Prediction of fragile points of coiled coils

Hideki Tanizawa^{1*}, Mieko Taniguchi², Ganga D. Ghimire³ and Shigeki Mitaku¹

¹*Department of Applied Physics, School of Engineering, Nagoya University,
Furocho, Chikusa-ku, Nagoya 464-8603, Japan*

²*Department of Biotechnology and Biomaterial Chemistry, Graduate School of Engineering, Nagoya
University, Furocho, Chikusa-ku, Nagoya, 464-8003 Japan*

³*Research Centre Juelich, Institute of Neurosciences and Biophysics, INB-2, Molecular Biophysics,
D-52425 Juelich, Germany*

*E-mail: tanizawa@bp.nuap.nagoya-u.jp

(Received December 12,2008; accepted January 23,2009; published online February 2,2009)

Abstract

A prediction system for identifying the region of flexible regions of the coiled coil was developed to determine the bending positions of the myosin rods using atomic force microscopy (AFM) and to analyze the molecular structures of proteins containing coiled coils. The prediction system comprises two modules: identification of heptad break points and prediction of fragile points in the coiled coil due to the hydrophilic core or hydrophobic outfield region. Here, we investigated the myosin rods using this prediction system. The results of AFM imaging showed four main flexible regions in a single myosin rod and of the 17 possible fragile points predicted, 16 were located in the four experimental bending regions. Next, we examined the enhanced fluctuation around these predicted fragile points using the B-factor for the three dimensional structure of coiled coil proteins from the SCOP database and found that the fluctuations in the hydrophilic core regions were significantly larger than those in the regions of the normal coiled coil. In contrast, the fluctuations in the hydrophobic outfield regions were reduced, suggesting a structural change of the coiled coils to balance these regions. Thus, the dynamic changes in the structure of the coiled coils around the fragile points may be related to the biological functions of the proteins. The prediction tool which developed in this work was incorporated in the SOSUicoil system which predicts the coiled coil regions.

Key Words: coiled coil, myosin, SOSUicoil, fragile point, atomic force microscopy

Area of Interest: Bioinformatics and Bio Computing, Nanotechnology of Single molecular

1. Introduction

Proteins containing a coiled coil structure have a variety of functions: formation of large, mechanically rigid structures such as hair, scales and feathers (keratin), as well as blood clots (fibrin), the cellular skeleton (intermediate filaments), molecular stalks (kinesin, influenza haemagglutinin) and levers (myosin). At the molecular level of coiled coils, patterns of heptad repeats have been observed that contain hydrophobic amino acids alternately located three and four residues apart [1]. The positions of these repeats are usually labeled alphabetically (a - g) with registers of "a" and "d" that represent the positions of hydrophobic residues. Regular heptad repeat patterns are sometimes broken by nonheptad inserts [2-4], generally described as heptad breaks. Four types of heptad breaks have been reported [4, 5]: (1) The stutter is the break in which three residues are deleted from the heptad repeats [1]. It has the tendency to loosen the supercoiling locally. (2) The stammer is the break in which four residues are deleted [5]. It has the tendency to tighten the supercoiling by decreasing the local pitch of the helix at this type of break. (3) The skip(+) is the break in which one residue is inserted [6]. The structure of a skip(+) forms a π -turn in the α -helix. It has a tendency to make a kink in the helix. (4) The skip(-), which we discovered recently, is another type of heptad breaks in which one residue is deleted [30].

In many cases, heptad breaks correlate well with the structural breaks of the coiled coil. There are also many heptad breaks present in the conserved regions of the coiled coil structure. However, the factors that determine the final structure of the coiled coil regions are still unclear. There are two possible factors for the structural change of the coiled coils: the protein environment, such as the ion concentration and the protein-protein interactions, and the amino acid sequence deviations of typical heptad repeats.

Coiled coil structures in several proteins, including influenza haemagglutinin [7], bZip transcription factors [8] and myosin [9], have been reported to have flexible regions, which we call fragile points. These fragile points are the hot spots of the dynamic motions of the coiled coil structure, indicating the significant function of these regions. Myosin is a well-known coiled coil protein that contains several fragile points. Its molecular structure includes a globular head and a long tail composed of a coiled coil [9-11]. The sliding of the myosin heads along the actin requires the bend of the coiled coil. Thus, the fragile points in the coiled coil region of the myosin rod may allow the myosin tail to bend near its center, thus moving the myosin head away from the thick filament and toward the actin filament [13]. In fact, sharp bends in the tail were observed by electron microscopy [9]. Furthermore, the detailed analysis of the amino acid sequence revealed that the myosin tail has four one-residue insertions (skip(+)-type heptad breaks) at the sequence positions of 351, 548, 745 and 970 [14]. Although a correlation between the positions of fragile points and skips has been observed by electron microscopy [15], the skip(+) breaks could not sufficiently explain the wide distribution of fragile points observed.

In this work, we studied the coiled coil structures, focusing on the fragile points that are closely related to the important biological functions of the proteins. First, we developed a software system to predict the possible fragile points from the amino acid sequences alone, assuming that the fragile points correspond to the heptad break points and the segments with weak hydrophathy between the core and the outfield regions. Then, using this software system, we observed the myosin tail regions in various experimental conditions and compared the B-factor of the coiled coil structures in the SCOP database between the predicted fragile points and the normal coiled coil regions. The results indicated that both the heptad breaks and the regions with weak hydrophilic core or hydrophobic outfield in the coiled coils are closely related to its fragility, even when the structure is the same as

the normal coiled coils. Moreover, we introduce the first software system for predicting the flexible region in coiled coil structures.

2. Materials and Methods

2.1 Experimental

2.1.1 Sample preparation

Myosin was prepared from rabbit skeletal muscle according to the modified method of Szent-Gyorgi[17, 18]. Briefly, myosin was purified in three precipitation cycles of a low-salt solution containing 50mM KCl and 10mM Tris-HCl buffer (pH 8.0). This solution was centrifuged at 4000 rpm and the precipitant was resolved in 0.5M KCl and 10mM Tris-HCl buffer (pH 8.0). The supernatant was dialyzed against 0.5M KCl and 10mM Tris-HCl buffer (pH 8.0) and the final solution was ultracentrifuged at $150,000 \times g$ for 1 h at 4°C. Purified myosin was stored in a stock solution (50% glycerol, 0.5M KCl, and 10mM Tris-HCl buffer (pH 8.0)) at -20°C until use.

Before atomic force microscopy (AFM), the myosin stock solution was first diluted with either the rigor or relaxing solutions. The MgATP-free rigor solution contained 0.5MKCl in 10mM Tris-HCl buffer (pH 8.0). The MgATP-controlled relaxing solutions contained 0.5M KCl, 10mM ethylene glycol-bis-(2-amino-ethyl ester) tetra acetic acid (EGTA), 10mM EDTA, 10mM Tris-HCl buffer (pH 8.0), 4mM ATP and varying concentrations of $MgCl_2$. The values of $p[MgATP]$ were calculated using the method of Reuben et al [19]. The diluted solution was incubated at 26°C for 10 min.

A drop of myosin solution (10pM, 1 μ l) was spotted onto the surface of mica and allowed to stand for 60 s, followed by washing with double distilled water to remove the unattached myosin and solvent. The sample was then dried in a clean bench for 30 min with an antistatic device (Static free SF-100, Japan) to remove excess glycerol and water; no residual solvent, buffer, or glycerol was detected on the final mica surfaces. The high ion strength in the solution (0.5M KCl) prevented the folding of the myosin molecules over each other, allowing for observation under near-native conditions [18].

2.1.2 Atomic Force Microscopy

All AFM images were collected using a SPA 400 (SII Nanotechnology, Inc., Chiba, Japan) set to the tapping mode, as described elsewhere [18]. An etched silicon single crystal cantilever (NCH-10T) was used for scanning. The cantilever, which oscillates vertically to scan the sample surface, was set at a resonance frequency of 3kHz and drive amplitude of 200mV. The samples were scanned in air at a speed of 3Hz and a resolution of 512×512 pixels at 26°C with low humidity to minimize moisture in the sample. The scanner was calibrated from images of a standard grid and mica surface in air. Each AFM image of a single myosin molecule was taken in the tapping mode with an image dimension of $200 \times 200 \text{ nm}^2$. Photoshop (Adobe Systems, Inc., San Jose, CA, USA) and NIH Image software (National Institute of Mental Health (NIMH), Bethesda, MD, USA) were used to visualize the AFM images and to calculate the bending positions in the myosin rod, respectively.

2.2 Theoretical

2.2.1 Identification of heptad breaks as possible fragile points

Although most current software systems, such as COILS [20] and Paircoil2 [21], can accurately predict heptad repeat regions of coiled coils, they are unable to accurately predict registers around the discontinuous regions of the heptad repeat, known as the heptad breaks. Therefore, a semi-automatic method has been adopted for identifying heptad breaks in these methods. That is, the heptad repeat register are first determined in COILS and Paircoil [22] software systems, and then the register around the discontinuous region of heptad repeats are manually defined based on the region surrounding the register [2].

Figure 1 The process for determining the appropriate register of heptad repeats and breaks
 (A) The templates for 34 types of registers including 2 typical heptad repeats and 32 heptad breaks are used for determining the appropriate registers.
 (B) Among the positions of the seven residues from “a” to “g”, the residues at the core region of “a” and “d” are hydrophobic in the typical heptad repeats, as shown in green in the wheel diagram.

For determining the appropriate register, we had developed a prediction system SOSUIcoil [30] previously. It based on the 34 types of templates, including two typical heptad repeat patterns and 32 templates for heptad breaks (Figure 1A). SOSUIcoil used only the hydrophobicity of the core region because it is the most important factor for the supercoiling. Therefore, we defined the evaluation function $S(i)$ for the typical heptad repeats as Eq. (1).

$$S(i) = H(i-14) + H(i-11) + H(i-7) + H(i-4) + H(i) + H(i+3) + H(i+7) + H(i+10) + H(i+14) \quad \dots \quad (1)$$

in which $H(i)$ represents the hydropathy index of Kyte and Doolittle [23] at the sequence position i . This evaluation function is devised for determining the phase of the heptad repeat of the hydrophobic residues using the window 29 residues from $(i-14)$ to $(i+14)$. When the center of the window is positioned at the core of the coiled coil, “a” or “d”, the evaluation function has a large positive value. In contrast, when the center of window is positioned at the outfield region of coiled coil, the evaluation function has a large negative value.

Since a periodic change in the evaluation function is essential for the stability of the coiled coil, continuous hydrophobic segments can not form coiled coil structures. Therefore, we calculated the minimum value of $S(i)$ in the five residue window around the i -th position, and the difference between the evaluation function and the minimum value was used for evaluating the coiled coil regions.

$$D(i) = \min \{S(i) - \{S(i+k)\} : (-2 \leq k \leq 2, k \neq 0)\} \dots\dots\dots (2)$$

The regions with high $D(i)$ values have large differences in hydrophobicity between the core and outfield region in Figure 1B, leading to the high probability of a coiled coil. The peaks of $D(i)$ correspond to either the register “a” or “d”, and the interval from “a” to the next “d” is three residue, while that from “d” to the next “a” is four. Therefore, there are two possible templates for the typical heptad repeats in which the center of template corresponds to “a” or “d”.

However, mismatched regions between two typical heptad repeats sometimes occur, indicating possible heptad breaks in these regions. The evaluation functions of templates for heptad breaks were obtained by modifying the evaluation function Eq. (1) for the typical heptad repeats.

$$\begin{aligned} S_{j,s}(i) = & H(i \mp 14 + \alpha_1(j,s)) + H(i \mp 11 + \alpha_2(j,s)) + H(i \mp 7 + \alpha_3(j,s)) + \\ & H(i \mp 4 + \alpha_4(j,s)) + H(i) + H(i \pm 3 + \alpha_5(j,s)) + H(i \pm 7 + \alpha_6(j,s)) + \\ & H(i \pm 10 + \alpha_7(j,s)) + H(i \pm 14 + \alpha_8(j,s)) \\ & \dots\dots\dots (3) \end{aligned}$$

The parameters from $\alpha_1(j,s)$ to $\alpha_8(j,s)$ correct for the shifts in the core registers of “a” and “d”. Table 1 shows the values of the corrections. The parameter j represents the types of heptad breaks: 1-8 correspond to skip, 9-12 correspond to stutter and 13-16 correspond to stammer. The parameter s represents the same type of two heptad breaks at once whose core registers are shifted in the opposite direction: s = “upper” for upper signs and s = “lower” for the lower signs.

Table 1 Correction terms $\alpha_i(j,s)$ for 32 heptad break patterns.The parameter s is “upper” or “lower” for the upper or lower sign, respectively.

		$\alpha_1(j,s)$	$\alpha_2(j,s)$	$\alpha_3(j,s)$	$\alpha_4(j,s)$	$\alpha_5(j,s)$	$\alpha_6(j,s)$	$\alpha_7(j,s)$	$\alpha_8(j,s)$
skip(+)	$j = 1$	∓ 1	0	0	0	0	0	0	0
	$j = 2$	∓ 1	∓ 1	∓ 1	0	0	0	0	0
	$j = 3$	0	0	0	0	± 1	± 1	± 1	± 1
	$j = 4$	0	0	0	0	0	0	± 1	± 1
skip(-)	$j = 5$	± 1	± 1	0	0	0	0	0	0
	$j = 6$	± 1	± 1	± 1	± 1	0	0	0	0
	$j = 7$	0	0	0	0	0	∓ 1	∓ 1	∓ 1
	$j = 8$	0	0	0	0	0	0	0	∓ 1
stutter	$j = 9$	∓ 1	0	0	0	0	0	0	0
	$j = 10$	∓ 1	0	∓ 1	0	0	0	0	0
	$j = 11$	0	0	0	0	± 1	0	± 1	0
	$j = 12$	0	0	0	0	0	0	± 1	0
stammer	$j = 13$	0	± 1	0	0	0	0	0	0
	$j = 14$	0	± 1	0	± 1	0	0	0	0
	$j = 15$	0	0	0	0	0	∓ 1	0	∓ 1
	$j = 16$	0	0	0	0	0	0	0	∓ 1

Among 32 types of templates, the template with the highest $S_{j,s}(i)$ values was chosen as the appropriate template. In the case of heptad breaks, the registers “a” and “d” were directly determined from the templates. The heptad break points were considered to be the possible fragile points.

2.2.2 Possible fragile points at hydrophilic core and hydrophobic outfield

The coiled coil structure is stabilized mainly by the hydrophobic interaction at the core regions and the affinity of the outfield region with water. In other words, the contrast in the hydropathy between the core and the outfield regions, namely the amphiphilicity of the helices, is the main cause of the coiled coil. Therefore, when the hydrophobicity of the core regions becomes very low or the hydrophobicity of the outfield region becomes high, the contrast of the hydropathy between the core and the outfield disappears and the stability of the coiled coil structure is reduced. We developed a software module for predicting fragile points based on the “hydrophilic core” or “hydrophobic outfield”. After determining the heptad repeat registers, the double average of the hydropathy index is calculated by the following equation:

$$F_x(j) = \sum_{i=j-4}^{j+4} \left(\sum_{k=i-4}^{i+4} H_x(k)/9 \right) / 9 \dots\dots\dots (4)$$

where the suffix x represents the regions of the “core” or “outfield”, and $H_x(k)$ and $F_x(j)$ are the hydropathy index value and the evaluation function, respectively. We used the double average values because the single average plots are still very notched. The threshold for the fragile points by the hydrophilic core was -1.0 , most likely resulting in a break in the coiled coil structure, and that for the outfield was -0.5 .

2.2.3 Evaluation of structure fluctuation by normalized B-factor

A median-based method [24] was used for the evaluation of the structure fluctuation by B-factor. First, the median of the C_{α} -B-factor in a chain was determined and then the median of absolute displacements (MAD) was determined from the median. An $M(i)$ value of each B-factor was calculated by the following equation:

$$M(i) = 0.6745 \times (x_i - \bar{x}) / MAD \quad \dots \dots \dots (5)$$

where x_i is the B-factor of the i -th residue and $\bar{x}$ is the median of the B-factors; the equation is multiplied by a factor of 0.6745 as the expected value of MAD is 0.6745σ for large sample sizes [24]. An $M(i)$ value of ≥ 3.5 is used as the outlier.

Normalized B-factors were calculated according to David K. Smith [25]. After removal of the outlier, the mean (μ_{noout}) and standard deviation (σ_{noout}) of the remaining C_{α} -B-factor in the chain were determined to calculate a normalized B-factor as follows:

$$B_{\text{normalized},i} = (x_i - \mu_{\text{noout}}) / \sigma_{\text{noout}} \quad \dots \dots \dots (6)$$

Thus, the normalized B-factors have zero mean and unit variance.

2.3 Data set for B-factor comparison

The data set for B-factor analysis was constructed from the SCOP database [26] (ver. 1.73) (proteins are listed in Table 2). Sequences with homologies $> 30\%$ were removed. Fragile points located in the middle of the coiled coil > 7 residues from the terminus were used to calculate the normalized B-factor.

Table 2 PDB codes of 54 data sets of coiled coil proteins

Each data set was derived from Protein Data Bank.

1fif	2fxo	1hbw	1ha0	1jun	2ocy	1aa0
2np0	1d7m	1aq5	1mg1	1jad	2b9c	1ebo
1flk	1wpa	1ajy	1mv4	1qu7	1qu1	2e7s
2ezp	1ezj	1m7l	1vsg	1ik9	2d3e	1ox3
2cpb	1r48	1qey	2vsg	1av1	1wyy	1lj2
1fdm	1hf9	1dip	1g5g	1mlj	1joc	1m5i
2b5u	2a93	1zta	2ch7	1fu1	1env	1l8d
1c1g	1pjf	1cii	1qoy	1dkg		

3. Results

3.1 Observation of bending positions by AFM

We observed the bending positions of rabbit skeletal myosin rods by AFM (indicated by blue triangles in Figure 2). In some cases, myosin had more than two bending positions in one molecule. Among more than 100 samples of myosin, half of them had a straight conformation without any

bending (X; Figure 2), while the other half had bends at various positions.

We measured the length between the head-tail junction and the bending positions at varying concentrations of MgATP (Figure 3); lengths longer than 120 nm were not measured as some of the ends of the tail (C-terminal region of the myosin rod) were unclear. We classified the distance distribution into four regions: 29 ± 2 nm (region I), 49 ± 5 nm (region II), 79 ± 5 nm (region III) and 105 ± 2 nm (region IV). As seen in Figure 3, there are void regions in which no bends were found, indicating that the bends are characteristic of the amino acid sequence of the myosin rod.

To confirm the fragile points in the myosin rod, we studied the distance distribution at varying concentrations of MgATP ($p[\text{MgATP}] = 4, 5, 6, 7$ and > 9). Bends most frequently occurred at $p[\text{MgATP}] = 5$, the concentration corresponding to muscle contraction. The frequency of bends depended on the MgATP concentration around regions (II), (III) and (IV) of the distance distribution. Region (II) had many bends at $p[\text{MgATP}] = 5$, but no bends at low concentrations of MgATP ($p[\text{MgATP}] > 9$). In contrast, fewer bends were observed at high MgATP concentration around region (III) and significant differences were found around region (IV), in which bends occurred only at $p[\text{MgATP}] = 5$, except for one bend at $p[\text{MgATP}] > 9$. However, in region (I), the bends of the coiled coil showed little dependence on the MgATP concentration.

Figure 2 AFM images of single myosin molecules

Lengths from head to bend were measured. Typical examples of five types of myosin molecules with respect to the position of bends are shown: (X) straight conformation; (I) bends at about 29 nm; (II) about 49 nm; (III) about 79 nm; and (IV) about 105 nm.

Figure 3 MgATP concentration dependence of bending positions of myosin rod.

Plot of the bending positions in the tail are shown for 5 p[MgATP] concentrations. The length was measured from the head. The position of the bends was grouped into 4 regions: (I) 29 ± 2 nm, (II) 49 ± 5 nm, (III) 79 ± 5 nm and (IV) 105 ± 2 nm.

3.2 Analysis of amino acid sequence of myosin by heptad breaks and helical amphiphilicity

Figure 4 shows the double averaged hydropathy indices of 9 residue windows. The myosin rod is mostly hydrophilic, and the average hydropathy index is about -1 . However, the core regions are comprised of hydrophobic residues and the average hydropathy index is around $+1$. Only the fragile regions have average hydropathy indices below -1 . In contrast, the outfield regions are very hydrophilic, with indices that are around -2 and rarely exceed -0.5 . The outfield regions of the average hydropathy above -0.5 showed good correlation with the fragile regions of the myosin rod. Therefore, we defined the “hydrophilic core” and “hydrophobic outfield” fragile regions by the thresholds of the average hydropathy indices of -1 and -0.5 , respectively.

Figure 5 shows the fragile points experimentally determined by AFM, as well as the “hydrophilic core” region, the “hydrophobic outfield” region, and the three types of heptad break points. There were six hydrophilic cores and four hydrophobic outfields. Heptad breaks were found at seven positions: three skips, three stutters and one stammer. We found three skips at 349, 532 and 968, two stutters at 718 and 757, which correspond to previous reports of four skips in the myosin rod (351, 548, 745 and 970) [14]. Because two stutters (deletion of 3 residues) are mathematically equivalent to the deletion of 6 residues, namely one skip (insertion of one residue). Therefore, the present result is actually consistent with the previous work.

Figure 4 Hydropathy indices of the myosin rod.

(A) Histogram of the observed bending positions.

(B) Averaged hydropathy indices of the myosin rod. The hydropathy index was double averaged by 9 residue windows. Each graph was averaged using the values from different regions. The top graph used all the regions, the middle graph used only the core region, and the bottom graph used only the outfield region. The area shaded in blue shows the regions of the average hydropathy below -1 in the core region. The dotted red line indicates the threshold value of -1 . The area shaded red shows the region of the average hydropathy above -0.5 in the outfield region. The dotted blue line indicates the threshold value of -0.5 .

The predicted fragile regions due to the hydropathy index values and the heptad breaks agreed well with the 4 regions of the experimental bending positions (see Figure 5). Among ten predicted fragile points from the hydropathy analysis, eight were located in regions (II) and (III) in Figure 5, correlating well with many experimental bending positions in those regions. Here, we introduced for the first time the prediction of fragile points by the hydrophilic core and hydrophobic outfield regions. The good correlation between the frequency of the experimentally observed bends and those of the anomalous hydropathy at the core and outfield regions suggests that the contribution of the anomalous hydropathy is large and therefore useful for predicting fragile points.

A **B** **Figure 5** Comparison of experimentally observed and predicted fragile points

(A) Histogram of observed bending positions of the myosin rod obtained from Figure 3. Lengths longer than 120 nm were not measured. Bends are shown in blue.

(B) Plot of the predicted fragile points and three discontinuous heptad repeats by SOSUicoil. The horizontal axis indicates the number of amino acid residues from the myosin head. The scale was adjusted to the length of the AFM data.

3.3 Prediction of fragile points in coiled coil proteins from SCOP database

We analyzed the amino acid sequences of the coiled coil from the SCOP database (ver. 1.73) [26] and studied the relationship between the predicted fragile regions and their structural characteristics. We first examined the statistical meaning of the threshold of F_x , -1 for hydrophilic core and -0.5 for hydrophobic outfield. Figure 6 shows the distribution of the F_x values at all residues in coiled coil regions. The distribution of F_x for the core and the outfield well agreed with

the Gaussian distribution, and the threshold values of core were beyond the double standard deviations and threshold values of outfield were beyond the 1.5 standard deviations. Therefore, the predicted fragile points have to be very rare points in the coiled coil structure.

Figure 6 Comparison of F_x values of coiled coil proteins

(Top) Histogram of F_x values of core region. The scores under -1 were shown in blue color.

(Bottom) Histogram of F_x values of outfield region. The scores above -0.5 were shown in red color.

Figure 7 and Figure 8 present molecular structures and schematic diagrams of proteins, showing the relationship between the helical regions and the predicted fragile regions, respectively. Hydrophilic core and hydrophobic outfield regions are shown in blue and red, respectively. In addition, heptad breaks are shown in brown (skip), green (stammer) and purple (stutter). The proteins were categorized by the structural characteristics of around the fragile regions into three groups: a kink (A), dynamic structural change (B) and straight coiled coil (C).

Examples of the first group are shown in Figure 7A. The kinks are accompanied by the hydrophilic core regions and skips were found near the kinks in 1WPA and 1GRJ. The two examples in the second group (Figure 7B) have reportedly different crystal structures depending on the environment. In the example of ribosomal protein S15 from *Thermus thermophilus* (1AB3) [27], the structure in the single molecule state shows more loop regions than in its complex with RNA, which becomes a highly structured form (1FJG) of 30S ribosomal subunit. The other example is influenza haemagglutinin (1EO8, 1HTM) [7], which changes its structure drastically as pH decreases; that is, in acidic conditions, the helical structure within the coiled coil breaks at the fragile point, exposing the hydrophobic part of the helix to the environment. This structural change probably induces the interaction with the membrane. Therefore, the structure of the coiled coil around the fragile regions is determined by the environment, such as the ionic conditions and interaction with other molecules, indicating the importance of the fragile points for the biological functions of proteins. The third group of proteins shows straight coiled coils around the predicted fragile points, some of which are located in the middle of the coiled coils (Figure 7C and Figure 8C).

Figure 7. Examples of predicted fragile points of various proteins.

Examples of coiled coil protein structures having fragile points. The predicted hydrophilic core and hydrophobic outfield are shown in blue and red, respectively. The overlapped region are shown in green. The C α atom in the predicted heptad breaks are shown in red (skip), green (stammer) and purple (stutter). The structures were rendered with Molscript [28]

(A) The fragile points of occludin (1WPA), tropomyosin (1MV4) and GreA transcript cleavage factor (1GRJ) placed in kinked positions of their crystal structures.

(B) Ribosomal protein S15 from *Thermus thermophilus* shows changes in its single structure (1AB3) when complexed with RNA (1FJ3). Fragile points were located on the terminus of the coiled coil in the complex structure and in the loop region in single structure.

A long connecting loop in influenza haemagglutinin at pH 7 (1EO8) refolds into a coiled coil at pH 5 after dissociation of the globular head subunit. The bottom half of the long α -helix diverges around the predicted fragile point (1HTM).

(C) Some of the fragile points of tumor suppressor gene product APC (1M5I), rotavirus NSP3 (1LJ2) and PUT3 (1AJY) were located on the straight position of coiled coil.

Figure 8 Predicted hydrophilic fragile regions and heptad breaks
Each protein corresponds to those shown in Figure 6.

3.4 Analysis of normalized B-factors around fragile regions of coiled coils

Analyzing the structure around the fragile regions is difficult to not only the change in the static structure but also the dynamic structure. In order to study the dynamic fluctuation of the coiled coils, we focused on the atomic displacement parameter, or B-factor. The B-factors from X-ray crystallographic give information on the mobility of each of the atoms in the structure [29]. This B-factor reflects the degree of thermal motion and static disorder of an atom in a protein crystal structure. Thus, we analyzed the normalized B-factor of the coiled coils around various types of fragile points. We normalized the B-factor to obtain the average value of the whole protein of zero. Figure 9 shows the results for the hydrophilic core, the hydrophobic outfield, the heptad break points and the normal coiled coil regions. The mean normalized B-factors of the hydrophilic core, the hydrophobic outfield regions and the normal coiled coils were 0.855, -0.640 and 0.363 , respectively. Interestingly, the differences in the distribution of the normalized B-factor indicate that the effects of the hydrophilic core and the hydrophobic outfield contradict the normal coiled coils. The hydrophilic core regions enhanced the structural fluctuation, forcing the coiled coil to become more fragile. In contrast, the hydrophobic outfield regions reduced the fluctuation, probably stabilizing the coiled coils. It should be noted that the hydrophilic core and the hydrophobic outfield regions are coupled in the type C coiled coil proteins (Figures 6 and 7), and the coiled coil structures around the coupled regions appears to be stable. On the other hand, in the type B coiled coil proteins, the hydrophilic core regions were not coupled with the hydrophobic outfield regions, but were apart from each other. In addition, the structural changes always occurred at the hydrophilic core region.

Figure 9. Comparison of normalized B-factors

The atom of C_{α} -B-factor around the fragile region, heptad breaks and other regions were compared. The proteins used for calculation were shown in section 2.3. Each B-factor was normalized. The full line shows the data around the hydrophobic fragile region, the dashed line shows the hydrophilic fragile region, and the dotted line shows the normal coiled coil.

4. Discussion

In this study, we investigated hot spots related to the structural changes in the coiled coil, namely the fragile points. Due to their subtle nature, investigating fragile points can be difficult as they are found in straight coiled coils in some conditions and in broken coiled coils in other conditions. A small distortion or kink in the coiled coil can lead to a large structural change in the coiled coil, which may be closely related to the function of the protein. Thus, developing a prediction system requires the determination of the break points of the heptad repeats. We found that six types of breaks of the heptad repeats: four types are characterized by the phase shift of the heptad repeats and two types are related to the decrease in the hydropathy contrast between the core and the outfield regions. Combining those six types of mechanisms, we could identify the fragile regions of the myosin rod and the breaks of the coiled coil, including kinks and the helical end. Furthermore, we showed that the enhanced fluctuation of coiled coils at the hydrophilic core regions is coupled with the reduced fluctuation at the hydrophobic outfield regions.

Proteins become functional through dynamic structural changes. The structural change is different between proteins, involving either side chains, relatively short segments or domains. The coiled coil structure is one of the simplest structures, but its dynamic structural change, namely the break of the coiled coil, is an essential part of the function of various proteins. For example, the molecular process of muscle contraction cannot be understood without elucidating the coupling between the dynamic breaks of the coiled coil structure of the myosin rod and the actin-myosin binding in terms of the physical properties of amino acid sequence. Thus, we previously developed a prediction system for the coiled coil regions SOSUIcoil [30], and here we developed supplementary prediction tool for fragile points of coiled coils. Combining the two modules, the prediction of coiled coil regions and the identification of the fragile points, results in a simultaneous prediction of the coiled coil and its possible dynamic structural change. As Figure 8 suggests, it is possible to predict also the mode of the structural change by determining the distance between the hydrophilic core and the hydrophobic outfield regions. The prediction tool which developed in this work was incorporated in the SOSUIcoil system which predicts the coiled coil regions.

Our strong point of the present study was to evaluate the prediction system by experimentally observing the myosin rods using AFM. The very good correlation between the break points of the myosin rods and the predicted fragile points confirmed the validity of the assumption in the prediction system. However, as the present study investigated only a few data sets, further studies are needed to determine the accuracy of the prediction system. Thus, its quantitative evaluation in the structural biology of other coiled coil proteins will be beneficial.

This work was supported in part by a Grant-in-Aid for the 21st Century COE “Frontiers of Computational Science” and for the Creative and Pioneering Research (B) from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

References

- [1] D. A. Parby, Fibrinogen: a preliminary analysis of the amino acid sequences of the portion of the alpha, beta and gamma-chains postulated to form the interdomainal link between globular regions of the molecule., *J. Mol. Biol.*, **120**, 545-551, 1978.
- [2] J. H. Brown, C. Cohen, and D. A. Parry, Heptad breaks in alpha-helical coiled coils: stutters and stammers., *Proteins*, **26**, 134-145, 1996.

- [3] M. R. Hicks, D. V. Holberton, C. Kowalczyk, and D. N. Woolfson, Coiled-coil assembly by peptides with non-heptad sequence motifs., *Fold. Des.*, **2**, 149-158, 1997.
- [4] A. Lupas, Coiled coils: new structures and new functions., *Trends. Biochem. Sci.*, **21**, 375-382, 1996.
- [5] M. Gruber and A. N. Lupas, Historical review: another 50th anniversary--new periodicities in coiled coils., *Trends. Biochem. Sci.*, **28**, 679-685, 2003.
- [6] A. D. McLachlan and J. Karn, Periodic features in the amino acid sequence of nematode myosin rod., *J. Mol. Biol.*, **164**, 605-626, 1983.
- [7] P. A. Bullough, F. M. Hughson, J. J. Skehel, and D. C. Wiley, Structure of influenza haemagglutinin at the pH of membrane fusion., *Nature*, **371**, 37-43, 1994.
- [8] D. Krylov, M. Olive, and C. Vinson, Extending dimerization interfaces: the bZIP basic region can form a coiled coil., *EMBO J.*, **14**, 5329-5337, 1995.
- [9] A. Elliott and G. Offer, Shape and flexibility of the myosin molecule., *J. Mol. Biol.*, **123**, 505-519, 1978.
- [10] H. S. Slayter and S. Lowey, Substructure of the myosin molecule as visualized by electron microscopy., *Proc. Natl. Acad. Sci. U S A*, **58**, 1611-1618, 1967.
- [11] M. Walker, P. Knight, and J. Trinick, Negative staining of myosin molecules., *J. Mol. Biol.*, **184**, 535-542, 1985.
- [12] D. H. Elliott, STRUCTURE AND FUNCTION OF MAMMALIAN TENDON., *Biol. Rev. Camb. Philos. Soc.*, **40**, 392-421, 1965.
- [13] H. E. Huxley, The mechanism of muscular contraction., *Science*, **164**, 1356-1365, 1969.
- [14] A. D. McLachlan and J. Karn, Periodic charge distributions in the myosin rod amino acid sequence match cross-bridge spacings in muscle., *Nature*, **299**, 226-231, 1982.
- [15] G. Offer, Skip residues correlate with bends in the myosin tail., *J. Mol. Biol.*, **216**, 213-218, 1990.
- [16] A. Szent-Györgyi, The mechanism of muscle contraction., *Proc. Natl. Acad. Sci. U S A*, **71**, 3343-3344, 1951.
- [17] M. Taniguchi and H. Ishikawa, In situ reconstitution of myosin filaments within the myosin-extracted myofibril in cultured skeletal muscle cells., *J. Cell Biol.* **92**, 324-332, 1982.
- [18] M. Taniguchi, et al., MgATP-induced conformational changes in a single myosin molecule observed by atomic force microscopy: periodicity of substructures in myosin rods., *Scanning*, **25**, 223-229, 2003.
- [19] J. P. Reuben, P. W. Brandt, M. Berman, and H. Grundfest, Regulation of tension in the skinned crayfish muscle fiber. I. Contraction and relaxation in the absence of Ca (pCa is greater than 9)., *J. Gen. Physiol.*, **57**, 385-407, 1971.
- [20] A. Lupas, M. V. Dyke, and J. Stock, Predicting coiled coils from protein sequences., *Science*, **252**, 1162-1164, 1991.
- [21] A. V. McDonnell, T. Jiang, A. E. Keating, and B. Berger, Paircoil2: improved prediction of coiled coils from sequence., *Bioinformatics*, **22**, 356-358, 2006.
- [22] B. Berger, et al., Predicting coiled coils by use of pairwise residue correlations., *Proc. Natl. Acad. Sci. U S A*, **92**, 8259-8263, 1995.

- [23] J. Kyte and R. F. Doolittle, A simple method for displaying the hydropathic character of a protein., *J. Mol. Biol.*, **157**, 105-132, 1982.
- [24] B. Iglewicz and D. C. Hoaglin, *How to detect and handle outliers*. ASQC Quality Press, 1993.
- [25] D. K. Smith, P. Radivojac, Z. Obradovic, A. K. Dunker, and G. Zhu, Improved amino acid flexibility parameters., *Protein Sci.*, **12**, 1060-1072, 2003.
- [26] A. G. Murzin, S. E. Brenner, T. Hubbard, and C. Chothia, SCOP: a structural classification of proteins database for the investigation of sequences and structures., *J. Mol. Biol.*, **247**, 536-540, 1995.
- [27] H. Berglund, A. Rak, A. Serganov, M. Garber, and T. Hard, Solution structure of the ribosomal RNA binding protein S15 from *Thermus thermophilus*., *Nat. Struct. Biol.*, **4**, 20-23, 1997.
- [28] P. J. Kraulis, MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures, *Journal of Applied Crystallography*, **24**, 946-950, 1991.
- [29] D. Ringe and G. A. Petsko, "Study of protein dynamics by X-ray diffraction.," *Methods Enzymol.*, **131**, 389-433, 1986.
- [30] H. Tanizawa, et al., A high performance prediction system of coiled coil domains containing heptad breaks: SOSUIcoil., *CBIJ*, **8**, 96-111, 2008.