Published in final edited form as:

Biochim Biophys Acta. 2008 September ; 1784(9): 1301–1309. doi:10.1016/j.bbapap.2008.05.010.

Stability of Dystrophin STR Fragments in Relation to Junction Helicity

Ahmed Mirza and Nick Menhart

Department of Biological, Chemical and Physical Sciences, Illinois Institute of Technology, 3101 S. Dearborn, Chicago IL 60616 USA

Abstract

Dystrophin is a rod shaped protein consisting of amino- and carboxy- terminal binding domains linked by a large central rod composed of 24 homologous copies of the STR motif and 4 non-homologous regions termed hinges. These hinges are proposed to confer local flexibility; conversely, the tacit implication is that the STR regions away from the hinges are comparatively rigid. This, and the repeating nature of this rod, has contributed to the view that the STR region of the rod is uniform and monolithic. However, we have produced various 2 STR fragments, chosen to have high and low α -helix content at their junctions with each other, and show that they exhibit markedly different stabilities. In contrast to a related protein, spectrin, these differences are not correlated with the calculated helicity, but appear to be an intrinsic property of the motifs themselves. A full understanding of how these properties vary along the length of the rod has implications for the engineering of these rods regions in exon-skipping and mini-dystrophin therapies.

Keywords

dystrophin; spectrin-type-repeat; domain; stability

Introduction

Dystrophin is a member of the spectrin family of proteins[1,2], a group of rod shaped structural proteins that stabilize a wide array of tissues. These proteins are recognized by being predominantly composed of a central rod domain consisting of multiple tandem copies of an approximately 110 amino acid motif termed the spectrin type repeat, or STR. Twenty-four such STRs comprise the approximately 120 nm long rod region of dystrophin. These rods are thought to serve as molecular reinforcing elements, providing mechanical stability to various tissues and cells. In general, these rod regions are flanked by various globular domains that are binding modules of some sort; these serve to determine the biological partners of each protein, and so to which cellular constituents they are linked. In the case of dystrophin, an N-terminal actin binding domain mediates an interaction with F-actin in the contractile apparatus of muscle tissue (although a secondary actin binding region has also been identified in part of the rod [3]); and several C-terminal globular domains mediate an interaction with a large multisubunit protein complex termed the DGC, or dystrophin glycoprotein complex, that is embedded in the sarcolemma. The mechanical linkage so afforded between the machinery of muscle action and the cell membrane is crucial for long term integrity of the myocyte: when dystrophin is

absent, as it is in Duchenne Muscular Dystrophy, DMD, myocytes deteriorate much more rapidly than normal, leading to muscle deterioration and inevitably, death typically by the third decade of life.

Although there are no atomic resolution structures available for the STRs of dystrophin, numerous X-ray and NMR structures of homologous motifs from other members of the spectrin family are available[1,4], and it is now understood that these motifs adopt a triple α -helical bundle in which the three helices are arranged in a zig-zag fashion. This arrangement places the N and C-terminus at opposite ends of this bundle, and tandem arrangement of these motifs produces the long rod-shaped structures of these proteins. However, a further conceptual question arises: since this zig-zag arrangement juxtaposes the last helix of one motif to the first helix of the following, can the arrangement not be viewed equivalently as an alternating series of long and short α -helices? This has been described as the “nested model”[5] and has long ago been proposed for the homologous protein spectrin. In contrast, the original conception of these STRs as independent domains inspired the “independent model”, which has been characterized as having a “beads on a string” behavior[6]. This characterization implies quite a bit of flexibility - indeed in some cases it has been modeled as an entropy spring. On the other hand the nested model implies a relatively more unified and rigid rod.

In dystrophin, the situation is even more complex: the central STR region is interrupted by nonhomologous regions, which together with two other similarly non homologous regions at each end of the rod, were interpreted as some sort of “spacers”. These regions are of varying length (47 to 99 residues; 18 for the mini-hinge), show no homology to each other or to any known protein motif, but are proline rich suggesting difficulty forming any secondary structure. Thus, they were proposed to act as hinge regions, imparting localized flexibility[7]. The implication is that apart from these hinges, the STRs form less flexible, rigid rods – perhaps a unified rigid region, as described in the nested model. On the other hand, protein engineering of dystrophin - mainly in the context of minidystrophin variants developed with an eye toward possible gene therapy products – has considered the STRs as editable units, to be excised or shuffled as need be[8,9]. Certain other biophysical studies have strengthened the case for both independent and nested models in various cases. In particular, AFM single molecule unfolding experiments, in which the rod is stretched to unzip the zig-zags of the bundles has found that in some cases, individual bundles open up one at a time (independent) while in other events multiple STRs unfold at once (concordant with a nested model)[10]. Both situations can be seen in replicate experiments using the same molecule, suggesting that the two models may represent limiting cases of a continuum, and that the truth lies somewhere in between. The molecular basis of the AFM unfolding has been studied by molecular dynamics simulations of two 2-STR rods[11]. In one case (α -actinin2:3) the junction region always unfolded before the rest of the molecule; whereas for the other (β -spectrin1:2), the linker never unfolded at slow pulling rates, but only sometimes at higher pulling rates – showing this heterogeneous behavior.

Recently, studies in the related protein, spectrin, have also suggested that both models may apply, in different regions of the rod[12]. In these studies, STRs linked by regions of high calculated helicity have been found to be more thermodynamically stable, and more cooperative in their folding properties whereas those linked by junctions of low helicity were of low stability, and comparatively independent of each other. These low helicity junctions were then proposed to form “hinges” of spectrin, akin to the *string* in the *beads-on-a-string* model, whereas the high helicity regions were proposed to form interacting units concordant with the nested model. Unfortunately, both motifs studied in the molecular dynamics simulations had very high helicity junctions – 100% by DSC, and with $\langle\langle Pa \rangle\rangle$ values of 1.19 for β -spectrin1:2, and 1.25 for α -actinin2:3 (Indeed they started the simulation in a completely helical conformation), so the behavior of low helicity junctions was unexamined. However,

other molecular dynamics studies have identified coordinated breaking of 4 H-bonds, freeing one entire turn of the junction region α -helix as a key event in AFM unfolding of tandem STRs of spectrin[13].

Thus, we decided to determine if junction helicity was linked to stability in dystrophin as it as in the spectrin study[12], with the aim of determining if these new type of “low junction helicity” hinges might occur in dystrophin, and add to any flexibility imparted by the classical “intervening sequence” hinges[7]. To do this, we identified the 3 highest and 3 lowest calculated helicity junctions and produced the six 2-STR rods spanning these junctions. We then measured the thermodynamic stability and protease sensitivity of these motifs, to determine if these measures correlated with junction properties. Large differences in properties were found, demonstrating that dystrophin also is comprised of low and high stability STRs – but in this case, no correlation with junction helicity was found.

Understanding the biophysical properties of the dystrophin rod – especially the distribution of flexible regions - is crucial, since these regions have been shown to be important for full efficacy in gene replacement studies in mouse models[9]; and since mechanical flexibility has been proposed to be important in dystrophin’s role as a protein shock-absorber for muscle tissue [14]. This is even more important as possible therapeutics for DMD are developed that consist of dystrophin modified in some way (minidystrophins for gene therapy; or *in vivo* edited dystrophins produced by exon skipping therapies). As these move closer to patients (both of these are currently in phase I clinical trials) information about the biophysical consequences of these edits on the properties of the rod will become increasingly important.

Methods

Helical propensity

A scan of the rod region of dystrophin (Genbank NM004006) was conducted to determine which junction regions were predicted to be the most helical, and which the least. For purposes of this study, we interpreted the junction to be an 8 residue window, 4 residues (or a bit more than one turn, if helical) on either side of the STR motif junctions. In assessing the junction regions, we used the alignment set out in the Genbank database, concordant with the original alignment of Koenig and Kunkel. We used two methodologies for measuring predicted helicity: the DSC method[15] and the classic Chou Fasman, CF, method[16]. We chose the DSC method, since this method was previously used in the spectrin work that inspired this study [12]. DSC simply yields a residue by residue category classification (i.e. helical, or other) of the sequence. The overall DSC score for each junction was considered to be the total number of predicted helical residues i.e. a number between 0 and 8 (although the lowest score actually observed was 1, not 0). However, as expected for a highly α -helical protein, many of these regions had very high scores, with 10 of 20 scoring 8/8 helical. – see figure 1. This highlights a drawback of the DSC method in ranking the most and least helical regions: it is a discrete metric, and yields many ties. To deal with this, we also used the classical CF method, since in addition to placing each residue in a secondary structure category, it also yields a metric of helical propensity, $\langle Pa \rangle$, for each residue. We could thus calculate an average propensity for the 8 residue junction as a whole, $\langle \langle Pa \rangle_8 \rangle$. In general, CF and DSC were in agreement, although there were notable exceptions.

In order to begin our study of the rod region (we are in the midst of a comprehensive scan of the entire rod) we initially chose to pursue constructs spanning the highest and lowest calculated helicity junction regions, as reflected by $\langle \langle Pa \rangle_8 \rangle$ scores, since as previously discussed the DSC scores yielded too many ties at the high end. Interestingly these 6 choices all have different DCS scores – the three highest $\langle \langle Pa \rangle_8 \rangle$ junctions have DSC scores of 8/8, 7/8 and 5/8, while the three lowest have 1/8 (the lowest observed), 3/8 and 6/8.

Cloning and protein expression

Two STR fragments spanning the selected junction regions were constructed and purified as previously described[17,18]. Briefly, DNA fragments corresponding to the STR coding regions were PCR amplified, cloned in the expression vector pGEX4-T. Small extensions of 5–8 amino acids were added to the ends of each region to deal with alternative alignments[19] that cause ambiguity about exactly where the STR motifs are delimited, as well as biophysical studies that have shown that in some (but not all) cases such extensions are needed to provide stability to the motifs. The exact boundaries selected are given in Table I. The proteins were then expressed in *E. coli* as GST (Glutathione-S-Transferase) tagged proteins, affording initial affinity purification on glutathione agarose. The GST tag was then released by limited digestion with thrombin, and the dystrophin fragment further purified to homogeneity by a combination of ion-exchange and hydrophobic interaction chromatography. Specific chromatographic conditions were unique to each protein and given in Table I.

All vectors were sequenced and found to be correctly constructed. Purified proteins were > 95% pure by SDS PAGE, and had their molecular masses confirmed by MALDI mass spectrometry. Solution molecular mass and aggregation state was assessed by dynamic light scattering, DLS, using a Malvern Instruments Zetasizer Nano S. For these experiments, proteins were prepared at 1 to 3 mg/ml in PBS, and dust removed by centrifugation at 13000 rpm in a microcentrifuge for 20 minutes immediately before each experiment. Scattered intensity profiles at 630 nm were acquired for 10 second periods and converted to correlograms of the autocorrelation signal as a function of time using the DTS Nano software provided by the manufacturer[20]. These correlograms were then transformed to size distributions using the same software, and both the raw distribution as a function of intensity, as well as the mass weighted distribution, is reported. The mean size is reported as the centroid of the mass distribution, and the variance reported as the width of a fitted Gaussian peak to this distribution. A minimum of 7 runs were performed for each protein, and for comparison, the standard globular proteins lysozyme (chicken, 14 kDa, Sigma L7651), ovalbumin (chicken, 45 kDa, Sigma A7641), and bovine serum albumin (cow, 65 kDa, Sigma A8022) were analyzed under identical conditions.

Thermodynamic stability

Thermodynamic stability was performed as previously described in detail[18]. Briefly, protein solutions at $A_{280} \sim 0.06$ were subject to high concentrations of urea (solvent denaturation) or increasing temperature, and monitored spectroscopically as they unfolded. For urea, both CD and fluorescence were used to monitor the unfolding, while for the thermal scans only CD was used due to the very large sensitivity of the intrinsic fluorescence signal to changes in temperature which swamped the comparatively small perturbations caused by the unfolding transition.

For the CD studies, the spectra were monitored between 240 and 200 nm in a flat 1mm quartz cell in thermal runs, or a 4mm square cell for urea titrations. The signal at 222nm, ϕ_{222} , representative of α -helical secondary structure was then used to follow the transition. The fluorescence spectra were obtained with 280 nm excitation, and emission monitored between 320 and 380 nm. The redshift, as monitored by λ_{mean} (the energy weighed first moment of the spectra) was used as the parameter to reflect unfolding, since this parameter is understood to be correlated to loss of the hydrophobic core[21]. In contrast, intensity changes can be positive, negative - or indeed unchanged - and do not reflect any underlying physical property.

Protein unfolding, as monitored changes in ϕ_{222} or λ_{mean} and fit to a 2-state model as previously described. The urea titrations yield ΔG^0 and $[\text{Urea}]_{\text{mid}}$, whereas the thermal titrations yield ΔH , T_m and C_p . All parameters obtained were by global fitting of $n \geq 3$ independent titrations.

Reversibility was demonstrated in the case of urea unfolding by dilution of the highest [urea] sample, generally 8.4 M, to a point well below the transition, generally 0.84 M, and observation the fluorescence spectra. In the case of thermal transitions, which were monitored by CD, the sample was returned to the beginning temperature, generally 13C, and the CD spectra re-measured.

In some cases, i.e. for D8:9 and especially D16:17, adequate (>90%) reversibility of the thermal unfolding was not obtained. We hypothesized that this was due to prolonged incubation of the unfolded sample at high temperatures due the time required to obtain the unfolded baseline region. Thus, in those cases, we conducted separate experiments in which the protein samples were scanned to various temperatures part way along the transition and immediately (in < 60 seconds) returned to 13C. In most cases we were able to confirm substantial reversibility most of the way through the transition.

Proteolytic stability

Stability was also measured by resistance to a non specific protease, Protienase K, PK, as described[17]. Briefly, a serial geometric dilution of PK in 2.5 µg samples of constant test protein in 10 µl PBS were prepared, and incubated at 37C for 30 minutes in PBS, then immediately denatured and analyzed by SDS-PAGE. Gels were stained by Coomassie Blue, photographed, and the images densitometrically analyzed using Scion image (a Wintel port of NIH Image). The fraction remaining at various PK doses were fit to a first order (exponential decay) model as a function of [PK], and an effective dose that gave 50% degradation under these conditions, PK₅₀, was reported.

To ensure reproducibility, a single bolus of PK was obtained (Worthington Biochemical, New Jersey), and dissolved at 0.10 mg/ml in 25 mM tris Cl pH 7.8/1 mM CaCl₂/50% glycerol and stored at -70C. Aliquots were removed for use on a weekly basis, and stored at -20C. All experiments were conducted in triplicate over a period of at least 3 months, and excellent agreement was obtained, as reported in the modest variance of PK₅₀ values.

Results

Five of the six target proteins were easily expressible in our system; the lone holdout was D7:8 which was difficult to produce. This protein was obtained in moderately low yield only after transferring cultures to 25C after induction, suggesting that *in vivo* folding in *E. coli* is challenging for this variant. However, using this lower temperature strategy, as well as standard conditions detailed above for the others we were able to produce all the target proteins. They all gave molecular masses within 0.5% of expected by MALDI, and exhibited α-helical CD spectra with double minima near 208 and 222 nm as expected, as shown in Figure 2. Proteins were also subject to solution molecular mass determination by dynamic light scattering, DLS, using a Malvern Nanosizer instrument. Protein solutions were tested at concentrations of 1 – 3 mg/ml, and fluctuation spectra acquired for multiple (10 to 20) 10 second runs and analyzed by software (DTS Nano) provided by the manufacture. These spectra were then converted to a correlogram showing the signal autocorrelation coefficient, on a time scale of 0.5 µsec to 10 seconds. The decay of this correlation coefficient reflects the diffusional characteristics of the molecule, and can be converted to a size equivalent sphere diameter that diffuses at an equivalent rate. The correlograms observed for our fragments exhibited decays on two time scales: an initial decay in the 10 microsecond regime; and a smaller tailing component in the millisecond to second range. This suggested that the samples contained molecules of two sizes, with the slow component representing small amounts of aggregate.

Size information is obtained by transforming the correlogram to a size distribution histogram consistent with the correlation values at all time scales. There are two important factors to

consider when interpreting this analysis. The first is that this conversion to size distribution from the time domain most directly produces a distribution reflecting scattering intensity as a function of size. However, scattering intensity is proportional to the square of the molecular weight, and so the sixth power of the size of the molecule. Thus very small amounts of large aggregates can be seen in (or even dominate the) distribution, even if they make up tiny fractions of the overall mass of the sample. For this reason the intensity distribution in Figure 3 is de-emphasized by presenting it as a light grey trace. A more appropriate presentation representative of the state of the molecular cohort is the mass weighted distribution, produced by re-weighting by size⁻³. This is shown in the black traces in Figure 3, and shows that >99% of every sample was found in a single size cohort centered at ~5nm.

The second caveat to consider is that this size distribution is calculated using a sphere-equivalent model that is appropriate for globular proteins, but perhaps less so for these rod shaped proteins. Indeed, we also measured apparent sizes for a few well characterized standard globular proteins, as shown in the bottom panel of figure 3. We can see that the ~ 5 nm size observed for our ~27 kDa proteins falls between the 4.7 nm size observed for Ovalbumin (45 kDa; chicken, Sigma) and 6.7 nm observed for bovine serum albumin (65 kDa, Sigma A-8022). Thus our proteins appear to be diffusing slower than expected for their molecular mass, as expected for rods compared to spheres.

Overall, this demonstrates that the target proteins are largely monomeric in solution and seem to be appropriately folded into non-globular shapes concordant with the expected triple α -helical bundles. The small amount of aggregation seen, evident in the intensity distribution traces (as well as being reflected in the tailing of the correlogram into time regimes longer than 10 μ sec) seems to be a property of STR motifs in general, particularly those of dystrophin. Anecdotal reports of STR aggregation - indeed by scattering as detected by the naked eye as cloudiness - have been provided from the earliest studies of these motifs[22]. However, the determination that these represent < 1% of the mass of the samples under our conditions means that techniques used in our subsequent thermodynamic analyses such as CD and fluorescence spectroscopy that yield signals proportional to concentration alone (i.e. not weighted by size³ as scattering is) will not be significantly affected.

In the thermodynamics of the proteins we start to see some significant differences, as reported in Table 2. Three of the targets, D4:5, D16:17 and D18:19 show well defined two state transitions by both thermal, figure 4, and urea, figure 5 experiments. These motifs appear to act as unified, single domains, concordant with the nested model. In contrast, two targets, D7:8 and D8:9 exhibit very broad curves in these denaturation experiments, with no cooperative transition being seen. Some sort of structure seems to be apparent, since **1.** the initial CD spectra (Figure 2) indicate some α -helical secondary structure, which was lost upon exposure to denaturing conditions; and **2.** a red-shift upon such exposure was seen, which was fully reversible, indicating some burying of the tryptophan residues. However, these fragments do not appear to be very stable. In some cases, the transition was so broad that we were not able to fit these data to the models described due to indeterminacy. This is the case with the D7:8 protein in the thermal experiment, and both D7:8 and D8:9 in the urea denaturation experiments. We were able to analyze the data for thermal denaturation of D8:9, which exhibited the very low value of $\Delta H = -120$ kJ/mol, corresponding to a concomitantly broad transition. This is also in agreement with the PK data, discussed below.

The final target, D6:7, shows what appears to be two-state (i.e. one transition) behavior in thermal denaturation, but shows two transitions in the urea denaturation: a low [urea], low intensity blue-shift transition coupled with loss of about half of the ϕ_{222} signal; followed by a more typical red-shift transition at higher urea and the loss of the remaining ϕ_{222} . While it is tempting to view this as two independent unfoldings of each of the two STRs (i.e. D6, and D7)

of this tandem construct, the fact that the initial transition results in a blueshift is problematic – any unfolding of a hydrophobic core, where the tryptophans are expected to be located, is expected to produce a red-shift due to increased solvent exposure and polarity of the environment around these fluorophores.

In the proteolysis experiments, the various targets exhibited widely varying behaviors, but were in general agreement with the classes described above for the denaturation experiments. The two unstable or misfolded constructs as judged by the denaturation experiments, D7:8 and D8:9, were very sensitive to PK, being 50% destroyed at ~0.15 ng PK to 2.5 μ g target, a ratio of ~16,000:1. The proteins that displayed well defined two state transitions, D4:5, D16:17 and D18:19, as well as the slightly unusual D6:7, were much more resistant to digestion, as shown in Figure 6 and Figure 7 and Table 2. Interestingly there was no evidence of a two state degradation for D6:7, producing intermediate meta-stable products, as has been seen with some other similar STR fragments. Instead, this degraded directly to fragments either too small to be resolved on the gel, or too heterogeneous to produce an abundant, identifiable intermediate. In this respect it is similar to the most stable target by this measure, D4:5, which also was digested without intermediates. In contrast, D16:17 and D18:18 *did* display some intermediate structure – yet their urea and thermal profiles were strictly two state.

Discussion

The main conclusion of this paper is that the dystrophin rod is not a uniform, monolithic whole, but that the STRs that constitute it exhibit a wide range of stabilities. Furthermore, these stabilities, while internally consistent across techniques, do not correlate with the nature of the junction region with respect to calculated helicity, as was the case with the homologous protein, spectrin. A summary of all the parameters reflective of stability is shown in figure 8. This figure is arranged so that the *more folded* or *more stable* state is toward the top; thus the energy parameters are negated: $-\Delta G$ and $-\Delta H$, not ΔG or ΔH , are plotted. The other parameters, T_m and $[\text{Urea}]_{\text{mid}}$, and $\log PK_{50}$ are as reported in table 2. No correlation is seen with either the Chou-Fasman $\langle\langle P\alpha \rangle\rangle_8$ parameter, which we used to select the most and least helical junctions, or the lower resolution DSC metric which was previously used with spectrin. In fact, the only correlation with helicity was a slight negative correlation, in which more helical junctions appear to have less negative free energies of folding, both ΔG and ΔH .

The parameter that varies the most was PK_{50} , which varied over more than 3 orders of magnitude, again with little correlation to helicity. In fact, the most resistant protein was D4:5, which has the lowest predicted helicity by both methods; conversely one of the high helicity targets, D8:9 had the greatest sensitivity to PK. On the other hand, all the data is globally consistent, as the two very sensitive PK targets also had the lowest stability by the thermodynamic measures - both too unstable to measure at all by urea denaturation, and only one of them having a measurable but low energy transition by thermal unfolding.

Protease sensitivity is interesting since this has provided the only empirical evidence for the so-called hinges of dystrophin, taking advantage of the fact that disordered regions are in general more rapidly degraded by proteases than tightly folded globular regions. Proteinase K is an especially suitable protease for use in the detection of flexible or unstable regions since it is much more active toward unfolded proteins than globular proteins[23]. Indeed, we have identified a cryptic hinge between D15 and D16 in a region away from the 4 canonical hinges on the basis of such excess sensitivity[17]. Such hinge regions have been correlated with efficacy of the engineered minidystrophin construct in mouse models of DMD[9]. As well, specific protease sensitivity sites in dystrophin have been linked to various pathologies involving viral infection and heart attack[24,25].

We show here that different motifs of dystrophin exhibit marked differences in stability and folding. Since dystrophin is monomeric, we cannot observe the clustering across lateral dimers as was done in the case of spectrin; but we have demonstrated that these properties vary widely along the dystrophin rod, and that the rod cannot be regarded as a uniform, monolithic whole. Perhaps significantly, D7:8 and D8:9 are the two least stable motifs by all measures, and are obviously adjacent in the native rod. This 7:8:9 region may form an extended flexible region of the rod; however this hypothesis awaits study of larger proteins encompassing this region as a whole.

Why do we care? The rod region has sometimes been considered to be the “boring” part of dystrophin, a monolithic whole that simply connects the two “business ends” of the molecule. In animal experiments, some (but not all) modified dystrophins with relatively large rod deletions were well tolerated and quite effective in reversing DMD like symptoms in mice. However, mice are only mildly affected by dystrophin absence (living a relatively normal lifespan, during which they are quite mobile and reproductively competent, for instance) so the demands on a replacement molecule may be much less stringent than in humans. In humans, naturally occurring rod deletions that do not cause truncation, in general, cause the milder Becker’s muscular dystrophy, BMD – which exhibits a wide range of clinical severity from nearly as bad as DMD to nearly benign. This argues that the specifics of rod region structure and properties are important. We have identified markedly unstable regions (i.e. the 7:8:9 region) that may act as a hinge, and other regions that are very stable; this indicated that the rod is far from uniform as might be inferred from its modular structure.

This is especially relevant for minidystrophin or exon skipping therapies which aim to mimic mild forms of BMD, and produce modified rods with as full functionality as possible. These modifications typically involve editing the rod at STR junctions; this has heretofore not taken into account any differences in STR properties that may have biological relevance. While producing any sort of viable replacement for dystrophin with simple restoration of the reading frame, staving off the rapid and severe clinical progression of DMD is certainly a high priority, we wish also to produce an optimal replacement, and to provide as complete a therapy as possible. Such efforts will undoubtedly have to take into account the differences in properties of different regions of the rod, which we have demonstrated to be substantial in this study.

Acknowledgments

This work was financially supported by the NIH / NIAMS through 5R01AR053970.

References Cited

1. Broderick MJ, Winder SJ. Spectrin, alpha-actinin, and dystrophin. *Adv Protein Chem* 2005;70:203–246. [PubMed: 15837517]
2. Hartwig JH. Actin-binding proteins. 1: Spectrin super family. *Protein Profile* 1995;2:703–800. [PubMed: 7584474]
3. Amann KJ, Renley BA, Ervasti JM. A cluster of basic repeats in the dystrophin rod domain binds F-actin through an electrostatic interaction. *J Biol Chem* 1998;273:28419–28423. [PubMed: 9774469]
4. Broderick MJ, Winder SJ. Towards a complete atomic structure of spectrin family proteins. *J Struct Biol* 2002;137:184–193. [PubMed: 12064945]
5. Calvert R, Kahana E, Gratzner WB. Stability of the dystrophin rod domain fold: evidence for nested repeating units. *Biophys J* 1996;71:1605–1610. [PubMed: 8874034]
6. Heinrich V, Ritchie K, Mohandas N, Evans E. Elastic thickness compressibility of the red cell membrane. *Biophys J* 2001;81:1452–1463. [PubMed: 11509359]
7. Koenig M, Kunkel LM. Detailed analysis of the repeat domain of dystrophin reveals four potential hinge segments that may confer flexibility. *J Biol Chem* 1990;265:4560–4566. [PubMed: 2407739]

8. Watchko J, O'Day T, Wang B, Zhou L, Tang Y, Li J, Xiao X. Adeno-associated virus vector-mediated minidystrophin gene therapy improves dystrophic muscle contractile function in mdx mice. *Hum Gene Ther* 2002;13:1451–1460. [PubMed: 12215266]
9. Harper SQ, Hauser MA, DelloRusso C, Duan D, Crawford RW, Phelps SF, Harper HA, Robinson AS, Engelhardt JF, Brooks SV, Chamberlain JS. Modular flexibility of dystrophin: implications for gene therapy of Duchenne muscular dystrophy. *Nat Med* 2002;8:253–261. [PubMed: 11875496]
10. Bhasin N, Law R, Liao G, Safer D, Ellmer J, Discher BM, Sweeney HL, Discher DE. Molecular extensibility of mini-dystrophins and a dystrophin rod construct. *J Mol Biol* 2005;352:795–806. [PubMed: 16139300]
11. Ortiz V, Nielsen SO, Klein ML, Discher DE. Unfolding a linker between helical repeats. *J Mol Biol* 2005;349:638–647. [PubMed: 15896349]
12. MacDonald RI, Cummings JA. Stabilities of folding of clustered, two-repeat fragments of spectrin reveal a potential hinge in the human erythroid spectrin tetramer. *Proc Natl Acad Sci U S A* 2004;101:1502–1507. [PubMed: 14747656]
13. Paramore S, Ayton GS, Voth GA. Extending a spectrin repeat unit. II: rupture behavior. *Biophys J* 2006;90:101–111. [PubMed: 16227505]
14. Ervasti JM. Dystrophin, its interactions with other proteins, and implications for muscular dystrophy. *Biochim Biophys Acta* 2007;1772:108–117. [PubMed: 16829057]
15. King RD, Saqi M, Sayle R, Sternberg MJ. DSC: public domain protein secondary structure predication. *Comput Appl Biosci* 1997;13:473–474. [PubMed: 9283763]
16. Chou PY, Fasman GD. Empirical predictions of protein conformation. *Annu Rev Biochem* 1978;47:251–276. [PubMed: 354496]
17. Menhart N. Hybrid spectrin type repeats produced by exon-skipping in dystrophin. *Biochim Biophys Acta* 2006;1764:993–999. [PubMed: 16716778]
18. Saadat L, Pittman L, Menhart N. Structural cooperativity in spectrin type repeats motifs of dystrophin. *Biochim Biophys Acta* 2006;1764:943–954. [PubMed: 16603424]
19. Winder SJ, Gibson TJ, Kendrick-Jones J. Dystrophin and utrophin: the missing links! *FEBS Lett* 1995;369:27–33. [PubMed: 7641878]
20. Nobbmann U, Connah M, Fish B, Varley P, Gee C, Mulot S, Chen J, Zhou L, Lu Y, Shen F, Yi J, Harding SE. Dynamic light scattering as a relative tool for assessing the molecular integrity and stability of monoclonal antibodies. *Biotechnol Genet Eng Rev* 2007;24:117–128. [PubMed: 18059629]
21. Royer CA. Probing protein folding and conformational transitions with fluorescence. *Chem Rev* 2006;106:1769–1784. [PubMed: 16683754]
22. Kahana E, Gratzer WB. Minimum folding unit of dystrophin rod domain. *Biochemistry* 1995;34:8110–8114. [PubMed: 7794924]
23. Hilz H, Wieggers U, Adamietz P. Stimulation of proteinase K action by denaturing agents: application to the isolation of nucleic acids and the degradation of 'masked' proteins. *Eur J Biochem* 1975;56:103–108. [PubMed: 1236799]
24. Andreoletti L, Venteo L, Douche-Aourik F, Canas F, Lorin de la Grandmaison G, Jacques J, Moret H, Jovenin N, Mosnier JF, Matta M, Duband S, Pluot M, Pozzetto B, Bourlet T. Active Coxsackieviral B infection is associated with disruption of dystrophin in endomyocardial tissue of patients who died suddenly of acute myocardial infarction. *J Am Coll Cardiol* 2007;50:2207–2214. [PubMed: 18061067]
25. Badorff C, Berkely N, Mehrotra S, Talhouk JW, Rhoads RE, Knowlton KU. Enteroviral protease 2A directly cleaves dystrophin and is inhibited by a dystrophin-based substrate analogue. *J Biol Chem* 2000;275:11191–11197. [PubMed: 10753926]

Figure 1. High-Low choices

Junction helicities were calculated by either the classical Chou-Fasman helical propensity, $\langle\langle Pa \rangle\rangle_8$, or the DSC method, which yields a # helical (out of 8 possible) score. In general the two methods agree, with a few notable exceptions. The selected targets were the junctions with the three highest and three lowest $\langle\langle Pa \rangle\rangle_8$ scores, and are indicated by the solid green squares (highest) or the half-solid red squares (lowest). Above the graph, a schematic diagram of the domain structure of dystrophin is shown, to locate the studied motifs. STRs are shown by the S-shaped structures, the canonical hinges by the wiggly lines, the mini-hinge by a much shorter such line. Rounding out the structure, at the N-terminus is found an actin binding domain, and at the C-terminus, a ZZ-domain, two EF-hands, and a WW domain.

Figure 2. CD and Fluorescence Spectra

Spectral properties of the target proteins. TOP: CD spectra at 25°C in PBS buffer, thin black trace; and as denatured by heating to 93°C (solid red circles) or by addition of urea to 8.4 M (hollow green triangles). BOTTOM: intrinsic fluorescence emission spectra of the target proteins (with excitation at 280 nm), in PBS buffer, thin black trace; or after denaturation by addition of urea to 8.4 M, hollow blue diamonds.

Figure 3. Size by DLS

Aggregation state was determined by Dynamic Light Scattering, DLS. On the left, the correlogram of the DLS signal is shown, with the correlation coefficient shown as a function of time. On the right, the size distribution resulting from this signal is shown. Error bars show ranges of at least 7 measurements. The dark (black) trace shows the distribution as a function of mass, and so reflects the true distribution. However, the initial transformation of the correlogram yields first the distribution as a function of intensity, which is shown in grey for each. Theory tell us that the intensity is proportional to the cube of the molecular size, and so the peaks at the large sizes (i.e. > 50 nm), interpreted as aggregates, reflect a vanishingly small portion (in all cases < 0.1%) of the total amount of protein, as discussed it the text. The lowest

panel contains shows three reference globular proteins: LYS: lysozyme, 14.4 kDa; OVA: ovalbumin, 45 kDa, nm; BSA bovine serum albumin 65 kDa. These are vertically offset from each other for clarity.

Figure 4. Thermal titrations

Proteins in PBS were heated at 1C/min as the CD signal was monitored. Molar ellipticity at 222 nm is presented, hollow circles, and is fit the two state model (red line) described in the text. After heating, the samples were cooled back to 13 C and the spectra re-measured, and presented as a solid symbol at 13C. Three proteins, D7:8, D8:9 and D18:19 were > 95 reversible by this measure. Other proteins were not completely reversible, so we conducted separate experiments to determine how far through the transition reversibility was evident, by scanning up to intermediate temperatures then cooling to 13C; detail in the text. These are indicated by solid symbols at the end of arrow which indicated the high temperature for that measurement.

Figure 5. Urea titrations

Urea denaturation curves were determined by Fluorescence (solid squares and upper traces in each graph) and CD (open squares and lower traces) as described, and fit to appropriate equations as described to determine the thermodynamic parameters ΔG and $[\text{Urea}]_{\text{mid}}$, characterizing each transition. Reversibility of the transition was determined by dilution of the last sample at 8.4M urea to 0.84 M (which is well below the transition in each case) and re-measuring the λ_{mean} ; this is shown by the hollow, yellow circles at 0.84 M.

Figure 6. Proteolysis gels

The target proteins were analyzed for resistance to PK by SDS-PAGE as described. Each assay consists of 2.5 μ g protein and some geometric PK dilution series, with the most concentrated PK to the left as indicated by the triangles above the gels. Standard experiments used a range of 68 to .068 ng and are indicated by larger black triangles. Two proteins, D7:8 and D8:9, exhibited extreme sensitivity, and a more dilute series was also run, PK range from 2.1 to 0.002 ng. These are indicated by the smaller white triangles. This dilution series is such that a 10-fold reduction in PK is achieved in two lanes, and the 31.6 dilution between the regular and low concentration series is such that the left most lane of the diluted series corresponds to the middle lane of the regular series. Each gel also has a standards lane, STD, with markers of 14, 27 and 40 kDa.

Figure 7. Proteolysis rates

Band intensity of the intact proteins were determined by densitometry, and fit to first order exponential decay model. Example gels are shown in figure 6; the different symbols (circle, triangle, square) represent replicate runs of this experiment. The left panel shows this fit on a semilog scale with $\log(\text{fraction remaining})$ vs. $[\text{PK}]$, demonstrating the appropriateness of a first order reaction model used to calculate the proteolytic sensitivity. D7:8 and D8:9 cannot be seen due to the scale used; we sacrificed this in the interest of emphasizing how much more sensitive they are... the fit can still be seen in left panels, where the semilog goes the other way, with a logarithmic x-axis. In this presentation, the PK_{50} can be seen clearly. It can be seen that the sensitivities vary over several orders of magnitude.

Figure 8. Summary graph

The parameters indicative of structural stability are plotted as a function of junction helicity to demonstrate the complete lack of correlation. Parameters include thermodynamic stability such as the ΔG (black squares) and $[Urea]_{mid}$ (blue triangles) from solvent denaturation, ΔH (red circles) and T_m (green pentagons) from thermal denaturation experiments, and the PK protease rate (magenta stars). In all cases, the parameters scale was set so that “more stable” is at the top of the graph i.e. the negatives, $-\Delta G$ (black squares) and $-\Delta H$ (red diamonds), were plotted. Both the DSC (filled symbols) and $\langle P_{\alpha} \rangle$ (open symbols) helicity metrics were examined. (In the right panel, a dotted vertical line was placed at $\langle P_{\alpha} \rangle = 1.05$, the classical CF α -helix cutoff.)

While substantial heterogeneity between the motifs was noted, it appears that the junction helicity is not linked to this variation.

Table 1

Description and characteristics of the target proteins

The amino and C terminal residues of each protein are shown. Italicized residues are artifacts of the restriction sites used for cloning into the expression vector. Bold residues are the first and last residues of the motifs as presented in the Genbank database. Underlined residues are the first and last in an alternative alignment. The 8 residue junction region is listed, as well as the DSC and <<Pa>>g values. Other quality control values such as the MALDI mass spec, sphere-equivalent hydrodynamic diameter by DLS, and CD signal at 222 nm (indicative of helical content) are given, as well as the protein's calculated pI and buffer pH used to purify it by ion exchange chromatography.

terminus	C-terminus	Junction Region	DSC	<<Pa>>g	N	FW	MALDI Da	Δ %	pI	purification buffer	DLS size nm	ϕ_{222}
SQKKRQITVDS	YIFDTLPPV	WLEYQNNI	1	0.981	236	27300	27376	0.28%	5.90	tris pH 8.5	5.31	+/-
SAREKELQTLIFDT	GGLEKTVSLPV	KLEEQMNK	8	1.251	240	27879	27876	-0.01%	5.84	tris pH 8.5	5.04	+/-
SEHCQKLEEQM	EVWACWHELPPV	ALKGGLEK	3	1.028	238	27756	27772	0.06%	5.30	piperazine pH 6.5	5.83	+/-
SARKKALKGGL	QSIQSAQETEPV	TLEEVWAC	7	1.213	234	27089	27108	0.07%	4.78	bistris pH 6.5	5.12	+/-
SVYMTEDMPLEI	EQKNILSEFQPV	RFDRSVEK	6	1.029	238	28006	28028	0.08%	6.19	tris pH 7.9	5.42	+/-
SDRKKRLEEQQ	DLAPGPPV	EIEAQIKD	5	1.218	234	26945	26929	-0.06%	5.01	piperazine pH 6.5	5.32	+/-

Thermodynamics and stability of the target proteins

The thermodynamic parameters derived from solvent (urea) and thermal denaturation, as well as PK sensitivity are presented for each target. All values are shown as the mean $\pm$ the standard deviation of at least 3 independent experiments. For proteins D6:7 and D16:17, only partial reversibility was demonstrated in the thermal denaturation experiments, so the fitted parameter values are subject to interpretation and should best be viewed as estimates; for this reason those results have been *italicized*; see text for details.

protein	Solvent Denaturation			DH kJ/mol	Thermal Denaturation		Cp J/mol K	Proteolysis PK _{S50} nmol					
	DG kJ/mol	[Urea]mM	Tm °C										
D4:5	-32.4	+/-	2.2	4.36	0.03	9	65.0	+/-	0.1	0.5	26.5	+/-	6.2
D6:7	-7.9	+/-	0.5	1.64	0.06	10	57.4	+/-	0.1	1.2	2.8	+/-	0.7
	-25.2	+/-	2.3	3.84	0.07								
D7:8	ND	ND	ND	ND	ND	6	53.8	ND	ND	ND	0.15	+/-	0.03
D8:9	ND	ND	ND	ND	ND	12	69.2	+/-	0.1	0.4	0.14	+/-	0.04
D16:17	-46.3	+/-	1.1	5.07	0.01	7	59.2	+/-	0.1	0.6	5.5	+/-	1.9
D18:19	-10.7	+/-	0.6	2.87	0.05			+/-	0.1	3.0	5.4	+/-	2.4