Writing up your results

BIOL3004 electives

In General

- every paper has a very distinct and clear aim
  - precise the focus of the paper in the title
- every paper is different
  - different proteins have different stories
    - for some the structure is the main focus
    - for some the evolution is more important
    - others concentrate on the function
  - good papers focus in on one topic but also cover all the other areas
  - personal style of author
    - However, rules of scientific writing apply

Step 1

- Discuss in your group
  - what information can you present
    - (methods, results & discussion)
  - how do you put your data into order
    - content and overall structure of your paper
  - what background information is needed
    - introduction
- Often it helps to draw a conceptional map
- Look at papers of related structures: how are they written?

Step 2

- Start a draft of your paper
  - Structured outline (title page, abstract, introduction, methods, results & discussion, conclusion)
- Order of writing is different between people, that's how I do it:
  - Start with introduction (background) and give a short summary (road map) of what is in the paper
  - prepare figures and tables in publication quality and include them into the draft
  - add methods, results and discussion section
    - methods = how did you do it
    - results = what data/information do you present
    - discussion = what does it mean and how do your results fit into the larger picture of knowledge
  - when the body of the paper is finished add
    - abstract = precise outline of key points in your paper
    - and conclusion = summary of key points and potential future work

Step 3

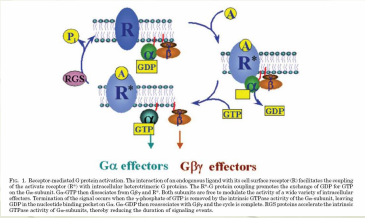
- proof reading
  - give the paper to anyone you can think of
    - your peers, your spouse, your grandmother
  - reader may not understand the topic but s/he can give you valuable feedback on grammar and logic
  - if it's a good paper your grandma will know it

What you have to discuss

Only a tentative pointer. You may have much more material to include.

Background

- if function is known describe context
  - often a figure is better than many words
    - copyright!!!!!!



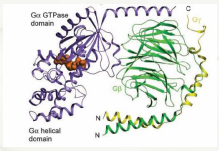
without permission from Endocrine Reviews 24(6):765–781

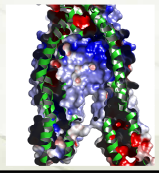
else

- Describe your gene
  - genomic organisation
  - highlight that function is not yet established
- to establish function in discussion
  - microarray experiments
  - est expression profiles
  - localisation data
  - genomic context
- include figures where appropriate
  - but only include figures if they really help to support your argument

Describe your structure

- In methods, give experimental data
  - compare with other papers
- Describe fold, class, structural elements, ligand binding sites, conserved residues, surface properties, &c
- figures of structure (+ details, surface, …)
- Add labels, arrows into figures to highlight features



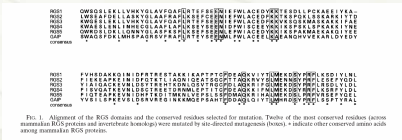


Compare your structure

- to structures with similar folds
  - how similar are they
    - by structure
    - by sequence
    - by function (relative to each other)
  - use figure to illustrate differences/ similarities

Describe sequence features

- conserved residues (or properties)
- non-conserved sequences with similar structures
- quality of alignment
- not the complete alignment is necessary if your protein is very long
  - But prepare a figure of the complete alignment as supplementary information
  - Use clear sequence labels (not gi|8952913814)



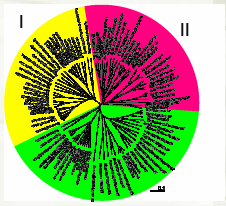
From: Kirk M. Druey, and John H. Kehrl PNAS 1997;94:12851-12856

Function in context of structure

- describe (likely) function
- Functional important features in structure
  - catalytic residues, binding sites, flexible sites, conserved regions
- how well is molecular function supported by cellular function?
  - expression data (=very noisy data!!)
- is the organismal distribution of the gene in agreement with its function, or has function possibly changed over time?

phylogeny

- show best tree possible (exclude sequences if necessary)
- clear labels
- indicate in text the reliability (bootstrapping)
- beautify tree
- in discussion: point out missing phyla/ organisms



Specifics

- Layout
  - *Journal of Molecular Biology* style
    - Title, Authors, Abstract, Introduction, Results, Discussion, Materials and Methods, References
    - use sub-headings to structure text
    - references and citations in JMB style
    - no word or figure limit, but write succinct and concise and only include figures clearly help your argument

Article No. jmbi.1999.3318 available online at http://www.idealibrary.com on IDEAL *J. Mol. Biol.* (1999) **294**, 1271–1285

JMB

40

X-ray Structure Determination of Human Profilin II: A Comparative Structural Analysis of Human Profilins

Ilana M. Nodelman[1], Gregory D. Bowman[1], Uno Lindberg[2] and Clarence E. Schutt[3]*

[1]*Department of Molecular Biology, Henry H. Hoyt Laboratory, Princeton University, Princeton NJ 08544, USA*

[2]*Department of Zoological Cell Biology, WGI, Arrhenius Laboratories for Natural Sciences, Stockholm University S-10691, Stockholm, Sweden*

[3]*Department of Chemistry Henry H. Hoyt Laboratory Princeton University Princeton NJ 08544, USA*

Human profilins are multifunctional, single-domain proteins which directly link the actin microfilament system to a variety of signalling pathways *via* two spatially distinct binding sites. Profilin binds to monomeric actin in a 1:1 complex, catalyzes the exchange of the actin-bound nucleotide and regulates actin filament barbed end assembly. Like SH3 domains, profilin has a surface-exposed aromatic patch which binds to proline-rich peptides. Various multidomain proteins including members of the Ena/VASP and formin families localize profilin:actin complexes through profilin:poly-L-proline interactions to particular cytoskeletal locations (e.g. focal adhesions, cleavage furrows). Humans express a basic (I) and an acidic (II) isoform of profilin which exhibit different affinities for peptides and proteins rich in proline residues. Here, we report the crystallization and X-ray structure determination of human profilin II to 2.2 Å. This structure reveals an aromatic extension of the previously defined poly-L-proline binding site for profilin I. In contrast to serine 29 of profilin I, tyrosine 29 in profilin II is capable of forming an additional stacking interaction and a hydrogen bond with poly-L-proline which may account for the increased affinity of the second isoform for proline-rich peptides. Differential isoform specificity for proline-rich proteins may be attributed to the differences in charged and hydrophobic residues in and proximal to the poly-L-proline binding site. The actin-binding face remains nearly identical with the exception of five amino acid differences. These observations are important for the understanding of the functional and structural differences between these two classes of profilin isoforms.

Introduction

In eukaryotic cells there exists a complex web of signal transduction pathways which function to convey signals from cell surface receptors to the proper intracellular targets. The actin microfilament system is one of the major targets of signalling cascades whose activation is essential for fundamental cellular processes including motility (Stossel, 1993), endo- and exocytosis (Perrin *et al.*, 1992), cytokinesis (Sanger *et al.*, 1989) and determination of cell shape (Small, 1988). Profilin plays a central role by integrating multiple signalling pathways and directly affecting actin filament dynamics. *Via* profilin, actin is linked to the phosphoinositide cycle and to a host of pathways in which specific proline-rich proteins play a role.

Profilin is an essential protein (Verheyen & Cooley, 1994; Witke *et al.*, 1993) which forms a 1:1 complex with monomeric actin. Profilin was first isolated from spleen and was thought to function as a sequestrer of monomeric actin (Carlsson *et al.*, 1977); however, subsequent biochemical studies have shown its role to be more complex. In the presence of capped F-actin barbed ends, profilin acts as a sequesterer and causes the depolymerization of actin filaments (Pollard & Cooper, 1984; Pantaloni & Carlier, 1993). In the absence of filament end cappers, profilin complexed to ATP-actin adds to the barbed ends of growing actin filaments (Pollard & Cooper, 1984; Pring *et al.*, 1992; Pantaloni & Carlier, 1993; Korenbaum *et al.*, 1998). Also, profilin catalyzes actin nucleotide exchange *in vitro*, thereby having the potential to increase the pool of ATP-actin necessary for barbed end assembly (Mockrin & Korn, 1980; Nishida, 1985; Goldschmidt-Clermont *et al.*, 1991b). In agreement with its role as a key regulator of actin dynamics, profilin localizes to regions in the cell undergoing active cytoskeletal remodelling *in vivo* (Buss *et al.*, 1992; Edamatsu *et al.*, 1992; Finkel *et al.*, 1994; Suetsugu *et al.*, 1998; Wills *et al.*, 1999).

Profilin binds to proline-rich stretches in a variety of proteins including members of the Ena/VASP (Reinhard *et al.*, 1995; Gertler *et al.*, 1996) and formin families (Manseau *et al.*, 1996; Watanabe *et al.*, 1997; Evangelista *et al.*, 1997; Chang *et al.*, 1997), drebrin (Mammato *et al.*, 1998), gephyrin (Mammato *et al.*, 1998), N-WASP (Suetsugu *et al.*, 1998), WIP (Ramesh *et al.*, 1997), dynamin I (Witke *et al.*, 1998), and the p85 subunit of PI3-kinase (Singh *et al.*, 1996). Members of the

...

The structure of profilins from human (Metzler *et al.*, 1993; Mahoney *et al.*, 1997; Mahoney *et al.*, 1999), bovine (Schutt *et al.*, 1993; Cedergren-Zeppezauer *et al.*, 1994), *Saccharomyces cerevisiae* (Eads *et al.*, 1998), *Acanthamoeba castellanii* (Vinson *et al.*, 1993; Fedorov *et al.*, 1994*a*), *Arabidopsis thaliana* (Thorn *et al.*, 1997) and birch pollen (Fedorov *et al.*, 1997) have previously been reported. Profilins have a common fold consisting of a central seven-stranded β-sheet flanked by N and C-terminal helices on one side and two short helices on the other side. X-ray structure determination of bovine profilin I in complex with bovine β-actin defined the actin binding site of profilin as residues from helix 3, helix 4, and β-strands 4, 5, and 6. Also, crystal structures of human profilin I bound to poly-L-proline peptides (Mahoney *et al.*, 1997; Mahoney *et al.*, 1999) have confirmed a previously described surface patch of aromatic residues (Trp3, Tyr6, Trp31, His133, Tyr139) as the poly-L-proline binding site (Björkegren *et al.*, 1993; Schutt *et al.*, 1993; Cedergren-Zeppezauer *et al.*, 1994; Thorn *et al.*, 1997). The uncomplexed structures of bovine profilin I and human profilin I are virtually identical with the actin- and poly-L-proline-complexed forms (Cedergren-Zeppezauer *et al.*, 1994; Mahoney *et al.*, 1997). Here, we present the first X-ray crystal structure of this second class of mammalian profilins, human profilin II. This structure reveals an aromatic extension of the poly-L-proline binding site which may explain the increased affinity of profilin II for proline-rich peptides. Comparison with profilin I shows the region surrounding the poly-L-proline site to have a distinct electrostatic and hydrophobic character which may explain differential ligand specificities.

Results

Structure determination

The asymmetric unit of human profilin II consists of four molecules related by approximate 222 non-crystallographic symmetry (NCS) with the central interaction of the tetramer formed by all four C-terminal helices (Figure 1). Monomers related by the NCS 2-fold axis parallel to the crystallographic *c*-axis (AC and BD) interact at β-strands 4 and 5, the turn connecting β-strands 4 and 5, and the C-terminal helix of each molecule. The AC and BD interfaces each bury a total of 1171 Å^2 of solvent-accessible surface area. A second NCS 2-fold axis (roughly parallel with Main-chain C$^\alpha$ atoms of A and B superimpose with an rms deviation of 0.26 Å while C and D C$^\alpha$ atoms superimpose with an rms deviation of 0.09 Å. C$^\alpha$ superposition of A or B onto C or D results in a slightly higher rms deviation of 0.47 Å. The first two amino acids are visible only in chains C and D, and the last two residues are disordered in all four chains. Also, the electron density is poor for residues 92-95 of the loop preceding β-strand 6 of chain B, whereas it is clearly defined for chains C and D. The presence of a second bound PEG 400 molecule and close packing contributes to differences in the local environments of this loop region in chains C and D. A sulfate molecule from the crystallization buffer is bound to each molecule in

Figure 1. Backbone worm representation of the asymmetric unit in human profilin II crystals. The four chains, A, B, C, and D are colored pink, brown, gray, and blue, respectively. A PEG 400 molecule (red and gray sticks) is buried by the C:D interface.

Structural overview of human profilins

Mammals express two isoforms of profilin in which each isoform class, I or II, is conserved between species. Human profilin I and II are 62% identical in primary sequence and share the same fold (Figure 2(a) and (b)). Residues 3 to 136 of human profilin I and II superimpose with an average C^{α} rms deviation of 1.13 Å (C^{α} atoms of chains A, B, C, and D were each superimposed onto crystal structures of human profilin I (PDB accession code: 1fik; 1fil; 1awi)). Slight deviations in the

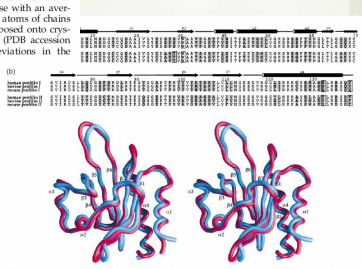


Figure 2. (a) Sequence alignment of mammalian profilins (human profilin I (Kwiatkowski & Bruns, 1988); human profilin II (Honoré *et al.*, 1993); bovine profilin I (Ampe *et al.*, 1988); bovine profilin II (Lambrechts *et al.*, 1995); mouse profilin I (Widada *et al.*, 1989); mouse profilin II (medline accession: AI322548)). Amino acid differences are shaded, and profilin I and putative profilin II poly-L-proline binding residues are boxed. Sequence alignment was generated using AMPS and ALSCRIPT (Barton *et al.*,1990; Barton *et al.*, 1993). (b) Superposition of human profilin I and II. Residues 3 to 136 of 1fik (red) were superimposed onto chain D of profilin II (blue). Secondary structural elements are labelled (α, alpha-helix; β, beta-strand).

Actin binding surface

In the X-ray structure of bovine profilin I complexed to β-actin, approximately 1050 Å^2 of the profilin I surface is excluded from solvent. Many profilin I residues which either directly interact with β-actin or act to position side-chains contacting actin are conserved in the profilin II isoform. Suetsugu *et al.* (1998) demonstrated dramatic loss of affinity for actin as a result of profilin His119 mutation to glutamic acid. In the profilin:β-actin complex, this histidine residue protrudes from the surface of profilin and sits in a pocket largely formed by Tyr133, Tyr169, Met355, and Phe375 of actin. Similar to His119, Phe59 contributes to the stabilization of the profilin:actin complex. Mutation of Phe59 to alanine results in the disruption of the π-stacking interaction with His173 of actin and a 14-fold decrease in affinity for actin (Schlüter *et al.*, 1998). In contrast to His119, the rotomer conformation of Phe59 appears to be stabilized by three

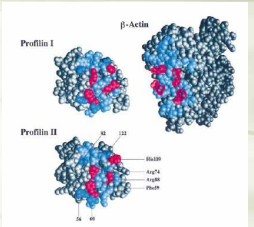
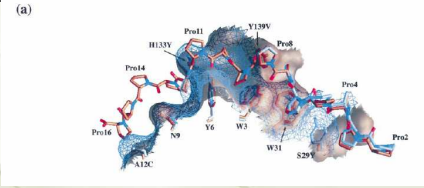


Figure 6. Conserved key residues and non-conserved isoform differences highlighted at the profilin:actin interface. The bovine profilin I:β-actin complex (top; 2btf) has been opened and rotated towards the observer to reveal the protein binding surfaces (light blue). Human profilin II (bottom) was superimposed onto bovine profilin I and translated. Residues colored red (Phe59, Arg74, Arg88 and His119) have been shown to contribute significantly to actin binding. The footprints of these residues are shown mapped onto β-actin as red transparent outlines. The differences between human profilin I and II at the actin-binding surface are colored dark blue (S56E, V60T, E82D, and L122T; first letter corresponds to profilin I, the second to profilin II). Although position 59 differs between human profilin I and II (Y59F) the essential character is conserved.

Poly-L-proline binding site

As determined by site-directed mutagenesis and later confirmed by crystallographic studies, residues in profilin I which directly interact with stretches of poly-L-proline are Trp3, Tyr6, Trp31, His133, and Tyr139 (Björkegren *et al.*, 1993; Mahoney *et al.*, 1997; Mahoney *et al.*, 1999). Superposition of human profilin II onto human profilin I complexed to a pentadeca-prolylpeptide (1cjf) allows for an analysis of the peptide:profilin II interactions (Figure 7(a)). Comparison of human profilin I poly-L-proline binding residues Trp3, Tyr6, Trp31, and His133 with the analogous positions in human profilin II reveals virtually identical rotomer conformations. In human profilin II, amino acid positions 133 and 139 are tyrosine and valine residues, respectively. Analogous to His133 of human profilin I, Tyr133 of human profilin II is



Acidic *versus* basic character of the human profilins

Human profilin I and II are often noted for their basic and acidic isoelectric points, respectively (Honoré *et al.*, 1993). Human profilin II has four acidic residues (Asp48, Glu56, Glu95, Asp138) and two basic residues (Lys28 and Lys68) which are not found in human profilin I. Likewise, human profilin I has four charged residues (Lys25, Lys37, Asp106, His133) which are not conserved in human profilin II. The majority of these non-conserved differences between the two isoforms are located in (Asp138-II, His133-I) or proximal to (Lys25-I, Lys28-II, Asp48-II proline peptide binding sit charge differences suggest contribute to the differentia the profilin isoforms for teins (Figure 8).

The two basic residues but not in profilin I, par

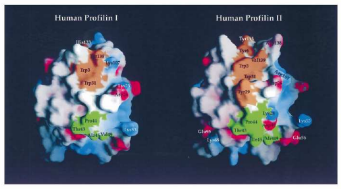


Figure 8. Charge and hydrophobic differences between human profilin I and II surrounding the poly-L-proline binding site. Electrostatic surface potentials of human profilin I (PDB accession code: 1fik) and II (chain D) were generated using GRASP (Nicholls *et al.*, 1991). The brown patch represents side-chains of poly-L-proline binding residues in human profilin I and analogous residues in human profilin II were colored brown and mapped to the surface. Also, the surface area of a flanking hydrophobic region differs between the two human isoforms (only side-chains are colored green and then mapped to the surface). The proteins are displayed in the same orientation. All visible charged residues which differ between the profilin isoforms have been labelled.

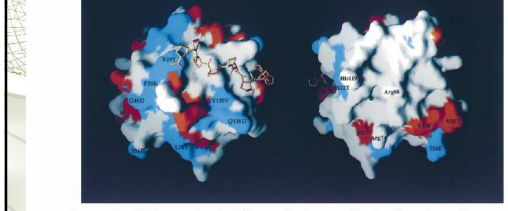


Figure 3. (a) and (b) Amino acid residue differences of the human profilins. A surface rendering of human profilin II (chain D) with colored regions highlighting amino acid residues which differ in human profilin I (blue, non-conserved amino acid changes; brown, conserved amino acid changes). Sites of the more dramatic differences are labelled, e.g. S29Y, with the first letter representing the amino acid in human profilin I followed by the residue number, ending with the amino acid present in human profilin II. Human profilin II was superimposed with human profilin I and bound proline peptide (PDB accession code: 1awi) to orient the poly-L-proline binding groove. Image (b) is rotated 180° about the *y*-axis with respect to (a). Image was generated using the program GRASP (Nicholls *et al.*, 1991).

Discussion

The structure determination of human profilin II has allowed for a comparative analysis of the two human profilin isoforms. Amino acid differences between the isoforms are seen throughout the surface of the protein with the exception of the majority of the actin and poly-L-proline binding faces. These differences contribute to locally unique chemical environments and topologies, which form the structural basis for functional distinctions between profilin I and II.

Mammalian profilin II isoforms display a higher affinity for proline-rich peptides than profilin I (Lambrechts *et al.*, 1997; Ionckheere *et al.*, 1999). of profilin I to phosphate. The general propensity of profilin to accommodate a negatively charged group in this pocket lends support to the idea that these residues participate in phosphoinositide binding.

Overall, profilin isoforms from many different species share the same global biochemical characteristics even with a considerable degree of sequence variation. The two human isoforms have a higher level of sequence similarity, and yet differences on the atomic level reflect the distinct signalling and biochemical pathways specific to the tissues in which each isoform is expressed. The finding that profilin I and profilin II associate with different functional complexes from brain lysates indicates that these isoforms have distinct roles in tissues where they are coexpressed (Witke *et al.*, 1998). Sequence variation between profilin I and II also suggests the ability to bind to proline-rich ligands is differentially regulated by phosphorylation. Profilin I is phosphorylated on Tyr139 in an EGF-stimulated assay which leads to the inability to bind to poly-L-proline (Björkegren-Sjögren *et al.*, 1997), whereas profilin II encodes a valine or phenylalanine residue in position 139. An atomic resolution structure of profilin in complex with a full-length proline-rich protein will be necessary to elucidate the structural determinants not evident in the structural studies of profilin:peptide complexes. The comparative analysis presented here will aid characterization of the profilin:ligand interface with respect to isoform specificity, and may provide insight into the basis for two distinct profilin isoforms.

Materials and Methods

Table 1. Data collection and statistics

A. *Data set*	
Space group	$P2_1$
Cell dimensions (Å; deg.)	$a = 71.43$
	$b = 44.80$
	$c = 89.60$
	$\beta = 98.59$
Molecules per asymmetric unit	4
Solvent content (%)	56
Resolution (Å)	50-2.2
No. of reflections	
Observed	430,124
Unique	28,050
$\langle I/\sigma(I)\rangle$	10.3
Completeness (%)	
Overall	96.9
Highest shell	83.5
R_{merge} [a] (%)	8.9
B. *Refinement statistics*	
Resolution (Å)	20-2.2
R_{cryst} [b] (%)	21.51
R_{free} [c] (%)	26.78
No. of water molecules	398
Average B-factor (Å^2)	
All atoms	23.00
Overall (protein)	17.57
Main-chain	17.40
Side-chain	17.78
Water molecules	27.82
rms deviations from ideality	
Bonds (Å)	0.007
Angles (deg.)	1.34

[a] $R_{merge} = \Sigma|I - \langle I\rangle|/\Sigma I$, where $\langle I\rangle$ is the average of symmetry-equivalent reflections and the summation extends over all observations for all unique reflections.
[b] $R_{cryst} = \Sigma||F_o| - |F_c||/\Sigma|F_o|$, where F_o and F_c are the observed and calculated amplitudes.
[c] R_{free} set contains 5.3 % of total reflections.

Structure determination

Bovine profilin I (PDB accession code 1pne) was employed as a molecular replacement search model, with all non-conserved residues modified to alanine. Molecular replacement solutions for all four molecules in the asymmetric unit were determined using AMoRe (Navaza, 1994) as part of the CCP4 suite of programs (Collaborative Computational Project, 1994). The ASU is a dimer of dimers with roughly 222 NCS symmetry. Cycles of building and refinement were performed using O (Jones *et al.*, 1991) and XPLOR with bulk solvent correction (Brünger, 1993). During the first cycles of refinement NCS restraints grouped all four molecules. In later stages NCS restraints grouped A to B and C to D, and in the final rounds of refinement all NCS restraints were lifted. Solvent molecules were added to stereochemically sensible positions and a sulfate group was placed in identical positions for each of the four chains. Also, two PEG 400 molecules of varying length were modelled as described in the text.

Analysis

Structures of mammalian profilin I used in structural comparisons with human profilin II were: PDB accession code 1fik (human profilin I crystallized in low salt; Fedorov, A.A., Pollard, T.D., Almo, S.C.), 1fil (human profilin I crystallized in high salt; Fedorov, A.A., Pollard, T.D., Almo, S.C.), 1cjf (Mahoney *et al.*, 1999), 1awi (Mahoney *et al.*, 1997), 1pne (Cedergren-Zeppezauer *et al.*, 1994), and 2btf (Schutt *et al.*, 1993). All solvent-accessible surface area calculations were performed using GRASP with a probe radius of 1.4 Å (Nicholls *et al.*, 1991). Protein secondary structure was analyzed using PROMOTIF (Hutchinson & Thornton, 1996) and protein superpositions were performed using O (Jones *et al.*, 1991).

References

Ampe, C., Markey, F., Lindberg, U. & Vandekerckhove, J. (1988). The primary structure of human platelet profilin: reinvestigation of the calf spleen profilin sequence. *FEBS Letters*, **228**, 17-21.

Barton, G. J. (1990). An efficient algorithm to locate all locally optimal alignments between two sequences allowing for gaps. *Comput. Appl. Biosci.* **9**, 729-734.

Barton, G. J. & Sternberg, M. J. (1993). Flexible protein sequence patterns. A sensitive method to detect weak structural similarities. *J. Mol. Biol.* **212**, 389-402.

Björkegren, C., Rozycki, M., Schutt, C. E., Lindberg, U. & Karlsson, R. (1993). Mutagenesis of human profilin locates its poly(L-proline)-binding site to a hydrophobic patch of aromatic amino acids. *FEBS Letters*, **333**, 123-126.

Rewards

- Please indicate on Wiki if you'd like to be considered for scholarship
  - completion of work for publication
  - scholarship includes registration to East Coast Protein Meeting to present your work
- Scholarships have no influence on marks