HHS Public Access

Author manuscript

J Biomol Struct Dyn. Author manuscript; available in PMC 2015 June 03.

Published in final edited form as:

J Biomol Struct Dyn. 2013 October ; 31(10): 1150–1159. doi:10.1080/07391102.2012.726531.

An N-terminal, 830-residue Intrinsically Disordered Region of the Cytoskeleton-regulatory Protein Supervillin Contains Myosin II- and F-actin- Binding Sites

Stanislav O. Fedechkin[a,#], **Jacob Brockerman**[a,#], **Elizabeth J. Luna**[b], **Michail Yu. Lobanov**[c], **Oxana V. Galzitskaya**[c,*], and **Serge L. Smirnov**[a,*]

[a]Department of Chemistry MS-9150, Western Washington University, 516 High Street, Bellingham, WA 98225-9150, USA

[b]Department of Cell and Developmental Biology, University of Massachusetts Medical School, 377 Plantation Street, Worcester, MA 01605, USA

[c]Institute of Protein Research, Russian Academy of Science 4 Institutskaya Street, Pushchino, Moscow Region 142290, Russia

Abstract

Supervillin, the largest member of the villin/gelsolin family, is a cytoskeleton-regulating, peripheral membrane protein. Supervillin increases cell motility and promotes invasive activity in tumors. Major cytoskeletal interactors, including filamentous actin and myosin II, bind within the unique supervillin amino-terminus, amino acids 1–830. The structural features of this key region of the supervillin polypeptide are unknown. Here, we utilize Circular Dichroism (CD) and bioinformatics sequence analysis to demonstrate that the N-terminal part of supervillin forms an extended intrinsically disordered region (IDR). Our combined data indicate that the N-terminus of human and bovine supervillin sequences (positions 1–830) represents an intrinsically disordered region, which is the largest IDR known to date in the villin/gelsolin family. Moreover, this result suggests a potentially novel mechanism of regulation of myosin II and F-actin via the intrinsically disordered N-terminal region of hub protein supervillin.

Introduction

Supervillin (Pestonjamasp, Pope, Wulfkuhle & Luna 1997; Pope et al. 1998) is a large (>1780 residues) eukaryotic protein from the villin/gelsolin superfamiliy of cytoskeleton regulators (Silacci et al. 2004). The protein binds the principal components of the cytoskeleton, filamentous actin and myosin II (Chen et al. 2003) as well as interacting with ~70 other signaling and cytoskeletal proteins (Gangopadhyay et al. 2004; Takizawa et al. 2006; Gangopadhyay et al. 2009; Smith, Fang & Luna 2010) and cholesterol- and

[*]Corresponding authors contact information: OVG : phone: +7.4967.318275; fax: +7.4967.318435; ogalzit@vega.protres.ru; SLS : phone: 1.360.650.2302; fax: 1.360.650. 2826; smirnov@chem.wwu.edu.

[#]These authors contributed equally to the work

A supplementary figure showing the absence of detectable signals specific to a folded protein from fragment M in 1D proton NMR is available from the authors directly and can be downloaded free of charge from the author's server at: http://bioinfo.protres.ru/papers/SVN_IDR_Supplementary_figure1.pdf

polyphosphoinositide-enriched membranes (Nebl et al. 2002). Supervillin regulates cell-substrate adhesion of platelets and tumor cells (Takizawa et al. 2006; Edelstein et al. 2012), localizes at cell-cell contacts (Pestonjamasp et al. 1997), and regulates myosin II-associated functions, including cell motility (Fang et al. 2010), cell spreading (Takizawa, Ikebe, Ikebe & Luna 2007), cytokinesis (Smith et al. 2010), and invadosome function in macrophages and tumor cells (Crowley, Smith, Fang, Takizawa & Luna 2009; Bhuwania et al. 2012). The supervillin C-terminus shares high sequence identity (~25%) with gelsolin and villin (Figure 1), which binds and bundles filamentous actin (F-actin) (Silacci et al. 2004). Supervillin also binds and bundles F-actin but, surprisingly, it utilizes its N-terminus instead of the villin-like C-terminus (Wulfkuhle et al. 1999; Chen et al. 2003). This functional difference between villin and supervillin is due to an amino acid change in the supervillin headpiece, which renders this structure incapable of binding to F-actin (Vardar et al. 2002; Brown, Vardar-Ulu & McKnight 2009). Another difference is that the actin-binding and filament-severing subdomain 1 in villin and gelsolin is replaced in supervillin with a much larger N-terminus (positions 1–830). This 830-residue polypeptide contains numerous cytoskeletal protein interaction sites, including three for F-actin and one for the subfragment 2 (S2) neck region of nonmuscle and smooth muscle myosin II (Chen et al. 2003; Takizawa et al. 2007). Myosin II binds within supervillin residues 1–174 (fragment M), and F-actin binds within fragments 172–342 (A1), 343–570 (A2) and 570–830 (A3) respectively (Figure 1) (Chen et al. 2003). The supervillin N-terminus is uniquely found in supervillins and their muscle isoforms, called archvillin and smooth muscle archvillin (Oh et al. 2003; Gangopadhyay et al. 2004), and thus may contain novel structural motifs and binding mechanisms. Therefore, it is important to structurally characterize the N-terminal supervillin region and its cytoskeleton-binding fragments. Here, we present CD spectra and bioinformatics analyses to show that this region of supervillin is predominately disordered in the unbound state.

Materials and Methods

Bacterial expression of bovine (*Bos taurus*) supervillin N-terminal fragments M and A1 was performed using the expression vectors and procedures described previously (Chen et al. 2003). In brief, we produced fragments M and A1 as GST-fusions from pGEX-6p-1 expression vectors, purified the proteins with a GST resin (GE Healthcare), and removed the GST tags with PreScission Protease (GE Healthcare). Fragment A1 was further purified using a Superdex™ 75 10/300 size-exclusion GL column (GE Healthcare). The same procedures applied to fragments A2 and A3 resulted in yields too low for CD data collection.

CD data were recorded with an AVIV 62 DS spectrometer (Boston University School of Medicine, Boston, MA) and a Jasco J-815 spectrophotometer (Fred Hutchinson Cancer Research Center, Seattle, WA), each with a 1 mm path-length cuvette.

Analysis of bovine and human (*Homo sapiens*) supervillin sequence was performed with IsUnstruct software (Lobanov & Galzitskaya 2011) based on the Ising model (Ising 1925). The server is available at the site: http://bioinfo.protres.ru/IsUnstruct (Lobanov, Sokolovskiy & Galzitskaya 2012, in press). PONDR-FIT and IUPred predictions for intrinsically disordered sequences were performed utilizing the corresponding public web-servers (http://

www.disprot.org/pondr-fit.php; http://iupred.enzim.hu) using default settings (Dosztanyi, Csizmok, Tompa & Simon 2005; Xue, Dunbrack, Williams, Dunker & Uversky 2010). We also searched for short molecular recognition features using the MoRFpred server at http://biomine-ws.ece.ualberta.ca/MoRFpred/index.html (Disfani et al. 2012). Potential ligand-binding sites were predicted with the ANCHOR program (http://anchor.enzim.hu) (Dosztanyi, Meszaros & Simon 2009). The compilation of phosphorylation sites in human supervillin was obtained from PhosphoSitePlus (http://www.phosphosite.org) (Hornbeck et al. 2012). The theoretical pI values were calculated with PROTPARAM server (Gasteiger E. 2005). Multiple sequences were aligned with the ClustalW server (Larkin et al. 2007; Goujon et al. 2010).

Analysis of the relative intensities of SDS-PAGE bands were performed with ImageJ software (Schneider, Rasband & Eliceiri 2012). CD spectral deconvolution was performed with the K2d server (http://www.embl.de/~andrade/k2d).

Results

Amino Acid Sequence

The N-terminal 830 residues of supervillin possess a low fraction of aromatic and non-polar residues and a high percentage of charged amino acids. In particular, this sequence has a low percentage of aromatic residues, Trp, Tyr, and Phe (≤4.3%), which is less than half the percentage of aromatics in the supervillin C-terminus (positions 831 to the end of the sequence) and in villin and gelsolin (≥8.7%) (Table 1). Furthermore, the incidence of charged side chains within residues 1–830 is also somewhat higher in the N-terminus of supervillin than in its C-terminus or in villin and gelsolin. These hydrophobicity-charge features of the supervillin N-terminus are typical for intrinsically disordered proteins (IDP), which lack the capacity to fold in an aqueous solvent (Uversky 2002; Uversky 2011). The Fisher T-test analysis (http://www.graphpad.com/quickcalcs/contingency1.cfm) of the charged vs. aromatics content (Table 1) of the supervillin *N*- and *C*- termini gives the two-tail P-value of <0.0001 indicating a statistically significant correlation. Performing the same analysis with aromatics content vs. size of the termini (Table 1) results in the two-tail P-value of 0.0016, which also is highly statistically significant.

Circular Dichroism (CD) spectra and NMR data

Unstructured polypeptides are difficult to produce in large quantities via bacterial expression as they are subject to active proteolysis, which compounds the challenges of their investigation. In our studies, we observed low yields and high susceptibility to proteolytic and spontaneous degradation of supervillin N-terminal fragments M, A1, A2 and A3 (Figure 1) (Chen et al. 2003), as is typical for an IDR. Bovine supervillin fragments M and A1 were produced in sufficient quantities for CD recording. The yield and purity of A2 and A3 fragments were insufficient for CD data collection.

The far-UV CD spectra for fragment M (Figure 2A) shows a single minimum at or below 200 nm and a small positive maximum above 240 nm. In addition, this spectrum lacks pronounced local minima specific to α-helix (~222 and ~208 nm) or pleated β-sheet (~218

nm) indicating the low percentages of these secondary structural elements. These features of the CD spectrum of fragment M suggest that the polypeptide adopts a predominantly random coil conformation in solution. A minor presence of helical elements within fragment M is suggested by a subtle, plateau-like feature of its CD spectrum at ~222 nm (Figure 2A) and can be attributed to < 20 residues within sequence M identified by the IsUnstruct/PONDR-FIT analysis as ordered (disorder score < 0.5, Figure 3).

The CD spectrum of bovine supervillin fragment A1 with two contaminants of higher molecular weight is shown on Figure 2B. Deconvolution of this spectrum with K2d CD spectrum simulation software (http://www.embl.de/~andrade/k2d) indicated the presence of ~60–65% of random coil and ~35% of α−/β− secondary structure elements. The A1 band (red arrow, Figure 2B) has a relative intensity approximately equal to that of the two high molecular weight impurity bands combined (black arrows, Figure 2B) as estimated with ImageJ software (Schneider et al. 2012). It was difficult to purify A1 from these two proteins. However, the impurities (Figure 2C, black arrows) were purified from mostly degraded A1. The CD spectrum of the two contaminating proteins combined (Figure 2C) is characteristic of a mixture of α-helix and β-sheet elements without a significant amount of random coil structure. This suggests that most of the α−/β− elements observed in the CD spectrum of A1 plus impurities (Figure 2B) are contributed by the two impurity bands, with A1 responsible for most of the random-coil signal.

The solution ^{1}H NMR spectrum of fragment M, although of low signal-to-noise level, is consistent with an unfolded polypeptide with no detectable resonances upfield of 0.8 ppm and downfield of 8.7 ppm (Supplementary Figure 1).

Order/disorder analysis

For reliable prediction of IDR sequences, computational methods employing different principles need to be combined (Ferron, Longhi, Canard & Karlin 2006). We analyzed human and bovine supervillin sequences using the stand-alone software IsUnstruct (Lobanov & Galzitskaya 2011) based on the Ising model of statistical physics (Ising 1925), as well as meta servers PONDR-FIT (Xue et al. 2010) and IUPred (Dosztanyi et al. 2005), which combine the results of multiple disorder predictors. Notably, all three prediction programs report very similar order/disorder trends over the entire sequence. The human and bovine graphs produced by these algorithms are very similar as these proteins share 78% of sequence identity within their N-terminal residues (position 1–830) and 95% within their C-termini (831–1792). All algorithms indicate that the N-terminal half of supervillin, which includes fragments M, A1, A2, and A3, is intrinsically disordered (Figure 3). Supervillin C-terminal sequences are highly homologous to regions responsible for generation of folding elements in the actin-binding proteins gelsolin and villin (Burtnick et al. 1997; Markus, Matsudaira & Wagner 1997; Vardar, Buckley, Frank & McKnight 1999; Smirnov, Isern, Jiang, Hoyt & McKnight 2007). However, gelsolin/villin residues involved in binding to actin are replaced in supervillin by highly conserved alterations in surface loops and by the replacement of the actin filament-severing domain 1 in gelsolin and villin with the unique supervillin N-terminus (Smith et al. 2010). The C-terminal headpiece domain of supervillin was shown to have a folded solution structure (Brown et al. 2009), and the supervillin C-

termini are classified by IsUnstruct and PONDR-FIT as decidedly ordered (Figures 3A and 3B) for both human and bovine polypeptides. The two programs indicate that the biggest difference between human and bovine supervillins lies at positions 550–600; these residues are unfolded for human and somewhat folded for bovine sequences (Figures 3A and 3B). Positions ~430–470, which are in the center of the region required for targeting to focal adhesions (Takizawa et al. 2006), show a local disagreement in the order/disorder prediction by the two algorithms; IsUnstruct reports the locus as disordered, whereas PONDR-FIT indicates a certain order. Another example of a minor discrepancy between IsUnstruct and PONDR-FIT analysis is seen for residues 950–1000. However, this region likely forms an interface between the disordered N-terminus and ordered C-terminus and thus can be analyzed separately.

The distribution of phosphorylation sites along the length of supervillin is consistent with intrinsic disorder in the N-terminal ~1000 amino acids. As shown in Figure 4, pooled mass spectroscopic analyses of proteins in human cells have identified 70 sites of amino acid modification, with all but a few localized within the first 1000 amino acids (http://www.phosphosite.org). Most of these phosphorylation sites also have been documented in mouse and rat supervillin (not shown). The correlation of phosphorylation with localized protein disorder has been attributed both to an increased accessibility of substrate sites and to phosphorylation-induced folding associated with ligand selectivity (Radhakrishnan et al. 1997; Gsponer, Futschik, Teichmann & Babu 2008; Fong, Shoemaker & Panchenko 2012).

Binding Site Predictions

ANCHOR predictions of the locations of regulated ligand-binding sites (Figure 3C) are associated with sequences within bovine supervillin that bind myosin II, F-actin or focal adhesions (Figure 5). The myosin II- and F-actin-binding site predictions were originally made by identifying ~170-amino acid recombinant fragments of the bovine supervillin N-terminus that bound directly to smooth muscle myosin II and actin filaments (Chen et al. 2003; Takizawa et al. 2006; Takizawa et al. 2007). Finer-scale predictions were made by partially proteolyzing these fragments and identifying highly conserved sequences within the largest proteolytic fragment that co-sedimented with actin filaments. The minimal myosin II-binding sequence was estimated by yeast two-hybrid analyses (Chen et al. 2003). More precise mapping will require results from ongoing point mutagenesis experiments. Nevertheless, the current level of resolution is consistent with a model in which myosin II and actin may induce localized folding as they interact with the supervillin N-terminus at or near the ligand-binding sites predicted by ANCHOR. As noted above, the focal adhesion-targeting sequence of supervillin (Takizawa et al. 2006) is also associated with a region containing predicted disordered ligand-binding sites.

Discussion

Proteins that are entirely unstructured or that contain substantial IDRs constitute an impressively large fraction of the eukaryotic proteome (Gsponer et al. 2008). An estimated ~12% of eukaryotic proteins are entirely unfolded and ~43% of eukaryotic proteins contain IDRs of 41 residues or longer (Bogatyreva, Finkelstein & Galzitskaya 2006). Such

disordered fragments often play key roles in the cell (Iakoucheva, Brown, Lawson, Obradovic & Dunker 2002; Ward, Sodhi, McGuffin, Buxton & Jones 2004; Liu et al. 2006). The molecular mechanisms by which IDRs interact with their binding partners are not well characterized at present. To help decipher these mechanisms, structural characterization of IDRs in isolation and in complexes with their binding partners are essential (Lobanov et al. 2010). It is of note here that dedicated Molecular Recognition Features (MORFs) often contain IDRs (Mohan et al. 2006), which may or may not undergo transition to higher structure upon binding to their globular protein targets (Tompa & Fuxreiter 2008).

Our combined CD experimental data and computational analysis of supervillin sequences predict that the N-terminus (positions 1–830) of human/bovine supervillin polypeptide is an IDR. This fragment is much larger than a 40-residue IDR reported within villin (Smirnov et al. 2007) and a 315-residue IDR in dematin (Chen, Jiang, Khan, Chishti & McKnight 2009), which makes the supervillin N-terminus the largest IDR documented for gelsolin/villin family of proteins. Moreover, the Disprot database of disordered proteins (release 5.9), has only 8 entries (out of 653 total) that contain more than 800 residues in their respective IDRs (Sickmeier et al. 2007). This observation further underscores the large size of the N-terminal, 830-residue IDR in human/bovine supervillin. This N-terminal disordered region of bovine supervillin interacts with at least 17 proteins (Chen et al. 2003; Takizawa et al. 2006; Takizawa et al. 2007; Crowley et al. 2009). Many other proteins interact with C-terminal sequences or at currently undefined sites (Nebl et al. 2002; Smith et al. 2010). The fact that about half of supervillin is disordered is consistent with the observation that hub proteins, those interacting with ≥10 partners, are often rich in IDR (Haynes et al. 2006).

Based on the disordered nature of the N-terminus, we can propose certain features of the mechanisms of binding of F-actin and myosin II by supervillin. Its sequence M binds the S2 fragment of nonmuscle and smooth muscle myosin II, whereas fragments A1, A2 and A3 bind F-actin (Figure 1) (Chen et al. 2003). Each larger sequence contains at least two very highly conserved smaller sequences that have been suggested as likely functional interaction sites (Chen et al. 2003). Actin and the S2 fragment of human nonmuscle (and smooth muscle) myosin II are very acidic, with theoretical pI values of 5.5 and 4.3 (4.5) respectively. Moreover, the surface potential of F-actin demonstrates a preponderance of negatively charged patches (Figure 6). The surface potential of the S2 fragment of human nonmuscle myosin II is likely to share this feature by analogy with muscle myosin II fragment S2 (PDB 2FXM, pI 4.8).

Bovine supervillin N-terminus (positions 1–830) is overall basic, with a theoretical pI of 8.6. This observation allows us to propose that electrostatic interactions may be important for supervillin binding to myosin II and F-actin. Other non-covalent forces (London dispersion, H-bonding etc) may also be important components of these interactions, e.g. (Bentivegna & Bresnick 1994). The theoretical pI values for the larger bovine supervillin fragments A1, A2 and A3 are 7.1, 5.7 and 9.5 respectively, with similar figures for the human sequences. These values indicate that only fragment A3 carries an overall positive charge at neutral pH. However, as shown in Figure 5, most of the shorter, highly conserved sequences within the larger fragments are more basic. The identities of these subfragments were suggested by the observation that only some truncated fragments of A1, A2 and A3 co-sediment with F-actin

(Chen et al. 2003). Based on the approximate lengths of the actin-binding fragments, Chen et al. proposed that bovine supervillin N-terminal sequences ~280–342, ~344–422 and ~700–830 within fragments A1, A2 and A3 respectively, contain the F-actin binding sites. Here we will name these proposed *a*ctin-*b*inding subfragments as AB1, AB2 and AB3 respectively (Figure 7). The theoretical pI values for these subfragments are as follows: AB1 (9.8), AB2 (8.4), AB3 (10.2) (Figure 7). Therefore, at neutral pH, the basic supervillin N-terminal sequences AB1, AB2 and AB3 are likely to have strong electrostatic attractions for the acidic actin filaments. The isoelectric point values of the respective N-terminal human supervillin subfragments AB1, AB2 and AB3 are similar: 9.4, 8.1 and 10.4 respectively. Therefore, we propose electrostatics as a primary basis for the interactions between F-actin and supervillin. This proposal differs most significantly from that made previously (Chen et al. 2003) by predicting that residues in AB1 that flank the highly conserved core of amino acids 291–319 within residues 291–342 will be required for binding to F-actin. It is remarkable that the F-actin binding sequences AB1 and AB2 are adjacent, with no sizable, acidic linker in between. This observation suggests that, in the cell, AB1 and AB2 in certain circumstances may act together and form a single F-actin-binding site AB1−AB2 of ~140 residues (pI 9.6).

We also propose that electrostatic forces are important for the binding of supervillin fragment M to the myosin II S2 domain. This proposal is consistent with the decrease in affinity of bovine supervillin fragment M to myosin II heavy chain and its fragment S2 with increasing levels of salt (Chen et al. 2003). Based on the homology of all currently available complete supervillin sequences (human, cow, mouse, frog and zebrafish), we suggest that two N-terminal subfragments within fragment M (positions 1–14 and 99–119) constitute the core of the myosin II -binding elements. We refer to these potential *m*yosin II-*b*inding supervillin subfragments as MB1 (1–14) and MB2 (99–119). Their sequences are 100% conserved in human, cow and mouse (>95% conservation in all the five species) and have pI values of 11.6 and 11.0 respectively (Figure 5), which would be very conductive for electrostatic interactions with the acidic coiled-coil of the myosin II S2 fragment. This is in reasonable agreement with the predicted functionality of bovine supervillin sequences 1 – 60 and 93 – 136, based on yeast two-hybrid and sedimentation binding analyses (Chen et al. 2003).

The human/bovine sequence identity within fragments MB1, MB2, AB1, AB2 and AB3 is 100%, 100%, 84%, 91% and 84% respectively, which is noticeably higher than 74%, the maximal value for the N-terminal regions that separate most of the basic regions. The exception is the sequence between MB1 and MB2 (90% sequence identity), which may be involved in binding another ligand, such as the N-terminus of myosin light chain kinase (Takizawa et al. 2007). The higher sequence conservation of the presumptive functional sites supports the hypothesis that the basic regions are the cores of the myosin II and F-actin-binding sites within the N-terminus of supervillin. Similar “pI vs. sequence identity” trends are observed within all examined vertebrate sequences, including human, bovine, mouse, zebrafish and xenopus supervillins.

To decipher the interface between F-actin and supervillin binding fragments AB1, AB2 and AB3, crystallization of the respective complexes may be attempted. However, given the

disordered nature of the supervillin fragments, crystals may not develop readily. Solution NMR investigation may also be difficult due to size limitations imposed by the bulky actin filaments. These experimental difficulties are quite likely for many complexes involving IDRs. Since sizable disordered fragments occur in a great fraction of eukaryotic proteins (Bogatyreva et al. 2006; Gsponer et al. 2008), a general approach allowing determination of interfaces with their respective binding partners is needed. We envision that computational modeling of complexes containing extended IDRs may offer groundbreaking advantages, especially after the emergence of dedicated, *IDR-focused, induced-fit docking* algorithms. The disordered subfragments of supervillin complexed with their target F-actin or myosin II elements may represent very suitable systems to advance such structural modeling of complexes with IDRs.

Conclusion

Our CD data indicate that the first 342 residues of supervillin form an intrinsically disordered region. The complementary bioinformatics analysis suggests that the entire N-terminus of supervillin likely represents a large IDR. Sequence analysis of this IDR in the context of the previously reported binding studies indicates that electrostatic forces most likely are important for the interactions of supervillin with F-actin and myosin II. We propose that the N-terminus of human/bovine supervillin contains two binding elements for myosin II (basic, conserved stretches of residues 1–14 and 99–119). We also show that the three larger F-actin binding sites proposed previously [AB1 (280–342), AB2 (343–422) and AB3 (700–830)] are basic, in line with the suggestion that electrostatic forces are significant for interactions of supervillin with acidic F-actin surfaces. Our data and hypothesis may help to guide future investigations of supervillin-binding interfaces and mechanisms of action in many of supervillin's roles including promoting the invadopodial activity of tumor cells.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

SF acknowledges RSP funding support from Western Washington University in 2011. For OVG and MYuL, the support came in part from the Russian Foundation for Basic Research (grant No. 11-04-00763), Russian Academy of Sciences (programs "Molecular and Cell Biology" (01200959110) and "Fundamental Sciences to Medicine"). SLS acknowledges the support from Murdock Charitable Trust, and EJL acknowledges support from the National Institutes of Health (GM033048) and the Department of Cell and Developmental Biology, UMASS Medical School. Useful feedback from Tara C. Smith (UMASS Medical School, Worcester, MA, USA) is thankfully acknowledged. The authors are grateful to Dr. C. James McKnight (Boston University Medical School, Boston, MA, USA) for assistance with initiating the project and fruitful discussions. We also express gratitude to Dr. Olga Gursky (Boston University Medical School, Boston, MA, USA) as well as Dr. Roland K. Strong and Tim Vanden Bos (Fred Hutchinson Cancer Research Center, Seattle, WA) for their help with CD recordings of our samples. Dr. Olga Platonova (STRUBI, Oxford, Great Britain), David Gruber (Western Washington University, Bellingham, WA, USA) and Kristen Brewster (Mt. Hood Community College, Gresham, OR, USA) are acknowledged for their assistance in expression and purification of the protein samples.

Abbreviations

IDR	intrinsically disordered region

IDP	intrinsically disordered protein; Hub protein
F-actin	filamentous actin
CD	circular dichroism

References

Bentivegna SS, Bresnick E. Inhibition of human O6-methylguanine-DNA methyltransferase by 5-methylcytosine. Cancer Res. 1994; 54:327–329. [PubMed: 8275462]

Bhuwania R, Cornfine S, Fang Z, Kruger M, Luna EJ, Linder S. Supervillin couples myosin-dependent contractility to podosomes and enables their turnover. J Cell Sci. 2012

Bogatyreva NS, Finkelstein AV, Galzitskaya OV. Trend of amino acid composition of proteins of different taxa. J Bioinform Comput Biol. 2006; 4:597–608. [PubMed: 16819805]

Brown JW, Vardar-Ulu D, McKnight CJ. How to arm a supervillin: designing F-actin binding activity into supervillin headpiece. J Mol Biol. 2009; 393:608–618. [PubMed: 19683541]

Burtnick LD, Koepf EK, Grimes J, Jones EY, Stuart DI, McLaughlin PJ, Robinson RC. The crystal structure of plasma gelsolin: implications for actin severing, capping, and nucleation. Cell. 1997; 90:661–670. [PubMed: 9288746]

Chen L, Jiang ZG, Khan AA, Chishti AH, McKnight CJ. Dematin exhibits a natively unfolded core domain and an independently folded headpiece domain. Protein Sci. 2009; 18:629–636. [PubMed: 19241372]

Chen Y, Takizawa N, Crowley JL, Oh SW, Gatto CL, Kambara T, Sato O, Li XD, Ikebe M, Luna EJ. F-actin and myosin II binding domains in supervillin. J Biol Chem. 2003; 278:46094–46106. [PubMed: 12917436]

Crowley JL, Smith TC, Fang Z, Takizawa N, Luna EJ. Supervillin reorganizes the actin cytoskeleton and increases invadopodial efficiency. Mol Biol Cell. 2009; 20:948–962. [PubMed: 19109420]

Disfani FM, Hsu WL, Mizianty MJ, Oldfield CJ, Xue B, Dunker AK, Uversky VN, Kurgan L. MoRFpred, a computational tool for sequence-based prediction and characterization of short disorder-to-order transitioning binding regions in proteins. Bioinformatics. 2012; 28:i75–i83. [PubMed: 22689782]

Dosztanyi Z, Csizmok V, Tompa P, Simon I. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. Bioinformatics. 2005; 21:3433–3434. [PubMed: 15955779]

Dosztanyi Z, Meszaros B, Simon I. ANCHOR: web server for predicting protein binding regions in disordered proteins. Bioinformatics. 2009; 25:2745–2746. [PubMed: 19717576]

Edelstein LC, Luna EJ, Gibson IB, Bray M, Jin Y, Kondkar A, Nagalla S, Hadjout-Rabi N, Smith TC, Covarrubias D, Jones SN, Ahmad F, Stolla M, Kong X, Fang Z, Bergmeier W, Shaw C, Leal SM, Bray PF. Human Genome-Wide Association and Mouse Knockout Approaches Identify Platelet Supervillin as an Inhibitor of Thrombus Formation under Shear Stress. Circulation. 2012

Fang Z, Takizawa N, Wilson KA, Smith TC, Delprato A, Davidson MW, Lambright DG, Luna EJ. The membrane-associated protein, supervillin, accelerates F-actin-dependent rapid integrin recycling and cell motility. Traffic. 2010; 11:782–799. [PubMed: 20331534]

Ferron F, Longhi S, Canard B, Karlin D. A practical overview of protein disorder prediction methods. Proteins. 2006; 65:1–14. [PubMed: 16856179]

Fong JH, Shoemaker BA, Panchenko AR. Intrinsic protein disorder in human pathways. Mol Biosyst. 2012; 8:320–326. [PubMed: 22012032]

Gangopadhyay SS, Kengni E, Appel S, Gallant C, Kim HR, Leavis P, DeGnore J, Morgan KG. Smooth muscle archvillin is an ERK scaffolding protein. J Biol Chem. 2009; 284:17607–17615. [PubMed: 19406750]

Gangopadhyay SS, Takizawa N, Gallant C, Barber AL, Je HD, Smith TC, Luna EJ, Morgan KG. Smooth muscle archvillin: a novel regulator of signaling and contractility in vascular smooth muscle. J Cell Sci. 2004; 117:5043–5057. [PubMed: 15383618]

Gasteiger E HC, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A. Protein Identification and Analysis Tools on the ExPASy Server. The Proteomics Protocols Handbook. 2005:571–607.

Goujon M, McWilliam H, Li W, Valentin F, Squizzato S, Paern J, Lopez R. A new bioinformatics analysis tools framework at EMBL-EBI. Nucleic Acids Res. 2010; 38:W695–699. [PubMed: 20439314]

Gsponer J, Futschik ME, Teichmann SA, Babu MM. Tight regulation of unstructured proteins: from transcript synthesis to protein degradation. Science. 2008; 322:1365–1368. [PubMed: 19039133]

Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis. 1997; 18:2714–2723. [PubMed: 9504803]

Haynes C, Oldfield CJ, Ji F, Klitgord N, Cusick ME, Radivojac P, Uversky VN, Vidal M, Iakoucheva LM. Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes. PLoS Comput Biol. 2006; 2:e100. [PubMed: 16884331]

Hornbeck PV, Kornhauser JM, Tkachev S, Zhang B, Skrzypek E, Murray B, Latham V, Sullivan M. PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. Nucleic Acids Res. 2012; 40:D261–270. [PubMed: 22135298]

Iakoucheva LM, Brown CJ, Lawson JD, Obradovic Z, Dunker AK. Intrinsic disorder in cell-signaling and cancer-associated proteins. J Mol Biol. 2002; 323:573–584. [PubMed: 12381310]

Ising E. Beitrag zur Theorie des Ferromagnetismus. Zeitschrift für Physik. 1925; 31:253–258.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. Clustal W and Clustal X version 2.0. Bioinformatics. 2007; 23:2947–2948. [PubMed: 17846036]

Liu J, Perumal NB, Oldfield CJ, Su EW, Uversky VN, Dunker AK. Intrinsic disorder in transcription factors. Biochemistry. 2006; 45:6873–6888. [PubMed: 16734424]

Lobanov MY, Galzitskaya OV. The Ising model for prediction of disordered residues from protein sequence alone. Phys Biol. 2011; 8:035004. [PubMed: 21572175]

Lobanov MY, Shoemaker BA, Garbuzynskiy SO, Fong JH, Panchenko AR, Galzitskaya OV. ComSin: database of protein structures in bound (complex) and unbound (single) states in relation to their intrinsic disorder. Nucleic Acids Res. 2010; 38:D283–287. [PubMed: 19906708]

Lobanov MY, Sokolovskiy IV, Galzitskaya OV. IsUnstruct: Prediction of the Residue Status to Be Ordered or Disordered in the Protein Chain by a method based on the Ising model. J Biomol Struct Dyn. 2012 in press.

Markus MA, Matsudaira P, Wagner G. Refined structure of villin 14T and a detailed comparison with other actin-severing domains. Protein Sci. 1997; 6:1197–1209. [PubMed: 9194180]

Mohan A, Oldfield CJ, Radivojac P, Vacic V, Cortese MS, Dunker AK, Uversky VN. Analysis of molecular recognition features (MoRFs). J Mol Biol. 2006; 362:1043–1059. [PubMed: 16935303]

Nebl T, Pestonjamasp KN, Leszyk JD, Crowley JL, Oh SW, Luna EJ. Proteomic analysis of a detergent-resistant membrane skeleton from neutrophil plasma membranes. J Biol Chem. 2002; 277:43399–43409. [PubMed: 12202484]

Oh SW, Pope RK, Smith KP, Crowley JL, Nebl T, Lawrence JB, Luna EJ. Archvillin, a muscle-specific isoform of supervillin, is an early expressed component of the costameric membrane skeleton. J Cell Sci. 2003; 116:2261–2275. [PubMed: 12711699]

Pestonjamasp KN, Pope RK, Wulfkuhle JD, Luna EJ. Supervillin (p205): A novel membrane-associated, F-actin-binding protein in the villin/gelsolin superfamily. J Cell Biol. 1997; 139:1255–1269. [PubMed: 9382871]

Pope RK, Pestonjamasp KN, Smith KP, Wulfkuhle JD, Strassel CP, Lawrence JB, Luna EJ. Cloning, characterization, and chromosomal localization of human supervillin (SVIL). Genomics. 1998; 52:342–351. [PubMed: 9867483]

Radhakrishnan I, Perez-Alvarado GC, Parker D, Dyson HJ, Montminy MR, Wright PE. Solution structure of the KIX domain of CBP bound to the transactivation domain of CREB: a model for activator:coactivator interactions. Cell. 1997; 91:741–752. [PubMed: 9413984]

Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nature Methods. 2012; 9:671–675. [PubMed: 22930834]

Sickmeier M, Hamilton JA, LeGall T, Vacic V, Cortese MS, Tantos A, Szabo B, Tompa P, Chen J, Uversky VN, Obradovic Z, Dunker AK. DisProt: the Database of Disordered Proteins. Nucleic Acids Res. 2007; 35:D786–793. [PubMed: 17145717]

Silacci P, Mazzolai L, Gauci C, Stergiopulos N, Yin HL, Hayoz D. Gelsolin superfamily proteins: key regulators of cellular functions. Cell Mol Life Sci. 2004; 61:2614–2623. [PubMed: 15526166]

Smirnov SL, Isern NG, Jiang ZG, Hoyt DW, McKnight CJ. The Isolated Sixth Gelsolin Repeat and Headpiece Domain of Villin Bundle F-Actin in the Presence of Calcium and Are Linked by a 40-Residue Unstructured Sequence. Biochemistry. 2007; 46:7488–7496. [PubMed: 17547371]

Smith TC, Fang Z, Luna EJ. Novel interactors and a role for supervillin in early cytokinesis. Cytoskeleton (Hoboken). 2010; 67:346–364. [PubMed: 20309963]

Takizawa N, Ikebe R, Ikebe M, Luna EJ. Supervillin slows cell spreading by facilitating myosin II activation at the cell periphery. J Cell Sci. 2007; 120:3792–3803. [PubMed: 17925381]

Takizawa N, Smith TC, Nebl T, Crowley JL, Palmieri SJ, Lifshitz LM, Ehrhardt AG, Hoffman LM, Beckerle MC, Luna EJ. Supervillin modulation of focal adhesions involving TRIP6/ZRP-1. J Cell Biol. 2006; 174:447–458. [PubMed: 16880273]

Tompa P, Fuxreiter M. Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. Trends Biochem Sci. 2008; 33:2–8. [PubMed: 18054235]

Uversky VN. Natively unfolded proteins: a point where biology waits for physics. Protein Sci. 2002; 11:739–756. [PubMed: 11910019]

Uversky VN. Intrinsically disordered proteins from A to Z. Int J Biochem Cell Biol. 2011; 43:1090–1103. [PubMed: 21501695]

Vardar D, Buckley DA, Frank BS, McKnight CJ. NMR structure of an F-actin-binding “headpiece” motif from villin. J Mol Biol. 1999; 294:1299–1310. [PubMed: 10600386]

Vardar D, Chishti AH, Frank BS, Luna EJ, Noegel AA, Oh SW, Schleicher M, McKnight CJ. Villin-type headpiece domains show a wide range of F-actin-binding affinities. Cell Motil Cytoskeleton. 2002; 52:9–21. [PubMed: 11977079]

Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. J Mol Biol. 2004; 337:635–645. [PubMed: 15019783]

Wulfkuhle JD, Donina IE, Stark NH, Pope RK, Pestonjamasp KN, Niswonger ML, Luna EJ. Domain analysis of supervillin, an F-actin bundling plasma membrane protein with functional nuclear localization signals. J Cell Sci. 1999; 112(Pt 13):2125–2136. [PubMed: 10362542]

Xue B, Dunbrack RL, Williams RW, Dunker AK, Uversky VN. PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. Biochim Biophys Acta. 2010; 1804:996–1010. [PubMed: 20100603]

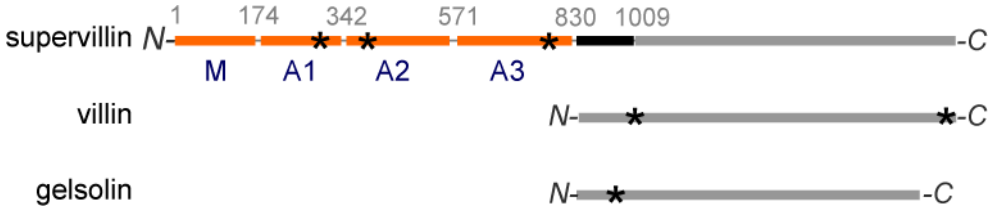

Figure 1.

Sequence alignment of supervillin, villin and gelsolin. N-terminal supervillin fragments are labeled according to their established functions: M (myosin II binding); A1, A2 and A3 (F-actin binding). Their respective sequence positions are indicated above the graphics (Chen et al. 2003). Approximate F-actin binding loci in all the proteins are marked with asterisks (*). The gray elements represent the high sequence homology areas within gelsolin/villin and C-terminus of supervillin. The black element on the supervillin diagram identifies the region lacking homology to gelsolin repeat 1. This fragment (positions 830–1009) includes a strong nuclear localization signal (Wulfkuhle et al. 1999).

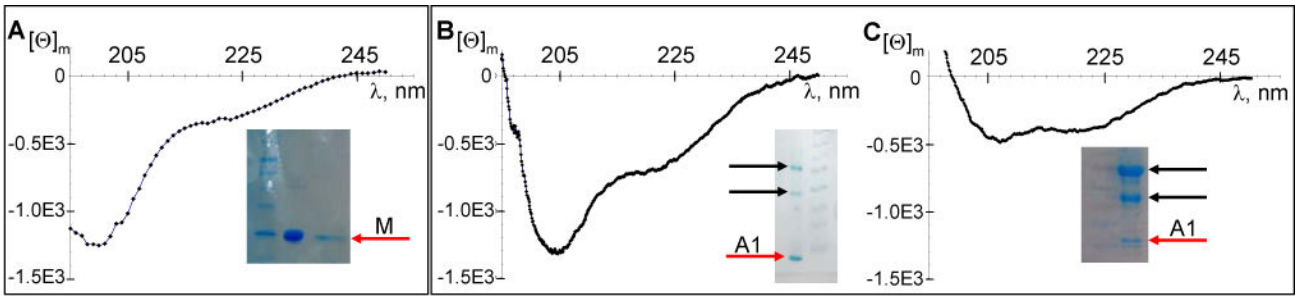

Figure 2.

Representative CD spectra ($[\Theta]_m$, deg•cm^2•dmol^{-1} vs. λ, nm) of bovine supervillin N-terminal fragments M and A1: **A**) Spectrum of fragment M recorded at 25 °C in 20 mM PIPES buffer with 50 mM NaCl, pH 7.5. Red arrow marks the SDS-PAGE band for purified fragment M. **B**) Spectrum of fragment A1 recorded at 23 °C in 10 mM sodium phosphate buffer, pH 7.5. Red arrow marks the major SDS-PAGE band corresponding to fragment A1 and black arrows identify the two proteins (A1-specific impurities) consistently showing up on the gels with A1. **C**) Spectrum of the A1-specific impurities identified with black arrows on SDS-PAGE inserts. The red arrow marks the degraded A1 polypeptide (<10% of total sample, ImageJ (Schneider et al. 2012)). Spectra B) and C) A1 are recorded at 23 °C in 10 mM sodium phosphate buffer at pH 7.0. Concentration of M and A1 samples were ~3–5 μM. The $[\Theta]_m$ values for panels B) and C) were calculated using the length of fragment A1 (171 residue) which precludes the quantitative comparison of the intensity of the CD signal for these samples.

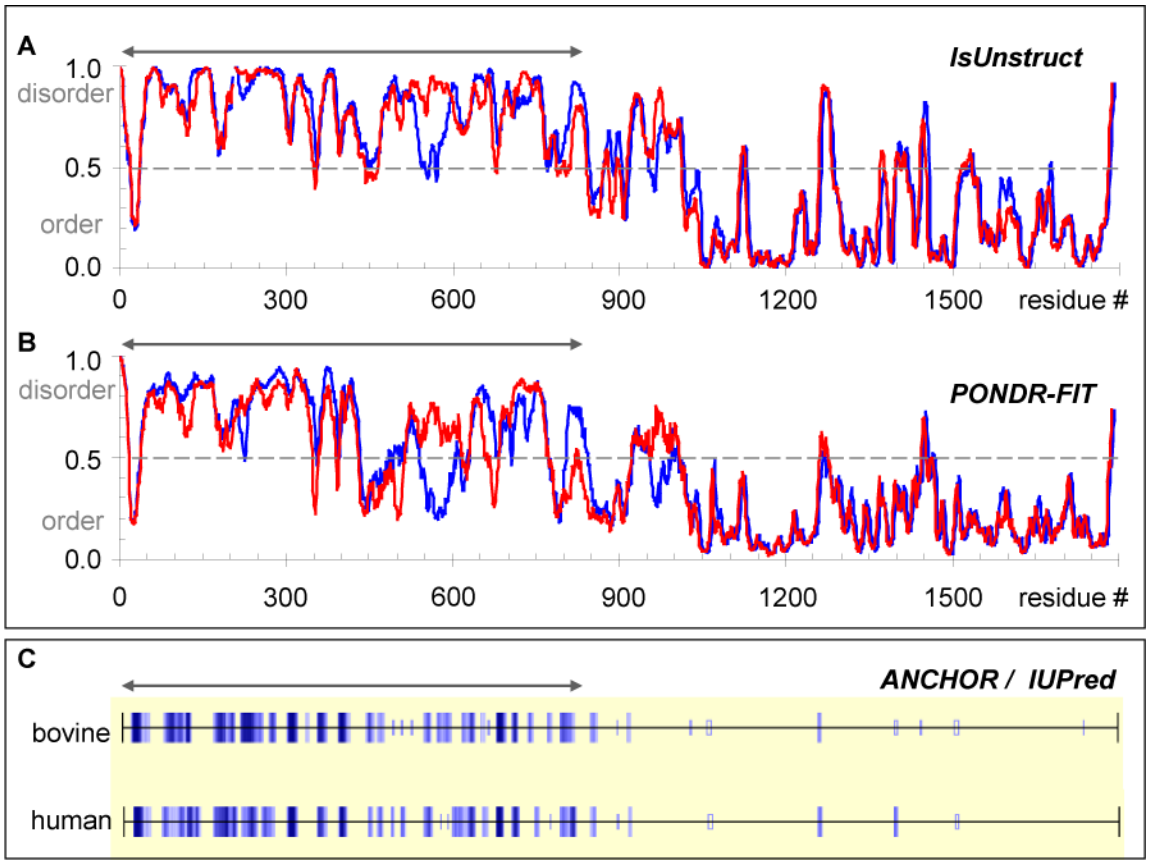

Figure 3.

IsUnstruct (**A**) and PONDR-FIT (**B**) analysis of human (red lines) and bovine (blue lines) supervillin sequences. Both IsUnstruct and PONDR-FIT scores run from 0 to 1 with values below 0.5 generally indicating ordered/folded regions and above 0.5 intrinsically unfolded/disordered regions. The dashed horizontal lines in the middle of the graphs separate the predicted ordered and disordered areas. (**C**) ANCHOR analysis of human supervillin ligand binding sites (blue loci) with IUPred helping to filter out unlikely or folded binding elements. ANCHOR predicts the presence of consistent numerous disordered ligand binding sites throughout the supervillin N-terminus (positions 1–830) with very few sites at the C-terminus (positions 831 onward). The gray double-end arrows on panels A), B) and C) identify the part of supervillin sequences (positions 1–830) investigated in this paper and comprising the polypeptides M, A1, A2 and A3 as shown on Figure 1.

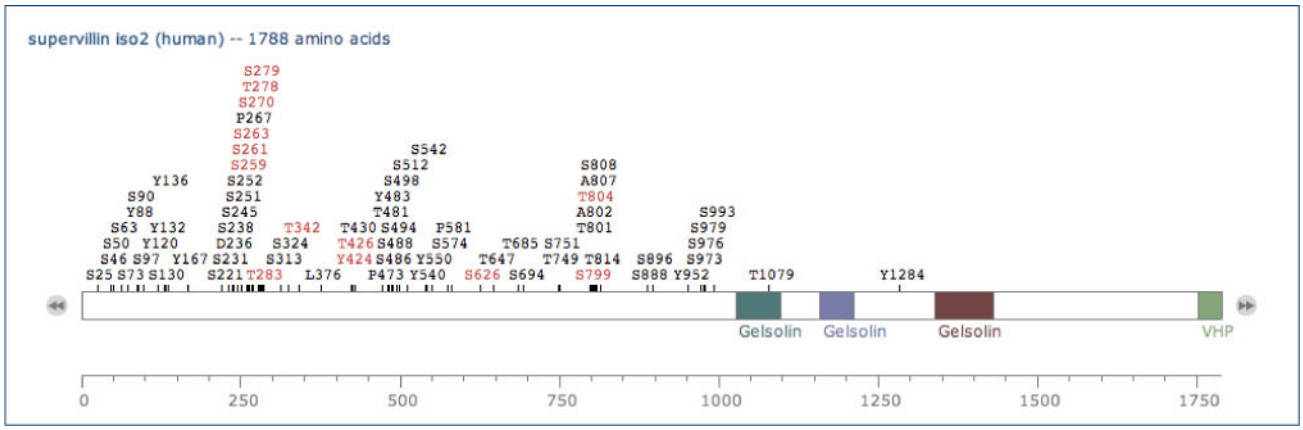

Figure 4.

Phosphorylated amino acids in human supervillin, isoform 2. Phosphorylated residues cluster in the N-terminal ~1000 residues (www.phosphosite.org).

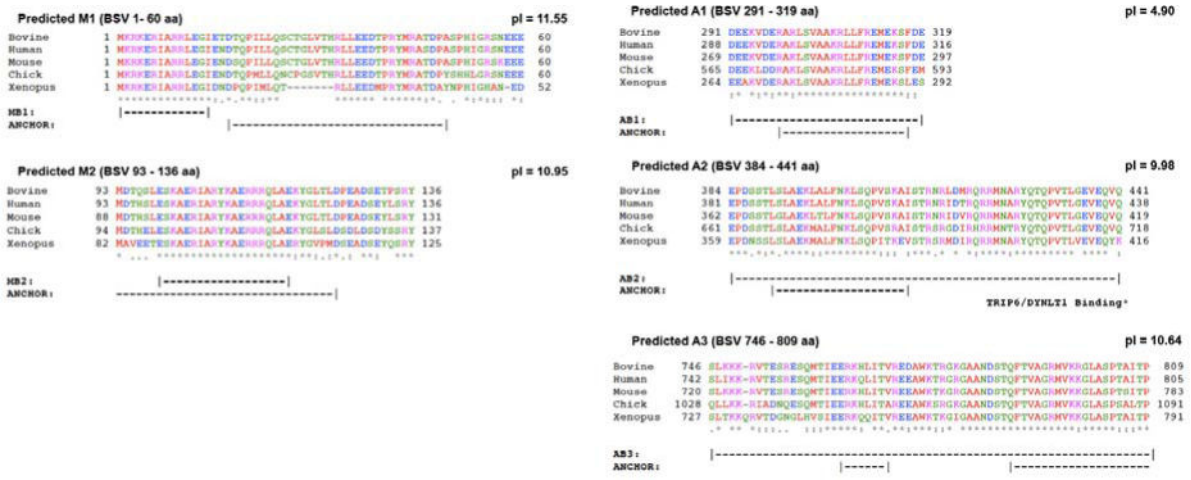

Figure 5.

Predicted myosin II- and F-actin binding sites in bovine supervillin (Chen et al. 2003; Takizawa et al. 2006; Takizawa et al. 2007) include sequences predicted to be ligand-binding sites by the ANCHOR program (Figure 3C). The binding subfragments (MB1, MB2, AB1, AB2 and AB3) were proposed based on their high pI and high sequence conservation as described in Figure 7. Highly conserved regions within the supervillin N-terminal binding sites for myosin II (M1, M2) or F-actin (A1, A2, A3) were aligned using CLUSTALW. Residue numbers are shown for supervillin sequences from *Bos taurus* (Bovine, NP_776615), *Homo sapiens* (Human, NP_003165), *Mus musculus* (Mouse, ADP02396.1), *Gallus gallus* (Chick, XP_418577.3), and *Xenopus (Silurana) tropicalis* (Xenopus, NP_001090765.1). The location of the focal adhesion-targeting site that binds to TRIP6 and DYNLT1 (Takizawa et al. 2006) also is shown.

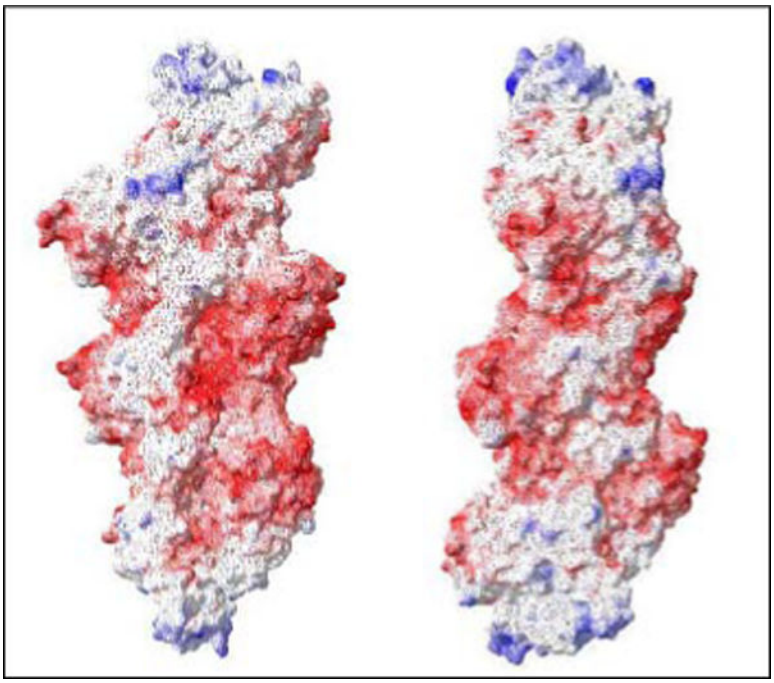

Figure 6.

Electric surface potential of F-actin (PDB 2Y83). The two views were generated by 90° rotation of the structure around its major (vertical) axis. Most of the non-neutral surfaces on F-actin are negative (red) with only minor presence of the positive patches (blue). The images were generated with Swiss-PdbViewer software (Guex & Peitsch 1997).

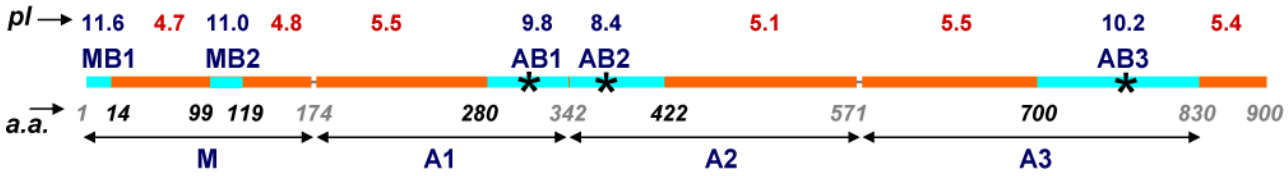

Figure 7.

Bovine supervillin N-terminus and its proposed myosin II- and F-actin- binding subfragments. F-actin binding loci in all the proteins are marked with asterisks (*). Basic F-actin-binding sequences AB1, AB2 and AB3 with their respective pI values are indicated (cyan). Similarly color-coded are the basic N-terminal subfragments MB1 and MB2 within fragment M. The amino acid (*a.a.*) sequence positions and the pI values for the subfragments are indicated below and above the graphics respectively.

Table 1

Comparison of the amino acid composition of human gelsolin, chicken villin, and supervillin (human and bovine) N- and C- termini. The values for the supervillin N-terminus are highlighted.

Protein/Fragment	% Aromatics (W,Y,F)		% Charged (K,R,D,E)		Size (a.a.)
gelsolin, human	8.7%	68 (a.a.)	23.5%	184 (a.a.)	782
villin, chicken	9.4%	78 (a.a.)	24.6%	203 (a.a.)	826
supervillin C-terminus, human	9.1%	87 (a.a.)	26.7%	256 (a.a.)	958
supervillin C-terminus, bovine	8.9%	86 (a.a.)	27.0%	260 (a.a.)	962
supervillin N-terminus, human	4.3%	**36 (a.a.)**	30.5%	**253 (a.a.)**	830
supervillin N-terminus, bovine	4.2%	**35 (a.a.)**	30.0%	**249 (a.a.)**	830