Published in final edited form as:
Structure. 2012 November 7; 20(11): 1929–1939. doi:10.1016/j.str.2012.08.024.

The Survival Motor Neuron protein forms soluble glycine zipper oligomers

Renee Martin[†], **Kushol Gupta**, **Nisha S. Ninan**[†], **Kay Perry**[*], and **Gregory D. Van Duyne**
Department of Biochemistry & Biophysics, Howard Hughes Medical Institute, Perelman School of Medicine

[†]Graduate Group in Biochemistry & Molecular Biophysics, University of Pennsylvania, Philadelphia, PA 19104

[*]NE-CAT and Department of Chemistry and Chemical Biology, Cornell University, Building 436E, Argonne National Laboratory, 9700 S. Cass Ave., Argonne, IL 60439

Abstract

The Survival Motor Neuron (SMN) protein forms the oligomeric core of a multi-protein complex that functions in spliceosomal snRNP biogenesis. Loss of function mutations in the *SMN* gene cause spinal muscular atrophy (SMA), a leading genetic cause of infant mortality. Nearly half of the known SMA patient missense mutations map to the SMN YG-box, a highly conserved oligomerization domain of unknown structure that contains a $(YxxG)_3$ motif. Here we report that the SMN YG-box forms helical oligomers similar to the glycine zippers found in transmembrane channel proteins. A novel network of tyrosine-glycine packing between helices drives formation of soluble YG-box oligomers, providing a structural basis for understanding SMN oligomerization and for relating defects in oligomerization to the mutations found in SMA patients. These results have important implications for advancing our understanding of SMN function and glycine zipper-mediated helix-helix interactions.

Keywords

SMN; Spinal Muscular Atrophy; oligomerization; snRNP assembly; glycine zipper

Spinal muscular atrophy (SMA) is the most common genetic cause of infant mortality (Markowitz et al., 2012). SMA patients exhibit muscle weakness and poor muscle tone resulting from degeneration of α-motor neurons in the spinal cord (Crawford and Pardo, 1996). In the most severe cases (type I SMA), life expectancy is limited to ~2 years due to difficulties in breathing and poor airway protection. In most SMA patients, deletions or mutations in the *survival motor neuron* (*SMN*) gene are responsible for the disease (Lefebvre et al., 1995).

To whom correspondence should be addressed. vanduyne@mail.med.upenn.edu.
Contact information: Gregory D. Van Duyne, University of Pennsylvania, Department of Biochemistry and Biophysics, 242 Anatomy-Chemistry Building, Philadelphia, PA 19104-6059, Phone: 215-898-3058, Fax: 215-573-4764, vanduyne@mail.med.upenn.edu

The authors declare no conflict of interest.

Most eukaryotes have a single copy of the *SMN* gene, which has been found to be essential in each model organism in which it has been studied. Humans are unusual in having two copies of *SMN*, located in a large inverted duplication on chromosome 5q. The telomeric *SMN1* gene is deleted or mutated in SMA patients, whereas the centromeric *SMN2* gene is generally unaffected and can be present in multiple copies (Lefebvre et al., 1995; Parsons et al., 1998). The two genes code for identical proteins of 294 amino acids, but a silent point mutation in *SMN2* leads to exon 7 skipping during pre-mRNA splicing of most transcripts (Lorson et al., 1999; Monani et al., 1999). The result is a modified and truncated SMN (SMNΔ7), which is rapidly degraded (Burnett et al., 2009). SMA patients therefore have less functional SMN due to homozygous mutations in *SMN1* and the wild-type SMN required for survival comes from low levels of correctly spliced *SMN2* transcripts.

SMN forms the oligomeric core of the multiprotein 'SMN complex' that is found both in the cytoplasm and nucleus of most cells (Paushkin et al., 2002). Although this complex may have multiple roles in different tissues, its only well-characterized function is the ordered assembly of uridine-rich small nuclear ribonucleoprotein particles, or U-snRNPs (Meister et al., 2001; Pellizzoni et al., 2002). During the initial stages of snRNP biogenesis, seven highly conserved Sm proteins are assembled to form a ring around the Sm site of U-snRNAs, resulting in a snRNP core (Will and Lührmann, 2001). In human cells, snRNA transcripts are exported from the nucleus to the cytoplasm, where the SMN complex catalyzes this core assembly reaction. Other steps in snRNP biogenesis include 5'-cap hypermethylation and 3'-end processing of the snRNA, internal snRNA modification, and addition of snRNP-specific proteins (Kiss, 2004). In addition to SMN, seven proteins referred to as 'Gemins', plus the protein Unrip, contribute to the cytoplasmic phase of snRNP biogenesis.

SMN is oligomeric (Liu et al., 1997; Lorson et al., 1998), forming complexes with unusually large hydrodynamic sizes. The SMN complex from HeLa cells, for example, has a ribosome-like sedimentation coefficient of ~40S on sucrose gradients (Paushkin et al., 2002). SMN has three highly conserved domains (Fig. 1A). A short segment near the N-terminus is responsible for binding with high affinity to Gemin2 (Liu et al., 1997). A central Tudor domain recognizes symmetric dimethylarginine (sDMA) modifications in arginine/glycine rich regions of a number of proteins involved in RNA processing, including the Sm proteins (Brahms et al., 2001; Bühler et al., 1999). A C-terminal domain referred to as the 'YG-box' is thought to be primarily responsible for oligomerization (Lorson et al., 1998; Talbot et al., 1997). The YG-box sequence contains three overlapping motifs: GxxxGxxxG, YxxxYxxxY, and SxxxSxxxSxxxT (Fig. 1B). The glycine motif is found in a class of trans-membrane (TM) helical oligomers (Kim et al., 2005), but the tyrosine and serine repeats are not generally found in membrane-spanning protein segments. There are currently no structural models of the SMN YG-box region and no sufficiently similar sequences in available databases to provide reliable predictions.

Amino acid substitutions arising from SMA patient mutations are located primarily in the Tudor and YG-box regions of SMN (summarized by Burghes and Beattie, 2009), with 12 of 25 identified point mutants clustered in the YG-box (Fig. 1B). Some of the YG-box variants have been shown to be defective in oligomerization (Lorson et al., 1998; Pellizzoni et al., 1999) and snRNP assembly (Shpargel and Matera, 2005; Wan et al., 2005), establishing a correlation between loss of oligomerization and a strong SMA phenotype. It is currently unknown what types of oligomers are formed by SMN or how many SMA patient mutations in this region cause defects in SMN's ability to self-associate.

To address these questions, we have established model systems for studying the structural and biochemical properties of SMN and SMN complex oligomers. We find that both SMN

and the isolated YG-box form oligomeric states ranging from dimers to octamers, all of which are smaller than that implied by size-exclusion chromatography experiments. The crystal structure of a dimeric YG-box construct reveals that this region of SMN forms glycine zipper oligomers related to those found in TM helical bundles. Unlike the TM oligomers, however, SMN uses novel tyrosine-glycine packing to stabilize inter-helical interactions, providing a structural explanation for the highly conserved tyrosine and glycine repeats found in diverse SMN orthologs. Alanine scanning mutagenesis of the conserved region of the SMN YG-box leads to a strong correlation between the oligomeric properties of the substituted YG-box and interactions observed in the YG-box dimer structure. We have also analyzed the oligomerization properties of each of the twelve SMA patient YG-box mutations. Although most SMA substitutions in this region affect oligomerization, some have properties similar to the wild-type controls. The results suggest that the SMN YG-box is not simply an oligomerization motif and that the conserved surfaces generated by YG-box helical oligomers are likely to be used directly in mediating interactions that are crucial for SMN's functions.

Results

The SMN-Gemin2 complex forms a distribution of oligomeric states

To determine what types of oligomers are formed by SMN, we examined the hydrodynamic properties of a SMN-Gemin2 complex that lacks the segment coded by exon 5 of *SMN* (Fig. 1A). The SMN Δexon5 (SMNΔ5) variant is a naturally occurring isoform (Gennarelli et al., 1995) that resembles SMN from fly and worm, which lack much of this 32-residue proline-rich region. The SMNΔ5-Gemin2 complex has improved solubility at physiological ionic strength and shows a reduced tendency to aggregate, making it more amenable to solution analyses than the full length SMN-Gemin2 complex. To simplify our biophysical investigations, we chose to study the minimal complex formed by SMNΔ5 and Gemin2 rather than larger complexes containing additional gemins (e.g., Gemin3 and Gemin8) where expression and purification of adequate soluble material has been challenging.

The SMNΔ5-Gemin2 complex elutes near the 670 kDa marker of analytical size-exclusion chromatography (SEC) columns (Fig. 2A). This size estimate implies that ten or more copies of the 59 kDa heterodimer are present in a globular oligomer or that a smaller oligomer adopts an elongated shape and consequently a large Stokes radius (R_s). When SEC analysis of this complex is coupled with multi-angle light scattering detection (SEC-MALS), the weight-average molecular weight (M_w) distribution calculated across the peak indicates that multiple species are present with masses that range from ~100 kDa to ~300 kDa (Fig. 2A). Thus, the sizes of the majority of SMN-Gemin2 oligomers present in solution are smaller than that implied by SEC alone.

To further investigate the nature of the SMN complex oligomers formed, we used sedimentation equilibrium ultracentrifugation (Fig. 2B). As anticipated, the resulting radial distributions of absorbance data could not be described by a single species in solution. The data could be adequately fit by a dimer-tetramer-octamer equilibrium model, which is consistent with the mass distribution observed by SEC-MALS. However, this model is not unique, since other equilibria can also be fit to the data. It is clear, however, that the SMNΔ5-Gemin2 complex adopts multiple oligomeric states in the nM-μM concentration regime and that these oligomers adopt highly extended conformations.

A model system to study YG-box oligomerization

The non-globular and heterogeneous nature of the SMN-Gemin2 complex has been a major obstacle to obtaining structural models. To understand the oligomers formed by SMN, we

wanted to be able to study the YG-box region in isolation. We found that most YG-box fusions with common purification tags were poorly soluble, as were synthetic YG-box peptides. However, several constructs fused to the carboxyl terminus of the maltose binding protein (MBP) were very soluble, allowing us to carry out experiments with the YG-box in the absence of other regions of SMN. Furthermore, we found that we could control the types of oligomers formed by changing the length of the MBP-SMN linker.

As shown in Fig. 3, MBP-YG$^{252–294}$ forms stable tetramers in solution based on SEC-MALS and SE analyses. Small-angle X-ray scattering (SAXS) data also indicate a tetramer of MBP-YG subunits in solution (Fig. S1). As a control for a non-oligomeric fusion, we used the corresponding YG-box sequence from SMNΔ7, the truncated isoform that is the primary product of the *SMN2* gene. MBP-YGΔ7$^{252–282}$ behaves exclusively as a monomer, as expected from previous studies in the context of full-length SMN (Lorson et al., 1998; Pellizzoni et al., 1999). Summaries of the hydrodynamic properties of MBP-YG$^{252–294}$ and MBP-YGΔ7$^{252–282}$ are given in Table S1.

Structure of the SMN YG-box

In addition to serving as a biochemical handle to study SMN oligomerization, the MBP-YG-box fusion provided an opportunity to attempt crystallization of a SMN YG-box oligomer. We therefore purified and characterized a series of MBP-YG fusions, where the MBP-SMN linker was shortened to alter the nature of the coupling between domains (Fig. 3A). Fusions starting from Asp252 through Asp257 formed tetramers or mixtures of oligomers that included tetramers, but did not form diffraction-quality crystals. Fusions from Ala258 to Met263 were either monomeric or formed dimers, and one of the dimeric fusions yielded well-diffracting crystals. We determined the structure of the MBP-YG$^{263–294}$ fusion using molecular replacement of MBP and refined the structure to 1.9 Å resolution. Crystallographic data are summarized in Table 1 and electron density for the refined model is shown in Fig. 4A.

In the MBP-YG$^{263–294}$ structure, the SMN YG-box forms a helical dimer with a right-handed crossing angle (Fig. 4B). The most highly conserved residues in the YG-box are well-ordered in the structure (Met263-Arg281), but the C-terminal 13 residues are disordered. The inter-subunit interactions in the MBP-YG dimer come almost entirely from the YG-box; there are no MBP-MBP contacts and only a small number of contacts between MBP and the dimeric partner YG-box (Fig. 4C). The conserved residues that are not part of the Y-, G-, or S-repeat motifs form hydrophobic contacts that are characteristic of helix-helix interactions. The tyrosine and glycine residues, however, form a novel network of inter-subunit interactions (Fig. 5). Each tyrosine side chain packs against the peptide bond π-electrons preceding the i+3 glycine residue in the opposing helix. Each glycine residue also packs against the corresponding glycine in the opposing helix using the peptide bond that follows C_α. Together, these interactions provide a structural basis for the unusual $(YxxG)_3$ motif. The close interaction of helical glycine residues is similar to that seen in TM helical dimers such as glycophorin (MacKenzie et al., 1997) and in some soluble proteins (Kleiger et al., 2002) but here there are three consecutive close Gly-Gly interactions whereas in GxxxG motif dimers there is normally one close contact with a larger right-handed crossing angle between helices. The conserved serine and threonine residues do not participate directly in the dimer interface. Instead, they form a network of hydrogen bonds within each helix, hydrogen-bonding to the carbonyl oxygens of the i-3 residues (Fig. 5).

The SMN YG-box dimer structure provides an explanation for the variable oligomerization states observed in the MBP-YG truncation series generated for crystallization trials. When the linker is sufficiently long, the YG-box is oligomeric but the flexible nature of the attachment presumably hinders crystallization. When the linker is shortened, steric effects

from MBP favor the formation of smaller oligomers (Fig. 3A). An additional consequence of close MBP-YG coupling is that the YG-box will adopt an appropriate helical phase to make oligomeric self-interactions for only certain fusions, since the YG-box forms a continuous extension of the MBP carboxyl-terminal helix (Fig. 4C). MBP therefore serves to limit the extent of oligomerization for both the tetrameric and dimeric MBP-YG fusions. When longer linkers were employed, a wider range of oligomeric states could be formed that mimic those observed in native SMN. The MBP-SMN$^{229–294}$ fusion, for example, forms oligomers spanning a molecular weight range that corresponds to tetramers through octamers (Fig. S2).

The SMN YG-box forms novel glycine zipper homo-oligomers

The GxxxGxxxG motif in the SMN YG-box sequence is found in a family of transmembrane (TM) helical oligomers referred to as 'glycine zippers' (Kim et al., 2005). Tetrameric, pentameric, and heptameric examples of this family are shown in Figure 6. SMN appears to be the first example of a soluble glycine zipper oligomer, but with an important difference compared to the TM structures. The conserved tyrosine residues of the SMN YG-box play a crucial structural role and, as described below, are for the most part intolerant to substitution. Although tyrosine is generally disfavored in the central region of the membrane, phenylalanine has been found to be statistically over-represented in the i-3 position relative to glycine in TM helices containing the GxxxG motif (Unterreitmeier et al., 2007), suggesting that TM helices may use similar types of aromatic-glycine interactions. Studies on model peptides containing the 'FxxGxxxG' motif have confirmed the strong, stabilizing effect of phenylalanine in this context (Unterreitmeier et al., 2007). To emphasize the important and conserved role of tyrosine in the YxxG repeating motif of the SMN YG-box, we call this oligomerization domain a 'YG-zipper'.

In TM glycine zipper oligomers, the conserved glycine residues (which can also be small side chains such as alanine, serine, or threonine) pack against side chains of an adjacent subunit in the helical bundle. The structural changes that relate the different oligomeric forms (i.e., tetramer to pentamer) are subtle, involving relatively small helical rotations coupled with displacements with respect to the axis of the helical bundle. Thus, the nature of the helix-helix interactions are similar in the different glycine zipper oligomers. The SMN YG-box may be capable of forming higher order oligomers using a similar scheme, but with tyrosine-glycine interactions playing a prominent role. Based on this idea, we constructed a model of a SMN YG-box tetramer in which the conserved YG-box tyrosines pack against the neighboring i+3 glycine residues in the adjacent helices (Fig. S3). As discussed below, glycine zipper helical bundles represent a plausible model for higher order SMN oligomers.

YG-box residues that are important for oligomerization

To test the relative importance of residues in the SMN YG-box for oligomerization, we substituted each amino acid from Gly261 through Asn283 with alanine, purified the respective MBP-YG$_{252–294}$ fusions, and analyzed their oligomeric properties. The results are summarized in Fig. 7A. For those alanine substitutions that replace residues that are present and well-ordered in the YG dimer structure, the oligomeric states are also indicated in Fig. 4B. The Met263, Leu264, Trp267, Tyr268, Gly271, Tyr272, Thr274, and Gly275 alanine substitutions all disrupt oligomerization, resulting in primarily monomeric MBP-YG fusions. Each of these residues is strongly conserved among SMN orthologs (Fig. 1B). Three additional alanine substitutions at Met269, Tyr276, and Tyr277 resulted in weakened oligomerization where the oligomeric species still exists but is in equilibrium with a significant population of monomer (M269A and Y276A) or monomer and dimer (Y277A).

With the exception of Tyr277, the residues sensitive to alanine substitution have clear roles in the YG-box dimer structure (Fig. 4B) and would be predicted to have similar roles in higher order YG-zipper oligomers (Figs. 6 & S3). Met263 and Trp267 from one helix form a hydrophobic cluster with Leu264 from the partner helix, explaining the importance of these side chains. The structural roles of Tyr268, Gly271, Tyr272, Gly275, Tyr276, and Gly279 follow from the Y-G interactions described above and shown in detail in Fig. 5. Loss of either of the first two tyrosine-glycine contacts is sufficient to disrupt oligomer formation, emphasizing the critical roles of these residues. The third pair involving Tyr276 and Gly279 appears to have more flexibility, with the Tyr276 substitution having only a small effect and the Gly279 substitution having no negative effect. As shown in Fig. 4B, the YG helices begin to diverge at this point in the dimer and the Y-G packing is less intimate for Tyr276 and Gly279. Indeed, worm SMN has alanine at the position corresponding to Gly279 and Tyr276 is leucine in the fly and fission yeast orthologs (Fig. 1B).

Thr274 has a unique role in the YG-box structure, being positioned to make multiple interactions. In addition to an intra-helical hydrogen bond to the carbonyl group of Ser270, Thr274 makes van der Waals contact with Tyr272 from the opposing helix, one of the critical Y-G partners near the center of the YG-box. This highly conserved threonine is also positioned to interact with Tyr277. As shown in Fig. 4B, Tyr277 does not participate directly in the helix-helix interface and would be largely solvent-exposed in a YG-box dimer. In the crystal structure, MBP restricts the Tyr277 side chain rotamers that are accessible. Rotation of the Tyr277 side chain about the C_α-C_β bond indicates that a hydrophobic cluster involving Tyr277, Thr274, and Tyr272 can be readily formed. This three-way interaction could explain why both Tyr277 and Thr274 influence oligomerization. The Y277A substitution was also the only YG-box mutant that we studied which produced a substantial amount of a dimeric MBP-YG-box species in the context of the Asp252-Asn294 construct. This observation is consistent with a minor role for Tyr277 in stabilizing a YG-box dimer, but a more prominent role in stabilizing higher order oligomers.

Alanine substitutions of several moderately well-conserved residues in the YG-box do not strongly affect oligomerization. Ile265, Ser266, Met269, Ser270, and His273 are each solvent-exposed in the dimer structure and play no clear role in mediating helix-helix interactions. Of these, only the M269A substitution has a minor effect in destabilizing the MBP-YG tetramer. This group includes two of the serine residues in the conserved YG-box serine motif. Ser270 forms a hydrogen-bond to the backbone carbonyl group of Ser266 on the same helix, but Ser266 does not make a corresponding hydrogen-bond to the MBP-derived residue preceding Met263. Thus, the conserved serines in the YG-box may contribute to local stabilization of the helix, but do not participate directly in oligomerization in either the dimer structure or a glycine zipper tetramer model.

Two less well-conserved residues, Met278 and Phe280, are also not required for oligomerization, consistent with their observed positions in the YG-box structural models. Met278 is solvent exposed, with leucine, glutamine, and glutamic acid represented in this position among the SMN sequences shown in Fig. 1B. Phe280 is positioned at the oligomerization interface, where it makes only minimal van der Waals contacts with Arg281. None of the residues outside of the ordered YG-box helix observed in the crystal structure are sensitive to alanine substitution, including Arg281 through Asn283. Consistent with these results, the Q282A substitution has been previously studied in full-length SMN and was found to be proficient at self-association and in interacting with wild-type SMN (Carrel et al., 2006).

Four of the alanine-substituted MBP-YG-box constructs formed oligomers larger than those observed for the wild-type control. G261A, S270A, G279A, and N283A all formed a

substantial number of species with molecular weights in the 200–350 kDa range in addition to an apparent tetramer at 180 kDa. For the substitutions at Gly261, Ser270, and Gly279, these effects would be consistent with interactions between YG-box oligomeric surfaces that have gained additional hydrophobic character. The effect of the N283A substitution was surprising and suggests that the polar residues at the carboxyl-terminal end of the YG-box also play a role in modulating the oligomeric state(s) of SMN.

In addition to the alanine-substitutions within the conserved region of the SMN YG-box discussed above, we purified and analyzed a series of C-terminal deletions to determine how many residues coded by exon 7 (Gly279-Asn294) are required for stable oligomer formation (Fig. 7B). We found that deletions up to Asn283 were tolerated, with MBP-YG$^{252–282}$ forming stable oligomers similar to the MBP-YG$^{252–294}$ construct. Deletion of Gln282 led to a primarily monomeric MBP-YG species and further truncations yielded high molecular weight aggregates that could not be characterized. Thus, the SMN YG-box requires at least three residues following conserved Gly279 for efficient oligomerization, with no apparent consequences when either of the three is substituted by alanine in the context of full-length exon 7. For correct cellular localization and function, however, it appears that at least seven residues following Gly279 are required, including the moderately well-conserved QNQKE motif starting at Gln282(Carrel et al., 2006; Hua and Zhou, 2004; Zhang et al., 2003).

Source of the oligomerization defect in SMNΔ7

The majority of SMN derived from the *SMN2* gene is translated from a transcript where exon 7 has been removed during pre-mRNA splicing (Gennarelli et al., 1995; Lorson et al., 1999; Monani et al., 1999). The resulting SMNΔ7 contains translated exons 1–6, plus a four residue EMLA sequence derived from the normally non-coding exon 8 (Fig. 7C). SMNΔ7 has been shown to be defective in self-interaction assays (Lorson et al., 1998; Pellizzoni et al., 1999) and shows a reduced hydrodynamic size compared to wild-type SMN by gel filtration (Pellizzoni et al., 1999). This isoform is also defective in snRNP assembly (Shpargel and Matera, 2005; Wan et al., 2005; Winkler et al., 2005), fails to complement loss of wild-type SMN (Carrel et al., 2006; Shpargel and Matera, 2005), and is less stable *in vivo* (Lorson and Androphy, 2000).

As expected, we found that MBP-YG$^{252–278}$-EMLA is defective in oligomerization (Fig. 3B). We were interested in understanding the molecular basis of this defect, which in principle could arise from amino acid substitutions, truncation, or a combination of the two. We therefore tested the oligomerization properties of a series of MBP-YG$^{252–294}$ mutants involving Gly279 through Asn294 (Fig. 7C). The MBP-YG(EMLA) construct contains SMN residues 252–294 with the four substitutions found in the SMNΔ7 variant. This protein is monomeric, with properties identical to that seen for MBP-YGΔ7. The EMLA substitutions are therefore sufficient to disrupt oligomerization in the absence of truncation. The MBP-YG(GMLA) construct retains the native Gly279 and has oligomerization properties similar to wild-type MBP-YG, suggesting that it is the G279E substitution that is primarily responsible for the defect. This interpretation would be consistent with the YG-box structural models, which indicate that the long, charged side chain of glutamic acid would not be readily accommodated in the oligomeric interface. Finally, we tested the MBP-YG$^{252–294}$ G279E point mutant and confirmed that it was monomeric. These results, when combined with the observation that C-terminal truncations up to Asn283 do not disrupt oligomerization of the YG-box, indicate that the G279E substitution is primarily responsible for defective oligomerization in SMNΔ7.

SMA patient mutations and YG-box oligomerization

Nearly half of the missense mutations identified in SMA patients are located in the SMN YG-box (reviewed in Burghes and Beattie, 2009 and listed in Table 2). The type I mutations Y272C and G279V have been shown to be defective in their ability to self-interact in a GST-pulldown assay, whereas the type III mutations S262I and T274I show milder reductions in self-association (Lorson et al., 1998). The Y272C mutant was also shown to be defective in oligomerization based on SEC analysis of purified protein (Pellizzoni et al., 1999). Neither the Y272C or G279V mutants rescue SMN depletion in zebrafish (Carrel et al., 2006) and both have sharply reduced half-lives (Burnett et al., 2009). Half of the substitutions shown in Table 2 (M263R, Y272C, M263T, S266P, T274I, and G275S) would be predicted to be defective in oligomerization based on the YG-box structure and the results of alanine scanning mutagenesis. On the other hand, the H273R substitution would be predicted to have little or no effect on oligomerization.

To test these predictions and to provide insight into the remaining patient mutations, we purified MBP-YG$^{252–294}$ fusions for each of the amino acid substitutions in Table 2 and determined their oligomeric states using SEC-MALS. The results are summarized in Table 2 and Fig. 8. With one exception, each of the mutations predicted to disrupt oligomerization was found to be monomeric. The T274I mutant showed only a minor oligomerization defect, with a reduction in the amount of tetramer and a corresponding increase in the amount of monomer observed on the sizing column. The T274I substitution replaces the threonine hydroxyl group with the hydrophobic ethyl group of isoleucine. Thus, the intra-helical hydrogen bond made by Thr274 would be lost (Fig. 5). However, the Thr274 methyl group would remain in the isoleucine substitution, likely preserving the hydrophobic contacts made to Tyr272 in the adjacent helix. The defect in oligomerization observed for the smaller T274A substitution suggests that these contacts play an important energetic role. These properties are consistent with the findings ofLorson et al. (1998), who showed that the T274I substitution largely retains the ability to self-associate and are consistent with recent studies involving the corresponding T205I substitution in fly SMN, which also retains the ability to self-associate (Praveen et al., 2012). As expected, the H273R mutant has properties similar to that of the wild-type YG-box. His273 is solvent-exposed and does not participate in the oligomerization interface.

Three patient mutations (L260S, S262I, and S262G) are located immediately N-terminal to Met263, the first YG-box residue in our structural model. If we assume that the YG-box helix extends several residues in that direction, then Leu260 would be located at the oligomerization interface and Ser262 would be on a solvent-exposed surface of the helix. Assuming that this is the case, the L260S substitution would be expected to be destabilizing and the S262 substitutions should have milder effects. This explanation is consistent with the observed oligomerization properties (Table 2) and with a previous report that SMN S262I is only mildly impaired in self-association (Lorson et al., 1998), but structural investigations that include more N-terminal residues in the YG-box, along with biochemical analyses of these upstream residues will be required to confirm this interpretation.

The patient mutations at Gly279 are particularly interesting. We find that in the context of the MBP-YG model system, G279V is defective at oligomerization, while G279C is capable of forming oligomers that are larger than the tetramers observed for the wild-type sequence. Modeling indicates that the bulky valine substitution at Gly279 would force Tyr276 to adopt an alternative rotamer, most likely resulting in loss of a productive interaction. A cysteine substitution should be better tolerated in this position, since greater flexibility exists in side chain positioning compared to valine. However, it is not clear why larger oligomers form for G279C or for the G279A substitution, which has similar properties.

Discussion

Our finding that the SMN YG-box forms glycine zipper oligomers provides important new insights into the nature of SMN oligomerization and raises a number of intriguing questions. The existing structural models of glycine zipper oligomers are all found in membrane proteins that function as channels (Kim et al., 2005). SMN appears to be the first example of a soluble protein that uses this class of homo-oligomerization motif. In view of the known structures of glycine zipper helical bundles, two distinct models for SMN oligomerization based on YG zippers can be envisioned. In one model, SMN could form symmetric helical bundles similar to those shown in Figure 6. This would imply that the SMN YG-box oligomerizes promiscuously, forming YG-zipper dimers, tetramers, and possibly other oligomeric forms. In an alternative model, higher order multimers could arise from weaker interactions between stable SMN dimers. Shape reconstructions from SAXS data for the MBP-YG$^{252–294}$ construct do not readily distinguish between these models (Fig. S1).

Aside from SMN, the $(YxxG)_3$ motif is found in only three other characterized human proteins: keratin-associated protein (Genbank NP_853650), the gap junction protein occludin (AAB00195), and translation initiation factor EIF4B (NP_001408). In the first two examples, the motif exists in a low-complexity region rich in tyrosine and glycine (this is also the case for several hypothetical proteins). Some additional human proteins also contain this motif, but without the final glycine. The $(YxxG)_3$ repeat in EIF4B was noted earlier by Talbot *et al.* (1997), who also identified the motif in a small number of RNA-binding proteins from diverse organisms. Thus, unlike the glycine zipper, which is highly represented in membrane proteins (Kim et al., 2005; Senes et al., 2000), the YG-zipper motif does not appear to be widely used as a biological oligomerization strategy for soluble proteins. This suggests that the nature of SMN oligomers may serve a unique functional or architectural role in the cell.

The EIF4B YG zipper motif occurs with the sequence $(YRDG)_3$, which in the context of the SMN YG-box dimer structure would place charged arginine and aspartic acid side chains in solvent-exposed positions where they could interact to stabilize a helical YG-zipper based dimer. Although it remains to be shown that EIF4B forms YG zipper dimers using this motif, it has been reported that a 100-residue region that includes the $(YRDG)_3$ sequence mediates dimerization of the protein and that this motif is required for oligomer formation (Méthot et al., 1996).

An important future goal will be to understand the nature of the interactions that give rise to the oligomeric heterogeneity found in SMN complexes. In addition to serving as a self-interaction module, the conserved surface of the SMN YG-box dimer may also be used as an interaction site for binding of other components in the SMN complex. Indeed, a construct containing the C-terminal 53 residues of SMN (i.e, only ten residues larger than that contained in our MBP-YG$^{252–294}$ construct) has been reported to be sufficient for interaction with Gemin8 (Otter et al., 2007). It is also tempting to consider that the unusual oligomeric properties of SMN are related to its role in maintaining the integrity of Cajal bodies and Gems, given that self-assocation appears to be a common feature of major nuclear body proteins (Hebert and Matera, 2000).

SMA patient mutations mapping to the SMN YG-box result in a range of oligomeric properties. For substitutions such as H273R, it seems unlikely that defective oligomerization *in vivo* is responsible for the SMA phenotype. In these cases as well as the YG-box substitutions with intermediate effects on oligomerization (e.g., T274I), the altered SMN subunits would be expected to form mixed oligomers with wild-type SMN arising from *SMN2*. SMN complexes in the cells of SMA patients with these YG-box substitutions are

therefore likely to contain both wild-type and mutant proteins, with ratios that depend on the *SMN2* copy number, cellular levels of the given mutant, and the energetic contribution to oligomer formation for the SMN variant vs. wild-type SMN. The importance of such heteromeric SMN complex formation in the context of SMA has been previously discussed (Burghes and Beattie, 2009) and a demonstration of this effect was shown by allelic complementation in transgenic mice of *SMN2* by *SMN1* containing either the A2G or A111G mutations (Monani et al., 2003; Workman et al., 2009).

The finding that some SMA patient YG-box variants are competent for oligomerization raises the question of what functional defects could be responsible for the SMA phenotypes. If we assume that protein levels are not reduced for these mutants (i.e. lower protein stability or alterations in splicing of the mutant allele are not responsible for the SMA phenotype), one explanation could be that the YG-box region of SMN also functions as an interaction surface and one or more interactions is compromised with YG-box substitutions such as H273R. Indeed, the fly T205I substitution (corresponding to the SMA patient T274I variant) has been shown to be stable *in vivo*, suggesting a functional defect responsible for the mild SMA phenotype observed that may not be related to oligomerization or stability (Praveen et al., 2012).

Part of the complexity in relating SMN oligomerization to functional roles *in vivo* comes from not knowing what all of those roles are and how they may differ in specialized cell types. Even in the case of snRNP assembly, the oligomeric state of SMN required for the core assembly steps is not known. It is possible that different oligomeric forms are involved in core assembly, snRNA 3'-processing and cap hypermethylation, nuclear import, nuclear shuttling, and maintaining the integrity of gems and Cajal bodies. The YG zipper model for SMN oligomerization offers intriguing possibilities to consider what these oligomeric forms might be and to direct future investigations to understand both cytoplasmic and nuclear roles of the SMN complex.

Experimental Procedures

Protein constructs

SMNΔ5-Gemin2 complex was produced by co-expression of SMNΔ5 from pCDFDuet (Novagen) and Gemin2 from pETDuet (Novagen) as a C-terminal fusion to the Mxe intein (NEB) containing a hexahistidine tag in strain BL21(DE3) at 37°C with IPTG induction. The complex was purified using Ni-NTA (Qiagen) and chitin (NEB) beads, according to vendor protocols. Eluted protein was further purified using MonoQ anion-exchange chromatography (GE Healthcare) using a sodium chloride gradient at pH 8 and gel filtration using a Superdex 200 (GE Healthcare) column in 20 mM TrisHCl, pH 7.5, 400 mM NaCl, 5 mM DTT. SMNΔ5-Gemin2 complex was > 95% pure and contained a 1:1 stoichiometric ratio of SMN:Gemin2 based on Coomassie-stained SDS-PAGE.

Most MBP-YG-box constructs were expressed as C-terminal Mxe-His_6 fusions in pET21a. C-terminal deletions of the MBP-YG-box fusions were constructed by insertion of stop codons at the appropriate locations. For all MBP fusions, the wild-type *E. coli* MBP sequence lacking the C-terminal lysine residue was used and proteins were expressed in BL21(DE3) cells at 18°C with IPTG induction. Constructs containing the Mxe-His_6 tag were purified using Ni-NTA and chitin beads using vendor protocols. MBP fusions lacking the Mxe-His_6 tag were purified on amylose beads (NEB) according to the vendor's protocol. All MBP fusions were further purified on a MonoQ column followed by gel filtration using a Superdex 200 column equilibrated in 20 mM TrisHCl, pH 7.4, 300 mM NaCl and 10 mM 2-mercaptoethanol. MBP-YG fusions were > 98% pure as judged by Commassie-stained SDS-PAGE.

Size-Exclusion Chromatography in-line with Multi-Angle Light Scattering (SEC-MALS)

Samples were analyzed at room temperature using a Superdex 200 10/300 analytical size-exclusion column (GE Healthcare) in-line with a DAWN HELEOS 18-angle multi-angle light scattering (MALS) detector and refractive interferometer (Wyatt Technology Corp.). Most experiments were conducted by injecting 1 nmol of MBP-YG fusion (45 μg) or SMNΔ5-Gemin2 complex (60 μg) onto the column. Protein concentrations of eluate typically ranged from 10–50 μg/ml within the ranges where M_w values could be calculated. SEC-MALS analyses of complexes containing SMA patient mutations were carried out by injecting nearly identical quantities of protein to facilitate meaningful comparison of M_w values and oligomeric states.

Analytical Ultracentrifugation

Sedimentation equilibrium experiments were performed with an XL-A analytical ultracentrifuge (Beckman) and a TiAn60 rotor with charcoal-filled epon centerpieces and quartz windows. Data were measured at 4°C after experimental verification that equilibrium had been reached. The radial absorbance data were globally fit with strict mass conservation using the best fit model based on good model statistics and visual criteria using SEDPHAT (Vistica et al., 2004).

Structure Determination

Crystalline plates of MBP-YG$^{263–294}$ grew after two days at 21°C by hanging drop vapor diffusion in 0.1 M TrisHCl, pH 8.0, 14% (w:v) PEG 3350, 10 mM $CaCl_2$, 0.1M KCl, 0.1M $(NH4)_2SO_4$ and 18% ethylene glycol and were flash frozen directly in liquid nitrogen. High resolution diffraction data were measured at the APS NE-CAT 24-ID-C beam line. Starting phases were obtained at 3Å by molecular replacement of MBP residues 5–369 (PDB 1OMP). Clear electron density for the conserved region of the YG-box was visible from the initial $2F_o$-F_c map. Iterative refinement with REFMAC (Murshudov et al., 1997) and model building with COOT (Emsley and Cowtan, 2004) converged to a final model at 1.9 Å with R_{work} and R_{free} values of 0.203 and 0.247, respectively. A composite omit map was generated to check the final model, which includes residues 1–369 of MBP and residues 263–281 of SMN. SMN residues 282–294 are disordered. Ninety-two percent of residues lie in the most favored region of a Ramachandran plot, and the remaining eight percent are in additional allowed regions.

Data deposition

Coordinates of the MBP-YG$^{263–294}$ structure have been deposited in the Protein Data Bank (code 4GLI).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Bill DeGrado for his insights into helical oligomers, Kathryn Sarachan for helpful discussions, and Robert Sharp for technical assistance. The APS NE-CAT beamline is supported by grants from the NCRR (2P41RR008630-17) and NIGMS (9 P41 GM103622-17). G.D.V. is an Investigator of the Howard Hughes Medical Institute.

Abbreviations

SMN	Survival Motor Neuron
SMA	Spinal Muscular Atrophy

References

Brahms H, Meheus L, de Brabandere V, Fischer U, Lührmann R. Symmetrical dimethylation of arginine residues in spliceosomal Sm protein B/B' and the Sm-like protein LSm4, and their interaction with the SMN protein. RNA. 2001; 7:1531–1542. [PubMed: 11720283]

Bühler D, Raker V, Lührmann R, Fischer U. Essential role for the tudor domain of SMN in spliceosomal U snRNP assembly: implications for spinal muscular atrophy. Hum Mol Genet. 1999; 8:2351–2357. [PubMed: 10556282]

Burghes AHM, Beattie CE. Spinal muscular atrophy: why do low levels of survival motor neuron protein make motor neurons sick? Nat Rev Neurosci. 2009; 10:597–609. [PubMed: 19584893]

Burnett BG, Muñoz E, Tandon A, Kwon DY, Sumner CJ, Fischbeck KH. Regulation of SMN protein stability. Mol Cell Biol. 2009; 29:1107–1115. [PubMed: 19103745]

Carrel TL, McWhorter ML, Workman E, Zhang H, Wolstencroft EC, Lorson C, Bassell GJ, Burghes AHM, Beattie CE. Survival motor neuron function in motor axons is independent of functions required for small nuclear ribonucleoprotein biogenesis. J Neurosci. 2006; 26:11014–11022. [PubMed: 17065443]

Crawford TO, Pardo CA. The neurobiology of childhood spinal muscular atrophy. Neurobiol Dis. 1996; 3:97–110. [PubMed: 9173917]

Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr. 2004; 60:2126–2132. [PubMed: 15572765]

Gennarelli M, Lucarelli M, Capon F, Pizzuti A, Merlini L, Angelini C, Novelli G, Dallapiccola B. Survival motor neuron gene transcript analysis in muscles from spinal muscular atrophy patients. Biochem Biophys Res Commun. 1995; 213:342–348. [PubMed: 7639755]

Hebert MD, Matera AG. Self-association of coilin reveals a common theme in nuclear body localization. Mol Biol Cell. 2000; 11:4159–4171. [PubMed: 11102515]

Hua Y, Zhou J. Survival motor neuron protein facilitates assembly of stress granules. FEBS Lett. 2004; 572:69–74. [PubMed: 15304326]

Kim S, Jeon T, Oberai A, Yang D, Schmidt JJ, Bowie JU. Transmembrane glycine zippers: physiological and pathological roles in membrane proteins. Proc Natl Acad Sci U S A. 2005; 102:14278–14283. [PubMed: 16179394]

Kiss T. Biogenesis of small nuclear RNPs. J Cell Sci. 2004; 117:5949–5951. [PubMed: 15564372]

Kleiger G, Grothe R, Mallick P, Eisenberg D. GXXXG and AXXXA: common alpha-helical interaction motifs in proteins, particularly in extremophiles. Biochemistry. 2002; 41:5990–5997. [PubMed: 11993993]

Lefebvre S, Bürglen L, Reboullet S, Clermont O, Burlet P, Viollet L, Benichou B, Cruaud C, Millasseau P, Zeviani M, et al. Identification and characterization of a spinal muscular atrophy-determining gene. Cell. 1995; 80:155–165. [PubMed: 7813012]

Liu Q, Fischer U, Wang F, Dreyfuss G. The spinal muscular atrophy disease gene product SMN its associated protein SIP1 are in a complex with spliceosomal snRNP proteins. Cell. 1997; 90:1013–1021. [PubMed: 9323129]

Lorson CL, Androphy EJ. An exonic enhancer is required for inclusion of an essential exon in the SMA-determining gene SMN. Hum Mol Genet. 2000; 9:259–265. [PubMed: 10607836]

Lorson CL, Hahnen E, Androphy EJ, Wirth B. A single nucleotide in the SMN gene regulates splicing and is responsible for spinal muscular atrophy. Proc Natl Acad Sci U S A. 1999; 96:6307–6311. [PubMed: 10339583]

Lorson CL, Strasswimmer J, Yao JM, Baleja JD, Hahnen E, Wirth B, Le T, Burghes AH, Androphy EJ. SMN oligomerization defect correlates with spinal muscular atrophy severity. Nat Genet. 1998; 19:63–66. [PubMed: 9590291]

MacKenzie KR, Prestegard JH, Engelman DM. A transmembrane helix dimer: structure and implications. Science. 1997; 276:131–133. [PubMed: 9082985]

Markowitz JA, Singh P, Darras BT. Spinal muscular atrophy: a clinical and research update. Pediatr Neurol. 2012; 46:1–12. [PubMed: 22196485]

Meister G, Bühler D, Pillai R, Lottspeich F, Fischer U. A multiprotein complex mediates the ATP-dependent assembly of spliceosomal U snRNPs. Nat Cell Biol. 2001; 3:945–949. [PubMed: 11715014]

Méthot N, Song MS, Sonenberg N. A region rich in aspartic acid, arginine, tyrosine, and glycine (DRYG) mediates eukaryotic initiation factor 4B (eIF4B) self-association and interaction with eIF3. Mol Cell Biol. 1996; 16:5328–5334. [PubMed: 8816444]

Monani UR, Lorson CL, Parsons DW, Prior TW, Androphy EJ, Burghes AH, McPherson JD. A single nucleotide difference that alters splicing patterns distinguishes the SMA gene SMN1 from the copy gene SMN2. Hum Mol Genet. 1999; 8:1177–1183. [PubMed: 10369862]

Monani UR, Pastore MT, Gavrilina TO, Jablonka S, Le TT, Andreassi C, DiCocco JM, Lorson C, Androphy EJ, Sendtner M, et al. A transgene carrying an A2G missense mutation in the SMN gene modulates phenotypic severity in mice with severe (type I) spinal muscular atrophy. J Cell Biol. 2003; 160:41–52. [PubMed: 12515823]

Murshudov G, Vagin A, Dodson E. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D Biol Crystallogr. 1997; 53:240–255. [PubMed: 15299926]

Otter S, Grimmler M, Neuenkirchen N, Chari A, Sickmann A, Fischer U. A comprehensive interaction map of the human survival of motor neuron (SMN) complex. J Biol Chem. 2007; 282:5825–5833. [PubMed: 17178713]

Parsons DW, McAndrew PE, Iannaccone ST, Mendell JR, Burghes AH, Prior TW. Intragenic telSMN mutations: frequency, distribution, evidence of a founder effect, and modification of the spinal muscular atrophy phenotype by cenSMN copy number. Am J Hum Genet. 1998; 63:1712–1723. [PubMed: 9837824]

Paushkin S, Gubitz AK, Massenet S, Dreyfuss G. The SMN complex, an assemblyosome of ribonucleoproteins. Curr Opin Cell Biol. 2002; 14:305–312. [PubMed: 12067652]

Pellizzoni L, Charroux B, Dreyfuss G. SMN mutants of spinal muscular atrophy patients are defective in binding to snRNP proteins. Proc Natl Acad Sci U S A. 1999; 96:11167–11172. [PubMed: 10500148]

Pellizzoni L, Yong J, Dreyfuss G. Essential role for the SMN complex in the specificity of snRNP assembly. Science. 2002; 298(80):1775–1779. [PubMed: 12459587]

Praveen K, Wen Y, Matera AG. A Drosophila Model of Spinal Muscular Atrophy Uncouples snRNP Biogenesis Functions of Survival Motor Neuron from Locomotion and Viability Defects. Cell Rep. 2012; 1:624–631. [PubMed: 22813737]

Senes A, Gerstein M, Engelman DM. Statistical analysis of amino acid patterns in transmembrane helices: the GxxxG motif occurs frequently and in association with betabranched residues at neighboring positions. J Mol Biol. 2000; 296:921–936. [PubMed: 10677292]

Shpargel KB, Matera AG. Gemin proteins are required for efficient assembly of Sm-class ribonucleoproteins. Proc Natl Acad Sci U S A. 2005; 102:17372–17377. [PubMed: 16301532]

Talbot K, Ponting CP, Theodosiou AM, Rodrigues NR, Surtees R, Mountford R, Davies KE. Missense mutation clustering in the survival motor neuron gene: a role for a conserved tyrosine and glycine rich region of the protein in RNA metabolism? Hum Mol Genet. 1997; 6:497–500. [PubMed: 9147655]

Unterreitmeier S, Fuchs A, Schäffler T, Heym RG, Frishman D, Langosch D. Phenylalanine promotes interaction of transmembrane domains via GxxxG motifs. J Mol Biol. 2007; 374:705–718. [PubMed: 17949750]

Vistica J, Dam J, Balbo A, Yikilmaz E, Mariuzza RA, Rouault TA, Schuck P. Sedimentation equilibrium analysis of protein interactions with global implicit mass conservation constraints and systematic noise decomposition. Anal Biochem. 2004; 326:234–256. [PubMed: 15003564]

Wan L, Battle DJ, Yong J, Gubitz AK, Kolb SJ, Wang J, Dreyfuss G. The survival of motor neurons protein determines the capacity for snRNP assembly: biochemical deficiency in spinal muscular atrophy. Mol Cell Biol. 2005; 25:5543–5551. [PubMed: 15964810]

Will CL, Lührmann R. Spliceosomal UsnRNP biogenesis, structure and function. Curr Opin Cell Biol. 2001; 13:290–301. [PubMed: 11343899]

Winkler C, Eggert C, Gradl D, Meister G, Giegerich M, Wedlich D, Laggerbauer B, Fischer U. Reduced U snRNP assembly causes motor axon degeneration in an animal model for spinal muscular atrophy. Genes Dev. 2005; 19:2320–2330. [PubMed: 16204184]

Workman E, Saieva L, Carrel TL, Crawford TO, Liu D, Lutz C, Beattie CE, Pellizzoni L, Burghes AHM. A SMN missense mutation complements SMN2 restoring snRNPs and rescuing SMA mice. Hum Mol Genet. 2009; 18:2215–2229. [PubMed: 19329542]

Zhang HL, Pan F, Hong D, Shenoy SM, Singer RH, Bassell GJ. Active transport of the survival motor neuron protein and the role of exon-7 in cytoplasmic localization. J Neurosci. 2003; 23:6627–6637. [PubMed: 12878704]

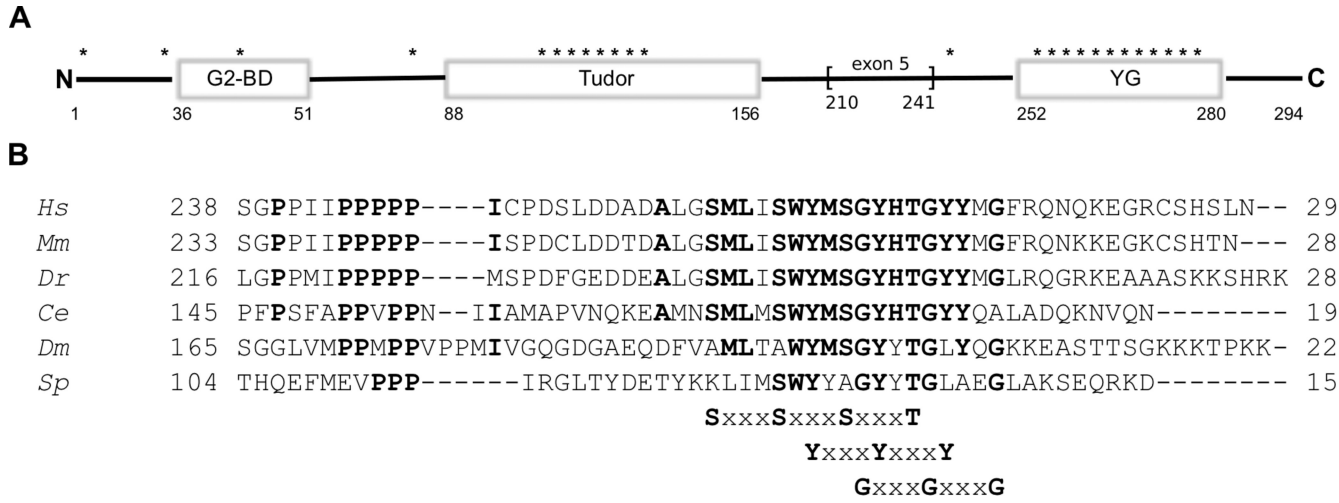

Figure 1.

Conserved regions of the SMN protein. a) Domain structure of SMN. The conserved Gemin2-binding, Tudor, and YG-box domains are indicated. Asterisks indicate the approximate locations of amino acid changes resulting from reported SMA patient mutations. b) Alignment of SMN YG-box sequences from diverse organisms. Residues identical in four out of the six sequences are in boldface type. The S, Y, and G repeat motifs are indicated. Hs=human, Mm=mouse, Dr=zebrafish, Ce=worm, Dm=fly, Sp=fission yeast.

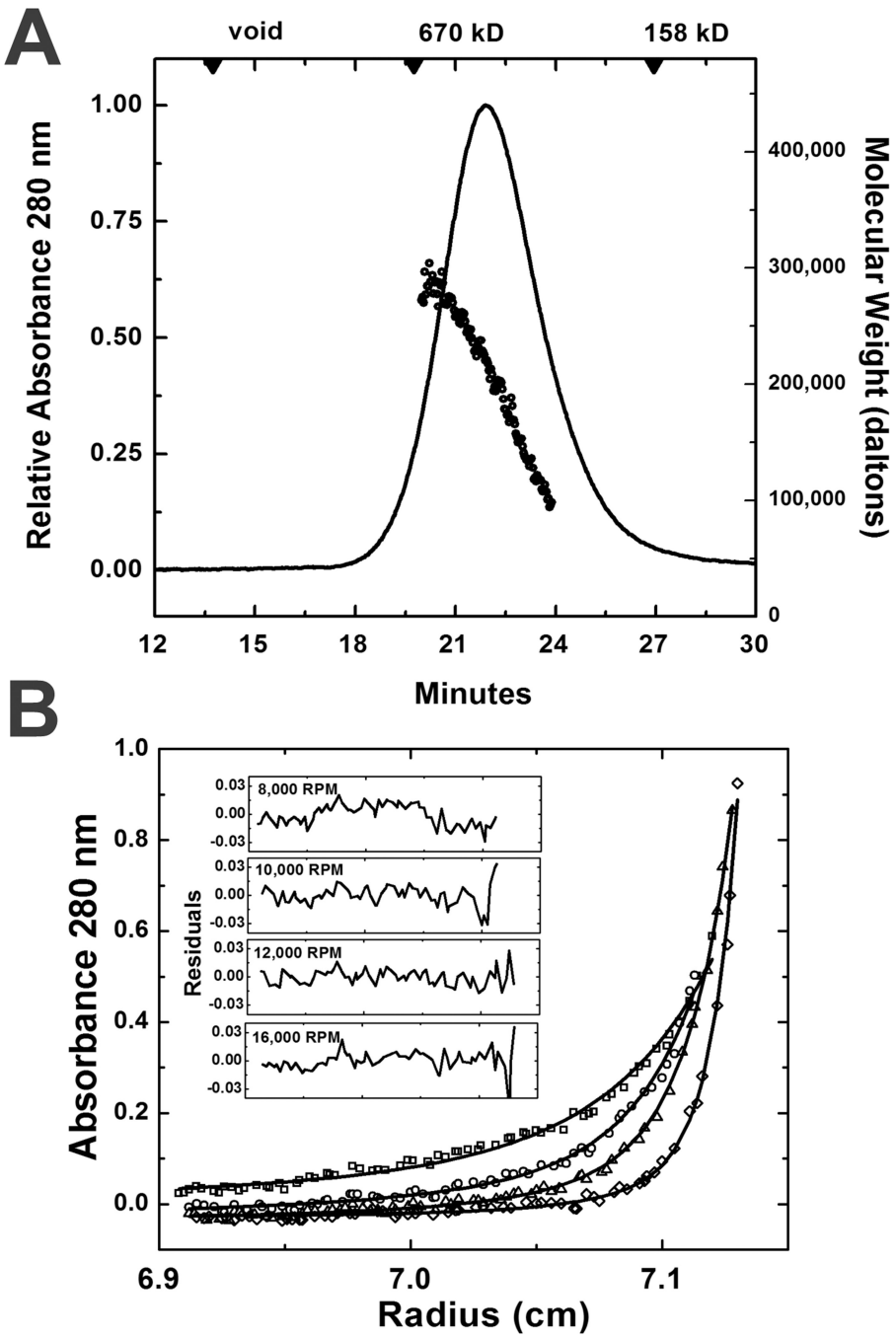

Figure 2.

Oligomerization of the SMNΔ5-Gemin2 complex. A) Size-exclusion chromatography (SEC) of SMNΔ-Gemin2, coupled with in-line multi-angle light scattering detection (SEC-MALS). The weight-averaged molecular weight (M_w) of the majority of SMN complex ranges from 100–300 kDa. The elution position of the 670 kDa thyroglobulin marker is indicated for reference. B) Sedimentation equilibrium carried out at four rotor speeds and two complex concentrations (3.0 and 4.3 μM) at 4°C. Radial absorbances for the 4.3 μM concentration are shown. The data cannot be fit well to a single species. The fit shown is for a dimer-tetramer-octamer equilibrium, with K_{4-8}= 3 μM and K_{2-4}=0.4 μM.

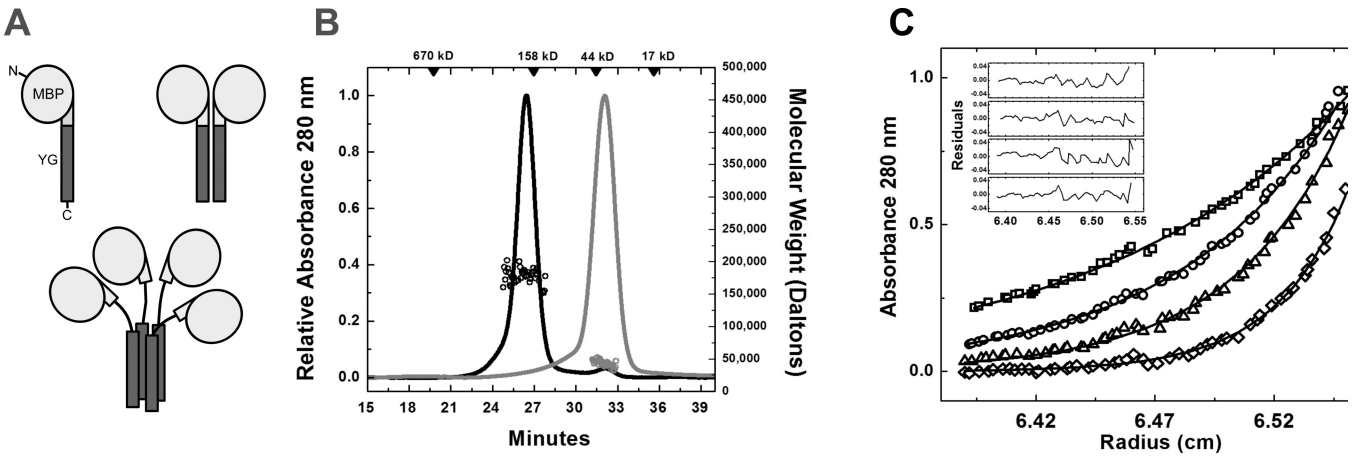

Figure 3.

The MBP-YG-box model system (see also Figs. S1 & Table S1). A) Schematic of MPB-YG-box fusions. Tightly coupled fusions with a small number of SMN residues between the MBP C-terminus and the conserved region of the YG-box are restricted to be monomers or dimers. Longer linkers allow higher oligomerization states. B) SEC-MALS of MBP-$YG^{252–294}$ (dark curve) and MBP-$YG\Delta7^{252–282}$ (light curve). The wild-type YG-box (46 kDa) forms stable tetramers (experimental M_w=183 kDa) but the corresponding construct containing the sequence from SMNΔ7 is monomeric (M_w=42 kDa). M_w is the average weight-averaged 25 molecular weight across the center of the peak. C) Sedimentation equilibrium of MBP-$YG^{252–294}$ carried out at two concentrations (6.3 and 12.6 μM) and four rotor speeds (the 6.3 μM concentration is shown). The fit shown is for a single species tetramer model with molecular weight = 190 kDa.

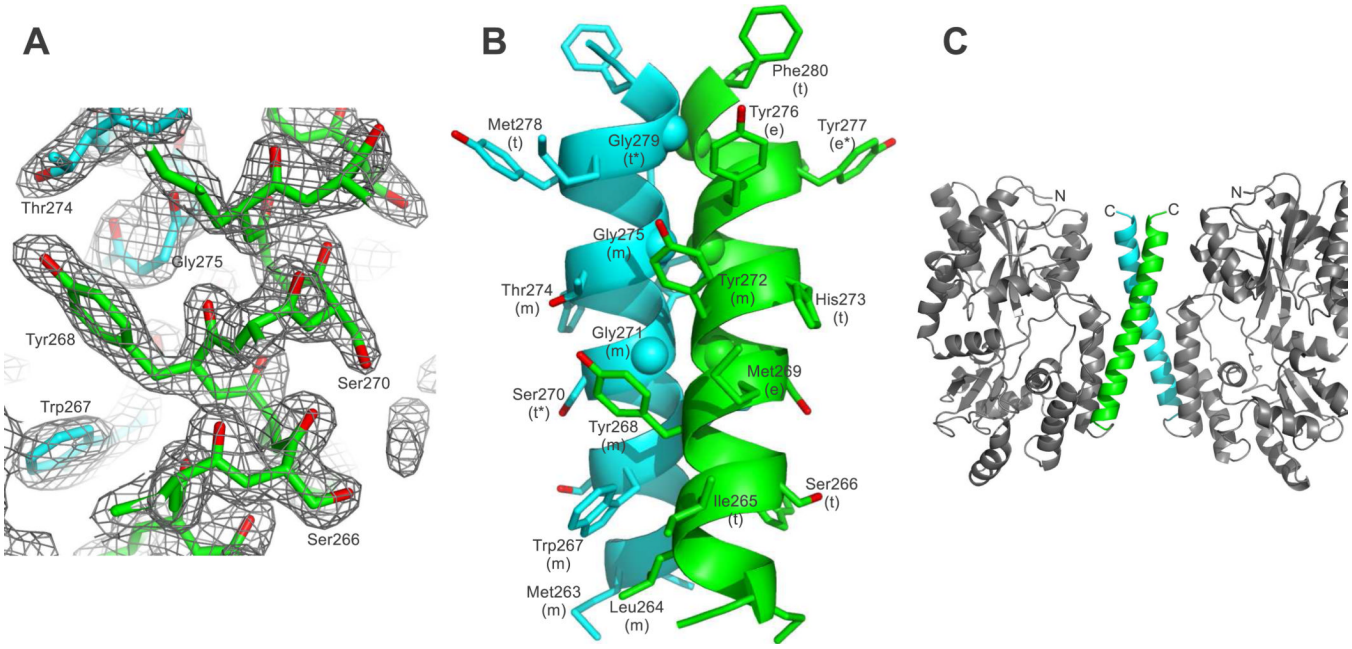

Figure 4.

Structure of the SMN YG-box (see also Fig. S2). A) Electron density of the SMN YG-box, near the center of the conserved region. A weighted, 1.9 Å resolution $2F_o$-F_c map contoured at 1.4σ is shown. B) Structure of the SMN YG-box dimer. The dimer is right-handed, with a −34° inter-helical dihedral angle. Conserved tyrosine side chains in the YxxxYxxxY motif (Y268, Y272, and Y276) pack against conserved i+3 glycine backbone atoms in the GxxxGxxxG motif (G271, G275, and G279). The α-carbon atoms of these glycine residues are drawn as a spheres. Oligomeric states of the MBP-$YG^{252–294}$ constructs with single alanine substitutions were determined by SEC-MALS analysis and are indicated in parentheses: t=primarily tetramer, m=primarily monomer, e=significant monomer and tetramer, t*=higher order oligomers in the 200–350 kDa range are also present, e*=monomer, dimer, and tetramer species present. C) Structure of the MBP-YG-box dimer. The SMN YG-box forms a continuous helix with the C-terminal helix of MBP (colored). The inter-subunit contacts in the dimer are almost exclusively between YG-box helices. There are no MBP-MBP interactions and few MBP-YG-box interactions.

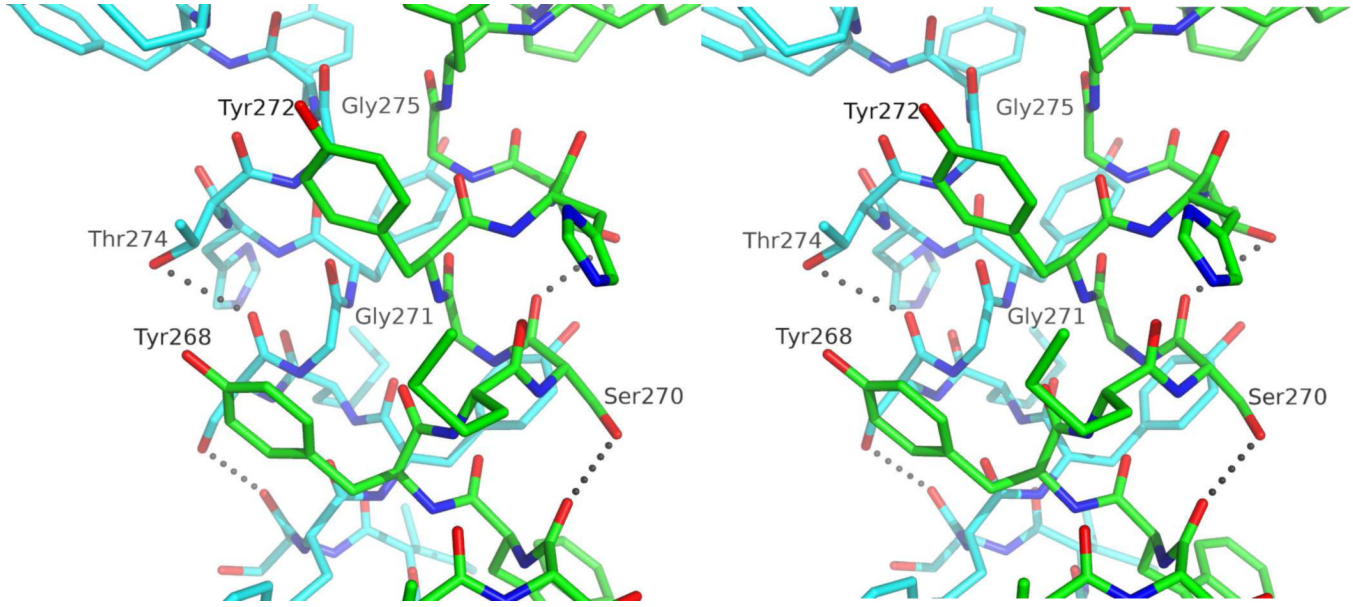

Figure 5.
Tyrosine-glycine interactions in the YG-zipper dimer. A stereo view of the dimer-dimer interface in the region including the Tyr268-Gly271 and Tyr272-Gly275 pairs is shown. Note the intimate Gly271-Gly271 and Gly275-Gly275 packing between helices.

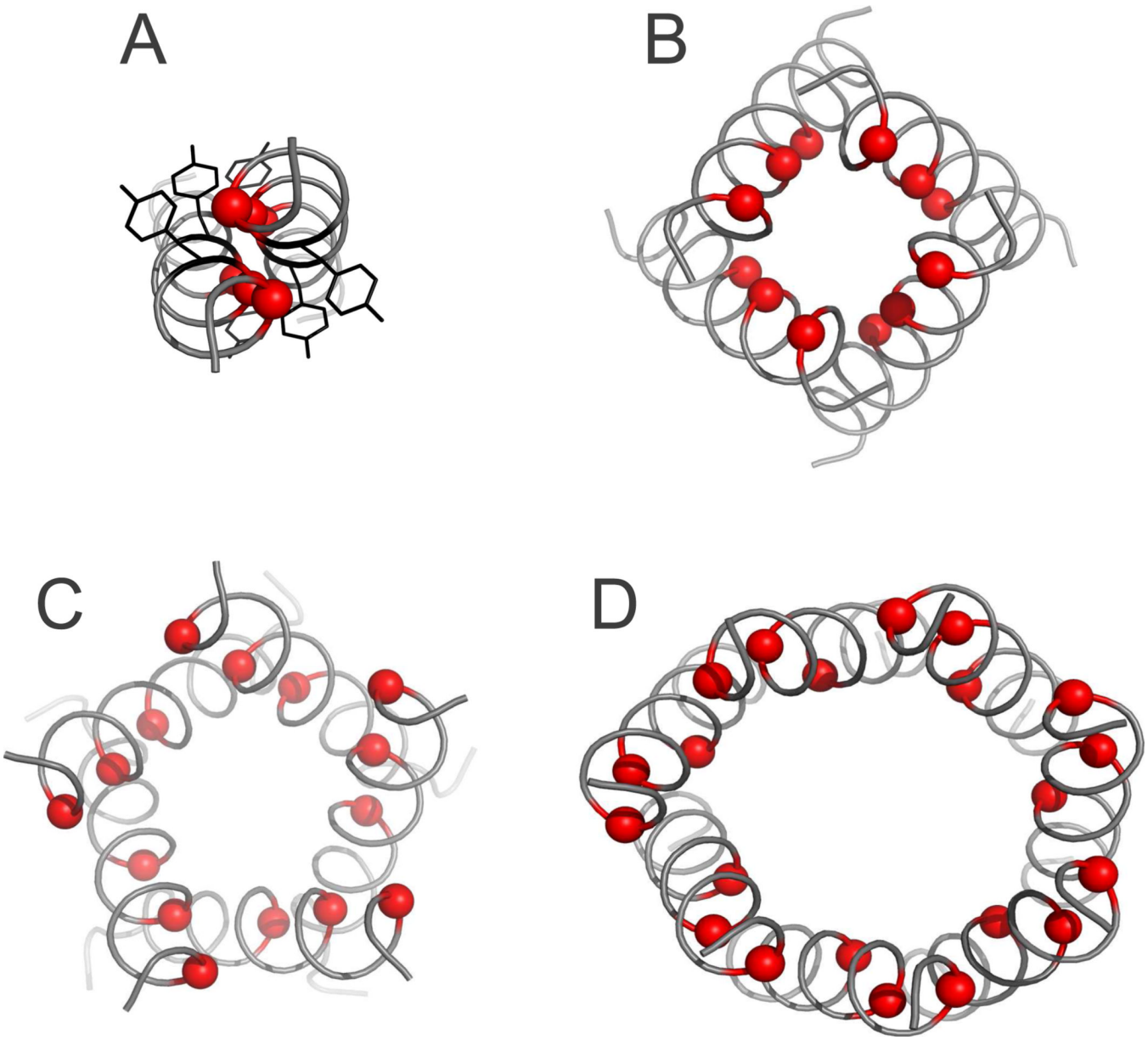

Figure 6.

Glycine zipper oligomers (see also Fig. S3). The SMN YG-box dimer is shown for comparison in (A). Examples of higher order glycine zipper oligomer structures are B) the pore-forming helices from the KcsA potassium channel tetramer (pdb 1BL8), C) mechanosensitive channel MscL pentamer (pdb 1MSL), and D) mechanosensitive channel MscS heptamer (pdb 1MXM). Conserved glycine Cα atoms are drawn as red spheres and for (A), the conserved tyrosine side chains are drawn as sticks.

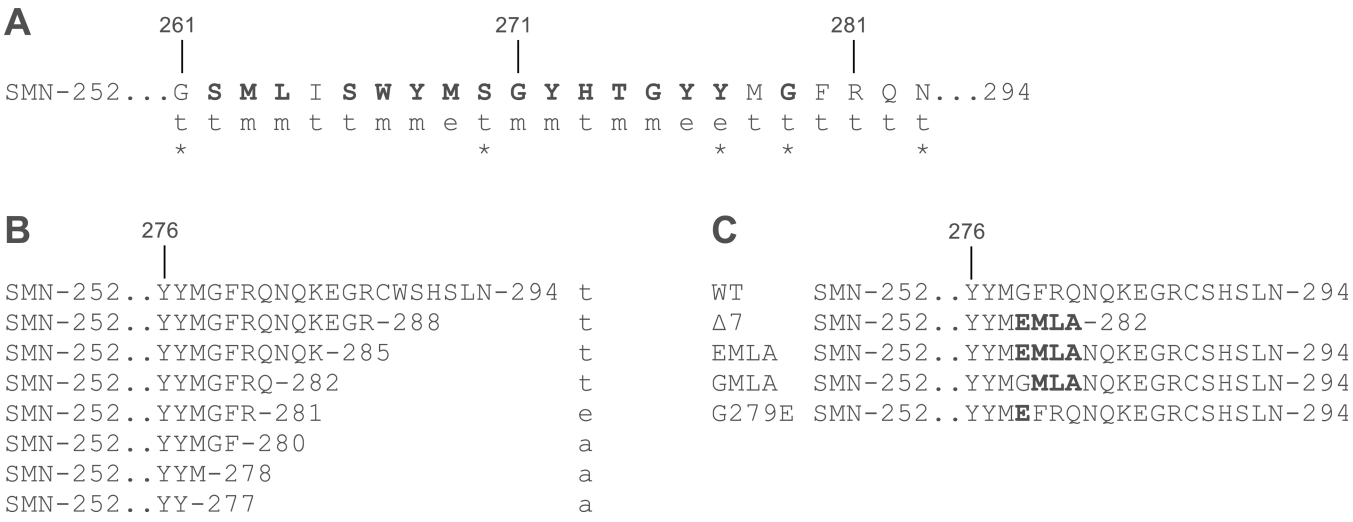

Figure 7.

Oligomerization of MBP-YG mutants. A) Effects of alanine substitution of residues from Gly261-Asn283 on oligomerization of MBP-YG$^{252–294}$. These results are also indicated in Fig. 4B. B) Oligomeric states of C-terminal truncations of MBP-YG$^{252–294}$. C) Oligomeric states of exon 7 substitutions and deletions, using the MBP-YG$^{252–294}$ construct. Oligomeric states were determined from SEC-MALS analysis of purified fusion proteins and are defined as in Fig. 4B (a=high molecular weight aggregates).

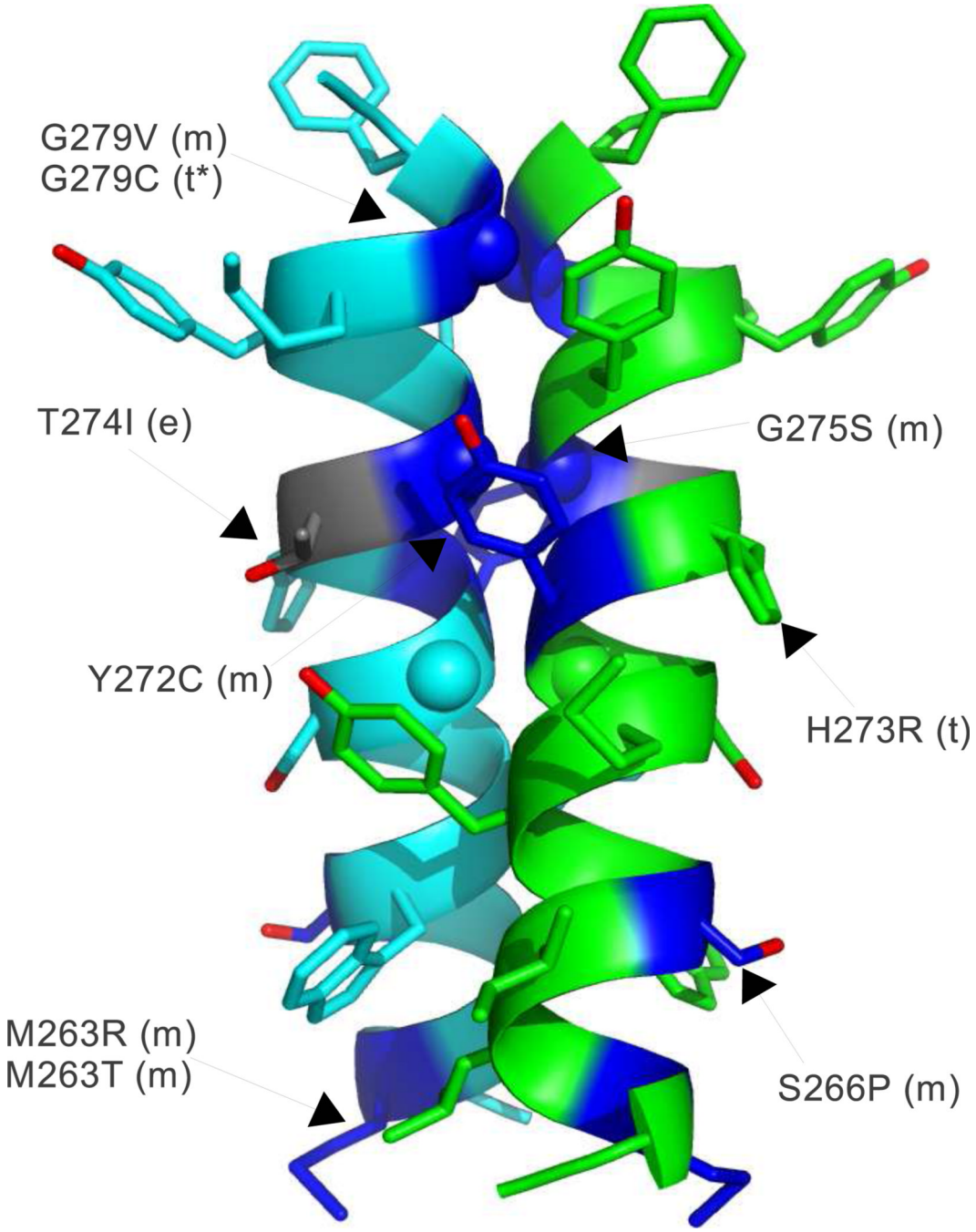

Figure 8.
Effects of SMA patient mutations on oligomerization. The results of SEC-MALS analyses of SMA patient substitutions using the MBP-YG$^{252–294}$ construct, mapped onto the YG-box dimer structure. Defective and intermediate oligomerization properties are colored blue and gray, respectively.

Table I

Crystallographic Data and Results

Diffraction Data	
Resolution (Å)	1.9
Space Group	C2
Unit Cell	a = 87.7, b = 65.9, c = 84.9 Å, β = 110.5°
Mosaicity (°)	0.6
Wavelength (Å)	0.9792
Completeness (%)	99.4 (98.8)
R_{merge}	0.126 (0.560)
Total Reflections	204,637
Unique Reflections	38,656
I/σ	19.7 (2.4)
Redundancy	5.3 (4.7)
Refinement	
R_{free}	0.247 (0.322)
R_{work}	0.203 (0.248)
Number of atoms	3,216
Protein	2,625
Solvent	591
Average B factors (Å^2)	
Protein	24.9
Solvent	27.4
Rmsd	
Bond lengths (Å)	0.02
Bond angles (°)	1.98

$R_{merge} = \Sigma |I_h - \langle I_h \rangle| / \Sigma I_h$, where $\langle I_h \rangle$ is the average intensity over symmetry equivalent measurements.

R-factor = Σ |Fobs – Fcalc| / Σ Fobs, where summation is data used in refinement.

The summation for R_{free} was calculated with 5% of the data.

Numbers in parentheses represent values in the highest-resolution shell.

Table 2

Oligomerization of MBP-YG$^{252–294}$ containing SMA patient mutations

Mutant	Type *(a)*	MBP-YG$^{252–294}$ SEC-MALS (20°C) *(b)*
WT	Normal	tetramer
L260S	II (2)	monomer
S262G	III (1)	tetramer
S262I	III (1)	tetramer+
M263T	II (1)	monomer
M263R	I (2)	monomer
S266P	II (2)	monomer
Y272C	I (2)	monomer
H273R	II (ND)	tetramer
T274I	III (1)	monomer-tetramer
G275S	III (ND)	monomer
G279C	I (ND)	tetramer+
G279V	I (ND)	monomer

(a) For each amino acid substitution, type refers to the SMA severity, which also depends on the number of copies of the *SMN2* gene present in the patient. The copy number associated the given mutation and type is given in parentheses, if known. ND=not determined. See Burghes (2009) for further details and references for the individual patient mutations.

(b) Oligomerization states were categorized based on relative peak heights from SEC separation together with measured M_W values. monomer=primarily monomer; tetramer: primarily tetramer; tetramer+: tetramer plus higher oligomerization states; monomer-tetramer: significant monomer and tetramer present. These results are mapped onto the YG box dimer structure in Fig. 8.