42), 12742–12749. <https://doi.org/10.1021/jp401113p>
 20. Ryoji, M.; Terazima, M. Transient thermal expansion of a protein in solution after photo-excitation of the chromophore: deoxymyoglobin. *Bull. Chem. Soc. Japan* **2003**, 76, 1707–1712.
 21. Zhang, C.; Shomali, A.; Guyer, R.; Keten, S.; Coasne, B.; Derome, D.; Carmeliet, J. Distinguishing heat and moisture effects on biopolymer mechanics. *Macromolecules* **2020**, 53, 1527–1535.
 22. Brovchenko, I., Burri, R. R., Krukau, A., Oleinikova, A., & Winter, R. (2008). Intrinsic thermal expansivity and hydrational properties of amyloid peptide Abeta42 in liquid water. *The Journal of chemical physics*, 129(19), 195101. <https://doi.org/10.1063/1.3012562>
 23. Xoconostle-Cázares, B., Xiang, Y., Ruiz-Medrano, R., Wang, H. L., Monzer, J., Yoo, B. C., McFarland, K. C., Franceschi, V. R., & Lucas, W. J. (1999). Plant paralog to viral movement protein that potentiates transport of mRNA into the phloem. *Science (New York, N.Y.)*, 283(5398), 94–98. <https://doi.org/10.1126/science.283.5398.94>
 24. Sashi, P., Singarapu, K. K., & Bhuyan, A. K. (2018). Solution NMR Structure and Backbone Dynamics of Partially Disordered Arabidopsis thaliana Phloem Protein 16-1, a Putative mRNA Transporter. *Biochemistry*, 57(6), 912–924. <https://doi.org/10.1021/acs.biochem.7b01071>
 25. Santi Swarupini, D., & Bhuyan, A. K. (2018). Amyloid fibrillation of an intrinsically disordered plant phloem protein AtPP16-1 under acidic condition. *Biophysical chemistry*, 237, 1–8. <https://doi.org/10.1016/j.bpc.2018.03.004>
 26. Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D., & Kay, L. E. (1994). Backbone

- dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by ^{15}N NMR relaxation. *Biochemistry*, 33(19), 5984–6003. <https://doi.org/10.1021/bi00185a040>
27. Cordier, F., & Grzesiek, S. (2002). Temperature-dependence of protein hydrogen bond properties as studied by high-resolution NMR. *Journal of molecular biology*, 317(5), 739–752. <https://doi.org/10.1006/jmbi.2002.5446>
 28. Hong, J., Jing, Q., & Yao, L. (2013). The protein amide $^1\text{H}(\text{N})$ chemical shift temperature coefficient reflects thermal expansion of the $\text{N-H}\cdots\text{O}=\text{C}$ hydrogen bond. *Journal of biomolecular NMR*, 55(1), 71–78. <https://doi.org/10.1007/s10858-012-9689-3>
 29. Veltri, T., de Oliveira, G., Bienkiewicz, E. A., Palhano, F. L., Marques, M. A., Moraes, A. H., Silva, J. L., Sorenson, M. M., & Pinto, J. R. (2017). Amide hydrogens reveal a temperature-dependent structural transition that enhances site-II Ca^{2+} -binding affinity in a C-domain mutant of cardiac troponin C. *Scientific reports*, 7(1), 691. <https://doi.org/10.1038/s41598-017-00777-6>
 30. Cierpicki, T., & Otlewski, J. (2001). Amide proton temperature coefficients as hydrogen bond indicators in proteins. *Journal of biomolecular NMR*, 21(3), 249–261. <https://doi.org/10.1023/a:1012911329730>
 31. Cierpicki, T., Zhukov, I., Byrd, R. A., & Otlewski, J. (2002). Hydrogen bonds in human ubiquitin reflected in temperature coefficients of amide protons. *Journal of magnetic resonance (San Diego, Calif.: 1997)*, 157(2), 178–180. <https://doi.org/10.1006/jmre.2002.2597>
 32. Abraham, M. H.; Abraham, R. J. A simple and facile NMR method for the determination of hydrogen bonding by amide N–H protons in protein models and other compounds. *New. J. Chem.* **2017**, 41, 6064–6066.
 33. Fleming, P. J., & Rose, G. D. (2005). Do all backbone polar groups in proteins form hydrogen bonds? *Protein science: a publication of the Protein Society*, 14(7), 1911–1917. <https://doi.org/10.1110/ps.051454805>
 34. Myers, J. K.; Pace, C. N. Hydrogen bonding stabilizes globular proteins. *Biophys. J.* **1996**, 71, 2033–2039.
 35. Takano, K., Yamagata, Y., Funahashi, J., Hioki, Y., Kuramitsu, S., & Yutani, K.

- (1999). Contribution of intra- and intermolecular hydrogen bonds to the conformational stability of human lysozyme(.). *Biochemistry*, 38(39), 12698–12708. <https://doi.org/10.1021/bi9910169>
36. Shi, Z., Krantz, B. A., Kallenbach, N., & Sosnick, T. R. (2002). Contribution of hydrogen bonding to protein stability estimated from isotope effects. *Biochemistry*, 41(7), 2120–2129. <https://doi.org/10.1021/bi011307t>
 37. Pace, C. N., Fu, H., Lee Fryar, K., Landua, J., Trevino, S. R., Schell, D., Thurlkill, R. L., Imura, S., Scholtz, J. M., Gajiwala, K., Sevcik, J., Urbanikova, L., Myers, J. K., Takano, K., Hebert, E. J., Shirley, B. A., & Grimsley, G. R. (2014). Contribution of hydrogen bonds to protein stability. *Protein science: a publication of the Protein Society*, 23(5), 652–661. <https://doi.org/10.1002/pro.2449>
 38. Wagner, G.; Pardi, A.; Wuthrich, K. Hydrogen Bond Length and ^1H NMR Chemical Shifts in Proteins. *J. Am. Chem. Soc.* **1983**, 105, 5948-5949.
 39. Veltri, T., de Oliveira, G., Bienkiewicz, E. A., Palhano, F. L., Marques, M. A., Moraes, A. H., Silva, J. L., Sorenson, M. M., & Pinto, J. R. (2017). Amide hydrogens reveal a temperature-dependent structural transition that enhances site-II Ca^{2+} -binding affinity in a C-domain mutant of cardiac troponin C. *Scientific reports*, 7(1), 691. <https://doi.org/10.1038/s41598-017-00777-6>
 40. Spera, S.; Bax, A. Empirical correlation between protein backbone conformation and C_α and C_β ^{13}C nuclear magnetic resonance chemical shifts. *J. Am. Chem. Soc.* **1991**, 113, 5490–5492.
 41. Asakawa, N.; Kurosu, H.; Ando, I. Structural study of peptides containing K-alanine residues by ab initio chemical shielding calculation. *J. Mol. Struct.* **1994**, 323, 279–285.
 42. Wishart, D. S., & Sykes, B. D. (1994). Chemical shifts as a tool for structure determination. *Methods in enzymology*, 239, 363–392. [https://doi.org/10.1016/s0076-6879\(94\)39014-2](https://doi.org/10.1016/s0076-6879(94)39014-2)
 43. Ösapay, K.; Case, D. A. A new analysis of proton chemical shifts in proteins. *J. Am. Chem. Soc.* **1991**, 113, 9436–9444.
 44. Fushman, D., Weisemann, R., Thüring, H., & Rüterjans, H. (1994). Backbone

- dynamics of ribonuclease T1 and its complex with 2'GMP studied by two-dimensional heteronuclear NMR spectroscopy. *Journal of biomolecular NMR*, 4(1), 61–78. <https://doi.org/10.1007/BF00178336>
45. Sarvazyan A. P. (1991). Ultrasonic velocimetry of biological compounds. *Annual review of biophysics and biophysical chemistry*, 20, 321–342. <https://doi.org/10.1146/annurev.bb.20.060191.001541>
 46. Naganathan, A. N.; Perez-Jimenez, R.; Muñoz, V.; Sanchez-Ruiz, J. M. Estimation of protein folding free energy barriers from calorimetric data by multi-model Bayesian analysis. *Phys. Chem. Chem. Phys.* **2011**, 13, 17064–17076.
 47. Cooper A. (1976). Thermodynamic fluctuations in protein molecules. *Proceedings of the National Academy of Sciences of the United States of America*, 73(8), 2740–2741. <https://doi.org/10.1073/pnas.73.8.2740>
 48. Eden, D., Matthew, J. B., Rosa, J. J., & Richards, F. M. (1982). Increase in apparent compressibility of cytochrome c upon oxidation. *Proceedings of the National Academy of Sciences of the United States of America*, 79(3), 815–819. <https://doi.org/10.1073/pnas.79.3.815>
 49. Richards F. M. (1977). Areas, volumes, packing and protein structure. *Annual review of biophysics and bioengineering*, 6, 151–176. <https://doi.org/10.1146/annurev.bb.06.060177.001055>
 50. Chalikian, T. V., Gindikin, V. S., & Breslauer, K. J. (1995). Volumetric characterizations of the native, molten globule and unfolded states of cytochrome c at acidic pH. *Journal of molecular biology*, 250(2), 291–306. <https://doi.org/10.1006/jmbi.1995.0377>
 51. Brookes, B. R.; Janezic, D.; Karplus, M. Harmonic analysis of large systems. *J. Comput. Chem.* **1995**, 16, 1522–1542.
 52. Whitmire, S. E., Wolpert, D., Markelz, A. G., Hillebrecht, J. R., Galan, J., & Birge, R. R. (2003). Protein flexibility and conformational state: a comparison of collective vibrational modes of wild-type and D96N bacteriorhodopsin. *Biophysical journal*, 85(2), 1269–1277. [https://doi.org/10.1016/S0006-3495\(03\)74562-7](https://doi.org/10.1016/S0006-3495(03)74562-7)
 53. Uversky V. N. (2009). Intrinsically disordered proteins and their environment: effects

- of strong denaturants, temperature, pH, counter ions, membranes, binding partners, osmolytes, and macromolecular crowding. *The protein journal*, 28(7-8), 305–325. <https://doi.org/10.1007/s10930-009-9201-4>
54. Kjaergaard, M., Nørholm, A. B., Hendus-Altenburger, R., Pedersen, S. F., Poulsen, F. M., & Kragelund, B. B. (2010). Temperature-dependent structural changes in intrinsically disordered proteins: formation of alpha-helices or loss of polyproline II?. *Protein science : a publication of the Protein Society*, 19(8), 1555–1564. <https://doi.org/10.1002/pro.435>
 55. Shkumatov, A. V., Chinnathambi, S., Mandelkow, E., & Svergun, D. I. (2011). Structural memory of natively unfolded tau protein detected by small-angle X-ray scattering. *Proteins*, 79(7), 2122–2131. <https://doi.org/10.1002/prot.23033>
 56. Ciasca, G., Campi, G., Battisti, A., Rea, G., Rodio, M., Papi, M., Pernot, P., Tenenbaum, A., & Bianconi, A. (2012). Continuous thermal collapse of the intrinsically disordered protein tau is driven by its entropic flexible domain. *Langmuir : the ACS journal of surfaces and colloids*, 28(37), 13405–13410. <https://doi.org/10.1021/la302628y>
 57. Luo, Y., Dinkel, P., Yu, X., Margittai, M., Zheng, J., Nussinov, R., Wei, G., & Ma, B. (2013). Molecular insights into the reversible formation of tau protein fibrils. *Chemical communications (Cambridge, England)*, 49(34), 3582–3584. <https://doi.org/10.1039/c3cc00241a>
 58. Langridge, T. D., Tarver, M. J., & Whitten, S. T. (2014). Temperature effects on the hydrodynamic radius of the intrinsically disordered N-terminal region of the p53 protein. *Proteins*, 82(4), 668–678. <https://doi.org/10.1002/prot.24449>
 59. Battisti, A., Ciasca, G., Grottesi, A., & Tenenbaum, A. (2017). Thermal compaction of the intrinsically disordered protein tau: entropic, structural, and hydrophobic factors. *Physical chemistry chemical physics : PCCP*, 19(12), 8435–8446. <https://doi.org/10.1039/c6cp07683a>
 60. Kumar, N., Shukla, S., Kumar, S., Suryawanshi, A., Chaudhry, U., Ramachandran, S., & Maiti, S. (2008). Intrinsically disordered protein from a pathogenic mesophile *Mycobacterium tuberculosis* adopts structured conformation at high temperature. *Proteins*, 71(3), 1123–1133. <https://doi.org/10.1002/prot.21798>

61. Brocca, S., Samalíková, M., Uversky, V. N., Lotti, M., Vanoni, M., Alberghina, L., & Grandori, R. (2009). Order propensity of an intrinsically disordered protein, the cyclin-dependent-kinase inhibitor Sic1. *Proteins*, 76(3), 731–746.
<https://doi.org/10.1002/prot.22385>
62. Hamdi, K., Salladini, E., O'Brien, D. P., Brier, S., Chenal, A., Yacoubi, I., & Longhi, S. (2017). Structural disorder and induced folding within two cereal, ABA stress and ripening (ASR) proteins. *Scientific reports*, 7(1), 15544.
<https://doi.org/10.1038/s41598-017-15299-4>

CHAPTER 4

Electrostatic effect on the folding of AtPP16-1

4.1. ABSTRACT

The finding that the IDP AtPP16-1 is stable at pH ~4 (net charge +14.7) but not at pH ~6 (net charge -2.4) hints at the key role of sequence charges to maintain structure, conformation, and dynamics of the protein. Screening the intrinsic charges leads to reversible acquisition of additional secondary and tertiary structures not only at pH 4.1 but also at 6.1, suggesting that the salt-induced structure acquisition is not critically dependent on net charge on the protein. Surprisingly, the charged-screened protein state undergoes reversible and cooperative folding in the presence of high urea. Such urea-induced reversible folding of proteins, globular or intrinsically disordered, is not known. The urea-refolded state may be structured to a large extent, but is likely to be an off-pathway abortive form.

4.2. INTRODUCTION

Several studies have suggested that charge distribution, and net and fractional charge per residue critically determine the size, shape and conformational fluctuations of IDPs.¹⁻⁵ The importance of charge patterning has been analyzed by both theoretical^{5,6} and experimental methods.^{2,7} Numerous influential reviews have also been published on the charge-IDP problem⁴ all of the literature is not listed here. It would appear that the vast set of data on this subject implicates conformer fluctuations, molecular dimension, and functional properties with changing charge pattern on the protein sequence, and many invoke popular theories of polymer physics.⁴ Despite the advances made, the effect of screening of IDP charges by solvent ions has not been worked out in sufficient details. The problem is analogous to stabilization of acid- or base-destabilized proteins to molten globule-like states by ions, although there are specific objectives of charge-screening of

IDPs. It is of interest to know if the charge-screened state is functionally active and whether the structural alterations brought about is reversible. One may also like to know the extent to which structural dynamics and chain fluctuations are perturbed. ampholyte theory and all the large increase in protein structure induced by salt irrespective of the net charge on the chain have been scarcely described.

This chapter describes the effect of solvent ions on the structure and conformation of AtPP16-1 at pH ~4 where the protein is most stable and pH ~6 where it is least stable. The secondary and tertiary structural alterations due screened charges are monitored by the usual CD and fluorescence methods. The protein indeed acquires substantial structure in the presence of NaCl, but surprisingly, the salt-stabilized state undergoes further reversible and cooperative folding induced by urea.

4.3. MATERIALS AND METHODS

Production and purification of AtPP16-1 has been described already. Samples for urea titration of the protein were prepared by mixing two stock solutions that were identical with respect to protein content (~8 μ M) and buffer (50 mM acetate, pH 4 or 20 mM PIPES, pH 6.5), except that one contained no urea (native protein solution) and the other did in the 6–9 M range (unfolded protein solution). If salt titration was to carry out, one of the stock solutions contained the highest concentration of salt to be used, generally 2 mM NaCl, and the other contained no salt. The two solutions were mixed appropriately to generate samples in the interval of 0.15 M urea or 0.1 M NaCl. This titration approach not only ensures uniformity of the samples except variable urea content, but also tests the reversibility of the unfolding reaction. Samples were equilibrated at 25°(±1)°C for ~1 h before measuring CD and fluorescence in succession. Far-UV CD (218 nm) readings were taken in a AVIV SF420 instrument by averaging ~30 points acquired in the time base mode, and fluorescence at the appropriate emission wavelength was noted from the average of two emission spectra (excitation at 280 nm) recorded in a JASCO FP–8300 spectro fluorometer. The pH and urea content of the samples were routinely confirmed by reading pH and refractive index after completing the experiment.

4.4. RESULTS AND DISCUSSION

4.4.1. A revisit to the pH dependence of *AtPP16-1* conformers

Since the pattern and distribution of charges in the sequences of IDPs are dominant determinants of conformations,^{1,2,4,5,8,9} the solvent pH is expected to influence structure, conformation, and stability of disordered proteins considerably. The pH-dependence of the conformation of *AtPP16-1* that we have been considering from time to time is reviewed in Figure 4.1). The conformations at the extremes of pH, most likely resulting from acid and base unfolding will be excluded from discussion here. The protein is visibly most stable at pH 4.1(± 1) and 8.3(± 1) corresponding to net charges of +14.7 and -6, respectively, but contains little secondary structure at pH 6.1(± 1) when the net charge content is -2.4. This result restates the importance of sequence charges in the determination of IDP conformation.

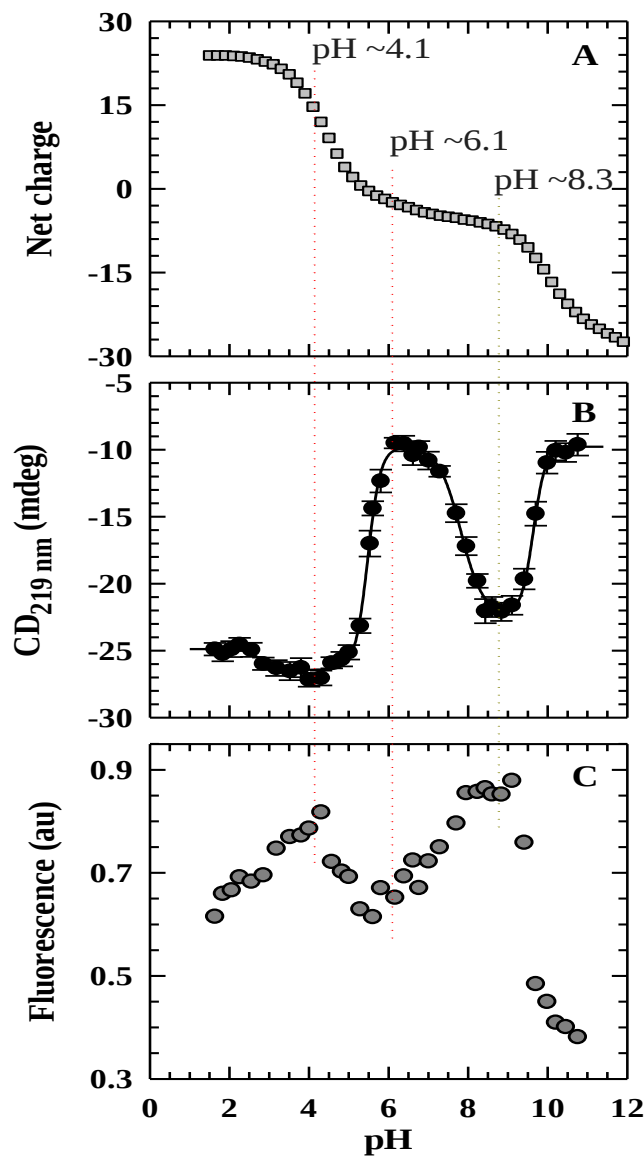


Figure 4.1. (A) Net charge on AtPP16-1 with pH. (B) Secondary structure content with pH. (C) Change in fluorescence at different pH. The solid line through data in (B) is drawn by a multi-pK_a model described in Chapter 5.

4.4.2. Folding and stability at pH 4.1.

One notices that the protein registers little secondary structure and quenched fluorescence in the region of instability. Using these two probes as markers we set out to determine the reversible folding and stability of AtPP16-1 at pH 4.1(±1) and 6.1(±1). It may appear odd

to do experiments at the latter pH where defined structural elements are absent, but the rationale is to modulate the action of protein charges by extrinsic ions and then examine the conformation(s). We present the urea-induced folding data at pH 4.1 first (Figure 4.2). The urea-dependence of fluorescence spectra produces two isosbestic points (Figure 4.2a), suggesting that the tertiary structure must unfold by at least one folding intermediate. The 14-nm red-shifted transition of the wavelength of maximum emission ($\lambda_{\max}$) shown in the *inset* indicates that tertiary structure unfolding results in surface exposure of aromatic chains. These results already suggest that AtPP16-1 at pH 4.1 has a defined tertiary structure that melts away with increments of urea. An inspection of the 331-nm fluorescence decrease with urea hints at deviation of the unfolding from a simple two-state process (Figure 4.2b), characteristic of small globular proteins. The results are affirmative of the involvement of at least one structural intermediate ($N \rightleftharpoons I \rightleftharpoons U$) which leads to fitting the data by the use of the three-state equation

$$S = (c_f + m_f x) + (c_{u1} + m_{u1} x) \frac{e^{\frac{-\Delta G_1 + m_{g1}}{RT}}}{1 + e^{\frac{-\Delta G_1 + m_{g1}}{RT}}} + (c_{u2} + m_{u2} x) \frac{e^{\frac{-\Delta G_2 + m_{g2}}{RT}}}{1 + e^{\frac{-\Delta G_2 + m_{g2}}{RT}}} \dots\dots\dots(1)$$

where c_{ui} and m_{ui} are baseline parameters for the intermediate ($i = 1$) and unfolded ($i = 2$) states, and ΔG_i and m_{gi} are free energy of unfolding and equilibrium m -values, respectively, for the first ($i = 1$) and second ($i = 2$) transitions. The iterated values of ΔG_i and m_{gi} are listed in Table 4.1.

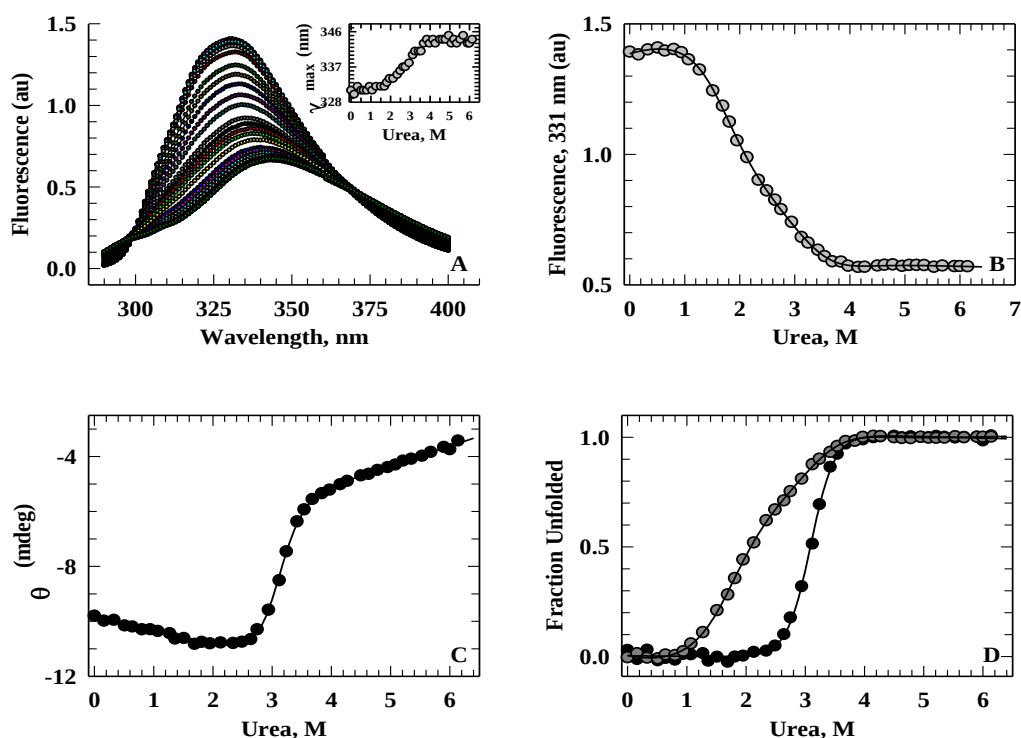


Figure 4.2. Urea denaturation of AtPP16-1 at pH 4.1, 50 mM sodium acetate, 25°C. (A) Fluorescence emission spectra (ex: 280 nm) at different urea concentrations. (B) Fluorescence-monitored denaturation by urea with the solid line drawn according to a three-state model ($N \rightleftharpoons I \rightleftharpoons U$, where I is an intermediate). (C) Far-UV CD-monitored denaturation is described by a two-state process, which is the overall $N \rightleftharpoons U$ transition. (D) Fractions of tertiary and secondary structure with urea are plotted to emphasize on the accumulation of an equilibrium folding intermediate.

Table 4.1. Urea-induced equilibrium unfolding

pH	CD		ΔG_1	Fluorescence		
	ΔG	m_g		m_{g1}	ΔG_2	m_{g2}
4.1(± 0.1)	9.11	3.00	3.00	1.86	5.59	1.80
6.3(± 0.1)	4.38	1.38	2.16	0.30	4.30	1.62

ΔG is in kcal mol^{-1} and m_g is given in $\text{kcal mol}^{-1} \text{M}^{-1}$.

Error in the determination of ΔG and m_g are ± 0.2 and ± 0.1 , respectively.

The secondary structure at pH 4.1 unfolds by a single cooperative transition yielding ΔG and m_g values of 9.11 kcal mol⁻¹ and 3.00 kcal mol⁻¹ M⁻¹ (Figure 2c). A comparison of the fraction of each type of structure unfolded with urea (Figure 2d) clearly indicates the operation of a minimal mechanism involving a structural intermediate ((N $\rightleftharpoons$ I $\rightleftharpoons$ U)). We call it a minimal mechanism because we have not searched for more of such intermediates. The result is however striking; this is perhaps one of few instances, if not the only one, where an IDP is seen to undergo equilibrium unfolding by populating a well-defined intermediate in which the tertiary structure is partially retained.

4.4.3. The excitement of charge screening

4.4.3.1. pH 6.3

The presence of little structure at pH 6.3 in spite of a net charge of -2.4 on the sequence is atypical of globular proteins. This must mean that charge-charge repulsion, as in pH 4.1, plays a significant role in the maintenance of IDP structure. To find out the action of extrinsic ions on the unstable conformation of AtPP16-1 at pH 6.3 we titrated the protein with NaCl in the 0–2 M range and monitored the levels of secondary structure (Figure 4.3a). Clearly, the secondary structure builds up with Na⁺ and Cl⁻ ions, suggesting that some mechanism other than minor charge screening leads to structure acquisition and folding. The events of charge screening and secondary structure formation are quick, and this must be a specific structural change because CD spectra of the salt-containing samples recorded ~65 hours later (Figure 4.3b) showed no drift in the ellipticity values. Similar results are obtained with NaCl dependence of fluorescence; the spectra (Figure 4.3c) and the intensity of fluorescence emission (Figure 4.3d) are time-invariant. Note that fluorescence of AtPP16-1 increases with tertiary structure formation. Charge screening, which is purely an electrostatic effect operational at low ionic strength, is unlikely to bring about the folding and structure formation to the extent observed. At higher ionic strength, the salting-out effect on uncharged amino acid side chains sets in. Polar side chains, which account for ~ 20% of the residues in the AtPP16-1, are likely to be salted-out at higher ionic strength. This effect has been proposed to control

conformational changes in IDPs,⁶ but there is no explanation at present as to how salting-out can restore structure and folding sequence.

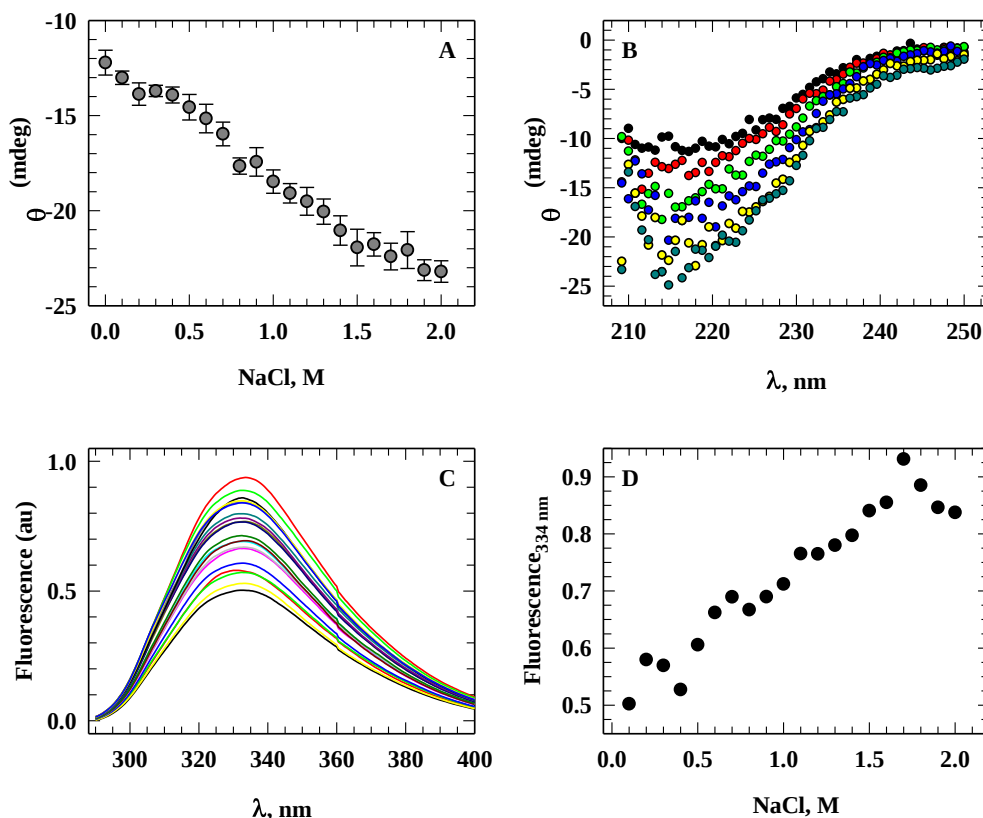


Figure 4.3. Charge-screening at pH 6.3. (A) NaCl titration of AtPP16-1 monitored by 218-nm CD. (B) The spectra of different salt-containing samples recorded ~65 hours later. The NaCl concentrations corresponding to the spectra from top to bottom are 0, 0.4, 0.8, 1.2, 1.6, and 2 M. (C) Fluorescence spectra at different NaCl concentrations showing the absence of an isosbestic point. (D) The 334-nm fluorescence intensity with NaCl.

4.4.3.2. pH 4.1

Going by the charge screening idea, a reverse of the pH 6.3 result above should be obtained at pH 4.1 where repealing the charge-charge repulsions may lead to structure loss and chain collapse. This does not happen. Salt addition rather leads to the formation

of additional secondary structure on the protein (Figure 4.4a), and the increase of fluorescence indicates addition of tertiary structure (Figure 4.4b). We note that the fluorescence spectrum across the 0–2 M NaCl scale hardly undergoes a shift in the emission wavelength. The salt effect on *AtPP16-1* at pH 4.1 (net charge +14.7) is typical of charge screening by which extrinsic anions transform a protein to a more compact structured form akin to ion stabilization of acid- or base-denatured states of globular proteins to the respective molten globule states.¹⁰

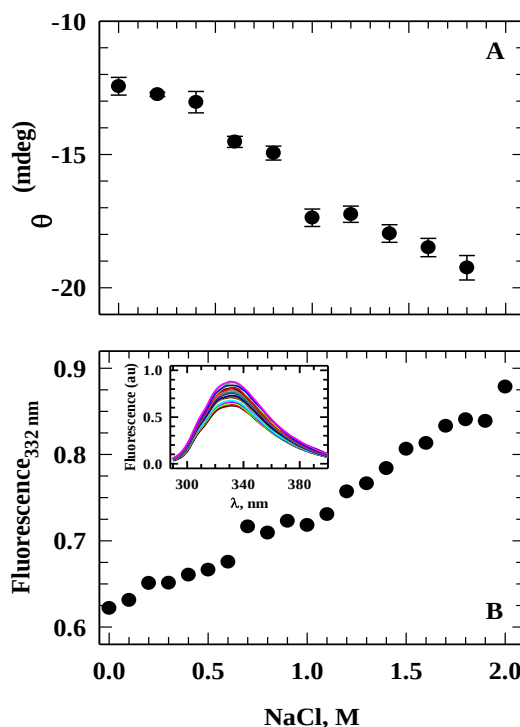


Figure 4.4. Charge-screening at pH 4.1 monitored by 218-nm CD (A) and 332-nm fluorescence (B).

4.4.4. The NaCl stabilized/structured proteins are off-pathway

The substantial structure acquisition and folding in the presence of NaCl give rise to the suspicion if the protein state so generated is native-like. To check on this, *AtPP16-1* in 2M NaCl was titrated with urea at pH 4.1, 25°C, and the results shown in Figure 4.5a,b are surprisingly disappointing. The 218-nm CD-monitored secondary structure undergoes

nice cooperative and reversible refolding transitions (Figure 4.5a), but the fluorescence-monitored tertiary structure unfolds with urea (Figure 4.5b). The results are even murkier at pH 6.3, 2M NaCl, where urea refolds both secondary and tertiary structures, the former is a cooperative and the latter is not (Figure 4.5c, d). It is only a procedural exercise to fit the transitions to the two-state equation

$$S = (c_f + m_f x) + (c_u + m_u x) \frac{e^{\frac{-\Delta G + m_g}{RT}}}{1 + e^{\frac{-\Delta G + m_g}{RT}}} \dots\dots\dots(2)$$

where the symbols have usual meaning. The fit parameters are indicated in the respective figures.

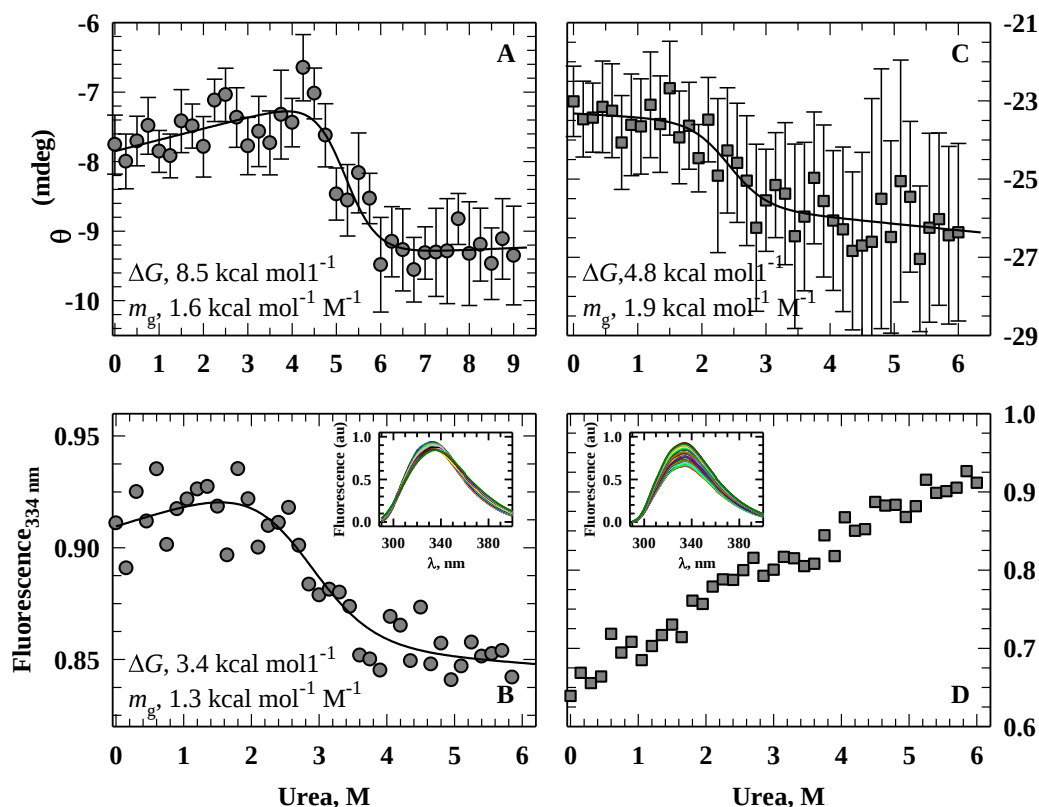


Figure 4.5. Urea-induced refolding-transitions of the 2 M-NaCl stabilized AtPP16-1. At pH 4.1, the 218-nm CD-monitored transition shows refolding of secondary structure (A), but the 334-nm fluorescence-monitored transition reveals unfolding of tertiary structure (B). At pH 6.3, the CD-monitored transition shows a cooperative refolding of secondary

structure (C), but the fluorescence transition is not cooperative, even though the increase in fluorescence is indicative of refolding of tertiary structure.

Urea induced refolding of a globular protein is unknown. Now that we see this in the case of a salt-stabilized IDP we seek an explanation; unfortunately we do not have data at this stage to offer an explanation. There are several possibilities including cooperative collapse and aggregation, and urea-induced protein stiffening and entropic stabilization.^{11,12} This area of salt-stabilized IDP refolding by the urea denaturant is left for future studies.

4.5. CONCLUSIONS

Charge-charge repulsion appears to maintain the folded conformation of AtPP16-1 around pH 4, but dwindling of net charge near the physiological pH (~6.5) leads to unfolding. Screening of the intrinsic charges by solvent ions leads to folding, stabilization, and structure acquisition irrespective of the net charge content. The structure, conformation, and dimensional properties of the charged-screened proteins are not known, but they undergo further acquisition of secondary structure in the presence of ~4–9 M range of urea.

4.6. REFERENCES

1. Das, R. K., and Pappu, R. V. (2013). Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proc. Natl. Acad. Sci. USA* **110**, 13392–13397.
2. Schuler, B., Soranno, A., Hofmann, H., and Nettels, D. (2016). Single-molecule FRET spectroscopy and the polymer physics of unfolded and intrinsically disordered proteins. *Annu. Rev. Biohys.* **45**, 207–231.
3. Sherry, K. P., Das, R. K., Pappu, R. V., and Barrick, D. (2017). Control of transcriptional activity by design of charge patterning in the intrinsically disordered Ram region of the Notch receptor. *Proc. Natl. Acad. Sci. USA* **114**, E9243–E9252.

4. Holehouse, A., and Pappu, R. V. (2018). Collapse transition of proteins and the interplay among backbone, sidechain, and solvent interactions. *Annu. Rev. Biohys.* **47**, 19–39.
5. Huihui, J., Firman, T., and Ghosh, K. (2018). Modulating charge patterning and ionic strength as a strategy to induce conformational changes in intrinsically disordered proteins. *J. Chem. Phys.* **149**, 085101.
6. Wohl, S., Jakubowski, M., and Zheng, W. (2021). Salt-dependent conformational changes of intrinsically disordered proteins. *J. Phys. Chem. Lett.* **12**, 6684–6691.
7. Sørensen, , C. S., and Kjaergaard, M. (2019). Effective concentrations enforced by intrinsically disordered linkers are governed by polymer physics. *Proc. Natl. Acad. Sci. USA* **116**, 23124–23131.
8. Uversky, V. N. (2002). Natively unfolded proteins: a point where biology waits for physics. *Protein Sci.* **11**, 739–756.
9. Mao, A. H., Crick, S. L., Vitalis, A., Chicoine, C. L., and Pappu, R. V. (2010). Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. *Proc. Natl. Acad. Sci. USA* **107**, 8183–8188.
10. Kumar, R., Prabhu, N. P., Rao, D. K., and Bhuyan, A. K. (2006). The alkali molten globule state of horse ferricytochrome c: observation of cold denaturation. *J. Mol. Biol.* **364**, 483–495.
11. Bhuyan, A. K. (2002). Protein stabilization by urea and guanidine hydrochloride. *Biochemistry* **41**, 13386–13394.
12. Kumar, R., Prabhu, N. P., Yadaiah, M., and Bhuyan (2004). Protein stiffening and entropic stabilization in the subdenaturing limit of guanidine hydrochloride. *Biohys. J.* **87**, 2656–2662.

Cysteine substitution mutagenesis of AtPP16-1 alters the average conformational ensemble throughout the backbone

5.1. ABSTRACT

Rapid evolution with mutation of intrinsically disordered proteins (IDPs) has drawn large attention recently. To examine the effect of single mutation of non-bulky amino acids on the energetic stability and average conformational ensemble of IDP, alanine-to-cysteine substitutions in the mRNA transporter protein AtPP16-1 (*Arabidopsis thaliana* Phloem Protein) were made. Mutations do not drastically change the stability of the protein, but the equilibrium folding pathway is altered by introducing more folding intermediates. The pH stability profile of the wild type IDP does not change much either. The tyrosine's of the mutants are largely buried compared to that of the wild type, and the degree of burial varies within the mutants themselves. Chemical shift index of mutation (CSIM) for the assigned ^1H and ^{15}N resonances reveal conformational changes throughout the backbone irrespective of not only the mutation position in the sequence, but also whether the substitution is in structured or disordered regions. This finding is consistent with the idea that single mutation can perturb IDP conformation globally unlike local changes found in most of the ordered and compact globular proteins.

Keywords:

Intrinsically disordered protein (IDP); mutation and evolution of IDP; chemical shift index of mutation; *Arabidopsis thaliana* phloem protein.

5.2. INTRODUCTION

Highly flexible proteins that do not fold to a fixed three dimensional structure and contain disordered regions lacking one or both of tertiary and secondary structures are called intrinsically unstructured or intrinsically disordered proteins (IDPs). They are of universal occurrence, found in all organisms from viruses to humans, and account for nearly 30% of eukaryotic cellular proteins. Even though IDPs are implicated in diabetes, cancer, neurodegenerative, and cardiovascular diseases, they perform some crucial functions such as molecular recognition, structural assembly, posttranslational protein modification, and DNA and RNA binding.¹⁻⁷ They do not or seldom take part in metabolism, biosynthesis, electron transport, and enzyme catalysis.

A major area of IDP research seeks to find out the composition of the amino acid sequence, evolution, and mutations with respect to their disordered regions of structure. Compared to ordered globular proteins, IDPs have low complexity sequence,⁸ undergo rapid evolution with high mutation rates,^{9,10} and play a role in evolutionary origins of cell type diversification.¹¹ With specific regard to IDP mutations, the intrinsically disordered regions (IDRs) are nonconservative and more vulnerable to undergo mutations.^{12,13} Patterns of mutation in IDPs are quite different from those in globular proteins, and the analysis of mutational effects has many facts. Not all mutations are deleterious. Those within IDRs that lead to aggregation of IDPs have been implicated in draining human conditions such as Alzheimer, Parkinson, and Huntington diseases.^{5,14} Mutations that induce various degrees of deviation in the conformation of the IDPs are expected to influence the functional attribute. Substitutions that affect the existing charge pattern distribution can alter the molecular dimension by expanding or collapsing the IDP chain, and thus affect the function.¹⁵ Expansion is also affected by widespread insertion and deletion mutations in IDPs.^{12,16} Another recent study suggests that the perturbation of the average dimensions of the conformational ensemble can occur not only by single mutation, but also double, quadruple and quintuple substitutions.¹⁷ In spite of these possible consequences of mutations, IDPs can still tolerate several potentially disruptive mutations.¹⁸

Clearly, studies on structure-function relationship of IDPs are hindered by rapid evolution¹² and amino acid mutations. More on the consequences of mutation needs to be done for a better understanding of the structure-function relationship of IDPs. Our interest in mutational studies arose from the finding that the solution NMR structure of the putative plant RNA transporter *Atpp16-1* (Figure 5.1) shows intrinsic disorder.¹⁹ The objective initially was to carry out Ala→Cys single mutations of the cysteine less protein so that extrinsic fluorophores could be attached to the cysteine side chain for mapping distances within the chain. After producing the mutants we grew concerned about altering the intrinsic conformational properties of the protein while modifying the cysteine side chain with fluorophore. We then discontinued with the fluorescence resonance transfer experiments and set out to examine the backbone conformational changes brought about by mutation of alanine's at different sequence positions. We were successful to produce only three single Ala→Cys mutants – A2C and A68C, both of which are in unstructured regions (Figure 5.1), and A61C located in the β -strand 2 (residues 56–62). Optical experiments suggest varying degrees of aromatic chain exposure, and the chemical shift indices of mutation (CSIM) establish perturbation throughout the backbone irrespective of the location of the mutation.



Figure 5.1. Ribbon diagram of one of the minor ensembles of conformers of AtPP16-1 showing the positions of A2, A61, and A68. The short β -strand immediately after the helix is not observed in 17 of the 20 conformers generated from NMR data ¹⁹. The long C-terminal tail also contains the H₆ tag.

5.3. MATERIALS AND METHODS

5.3.1. Site-directed mutagenesis and protein purification

The pET28a (+) vector-cloned sequence template of AtPP16-1 that already has *Nco I* and *Xho I* restriction sites was used for PCR amplification of the following primers for site directed mutagenesis.

A2C F 5' AGATATACCATGTGTGTTGGAATCCTTGAGGTTAGTG 3'
 R 5'GATTCCAACACACATTGGTATATCTCCTTCTTAAAGTTAAAC 3'

A61C F 5' GAAATGGAGATTGTGAGTTTCCTGGCTCCGGC 3'
 R 5' AGGAAACTCACATCTCCATTTC AATTTATCATTCCATGTC 3'

A68C F 5' CTCCGGCTTGCGACTACAAACTCATCGTC 3'
 R 5' TGTAGTCGCAGCCGGAGCCAGGAAACTC 3'

The desired mutations were confirmed by DNA sequencing. The procedures for expression and purification of Atpp16-1 have already been described.¹⁹ Briefly, cDNA is initially cloned into pTZ57R/T vector, sub cloned into pET28a (+) vector, and transformed into BL21(DE3) RIL cells for protein expression. Cells are grown in Luria Bertani broth containing kanamycin (50 μ g/mL) and chloramphenicol (50 μ g/mL) at 37°C up to 0.6 optical density at 600 nm before allowing protein expression at 16°C by the addition of IPTG (0.1 mM). After ~8 h of growth, cells are harvested and suspended in phosphate buffer saline. All purification steps are carried out at 4°C. Cells are lysed in 20 mM Tris-HCl, 50 mM NaCl, 5 mM imidazole, pH 8. The supernatant is collected by a 20-min centrifugation at 14000 rpm, passed through a Ni-NTA column equilibrated with

the lysis buffer, washed with the same lysis buffer, and eluted with 20 mM Tris-HCl, 50 mM NaCl, 150 mM imidazole, pH 8. The eluted fractions are dialyzed overnight against 7 mM sodium acetate at pH 4; the part of the protein precipitated during dialysis is removed by centrifugation, and the purity is checked by SDS-PAGE. For production of isotope-labelled protein, all steps are identical except that the cells are grown in M9 medium containing $^{15}\text{NH}_4\text{Cl}$ or both $^{15}\text{NH}_4\text{Cl}$ and ^{13}C -glucose.

5.3.2. pH titration

The buffer system consisted of 10 mM glycine, 4 mM each of sodium acetate, HEPES and PIPES, 8 mM Tris, and 2 mM CAPS. The initial protein solution (8 μM) prepared in this buffer was aliquoted as samples of 0.5 mL each, whose pH was set to different pH values in the range 1.4 – 12.6 using minimal volumes of HCl, acetic acid, and NaOH. The dilution of the initial protein solution in the process of pH adjustment was considered negligible. Following equilibration at 25°C for 2 h, the 218-nm CD value and fluorescence spectra (280-nm excitation) were measured. Each CD value at 218 nm was obtained as an average of ~30 readings measured in the time-base mode, and each fluorescence spectrum was an average of two or four scans.

5.3.3. Equilibrium unfolding

The protocol we use requires mixing the two end-state solutions, one containing the protein in the native state and the other in the unfolded state. The two stock solutions are identical, 10 μM in protein concentration and prepared in 40 mM sodium acetate at pH 4, except that the native protein solution contains no urea while the unfolded contains the highest urea employed (5 or 6 M). This procedure serves to test the reversibility of the unfolding reaction. Samples of the wild type, A2C, A61C, and A68C proteins in different urea prepared by mixing the respective end-state solutions were equilibrated at 24($\pm$ 1)°C for ~2 hours before measurements at the same temperature. Tryptophan fluorescence spectra were measured in a JASCO FP-8300 fluorometer, and far-UV CD spectra or 218-nm constant wavelength CD measurements were done using JASCO J1500 instruments. As a routine workup, the urea content and pH of the samples were determined again after

the measurements. Since the mutants of AtPP16-1 at 25°C were found to be unstable after ~15 h of sample preparation, freshly prepared samples of each mutant protein were used for each of fluorescence and CD measurements.

5.3.4. NMR Spectroscopy

Protein samples for NMR experiments were prepared in a 10% D₂O buffer of 20 mM sodium acetate, pH 4.1, and contained typically 100 µM uniformly ¹⁵N-labeled or ¹⁵N/¹³C doubly-labeled wild type, A2C, A61C, and A68C proteins. The spectral widths for ¹H and ¹⁵N in 2D [¹H–¹⁵N] HSQC spectra were 11 and 36 ppm, respectively, with corresponding carrier frequencies positioned at 4.7 and 117 ppm. Since the mutant proteins appeared unstable showing no resonance dispersion in ~20 h after the [¹H–¹⁵N] HSQC spectrum, 3D HNC(O), HNCA, HNCACB and CBCA(CO)NH experiments were performed in a restricted sense by non-uniform sampling (NUS). The mutant instability also required the use of fresh samples after every 3D experiment. All experiments were done at 298 K in a 500 MHz Bruker (Ascende III) spectrometer equipped with a 5 mm triple resonance probe (TXI). Data were processed initially using TopSpin 3.1 (Bruker) and NMRpipe²⁰ software before transferring them to Sparky for peak picking and assignments.

5.4. RESULTS AND DISCUSSION

5.4.1. Secondary and tertiary structures of Ala→Cys mutants

The amount of secondary structure in the mutant proteins do not appear to match that of the wild type AtPP16-1; while both A2C and A68C demonstrate slightly lower MRE relative to that of the wild type, A61C registers significantly higher MRE (Figure 5.2). The A61 residue is contained in the β-strand 2, but how its substitution with a cysteine can lead to more β content is not clear. It is possible that the β-strand 2 is relatively more constrained in the wild type protein; the substitution of A61 with the cysteine introduces some flexibility so as to favor β-structure, a suggestion emanating from molecular

dynamics simulation studies.²¹ There also are instances of stabilization of protein scaffolds, the (β/α) barrel and the $CS\alpha\beta$ structure for example, by establishing disulfide bridges between secondary structural elements.^{22,23} The A61C mutant can register more β -content if two cysteines establish a disulfide linkage intermolecularly. This scenario refers to protein dimerization, and possibly aggregation, some evidence of which was obtained from NMR spectra and equilibrium experiments that indicated that the mutant proteins tend to aggregate over tens of hours of incubation at room temperature. The spectra reported in Figure 5.2a were however recorded with freshly prepared protein samples stored at 4°C that were exposed to room temperature only to record the spectra. We conclude that the β -content of A61C increases probably due to an increase in flexibility of the β -strand 2 as a result of the mutation.

As a marker of tertiary structure, the 280-nm excited fluorescence spectra recorded for both wild-type and mutant proteins are shown in Figure 5.2b. While all four proteins emit at 334 nm, which is due to tryptophan fluorescing, the Ala→Cys mutants also demonstrate tyrosine emissions appearing as a shoulder at ~310 nm. The shoulder corresponding to the A2C is more distinct relative to that of A61C, and appears as a clear emission peak in the case of A68C. In general, the band width of the excitation light is sufficient to excite the tyrosine phenolic side chain whose absorption maximum is ~275 nm, but the result that the mutants show tyrosine fluorescence while the wild-type does not indicates that the tyrosine's in the mutant proteins are largely buried so that their fluorescence is not quenched. Even the three mutants are different from each other by the criterion of tyrosine burial, the tyrosine fluorescence of A68C is quenched the least. Tyrosine emission is well known to be highly sensitive to the environment, making it a very useful inherent probe to study structure and dynamics of proteins.²⁴ We see that the tertiary structure of the wild-type AtPP16-1 is perturbed sufficiently in the cysteine mutants, the A68C being the most perturbed.

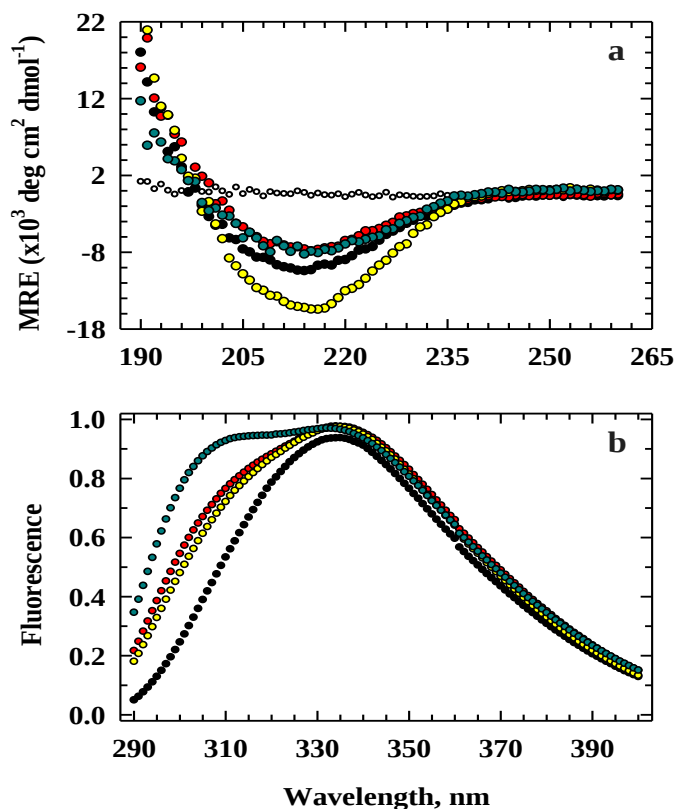


Figure 5.2. Far-UV CD spectra (a), and fluorescence emission (excitation at 280 nm) spectra (b) of wild type (black), A2C (red), A61C (yellow), and A68C (dark cyan) in 10 mM sodium acetate, pH 4.

5.4.2. Stability and folding

To determine the stability of the mutants we probed the protein conformation with urea by measuring both 218-nm CD signals and fluorescence spectra (Figure 5.3). The secondary structure of both wild type and mutants unfold in a single step, and the results satisfactorily describe a two-state transition. The A61C mutant however stands out by a significantly curved pre-transition baseline where the CD signal is ~70% more negative at ~1.5 M urea relative to the value in 0 M urea or the native state. Although curved pretransition in the secondary structure unfolding is fairly common – ferrocyclochrome *c* for example displays this,²⁵ the prominence of the curvature observed here is however

exceptional. An increase in ellipticity to this extent far outweighs any solvent effect that may render the CD signal more negative with respect to that of the native state. The reason for this increase could be urea-induced additional β -content in the A61C mutant, since it is known that urea-denatured state ensembles contain extensive secondary structure commensurate with hydrophobicity of the protein.²⁶ The hydrophobicity need not be rider here because all mutants and the wild-type have almost identical number of hydrophobic surfaces; we suspect, it is rather the pointed mutation in the β -strand 2 which is responsible for an increase in the secondary structure in denaturing levels of urea. It is already posited in the preceding section that β -strand 2 turns flexible as a result of mutation, and the flexibility may favor extension of the β -strand. We have not analyzed this problem experimentally or computationally, hence the suggestion is open to scrutiny.

The secondary structure unfolding curves were fitted (Figure 5.3) to the following two-state model ($N \rightleftharpoons U$) for transition between the folded and unfolded states

$$S = (c_f + m_f x) + (c_u + m_u x) \frac{e^{\frac{-\Delta G + m_g}{RT}}}{1 + e^{\frac{-\Delta G + m_g}{RT}}} \quad (1)$$

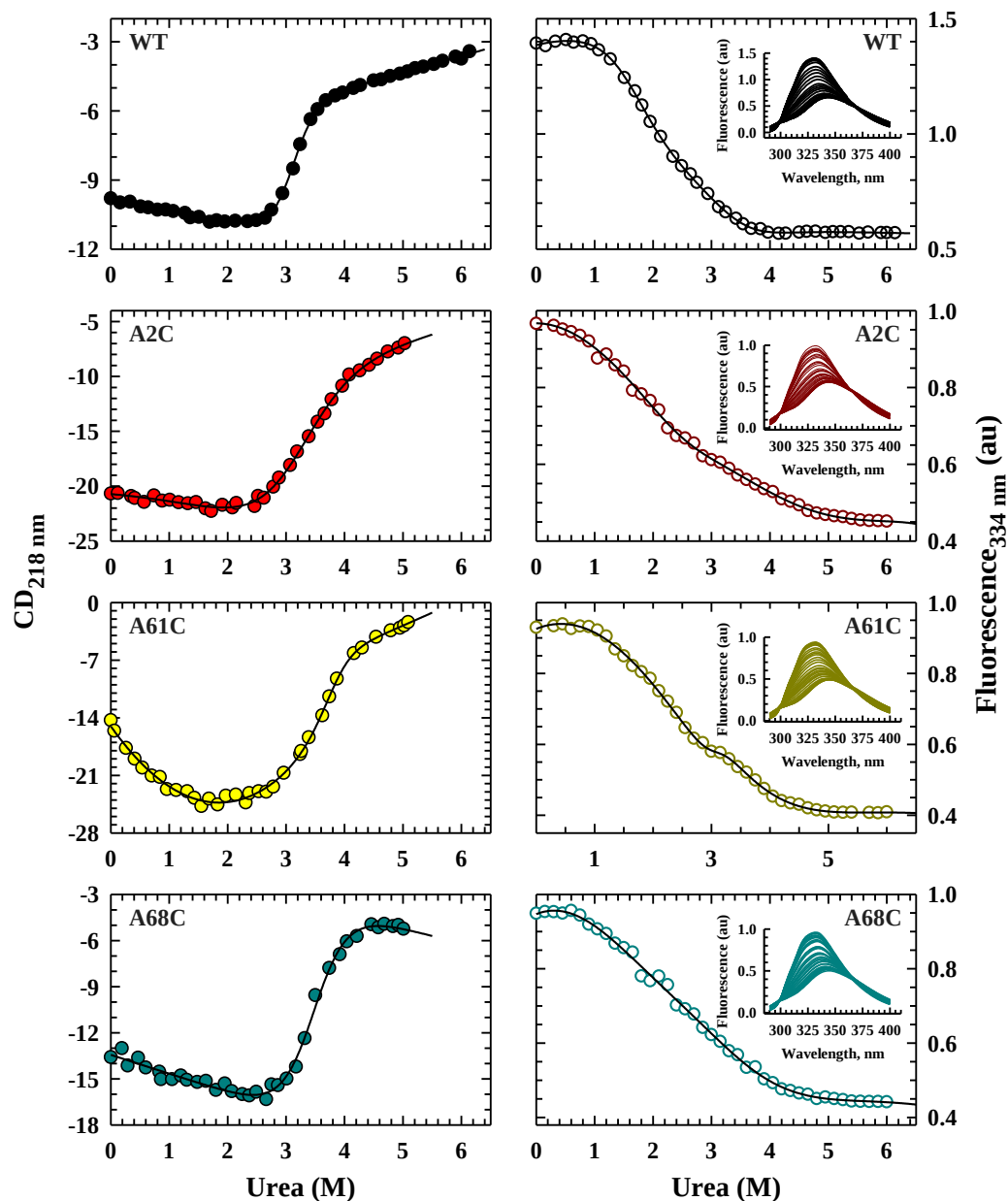


Figure 5.3. Equilibrium unfolding of secondary and tertiary structures the wild type and mutants of AtPP16-1 by urea in 50 mM sodium acetate, pH 4. The CD-monitored transitions are fitted to equation 1, and the fluorescence-monitored ones to equation 2. The urea dependence of fluorescence spectra are shown in the *insets*.

where c_f and m_f are intercept and slope of the native-state baseline, c_u and m_u are intercepts and slopes of the unfolded-state baseline, ΔG is the free energy of unfolding, m_g represents the equilibrium m -value (transition slope), and x is the molar concentration of urea. A second order polynomial was used to fit the curved pretransition baseline of A61C. The equilibrium between the two states is established by the procedure of performing the experiment in which the native– and unfolded–state protein solutions were mixed appropriately to generate the transition curve. The values of ΔG and m_g obtained from the fits (Table 5.1) establishes the stability order A61C>wild-type>A68C>A2C, suggesting that the mutation pointed at β –strand 2 stabilizes the protein by ~ 2 kcal mol^{−1} with reference to the wild-type. As mentioned above, one of the factors contributing to this increased stability originate from the mutation-favored extension of the β content. This result however is not consistent with the thinking that single mutation in a structured part, the β –strand 2 here, of an IDP is destabilizing.²⁷

Table 5.1. Urea-induced equilibrium unfolding parameters

Protein	CD		Fluorescence			
	ΔG	m_g	ΔG_1	m_{g1}	ΔG_2	m_{g2}
Wild type	9.11	5.48	3.27	1.86	5.59	1.65
A2C	4.38	1.38	2.16	0.30	4.30	1.62
A61C	10.51	2.75	2.00	0.56	13.04	4.23
A68C	7.20	2.07	1.50	0.61	2.27	0.82

ΔG is in kcal mol^{−1} and m_g is given in kcal mol^{−1} M^{−1}.

Error in the determination of ΔG and m_g are ± 0.2 and ± 0.1 , respectively.

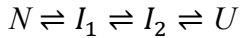
The urea dependence of unfolding of tertiary structures of both wild-type and mutants is quite distinct from that of secondary structure. The overlay of the fluorescence spectra with increasing levels of urea (Figure 5.3, *inset*) clearly shows a cooperative red shift, suggesting that the unfolding of tertiary structure is characterized by a greater

solvent exposure of the aromatic residues irrespective of wild type or mutants. The transitions are shallow in general, shallower for the mutants with respect to the wild-type. Each transition involves at least three states – N , U , and I , where I is an ensemble of intermediate states. A closer look at the shallowness of the transitions raises the possibility of occurrence of more than one intermediate, but the resolution of the experiment and the signal-to-noise do not permit determination of how many intermediates exist. Although each transition in principle can be fitted to a model having more than one intermediate species, they may be overly determined. We therefore describe each transition by a minimal three-state mechanism $N \rightleftharpoons I \rightleftharpoons U$

$$S = (c_f + m_f x) + (c_{u1} + m_{u1} x) \frac{e^{\frac{-dG_1 + m_{g1}}{RT}}}{1 + e^{\frac{-dG_1 + m_{g1}}{RT}}} + (c_{u2} + m_{u2} x) \frac{e^{\frac{-dG_2 + m_{g2}}{RT}}}{1 + e^{\frac{-dG_2 + m_{g2}}{RT}}} \quad (2)$$

in which c_{ui} and m_{ui} are baseline parameters for the intermediate ($i = 1$) and unfolded ($i = 2$) states, and dG_i and m_{gi} are unfolding free energy and equilibrium m -values, respectively, for the first ($i = 1$) and second ($i = 2$) sequential transitions. Note that the unfolded state mentioned here is specific to the unfolded tertiary structure, not the global unfolded state. The extracted values of dG_i and m_{gi} (Table 1) demonstrate that the $N \rightleftharpoons I$ transition is lower in energy than the $I \rightleftharpoons U$ transition, and the stability of I follows the order A61C > wild-type > A2C > A68C. The available data does not allow a discussion on the details of tertiary interactions responsible for stabilizing the intermediates.

One can nevertheless project a scheme for the unfolding of ArPP16-1 irrespective of mutations by including two intermediate states corresponding to the two resolved tertiary structure transitions,



where the intermediates I_1 and I_2 occur in the unfolding of both wild-type and mutants, although the structure of an intermediate varies within the protein species. The details of equilibrium unfolding of disordered proteins are less known at present, but the occurrence of several variably stable intermediates may be a common phenomenon to them. As mentioned, a general probe like fluorescence may not resolve the intermediates confidently especially when shallow transitions occur within a lower range of a denaturant.

The unfolding results of single mutation of IDP presented here do not indicate a drastic change in the stability of the protein, which is dissimilar to a dramatic change in stability generally observed with single mutations of ordered proteins.^{28–32} This is especially true when substitutions are made in the hydrophobic core, consistent with the fact that even a subtle change in the tightly packed core can be destabilizing.³³ There is no well-defined core in IDP, except for occasional hydrophobic clusters, which does not allow projecting the principles of mutation in ordered proteins to disordered ones. We conclude that single mutations in IDP does not change the stability greatly; the folding-unfolding pathway can however be modified.

5.4.3. pH-dependent changes in the structure

The working condition of AtPP16-1 at pH 4 stems from many earlier observations that the protein is most stable at this pH. It should be the working pH in vivo as well because the protein unfolds reversibly in the pH range 6– 7 even in a dilute solution. It is interesting to study the pH titration curve, the experimental data for which are shown in Figure 5.4 (FigureCC.jnb) for both wild-type and mutant proteins. Five transitions across the 1–13 range of pH can be identified by inspection, of which the lowest amplitude one occurring at pH <2.5 and which is likely due to the ionization of the α -COOH group was excluded from further analysis to avoid overestimation of parameters. The rest of the transitions were fitted according to

$$S = \sum_i S_i + \frac{\Delta S_i}{1 + 10^{n_i(\text{pH} - \text{p}K_i)}}, \quad i=1,2,3,4 \quad (3)$$

in which ΔS_i , n_i , and $\text{p}K_i$ are change in the signal amplitude, number of protons titrated, and the corresponding pK-value, respectively, for the i^{th} transition, and the iterated fit parameters are listed in Table 5.2.

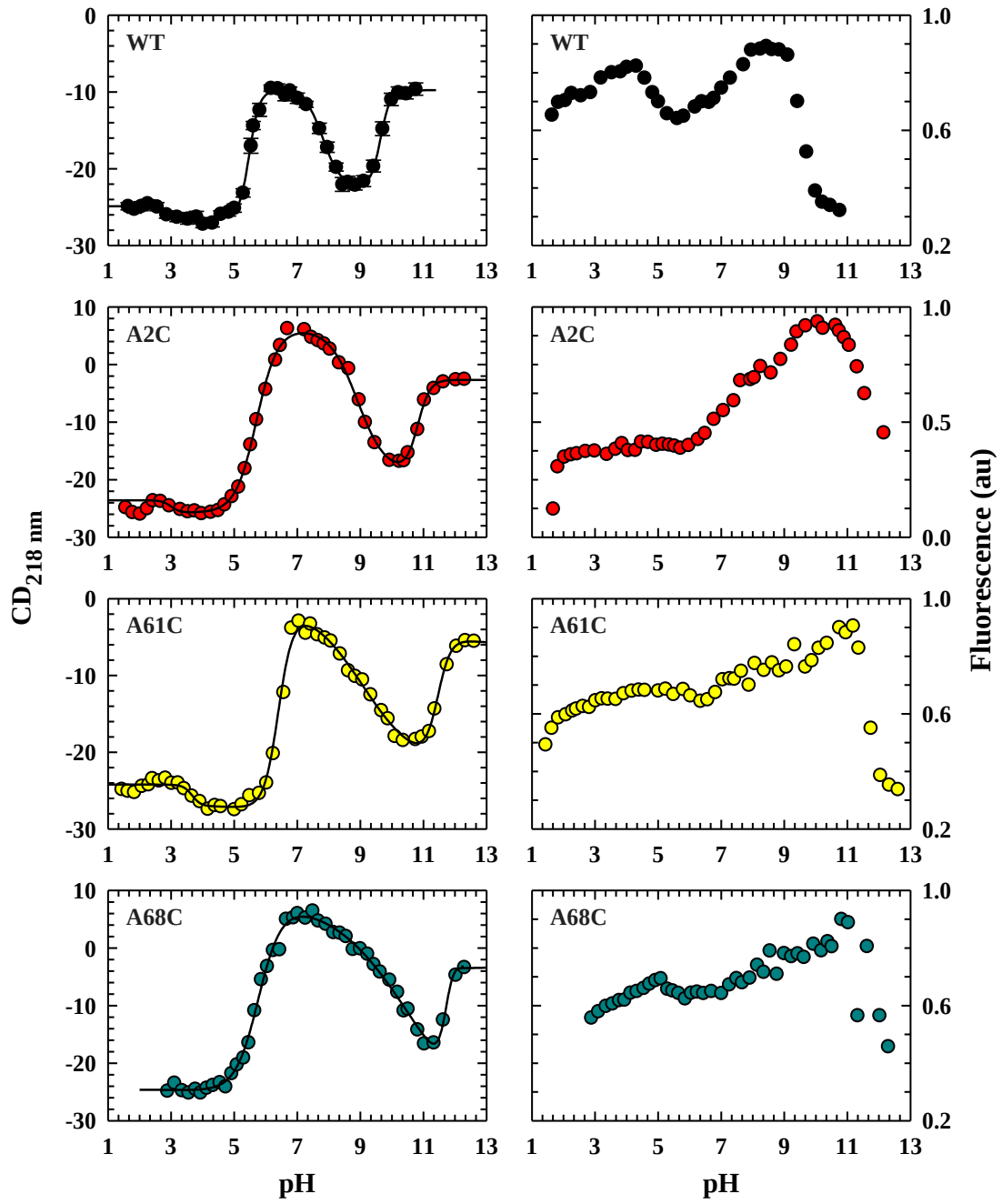


Figure 5.4. pH titration curves for the wild type and mutants of *AtPP16-1*. Data below $\sim$ pH 2.5 in the CD-monitored curves have been excluded from the fits according to equation 3.

Table 5.2. pH-induced conformational changes and the number of protons titrated

AtPP16-1	pK_1	pK_2	pK_3	pK_4	n_1	n_2	n_3	n_4
Wild type	2.78	5.48	7.82	9.64	2.0	2.7	1.6	2.8
A2C	3.00	5.71	8.91	10.83	3.0	1.3	2.2	3.0
A61C	3.60	6.40	9.10	11.46	3.0	1.9	0.5	2.1
A68C	–	5.77	9.52	11.72	–	0.6	1.2	3.1

Before contrasting the titration characteristics of the four protein species it is useful to assign the four pK values to probable titratable groups by considering the parameters for the wild-type alone. It has been known for a long time that accurate accounting for pK values in protein titration is not an easy task³⁴ because a number of factors, including electrostatic and dispersion interactions, the dielectric of the resident environment of titratable groups, and ionic strength of the medium, influence their dissociation. The interactions and the environment polarity may generally be determined if crystal structure is available, but the titrations are carried out in aqueous solutions where the same interactions and the environment need not exist. Regarding disordered proteins, intramolecular Coulombic attraction and repulsion are known to significantly influence structure and conformations of disordered proteins, and the number and distribution of opposite charges along the amino acid sequence are thought to determine the chain compaction and conformations.^{35,36}

There are 14 aspartic acid (E) and 11 glutamic acid (D) residues in AtPP16-1, and we tentatively assign the pK_1 (=2.78) of the wild-type (Table 5.2) to two aspartate carboxyl's in K44–E45, K96–E97, and E104–K105 stretches on the basis of possible electrostatic interactions between the carboxyl's and side chain amines, which can lower the intrinsic pK s of the carboxyl's and stabilize the protein conformation. Assigning two aspartates can account for the two-proton titration, $n_1=2$ (Table 5.2). The value of pK_2

(=5.48) appears fairly higher than the generally observed pK (~ 4) for carboxyl groups of glutamic acid in protein. This difference may be accounted for if the occurrence of electrostatic repulsion of an aspartate side chain with a nearby anion can be invoked. By perusing the sequence one finds the stretch D85–D86 in which short-range Coulombic repulsion between the ionized carboxyl's can shift the intrinsic pK of the carboxyl's to the observed value of 5.48. Assigning pK_2 to D85 and D86 satisfies the two-proton titration, $n_2=2$; the number 2.7 appearing for n_2 in Table 5.2 is in error of data taking and analysis. The value of pK_3 (=7.82) is higher than the intrinsic pK for histidine side chain and lower than those for K, Y, and R. One can take help of $n_3=2$, within experimental error, to spot stretches D55–K56, D69–Y70 and D139–R140 in which cation-anion Coulombic attraction can anomalously down-shift the intrinsic pK s of the basic side-chains, the dissociation of two of which can account for pK_3 . The pK_4 (=9.64) with $n_4=3$ should again be assigned to three basic side chain group whose dissociation is down-shifted. Although one or more of unused basic side chains mentioned above may be involved, but AtPP16-1 has 13 serine's and the it is tempting to invoke the serine richness. Provided the serine hydroxyl interacts with the side chain of the contiguous residue in stretches S19–D20, S94–D95, D119–S120, and H152–S153, the serine hydroxyl may dissociate at a lower pH. The possibility of involvement of other basic residues also cannot be dismissed. These assignments must be considered tentative, and are based only on charges contiguous along the linear sequence; possible long-range electrostatic interactions and the influence of other factors listed above are not analyzed.

Each of the Ala→Cys mutants shows a shift of the titration curve to the right (Figure 4), which is not due error of pH measurement. With reference to the wild type, the shifts for pK_3 and pK_4 are greater than those for pK_1 and pK_2 (Table 5.2), and the number of protons titrated in each transition also changes. The results suggest widespread perturbation in tertiary interactions to alter the strength of electrostatic interactions so as to change the pK shifts. Accounting for the difference in the number of titrating protons is however problematic, one can only suppose that the altered tertiary interactions, as shown by the fluorescence probes in Figure 5.4, add or delete dissociating groups having the same pK . Unlike the cysteine less wild type protein, the cysteines of the mutants may

also contribute to the respective titration curve, but we do not have data to discuss this possibility.

5.4.4. ^1H , ^{15}N backbone assignment of the cysteine mutants

Since we already have the solution structure of the wild-type *AtPP16-1*, the initial conception of this work was to determine the 3D structures of all the mutants and calculate their spectral densities. This ambition suffered a blow when it was found that the NMR spectrum of every cysteine mutant ($\sim 100\ \mu\text{M}$) loses dispersion after ~ 15 hours of experiment at $24(\pm)^\circ\text{C}$. For example, at the end of recording a 2D CPMG or a 3D HNCA spectrum in the non-uniform sampling mode, which sufficiently shortens the data collection time, a HSQC spectrum shows complete loss of chemical shift dispersion as though the protein is unfolded or aggregated. We did not sense the occurrence of this phenomenon during the optical experiments described above, perhaps because the protein concentration was one-tenth of that of NMR and the time taken to complete the measurements was ~ 5 hours at the most. Nonetheless, tests and optimizations at $24(\pm)^\circ\text{C}$ confirmed the feasibility of performing HSQC experiments alone. With this limitation we set out to assign the backbone ^1H and ^{15}N resonances to the extent possible.

The overlay of [^1H - ^{15}N] HSQC spectra of the wildtype and cysteine mutants recorded under identical conditions (Figure 5.5), HSQC-Mutants-Overlaid.ppt) blotches widespread changes in chemical shifts from one to another protein. For reference, the wild-type spectrum with resonance assignments in the sparsely crowded regions is also plotted. Apart from the changes in the chemical shifts, the resonance dispersion in the mutant spectra has decreased, suggesting more conformational disorder in the mutants. The disorder is not introduced locally to the cysteine mutation sites, but across the backbone.

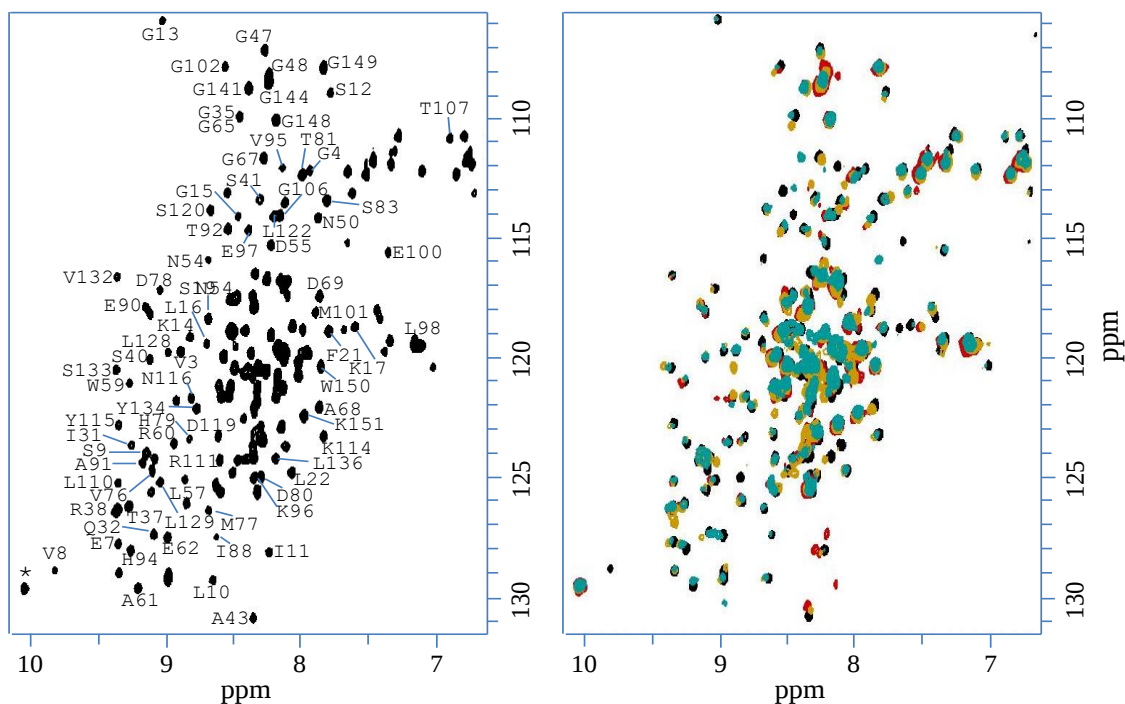


Figure 5.5. [^1H - ^{15}N] HSQC spectra of the wild type and overlaid with those of the mutants (right). The correlation peaks are color coded as black (wild type), red (A2C), yellow (A61C), and dark cyan (A68C). Some assignments in the dispersed regions of the wild type spectrum (left) are labeled.

To determine the extent of shift of the ^1H and ^{15}N backbone resonances of the mutants with reference to those of the wild type, the initially processed data were moved to Sparky software for peak picking and assignment. The difference of the chemical shift of an assigned mutant resonance from the corresponding chemical shift of the wild type protein gives the chemical shift index of mutation (CSIM) for that resonance. CSIM is primarily determined by the extent to which the secondary (local) magnetic fields of the wild-type backbone groups is altered through mutation-induced changes in the backbone dihedral angles, ϕ and ψ . The general practice in mutational studies is to calculate the conventional chemical shift index (CSI), defined as the difference in the chemical shift of a protein resonance with respect to the shift of the same resonance disposed in a predefined random coil, for the wild type and mutants individually. Since the interest

here is to determine the extent of disorder with reference to that of the wild type, CSIM was calculated for all resolved resonances of each of the three mutants.

Figure 5.6 shows CSIM for ^{15}N resonances across the backbone. Values larger than +0.5 or -0.5 were arbitrarily discarded from consideration because they are very likely due to wrong assignments. Clearly, mutation-induced disorder in the backbone conformation is least for the A2C protein, apparently due to N-terminal location of the mutation, and most for the A68C. The mean CSIM values are +0.069 (SD, 0.12) and -0.069 (SD, 0.089) for A2C, +0.103 (SD, 0.121) and -0.153 (SD, 0.151) for A61C, and 0.122 (SD, 0.127) and -0.176 (SD, 0.154) for A68C. Identical results are obtained with CSIM for the amide ^1H resonances – backbone conformation is perturbed most for the A68C mutant.

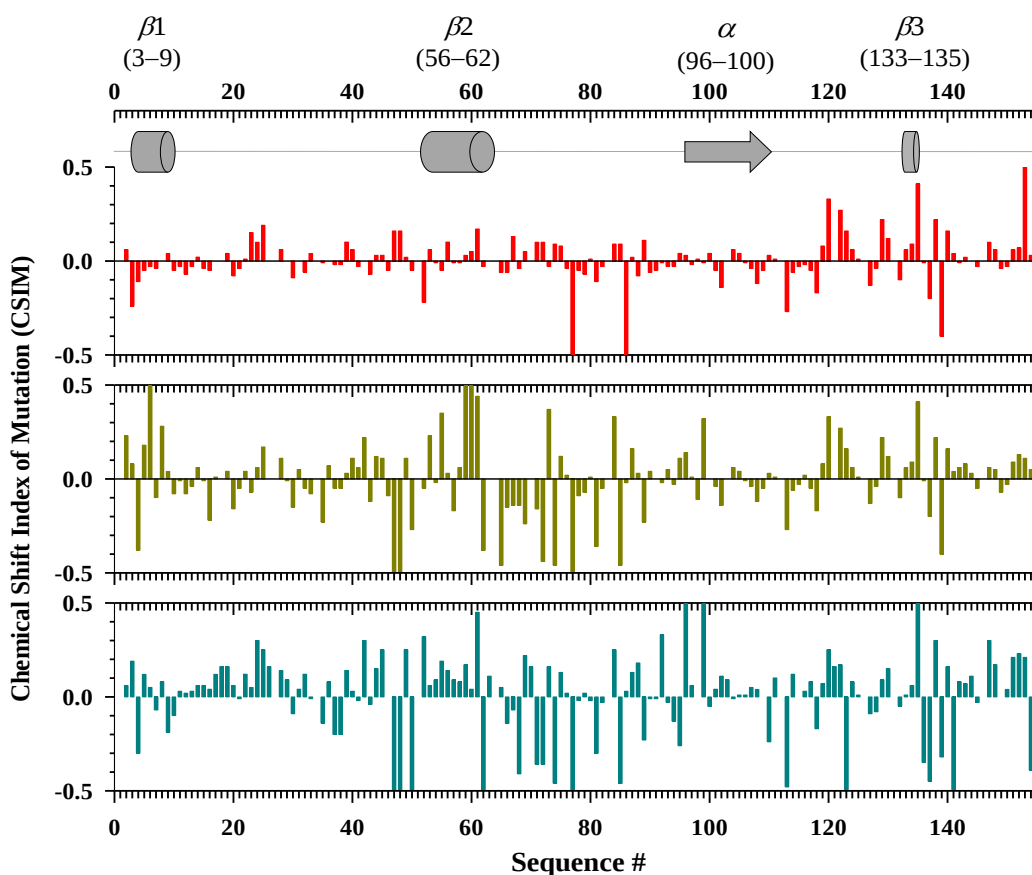


Figure 5.6. Widespread changes in backbone conformation of A2C (top), A61C (middle), and A68C (bottom) assessed by ^{15}N chemical shift index of mutation. Sequence stretches corresponding to secondary structure elements are shown on top.

The observation that Ala→Cys single mutations perturb the IDP conformation all along and throughout the backbone is inconsistent with the long known result for structurally ordered proteins that single mutation does not produce widespread structural changes.^{37–40} Single mutation in ordered proteins brings about only local changes, especially when substitutions are made in the flexible surface overhangs. Even if the local protein environment is averse to the mutated residue, the residue introduced is forced to adapt to the local surrounding without producing extensive changes.⁴¹ Since IDPs do not have a well-defined ordered structure, the adaptive accommodation of a substitution to an already fluctuating environment is irrelevant. The effect rather manifests as conformational changes in all parts of the molecule, which is shown in the case of AtPP16-1 as changes in the backbone dihedral angles all along the protein chain.

5.5. CONCLUSIONS

The nucleic acid carrier plant phloem protein AtPP16-1 is intrinsically disordered and most stable at pH 4. The Ala→Cys mutants A2C, A61C, and A68C of the protein do not show radical changes in stability and pH response. The mutants differ within themselves and from the wild type in terms of both secondary and tertiary structures. Although A2C and A68C appear to have similar extent of secondary structure, slightly lower than that of the wild type, the A61C protein acquires extended secondary structure which may be due to the residence of A61 in β -strand 2. The mutations also produce different tertiary structures with respect to that of the wild type, evident from more quenched fluorescence for most part of the pH scale. Specifically, the tyrosine emission peak, which overlaps with the tryptophan's in the wild type protein, appears as a distinct shoulder centered at ~ 310 nm for all three mutants, suggesting greater burial of tyrosine's in the mutants. The mutations perturb the backbone conformation of the wild type protein extensively,

virtually all along the backbone which might be the outcome of non-adaptive mutation. The mutated residue is not adapted to the local environment because most of the disordered regions are sparse in tertiary interactions, loosely packed, and less compact, hence IDPs do not offer an adaptive environment for the substitution.

5.6. REFERENCES

1. Tompa P. (2002). Intrinsically unstructured proteins. *Trends in biochemical sciences*, 27(10), 527–533. [https://doi.org/10.1016/s0968-0004\(02\)02169-2](https://doi.org/10.1016/s0968-0004(02)02169-2)
2. Dunker, A. K., Brown, C. J., Lawson, J. D., Iakoucheva, L. M., & Obradović, Z. (2002). Intrinsic disorder and protein function. *Biochemistry*, 41(21), 6573–6582. <https://doi.org/10.1021/bi012159>
3. L.M. Iakoucheva, P. Radivojac, C.J. Brown, T.R. O’Connors, J.G. Sikes, Z. Obradovic, A.K. Dunker, The importance of intrinsic disorder for protein phosphorylation. *Nucleic acids research*, 32(3), 1037–1049. <https://doi.org/10.1093/nar/gkh253>
4. Dyson, H. J., & Wright, P. E. (2005). Intrinsically unstructured proteins and their functions. *Nature reviews. Molecular cell biology*, 6(3), 197–208. <https://doi.org/10.1038/nrm1589>
5. Vacic, V., & Iakoucheva, L. M. (2012). Disease mutations in disordered regions-exception to the rule?. *Molecular bioSystems*, 8(1), 27–32. <https://doi.org/10.1039/c1mb05251a>
6. Darling, A. L., & Uversky, V. N. (2018). Intrinsic Disorder and Posttranslational Modifications: The Darker Side of the Biological Dark Matter. *Frontiers in genetics*, 9, 158. <https://doi.org/10.3389/fgene.2018.00158>
7. Forcelloni, S., & Giansanti, A. (2020). Mutations in disordered proteins as early indicators of nucleic acid changes triggering speciation. *Scientific reports*, 10(1), 4467. <https://doi.org/10.1038/s41598-020-61466-5>

8. Radivojac, P., Iakoucheva, L. M., Oldfield, C. J., Obradovic, Z., Uversky, V. N., & Dunker, A. K. (2007). Intrinsic disorder and functional proteomics. *Biophysical journal*, 92(5), 1439–1456. <https://doi.org/10.1529/biophysj.106.094045>
9. Shafee, T., Bacic, A., & Johnson, K. (2020). Evolution of Sequence-Diverse Disordered Regions in a Protein Family: Order within the Chaos. *Molecular biology and evolution*, 37(8), 2155–2172. <https://doi.org/10.1093/molbev/msaa096>
10. Kastano, K., Erdős, G., Mier, P., Alanis-Lobato, G., Promponas, V. J., Dosztányi, Z., & Andrade-Navarro, M. A. (2020). Evolutionary Study of Disorder in Protein Sequences. *Biomolecules*, 10(10), 1413. <https://doi.org/10.3390/biom10101413>
11. Niklas, K. J., Dunker, A. K., & Yruea, I. (2018). The evolutionary origins of cell type diversification and the role of intrinsically disordered proteins. *Journal of experimental botany*, 69(7), 1437–1446. <https://doi.org/10.1093/jxb/erx493>
12. Brown, C. J., Johnson, A. K., & Daughdrill, G. W. (2010). Comparing models of evolution for ordered and disordered proteins. *Molecular biology and evolution*, 27(3), 609–621. <https://doi.org/10.1093/molbev/msp277>
13. Brown, C. J., Johnson, A. K., Dunker, A. K., & Daughdrill, G. W. (2011). Evolution and disorder. *Current opinion in structural biology*, 21(3), 441–446. <https://doi.org/10.1016/j.sbi.2011.02.005>
14. Babu M. M. (2016). The contribution of intrinsically disordered regions to protein function, cellular complexity, and human disease. *Biochemical Society transactions*, 44(5), 1185–1200. <https://doi.org/10.1042/BST20160172>
15. Das, R. K., Ruff, K. M., & Pappu, R. V. (2015). Relating sequence encoded information to form and function of intrinsically disordered proteins. *Current opinion in structural biology*, 32, 102–112. <https://doi.org/10.1016/j.sbi.2015.03.008>
16. Light, S., Sagit, R., Sachenkova, O., Ekman, D., & Elofsson, A. (2013). Protein expansion is primarily due to indels in intrinsically disordered

- regions. *Molecular biology and evolution*, 30(12), 2645–2653.
<https://doi.org/10.1093/molbev/mst157>
17. Munshi, S., Rajendran, D., Ramesh, S., Subramanian, S., Bhattacharjee, K., Kumar, M. R., & Naganathan, A. N. (2020). Controlling Structure and Dimensions of a Disordered Protein via Mutations. *Biochemistry*, 59(2), 171–174. <https://doi.org/10.1021/acs.biochem.9b00678>
 18. Hultqvist, G., Åberg, E., Camilloni, C., Sundell, G. N., Andersson, E., Dogan, J., Chi, C. N., Vendruscolo, M., & Jemth, P. (2017). Emergence and evolution of an interaction between intrinsically disordered proteins. *eLife*, 6, e16059. <https://doi.org/10.7554/eLife.16059>
 19. Sashi, P., Singarapu, K. K., & Bhuyan, A. K. (2018). Solution NMR Structure and Backbone Dynamics of Partially Disordered Arabidopsis thaliana Phloem Protein 16-1, a Putative mRNA Transporter. *Biochemistry*, 57(6), 912–924. <https://doi.org/10.1021/acs.biochem.7b01071>
 20. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., & Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *Journal of biomolecular NMR*, 6(3), 277–293. <https://doi.org/10.1007/BF00197809>
 21. Daggett, V., & Levitt, M. (1993). Protein unfolding pathways explored through molecular dynamics simulations. *Journal of molecular biology*, 232(2), 600–619. <https://doi.org/10.1006/jmbi.1993.1414>
 22. Rotrekl, V., Nejedlá, E., Kucera, I., Abdallah, F., Palme, K., & Brzobohatý, B. (1999). The role of cysteine residues in structure and enzyme activity of a maize beta-glucosidase. *European journal of biochemistry*, 266(3), 1056–1065. <https://doi.org/10.1046/j.1432-1327.1999.00948.x>
 23. Zasloff M. (2002). Antimicrobial peptides of multicellular organisms. *Nature*, 415(6870), 389–395. <https://doi.org/10.1038/415389a>
 24. Antosiewicz, J. M., & Shugar, D. (2016). UV-Vis spectroscopy of tyrosine side-groups in studies of protein structure. Part 2: selected applications. *Biophysical reviews*, 8(2), 163–177. <https://doi.org/10.1007/s12551-016-0197-7>

25. Bhuyan, A. K., & Udgaonkar, J. B. (2001). Folding of horse cytochrome c in the reduced state. *Journal of molecular biology*, 312(5), 1135–1160. <https://doi.org/10.1006/jmbi.2001.4993>
26. Nick Pace, C., Huyghues-Despointes, B. M., Fu, H., Takano, K., Scholtz, J. M., & Grimsley, G. R. (2010). Urea denatured state ensembles contain extensive secondary structure that is increased in hydrophobic proteins. *Protein science : a publication of the Protein Society*, 19(5), 929–943. <https://doi.org/10.1002/pro.370>
27. Huang, H., & Sarai, A. (2012). Analysis of the relationships between evolvability, thermodynamics, and the functions of intrinsically disordered proteins/regions. *Computational biology and chemistry*, 41, 51–57. <https://doi.org/10.1016/j.compbiolchem.2012.10.001>
28. Sun, D. P., Sauer, U., Nicholson, H., & Matthews, B. W. (1991). Contributions of engineered surface salt bridges to the stability of T4 lysozyme determined by directed mutagenesis. *Biochemistry*, 30(29), 7142–7153. <https://doi.org/10.1021/bi00243a015>
29. Daopin, S., Alber, T., Baase, W. A., Wozniak, J. A., & Matthews, B. W. (1991). Structural and thermodynamic analysis of the packing of two alpha-helices in bacteriophage T4 lysozyme. *Journal of molecular biology*, 221(2), 647–667. [https://doi.org/10.1016/0022-2836\(91\)80079-a](https://doi.org/10.1016/0022-2836(91)80079-a)
30. Dao-pin, S., Anderson, D. E., Baase, W. A., Dahlquist, F. W., & Matthews, B. W. (1991). Structural and thermodynamic consequences of burying a charged residue within the hydrophobic core of T4 lysozyme. *Biochemistry*, 30(49), 11521–11529. <https://doi.org/10.1021/bi00113a006>
31. Fusi, P., Goossens, K., Consonni, R., Grisa, M., Puricelli, P., Vecchio, G., Vanoni, M., Zetta, L., Heremans, K., & Tortora, P. (1997). Extreme heat- and pressure-resistant 7-kDa protein P2 from the archaeon *Sulfolobus solfataricus* is dramatically destabilized by a single-point amino acid substitution. *Proteins*, 29(3), 381–390. [https://doi.org/10.1002/\(sici\)1097-0134\(199711\)29:3<381::aid-prot11>3.0.co;2-j](https://doi.org/10.1002/(sici)1097-0134(199711)29:3<381::aid-prot11>3.0.co;2-j)

32. Roberts, J. A., Digby, H. R., Kara, M., El Ajouz, S., Sutcliffe, M. J., & Evans, R. J. (2008). Cysteine substitution mutagenesis and the effects of methanethiosulfonate reagents at P2X2 and P2X4 receptors support a core common mode of ATP action at P2X receptors. *The Journal of biological chemistry*, 283(29), 20126–20136. <https://doi.org/10.1074/jbc.M800294200>
33. Varadarajan, R., & Richards, F. M. (1992). Crystallographic structures of ribonuclease S variants with nonpolar substitution at position 13: packing and cavities. *Biochemistry*, 31(49), 12315–12327. <https://doi.org/10.1021/bi00164a005>
34. Tanford, C., & Roxby, R. (1972). Interpretation of protein titration curves. Application to lysozyme. *Biochemistry*, 11(11), 2192–2198. <https://doi.org/10.1021/bi00761a029>
35. Mao, A. H., Crick, S. L., Vitalis, A., Chicoine, C. L., & Pappu, R. V. (2010). Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 107(18), 8183–8188. <https://doi.org/10.1073/pnas.0911107107>
36. Das, R. K., & Pappu, R. V. (2013). Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proceedings of the National Academy of Sciences of the United States of America*, 110(33), 13392–13397. <https://doi.org/10.1073/pnas.1304749110>
37. De Filippis, V., Sander, C., & Vriend, G. (1994). Predicting local structural changes that result from point mutations. *Protein engineering*, 7(10), 1203–1208. <https://doi.org/10.1093/protein/7.10.1203>
38. Takano, K., Funahashi, J., Yamagata, Y., Fujii, S., & Yutani, K. (1997). Contribution of water molecules in the interior of a protein to the conformational stability. *Journal of molecular biology*, 274(1), 132–142. <https://doi.org/10.1006/jmbi.1997.1365>
39. Marchand, A., Winther, A. M., Holm, P. J., Olesen, C., Montigny, C., Arnou, B., Champeil, P., Clausen, J. D., Vilsen, B., Andersen, J. P., Nissen, P., Jaxel, C., Møller, J. V., & le Maire, M. (2008). Crystal structure of D351A and P312A

- mutant forms of the mammalian sarcoplasmic reticulum Ca(2+) -ATPase reveals key events in phosphorylation and Ca(2+) release. *The Journal of biological chemistry*, 283(21), 14867–14882. <https://doi.org/10.1074/jbc.M710165200>
40. Roy, I. R., Sutton, C. K., & Berndsen, C. E. (2019). Resilience of BST-2/Tetherin structure to single amino acid substitutions. *PeerJ*, 7, e7043. <https://doi.org/10.7717/peerj.7043>
41. C.A. Laughton C. A. (1994). Prediction of protein side-chain conformations from local three-dimensional homology relationships. *Journal of molecular biology*, 235(3), 1088–1097. <https://doi.org/10.1006/jmbi.1994.1059>

List of Publications

1. **Santi Swarupini, D.**, and Bhuyan, A. K. (2018). Amyloid fibrillation of an intrinsically disordered plant phloem protein AtPP16-1 under acidic condition. *Biophysical chemistry*, 237, 1–8.
<https://doi.org/10.1016/j.bpc.2018.03.004>
2. **Santi Swarupini, D.**, Joshi, K., and Bhuyan, A. K. (2018). Negative Thermal Expansion of an Intrinsically Disordered Protein. (accepted *Chemical Physics*, 2022)

To Be Communicated

1. **Santi Swarupini, D.**, and Bhuyan, A. K. Electrostatic effect on the folding of AtPP16-1
2. **Santi Swarupini, D.**, and Bhuyan, A. K. Cysteine substitution mutagenesis of a plant disordered protein alters the average conformational ensemble throughout the backbone.

Conferences/ Presentations

- Negative Thermal Expansion of an Intrinsically Disordered Protein AtPP16-1.

Oral Presentation: 18th annual in-house symposium CHEMFEST-2021, Held at school of chemistry, University of Hyderabad. (Best oral presentation award).

- NMR structural studies of mutations in AtPP16-1

Poster Presentation: 16th annual in-house symposium CHEMFEST-2019, Held at school of chemistry, University of Hyderabad.

Structure, stability, and dynamics of an intrinsically disordered protein AtPP16-1 and its mutants

by

Didla Santi Swarupini

Submission date: 07-Feb-2022 01:26PM (UTC+0530)

Submission ID: 1756717988

File name: PhD-Thesis-16CHPH05.docx (4.16M)

Word count: 17942

Character count: 97162

Structure, stability, and dynamics of an intrinsically disordered protein AtPP16-1 and its mutants

ORIGINALITY REPORT

28%

SIMILARITY INDEX

1%

INTERNET SOURCES

28%

PUBLICATIONS

0%

STUDENT PAPERS

PRIMARY SOURCES

1

D. Santi Swarupini, Abani K. Bhuyan. "Amyloid fibrillation of an intrinsically disordered plant phloem protein At PP16-1 under acidic condition", Biophysical Chemistry, 2018

Publication

25%

Abani K. Bhuyan

ABANI K. BHUYAN, Ph.D.
Professor
University of Hyderabad
School of Chemistry
Hyderabad - 500 046.

2

Pulikallu Sashi, Kiran K. Singarapu, Abani K. Bhuyan. "Solution NMR Structure and Backbone Dynamics of Partially Disordered Phloem Protein 16-1, a Putative mRNA Transporter", Biochemistry, 2018

Publication

1%

3

Halavath Ramesh, Noorul Huda, Mujahid Hossain, Abani K. Bhuyan. "Food Additives Benzoic Acid and its Methyl and Propyl Parahydroxy Derivatives Aggregate Proteins", ACS Food Science & Technology, 2021

Publication

<1%

4

Systems Biology of Free Radicals and Antioxidants, 2014.

Publication

<1%

5

Pulikallu Sashi, Dasari Ramakrishna, Abani K. Bhuyan. "Dispersion Forces and the Molecular Origin of Internal Friction in Protein", Biochemistry, 2016

Publication

<1 %

6

Pulikallu Sashi, Kiran Kumar Singarapu, Abani K Bhuyan. " Solution NMR Structure and Backbone Dynamics of the Partially Disordered Phloem Protein 16-1, A Putative mRNA Transporter ", Biochemistry, 2018

Publication

<1 %

7

Francois-Xavier Theillet, Andres Binolfi, Tamara Frembgen-Kesner, Karan Hingorani et al. "Physicochemical Properties of Cells and Their Effects on Intrinsically Disordered Proteins (IDPs)", Chemical Reviews, 2014

Publication

<1 %

8

dipot.ulb.ac.be

Internet Source

<1 %

9

www.rcsb.org

Internet Source

<1 %

10

Kirthi Joshi, Abani K. Bhuyan. "Quasi-native transition and self-diffusion of proteins in water-glycerol mixture", Biophysical Chemistry, 2020

Publication

<1 %

Submitted to University of Leeds

11

Student Paper

<1 %

12

research.wsulibs.wsu.edu:8080

Internet Source

<1 %

13

Allyn J. Schoeffler, Allison M. Joubert, Fenggang Peng, Farheen Khan, Chin-Chi Liu, Vince J. LiCata. "Extreme free energy of stabilization of Taq DNA polymerase", Proteins: Structure, Function, and Bioinformatics, 2004

Publication

<1 %

14

arizona.openrepository.com

Internet Source

<1 %

15

hdl.handle.net

Internet Source

<1 %

16

Bikash R. Sahoo, Christopher L. Souders, Takahiro W. Nakayama, Zhou Deng et al. "Conformational Tuning of Amylin by Charged Styrene-maleic-acid Copolymers", Cold Spring Harbor Laboratory, 2021

Publication

<1 %

17

Weontae LEE. "Solution structure of the variant-3 neurotoxin from Centruroides sculpturatus Ewing", European Journal of Biochemistry, 1/1994

Publication

<1 %

Exclude quotes On

Exclude matches

< 14 words

Exclude bibliography On