Biophysics and Physicobiology

Note

Computational study of the impact of nucleotide variations on highly conserved proteins: In the case of actin

Ha T. T. Duong[1,2], Hirofumi Suzuki[3], Saki Katagiri[1], Mayu Shibata[1], Misae Arai[1], Kei Yura[1,3,4]

[1] *Graduate School of Humanities and Sciences, Ochanomizu University, Bunkyo, Tokyo 112-8610, Japan*

[2] *Institute of Genome Research, Vietnam Academy of Science and Technology, Cau Giay, Ha Noi 11355, Vietnam*

[3] *Graduate School of Advanced Science and Engineering, Waseda University, Shinjuku, Tokyo 169-8555, Japan*

[4] *Center for Interdisciplinary AI and Data Science, Ochanomizu University, Bunkyo, Tokyo 112-8610, Japan*

   Sequencing of individual human genomes enables studying relationship among nucleotide variations, amino acid substitutions, effect on protein structures and diseases. Many studies have found general tendencies, for instance, that pathogenic variations tend to be found in the buried regions of the protein structures, that benign variations tend to be found on the surface of the proteins, and that variations on evolutionary conserved residues tend to be pathogenic. These tendencies were deduced from globular proteins with standard evolutionary changes in amino acid sequences. In this study, we investigated the variation distribution on actin, one of the highly conserved proteins. Many nucleotide variations and three-dimensional structures of actin have been registered in databases. By combining those data, we found that variations buried inside the protein were rather benign and variations on the surface of the protein were pathogenic. This idiosyncratic distribution of the variation impact is likely ascribed to the extensive use of the surface of the protein for protein-protein interactions in actin.

──────────◀ *Significance* ▶──────────

Distribution of amino acid variation sites on actin, one of the highly conserved proteins, was investigated. Benign variations were found in the buried portion of the protein three-dimensional structure and pathogenic variations were found on the surface of the protein. This distribution is idiosyncratic and is opposite of the ones found in other proteins.

Introduction

   Advancement of genome sequencing technology has realized complete reading of human genome sequence [1]. Along with this achievement, the technology enabled to read individual genome sequences and the individual sequences in different regions of the world are being read [2,3]. The sequencing of different people unveiled the differences in genome sequences among individuals. The number of variations sums up to more than 190,000 cases and the variants cover more than 0.1% of an individual genome [2]. Some of those variations are considered to be related to diseases [4]. This assumption comes with the study of correlation between patients and sequence variations in genes which may be related to the cause of disease. Genome wide association study (GWAS) is one of the extensive studies for the correlation [5], and started to identify correlation between many diseases and nucleotide variations. In parallel to GWAS, a relationship

Corresponding author: Kei Yura, Graduate School of Advanced Science and Engineering, Waseda University, 3-4-1 Okubo, Shinjuku, Tokyo 169-8555, Japan. ORCID iD: https://orcid.org/0000-0003-0197-5478, e-mail: yura.kei@aoni.waseda.jp

between nucleotide variation and inherited disease has been studied and a number of causative variations are now catalogued in OMIM database [6].

The data of nucleotide variation between patients and non-patients of specific diseases make new data source for predicting the disease-causing variation in a gene. A number of attempts to predict the impact of variations on protein coding genes have been conducted [7-12]. It is a natural process to incorporate these pieces of variation information in testing patient specimens to acquire suggestive information of the clinical significance. American College of Medical Genetics and Genomics (ACMG) set a guideline for the interpretation of sequence variation in 2015 [13], which stated that the clinical significance of the variation should be categorized into pathogenic, likely pathogenic, benign, likely benign and uncertain significance. The criteria for the classification include computer prediction on the impact of variations to some extent. The classification results are stored, for example, in ClinVar [14]. Accurate computer prediction of the category of variation is in high demand, due to the fact that the number of the variation of uncertain significance (VUS) is high and once the impact or predisposition of the variation is known, the contribution to the human health is significant.

Computational prediction of the classification of missense variation has been carried out by many groups and they commonly found weak relationship between 1) accessibility of the affected residue by the variation and the pathogenicity, and between 2) conservation of the residue type during the evolution and the pathogenicity [7,9,11,15]. The assumption from the weak relationship was that the disease-causing missense variation tends to reside in the core of the protein structure, in other words the variation has a tendency to destabilize the protein structure, and that the well conserved residues during the evolution should be functionally and/or structurally important, hence the variation of the site has a certain impact on protein function which may lead to the disease. This assumption works to some extent and the prediction of the impact of missense variation has been performed, for example, by Terui et al [11]. However, there still remain a plenty of variations which are annotated as VUS. Especially, VUS remains in a highly conserved genes, because if the assumption holds, all the variations turn out pathogenic, which is not the case.

In this study, therefore, we investigated actin genes, one of the highly conserved gene families and tried to find a way for improvement of the prediction. Actin genes and proteins have been studied in numerous aspects [16-20]. Actin protein is an ATPase enzyme, but it mainly works as a structural protein. Actin protein can be found in muscle contraction, cell motility, cell division, hearing hair cell structure and other cellular processes. Actin family has actin-related proteins and actin-like proteins. Actin-related proteins are anciently diverged from actin and together with actin work in the cytoskeleton formation [21]. Actin-like protein is further distant to actin protein than actin-related protein in sequence identity and may function in cytoskeleton, but the detail remains to be known [22]. According to Ensembl database [23], human genome has seven actin genes (ACTC1, ACTA1, ACTA2, ACTB, ACTBL2, ACTG1, and ACTG2) coding for actin α, α1, α2, β, β-like 2, γ1 and γ2, respectively. The sequence identity among human actin proteins is more than 90%. Human and yeast actin proteins share the sequence identity of about 90%, too. Variations in human actin are known to be related to a couple of diseases [6,24]. Variations on ACTC1 and ACTA1 genes which code for cardiac muscle α-actin and skeletal muscle α-actin, respectively, cause different types of myopathies. ACTA2 gene which codes for smooth muscle aortic α-actin was mapped as a causative gene for aortic aneurysm and multisystemic smooth muscle dysfunction syndrome. ACTB codes for β-actin, essential for cytoplasmic function, and its variations cause Baraitser-Winter syndrome. ACTBL2 (β-like 2) was not associated with disease so far. Variations on ACTG1 and ACTG2 genes which code for γ1-actin and γ2-actin, respectively, were associated with deafness and megacystis-microcolon-intestinal hypoperistalsis syndrome. These characteristics above make actin protein the best target to investigate the distribution of variation in well conserved proteins. An additional intriguing aspect was found in ACTG1, which is expressed ubiquitously, but the impact of variations severely appears only in brain and auditory system [19,25].

Methods

Selection of Human Actin Genes and Proteins
Human genome (GRCh38.p13) in Ensembl genome browser [23] was searched for human actin genes by a key word of "actin." The DNA sequences of all the actin genes were retrieved from Ensembl genome browser and translated by an inhouse program. The genome location of the genes was also retrieved from Ensembl genome browser and NCBI database following the link from Ensembl to CCDS ID [26]. Expression level of actin genes in different tissues was retrieved from Gene database in NCBI where the result of gene expression experiments [27] are stored. Protein Data Bank Japan (PDBj) [28] was searched for the three-dimensional structure of actin proteins. The amino acid sequences of actin proteins were used as query sequences at the homology search tool in the PDBj website. Protein-protein interaction sites were extracted from the three-dimensional structure data, when the structure of actin protein was solved with other proteins. The solvent-accessible surface areas of each residue of actin in complex state and in monomer state were calculated by an inhouse program which implemented the calculation method used by Shrake and Rupley [29]. When the difference of the relative accessibilities of the same amino acid residue were greater than 0.003, the residue was annotated as protein-protein

interaction site.

Selection of Variations in Human Actin Genes

ClinVar [14] was used for finding nucleotide variations in human actin genes. An inhouse program was written to extract variations limited to single nucleotide substitutions. ClinVar has annotation of clinical significance on each variation. There are five different types in clinical significance, namely benign, likely-benign, pathogenic, likely-pathogenic and VUS. To make the analysis simple, we grouped benign and likely-benign into one type, and pathogenic and likely-pathogenic in one different type, hence we used three types of categories in this study. In case there is a conflict in clinical significance on a single nucleotide site, which sometimes happened between different genes, we resolved the conflict by choosing the most serious pathogenicity.

To find a correlation between the variation site and three-dimensional structure, the variations were mapped to actin protein sequence and protein three-dimensional structures by ALAdeGAP [30]. The solvent-accessible surface area of the residue with variation was calculated by the inhouse program. The program is executable at http://cib.cf.ocha.ac.jp/bitool/ASA/.

To find a variation hotspot in actin protein, variability of amino acid residue site was defined as a mean of the number of variations in the window size of five along the amino acid sequence. The variabilities of pathogenic variations and VUS were calculated in this study. The hotspot analysis could be performed on protein 3D structure rather than on amino acid sequence. As far as the window size was small, however, the residues in the window range were clustered in three dimensions and the result of the analysis on sequence was expected to be similar to the one on 3D structure. In addition, the analysis of 3D structure had at least two parameters, namely the radius of a hotspot patch and the accessibility threshold for the surface, which made analysis complicated. For those reasons, we employed the sequence-wise analysis for finding hotspot of variation in this study.

Results and Discussion

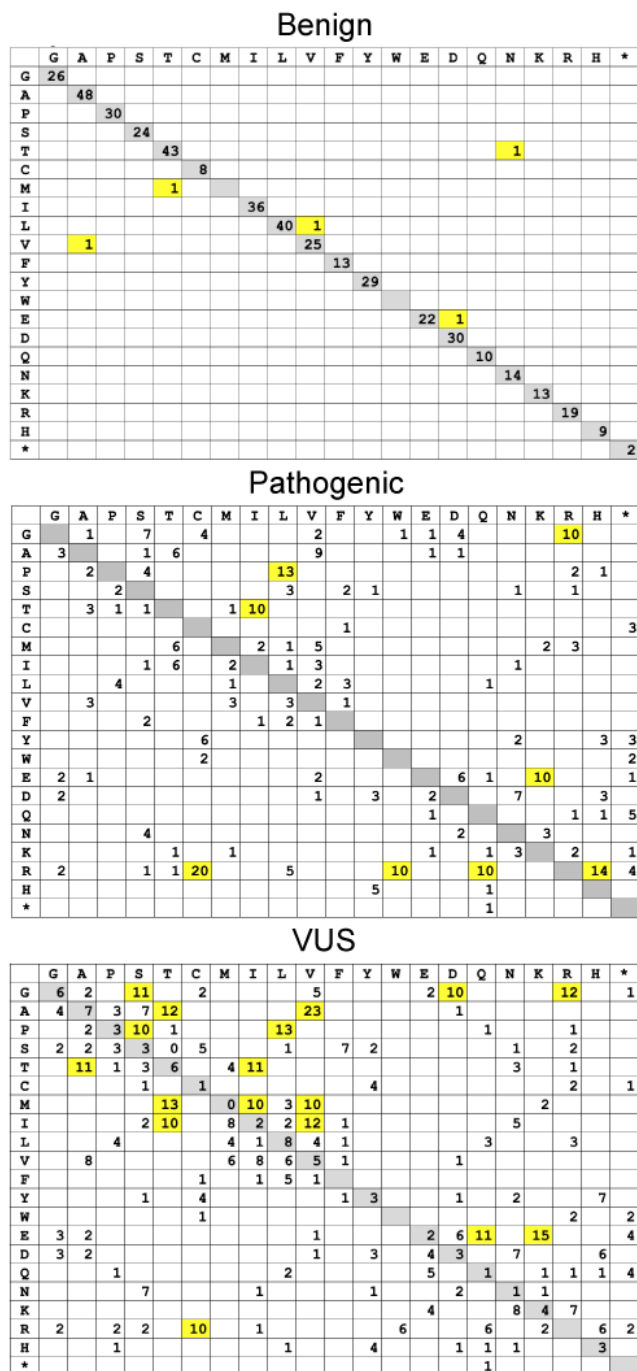### Actin Genes, Variations and Structure Data

We searched the public databases for human actin genes, nucleotide variations, protein three-dimensional (3D) structures, and gene expression. All these data were integrated and were stored in Supplementary Table S1. According to Ensembl database [23], human genome had seven actin genes and seven actin-like genes (Table 1). The length of actin proteins was within the range of 375 and 377, and they were almost the same. Actin-like proteins were longer than actin proteins except actin-like-10 protein. Actin-like proteins had extension on the N-terminal side and insertions in loop regions of actin proteins (Supplementary Table S1).

Table 1 Actin proteins identified in human genome

Protein	Gene	Chromosome	chain	start	end	length	UniProt
Actin alpha 1, skeletal muscle	ACTA1	1	R	229,431,499	229,433,115	377	P68133
Actin alpha 2, smooth muscle	ACTA2	10	R	88,935,223	88,948,930	377	P62736
Actin alpha cardiac muscle 1	ACTC1	15	R	34,790,412	34,794,808	377	P68032
Actin beta	ACTB	7	R	5,527,748	5,529,657	375	P60709
Actin beta like 2	ACTBL2	5	R	57,481,577	57,482,707	376	Q562R1
Actin gamma 1	ACTG1	17	R	81,510,690	81,512,354	375	P63261
Actin gamma 2, smooth muscle	ACTG2	2	F	73,901,312	73,919,575	376	P63267
Actin like 6A	ACTL6A	3	F	179,563,093	179,588,010	429	O96019
Actin like 6B	ACTL6B	7	R	100,643,246	100,656,354	426	O94805
Actin like 7A	ACTL7A	9	F	108,862,323	108,863,630	435	Q9Y615
Actin like 7B	ACTL7B	9	R	108,854,683	108,855,930	415	Q9Y614
Actin like 8	ACTL8	1	F	17,823,009	17,826,519	366	Q9H568
Actin like 9	ACTL9	19	R	8,697,451	8,698,701	416	Q8TC94
Actin like 10	ACTL10	20	F	33,667,498	33,668,235	245	Q5JWF8

Based on the gene locations on the human genome, we identified variations out of ClinVar database [14]. We limited the search of variations to single nucleotide substitutions, which enabled us to conduct one to one correspondence to amino acid sequence. On all 14 actin and actin-like genes, we found 803 variations (Supplementary Table S1) of which only 24 (3%) were from actin-like genes. These variations were classified into three different categories as described in the method section. These data included synonymous, nonsense and missense variations. We translated the variant genes

and deduced the changes in amino acid types (Figure 1). When synonymous variations, which have no impact on amino acid sequence of the protein, were omitted from the count, five cases (0.6%) were in benign category, 303 (37.7%) cases were pathogenic and 495 (61.7%) were VUS. This skewed distribution of clinical significance, especially the small number of benign cases and large number of VUS, is idiosyncratic. For instance, in BRCA1 gene of which variations are related to breast cancer [31], 18.8% cases were benign variations, 31.5% cases were pathogenic and 49.7% were VUS according to ClinVar. The comparison underlined both low number of benign variations and high number of VUS in actin.

Benign

	G	A	P	S	T	C	M	I	L	V	F	Y	W	E	D	Q	N	K	R	H	*
G	26																				
A		48																			
P			30																		
S				24																	
T					43												1				
C						8															
M					1																
I								36													
L									40	1											
V		1								25											
F											13										
Y												29									
W																					
E														22	1						
D															30						
Q																10					
N																	14				
K																		13			
R																			19		
H																				9	
*																					2

Pathogenic

	G	A	P	S	T	C	M	I	L	V	F	Y	W	E	D	Q	N	K	R	H	*
G		1		7		4				2		1	1	4					10		
A	3			1	6					9				1	1						
P		2		4					13										2	1	
S			2						3		2	1					1	1			
T		3	1	1				1	10												
C												1									3
M				6				2	1	5								2	3		
I			1	6			2		1	3							1				
L			4				1	2		3						1					
V		3					3	3	1												
F			2					1	2	1											
Y				6													2		3	3	
W				2																	2
E	2	1								2				6	1		10				1
D	2								1			3		2			7		3		
Q										1						3		1	1		5
N				4											2			3			
K				1		1								1		3		2			1
R	2		1	20	1	5							10			10				14	4
H												5			1						
*															1						

VUS

	G	A	P	S	T	C	M	I	L	V	F	Y	W	E	D	Q	N	K	R	H	*
G	6	2	11			2				5				2	10				12		1
A	4	7	3	7	12					23					1						
P		2	3	10	1				13							1			1		
S	2	2	3	3	0	5			1		7	2					1				2
T		11	1	3		6		4	11								3		1		
C				1		1				4							2				1
M				13	0			10	3	10							2				
I			2		10		8	2	2	12	1							5			
L			4				4	1	8	4	1					3			3		
V		8					6	8	6	5	1				1						
F								1	1	5		1									
Y				1		4					1				1		2			7	
W						1													2		2
E	3	2								1				2	6	11		15			4
D	3	2								1				3	4	3	7			6	
Q			1						2					5		1		1	1	1	4
N				7				1				1			2			1		1	
K														4			8	4	7		
R	2		2	2		10		1					6			6	2			6	2
H			1			1			4						1	1	1		3		
*															1						

Figure 1 The number of amino acid substitutions in actin and actin-like proteins based on the variation data in ClinVar. The vertical axis is the original amino acid types and the horizontal axis is the result of variations. The variation is categorized into three as stated in method section. Each number in a box represents a reported case of amino acid substitutions due to changes in nucleotide sequence in actin gene. Boxes on the diagonal axis of each chart are in grey to emphasize synonymous variations. Boxes in yellow indicate variations in high number in each category.

A number of protein 3D structures of actin have been stored in Protein Databank Japan (PDBj) [28]. Most of the actin entries were either actin fibers (PDB ID: 3J82[32], 3JBK[33], 3LUE[34], 5JLH[35], 6ANU[36], 6CXJ[37], 6G2T[37], 6UK4[38], 6VAO[39], 6VEC[40], 7CCC[41]), or complex with other proteins such as myosin (5JLH[35]), tropomyosin (5JLH[35], 6CXJ[37], 6G2T[37]), myosin-binding proteins (6G2T[37]), cofilin (6UC4[38], 6VAO[39]), gelsolin

(5UBO[42]), spectrin (6ANU[36]), phosphatase (7NZM[43]), acetyltransferase (6NBW[44]), and transcription activator (6LTJ[45]). The data indicated that actin protein had many partners to interact. When variation data were mapped to the 3D structure of actin (6NBW chain A), atypical tendency appeared (Figure 2). In most of the cases reported previously,
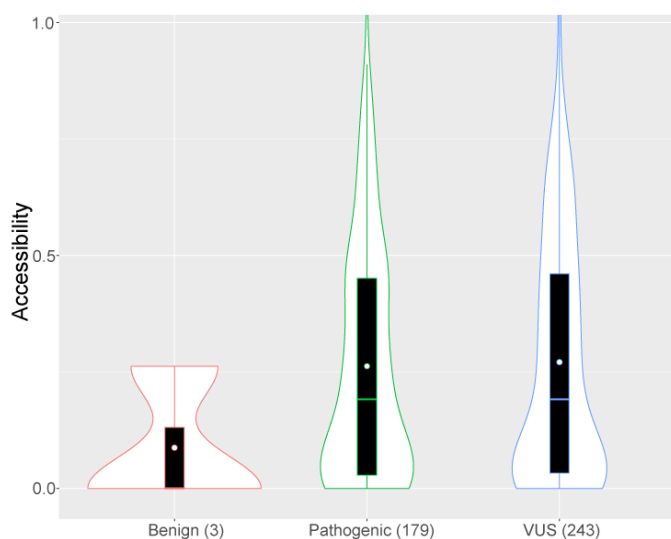
Figure 2 Relationship between variation type and amino acid residue accessibility depicted by violin plot. The number of variations is not equal to the ones in Figure 1, because some of the variations could not be mapped to 3D structure of actin (PDB ID: 6NBW chain A). Red, green and blue shapes represent density of the data depicted by kernel density estimate with normal distribution. A black box represents the range between the first and the third quartiles, a bar in the black box is the median, and a white dot is the mean value. The number in the parentheses is the count in each significance. Note that the number of benign variations is small.

pathogenicity and/or the number of variations were negatively correlated to relative accessible surface area (accessibility) of the residues. Benign variations tended to appear on the surface and pathogenic variations in the core [7,11,15]. In the case of actin, however, the accessibility distribution of benign variation was lower than that of pathogenic variations. In addition, the distribution of pathogenic variation was similar to that of VUS. The locations of five benign variations are shown in Figure 3. Out of five variations, three cases were able to be mapped on the 3D structure of human ACTB, but two cases were not, because they were on a disordered loop. Two out of the three variations were mapped on residues that were completely buried. These two variations were found in ACTC1 and ACTA2, hence there was a possibility that the structure including the two residues was different between ACTB and ACTC1/ACTA2. However, that was highly unlikely, because the sequence identity around the sites was almost 100%. The changes in amino acid residue types from Thr to Asn at 89 and Met to Thr at 176 likely have significant effect due to the structural and chemical differences of amino acid residues, hence being benign of these two residue sites is unexpected in the current knowledge. Other three sites were on the surface of the protein, but were at the site of actin interaction. The change in amino acid types in these cases are minor, namely substitution to a chemically

similar amino acid type, hence being benign is comprehensible. It should be noted that the number of benign variations were only five in the current study. The small number of benign variations precludes us from reaching general characteristics of benign variations in actin proteins.
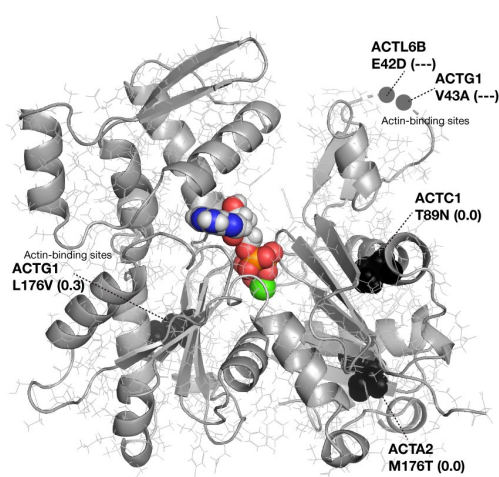
Figure 3 Locations of benign variations in actin 3D structure. The protein is shown by ribbon model with the side chains in line. Three benign variation sites are shown by space filling model in black. Two benign variation sites are located on the disordered loop depicted by dotted line. Each variation is shown with a gene name where the variant was found, residue number sandwiched by original and variant amino acid types, and accessibility of the original residue. A colored space filling model in the center is Mg^{2+}-ATP. The orientation of actin in this figure is named Front throughout this manuscript.

Protein-Protein Interaction Sites in Actin

Wealth of 3D structure data of actin in complex with other proteins in PDBj enabled mapping protein-binding sites on actin protein surface (Figures 4A-C). Actin utilized the surface opposite to the ATP-binding site for fiber formation
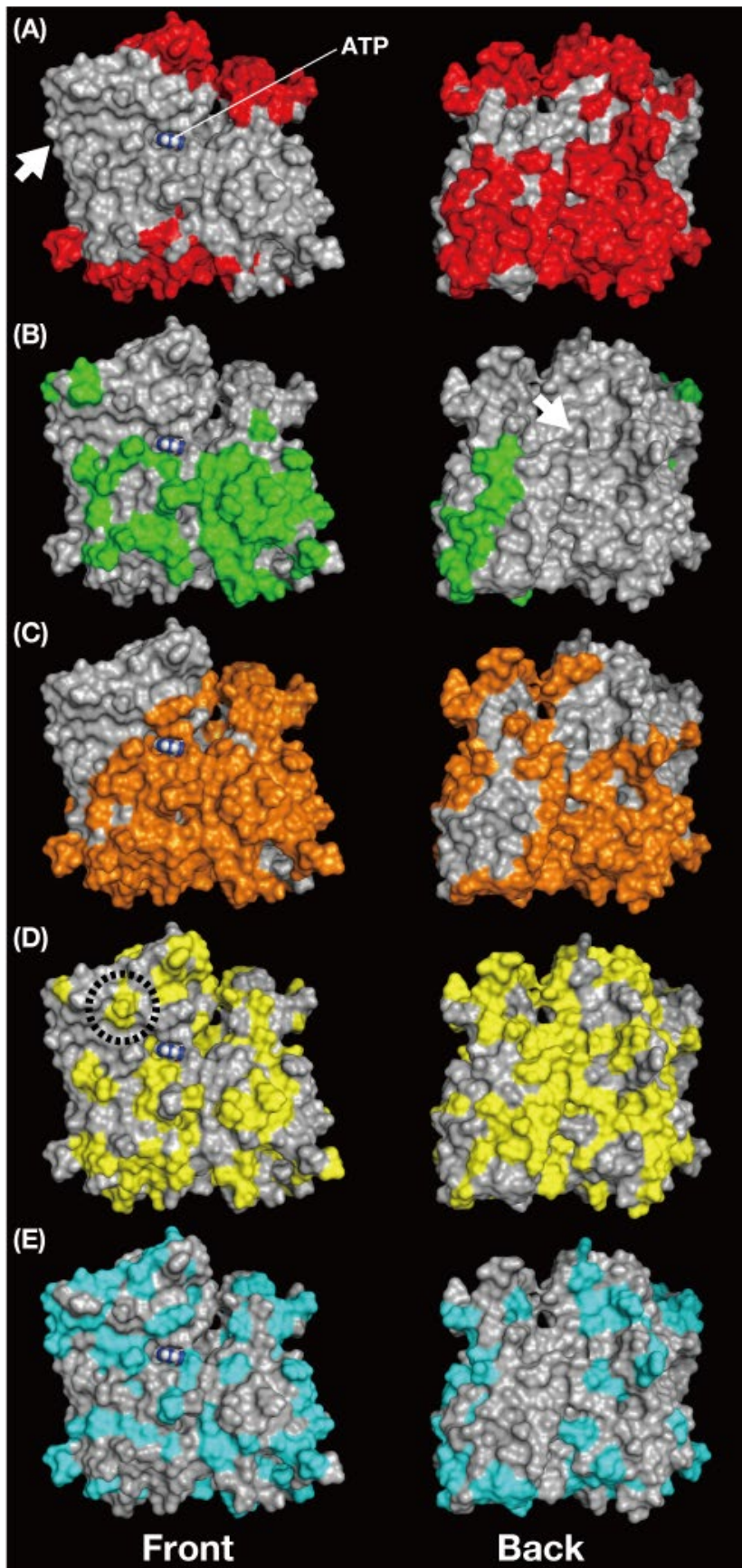
Figure 4 Protein-binding sites and variation sites on the surface of actin. The orientation of the molecule in the left side (Front) is the same orientation as the one in Figure 3. The orientation in the right (Back) is a 180° rotation of the left. A molecule in space filling model is ATP. (A) Actin-binding sites are colored in red on actin surface. The binding site were derived from the following PDB entries; 3J82, 3JBK, 3LUE, 5JLH, 6ANU, 6CXJ, 6G2T, 6LTJ, 6UK4, 6VAO, 6VEC, and 7CCC. (B) Myosin and tropomyosin-binding sites were derived from 5JLH, 6CXJ, and 6G2T. (C) Binding sites of other proteins from 1LOT, 3BYH, 3JBK, 3LUE, 5UBO, 6ANU, 6LTJ, 6NBW, 6UC4, 6VAO, 6VEC, 7CCC, and 7NZM. (D) Pathogenic variation sites. (E) VUS sites. White arrows in (A) and (B) are the surface where no binding sites were assigned. Black dotted circle in (D) is the residue cluster discussed in the text.

basically (Figure 4A). We will call the surface where ATP binds 'Front' and the opposite side 'Back' hereafter. Actin fiber, therefore, had an open Front for other proteins, and Front was actually used for an interface of myosin and tropomyosin (Figure 4B). Interfaces for other proteins including myosin-binding protein, gelsolin, cofilin, spectrin, phosphatase, acetyltransferase and transcription activator had significant overlap with either actin-binding sites or myosin-tropomyosin binding sites (Figure 4C), which clearly suggests a non-simultaneous role of these proteins. Superposition of these protein-binding sites left almost no free surface on actin, indicating that almost all surface of the protein is occupied by functionally important regions. The slight exception is the top left regions in Front view and the center in Back view where no binding sites were mapped (A white arrow in Figures 4A and B). Figure 4D tells pathogenic variation sites which were distributed on the whole surface of actin. The free surface on the Front was free from pathogenic variations, but filled with VUS (Figure 4E) except for a single knob (Arg238 and Arg254) shown by a black dotted circle in Figure 4E. Variation on Arg238 was found in ACTA2 coding for smooth muscle actin which was highly expressed at endometrium/prostate. Variation on Arg254 was found in ACTA1, ACTA2, ACTG1 and ACTL6B, each of which were expressed at heart, endometrium/prostate, ubiquitous and brain, respectively. Other than ACTG1, the proteins were expressed in a specific organ and had clinical impact at those organs. These pieces of evidence suggest that there is an unknown factor that interacts with this knob of actin in those specific organs and the variants may disturb the interactions. In Figure 4E, there were VUS free surface in Front right and in Back left. These surface areas were filled with pathogenic variations and the interface for other proteins. Disruption of interaction with other proteins seems to be critical for the function of actin.
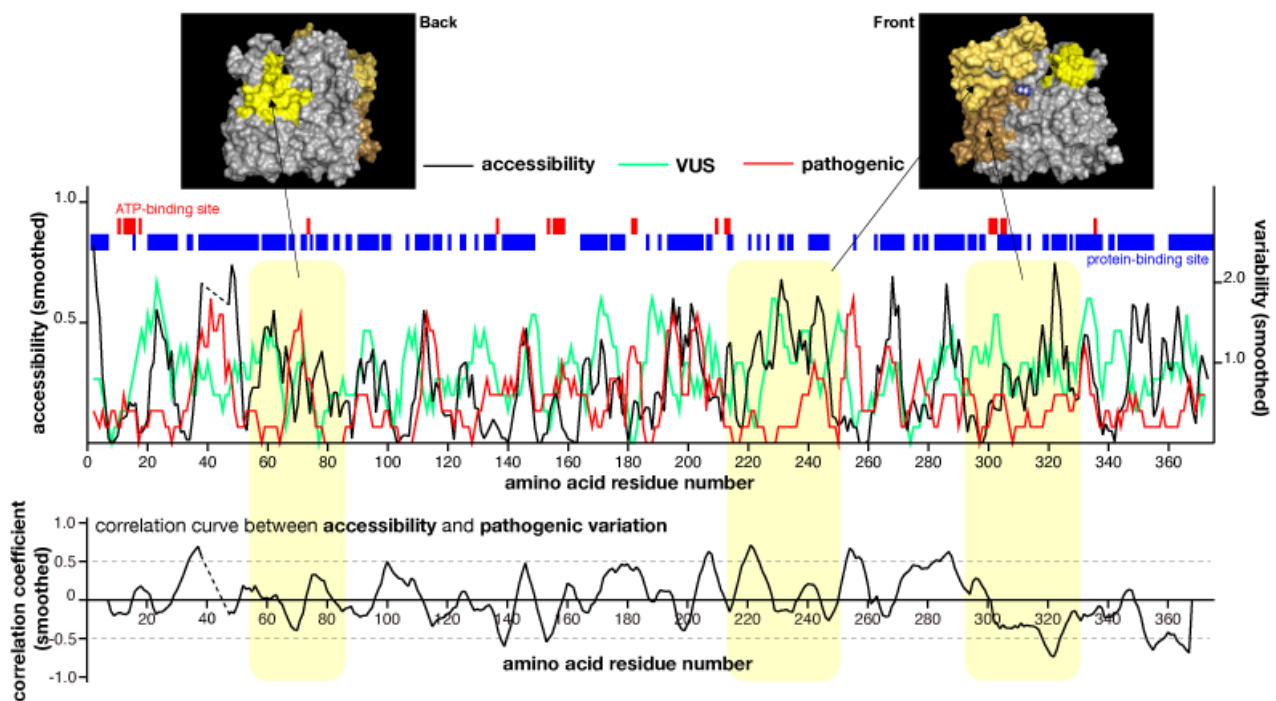
Figure 5 Relationship among amino acid variation, accessibility and binding sites on actin. Horizontal axes of the graphs are residue numbers. Vertical axes on the left are the smoothed accessibility/correlation coefficient given in black in the graphs. Smoothing was carried out by window size three in accessibility and window size seven in correlation coefficient without weight. A vertical axis on the right is the smooth variability given in red and green in the graph. Variability is defined as a mean of the number of variations in the window size of five. Red box indicates ATP-binding site calculated on 6NBW and blue boxes are protein-binding sites give in Figures 4 A-C. Yellow shaded regions with 3D structure are discussed in the text.

Relationship Among Accessibility, Interface and Variations

General scheme of variation-disease relationship tells that low accessibility residues tend to be related with disruption of protein function [46] and diseases [7]. This scheme does not hold in actin as shown in Figure 5. If the general scheme holds, accessibility (black line) and pathogenicity (red line) should have negative correlation. The correlation curve given in the figure shows that the correlation between the two lines is fluctuating between the positive and the negative values.

We focused on three regions shown in yellow highlight in Figure 5. The first yellow region (around residue number 70) has high pathogenicity in relatively low accessibility region and low pathogenicity at high accessibility region. The surface in this region is a binding site of other proteins (Figure 4 A-C), but the region likely plays a less important role in the interactions. The second yellow region (around residue number 230) has a widely accepted tendency except Arg238 and Arg254. The third yellow region (around residue number 310) also tends to have a generally accepted tendency, although the region is a binding site for myosin/tropomyosin and other proteins (Figures 4B and C). Other than these regions and the termini, actin has a tendency of positive correlation between accessibility and pathogenicity. Therefore, the correlation between accessibility and pathogenicity in actin protein was a mixture of positive and negative relationship. Positive relation between pathogenic variation and accessibility can be accounted for by the importance of the surface of actin for protein-protein interactions. Changes in amino acid residue types in the interface likely affect the interactions with other proteins to some extent, which end in pathogenic impact. The negative relation between pathogenic variation and accessibility can be accounted for by the results of previous studies [7,9,11,15]. However, the regions where weak negative relationship was observed were not necessarily interaction-free regions as shown in Figure 5. There are many interaction sites as shown by the blue box. The only interpretation of this negative relationship is that the interactions in these regions play minor roles in interactions and variations has minor impact to the interactions.

Conclusion

A substantial portion of atypical characteristics of variation sites on actin can be accounted for by protein-protein interactions. Actin protein is highly conserved throughout the sequence and it is probably due to the functional importance of the surface of the protein. Variations on surface residues may affect protein-protein interactions and result in pathogenic impact.

Assessing clinical significance of variations in highly conserved proteins like actin has been a difficult task. Conventional computational methods to predict the significance relies on an empirical rule that variations on evolutionary conserved residues tend to have significant impact [9-11]. In this study, we tested the variations on highly conserved actin proteins. The tendencies of pathogenic variation in actin proteins are the following: 1) variations on protein-binding surface tend to be pathogenic, and 2) unknown interactions between actin and other proteins may prevent accurate prediction of pathogenicity. Actin is a multigene family and each gene may have different protein-protein interactions. To improve the assessment of clinical significance of the variations on actin and other highly conserved proteins, rigorous and specific protein interaction data should be incorporated in the prediction scheme.

Conflict of Interest

The authors declares that they have no conflict of interest.

Author Contributions

H.D., and K.Y. designed this study. H.S. gathered the data of protein 3D structure. S.K., M.S. and M.A. gathered the data of sequences and variations. H.D., S.K., M.S. and M.A. performed the research. All the authors contributed to the writing of the manuscript.

Acknowledgements

References

[1] Nurk, S., Koren, S., Rhie, A., Rautiainen, M., Bzikadze, A. V., Mikheenko, A., et al. The complete sequence of a human genome. Science 376, 44-53 (2022). https://doi.org/10.1126/science.abj6987

[2] Yamaguchi-Kabata, Y., Nariai, N., Kawai, Y., Sato, Y., Kojima, K., Tateno, M., et al. iJGVD: An integrative Japanese genome variation database based on whole-genome sequencing. Hum. Genome Var. 2, 15050 (2015). https://doi.org/10.1038/hgv.2015.50

[3] Auton, A., Brooks L. D., Durbin, R. M., Garrison E. P., Kang, H. M., Korbel J. O., et al. A global reference for human genetic variation. Nature 526, 68-74 (2015). https://doi.org/10.1038/nature15393

[4]   Schork, N. J., Murray, S. S., Frazer, K. A., Topol, E. J. Common vs. rare allele hypotheses for complex diseases. Curr. Opin. Genet. Dev. 19, 212-219 (2009). https://doi.org/10.1016/j.gde.2009.04.010

[5]   Cirulli, E. T., Goldstein, D. B. Uncovering the roles of rare variants in common disease through whole-genome sequencing. Nat. Rev. Genet. 11, 415-425 (2010). https://doi.org/10.1038/nrg2779

[6]   Amberger, J. S., Bocchini, C. A., Scott, A. F., Hamosh, A. OMIM.org: Leveraging knowledge across phenotype–gene relationships. Nucleic Acids Res. 47, D1038-D1043 (2018). https://doi.org/10.1093/nar/gky1151

[7]   Wang, Z., Moult, J. SNPs, protein structure, and disease. Hum. Mutat. 17, 263-270 (2001). https://doi.org/10.1002/humu.22

[8]   Ng, P. C., Henikoff, S. Predicting deleterious amino acid substitutions. Genome Res. 11, 863-874 (2001). https://doi.org/10.1101/gr.176601

[9]   Kono, H., Yuasa, T., Nishiue, S., Yura, K. coliSNP database server mapping nsSNPs on protein structures. Nucleic Acids Res. 36, D409-D413 (2007). https://doi.org/10.1093/nar/gkm801

[10] Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., et al. A method and server for predicting damaging missense mutations. Nat. Methods 7, 248-249 (2010). https://doi.org/10.1038/nmeth0410-248

[11] Terui, H., Akagi, K., Kawame, H., Yura, K. CoDP: Predicting the impact of unclassified genetic variants in *MSH6* by the combination of different properties of the protein. J. Biomed. Sci. 20, 25 (2013). https://doi.org/10.1186/1423-0127-20-25

[12] Glusman, G., Rose, P. W., Prlić, A., Dougherty, J., Duarte, J. M., Hoffman, A. S., et al. Mapping genetic variations to three-dimensional protein structures to enhance variant interpretation: A proposed framework. Genome Med. 9, 113 (2017). https://doi.org/10.1186/s13073-017-0509-y

[13] Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., et al. Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet. Med. 17, 405-424 (2015). https://doi.org/10.1038/gim.2015.30

[14] Landrum, M. J., Lee, J. M., Benson, M., Brown, G. R., Chao, C., Chitipiralla, S., et al. ClinVar: Improving access to variant interpretations and supporting evidence. Nucleic Acids Res. 46, D1062-D1067 (2017). https://doi.org/10.1093/nar/gkx1153

[15] Gojobori, T., Ikeo, K., Katayama, Y., Kawabata, T., Kinjo, A. R., Kinoshita, K., et al. VaProS: A database-integration approach for protein/genome information retrieval. J. Struct. Funct. Genomics 17, 69-81 (2016). https://doi.org/10.1007/s10969-016-9211-3

[16] Perrin, B. J., Ervasti, J. M. The actin gene family: Function follows isoform. Cytoskeleton 67, 630-634 (2010). https://doi.org/10.1002/cm.20475

[17] Dominguez, R., Holmes, K. C. Actin structure and function. Annu. Rev. Biophys. 40, 169-186 (2011). https://doi.org/10.1146/annurev-biophys-042910-155359

[18] Blanchoin, L., Boujemaa-Paterski, R., Sykes, C., Plastino, J. Actin dynamics, architecture, and mechanics in cell motility. Physiol. Rev. 94, 235-263 (2014). https://doi.org/10.1152/physrev.00018.2013

[19] Drummond, M. C., Belyantseva, I. A., Friderici, K. H., Friedman, T. B. Actin in hair cells and hearing loss. Hear. Res. 288, 89-99 (2012). https://doi.org/10.1016/j.heares.2011.12.003

[20] Oosawa, F. My various thoughts on actin. Biophys. Physicobiol. 15, 151-158 (2018). https://doi.org/10.2142/biophysico.15.0_151

[21] Schafer, D. A., Schroer, T. A. Actin-related proteins. Annu. Rev. Cell Dev. Biol. 15, 341-363 (1999). https://doi.org/10.1146/annurev.cellbio.15.1.341

[22] Gupta, C. M., Ambaru, B., Bajaj, R. Emerging functions of actins and actin binding proteins in trypanosomatids. Front. Cell Dev. Biol. 8, 587685 (2020). https://doi.org/10.3389/fcell.2020.587685

[23] Cunningham, F., Allen, J. E., Allen, J., Alvarez-Jarreta, J., Amode, M R., Armean I. M., et al. Ensembl 2022. Nucleic Acids Res. 50, D988-D995 (2021). https://doi.org/10.1093/nar/gkab1049

[24] Parker, F., Baboolal, T. G., Peckham, M. Actin mutations and their role in disease. Int. J. Mol. Sci. 21, 3371 (2020). https://doi.org/10.3390/ijms21093371

[25] Zhu, M., Yang, T., Wei, S., DeWan, A. T., Morell, R. J., Elfenbein, J. L., et al. Mutations in the γ-actin gene (*ACTG1*) are associated with dominant progressive deafness (DFNA20/26). Am. J. Hum. Genet. 73, 1082-1091 (2003). https://doi.org/10.1086/379286

[26] Pujar, S., O'Leary, N. A., Farrell, C. M., Loveland, J. E., Mudge, J. M., Wallin, C., et al. Consensus coding sequence (CCDS) database: A standardized set of human and mouse protein-coding regions supported by expert curation. Nucleic Acids Res. 46, D221-D228 (2017). https://doi.org/10.1093/nar/gkx1031

[27] Fagerberg, L., Hallström, B. M., Oksvold, P., Kampf, C., Djureinovic, D., Odeberg, J., et al. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. Mol.

Cell. Proteomics 13, 397-406 (2014). https://doi.org/10.1074/mcp.M113.035600

[28] Bekker, G.-J., Yokochi, M., Suzuki, H., Ikegawa, Y., Iwata, T., Kudou, T., et al. Protein Data Bank Japan: Celebrating our 20th anniversary during a global pandemic as the Asian hub of three-dimensional macromolecular structural data. Protein Sci. 31, 173-186 (2022). https://doi.org/10.1002/pro.4211

[29] Shrake, A., Rupley, J. A. Environment and exposure to solvent of protein atoms. Lysozyme and insulin. J. Mol. Biol. 79, 351-364 (1973). https://doi.org/10.1016/0022-2836(73)90011-9

[30] Hijikata, A., Yura, K., Noguti, T., Go, M. Revisiting gap locations in amino acid sequence alignments and a proposal for a method to improve them by introducing solvent accessibility. Proteins 79, 1868-1877 (2011). https://doi.org/10.1002/prot.23011

[31] Terui-Kohbata, H., Egawa, M., Yura, K., Yoshida, M. Knowledge and attitude of hereditary breast cancer among Japanese university female students. J. Hum. Genet. 65, 591-599 (2020). https://doi.org/10.1038/s10038-020-0743-9

[32] Hanč, P., Fujii, T., Iborra, S., Yamada, Y., Huotari, J., Schulz, O., et al. Structure of the complex of F-actin and DNGR-1, a c-Type lectin receptor involved in dendritic cell cross-presentation of dead cell-associated antigens. Immunity 42, 839-849 (2015). https://doi.org/10.1016/j.immuni.2015.04.009

[33] Kim, L. Y., Thompson, P. M., Lee, H. T., Pershad, M., Campbell, S. L., Alushin, G. M. The structural basis of actin organization by Vinculin and Metavinculin. J. Mol. Biol. 428, 10-25 (2016). https://doi.org/10.1016/j.jmb.2015.09.031

[34] Galkin, V. E., Orlova, A., Salmazo, A., Djinovic-Carugo, K., Egelman, E. H. Opening of tandem calponin homology domains regulates their affinity for F-actin. Nat. Struct. Mol. Biol. 17, 614-616 (2010). https://doi.org/10.1038/nsmb.1789

[35] Ecken, J. v. d., Heissler, S. M., Pathan-Chhatbar, S., Manstein, D. J., Raunser, S. Cryo-EM structure of a human cytoplasmic actomyosin complex at near-atomic resolution. Nature 534, 724-728 (2016). https://doi.org/10.1038/nature18295

[36] Avery, A. W., Fealey, M. E., Wang, F., Orlova, A., Thompson, A. R., Thomas, D. D., et al. Structural basis for high-affinity actin binding revealed by a β-III-spectrin SCA5 missense mutation. Nat. Commun. 8, 1350 (2017). https://doi.org/10.1038/s41467-017-01367-w

[37] Risi, C., Belknap, B., Forgacs-Lonart, E., Harris, S. P., Schröder, G. F., White, H. D., et al. N-terminal domains of cardiac myosin binding protein C cooperatively activate the thin filament. Structure 26, 1604-1611 (2018). https://doi.org/10.1016/j.str.2018.08.007

[38] Devlin, J. R., Alonso, J. A., Ayres, C. M., Keller, G. L. J., Bobisse, S., Vander Kooi, C. W., et al. Structural dissimilarity from self drives neoepitope escape from immune tolerance. Nat. Chem. Biol. 16, 1269-1276 (2020). https://doi.org/10.1038/s41589-020-0610-1

[39] Huehn, A. R., Bibeau, J. P., Schramm, A. C., Cao, W., Cruz, E. M. D. L., Sindelar, C. V. Structures of cofilin-induced structural changes reveal local and asymmetric perturbations of actin filaments. Proc. Natl. Acad. Sci. U.S.A. 117, 1478-1484 (2020). https://doi.org/10.1073/pnas.1915987117

[40] Schwebach, C. L., Kudryashova, E., Zheng, W., Orchard, M., Smith, H., Runyan, L. A., et al. Osteogenesis imperfecta mutations in plastin 3 lead to impaired calcium regulation of actin bundling. Bone Res. 8, 21 (2020). https://doi.org/10.1038/s41413-020-0095-2

[41] Mwangangi, D. M., Manser, E., Robinson, R. C. The structure of the actin filament uncapping complex mediated by twinfilin. Sci. Adv. 7, eabd5271 (2021). https://doi.org/10.1126/sciadv.abd5271

[42] Grintsevich, E. E., Ge, P., Sawaya, M. R., Yesilyurt, H. G., Terman, J. R., Zhou, Z. H., et al. Catastrophic disassembly of actin filaments via Mical-mediated oxidation. Nat. Commun. 8, 2183 (2017). https://doi.org/10.1038/s41467-017-02357-8

[43] Yan, Y., Harding, H. P., Ron, D. Higher-order phosphatase–substrate contacts terminate the integrated stress response. Nat. Struct. Mol. Biol. 28, 835-846 (2021). https://doi.org/10.1038/s41594-021-00666-7

[44] Rebowski, G., Boczkowska, M., Drazic, A., Ree, R., Goris, M., Arnesen, T., et al. Mechanism of actin N-terminal acetylation. Sci. Adv. 6, eaay8793 (2020). https://doi.org/10.1126/sciadv.aay8793

[45] He, S., Wu, Z., Tian, Y., Yu, Z., Yu, J., Wang, X., et al. Structure of nucleosome-bound human BAF complex. Science 367, 875-881 (2020). https://doi.org/10.1126/science.aaz9761

[46] Sasaki, Y., Kozaki, A., Ohmori, A., Iguchi, H., Nagano, Y. Chloroplast RNA editing required for functional acetyl-CoA carboxylase in plants. J. Biol. Chem. 276, 3937-3940 (2001). https://doi.org/10.1074/jbc.M008166200