



Article

# Structural Characterization of the CD44 Stem Region for Standard and Cancer-Associated Isoforms

Kun-Lin Chen <sup>1</sup>, Deng Li <sup>2</sup>, Ting-Xuan Lu <sup>3</sup> and Shu-Wei Chang <sup>2,\*</sup>

<sup>1</sup> Department of Engineering Science and Ocean Engineering, National Taiwan University, Taipei 10617, Taiwan; pen39251099@gmail.com

<sup>2</sup> Department of Civil Engineering, National Taiwan University, Taipei 10617, Taiwan; uplideng@gmail.com

<sup>3</sup> Department and Graduate Institute of Pharmacology, National Taiwan University, Taipei 10617, Taiwan; tingxuanlu19@gmail.com

\* Correspondence: changsw@ntu.edu.tw; Tel.: +886-2-3366-4245

Received: 7 November 2019; Accepted: 31 December 2019; Published: 3 January 2020



**Abstract:** CD44 is widely expressed in most vertebrate cells, whereas the expression of CD44v6 is restricted to only a few tissues and has been considered to be associated with tumor progression and metastasis. Thus, CD44v6 has been recognized as a promising prognostic biomarker and therapeutic target for various cancers for more than a decade. However, despite many experimental studies, the structural dynamics and differences between CD44s and CD44v6, particularly in their stem region, still remain elusive. Here, a computational study was conducted to address these problems. We found that the stem of CD44s adopted predominantly two conformations, one featuring antiparallel  $\beta$ -sheets and the other featuring parallel  $\beta$ -sheets, whereas the stem of CD44v6 adopted mainly one conformation with relatively highly suppressed  $\beta$ -sheet contents. Moreover, Phe215 was found to be essential in the  $\beta$ -sheets of both CD44s and CD44v6. We finally found intramolecular Phe215–Trp224 hydrogen-bonding interactions and hydrophobic interactions with Phe215 that cooperatively drove conformational differences upon the addition of the v6 region to CD44. Our study elucidated the structural differences between the stem regions of CD44s and CD44v6 and thus can offer useful structural information for drug design to specifically target CD44v6 in promising clinical applications.

**Keywords:** CD44; CD44v6; receptor tyrosine kinase (RTK); metastasis; tumor progression;  $\beta$ -sheets; molecular dynamics simulations

## 1. Introduction

CD44 is a member of the cartilage link protein family, with hyaluronan (HA) as its primary ligand [1,2]. CD44 is a type I transmembrane receptor protein that engages in many cellular processes, such as regulating cell–cell and cell–matrix adhesion and controlling cell proliferation, differentiation, migration, and survival [3]. CD44 proteins have a common structure consisting of four major parts: The extracellular HA-binding domain (HABD), the stem region, the transmembrane segment, and the cytoplasmic region [4]. The HABD confers binding capability to CD44 mainly for HA, but also for collagen, laminin, and fibronectin [5–7].

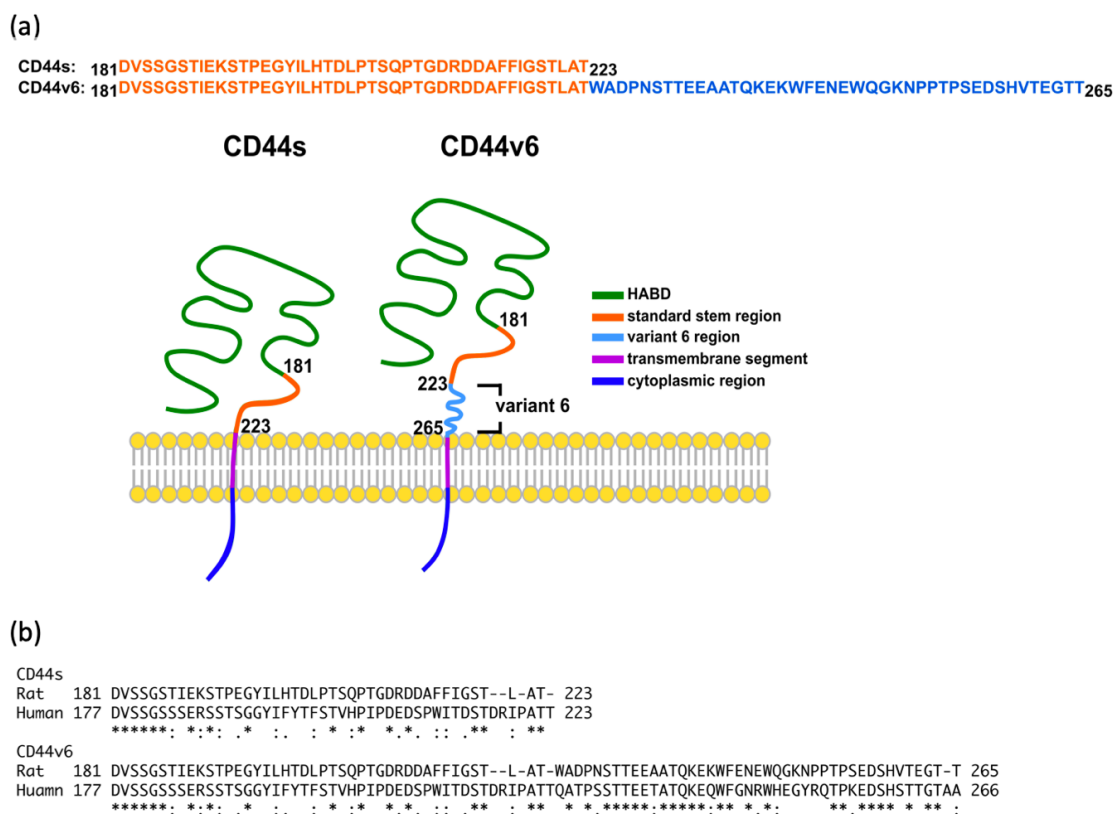
CD44 proteins are encoded by a single and highly conserved gene consisting of 20 exons [8]. Exons 1–5, 16–18, and 20 encode the smallest, the standard, and the hematopoietic isoform (“CD44s”), which is ubiquitously expressed in many vertebrate cells, including leukocytes, endothelial cells, neurons, and fibroblasts [4,9,10]. Ten variant exons, 6–15 (commonly called “v1–v10”), which are in the middle of the CD44 gene, can be alternatively spliced, yielding a wide variety of CD44 variant (CD44v) isoforms with different functions and structures in the stem region [8,11]. The heterogeneity of CD44

isoforms can be increased through decoration with *N*-glycans, *O*-glycans, and glycosaminoglycans, such as heparan sulfate and chondroitin sulfate [12–15].

Unlike CD44s, the expression of CD44v isoforms is restricted to only a few tissues [16,17]. CD44v isoforms have attracted keen interest because of their expression in a majority of tumors and the induction of the metastatic phenotype [9,18,19]. There is now abundant evidence for the crucial role of CD44v expression in the aggressive stages of various cancers as well as for its expression in cancer-initiating cells (also known as cancer stem cells) [20–25]. Among CD44v isoforms, CD44 variant 6 (CD44v6) plays a major role in cancer progression because its expression level correlates with a poor prognosis in many cancers, including colorectal cancer, lung cancer, hepatocellular carcinoma, prostate cancer, pancreatic cancer, endometrial cancer, and head and neck squamous cell carcinoma [26–28]. Furthermore, a recent study identified that CD44v6 is a marker of cancer stem cells in colon cancer and accounts for the metastatic phenotype of the tumors [29]. Therefore, CD44v6 is considered a promising prognostic biomarker and therapeutic target for various cancers [26,30–32]. CD44v6 proteins contain an exon v6-encoded region that can interact with the other major cytokines produced by the tumor microenvironment and with receptor tyrosine kinases (RTKs), such as c-Met, Ron, and vascular endothelial growth factor receptor 2 (VEGFR-2): Therefore they function as coreceptors, which plausibly explains how CD44v6 drives tumor progression and metastasis [33–35]. An alanine-scanning mutational analysis was used to determine that three amino acids (AAs) in the v6 sequence are required for CD44v6 to act as a coreceptor for RTKs [36].

Numerous experimental and computational studies on CD44 have been conducted. In addition to numerous experiments concentrating on its pathological functions and related signaling pathways [21,37], much work has focused on the molecular-level details of its structure and dynamics, particularly on HABD [38–42]. Several CD44 structures identified using X-ray crystallography or nuclear magnetic resonance (NMR) imaging have been deposited to the Protein Data Bank (PDB) [43–46]. Understanding of the nature of CD44 and its isoforms has considerably progressed. However, to the best of our knowledge, molecular-level insight into the structure and dynamics of CD44s and CD44v6 is largely missing because of the lack of suitable experimental techniques that can capture the structural information of the dynamic region within a protein at fine spatial and temporal resolution [47]. Because of the lack of information regarding the structural differences between CD44s and CD44v6, achieving a structure-based rational drug design remains difficult, which slows down the development efficiency of strategies to specifically target CD44v6 in clinical applications [48,49].

Therefore, in the present study, we aimed to explore the conformational dynamics of the stem region of CD44s and CD44v6 and to characterize the differences between them. We conducted all-atom explicit-solvent molecular dynamics (MD) simulations to explore the structure of the CD44s and CD44v6 stem regions, which is the only region that differs between these two isoforms in terms of AA sequences (Figure 1a). MD simulation uses physics-based force fields and is a computational approach that can provide valuable insight into protein dynamic characteristics at an atomic level, which is extremely difficult to achieve with any experimental technique [47,50,51]. Consequently, MD simulation serves as a powerful tool to complement experimental techniques for characterizing the conformational ensembles of the CD44s and CD44v6 stem regions.



**Figure 1.** (a) Sequence comparison of CD44s and CD44v6 isoform of rat. Different regions are shown in different colors. (b) Sequence comparison of human and rat. An \* (asterisk) indicates positions which have a single, fully conserved residue. A: (colon) indicates conservation between groups of strongly similar properties. A. (period) indicates conservation between groups of weakly similar properties.

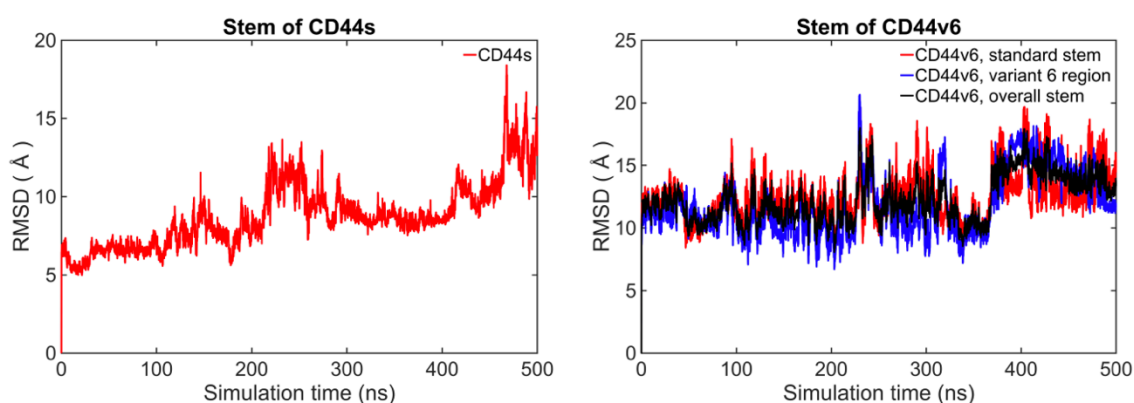
Since the CD44s and CD44v6 stem regions have not been structurally solved using X-ray crystallography or NMR imaging, we generated 3D models of these proteins by using the threading method. Long-timescale MD simulations were employed to sample their conformations. The sampled conformations were analyzed for further characterization. We termed the “CD44s stem region” the “standard stem region”, and the “CD44v6 stem region” is composed of the “standard stem region” and “variant 6 region”. The AA sequences of the standard stem region of CD44s and CD44v6 were exactly the same (Figure 1a). In previous studies, rat models were widely used in CD44-related research. Figure 1b provide the sequence comparison of human and rat. Previous studies have successfully proved the significance of CD44v6 in rat models [52–54]. Thus, we began our study with rat sequence.

## 2. Results

### 2.1. CD44s Stem Region Adopted Predominantly Two Conformations, Whereas CD44v6 Mainly Adopted One

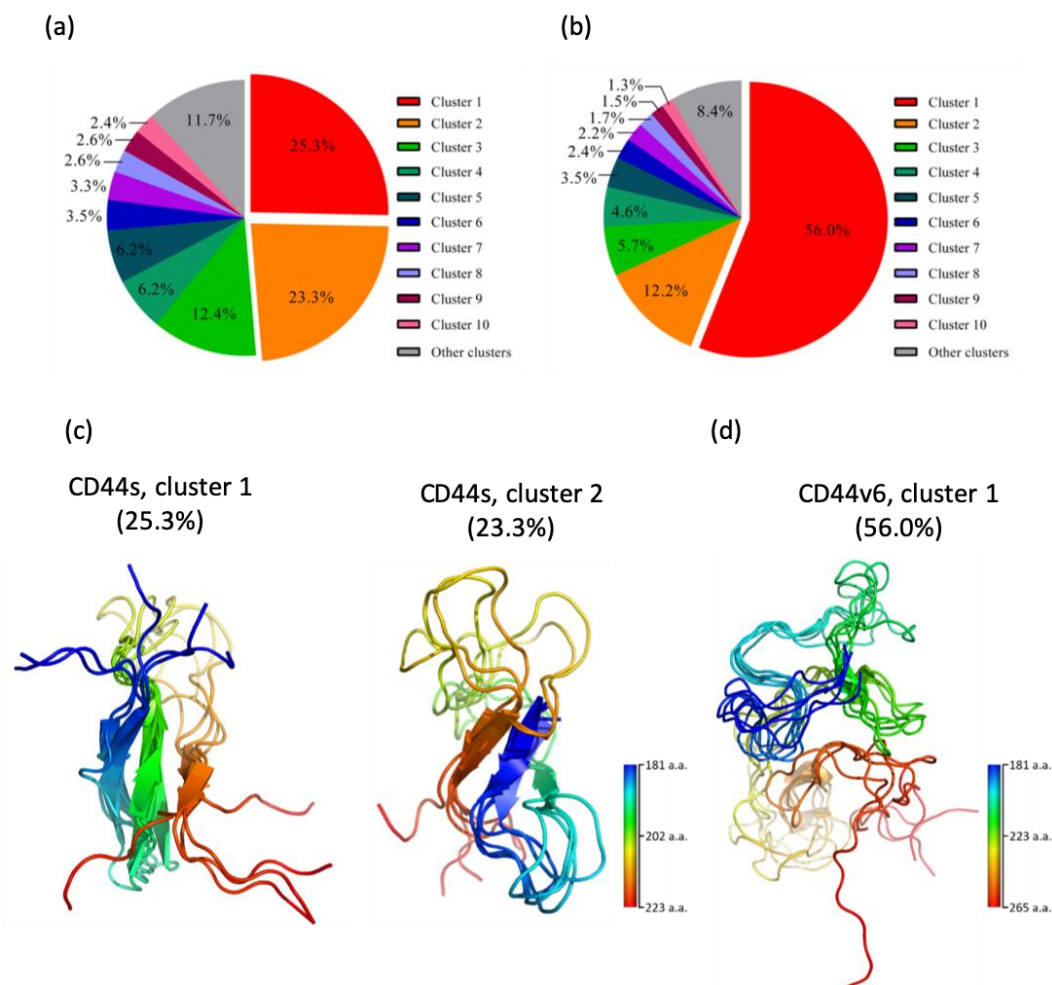
Due to the low sequence similarity (<30%) of solved proteins in the PDB (Figure S1), the threading method was employed to obtain a rough model of target sequences instead of homology modeling [55,56]. After sending the protein sequence to the I-TASSER, a popular server system that develops 3D models by using multiple threading alignments [57], five rough 3D models were obtained. The best C-scores (the confidence score of the predicted model) of the first two models for the CD44s stem region were −2.04 and −4.10, and those for the CD44v6 stem region were −3.31 and −3.69. The four models were further assessed using VERIFY3D to assess the protein models with three-dimensional profiles: 100% and 69.77% of the residues in the two models of the CD44s stem region exhibited an average 3D–1D score of >0.2, whereas percentages of 48% and 29% were obtained for the CD44v6 stem region. The models with lower percentages were discarded.

The selected models of the CD44s and CD44v6 stem regions were then energy-minimized. Subsequently, two atomistic MD simulations were performed separately for a long time duration of 500 ns on the two systems—the stem region of CD44 and the stem region of CD44v6—to explore their conformational dynamics and structural properties. The root mean square deviations (RMSDs) were monitored over a 500 ns trajectory to view the extent of conformational sampling overall. The RMSDs relative to average structure are shown in Figure 2. The RMSDs of the standard stem region for CD44s and CD44v6 both mainly ranged from 6 Å to 16 Å. A large fluctuation in the RMSD value of the backbone atoms for both cases indicated dynamic behavior of the standard stem region for both CD44s and CD44v6. Notably, no considerable variation in the RMSD of the standard stem region backbone of both CD44s and CD44v6 was observed from 330 to 470 ns.



**Figure 2.** The RMSD comparison between the CD44s stem region and the CD44v6 stem region based on average structure.

To characterize the conformational ensemble sampled using MD for detailed investigation, we performed a clustering analysis on the standard stem region in the CD44s and CD44v6 systems. In this calculation, we classified the sampled structures into distinct clusters on the basis of their RMSD values with a cutoff of 8.5 Å. The population distributions of the first 10 most populated clusters are shown in Figure 3a,b. Through this analysis, we found two predominant conformations in our simulations of the CD44s stem region, with population distributions of 25.3% (cluster 1) and 23.3% (cluster 2), which together represented approximately 50% of all conformations. In contrast, a single major conformation was detected for the CD44v6 stem region, with 56.0% (cluster 1) of the population of all conformations. The structures obtained from the clustering analysis are presented in Figure 3c,d. It is apparent that the most populated clusters of the CD44s stem region, cluster 1 and cluster 2, were both mixed with loops and  $\beta$ -sheet secondary structures, whereas the most populated cluster of the CD44v6 stem region, cluster 1, exhibited much less  $\beta$ -sheet content. A significant structural difference between cluster 1 and cluster 2 in CD44s was the  $\beta$ -sheet arrangement. In cluster 1,  $\beta$ -sheets were arranged in an antiparallel manner consisting of three  $\beta$ -strands. In contrast, in cluster 2,  $\beta$ -sheets were arranged in a parallel manner and consisted of two  $\beta$ -strands. These observations are in agreement with the calculation indicating that cluster 1 had a higher population than cluster 2 insofar as the antiparallel  $\beta$ -sheet was more abundant and energetically favored than the parallel  $\beta$ -sheet was [58,59]. Due to the aforementioned results, we selected the most populated clusters of the CD44s stem region (cluster 1 and cluster 2) and cluster 1 in the CD44v6 stem region for further analyses.



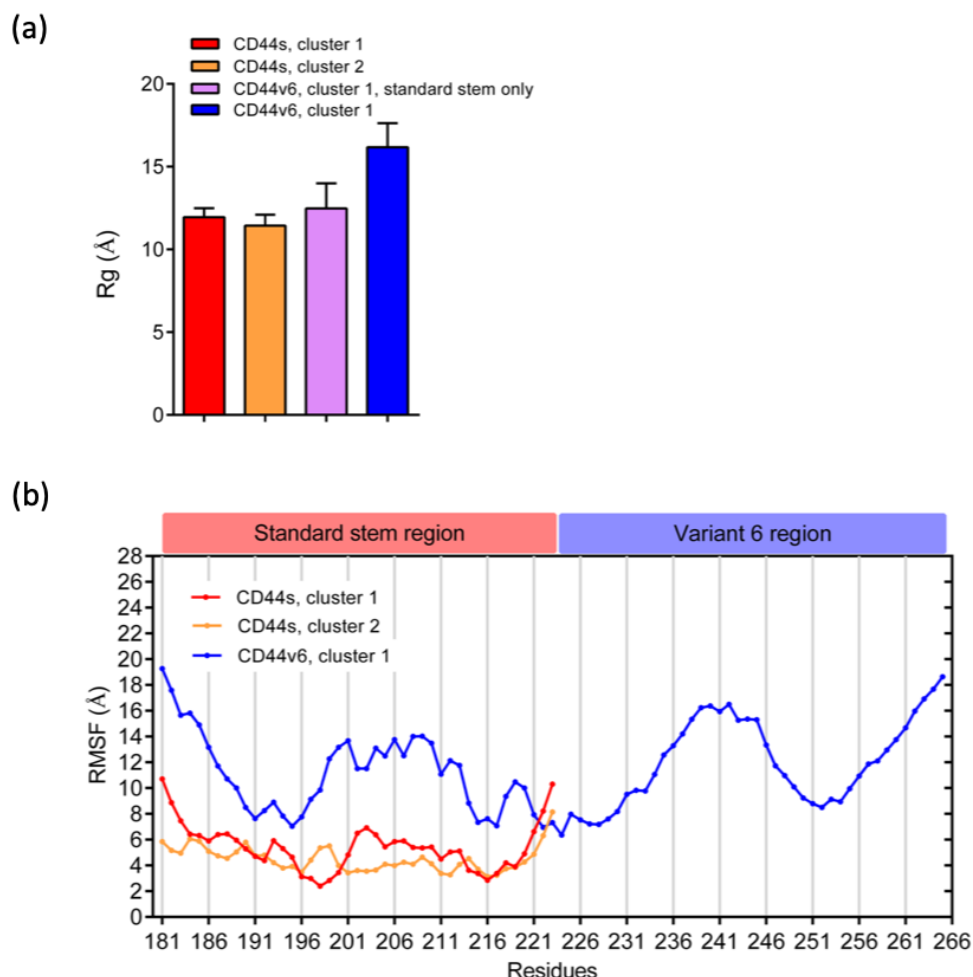
**Figure 3.** Clustering analysis of the CD44s and CD44v6 stem regions. (a) Population distribution of each cluster of the CD44s stem region. (b) Population distribution of each cluster of the CD44v6 stem region. (c) Structures of the most populated clusters of the CD44s stem region, cluster 1 and cluster 2. (d) Structures of the most populated cluster of the CD44v6 stem region, cluster 1. Ensemble structures are rainbow-colored from blue at the N-terminus to red at the C-terminus.

## 2.2. The Standard Stem Region of CD44v6 Was Much More Flexible than CD44s

We calculated the radius of gyration (Rg) to understand the levels of compaction of the CD44s and CD44v6 stem regions. Rg is defined as the mass-weighted root mean square distance of a collection of atoms from their center of mass [60]. Therefore, this analysis could provide the overall dimensions of the protein. The Rg values of the standard stem region in CD44s and CD44v6 were quite similar and were all approximately 12 Å, as shown in Figure 4a. Although the conformational differences between CD44s and CD44v6 in the standard stem region were evident, as mentioned previously, the Rg results indicated that the v6 region did not affect the standard stem region by markedly changing its size. However, considering the overall CD44v6 stem region, we could see an increase in the Rg value to an average of 16.17 Å, as expected due to an increase in the number of AAs.

To determine whether the addition of v6 affected the dynamic behavior of the standard stem on particular residues, we calculated the C $\alpha$  root mean square fluctuation (RMSF) of their stem region to measure the flexibilities of each residue (Figure 4b). The CD44v6 stem region exhibited an overall higher RMSF than did the CD44s stem region, suggesting a perturbation effect upon the addition of v6 to the stem region. Notably, for cluster 1 of CD44s, two minimums at the residue regions of 196–200 and 214–217 were observed, indicating the low flexibility of these two regions. In contrast, the overall RMSF of CD44s cluster 2 was relatively flat and slightly lower than that of cluster 1. In

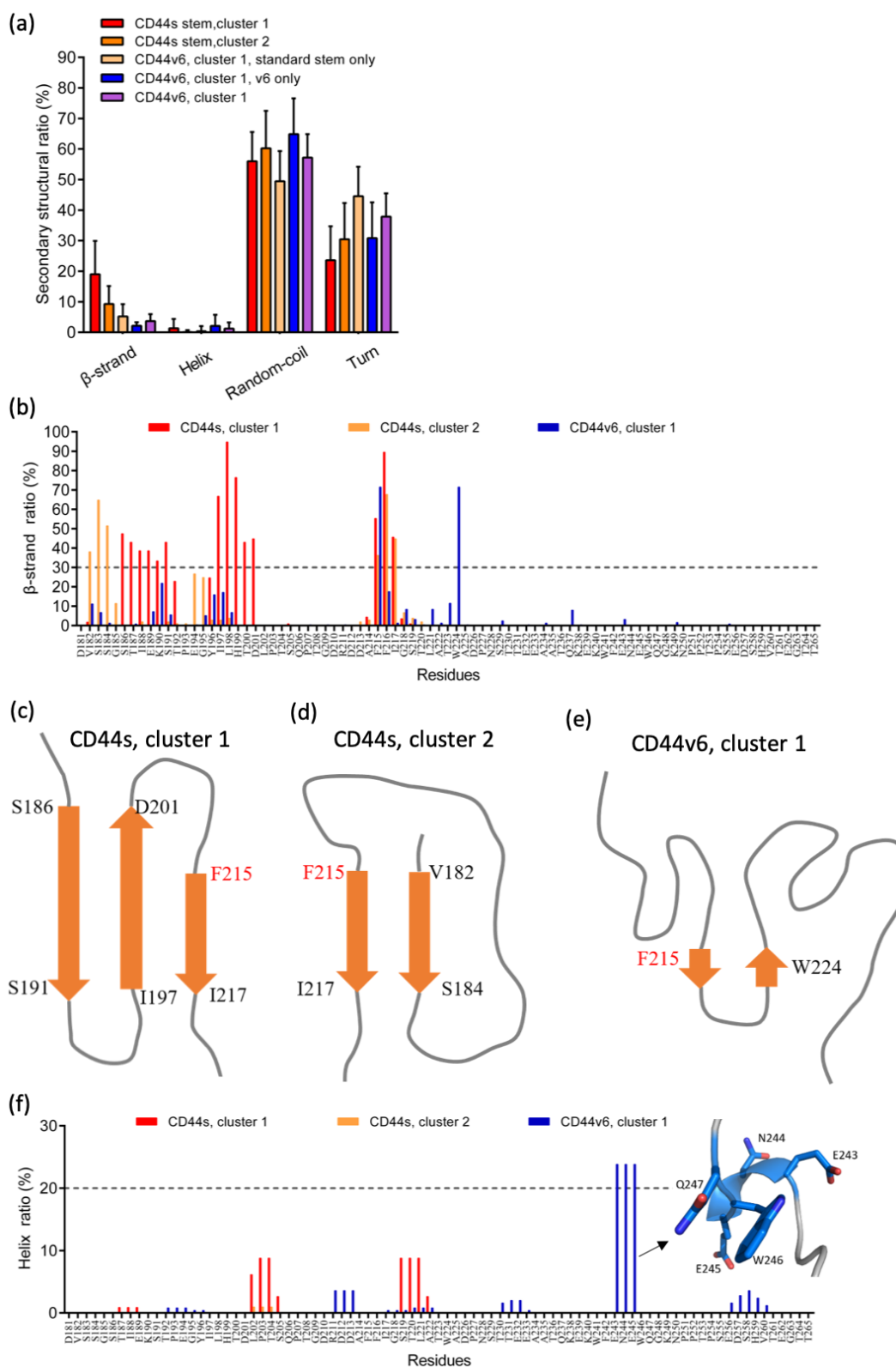
CD44v6 cluster 1, local minimums were clearly observed in the residue regions of 190–196, 215–217, 222–224, and 250–254, indicating the low flexibility of these regions. Taken together, these results demonstrated that when v6 was present, there was a gain in flexibility and an altered fluctuation of residues in the standard stem region without an effect on the overall compactness.



**Figure 4.** Analysis of (a) the radius of gyration (Rg) and (b) the root mean square fluctuation (RMSF) for the C $\alpha$  atoms of the most populated clusters.

### 2.3. Phe215 Displayed a High Tendency to Form $\beta$ -Strands in CD44s and CD44v6, Whereas the CD44v6 Stem Region Exhibited Fewer $\beta$ -Sheets than CD44s Did

To quantitatively investigate the observed  $\beta$ -sheet structures, we calculated the  $\beta$ -sheet average ratio for all residues and the residue-based composition ratio of CD44s and CD44v6 by using the protein secondary structure assignment algorithm STRIDE [61]. The ratio of the other secondary structures (helix, coil, and turn) was also analyzed, as shown in Figure 5a. The standard stem region of CD44v6 cluster 1 exhibited much fewer  $\beta$ -sheets compared to CD44s, as shown in Figure 3c,d. The  $\beta$ -sheet contents for cluster 1 and cluster 2 of CD44s were 19.0% and 9.3%, respectively, and the ratio was reduced to 5.2% with the addition of the v6 region. In random-coil contents, a similar tendency of a decrease in the ratio upon the addition of the v6 region was observed. In contrast, the turn contents for cluster 1 and cluster 2 of CD44s were 23.6% and 30.5%, which increased to 44.6% when the v6 region was present. All of the ratios of the helical structure except for the v6 region were considerably small (<1.5%). These results indicated that the introduction of v6 could considerably inhibit  $\beta$ -sheet and coil formation and promote turn formation in the standard stem region.



**Figure 5.** Analysis with secondary structure and regular secondary structure ( $\beta$ -sheet and helix) composition on residue base. (a) Average ratio of each secondary structure (including helix,  $\beta$ -sheet, coil, and turn) for all residues and (b)  $\beta$ -sheet ratio as a function of amino acid residue. Illustrations of the  $\beta$ -strand arrangements of (c) the CD44s stem region, cluster 1; (d) the CD44s stem region, cluster 2; and (e) the CD44v6 stem region, cluster 1. (f) Helix ratio as a function of amino acid residue.

Figure 5b shows the propensities of  $\beta$ -sheets as a function of AA residues. The residue-based  $\beta$ -sheet ratios displayed quite different distributions. In CD44s cluster 1, AAs 186–191, 197–201, and 215–217 indicated a strong tendency (>30%) to form  $\beta$ -sheets. We found that these regions favored the low fluctuated residues in the RMSF plot (Figure 4b), which allowed us to elucidate the structurally stable region in the conformation. The antiparallel  $\beta$ -sheet structure that consisted of the three  $\beta$ -strands is depicted in Figure 5c. In CD44s cluster 2, only AAs 182–184 and 215–217 exhibited >30%  $\beta$ -sheet tendency, and the two  $\beta$ -strands assembled into a parallel  $\beta$ -sheet, as depicted in Figure 5d. Although two regions formed  $\beta$ -sheet structures, the RMSF analysis did not indicate relatively low flexibility in these regions, as was apparent in CD44s cluster 1 and CD44v6 cluster 1. This behavior was observed because parallel  $\beta$ -sheets are generally less stable than antiparallel  $\beta$ -sheets [62]. In CD44v6 cluster 1, only Phe215 and Trp224 exhibited high  $\beta$ -sheet propensity (both 71.43%), together forming a substantially short antiparallel  $\beta$ -sheet, as depicted in Figure 5e. These two residues were located in the local minimum of the RMSF analysis, which explained the local structural stability. These analyses identified the crucial residues involved in the  $\beta$ -sheet structures that affected the local structure flexibility and were consistent with the RMSF investigation results.

Notably, the involvement of residue Phe215 in the  $\beta$ -sheet structure was conserved in both CD44s and CD44v6, implying a crucial role of Phe215 in the cooperativity of  $\beta$ -sheets. The high  $\beta$ -sheet propensity of phenylalanine has been observed in previous studies [63,64]. Notably, Phe215 was situated at the first position in all three  $\beta$ -sheets. In addition, although the overall helical contents were considerably low for all cases (Figure 5a), we found that Glu243, Asn244, and Glu245 within v6 exhibited a considerably higher helical tendency compared to the other residues, as shown in Figure 5f. These three residues were adjacent only to the three AAs 245EWQ247, which were identified by mutational analysis to be critical for the coreceptor function of CD44v6, as described in the introduction [36]. The results present a structural view around 245EWQ247 within the v6 sequence, which is essential for the collaboration between CD44v6 and RTKs, which induces the cancer metastatic process [33].

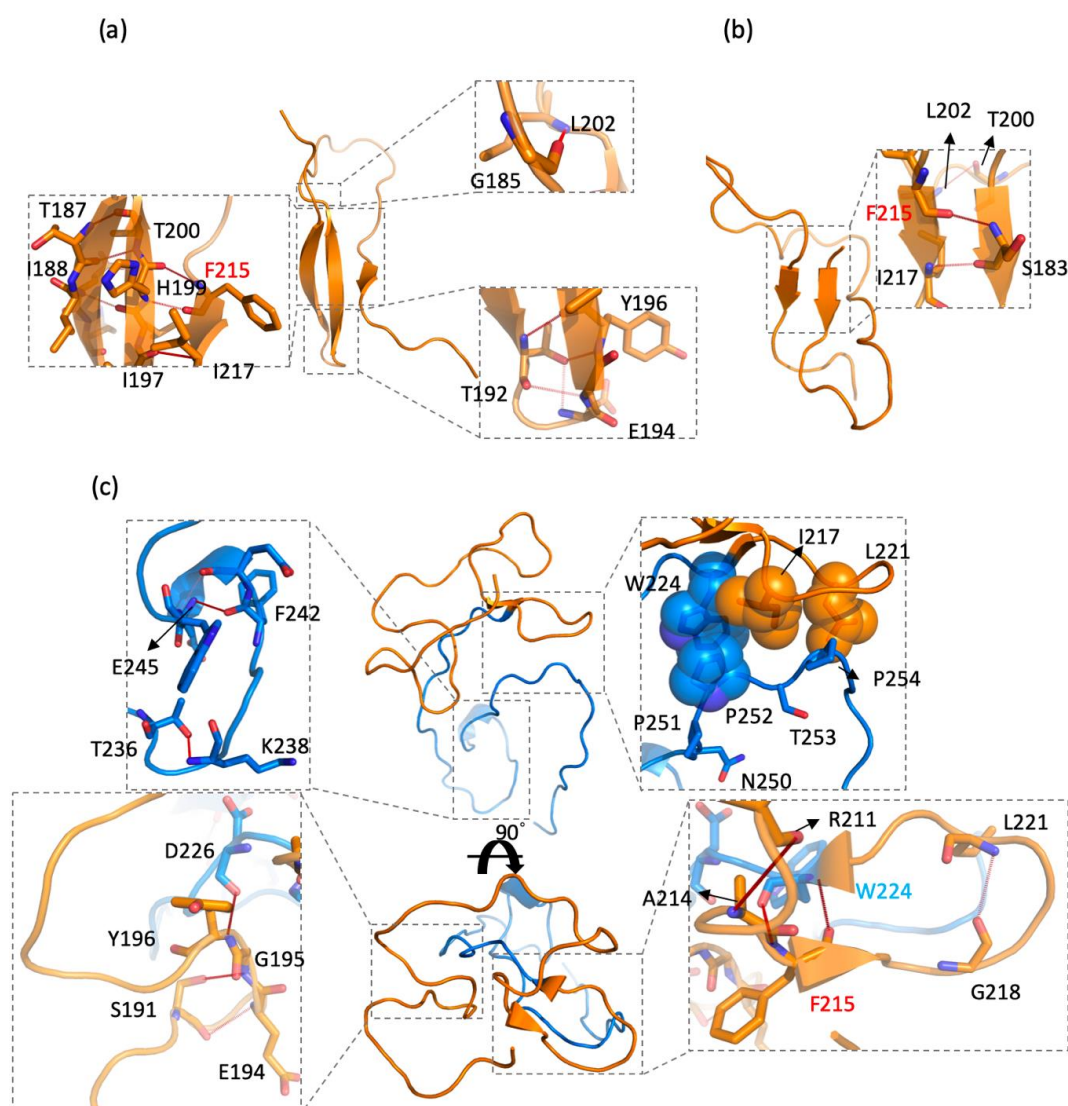
#### 2.4. CD44s and CD44v6 Utilized Phe125 with Different Interaction Partners to Form Hydrogen Bonds (Hbonds) to Stabilize $\beta$ -Sheets

Hydrogen-bonding interactions have long been considered crucial in the stability of a protein structure [65,66]. To identify the key residues for hydrogen bonds (Hbonds), we have presented the Hbonds interaction pairs with an occupancy of >35% in CD44s and CD44v6 in Tables S1–S3. Among all of the residues, Phe215 was the only AA that formed Hbonds in CD44s and CD44v6, but with different interaction partners. The main chain of Phe215 formed persistent Hbonds with His199 (occupancy of 88.60%) in CD44s cluster 1, Ser183 (occupancy of 65.71%) in CD44s cluster 2, and Trp224 (occupancy of 75.40% and 63.10%). The strong occupancies of Phe215 in the Hbonds indicated the crucial role of the residue in stabilizing  $\beta$ -sheet structures, which was consistent with the aforementioned observations. Moreover, the results may suggest that Trp224 in the v6 region competed with His199 and Ser183 for the formation of Hbonds with Phe215, which caused the  $\beta$ -sheet structures, disrupting the standard stem region. A structural view of Hbond interactions is shown in Figure 6.

Although we performed an Hbonds analysis for both the side and main chains of all residues, a majority of persistent Hbonds pairs were between the main chains in both CD44s and CD44v6, revealing that backbone–backbone chain Hbonds were the main conformation stabilizing forces. As presented in Tables S1 and S2, more than half of the Hbonds facilitated  $\beta$ -sheet formation. In CD44s cluster 1 (Figure 6a), 6 out of 11 persistent Hbonds were involved in a  $\beta$ -sheet structure, whereas in CD44s cluster 2, two out of three Hbonds were involved (Figure 6b), indicating a major contribution of Hbonds in CD44s. Notably, in CD44v6 cluster 1 (Figure 6c), three persistent Hbonds between the standard stem region and the v6 region prevented the two regions from separating completely but allowed for individual activity in each region. Additionally, a set of stable Hbonds between Glu194 and Ser191, Gly195 and Ser191, and Tyr196 and Asp226 considerably contributed to the stabilization of



the loop that encompassed AAs 191–196, as shown in Figure 6c, thus leading to less flexibility in AAs 191–196, as analyzed in the RMSF plot (Figure 4b).

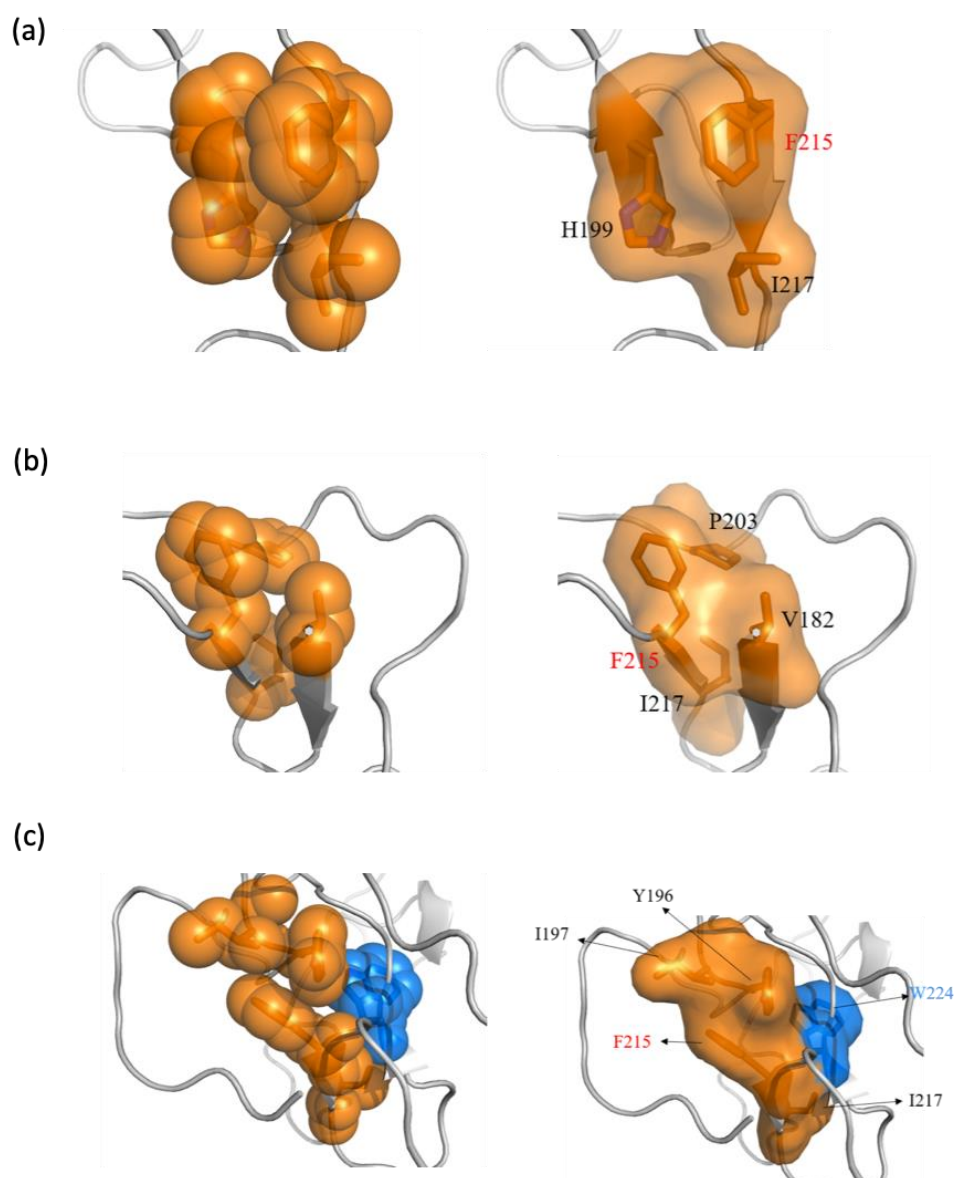


**Figure 6.** Detailed hydrogen bonds and intramolecular interaction networks: (a) the CD44s stem region, cluster 1 (also see Table S1); (b) the CD44s stem region, cluster 2 (also see Table S2); (c) the CD44v6 stem region, cluster 1 (also see Table S3). The standard stem region is colored orange, and the v6 region is colored blue. The key residues involved in interactions are shown in sticks. The key residues involved in the hydrophobic contacts with Pro252 are shown in spheres. Hydrogen bonds are depicted with red dash lines.

In the RMSF plot, the cause of the region of AAs (250–254) in CD44v6 that appeared to be less flexible is still unknown, because there were no helices,  $\beta$ -sheets, or even any persistent Hbonds in the region. To address this problem, we thoroughly investigated these five residues, 250NPPTP254. It was actually a proline-rich region, which is generally less flexible because proline exhibits a unique side chain that wraps around to form a covalent bond with the backbone, severely limiting backbone motility [67,68]. In addition, considering the hydrophobic side chain of proline, we calculated the hydrophobic interaction with Pro251, Pro252, and Pro254. However, persistent hydrophobic contacts (occupancy of >35%) were only formed with Pro252, which was 49.2% for Pro252–Ile217, 40.5% for Pro252–Trp224, and 38.1% for Pro215–Leu221. The hydrophobic interaction network with 250NPPTP254 is shown in Figure 6c.

### 2.5. Phe215 in CD44v6 Interacted with Surrounding Hydrophobic Residues More than CD44s, Leading to Becoming Blocked

Due to the aromatic and highly hydrophobic side chain of phenylalanine, it is highly likely that hydrophobic and aromatic stacking interactions with Phe215 contribute to consolidating the structure [69]. Thus, we performed a hydrophobic/aromatic interaction analysis on Phe215. The identified hydrophobic and aromatic residues are listed in Table S4. The hydrophobic and aromatic interaction network with Phe215 is shown in Figure 7. The analysis revealed that, compared to Phe215 in CD44s, the Phe215 in CD44v6 interacted more with hydrophobic residues, two of which contained an aromatic ring, namely Tyr196 and Trp224. Notably, Trp224 in v6 not only interacted with Phe215 through the aforementioned strong main chain Hbonds but also through side chain aromatic interaction, reflecting the strong tendency of Trp224 to interact with Phe215.



**Figure 7.** The hydrophobic and aromatic interaction network with the key residue Phe215 in (a) the CD44s stem region, cluster 1; (b) the CD44s stem region, cluster 2; and (c) the CD44v6 stem region, cluster 1. The residues involved in the hydrophobic and aromatic interactions are colored orange and are shown in sticks with spheres (left) and surfaces (right).

To evaluate the contribution of hydrophobic and aromatic stacking interactions, we computed the interaction energy (sum of columbic and van der Waals energies) between the key residues in Table S4 and Phe215. As shown in Table 1, the Phe215 in CD44v6 exhibited the lowest average interaction energy of  $-2.476$  kcal/mol, which was primarily contributed by the van der Waals effect. Thus, the interaction mode with the surrounding hydrophobic residues for Phe215 was more favored in CD44v6. We also computed the buried percentage of the solvent-accessible surface area (SASA) of Phe215, which is a measure of the AA surfaces exposed to the solvent. As shown in Table 2, the Phe215 in CD44v6 exhibited a higher buried SASA (70.96%) percentage than did Phe215 in CD44s because of shielding from water exposure by the surrounding hydrophobic residues. Taken together, these results reveal that both hydrogen bonding and hydrophobic interactions cooperatively drive the local interaction network around Phe215, leading to a conformational change in the stem region upon the addition of the v6 region.

**Table 1.** Interaction energy (kcal/mol) between the side chain of Phe215 and the residues that were involved in the persistent hydrophobic contacts listed in Table S4. All of the values are presented as mean  $\pm$  SD.

	van der Waals Energy	Electrostatic Energy	Interaction Energy
CD44s, cluster 1	$-1.196 \pm 1.058$	$-0.1625 \pm 0.3979$	$-1.358 \pm 1.243$
CD44s, cluster 2	$-1.534 \pm 0.8957$	$0.1097 \pm 0.2475$	$-1.424 \pm 0.9442$
CD44v6, cluster 1	$-2.394 \pm 1.445$	$-0.08138 \pm 0.2759$	$-2.476 \pm 1.485$

**Table 2.** Solvent-accessible surface area (SASA) for the side chain of the Phe215 in CD44s and CD44v6. All values are presented as mean  $\pm$  SD.

	SASA ( $\text{\AA}^2$ )	Buried Percentage (%)
CD44s, cluster 1	$114.30 \pm 31.24$	$39.85 \pm 16.44\%$
CD44s, cluster 2	$92.19 \pm 31.64$	$51.48 \pm 16.65\%$
CD44v6, cluster 1	$55.18 \pm 36.08$	$70.96 \pm 18.99\%$

### 3. Materials and Methods

#### 3.1. Protein Model Preparation

Since the stem region of CD44s and CD44v6 have not been structurally characterized with NMR or X-ray crystallography, it was necessary to generate and validate 3D models of these proteins. In order to build a three-dimensional model of protein structures, amino acid sequences of the stem region of CD44s and CD44v6 were retrieved from the Uniprot database (P26051). The amino acid sequences for variant exon 6 were identified from previous studies [70,71]. The sequences for the stem region of CD44s and CD44v6 were 181DVSSGSTIEKSTPEGYILHTDLPTSQPTGDRDDAFFIGSTLAT223 and 181DVSSGSTIEKSTPEGYILHTDLPTSQPTGDRDDAFFIGSTLATWADPNSTTE EAATQKEKWFENE WQGNPP TPSEDSHVTEGTT265 (the v6 region is underlined), respectively. The sequences were used as inputs for the Iterative Threading ASSEMBLY Refinement (I-Tasser) server, which combines various techniques including threading, ab initio modeling, and structure refinement approaches [57,72,73]. I-TASSER produced five models for both of the sequences submitted, and only the first two models with the best C-scores for both stem regions were retained. The structural quality of the provided models were assessed using the Verify3D program (<http://servicesn.mbi.ucla.edu/Verify3D/>) [74,75].

#### 3.2. Molecular Dynamics Simulations

The validated model of the stem region of CD44s and CD44v6 was subjected to all-atom explicit-solvent MD simulations using NAMD 2.13b2 [76] with a CHARMM36 force field [77–80]. Two systems were set up: One was the solvated stem region of CD44s, and the other was the solvated

stem region of CD44v6. Both of the systems were solvated in a rectangular box of TIP3P water molecules using the solvate plug-in of visual molecular dynamics (VMD) [81], which extended at least 15 Å from the surface of the protein, and the systems were neutralized with NaCl using the autoionize plug-in of VMD. The salt (NaCl) concentration was set to 0.15 mol/L. Both of the systems were energy-minimized for 1000 steps followed by a 500 ns MD simulation performed at 310 K and 1 atm under the isothermal-isobaric (NPT) ensemble. The systems were simulated in periodic boundary conditions using the Langevin algorithm to maintain the temperature at 310 K and the Langevin piston Nose–Hoover method to keep the pressure constant at 1 atm. The particle mesh Ewald (PME) method with a grid size of less than 1 Å was used to calculate the electrostatic interactions [82]. The SHAKE algorithm was applied to constrain all of the covalent bonds with hydrogen, and the time step was set to 2 fs [83]. Van der Waals interactions were computed using a cutoff of 12 Å with a switching function starting at 10 Å. Nonbonded interactions were fully scaled starting with 1–4 bonded atoms. The first 50 ns over the simulation trajectory was treated as equilibration. The trajectory from 50 ns to 500 ns was recorded every 1 ns, and thus 450 frames were collected for analysis.

### 3.3. Data Analysis

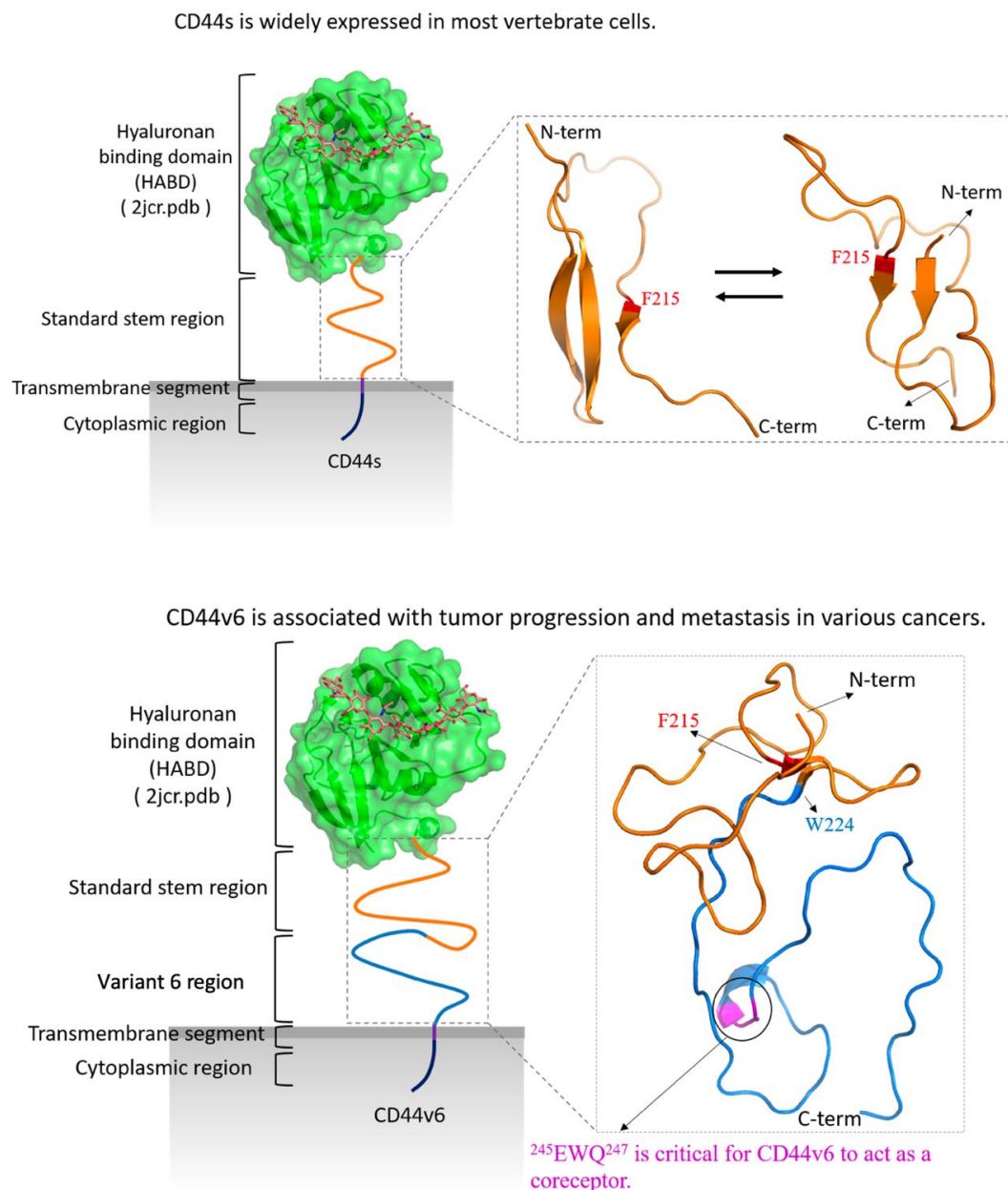
Root mean square deviations (RMSDs) were calculated for the entire simulation trajectory with reference to their first frames for both proteins. Extracted protein snapshots were then clustered using the clustering tool in VMD at a RMSD cutoff value of 8.5 Å, based on backbone atoms of the standard stem region (i.e., AAs 181–223) for CD44s and CD44v6. The root mean square fluctuations (RMSFs) and radius of gyration (Rg) were calculated by using the coordinates of the C $\alpha$  atoms. Secondary structural content was calculated by STRIDE [58] with the Timeline plug-in of the VMD package. The criteria to consider a hydrogen bond were a donor–acceptor distance less than 3.5 Å and a donor hydrogen–acceptor angle less than 45°. Hydrophobic and aromatic contacts were considered to exist if the distances between atoms in the side chains of hydrophobic/aromatic and hydrophobic/aromatic residues (Val, Leu, Met, Ile, Pro, Trp, Phe, His, and Tyr) were smaller than 5 Å. The solvent-accessible surface area (SASA) was measured with a spherical probe of 1.4 Å, mimicking a water molecule. The buried percentage of SASA was obtained by the formula (the buried SASA)/190.02  $\times$  100%, where 190.02 is the fully exposed SASA of the side chain of phenylalanine in our calculation. All structural images were generated using PyMOL (PyMOL Molecular Graphics System, Version 1.3, Schrödinger, LLC).

## 4. Conclusions

Our results revealed that the differences in amino acid sequences between CD44s and CD44v6 led to different structures. By performing long-timescale all-atom MD simulations, we found that the stem region of CD44s adopted predominantly two conformations: (1) antiparallel  $\beta$ -sheets with a population of 25.3% and (2) parallel  $\beta$ -sheets with a slightly lower population of 23.3%. However, the CD44v6 stem region adopted mainly one conformation with a population of 56%, and the  $\beta$ -sheet contents were highly suppressed compared to the CD44s stem region. Notably, when v6 was present, there was a gain in flexibility and an altered fluctuation of residues in the standard stem region that did not influence overall compactness. It is worth noting that the insertion of other variants might also alter the CD44s stem region. The computational approach proposed in this work could be applied to resolve the molecular structures of other variants. We have also performed a molecular simulation with the insertion of v7 and the results confirm that the  $\beta$ -sheets contents were also highly suppressed (Figure S2 and Table S5 in Supplementary Materials).

In addition, Glu243, Asn244, and Glu245 within v6 exhibited a relatively higher helical ratio than did the other residues. Their proximity to 245EWQ247 was identified as being essential for CD44v6 to act as a coreceptor. Notably, we found that Phe215 was the only residue that participated in the formation of  $\beta$ -sheets in both CD44s and CD44v6, highlighting its crucial role in stabilizing the structure. The residue Phe215 formed persistent Hbonds with either His199 or Ser183 in CD44s. However,

Phe215 formed persistent Hbonds with Trp224 in CD44v6. The observations indicated that the residue Trp224 in the v6 region played a competing role against His199 and Ser183 to form strong Hbonds with Phe215, which led to a completely different arrangement of  $\beta$ -sheet structures. The interaction energy also indicated that the Phe215 in CD44v6 interacted with its surrounding hydrophobic residues more energetically and favorably than did the Phe215 in CD44s. Therefore, intramolecular Phe215–Trp224 hydrogen-bonding and hydrophobic interactions with Phe215 cooperatively drove the conformational differences upon addition of the v6 region to CD44 (Figure 8).



**Figure 8.** Schematic for the molecular structures and biological characteristics of CD44s (top) and CD44v6 (bottom). The structure of the HABD (hyaluronan-binding domain) was retrieved from the Protein Data Bank (PDB) (ID: 2JCR), and hyaluronan is shown in sticks. The identified key residue Phe215 is colored red in both CD44s and CD44v6. The standard stem region is colored orange, and the v6 region is colored blue. The amino acids (AAs) 245EWQ247, which are required for CD44v6 to act as the coreceptor for receptor tyrosine kinases (RTKs), including c-Met, Ron, and VEGFR-2, are colored magenta.

In summary, this is the first study that has reported molecular-level details of the stem regions of CD44s and CD44v6. The results obtained from this study will be useful for future studies on structure-based drug design to precisely target CD44v6 and yield promising clinical strategies.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/1422-0067/21/1/336/s1>.

**Author Contributions:** K.-L.C. and S.-W.C. designed the study and project. K.L.C. performed the computational model. K.L.C. and S.-W.C. interpreted the results. D.L. validated the results. K.L.C. drafted the manuscript. D.L. and T.-X.L. performed editing. K.L.C., D.L., T.-X.L., and S.-W.C. discussed and approved the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Ministry of Science and Technology (MOST 108-2636-E-002-016 and MOST 108-2218-E-007-037), Taiwan (R.O.C), and National Taiwan University (NTU-CC-108L891102 and NTU-CC-109L891002).

**Acknowledgments:** The authors appreciate the financial support from the Ministry of Science and Technology (MOST 108-2636-E-002-016 and MOST 108-2218-E-007-037), Taiwan (R.O.C), and National Taiwan University (NTU-CC-108L891102 and NTU-CC-109L891002). We are grateful to the National Center for High-Performance Computing for computer time and facilities.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

HA	Hyaluronan
HABD	HA-binding domain
RTKs	Receptor tyrosine kinases
VEGFR-2	vascular endothelial growth factor receptor 2
NMR	Nuclear magnetic resonance
PDB	Protein Data Bank
MD	Molecular dynamics
RMSD	Root mean square deviation
Rg	Radius of gyration
RMSF	Root mean square fluctuation
SASA	Solvent-accessible surface area

## References

1. Stamenkovic, I.; Amiot, M.; Pesando, J.M.; Seed, B. A lymphocyte molecule implicated in lymph node homing is a member of the cartilage link protein family. *Cell* **1989**, *56*, 1057–1062. [[CrossRef](#)]
2. Aruffo, A.; Stamenkovic, I.; Melnick, M.; Underhill, C.B.; Seed, B. CD44 is the principal cell surface receptor for hyaluronate. *Cell* **1990**, *61*, 1303–1313. [[CrossRef](#)]
3. Ponta, H.; Sherman, L.; Herrlich, P.A. CD44: From adhesion molecules to signalling regulators. *Nat. Rev. Mol. Cell Biol.* **2003**, *4*, 33. [[CrossRef](#)] [[PubMed](#)]
4. Dzwonek, J.; Wilczynski, G.M. CD44: Molecular interactions, signaling and functions in the nervous system. *Front. Cell. Neurosci.* **2015**, *9*, 175. [[CrossRef](#)] [[PubMed](#)]
5. Banerji, S.; Day, A.J.; Kahmann, J.D.; Jackson, D.G. Characterization of a Functional Hyaluronan-Binding Domain from the Human CD44 Molecule Expressed in *Escherichia coli*. *Protein Expr. Purif.* **1998**, *14*, 371–381. [[CrossRef](#)] [[PubMed](#)]
6. Ishii, S.; Ford, R.; Thomas, P.; Nachman, A.; Steele, G., Jr.; Jessup, J. CD44 participates in the adhesion of human colorectal carcinoma cells to laminin and type IV collagen. *Surg. Oncol.* **1993**, *2*, 255–264. [[CrossRef](#)]
7. Jalkanen, S.; Jalkanen, M. Lymphocyte CD44 binds the COOH-terminal heparin-binding domain of fibronectin. *J. Cell Biol.* **1992**, *116*, 817–825. [[CrossRef](#)]
8. Sreaton, G.R.; Bell, M.V.; Jackson, D.G.; Cornelis, F.B.; Gerth, U.; Bell, J. Genomic structure of DNA encoding the lymphocyte homing receptor CD44 reveals at least 12 alternatively spliced exons. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 12160–12164. [[CrossRef](#)]
9. Naor, D.; Wallach-Dayana, S.B.; Zahalka, M.A.; Sionov, R.V. Involvement of CD44, a molecule with a thousand faces, in cancer dissemination. In *Hyaluronan in Cancer Biology*; Elsevier: Amsterdam, The Netherlands, 2009; pp. 127–146.

10. Acharya, P.S.; Majumdar, S.; Jacob, M.; Hayden, J.; Mrass, P.; Wening, W.; Assoian, R.K.; Puré, E. Fibroblast migration is mediated by CD44-dependent TGF $\beta$  activation. *J. Cell Sci.* **2008**, *121*, 1393–1402. [[CrossRef](#)]
11. Stamenkovic, I.; Aruffo, A.; Amiot, M.; Seed, B. The hematopoietic and epithelial forms of CD44 are distinct polypeptides with different adhesion potentials for hyaluronate-bearing cells. *EMBO J.* **1991**, *10*, 343–348. [[CrossRef](#)]
12. Bartolazzi, A.; Nocks, A.; Aruffo, A.; Spring, F.; Stamenkovic, I. Glycosylation of CD44 is implicated in CD44-mediated cell adhesion to hyaluronan. *J. Cell Biol.* **1996**, *132*, 1199–1208. [[CrossRef](#)] [[PubMed](#)]
13. Bennett, K.L.; Modrell, B.; Greenfield, B.; Bartolazzi, A.; Stamenkovic, I.; Peach, R.; Jackson, D.G.; Spring, F.; Aruffo, A. Regulation of CD44 binding to hyaluronan by glycosylation of variably spliced exons. *J. Cell Biol.* **1995**, *131*, 1623–1633. [[CrossRef](#)] [[PubMed](#)]
14. Bennett, K.L.; Jackson, D.G.; Simon, J.C.; Tanczos, E.; Peach, R.; Modrell, B.; Stamenkovic, I.; Plowman, G.; Aruffo, A. CD44 isoforms containing exon V3 are responsible for the presentation of heparin-binding growth factor. *J. Cell Biol.* **1995**, *128*, 687–698. [[CrossRef](#)] [[PubMed](#)]
15. Greenfield, B.; Wang, W.-C.; Marquardt, H.; Piepkorn, M.; Wolff, E.A.; Aruffo, A.; Bennett, K.L. Characterization of the heparan sulfate and chondroitin sulfate assembly sites in CD44. *J. Biol. Chem.* **1999**, *274*, 2511–2517. [[CrossRef](#)] [[PubMed](#)]
16. Ruiz, P.; Schwärzler, C.; Günthert, U. CD44 isoforms during differentiation and development. *Bioessays* **1995**, *17*, 17–24. [[CrossRef](#)] [[PubMed](#)]
17. Goodison, S.; Urquidi, V.; Tarin, D. CD44 cell adhesion molecules. *Mol. Pathol.* **1999**, *52*, 189. [[CrossRef](#)]
18. Günthert, U.; Hofmann, M.; Rudy, W.; Reber, S.; Zöller, M.; Haußmann, I.; Matzku, S.; Wenzel, A.; Ponta, H.; Herrlich, P. A new variant of glycoprotein CD44 confers metastatic potential to rat carcinoma cells. *Cell* **1991**, *65*, 13–24. [[CrossRef](#)]
19. Hofmann, M.; Rudy, W.; Zöller, M.; Tölg, C.; Ponta, H.; Herrlich, P.; Günthert, U. CD44 splice variants confer metastatic behavior in rats: Homologous sequences are expressed in human tumor cell lines. *Cancer Res.* **1991**, *51*, 5292–5297.
20. Pries, R.; Wittkopf, N.; Trenkle, T.; Nitsch, S.M.; Wollenberg, B. Potential stem cell marker CD44 is constitutively expressed in permanent cell lines of head and neck cancer. *In Vivo* **2008**, *22*, 89–92.
21. Zöller, M. CD44: Can a cancer-initiating cell profit from an abundantly expressed molecule? *Nat. Rev. Cancer* **2011**, *11*, 254. [[CrossRef](#)]
22. Dalerba, P.; Dylla, S.J.; Park, I.-K.; Liu, R.; Wang, X.; Cho, R.W.; Hoey, T.; Gurney, A.; Huang, E.H.; Simeone, D.M. Phenotypic characterization of human colorectal cancer stem cells. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 10158–10163. [[CrossRef](#)] [[PubMed](#)]
23. Lee, C.J.; Dosch, J.; Simeone, D.M. Pancreatic cancer stem cells. *J. Clin. Oncol.* **2008**, *26*, 2806–2812. [[CrossRef](#)] [[PubMed](#)]
24. Zhang, S.; Balch, C.; Chan, M.W.; Lai, H.-C.; Matei, D.; Schilder, J.M.; Yan, P.S.; Huang, T.H.; Nephew, K.P. Identification and characterization of ovarian cancer-initiating cells from primary human tumors. *Cancer Res.* **2008**, *68*, 4311–4320. [[CrossRef](#)] [[PubMed](#)]
25. Zeilstra, J.; Joosten, S.; Van Andel, H.; Tolg, C.; Berns, A.; Snoek, M.; Van De Wetering, M.; Spaargaren, M.; Clevers, H.; Pals, S. Stem cell CD44v isoforms promote intestinal cancer formation in Apc (min) mice downstream of Wnt signaling. *Oncogene* **2014**, *33*, 665. [[CrossRef](#)]
26. Yan, Y.; Zuo, X.; Wei, D. Concise review: Emerging role of CD44 in cancer stem cells: A promising biomarker and therapeutic target. *Stem Cells Transl. Med.* **2015**, *4*, 1033–1043. [[CrossRef](#)]
27. Chen, C.; Zhao, S.; Karnad, A.; Freeman, J.W. The biology and role of CD44 in cancer progression: Therapeutic implications. *J. Hematol. Oncol.* **2018**, *11*, 64. [[CrossRef](#)]
28. Ma, L.; Dong, L.; Chang, P. CD44v6 engages in colorectal cancer progression. *Cell Death Dis.* **2019**, *10*, 30. [[CrossRef](#)]
29. Todaro, M.; Gaggianesi, M.; Catalano, V.; Benfante, A.; Iovino, F.; Biffoni, M.; Apuzzo, T.; Sperduti, I.; Volpe, S.; Cocorullo, G. CD44v6 is a marker of constitutive and reprogrammed cancer stem cells driving colon cancer metastasis. *Cell Stem Cell* **2014**, *14*, 342–356. [[CrossRef](#)]
30. Leuci, V.; Casucci, M.; Grignani, G.; Rotolo, R.; Rossotti, U.; Vigna, E.; Gammaitoni, L.; Mesiano, G.; Fiorino, E.; Donini, C. CD44v6 as innovative sarcoma target for CAR-redirectioned CIK cells. *Oncoimmunology* **2018**, *7*, e1423167. [[CrossRef](#)]

31. Kennedy, P.J.; Sousa, F.; Ferreira, D.; Pereira, C.; Nestor, M.; Oliveira, C.; Granja, P.L.; Sarmento, B. Fab-conjugated PLGA nanoparticles effectively target cancer cells expressing human CD44v6. *Acta Biomater.* **2018**, *81*, 208–218. [[CrossRef](#)]
32. Odenthal, J.; Rijpkema, M.; Bos, D.; Wagena, E.; Croes, H.; Grenman, R.; Boerman, O.; Takes, R.; Friedl, P. Targeting CD44v6 for fluorescence-guided surgery in head and neck squamous cell carcinoma. *Sci. Rep.* **2018**, *8*, 10467. [[CrossRef](#)] [[PubMed](#)]
33. Orian-Rousseau, V. CD44, a therapeutic target for metastasising tumours. *Eur. J. Cancer* **2010**, *46*, 1271–1277. [[CrossRef](#)] [[PubMed](#)]
34. Orian-Rousseau, V.; Morrison, H.; Matzke, A.; Kastilan, T.; Pace, G.; Herrlich, P.; Ponta, H. Hepatocyte growth factor-induced Ras activation requires ERM proteins linked to both CD44v6 and F-actin. *Mol. Biol. Cell* **2007**, *18*, 76–83. [[CrossRef](#)] [[PubMed](#)]
35. Tremmel, M.; Matzke, A.; Albrecht, I.; Laib, A.M.; Olaku, V.; Ballmer-Hofer, K.; Christofori, G.; Héroult, M.; Augustin, H.G.; Ponta, H. A CD44v6 peptide reveals a role of CD44 in VEGFR-2 signaling and angiogenesis. *Blood* **2009**, *114*, 5236–5244. [[CrossRef](#)] [[PubMed](#)]
36. Matzke, A.; Herrlich, P.; Ponta, H.; Orian-Rousseau, V. A five-amino-acid peptide blocks Met-and Ron-dependent cell migration. *Cancer Res.* **2005**, *65*, 6105–6110. [[CrossRef](#)] [[PubMed](#)]
37. Orian-Rousseau, V. CD44 acts as a signaling platform controlling tumor progression and metastasis. *Front. Immunol.* **2015**, *6*, 154. [[CrossRef](#)]
38. Vuorio, J.; Vattulainen, I.; Martinez-Seara, H. Atomistic fingerprint of hyaluronan–CD44 binding. *PLoS Comput. Biol.* **2017**, *13*, e1005663. [[CrossRef](#)]
39. Guvench, O. Revealing the mechanisms of protein disorder and N-glycosylation in CD44-hyaluronan binding using molecular simulation. *Front. Immunol.* **2015**, *6*, 305. [[CrossRef](#)]
40. Favreau, A.J.; Faller, C.E.; Guvench, O. CD44 receptor unfolding enhances binding by freeing basic amino acids to contact carbohydrate ligand. *Biophys. J.* **2013**, *105*, 1217–1226. [[CrossRef](#)]
41. Jamison, F.W., II; Foster, T.J.; Barker, J.A.; Hills, R.D., Jr.; Guvench, O. Mechanism of binding site conformational switching in the CD44–hyaluronan protein–carbohydrate binding interaction. *J. Mol. Biol.* **2011**, *406*, 631–647. [[CrossRef](#)]
42. Patel, S.; Shaikh, F.; Devaraji, V.; Radadiya, A.; Shah, K.; Shah, A.; Rawal, R. Insights into the structural perturbations of spliced variants of CD44: A modeling and simulation approach. *J. Biomol. Struct. Dyn.* **2017**, *35*, 354–367. [[CrossRef](#)] [[PubMed](#)]
43. Teriete, P.; Banerji, S.; Noble, M.; Blundell, C.D.; Wright, A.J.; Pickford, A.R.; Lowe, E.; Mahoney, D.J.; Tammi, M.I.; Kahmann, J.D. Structure of the regulatory hyaluronan binding domain in the inflammatory leukocyte homing receptor CD44. *Mol. Cell* **2004**, *13*, 483–496. [[CrossRef](#)]
44. Banerji, S.; Wright, A.J.; Noble, M.; Mahoney, D.J.; Campbell, I.D.; Day, A.J.; Jackson, D.G. Structures of the Cd44–hyaluronan complex provide insight into a fundamental carbohydrate-protein interaction. *Nat. Struct. Mol. Biol.* **2007**, *14*, 234. [[CrossRef](#)] [[PubMed](#)]
45. Takeda, M.; Ogino, S.; Umemoto, R.; Sakakura, M.; Kajiwara, M.; Sugahara, K.N.; Hayasaka, H.; Miyasaka, M.; Terasawa, H.; Shimada, I. Ligand-induced structural changes of the CD44 hyaluronan-binding domain revealed by NMR. *J. Biol. Chem.* **2006**, *281*, 40089–40095. [[CrossRef](#)] [[PubMed](#)]
46. Liu, L.-K.; Finzel, B. High-resolution crystal structures of alternate forms of the human CD44 hyaluronan-binding domain reveal a site for protein interaction. *Acta Crystallogr. Sect. F Struct. Biol. Commun.* **2014**, *70*, 1155–1161. [[CrossRef](#)]
47. Hollingsworth, S.A.; Dror, R.O. Molecular dynamics simulation for all. *Neuron* **2018**, *99*, 1129–1143. [[CrossRef](#)]
48. Śledź, P.; Caflisch, A. Protein structure-based drug design: From docking to molecular dynamics. *Curr. Opin. Struct. Biol.* **2018**, *48*, 93–102. [[CrossRef](#)]
49. Durrant, J.D.; McCammon, J.A. Molecular dynamics simulations and drug discovery. *BMC Biol.* **2011**, *9*, 71. [[CrossRef](#)]
50. Rauscher, S.; Pomès, R. Molecular simulations of protein disorder. *Biochem. Cell Biol.* **2010**, *88*, 269–290. [[CrossRef](#)]
51. Karplus, M.; McCammon, J.A. Molecular dynamics simulations of biomolecules. *Nat. Struct. Mol. Biol.* **2002**, *9*, 646. [[CrossRef](#)]



52. Matzke-Ogi, A.; Jannasch, K.; Shatirishvili, M.; Fuchs, B.; Chiblak, S.; Morton, J.; Tawk, B.; Lindner, T.; Sansom, O.; Alves, F.; et al. Inhibition of Tumor Growth and Metastasis in Pancreatic Cancer Models by Interference with CD44v6 Signaling. *Gastroenterology* **2016**, *150*, 513–525. [[CrossRef](#)] [[PubMed](#)]
53. Takazoe, K.; Foti, R.; Tesch, G.H.; Hurst, L.A.; Lan, H.Y.; Atkins, R.C.; Nikolic-Paterson, D.J. Up-regulation of the tumour-associated marker CD44V6 in experimental kidney disease. *Clin. Exp. Immunol.* **2000**, *121*, 523–532. [[CrossRef](#)] [[PubMed](#)]
54. Hallouin, F.; Goupille, C.; leCabellec, M.; Bara, J.; lePendou, J. Expression of A and H blood-group and of CD44 antigens during chemical rat colonic carcinogenesis. *Glycoconj. J.* **1997**, *14*, 801–808. [[CrossRef](#)] [[PubMed](#)]
55. Khor, B.Y.; Tye, G.J.; Lim, T.S.; Choong, Y.S. General overview on structure prediction of twilight-zone proteins. *Theor. Biol. Med. Model.* **2015**, *12*, 15. [[CrossRef](#)] [[PubMed](#)]
56. Padmanabhan, S. *Handbook of Pharmacogenomics and Stratified Medicines/Edited by Sandosh Padmanabhan, Institute of Cardiovascular and Medical Sciences; University of Glasgow: Glasgow, UK, 2014.*
57. Roy, A.; Kucukural, A.; Zhang, Y. I-TASSER: A unified platform for automated protein structure and function prediction. *Nat. Protoc.* **2010**, *5*, 725–738. [[CrossRef](#)] [[PubMed](#)]
58. Perczel, A.; Gaspari, Z.; Csizmadia, I.G. Structure and stability of beta-pleated sheets. *J. Comput. Chem.* **2005**, *26*, 1155–1168. [[CrossRef](#)]
59. Zhao, Y.L.; Wu, Y.D. A theoretical study of beta-sheet models: Is the formation of hydrogen-bond networks cooperative? *J. Am. Chem. Soc.* **2002**, *124*, 1570–1571. [[CrossRef](#)]
60. Lobanov, M.Y.; Bogatyreva, N.S.; Galzitskaya, O.V. Radius of gyration as an indicator of protein structure compactness. *Mol. Biol.* **2008**, *42*, 623–628. [[CrossRef](#)]
61. Frishman, D.; Argos, P. Knowledge-based protein secondary structure assignment. *Proteins* **1995**, *23*, 566–579. [[CrossRef](#)]
62. Ranganathan, S.; Nakai, K.; Schonbach, C. *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*; Elsevier: Amsterdam, The Netherlands, 2018.
63. Minor, D.L.; Kim, P.S. Measurement of the Beta-Sheet-Forming Propensities of Amino-Acids. *Nature* **1994**, *367*, 660–663. [[CrossRef](#)]
64. Street, A.G.; Mayo, S.L. Intrinsic beta-sheet propensities result from van der Waals interactions between side chains and the local backbone. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 9074–9076. [[CrossRef](#)] [[PubMed](#)]
65. Dobson, C.M. Protein folding and misfolding. *Nature* **2003**, *426*, 884. [[CrossRef](#)] [[PubMed](#)]
66. Hubbard, R.E.; Kamran Haider, M. Hydrogen bonds in proteins: Role and strength. *eLS* **2010**. [[CrossRef](#)]
67. Ho, B.K.; Coutsias, E.A.; Seok, C.; Dill, K.A. The flexibility in the proline ring couples to the protein backbone. *Protein Sci.* **2005**, *14*, 1011–1018. [[CrossRef](#)] [[PubMed](#)]
68. Williamson, M.P. The structure and function of proline-rich regions in proteins. *Biochem. J.* **1994**, *297*, 249. [[CrossRef](#)] [[PubMed](#)]
69. Burley, S.; Petsko, G.A. Aromatic-aromatic interaction: A mechanism of protein structure stabilization. *Science* **1985**, *229*, 23–28. [[CrossRef](#)] [[PubMed](#)]
70. Tolg, C.; Hofmann, M.; Herrlich, P.; Ponta, H. Splicing Choice from 10 Variant Exons Establishes Cd44 Variability. *Nucleic Acids Res.* **1993**, *21*, 1225–1229. [[CrossRef](#)]
71. Sreaton, G.R.; Bell, M.V.; Bell, J.I.; Jackson, D.G. The Identification of a New Alternative Exon with Highly Restricted Tissue Expression in Transcripts Encoding the Mouse Pgp-1 (Cd44) Homing Receptor - Comparison of All 10 Variable Exons between Mouse, Human, and Rat. *J. Biol. Chem.* **1993**, *268*, 12235–12238.
72. Yang, J.Y.; Yan, R.X.; Roy, A.; Xu, D.; Poisson, J.; Zhang, Y. The I-TASSER Suite: Protein structure and function prediction. *Nat. Methods* **2015**, *12*, 7–8. [[CrossRef](#)]
73. Yang, J.Y.; Zhang, Y. I-TASSER server: New development for protein structure and function predictions. *Nucleic Acids Res.* **2015**, *43*, W174–W181. [[CrossRef](#)]
74. Bowie, J.U.; Luthy, R.; Eisenberg, D. A Method to Identify Protein Sequences That Fold into a Known 3-Dimensional Structure. *Science* **1991**, *253*, 164–170. [[CrossRef](#)] [[PubMed](#)]
75. Luthy, R.; Bowie, J.U.; Eisenberg, D. Assessment of Protein Models with 3-Dimensional Profiles. *Nature* **1992**, *356*, 83–85. [[CrossRef](#)]
76. Phillips, J.C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R.D.; Kale, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781–1802. [[CrossRef](#)]

77. Best, R.B.; Zhu, X.; Shim, J.; Lopes, P.E.M.; Mittal, J.; Feig, M.; MacKerell, A.D. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone phi, psi and Side-Chain chi(1) and chi(2) Dihedral Angles. *J. Chem. Theory Comput.* **2012**, *8*, 3257–3273. [[CrossRef](#)] [[PubMed](#)]
78. Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B.L.; Grubmuller, H.; MacKerell, A.D. CHARMM36m: An improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **2017**, *14*, 71–73. [[CrossRef](#)] [[PubMed](#)]
79. MacKerell, A.D.; Bashford, D.; Bellott, M.; Dunbrack, R.L.; Evanseck, J.D.; Field, M.J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616. [[CrossRef](#)]
80. MacKerell, A.D.; Feig, M.; Brooks, C.L. Improved treatment of the protein backbone in empirical force fields. *J. Am. Chem. Soc.* **2004**, *126*, 698–699. [[CrossRef](#)]
81. Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph. Model* **1996**, *14*, 33–38. [[CrossRef](#)]
82. Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald—an N.Log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092. [[CrossRef](#)]
83. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H.J. Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).