J Mol Model (2011) 17:2313–2323
DOI 10.1007/s00894-011-1010-y

ORIGINAL PAPER

# Nanomechanics of Ig-like domains of human contactin (BIG-2)

**Karolina Mikulska · Łukasz Pepłowski · Wiesław Nowak**

Received: 1 October 2010 / Accepted: 6 February 2011 / Published online: 29 March 2011


**Abstract** Contactins are modular extracellular cell matrix proteins that are present in the brain, and they are responsible for the proper development and functioning of neurons. They contain six immunoglobulin-like IgC2 domains and four fibronectin type III repeats. The interactions of contactin with other proteins are poorly understood. The mechanical properties of all IgC2 domains of human contactin 4 were studied using a steered molecular dynamics approach and CHARMM force field with an explicit TIP3P water environment on a 10-ns timescale. Force spectra of all domains were determined computationally and the nanomechanical unfolding process is described. The domains show different mechanical stabilities. The calculated maxima of the unfolding force are in the range of 900–1700 pN at a loading rate of 7 N/s. Our data indicate that critical regions of IgC2 domains 2 and 3, which are responsible for interactions with tyrosine phosphatases and are important in nervous system development, are affected by even weak mechanical stretching. Thus, tensions present in the cell may modulate cellular activities related to contactin function. The present data should facilitate the interpretation of atomic force microscope single-molecule spectra of numerous proteins with similar IgC2 motives.

**Keywords** Contactin · Immunoglobulin-like domain · Cell adhesion molecule · Steered molecular dynamics

**Abbreviations**

| | |
|---|---|
| AFM | Atomic force microscope |
| Big-2 | Contactin 4 |
| CNTN4 | Contactin 4 |
| FnIII | Fibronectin type III-like domain |
| IgC2 | Immunoglobulin-like C2-type domain |
| PDB | Protein data bank |
| MD | Molecular dynamics |
| SMD | Steered molecular dynamics |
| NHb | Total number of hydrogen bonds |
| PTPRG | Protein tyrosine phosphatase, receptor type, gamma |
| PTPRZ | Protein tyrosine phosphatase, receptor type, zeta |
| DSCAM | Down's syndrome cell adhesion molecule |
| APP | Amyloid precursor protein |
| CAM | Cell adhesion molecule |

## Introduction

Contactins are a subgroup of proteins belonging to the immunoglobulin superfamily. They are axonal cell adhesion molecules (CAMs) characterized by the presence of six Ig-like domains C2 type (IgC2) with disulfide bonds, four fibronectin type III-like repeats (FnIII), and a glycosylphosphatidylinositol (GPI)-anchoring domain. The contactin family includes six members: CNTN1/F3, CNTN2/TAG-1, CNTN3/BIG-1, CNTN4/BIG-2, CNTN5/NB-2 and CNTN6/NB-3 [1]. All contactins can be found in the extracellular matrix of cells. They are anchored in the cell membrane with a GPI-anchor moiety [2]. These proteins, present in various regions of the brain, affect nervous system function. Contactins have been shown to be involved in axon growth, guidance, and fasciculation [3–5], and they play a role in synaptic plasticity [6]. The first

**Electronic supplementary material** The online version of this article (doi:10.1007/s00894-011-1010-y) contains supplementary material, which is available to authorized users.

K. Mikulska · Ł. Pepłowski · W. Nowak (✉)
Theoretical Molecular Biophysics Group, Institute of Physics, N. Copernicus University,
Grudziadzka 5,
87–100 Torun, Poland
e-mail: wiesiek@fizyka.umk.pl

two contactins were identified in the late 1980s [7, 8], and they were studied intensively. The specific roles of CNTN3 and CNTN4 have not yet been clarified, although both can promote neurite outgrowth [3, 9, 10].

In this work, we study the major domains of the CNTN4 (or BIG-2) molecule, which plays an essential role in the formation, maintenance, and plasticity of neuronal networks [11]. CNTN4 was identified in 1995 by PCR cloning with degenerate primers based on homologous amino acid sequences in F3 and TAG-1 [3]. This contactin seems to be particularly important, since disrupting its gene causes developmental delay and mental retardation ("3p deletion syndrome"). 3p deletion syndrome is a rare contiguous-gene disorder involving the loss of the telomeric portion of the short arm of chromosome 3. It is characterized by developmental delay, growth retardation, dysmorphic features [12, 13] and spinocerebral ataxia [14]. Moreover, some reports suggest that CNTN4 mutations may be relevant to autism spectrum disorder pathogenesis [11, 12, 15]. Interestingly, CNTN4 is one of the axon guidance molecules that are crucial to the formation and maintenance of the functional odor map in the olfactory bulb [16]. Recently, a chick ortholog of BIG-2 was shown to be a binding partner of amyloid precursor protein (APP), which plays a central role in Alzheimer's disease [17].

Molecular signaling is related to the structure and mechanical stability of a protein. Thus, understanding the mechanical stability of modular CNTN4 is important. The nanomechanical properties of human CNTN4 were studied by means of single-molecule atomic force microscopy (AFM) [18, 19]. Two [19] or even three [18] groups of mechanical unfolding events were identified. It has been suggested that these correspond to the unfolding of individual FNIII and IgC2 domains. The maximum forces were on the order of 70 pN for FNIII and 45 pN for IgC2. It is not clear whether the wide distribution of IgC2 unfolding forces is related to distinct mechanical properties of the individual domains, or to experimental conditions, such as the random orientation of CNTN4 molecules attached to the mica surface. The present simulations provide numerical data which help to solve this issue.

The Ig-like domains are probably the most widespread domains, at least in animals. All Ig-like domains appear to be involved in binding functions. Their very similar folds are assigned into four different types: C1, C2, V, and H [20–22]. Classical Ig-like domains are composed of 7–10 β-strands distributed between two sheets with a Greek key β-barrel topology. The general shape of Ig-like domains is well conserved, but they can differ significantly in their size, due to the high variability of the loops [23].

Despite its importance, only a few MD simulations of unfolding solvated Ig-like domains have been reported in the last few years. Perhaps the first studies were devoted to the famous giant muscle protein titin, which is rich in these types of domains [24–27]. However, the mechanical behavior of the Ig-like domains has never been studied for any contactin until now. The work presented here is the first systematic mechanical unfolding analysis of the Ig-like domains present in CNTN4. Our data show that the nanomechanical properties of six IgC2 domains are not identical. Internal domains 2 and 3, which participate in interactions with enzymes regulating cell activity, exhibit the highest mechanical stability. A better understanding of IgC2 domains may aid in the development of protein-based nanomaterials with desirable mechanical properties. It is also important when assessing the role of mechanical cell deformation in the proper development of neurites.

## Materials and methods

The 3D structure of human CNTN4 is not yet known. Only four immunoglobulin-like domains of mouse had been registered in the PDB by December 2009. In order to investigate the mechanical properties of CNTN4 during stretching, we constructed six models of each immunoglobulin-like domain of human contactin 4 (accession number: Q8IWV2, isoform 1). At present, three isoforms of CNTN4 are described in the RefSeq database at NCBI. In our work we used only the longest variant of this molecule, because only this isoform is highly expressed in the human brain, particularly in cerebellum, thalamus, amygdala, and cerebral cortex [5, 28]. It is 1026 amino acids long (113.454 kDa). An intermediate-sized isoform does not appear to be expressed in the central nervous system [5, 28, 29]. A third isoform—a truncated variant lacking all six Ig-like domains and two of the four FnIII domains—has been shown to have low levels of expression in the brain [12, 28].

Models were made using a protein modeling server—the SWISS-MODEL Repository [30–33]—and Accelrys software (Accelrys Inc. 2008). Homology models were built using cell adhesion molecules rich in Ig domains from chicken and *Drosophila melanogaster*; see Table 1 and Fig. 1. Models were based on immunoglobulin superfamily (IgSF) adhesion molecules: contactins (CNTN2) and the Down's syndrome cell adhesion molecule (DSCAM), which are defined by the presence of domains in the extracellular region that have sequence similarities to the variable or constant domains of Ig molecules. All contactins show 40–60% identity with each other at the amino acid sequence level [5].

All models were solvated using a 0.7 nm layer of the TIP3P water model [34] in each dimension. A cutoff of 12 Å for nonbonded interactions was applied. Langevin dynamics and a Langevin piston algorithm were used to maintain the temperature at 300 K and a pressure of 1 atm. For each model, we created two variants where the parameters for electrostatic forces were changed. Initial structures are

**Table 1** Structures used to build homology models of the IgC2 domains of CNTN4

| Amino acid numbers | PDB code | Organism | Protein |
|---|---|---|---|
| 24–403 | 1CS6 A | Chicken | CNTN2, Ig domains |
| 404–591 | 3DMK B | *Drosophila melanogaster* | DSCAM, Ig domains |

presented in Fig. 2, and their folds, which are useful for further analysis, are shown in Fig. 3. We have performed three full electrostatics and three multiple time-step steered molecular dynamics (SMD) simulations for each of the IgC2 domains with disulfide bonds. Every domain has a single disulfide bond between two cysteines located in neighboring β-sheets. When the multiple time-step method was employed, time steps of 1 fs for bonded, 2 fs for short-range nonbonded, and 4 fs for long-range electrostatic forces were used.

All simulations were performed using software package NAMD 2.7b2 Infiniband [35] with the all-atom CHARMM27 force field [36]. The Visual Molecular Dynamics (VMD) software package [37] (version 1.8.7) and home-made scripts were used to prepare input files and to analyze output trajectories. We performed 0.2 ns of water equilibration, 10,000 steps of minimization, 0.35 ns of heating from 0 K up to 300 K, and 0.15 ns of equilibration of the whole system before each main SMD simulation. A constant-velocity SMD method was used to stretch all domains along their long axes. The force vector connected the Cα atoms of the N- and C-terminal residues. This was calculated after water equilibration. In this technique, a virtual harmonic force is applied to one end (the C-terminus) of the protein, which is simultaneously fixed at the other end (the N-terminus). Structures were stretched at a constant speed of 0.025 Å/ps with a spring constant of 4 kcal $mol^{-1}$ $Å^{-2}$ (278 pN $Å^{-1}$). We ran thirty-six SMD simulations of six IgC2 domains (total duration 360 ns). The systems include about 9400 to 10,900 atoms containing ions and water. A concentration of 150 mM NaCl was used for each system. This is similar to the salt concentration in the physiological environment and comparable to those observed in the extracellular matrix (100 mM).

## Results and discussion

### SMD unfolding of Ig domains

From the Cα-overlapped IgC2 domains (see Fig. 4), we can see that the pulling directions of the domains were nearly parallel. The largest angle between these vectors does not exceed 20°.

The intrinsic elasticity of protein domains acts against the pulling force. The forces that arise during the mechanical unfolding of isolated IgC2 domains of CNTN4 are presented in Fig. 5. The first 5 ns of SMD are shown for each domain, and the results for all six simulations are collected in the same figure. The time axis is used to represent extension, since the extension of a domain is proportional to time in each case studied here. Our data indicate that unfolding processes are multistep and not identical for each IgC2 domain. We observe one maximum of the force at 0.5 ns (1 ns for $IgC2_2$), and several distinct, lower maxima in the region of 1–3 ns (see Fig. 5). The individual force spectra for the same domain are rather similar. The time at which the second (or further) maxima occurs is more variable. For the IgC2 domains 1, 3 or 5, that maximum can appear at different times (extensions), with a difference of up to 1.5 ns in their appearance times. The $IgC2_3$ and $IgC2_5$ domains show qualitatively distinct unfolding characteristics in the regions of 1.5–2.5 ns and 2–3 ns, respectively. In these regions, the spread of maximum forces for these two domains is particularly large. This indicates possible domain-dependent unfolding mechanisms.

The first, main maximum observed for each stretching curve does not exceed 1750 pN ($IgC2_2$). The lowest one is around 950 pN (for $IgC2_6$). All of these values are collected in Table 2. The results of our atomic force microscope (AFM) experiments, published in [18], show multiple maxima attributed to the unfolding of six IgC2 domains and four FNIII domains. The measured maximum forces for CNTN4 unfolding are much lower, in the range of 40–70 pN. This discrepancy between the SMD and AFM results was expected, since the stretching process in the simulations was substantially faster ($10^4$–$10^6$ times) than the experimental timescale. It is known that the higher loading rates of AFM (or SMD) experiments result in higher barriers [38]. In the experiment, the details of the domain unfolding mechanism are obscured by a high level of noise.

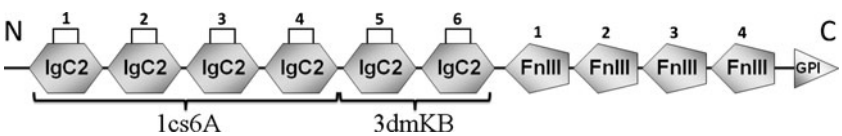


**Fig. 1** Multidomain structure of CNTN4, with the PDB codes of the 3D structures used to build the particular Ig domain

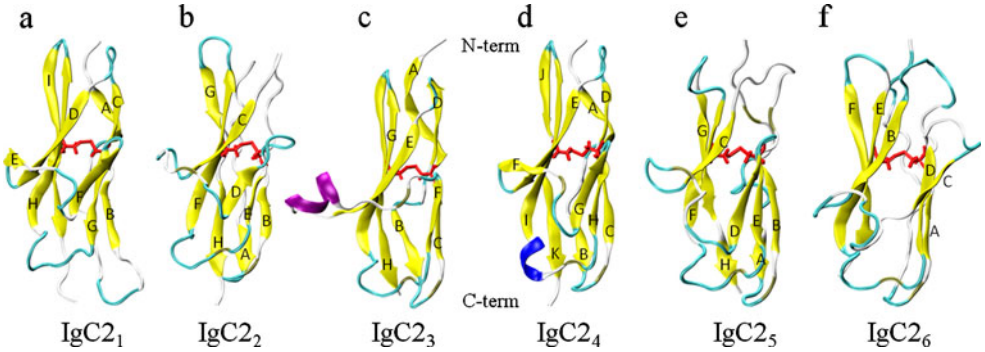


**Fig. 2** A comparison of the folds of six modeled IgC2 domains of CNTN4. Disulfide bonds are shown in *red*. The figure was prepared using VMD [37]

Since we had all-atom SMD data, we were able to investigate the unfolding process of each individual domain in full detail. In order to keep this paper concise, the results of analyzing only three of the six domains (2, 3, 5) are presented here. We selected these domains as it was recently shown that domains 2 and 3 interact directly with important signal enzymes: protein tyrosine phosphatases [39]. Also, since the 3D fold of the domain IgC2$_5$ was constructed based on the *Drosophila* template (and not on the chicken template, as for domains 1–4), its unfolding scenario is presented here in greater detail.

## The impact of the mechanical unfolding of the IgC2$_2$ domain on the tyrosine phosphatase–contactin interaction site

It was very recently discovered that the contactins CNTN1 and CNTN4 interact with the tyrosine phosphatases PTPRZ

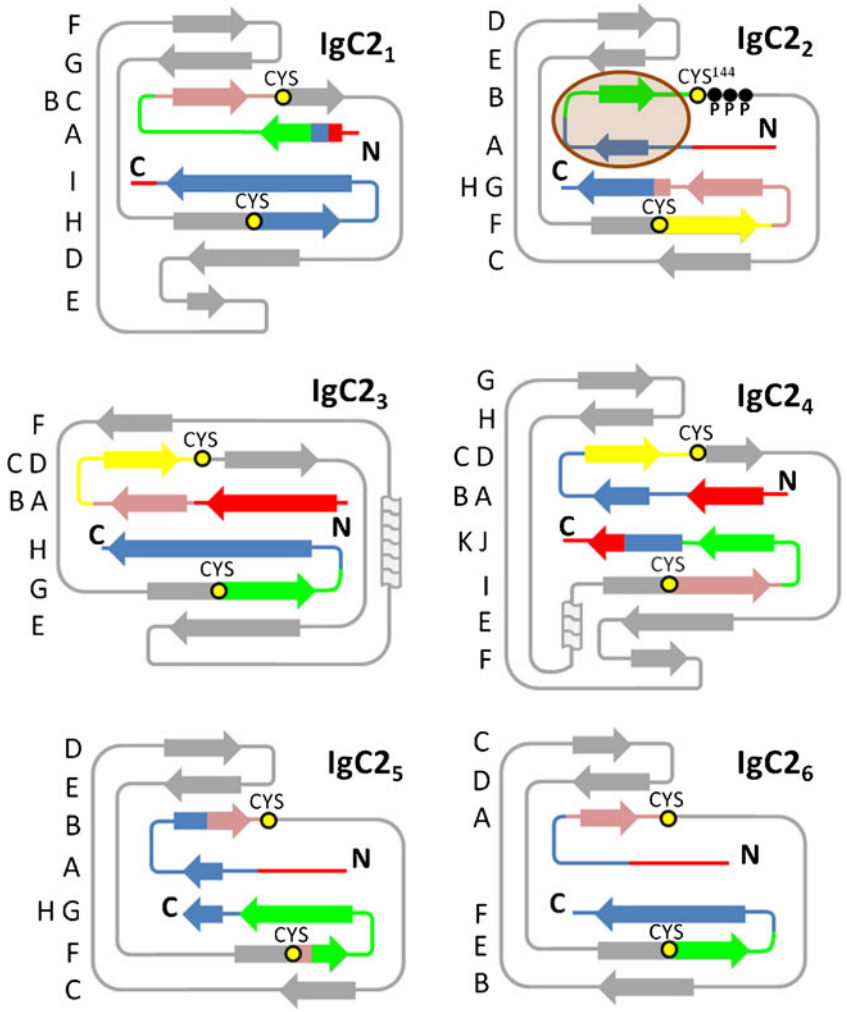


**Fig. 3** Schematic folds for IgC2 domains of CNTN4. *Color coding* corresponds to data in Figs. 6, 7, 8, and represents consecutive events during SMD unfolding. A *dark circle* in the IgC2$_2$ scheme represents an interaction region with phosphatase. *P* proline residue

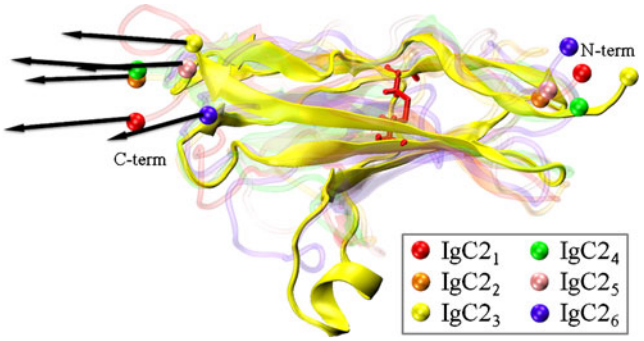


**Fig. 4** Pulling directions in each IgC2 domain of CNTN4

and PTPRG [39]. This interaction is crucial to the outgrowth of neurites, so it plays a role in nervous system development. The X-ray structure of the complex of four Ig repeats of murine CNTN4 with the PTPRG carbonic anhydrase domain has been determined [39]. Crystallographic data show that the receptor protein tyrosine phosphatase interacts with the IgC2$_2$ and IgC2$_3$ domains. The harpin loop of PTPRG protrudes mainly into the 129–142 region of the IgC2$_2$ domain. Since the sequences and 3D structures of human and murine CNTN4 are quite similar, we were able to check whether cellular mechanical strain affects interactions of CNTN4 with the receptor of phosphatase.

The scenario for the mechanical unfolding of IgC2$_2$ is presented in Fig. 6. The unfolding is clearly a stepwise process; distinct phases are represented by different colors in Fig. 6.

The initial stretching leads to a very high maximum force of 1700 pN (Table 2). Careful analysis of the origin of this force shows that it is generated mainly by A–H and H–F β-strand interactions. Initially up to eight hydrogen bonds were observed in this region, but after passing the first maximum at 1 ns (Fig. 5), only 3 or 2 H-bonds remained. We noted that the Arg129 side chain is oriented towards the interior of the IgC2$_2$ domain. This is an unusual orientation, since all the other arginines have side chains that are oriented towards the solvent, as expected. The Arg129 forms two hydrogen bonds with carbonyl oxygens from the backbone (Cys144 and Leu145). The scission of these bonds correlates perfectly well with the first maximum at 1 ns (Fig. 5). It is interesting to note that this particular Arg129 is present in mouse CNTN4 as well [39]. The Arg129 is located in the specific fragment of the IgC2$_2$ domain that is responsible for the interactions of contactin with PTPRG. It is also buried inside the murine IgC2$_2$ domain, and undergoes electrostatic interactions with the carbonyl oxygens of Pro146, Pro147 and Pro148 (see Fig. 3). Such a sequence of three proline residues is rather uncommon in protein architecture. It may be that this triad has a structural role in maintaining PTPRG–CNTN4 contacts, since this motif is conserved in IgC2$_2$ domains in humans and mice but is completely absent from the other IgC domains, which do not interact with receptor phosphatases.

The second unfolding force maximum of 1250 pN for the IgC2$_2$ domain is observed at 1.7 ns. It corresponds to

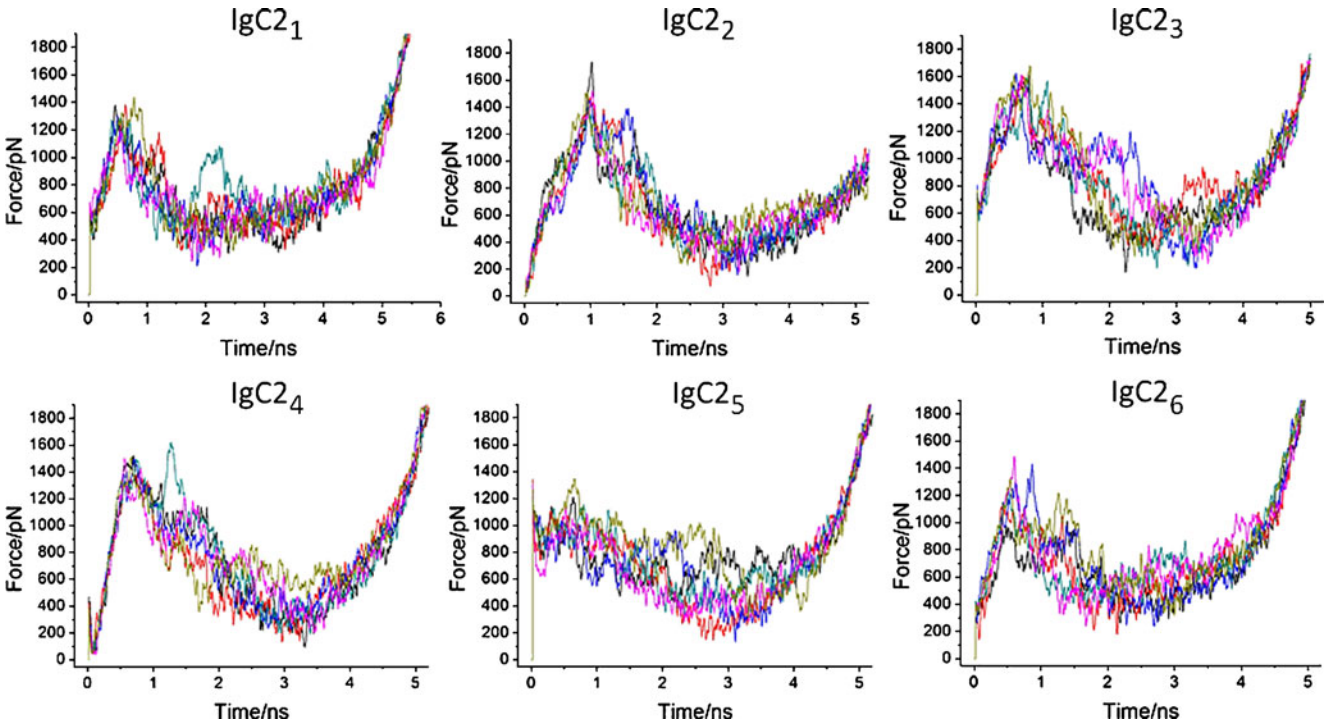


**Fig. 5** Force vs. time plots from SMD simulations for each IgC2 domain of CNTN4

**Table 2** The results for $F_{max}$ in different SMD simulations of the IgC2 domains. All values are in pN. The last row shows average values of $F_{max}$ and the variances of these averages, $\sigma F_{max}$

| $IgC2_1$ | | $IgC2_2$ | | $IgC2_3$ | | $IgC2_4$ | | $IgC2_5$ | | $IgC2_6$ | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $F_{max}$ | $\sigma F_{max}$ | $F_{max}$ | $\sigma F_{max}$ | $F_{max}$ | $\sigma F_{max}$ | $F_{max}$ | $\sigma F_{max}$ | $F_{max}$ | $\sigma F_{max}$ | $F_{max}$ | $\sigma F_{max}$ |
| 1381 | | 1737 | | 1632 | | 1524 | | 1309 | | 963 | |
| 1376 | | 1476 | | 1597 | | 1382 | | 1343 | | 1147 | |
| 1301 | | 1438 | | 1625 | | 1500 | | 1279 | | 1297 | |
| 1279 | | 1423 | | 1597 | | 1495 | | 1299 | | 1235 | |
| 1177 | | 1520 | | 1617 | | 1502 | | 1309 | | 1487 | |
| 1438 | | 1511 | | 1680 | | 1510 | | 1267 | | 1325 | |
| 1325 | 93 | 1518 | 114 | 1625 | 31 | 1486 | 52 | 1301 | 27 | 1242 | 177 |

the resistance of strand B, which interacts with the $IgC2_2$ core through β-strand E. Only a few H-bonds are broken here, such as two backbone Met140–Ile181 interactions localized mainly close to the fixed N-terminus of this module (Fig. 6, green). In phases (4) and (5) of Fig. 6, several lower maxima are observed, and these events are related to the gradual destruction of β-strand interactions. The last phase of SMD simulation (3.5–7 ns, Fig. 6) corresponds to quasi-harmonic stretching of the domain backbone. The domain is not fully extended, since the disulfide bridge protects the internal part of the protein from unfolding, and a residual secondary β-structure is still observed.

## Mechanical unfolding of the $IgC2_3$ domain

The third IgC2 domain of CNTN4 also participates in interactions with protein tyrosine phospatase PTPRG [39]. In contrast to $IgC2_2$, it has a substantial number of charged amino acids (2D, 9E, 7R, 7K). Therefore, at least six unfolding events related to the breaking of salt bridges are seen in Fig. 7.

In region (1) (red in Fig. 7), after the salt bridge Glu222 (strand A)–Arg301(H-G loop region) is broken, the sequential breaking of three hydrogen bonds is seen. In this way, a complex of B-H β-strands is broken. At the same time, another salt bridge between Arg301 and Glu224 (strand A) gets stretched. It is released just before the main force maximum at 0.4 ns is reached (Fig. 7). The sudden drop in the force is related to the synchronous breaking of at least 4–5 H-bonds between the H-G and A-D strands. In region (2), shown in blue in Fig. 7, further unfolding of the G-E region and the H-G interface is observed. Regions (3) and (4) correspond to much easier stretching of G and B stands. At 1.4 ns, the breakage of the salt bridge Glu294(G)–Arg307(H) contributes to a sudden elongation of this domain. From the fourth nanosecond on (force of 700 pN), only the stretching of all of the unfolded backbone regions of the $IgC2_3$ domain is observed. An important observation is that the critical region Gly221–Glu228, which is known to be responsible

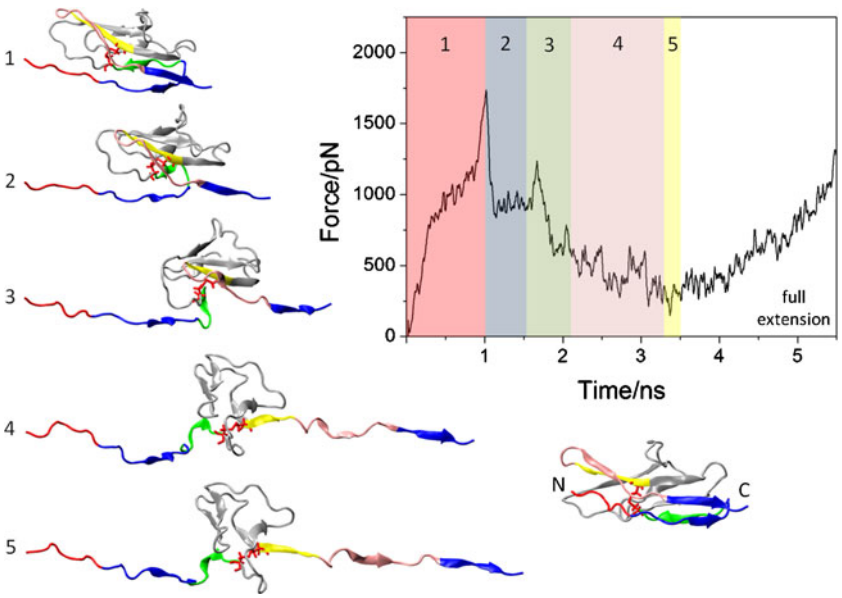


**Fig. 6** Sequence of events during the mechanical unfolding of the $IgC2_2$ domain of CNTN4. The *colors* in the plot are mapped to the ribbon model of the protein in order to indicate which regions unfold during each phase of stretching

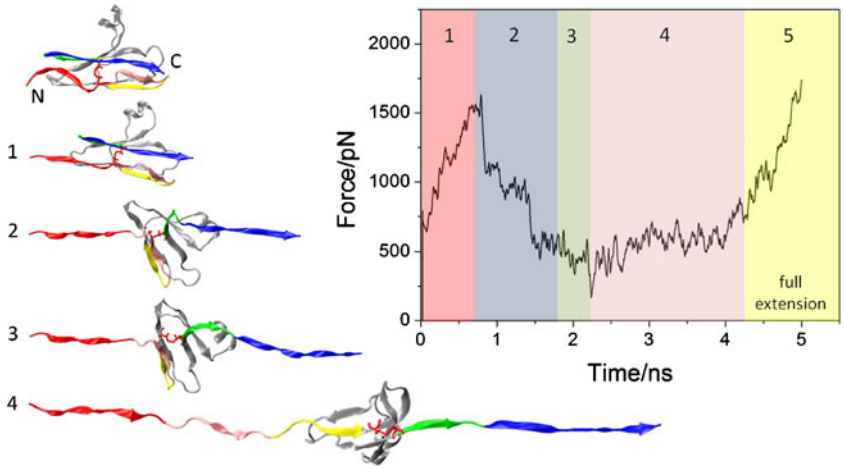


**Fig. 7** The sequence of events in the mechanical unfolding of the $IgC2_3$ domain

for interactions with PTPRG [39], is affected by the external force at a very early stage. Even relatively small stretching forces at 0.3 ns (600 pN in our SMD) modify the structure of this cavity. Thus, it is possible to speculate that mechanical strain in the ECM regions of neurite cells may modulate the interactions of phosphatase with the CNTN4. Our data indicate, for the first time, that this particular CNTN4–PTPRG contact is sensitive to even quite weak stretching, and it is not protected by the architecture of the IgC2 module 3. This "mechanical sensitivity" of CNTN4 is probably not accidental; it likely plays some physiological role in signal transduction.

## Mechanism of the unfolding of the $IgC2_5$ domain

The fifth IgC2 domain has (just like $IgC2_3$) a substantial number of charged residues (K: 10, R: 5, E: 5, D: 4), but their stabilizing effect is much weaker. In the force spectrum (Fig. 8a), there is no dominant initial maximum since the β-strand at the N-terminus is absent in this domain. In the first phase of stretching, the tension increases in the region of the positively charged residues Arg410–Lys414–Arg415, leading to the formation of a short loop. Reorganization of flanking hydrogen bonds is observed in the regions of the N- and C-termini. The residues Arg415 and Glu428 form a new salt bridge (0.5–1.2 ns, Fig. 8a). In region (2), strong interactions between Glu428, Thr417 and Lys430 contribute to large force values. At 1.55 ns, five hydrogen bonds between G and F strands start to break. This process ends at 2.37 ns, and then gradual stretching of the whole protein backbone is observed (phase 4, Fig. 8a).

Interestingly, the $IgC2_5$ domain shows an alternative unfolding path, with a lower mechanical stability than the other CNTN4 domains. The maximum force in this path is 1250 pN, and the force spectrum does not exhibit a particularly strong maximum. The unfolding starts from the N-terminus (fixed atom) region, and three relaxation processes are observed in the first phase (see Fig. 8b). These are all related to the scission of hydrogen bonds between A-G strands and major rearrangements of the backbone in the region 410–417 of the N-terminus. Three positively charged amino acids are packaged in this short region (Arg410, Lys414, Arg415) in a pseudo-helical structure. Phases (2) and (3) are very short; in a period shorter than 1 ns (during 0.7–1.2 ns), the C-terminal fragment of the domain $IgC2_5$ is stretched, and the virtual force acts directly on the S–S bridge. In contrast to that, the stretching of the N-terminal part (phases 4, 5 and 6) is a much longer process and takes 4 ns. The multiple maxima in this plot correspond to the unfolding of the N-terminal β-strands B, H, G, and F.

We observe that surface salt bridges contribute to the stability of the $IgC2_5$ domain. Thus, the mechanical strength of this part of CNTN4 may be modulated by the local pH. In principle, the strength of the salt bridge (3–4 kcal $mol^{-1}$) is similar to or smaller than that of the H-bond (3–10 kcal $mol^{-1}$), and such events should be distinguishable in force spectra. However, the level of noise in our trajectories (fluctuations in the protein structure at 300 K) is too high to allow us to separate the features in the force spectrum that could be related to H-bond breaking from those due to the breakage of an electrostatic interaction between salt bridge partners.

## A comparison with experimental data

In our previous paper, the mechanical unfolding of single human CNTN4 molecules was studied using the AFM technique [18]. Individual maxima observed in AFM spectra are associated with the unfolding events of individual IgC2 or FNIII domains. The experimental length of IgC2 domains (shorter modules), 19.4±1.6 nm, correlates well with the average extension of 16 nm calculated here with the SMD method. A comparison of the scaled $IgC2_2$ unfolding curve with the experimental one is shown in Fig. 9. The shapes of both spectra are similar, but the

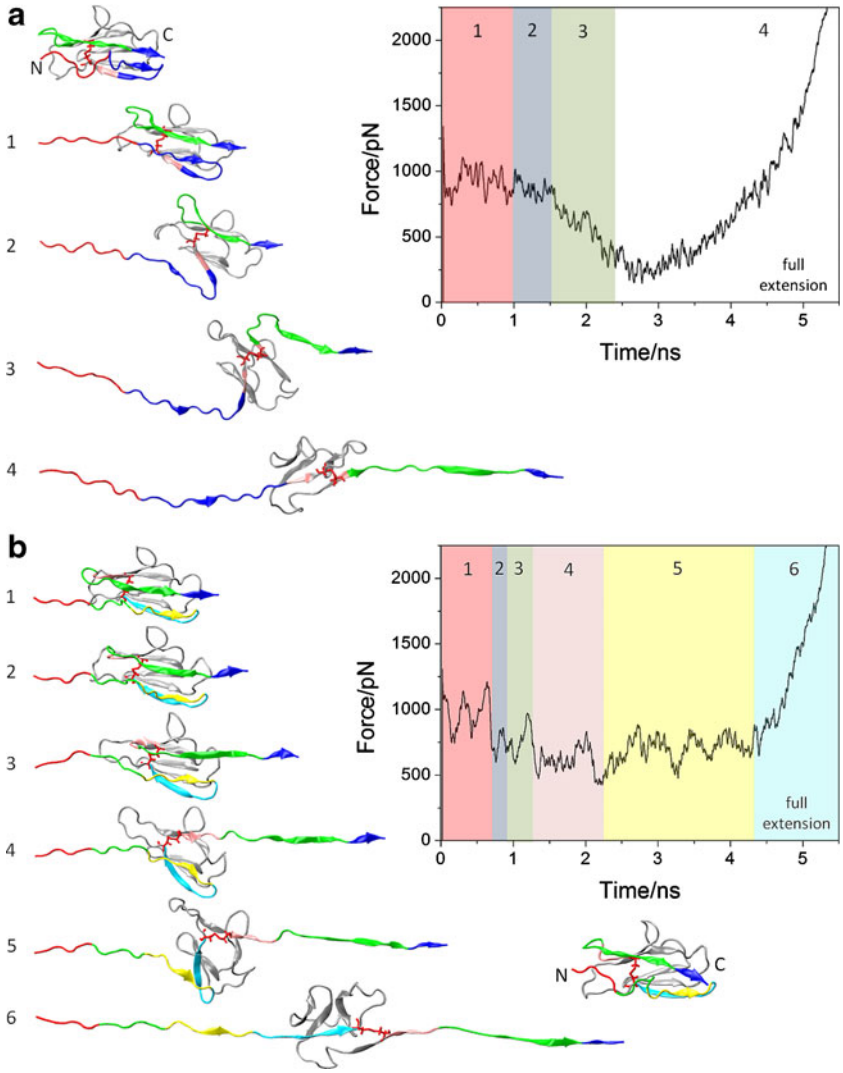


**Fig. 8a–b** The sequence of events during the mechanical unfolding of the $IgC2_5$ domain. **a** Example of a dominant path, **b** example of an alternative unfolding path

SMD force maxima are a factor of 15 higher than the experimental ones. Events corresponding to individual β-strand detachments are seen in both spectra, but the small maxima observed in the experimental curve are a similar height to the noise level that is characteristic of the cantilevers used in the experiment, and so are difficult to interpret. Therefore, the SMD data provide us with a more detailed picture of mechanical protein unfolding.

As already mentioned, a quantitative agreement between SMD simulations and AFM single-molecule spectroscopy is difficult to achieve. In addition to the approximate nature of the classical MD CHARMM forcefield, optimized for equilibrium geometries of amino acids, the high loading rate of SMD simulations is the main factor responsible for this overestimation of forces. The reason for this discrepancy has been discussed in numerous papers [40, 41]. It is worth noting that varying the loading rate may even affect a number of intermediates found during the mechanical decomposition of protein modules [41].

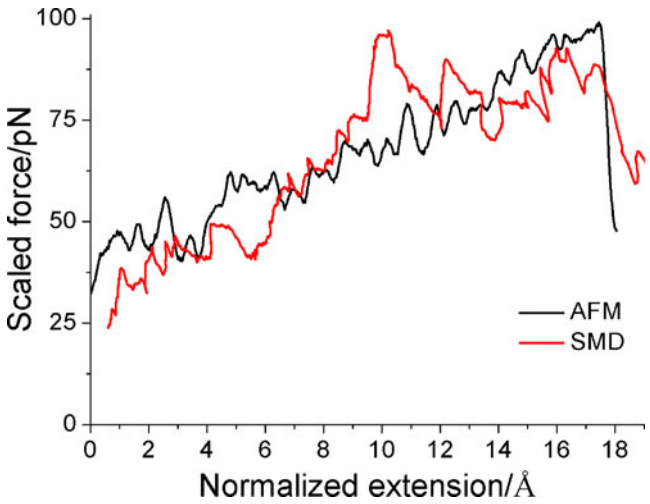


**Fig. 9** A comparson of the SMD force curve for $IgC2_2$ domain unfolding with an experimental AFM curve [18]. The SMD forces are scaled down by a factor of 15, and the stretching length has been normalized (by a factor of 2.3)

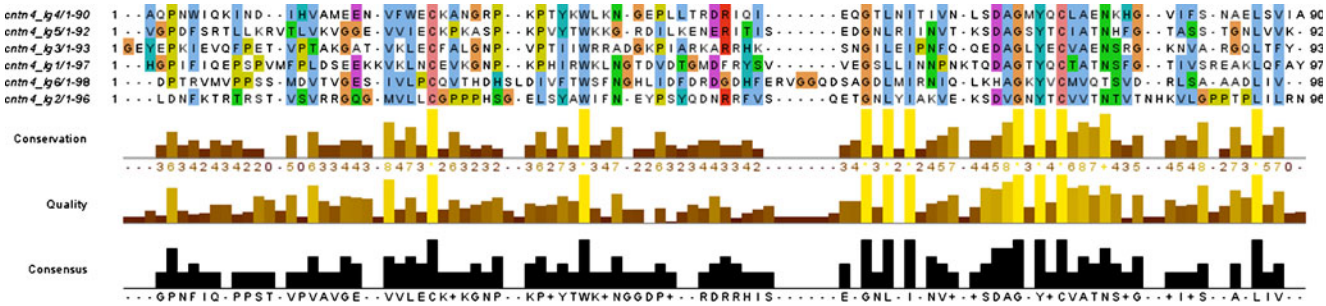


**Fig. 10** Multiple alignment of IgC2 domains of CNTN4, prepared using JALVIEW [42]

## Conserved residues and the mechanical stability of IgC2 domains

Despite the very similar 3D structures of all of the IgC2 domains studied here (see Fig. 2), the pairwise sequence similarity is not that high (27–43%). A comparison of sequences (Fig. 10) indicates that around 15 residues out of about 100 are well conserved in IgC2 domains of CNTN4. The natural question is then: what is their functional role? Two conserved cysteines are necessary to maintain S–S bridges. The locations of the other residues are shown in Fig 11. It seems that they are located in the same spatial region of the module. These residues flank the well-conserved Tyr residue. It is tempting to hypothesize that this Tyr may play a catalytic role. Despite our search of the biochemical literature, we could not find any reports on the catalytic activity of IgC2 domains.

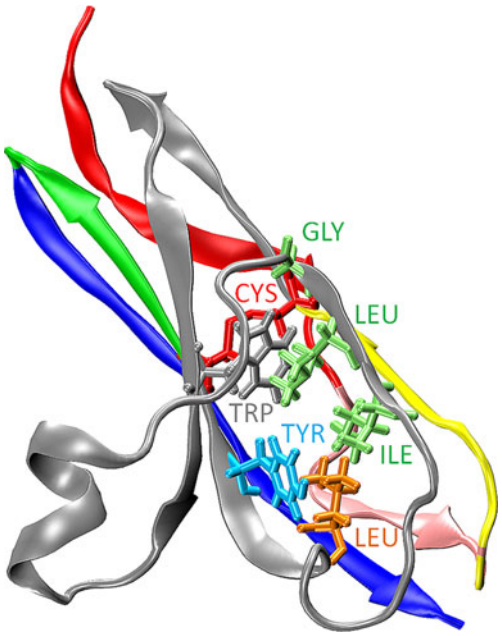


**Fig. 11** Characteristic, well-conserved amino acids present in each IgC2 domain of CNTN4. The $IgC2_3$ domain is shown

Checking the localization of the residues conserved in all six Ig domains, we have noticed that seven residues are fully conserved in the internal region located between the Cys–Cys bridges, while only two are outside this zone. The length of the internal region varies between 48 and 58 aminoacids, and is comparable to the sum of the amino acids present in both N- and C-end fragments of Ig (43–48). This fact may suggest that the S–S bridge does not allow for mechanical disturbances to the "internal" region, and that this zone has some physiological functions that are yet to be discovered. In our stretching experiments, this "internal" core was not studied. As we can see from Fig. 2, the fold in this region is pretty much different for each IgC2 domain, but this variance in 3D structure could be an artifact related to the homology modeling applied here.

## Conclusions

The mechanical unfolding paths of the six IgC2 domains of human CNTN4 studied here are generally similar within each domain landscape. We see no clear intermediates during these SMD simulations on a 10-ns timescale. The common mechanism is as follows. In the first phase, the N-terminal part detaches from the protein globule, and the C-terminal fragment is released in the second phase. These events are related to the maximum forces observed during unfolding. The differences in the force spectra of individual domains can be attributed to the distinct topologies of the IgC2 folds rather than to different amino acid compositions. However, the most stable domain, 3 (with an average $F_{max}$ of 1625 pN), is exceptionally rich in salt bridges (it has 11), which contribute to its mechanical resistance. Domain 2 is also quite stable, with an average maximum force of 1588 pN. The other "flanking domains" are more prone to stretching ($F_{max} \approx 1242–1325$ pN). We hypothesize that the particular stability of domains 2 and 3 is related to their physiological function. These modules directly interact with the extracellular parts of tyrosine phosphatases—important enzymes that regulate neuron development [39]. Even a

small mechanical deformation of CNTN4 would modulate phosphatase activity. This link with phosphorylation processes may explain why deletions in the CNTN4 gene at the loci 3p25–p26 lead to autism spectrum disorder [12].

Among all of the IgC2 domains studied, only seven residues (except for the Cys–Cys bridge) are fully conserved. These are mainly located close to the conserved Tyr residue. Possible catalytic activity of CNTN4 related to this region may be expected.

Hopefully, our study provides unique reference data on the mechanical stabilities of ubiquitous Ig-type domains in medically important proteins. Despite a numerical discrepancy between SMD and AFM force spectra, the qualitative features of IgC2 atomistic unfolding will aid in the interpretation of AFM single-molecule experiments. Further studies of the nanomechanics of FNIII repeats are currently in progress in our lab.

**Acknowledgments** This research was supported by Polish Ministry of Education and Science (grant nos. N202 262038 and N519 578138). The authors whish to thank the Computational Center TASK in Gdansk, where the calculations were performed. The authors would also like to offer their thanks for the nationwide license for Accelrys software.



## References

1. Shimoda Y, Watanabe K (2009) Contactins: emerging key roles in the development and function of the nervous system. Cell Adhes Migrat 3:64–70
2. Berglund E, Murai K et al (1999) Ataxia and abnormal cerebellar microorganization in mice with ablated contactin gene expression. Neuron 24:739–750
3. Yoshihara Y, Kawasaki M, Tamada A, Nagata S, Kagamiyama H, Mori K (1995) Overlapping and differential expression of BIG-2, BIG-1, TAG-1, and F3: four members of an axon-associated cell adhesion molecule subgroup of the immunoglobulin superfamily. J Neurobiol 28:51–69
4. Ogawa J, Kaneko H, Masuda T, Nagata S, Hosoya H, Watanabe K (1996) Novel neural adhesion molecules in the contactin/F3 subgroup of the immunoglobulin superfamily: isolation and characterization of cDNAs from rat brain. Neurosci Lett 218:173–176
5. Kamei Y, Takeda Y, Teramoto K, Tsutsumi O, Taketani Y, Watanabe K (2000) Human NB-2 of the contactin subgroup molecules: chromosomal localization of the gene (CNTN5) and distinct expression pattern from other subgroup members. Genomics 69:113–119
6. Murai K, Misner D, Ranscht B (2002) Contactin supports synaptic plasticity associated with hippocampal long-term depression but not potentiation. Curr Biol 12:181–190
7. Ranscht B (1988) Sequence of contactin, a 130-kD glycoprotein concentrated in areas of interneuronal contact, defines a new member of the immunoglobulin supergene family in the nervous system. J Cell Biol 107:1561–1573
8. Furley A, Morton S, Manalo D, Karagogeos D, Dodd J, Jessell T (1990) The axonal glycoprotein TAG-1 is an immunoglobulin superfamily member with neurite outgrowth-promoting activity. Cell 61:157–170
9. Connelly M, Grady R, Mushinski J, Marcu K (1994) PANG, a gene encoding a neuronal glycoprotein, is ectopically activated by intracisternal A-type particle long terminal repeats in murine plasmacytomas. Proc Natl Acad Sci USA 91:1337–1341
10. Yoshihara Y, Kawasaki M, Tani A, Tamada A, Nagata S, Kagamiyama H, Mori K (1994) BIG-1: a new TAG-1/F3-related member of the immunoglobulin superfamily with neurite outgrowth-promoting activity. Neuron 13:415–426
11. Roohi J, Montagna C et al (2009) Disruption of contactin 4 in three subjects with autism spectrum disorder. Br Med J 46:176–182
12. Fernandez T, Morgan T, Davis N, Klin A, Morris A, Farhi A, Lifton R (2004) Disruption of contactin 4 (CNTN4) results in developmental delay and other features of 3p deletion syndrome. Am J Hum Genet 74:1286–1293
13. Pohjola P, de Leeuw N, Penttinen M, Kääriäinen H (2010) Terminal 3p deletions in two families—correlation between molecular karyotype and phenotype. Am J Med Genet A 152:441–446
14. Miura S, Shibata H et al (2006) The contactin 4 gene locus at 3p26 is a candidate gene of SCA16. Neurology 67:1236–1241
15. Glessner J, Wang K et al (2009) Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. Nature 459:569–573
16. Kaneko-Goto T, Yoshihara S, Miyazaki H, Yoshihara Y (2008) BIG-2 mediates olfactory axon convergence to target glomeruli. Neuron 57:834–846
17. Osterfield M, Egelund R, Young L, Flanagan J (2008) Interaction of amyloid precursor protein with contactins and NgCAM in the retinotectal system. Development 135:1189–1199
18. Strzelecki J, Mikulska K, Lekka M, Kulik A, Balter A, Nowak W (2009) AFM force spectroscopy and steered molecular dynamics simulation of protein contactin 4. Acta Phys Pol A 116:S156–S159
19. Dabrowska A, Lebed K, Lekka M, Lekki J, Kwiatek W (2006) A comparison between the unfolding of fibronectin and contactin. J Phys Condens Matter 18:10157–10164
20. Bork P, Holm L, Sander C (1994) The immunoglobulin fold: structural classification, sequence patterns and common core. J Mol Biol 242:309–320
21. Halaby D, Poupon A, Mornon J (1999) The immunoglobulin fold family: sequence analysis and 3D structure comparisons. Protein Eng Des Sel 12:563–571
22. Halaby D, Mornon J (1998) The immunoglobulin superfamily: an insight on its tissular, species, and functional diversity. J Mol Evol 46:389–400
23. Volbeda A, Hol W (1989) Crystal structure of hexameric haemocyanin from *Panulirus interruptus* refined at 3.2 Å resolution. J Mol Biol 209:249–279
24. Stacklies W, Vega M, Wilmanns M, Gräter F (2009) Mechanical network in titin immunoglobulin from force distribution analysis. PLoS Comput Biol 5:1–11
25. Best R, Fowler S, Toca Herrera J, Steward A, Paci E, Clarke J (2003) Mechanical unfolding of a titin Ig domain: structure of transition state revealed by combining atomic force microscopy, protein engineering and molecular dynamics simulations. J Mol Biol 330:867–877
26. Gao M, Wilmanns M, Schulten K (2002) Steered molecular dynamics studies of titin I1 domain unfolding. Biophys J 83:3435–3445
27. Lu H, Isralewitz B, Krammer A, Vogel V, Schulten K (1998) Unfolding of titin immunoglobulin domains by steered molecular dynamics simulation. Biophys J 75:662–671

28. Zeng L, Zhang C et al (2002) A novel splice variant of the cell adhesion molecule contactin 4 (CNTN4) is mainly expressed in human brain. J Hum Genet 47:497–499
29. Hansford L, Smith S, Haber M, Norris M, Cheung B, Marshall G (2000) Cloning and characterization of the human neural cell adhesion molecule, CNTN4 (alias BIG-2). Cytogenet Genome Res 101:17–23
30. Guex N, Peitsch M (1997) SWISS-MODEL and the Swiss-Pdb Viewer: an environment for comparative protein modeling. Electrophoresis 18:2714–2723
31. Kiefer F, Arnold K, Kunzli M, Bordoli L, Schwede T (2009) The SWISS-MODEL Repository and associated resources. Nucleic Acids Res 37:D387–D392
32. Arnold K, Bordoli L, Kopp J, Schwede T (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics 22:195–201
33. Schwede T, Kopp J, Guex N, Peitsch M (2003) SWISS-MODEL: an automated protein homology-modeling server. Nucleic Acids Res 31:3381–3385
34. Jorgensen W, Chandrasekhar J, Madura J, Impey R, Klein M (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79:926–935
35. Phillips J, Braun R et al (2005) Scalable molecular dynamics with NAMD. J Comput Chem 26:1781–1802
36. MacKerell A Jr, Bashford D et al (1998) All-atom empirical potential for molecular modeling and dynamics studies of protein. J Phys Chem B 102:3586–3616
37. Humphrey W, Dalke A, Schulten K (1996) VMD: Visual Molecular Dynamics. J Mol Graph 14:33–38
38. Merkel R, Nassoy P, Leung A, Ritchie K, Evans E (1999) Energy landscapes of receptor–ligand bonds explored with dynamic force spectroscopy. Nature 397:50–53
39. Bouyain S, Watkins DJ (2010) The protein tyrosine phosphatases PTPRZ and PTPRG bind to distinct members of the contactin family of neural recognition molecules. Proc Natl Acad Sci USA 107:2443–2448
40. Evans E, Ritchie K (1997) Dynamic strength of molecular adhesion bonds. Biophys J 72:1541–1555
41. Mitternacht S, Luccioli S, Torcini A, Imparato A, Irback A (2009) Changing the mechanical unfolding pathway of FnIII10 by tuning the pulling strength. Biophys J 96:429–441
42. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ (2009) Jalview Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics 25:1189–1191