Article

Analysis of Procollagen C-Proteinase Enhancer-1/Glycosaminoglycan Binding Sites and of the Potential Role of Calcium Ions in the Interaction

Jan Potthoff [1,2,†]**, Krzysztof K. Bojarski** [1,†] **, Gergely Kohut** [3,4] **, Agnieszka G. Lipska** [1] **, Adam Liwo** [1]**, Efrat Kessler** [5]**, Sylvie Ricard-Blum** [6] **and Sergey A. Samsonov** [1,*]

[1]   Faculty of Chemistry, University of Gdańsk, ul. Wita Stwosza 63, 80-308 Gdańsk, Poland; potthoff.jan@googlemail.com (J.P.); krzysztof.bojarski@ug.edu.pl (K.K.B.); lypstyk@gmail.com (A.G.L.); adam.liwo@ug.edu.pl (A.L.)
[2]   Institute of Chemistry and Biochemistry, Free University of Berlin, Takustr. 3, 14195 Berlin, Germany
[3]   Institute of Materials and Environmental Chemistry, Research Centre for Natural Sciences, Hungarian Academy of Sciences, Magyar tudósok körútja 2, 1117 Budapest, Hungary; kohut.gergely@ttk.mta.hu
[4]   MTA-ELTE Research Group of Peptide Chemistry, Hungarian Academy of Sciences, Eötvös Loránd University, Budapest 112, P.O. Box 32, 1518 Budapest, Hungary
[5]   Maurice and Gabriela Goldschleger Eye Research Institute, Tel-Aviv University Sackler Faculty of Medicine, Sheba Medical Center, Tel-Hashomer, 52621 Ramat Gan, Israel; ekessler@post.tau.ac.il
[6]   Institute of Molecular and Supramolecular Chemistry and Biochemistry, UMR 5246, CNRS, INSA Lyon, CPE, University Claude Bernard Lyon 1, Univ Lyon, 69622 Villeurbanne CEDEX, France; sylvie.ricard-blum@univ-lyon1.fr
*   Correspondence: sergey.samsonov@ug.edu.pl; Tel.: +48-58-523-5166
†   These authors contributed equally to this work.

check for updates

Abstract: In this study, we characterize the interactions between the extracellular matrix protein, procollagen C-proteinase enhancer-1 (PCPE-1), and glycosaminoglycans (GAGs), which are linear anionic periodic polysaccharides. We applied molecular modeling approaches to build a structural model of full-length PCPE-1, which is not experimentally available, to predict GAG binding poses for various GAG lengths, types and sulfation patterns, and to determine the effect of calcium ions on the binding. The computational data are analyzed and discussed in the context of the experimental results previously obtained using surface plasmon resonance binding assays. We also provide experimental data on PCPE-1/GAG interactions obtained using inhibition assays with GAG oligosaccharides ranging from disaccharides to octadecasaccharides. Our results predict the localization of GAG-binding sites at the amino acid residue level onto PCPE-1 and is the first attempt to describe the effects of ions on protein-GAG binding using modeling approaches. In addition, this study allows us to get deeper insights into the in silico methodology challenges and limitations when applied to GAG-protein interactions.

Keywords: procollagen C-proteinase enhancer-1; glycosaminoglycans; computational analysis of protein-glycosaminoglycan interactions; calcium ions; fragment-based docking

1. Introduction

Glycosaminoglycans (GAGs) are anionic periodic linear polysaccharides, which are composed of periodic disaccharide units [1] and play a key role in many biologically relevant processes by interacting with their numerous and diverse protein targets such as cytokines and growth factors in the extracellular matrix [2–5]. However, the molecular mechanisms underlying GAG-mediated interactions are not fully understood, and experimental techniques alone are not sufficient for gaining

insights into them [6]. Molecular modeling approaches are not only complementary to experiments, but also provide additional and crucial details, which are experimentally inaccessible. In our previous work, we successfully applied molecular docking and molecular dynamics methodologies in order to model protein-GAG interactions. In particular, we have modeled the effects of GAG binding on chemokines [7,8], growth factors [9,10] and other proteins [11,12], which allowed us to investigate the fundamental questions related to these interactions such as their specificity, the role of multipose character of GAG binding and polarity of binding poses of these periodic molecules.

In this work, we model interactions of GAGs with procollagen C-proteinase enhancer-1 (PCPE-1, encoded by gene *PCOLCE*), a glycoprotein which plays an important role in the assembly of the extracellular matrix [13,14]. Lacking proteolytic activity on its own, PCPE-1 enhances C-terminal procollagen processing, mediated by tolloid-like proteinases such as bone morphogenetic protein 1 (BMP-1) and mammalian tolloid (mTLD) designated BMP-1/tolloid-like proteinases (BTPs) [14–17]. PCPE-1 expression is upregulated in fibrosis [18,19]. PCPE-1 comprises two complement, sea urchin protein Uegf, BMP-1 (CUB) domains [20] and a netrin-like (NTR) domain [21]. Although neither an X-ray nor an NMR structure is available for full-length PCPE-1, X-ray structure of CUB1-CUB2 domains (PDB ID: 6FZV, 2.7 Å) in a complex with C-propeptide of procollagen [22] and NMR structure of the NTR domain (PDB ID: 1UAP) are available [23]. In the structure of the active CUB1-CUB2 fragment of PCPE-1 bound to the C-propeptide trimer of procollagen III (CPIII), two Ca^{2+} ions participate in the formation of the interface between the CUB1-CUB2 domains and the procollagen III molecule [22]. Often, CUB domains bind Ca^{2+}, and Ca^{2+} coordination involves acidic amino acid residues (i.e., Tyr-Glu-Asp-Asp motif) [24]. A conserved calcium binding site has indeed been identified in the CUB1 domain of PCPE-1, and mutational analysis of this site confirmed that PCPE-1 stimulating activity requires a calcium binding motif in the CUB1 domain, which is highly conserved among CUB-containing proteins [25]. A low-resolution structure of the full-length PCPE-1 protein was proposed based on small angle X-ray scattering (SAXS), analytical ultracentrifugation and transmission electron microscopy, showing that PCPE-1 is a rod-like molecule, with a length of 150 Å [26]. PCPE-1 binds to heparin (HP) as shown using affinity chromatography [27] and surface plasmon resonance (SPR) binding assays [28], and the binding is mediated by the NTR domain. Heparan sulfate (HS) and dermatan sulfate (DS) but not chondroitin sulfate (CS) inhibit PCPE-1-HP interactions. HP also binds to BMP-1 [29]. HS could thus potentially act as a scaffold to assemble BMP-1, PCPE-1 and procollagen together at the cell surface [28]. Therefore, the characterization of PCPE-1/GAG interactions at the atom level is important for the detailed understanding of PCPE-1 functions.

The aim of this work is to get deeper insights into PCPE-1/GAG interactions using both SPR inhibition assays and in silico techniques to complement the experimental data obtained in the previous [28] and present work. Modeling approaches were used to build structural models of full-length PCPE-1 and to determine GAG specific binding to PCPE-1 and its domains. We analyzed the binding of PCPE-1 to GAGs of different types, lengths and sulfation patterns, which were rationally and systematically chosen to match those used in experiments. We also investigate the potential role of Ca^{2+} in these interactions [28] and evaluate the challenges of in silico methodology to study protein-GAG interactions [30]. The results reported here contribute to the understanding of the biologically relevant PCPE-1/GAG interaction and, for the first time, systematically predict the structural positions and the effects of Ca^{2+} ions on protein-GAG complexes.

2. Results and Discussion

2.1. Experimental Results

We have previously shown that DS, HS and HP but not CS inhibited the binding of soluble PCPE-1 to immobilized HP [28]. Here, we investigated the effect of HP oligosaccharides of various length as inhibitors of PCPE-1 binding to HP in order to determine the optimal size of HP required to bind to PCPE-1. There was a trend towards an increase in inhibition of PCPE-1-HP interaction with the length

of HP oligosaccharides from dp2 to dp8, and then from dp14 to full-length HP chains (Figure 1). HP decasaccharides and dodecasaccharides (dp10 and dp12, respectively) inhibited the binding of PCPE-1 to HP to a lesser extent than the HP octasaccharide (dp8). The oligosaccharides used for inhibition experiments were separated according to their size and not to their sulfation pattern and/or charges. They thus contain a mixture of oligosaccharides of the same size displaying a different number of sulfate groups in different positions of their sequences resulting in different binding motifs with likely different inhibitory efficiencies. This heterogeneity might be more pronounced in dp10 and dp12, leading to a lower global inhibition by these oligosaccharides than by the octasaccharide.
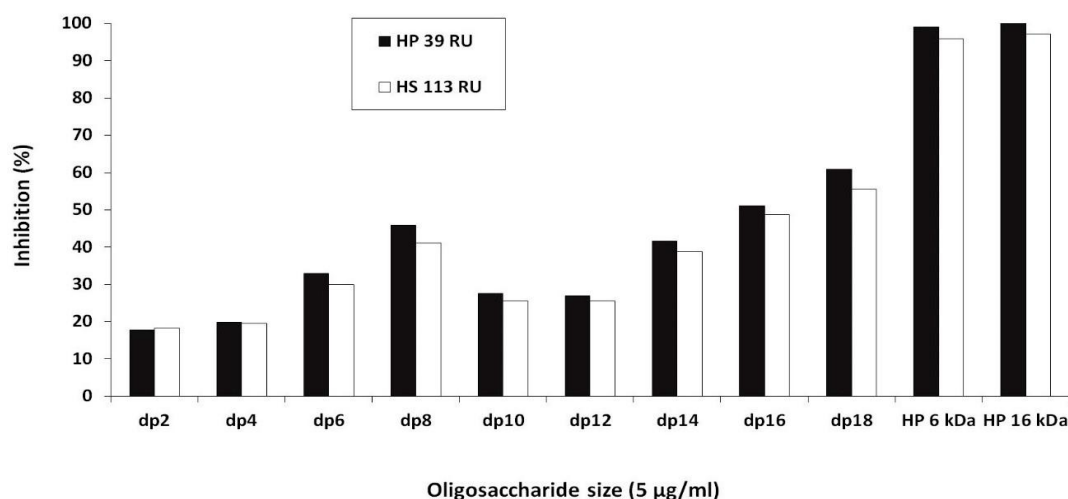

Figure 1. Surface plasmon resonance (SPR) inhibition assays. Inhibition of the binding of recombinant human procollagen C-proteinase enhancer-1 (PCPE-1) to biotinylated heparin (HP) and heparan sulfate (HS) captured on a streptavidin sensor chip (39 and 113 resonance units (RU) respectively) by HP oligosaccharides of different degrees of polymerization (dp2-dp18) and by HP (6 and 16 kDa) at a concentration of 5 µg/mL.

Then we applied the in silico approaches we have previously developed to analyze the binding of PCPE-1 to GAGs at the atomic level and to determine if these interactions were exclusively electrostatic-driven or if other factors modulate the binding strength.

2.2. Modeling the Full Structure of PCPE-1

We created two ensembles of full-length PCPE-1 structures using the UNRES (from UNited RESidue) coarse-grained (CG) approach to determine the structure of the linker located between the CUB1-CUB2 and the NTR domains. In the first one, the structures of the linkers were optimized, and the domain structures were restrained, while in the second one, SAXS derived restraints were used additionally in order to reproduce the experimental data [26] (see Section 3.4 for more details). Five most probable structural models were obtained for both ensembles. For HP binding analysis we used the first three models obtained without SAXS restraints and one model obtained with SAXS restraints (SAXS Model) (Table 1). The radii of gyration of the models obtained without SAXS restraints were significantly lower than those of the elongated structures restrained using SAXS data. As expected, the SAXS Model had a radius of gyration in agreement with the experimental value calculated using SAXS (41 ± 3 Å versus 43 ± 1 Å [26]. The obtained model was also consistent with the length of the protein determined experimentally (150 Å). Poisson-Boltzmann surface area (PBSA) calculations applied to these 4 models suggest that potential binding regions for HP were located in the NTR domain for Model 3, at the interface of the linker and the NTR domain for Model 2 and SAXS Model, and at the common interface of all domains (CUB1-CUB2, linker and NTR) in Model 1 (Figure 2, Supplementary Figure S1).

Table 1. Models of the full-length PCPE-1 obtained using UNRES coarse-grained (CG) simulations.

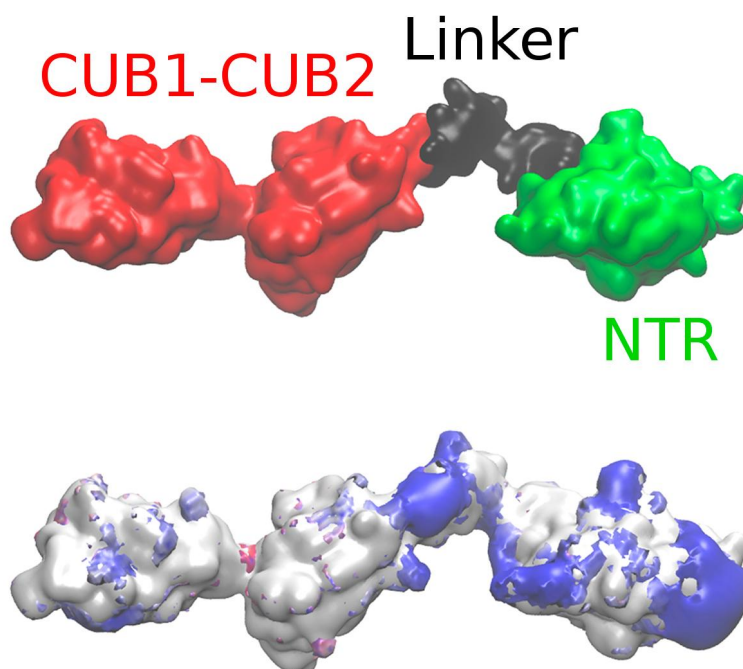| Model	Restraints	Probability	Radius of Gyration (Å)
1		34	22.2
2		32	24.8
3	CUB1-CUB2, NTR domains	18	22.6
4		8	22.8
5		8	22.6
1		39	43.5
2		21	44.5
3	CUB1-CUB2, NTR domains + SAXS-based	17	43.5
4		14	43.3
5		9	43.9

Figure 2. Small angle X-ray scattering (SAXS) Model (upper panel). Netrin-like (NTR) domain: green; CUB1-CUB2: red; the interdomain linker between the CUB2 and NTR domains: black. Positive electrostatic potential isosurfaces (2.0 kcal/mol · e^{-1}) in the absence of Ca^{2+} ions obtained by Poisson-Boltzmann surface area (PBSA) calculations (bottom panel).

2.3. PCPE-1 Interactions with Glycosaminoglycans

We modeled and analyzed the binding of PCPE-1 and its domains, NTR and CUB1-CUB2, with the following GAGs: chondroitin sulfate-6 (CS6) made of two GalNAc6S-GlcA or three disaccharide units (dp4 and dp6, respectively), dermatan sulfate comprised of three GalNAc6S-IdoA disaccharide units (dp6), and heparin (HP) made of one, two and three GlcNS6S-IdoA2S disaccharide units (dp2, dp4, and dp6 respectively). These GAGs were selected for the following reasons: to compare the in-silico data with the experimental ones previously obtained with these GAGs [28] and to investigate the effects of epimerization, length and sulfation pattern of GAGs on binding. Conventional docking approaches are severely limited in terms of the size of GAGs and can be effectively used only for the GAGs with a length up to dp6 [31]. Therefore, we used HP oligosaccharides of different lengths, from dp2 to dp6, to determine the effect of the GAG length on the binding to PCPE-1. Since HP is the

strongest binder, the results obtained with HP oligosaccharides of different lengths should be the most representative. Furthermore, the GAGs studied here were selected in order to systematically evaluate the changes in binding to PCPE-1 according to the GAG length (dp4-dp6 for CS6 and dp2-dp6 for HP), the epimerization of glucuronic acid (CS6 dp6 and DS dp6), the increase in the number and position of sulfated groups (i.e., the sulfation pattern) and the net charge of the oligosaccharides (CS6, DS and HP).

Several clusters of docking solutions were obtained for each GAG tested. The polarity of the binding poses was analyzed because the orientation of the GAG chain was shown to be non-random for the IL-8 chemokine [7] and determinant for the binding specificity of the C-X-C motif chemokine ligand 14 [8], suggesting an important functional role of GAG polarity in their interactions with proteins. Then, for the most diverse binding poses within these clusters, molecular dynamics (MD) simulations were performed with binding free energy post-processing calculations and per residue binding free energy decomposition. We would like to emphasize that choosing a proper procedure of pose selection for such analysis is very challenging, since it is unclear how many clusters and solutions within each cluster should be representative, which part of the trajectory should be analyzed in terms of the free energy, if only the best scored pose from a cluster or all the poses should be taken into account for the further calculations, and how to weight their contributions in the latter case. The answers to these questions are dependent on the molecular systems and on the particular goal of the modeling study. These methodology-oriented aspects of protein-GAG modeling will be further discussed below.

2.3.1. The NTR-Domain

Among the found clusters of docking solutions, for CS6 dp4 and HP dp2, dp4 and dp6, one major cluster was observed, while there was more uniform distribution of the solutions between several clusters for CS6 dp6 and DS dp6 (Table 2). This suggests that for those molecules, especially for CS6 where the clusters are especially diverse, multipose binding might be quite probable. GAG multipose binding was previously identified both experimentally and computationally for TIMP-3, which is homologous with the NTR domain [11]. Most solutions were localized near the C-terminal α-helix of the NTR domain except for CS6 dp6 (Figures 3 and 4). The size of the clusters obtained by molecular docking was not correlated with their corresponding free binding energies calculated from the MD simulation. This means that molecular docking alone was not able to properly score the solutions, although the Autodock 3 (AD3) scoring function is one of the most successful scoring schemes when applied to GAG complexes [10]. Similarities of the binding regions for the docking solutions post-processed by MD-based binding free energy decomposition per residue are reflected in Tables 3 and 4 for the obtained clusters and for each GAG ligand respectively. According to the binding free energy values obtained for the NTR domain bound to GAGs compared to the experimental complexes from the PDB [31] and given that no dissociation of these complexes was observed, we assume that the binding of the analyzed GAGs to NTR is stable. The binding strength, evaluated by the calculation of free binding energy, of CS6 dp4 and CS6 dp6 did not significantly differ, but the cluster location of CS6 dp6 differed from those of CS dp4, DS dp6 and HP dp2, dp4 and dp6. Only the third biggest cluster for CS6 dp6 was located in the region overlapping with those of other analyzed GAGs. CS6 dp6 and longer CS6 oligosaccharides might thus bind NTR differently from CS6 dp4 and other GAGs. Therefore, although the binding strength was similar for CS6 and DS, their preferred binding sites were distinct for these two GAGs, which differ only in the epimerization of glucuronic acid. This could potentially explain the results from surface plasmon resonance binding assays, which showed that CS6 did not inhibit PCPE-1 binding to HP whereas DS did [28]. Whereas DS competes with HP for the same binding site on PCPE-1, CS6 binds to a different region, which would allow HP oligosaccharides to remain bound to the NTR domain. Similar computational approaches were successfully applied to demonstrate the experimentally proven differences in binding strength between DS and CS6 interacting via the same binding pose to IL-8 [7,32] In contrast, the binding differences for those GAGs were related to certain differences in the binding pose for CXCL14 [8]. This suggests that for protein-GAG complexes, the predictive power of the computational methods is dramatically

dependent on the protein involved and the distribution of the clusters on its surface, which is, in turn, also sensitive to a particular clustering procedure. HP binds the NTR domain stronger than CS6 and DS, while its increase in length stabilizes the interaction suggesting a key role of electrostatic interactions, although few hydrophobic amino acid residues (leucine and valine) and polar, uncharged, amino acid residues (asparagine and glutamine) were predicted to interact with the analyzed GAGs (Figure 4). Most clusters revealed a bias towards specific polarity of GAG binding poses, although this trend was less pronounced for HP dp6 and DS dp6 (Table 2). This suggests that our docking approach is able to distinguish GAG polarity, which is an important methodological finding and will allow us to investigate one of the potential parameters underlying the specificity of protein-GAG interactions [8].

Table 2. Molecular docking molecular dynamics (MD)-based analysis summary for NTR-GAG interaction.

GAG	[1] m, ε	[2] #	[3] Size	[4] ΔG (kcal/mol)	[5] Polarity
CS6, dp4	3, 2	1	19	−42.0 ± 6.6; −48.3 ± 7.7; −41.3 ± 6.6	17/2
		2	6	−30.1 ± 16.0; −63.3 ± 7.1	6/0
		3	4	−34.4 ± 9.6; −38.4 ± 8.6	2/2
		4	3	−46.7 ± 10.5	3/0
CS6, dp6	3, 2	1	3	−56.6 ± 9.0	3/0
		2	3	−33.9 ± 9.2	3/0
		3	3	−36.8 ± 7.1; −64.2 ± 11.8	3/0
DS, dp6	3, 2	1	6	−35.5 ± 6.3; −41.5 ± 6.8	5/1
		2	4	−36.7 ± 6.6	4/0
		3	3	−63.7 ± 8.3	3/0
		4	3	−37.8 ± 8.2	2/1
HP, dp2	3, 2	1	25	−44.9 ± 9.3; −41.1 ± 7.3; −23.1 ± 7.6	25/0
		2	12	−27.9 ± 9.2	12/0
		3	9	−42.0 ± 9.0	9/0
		4	3	−27.7 ± 8.9; −28.7 ± 5.6	3/0
HP, dp4	3, 2	1	32	−39.0 ± 7.2; −29.4 ± 10.4	21/11
		2	3	−53.9 ± 7.2	3/0
		3	3	−50.6 ± 11.5; −57.4 ± 8.6	2/1
HP, dp6	3, 2	1	15	−69.5 ± 7.8; −56.7 ± 7.4; −43.5 ± 9.7; −54.0 ± 14.8; −80.5 ± 10.7	9/6
		2	7	−68.3 ± 11.0; −44.7 ± 7.5; −57.1 ± 10.6; −55.4 ± 9.1	4/3
		3	6	−50.1 ± 10.0; −44.6 ± 9.5; −65.7 ± 11.5; −61.8 ± 13.6	4/2

[1] DBSCAN parameters *m*, the minimal neighborhood size, and *ε*, neighborhood search radius [33]; [2] cluster number; [3] cluster size (number of solutions); [4] free energy of binding obtained by MM-GBSA; [5] the polarity of a GAG binding pose was defined as its preferred orientation in relation to the reducing and non-reducing end.
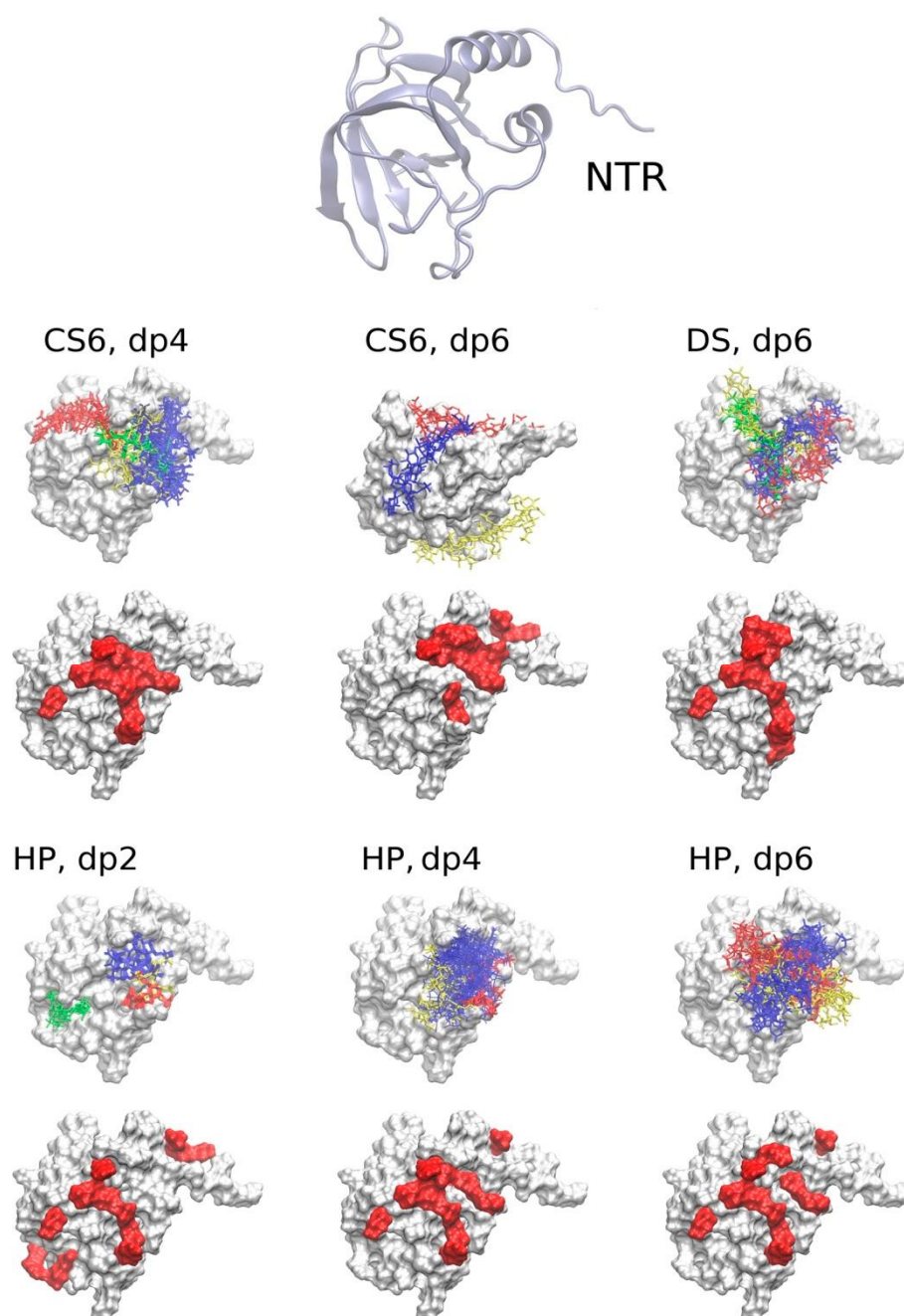
Figure 3. Molecular docking and molecular mechanics-generalized born surface area (MM-GBSA) for NTR-glycosaminoglycan (GAG) complexes. The structure of the NTR domain is shown in cartoon representation at the top. For each GAG, the analyzed clusters of docking solutions are shown in blue, red, yellow and green (from the most to the less populated cluster); the top 10 residues binding to GAGs according to MM-GBSA calculations averaged per GAG are highlighted in red surface. Note that the clusters for CS6 dp6 are shown for a different protein spatial orientation to allow for a better visualization. In addition, averaging the per-residue energy for very different clusters could be misleading as shown for CS6 dp6: the residues shown in red do not overlap with the surface patches where the most representative clusters of solutions are located.

```
  318   C P K Q C R R T G T L Q S N F C A S S L V V T A T V K S M V R E P G E G L A V T   357
CS6 dp4             * *     * *
CS6 dp6               *       * * *
DS  dp6         *     * *
HP  dp2             * *                                            *
HP  dp4             * *       * *
HP  dp6             * *       * *

  358   V S L I G A Y K T G G L D L P S P P T G A S L K F Y V P C K Q C P P M K K G V S   397
CS6 dp4                                                                           *
CS6 dp6
DS  dp6                                                                         *
HP  dp2                                                                  *       *
HP  dp4                                                                         *
HP  dp6                                                                         *

  398   Y L L M G Q V E E N R G P V L P P E S F V V L H R P N Q D Q I L T N L S K R K C   437
CS6 dp4                                       * *   *   *   *
CS6 dp6                                           *   * *     *             * *
DS  dp6                                     * *   * * * *
HP  dp2                                     * *   *                         * *
HP  dp4                                     * *   *     *                   *
HP  dp6                                     * *   *   *                     *
```

Figure 4. NTR amino acid residues identified in the top 10 for binding GAGs according to MM-GBSA calculations per cluster are labeled as an asterisk.

Table 3. Similarity of GAG binding poses for the NTR domain as of common amino acid residues identified in the top 10 for binding according to MM-GBSA calculations per cluster.

GAG	CS6, dp4				CS6, dp6			DS, dp6				HP, dp2				HP, dp4			HP, dp6		
CS6, dp4	10	7	7	6	6	4	7	5	6	6	4	7	5	5	5	6	6	6	7	7	7
	7	10	6	9	3	2	6	6	6	7	5	6	7	6	6	5	8	7	8	7	9
	7	6	10	5	4	3	5	6	5	7	4	7	4	6	5	4	5	7	7	7	6
	6	9	5	10	3	2	5	6	6	6	4	5	7	5	5	6	9	6	7	6	8
CS6, dp6	6	3	4	3	10	7	5	2	5	3	1	4	3	1	1	5	3	2	5	5	4
	4	2	3	2	7	10	5	2	4	2	1	5	2	1	1	4	2	1	4	3	3
	7	6	5	5	5	5	10	5	5	4	5	6	5	4	4	7	5	4	7	6	7
DS, dp6	5	6	6	6	2	2	5	10	5	6	6	6	6	7	6	5	6	8	7	7	7
	6	6	5	6	5	4	5	5	10	5	5	4	6	4	4	6	7	5	7	6	6
	6	7	7	6	3	2	4	6	5	10	6	6	6	6	6	3	5	7	7	8	6
	4	5	4	4	1	1	5	6	5	6	10	4	4	5	5	3	4	6	5	5	5
HP, dp2	7	6	7	5	4	5	6	6	4	6	4	10	5	5	5	5	5	6	7	7	7
	5	7	4	7	3	2	5	6	6	6	4	5	10	5	5	7	7	5	7	7	7
	5	6	6	5	1	1	4	7	4	6	5	5	5	10	6	3	5	7	6	6	6
	5	6	5	5	1	1	4	6	4	6	5	5	5	6	10	3	5	6	6	6	6
HP, dp4	6	5	4	6	5	4	7	5	6	3	3	5	7	3	3	10	7	3	6	5	6
	6	8	5	9	3	2	5	6	7	5	4	5	7	5	5	7	10	6	7	6	8
	6	7	7	6	2	1	4	8	5	7	6	6	5	7	6	3	6	10	6	7	7
HP, dp6	7	8	7	7	5	4	7	7	7	7	5	7	7	6	6	6	7	6	10	9	9
	7	7	7	6	5	3	6	7	6	8	5	7	7	6	6	5	6	7	9	10	8
	7	9	6	8	4	3	7	7	6	6	5	7	7	6	6	6	8	7	9	8	10

Each line/column in front/below each GAG reflects a separate cluster, for which average values were taken into account.

Table 4. Similarity of GAG binding poses for the NTR domain as of the number of common amino acid residues identified in the top 10 for binding according to MM-GBSA calculations per GAG.

GAG	CS6, dp4	CS6, dp6	DS, dp6	HP, dp2	HP, dp4	HP, dp6
CS6, dp4	10	5	7	6	9	8
CS6, dp6	5	10	4	4	6	6
DS, dp6	7	4	10	6	7	7
HP, dp2	6	4	6	10	7	7
HP, dp4	9	6	7	7	10	9
HP, dp6	8	6	7	7	9	10

2.3.2. CUB1-CUB2 Domains

Although there is no experimental evidence suggesting that CUB1-CUB2 domains of PCPE-1 directly interact with GAGs, the differences in binding of NTR and full-length PCPE-1 to HP and HS [28] indicate that CUB1-CUB2 domains could affect GAG binding to the full-length PCPE-1 protein. Therefore, we analyzed the potential binding of these domains to GAGs using the same procedure as above.

All the predicted binding poses were either located in the cleft region between the CUB domains or bridged both CUB domains (Supplementary Figures S2 and S3). In both cases, such potential binding would lead to restricted movements of the CUB domains relative to each other, which, in turn, would affect the overall flexibility of PCPE-1 and its ability to recognize and to bind its partners. Calculated GAG free binding energies were essentially less favorable than those calculated for the NTR domain (Supplementary Table S1), which is consistent with NTR being responsible for GAG binding in PCPE-1. No binding poses of the analyzed GAGs or the structures that can be obtained from them by GAG chain elongation were found to be in close proximity to the Ca^{2+} binding sites or at the interface with procollagen peptides [23]. In a number of cases, the binding poses predicted by molecular docking were unstable (ΔG higher than −15 kcal/mol), and the GAG dissociated from the protein. Such behavior was typically observed for HP oligosaccharides and is explained by the repulsion of these highly charged molecules by the negatively charged residues of the CUB1-CUB2 domain. Poses corresponding to the binding of CS6 and DS, which are less negatively charged than HP, were globally more stable. However, some binding poses were very stable and comparable with those found in the NTR domain (e.g., cluster 2, solution 2 for CS6 dp4). In such cases, bound GAGs protruded deeply into the cleft between the CUB1 and CUB2 domains forming strong van der Waals interactions in addition to the electrostatic interactions, which are believed to be the driving force in the formation of protein-GAG complexes [31,34]. As reported for the NTR domain, highly significant differences in free energy were found for GAGs within and beyond the same clusters. One major cluster was found for CS6 of various length in contrast to what was observed for other GAG analyzed. No correlation was found between the size of clusters and their free binding energies. The comparison between the observed clusters and the data averaged for different GAGs in terms of the most important protein binding residues showed high similarities for all GAGs, suggesting weak and rather unspecific binding to CUB1-CUB2 domains (Supplementary Figure S2, Supplementary Tables S2 and S3). Interestingly, for CS6 dp4 the differences between the clusters were more prominent than the differences of these clusters with those obtained for other GAGs. The increase in length of HP from dp2 to dp6 did not modify the potential interaction pattern with the CUB1-CUB2 domains. All clusters revealed strong polarity preferences except for the DS clusters.

2.3.3. Full PCPE-1

GAG binding was characterized with full-length PCPE-1 models obtained using UNRES CG simulations and HP dp6 as a ligand. Binding to Model 1, which was the most probable model among

the ones obtained without the SAXS-based restraints, was significantly stronger than to Models 2, 3 and the SAXS Model (Table 5, Supplementary Table S4), as well as to the NTR domain (t-test, *p*-value < 0.05). Binding to Model 3 was also significantly stronger than to Model 2 and to the NTR domain. All clusters of HP dp6 solutions obtained for Model 1 were located in the region formed by the same residues of the NTR domain, the linker and the CUB1-CUB2 domains. For Model 2, the first cluster was located differently from the second and the third clusters. All the clusters correspond to the residues belonging predominantly to the NTR domain and the linker, but also partially to the CUB1-CUB2 domains. For Model 3, only NTR residues contributed to the binding of HP dp6. Cluster 1 was the most representative for Model 3 according to the molecular docking results, although not the most favorable according to MM-GBSA calculations, which again points to the essential differences in molecular docking and MD-based scoring. In the SAXS Model, GAG binding occurred at the NTR/linker interface. In all cases, the clusters were located in PCPE-1 patches corresponding to the positive electrostatic potential shown in Figure 2 and Figure S1.

Table 5. Molecular docking MD-based analysis summary for PCPE-1 SAXS Model/HP dp6 interaction.

[1] m, ε	[2] #	[3] Size	[4] ΔG, kcal/mol	[5] Top$_{MM\text{-}GBSA}$ 10 Residues for GAG Binding	[6] Polarity
2, 2.64	1	4	−62.4.8 ± 19.0; −54.9 ± 9.1 −49.6 ± 18.6	R435, K436, R275, R288, K279, K299, K365, K434, N331, K295	4/0
	2	3	−50.1 ± 9.7; −79.0 ± 17.0; −38.1 ± 9.4	K436, R435, K365, K299, K434, K271, K295, R288, K165, K279	3/0
	3	3	−30.8 ± 10.7; −36.0 ± 7.8; −42.3 ± 10.6	K299, K436, K279, K365, K271, K434, K295, K165, Q282, R435	2/1

[1] DBSCAN parameters *m*, the minimal neighborhood size, and ε, neighborhood search radius [33]; [2] cluster number; [3] cluster size; [4] free energy of binding obtained by MM-GBSA; [5] residues identified in the top 10 for binding according to MM-GBSA calculations per cluster ordered by the impact (starting from the most favorable one). [6] The polarity of a GAG binding pose was defined as its preferred orientation in relation to the reducing and non-reducing end.

2.4. The Potential Role of Ca^{2+} in PCPE-1 Interactions with Glycosaminoglycans

2.4.1. Prediction of Ca^{2+} Binding Sites

According to the experimental data, the interactions between both full-length PCPE-1 and the NTR domain with HP and HS are cation-dependent [28]. Therefore, we attempted to analyze the impact of Ca^{2+} ions on HP binding in silico, which allowed us to evaluate the available computational tools in terms of sensitivity and prediction power to account for divalent ions in such calculations. As a first step, we applied three different approaches (see Section 3.7 for details) to annexin V protein, which has 9 experimentally identified occupied Ca^{2+} binding sites, some of which are occupied upon HP binding [35]. The IonCom server predicted correctly eight out of nine experimentally known binding sites, while FoldX and MD approaches correctly predicted six binding sites (Table 6).

Table 6. Ca^{2+} predictions for annexin V and PCPE-1 domains: number of the binding sites predicted are provided.

Protein	PDB ID	Experimental Structure	Method		
			FoldX	IonCom	[1] MD
Annexin V	1G5N	9	6	8	6
NTR	1UAP	0	0	0	2
					3
					1
					1
					1
CUB1-CUB2	6FZV	2	2	2	2
					2
					1
					2
					1

[1] Five repetitions of the MD simulations were performed for PCPE-1 domains.

Furthermore, we performed MM-GBSA calculations to estimate if the strength of the Ca^{2+} binding in these experimentally known binding sites correlated with the predictions (Table 7). As shown in the table, the total energies of interactions were positive despite the fact that all the ions were stable during the entire MD simulation performed in explicit solvent. This reflects the fact that the implicit continuous solvent model in MM-GBSA fails to properly account for the strength of binding for these divalent ions in terms of the full binding free energy. At the same time, in vacuo electrostatic energy was highly negative and could be meaningful for comparing binding sites since the studied interactions were electrostatically driven. A t-test performed for the in vacuo electrostatic energy values did not point out any statistical differences between the sites, which were properly predicted and the ones which the MD-based approach failed to predict.

Table 7. MM-GBSA free energy calculations (per Ca^{2+} ion) for the experimentally known Ca^{2+} binding sites in annexin V.

Ca^{2+} Number (X-Ray)	[1] ΔG, kcal/mol	[2] ΔG_{ele}, kcal/mol	[3] FoldX	[3] IonCom	[3] MD
319	57.2 ± 4.7	-310.4 ± 10.3	+	+	+
320	47.5 ± 4.8	-264.8 ± 15.7	+	+	−
321	36.5 ± 3.5	-296.0 ± 10.9	−	+	+
322	59.7 ± 4.9	-380.5 ± 9.5	+	+	−
323	36.4 ± 3.5	-332.4 ± 7.9	−	−	+
324	62.4 ± 4.4	-376.6 ± 8.1	+	+	+
325	47.5 ± 6.1	-413.2 ± 13.0	+	+	+
326	39.3 ± 3.7	-312.2 ± 9.2	−	+	+
327	59.2 ± 4.7	-302.3 ± 8.6	+	+	−

[1] and [2]: ΔG and ΔG_{ele} stand for the total and in vacuo electrostatic MM-GBSA free energies, respectively. [3] Plus and minus reflect whether the method was capable of predicting the corresponding experimentally detected binding site correctly.

We applied three ion-binding site prediction methods to the NTR and the CUB1-CUB2 domains of PCPE-1 (Table 6). Neither FoldX nor IonCom found any Ca^{2+} binding site for the NTR domain, while

the MD approach identified from one to three binding sites, one of which being consistent through all five repetitions of MD simulations. The fact that these methods did not agree with MD simulations could be due to conformational changes of negatively charged amino acid side chains during the MD simulation, allowing them to come close to each other and to coordinate calcium ions. FoldX and IonCom used static structures, which prevents the dynamics required for the coordination of Ca^{2+}. For CUB1-CUB2 domains, all methods were consistent and predicted two Ca^{2+} binding sites identical to those found in the CUB1-CUB2 domain complexed with procollagen (PDB ID: 6FZV). This means that CUB1-CUB2 domains in PCPE-1 could be already prebound to Ca^{2+} ions when the interaction with the procollagen is established. Furthermore, we compared the predicted Ca^{2+} binding sites for PCPE-1 domains in terms of electrostatic energies obtained from MM-GBSA calculations with the corresponding energies for annexin V in order to estimate their strength (Table 8). CUB1-CUB2 binding sites were energetically comparable with those of annexin V. The Ca^{2+} binding site in CUB2 was stronger than in CUB1 as suggested by more favorable electrostatic energies and by the fact that the Ca^{2+} binding site in CUB2 was identified by MD in all five MD replicas, while the Ca^{2+} binding site of CUB1 was correctly identified in three MD simulations. The Ca^{2+} binding sites predicted for the NTR domain were significantly weaker, and only one of them was found in all MD replicas. This suggests that this site may be unoccupied when the NTR domain is in solution and not bound to a GAG. The experimental evidence that the NTR domain binding to HP is dependent on divalent cations [28] leads to the hypothesis that Ca^{2+} ions could potentially bind within the interface of the NTR-HP complex.

Table 8. MM-GBSA free energy calculations (per Ca^{2+} ion) for the predicted Ca^{2+} binding sites in PCPE-1 domains and corresponding Ca^{2+} binding site occupancy in 100 ns MD simulation.

PCPE-1 Domain	Ca^{2+} Site	[2] ΔG_{ele}, kcal/mol	Site Occupancy, ns
NTR, MD1	E405, E406, N407	-116.5 ± 20.4	65
	G367, D370	-58.3 ± 15.7	40
NTR, MD2	E405, E406, N407	-125.8 ± 14.7	85
	D314/N-terminus of NTR	-38.6 ± 19.0	35
	G367, D370	-49.6 ± 14.2	90
NTR, MD3	E405, E406, N407	-120.3 ± 15.9	75
NTR, MD4	E405, E406, N407	-123.4 ± 12.2	45
NTR, MD5	E405, E406, N407 [1] E405, E406, N407/G367, D370	-51.7 ± 13.7; -156.2 ± 33.6	65 25
CUB1-CUB2 (X-ray, PDB ID: 6FZV)	E85, Y92, D93, D134	-363.2 ± 10.8	100
	Y180, E208, D216, D258	-466.5 ± 12.5	100
CUB1-CUB2, MD1	E85, Y92, D93, D134	-389.9 ± 19.3	25
	Y180, E208, D216, D258	-302.9 ± 24.3	90
CUB1-CUB2, MD2	E85, Y92, D93, D134	-368.3 ± 18.7	85
	Y180, E208, D216, D258	-371.9 ± 11.8	90
CUB1-CUB2, MD3	Y180, E208, D216, D258	-389.2 ± 18.6	85
CUB1-CUB2, MD4	E85, Y92, D93, D134	-293.7 ± 9.0	75
	Y180, E208, D216, D258	-374.0 ± 16.1	95
CUB1-CUB2, MD5	Y180, E208, D216, D258	-521.4 ± 10.8	95

[1] In the course of this simulation, G367 and D370 moved towards E405, E406 and N407 to coordinate Ca^{2+}. (MD: molecular dynamics, 1–5: replicas). [2] ΔG_{ele} stands for the in vacuo electrostatic MM-GBSA free energy.

We calculated electrostatic potential isosurfaces for the NTR and CUB1-CUB2 domains in the presence and in the absence of Ca^{2+} ions using the PBSA approach to predict how the electrostatic properties of the protein were affected by Ca^{2+} ions binding, which, in turn, could have an impact on

GAG binding. For this, we used two Ca^{2+} binding sites corresponding to the X-ray structure (PDB ID: 6FZV) of the CUB1-CUB2 domain, and the weak Ca^{2+} binding site predicted in the NTR domain (Figure 5). Major differences in both positive and negative electrostatic potential shape were observed for the NTR domain. This is not only explained by the direct effect of Ca^{2+} positive charge but also by the fact that E405, E406 and N407 were moved closer to each other to coordinate the cation, which also affects the topology of the isosurface. For CUB1-CUB2 domains, the largest positively charged patch of the potential isosurface was not noticeably affected by the presence Ca^{2+} ions. To sum up, the predicted 3 Ca^{2+} binding sites for both PCPE-1 domains, when occupied, could potentially affect GAG binding. This potential effect is analyzed below.
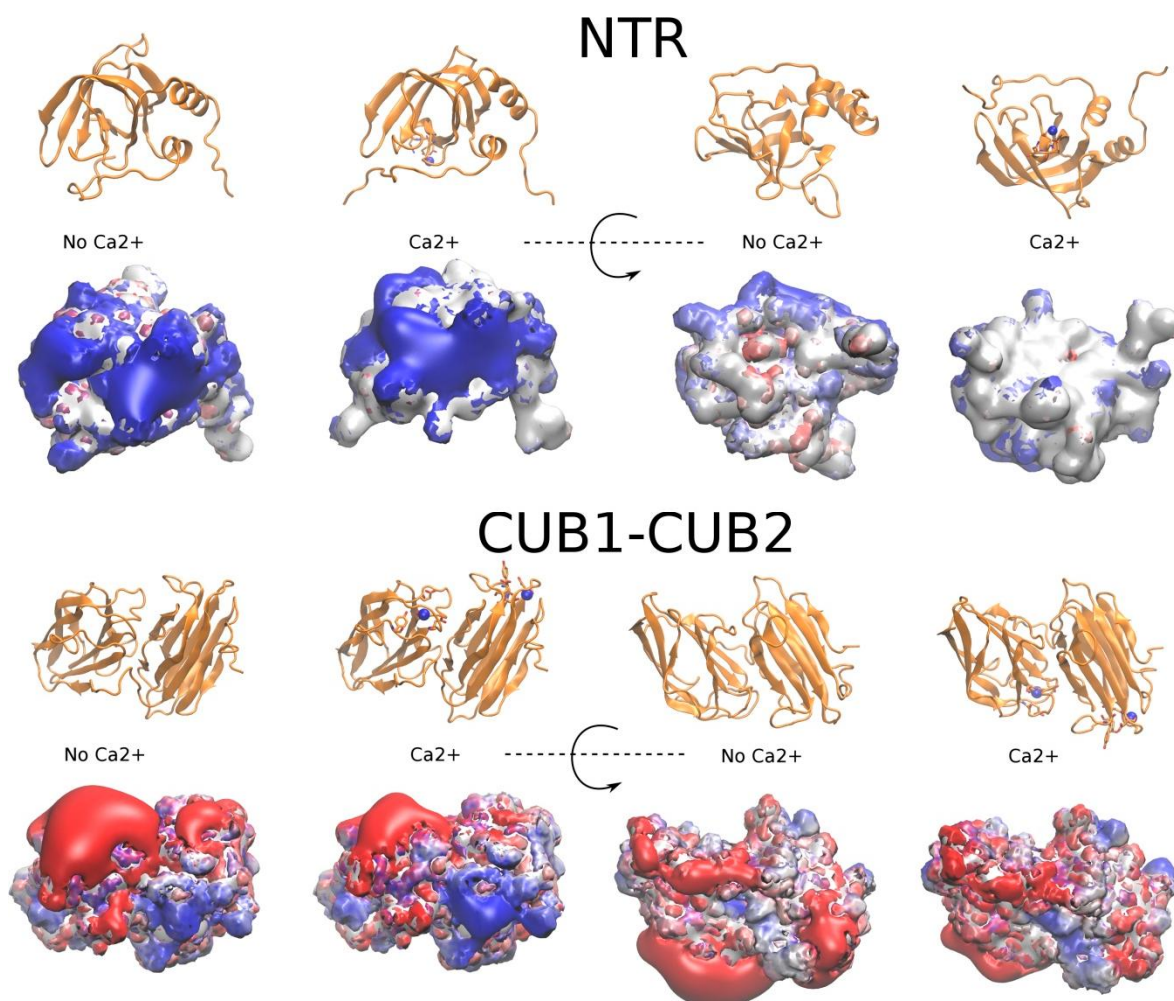
Figure 5. Electrostatic potential isosurfaces (blue, positive; red, negative) of NTR (-2.5 kcal/mol·e^{-1} and 1.0 kcal/mol·e^{-1}) and CUB1-CUB2 (-3 kcal/mol·e^{-1} and 3 kcal/mol·e^{-1}) domains in the presence and in the absence of Ca^{2+} ions obtained by PBSA calculations. Protein domains are shown in cartoon with the residues coordinating Ca^{2+} ions in licorice representation; Ca^{2+} ions: blue spheres.

2.4.2. PCPE-1 Interactions with Glycosaminoglycans in the Presence of Ca^{2+} Ions

In order to analyze the potential impact of Ca^{2+} ions on PCPE-1-GAG interactions, HP dp2, dp4 and dp6 were docked onto NTR, CUB1-CUB2 and the full-length protein in the presence of Ca^{2+} ions, followed by MD-based analysis. Two Ca^{2+} ions were prebound to CUB1-CUB2 and one to the NTR domain. Despite the differences in electrostatic properties of these domains described above, no significant changes in docking results or binding free energies were observed when compared with the domains without prebound Ca^{2+} (Supplementary Figures S4 and S5 and Supplementary Tables S5

and S6). In both cases, the binding occurred in the regions distant from the Ca^{2+} ion binding sites. Only one cluster (HP dp6, cluster 3 in the NTR domain) was found to be close to the Ca^{2+} ion. In the corresponding binding poses the ion was coordinated by a sulfate group from the terminal GlcNS(6S) residue. However, the binding poses from this cluster were significantly less favorable than those located distantly from the Ca^{2+} ion.

We also analyzed the impact of Ca^{2+} ions on the GAG binding to the full-length PCPE-1 (Table 9, Supplementary Table S7, Figure 6, Supplementary Figure S6). For Models 1 and 2 and the SAXS Model, we did not observe any effect of Ca^{2+} ions on the location of the structural clusters nor the direct participation of the Ca^{2+} ions in binding HP dp6. For Model 3, there was a significant difference between the clusters observed in the absence and in the presence of Ca^{2+} ions (e.g., cluster 2) related to the essential changes of the electrostatic potential on the protein surface in the presence of Ca^{2+}. In terms of the free energies of binding, the presence of Ca^{2+} ions did not significantly affect binding to Models 1 or 2 or the SAXS Model. For cluster 2 of Model 3, which was relocated in the presence of Ca^{2+} ions, the free energy of binding became less favorable than in the absence of Ca^{2+} ions.

Table 9. Molecular docking MD-based analysis summary for PCPE-1 SAXS Model/Ca^{2+}/HP dp6 interaction.

[1] m, ε	[2] #	[3] Size	[4] ΔG, kcal/mol	[5] Top$_{MM-GBSA}$ 10 Residues for GAG Binding	[6] Polarity
2, 2.8	1	6	−58.8 ± 12.2; −56.0 ± 19.7; −58.6 ± 13.7	R435, K436, K434, K365, K299, K279, K295, R288, P438, K271	5/1
	2	4	−79.5 ± 15.6; −32.6 ± 11.0; −42.8 ± 10.0	K436, R435, K279, R288, K365, K299, K434, Q282, G281, K287	3/1
	3	3	−30.8 ± 10.7; −70.7 ± 13.2; −56.6 ± 18.4	R435, K299, K436, K365, K434, R275, K279, K295, K271, K305	3/0

[1] DBSCAN parameters *m*, the minimal neighborhood size. and *ε*, neighborhood search radius [33]; [2] cluster number; [3] cluster size; [4] free energy of binding obtained by MM-GBSA; [5] residues identified in the top 10 for binding according to MM-GBSA calculations per cluster ordered by the impact (starting from the most favorable one). [6] The polarity of a GAG binding pose was defined as its preferred orientation in relation to the reducing and non-reducing end.
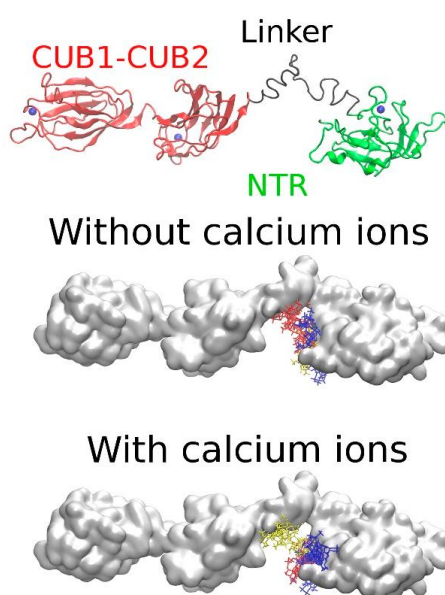

Figure 6. Molecular docking results for the models of the full-length PCPE-1 SAXS Model in the absence and presence of Ca^{2+} ions and HP dp6. The clusters of docking solutions are shown in blue, red and yellow (from the most to the least populated clusters). NTR domain: green; CUB1-CUB2: red; the interdomain linker between the CUB2 and NTR domains: black.

To summarize our attempts to determine the impact of Ca^{2+} on GAG binding to PCPE-1, our approach to consider Ca^{2+} ions as a part of the protein did not detect any favorable effect of these divalent ions on GAG binding in contrast to the experimental data [28]. Therefore, we hypothesize that Ca^{2+} ions might bind to GAGs rather than to PCPE-1 prior to the complex formation. The binding of divalent ions to GAGs has been experimentally reported for Zn^{2+}, Mn^{2+}, Cu^{2+}, Ca^{2+}, Co^{2+}, Na^{+}, Mg^{2+}, Fe^{3+}, Ni^{2+}, Al^{3+} and Sr^{2+} ions [36,37]. The crucial role of cations in protein-GAG interactions was experimentally shown for amyloid precursor protein [38], HP cofactor II [39], endostatin [40,41], FGF1 and IL-7 [42]. Experimental data obtained using mass spectrometry [43], NMR [44], gel-filtration chromatography [45,46] and infrared spectroscopy [47] indicate that GAGs interact with divalent ions, and that these interactions affect GAG structure and conformational properties. Divalent ions will be integrated in GAG structure in our future work on protein-GAG complexes. Another potential role of Ca^{2+} ions could be to stabilize PCPE-1 structure, which would affect its interactions with GAGs.

2.4.3. Predicting Longer GAG Binding Poses Using the Fragment-Based Approach

We calculated the binding poses of long (dp11) HP chains on full-length PCPE-1 in the absence and the presence of three Ca^{2+} ions, two bound to CUB1-CUB2 and one to the NTR domain (Table 10, Table S8). HP dp11 was the longest GAG that we managed to assemble in these docking experiments, and it was used to model the scenario when the GAGs longer than dp6 are bound to the protein. The increase in length of the HP chain from dp6 to dp11 significantly stabilized (*p*-value < 0.05) the interactions with PCPE-1 Models 1 and 2. The addition of Ca^{2+} strengthened the interactions of HP dp11 with PCPE-1 for Models 1, 2 and the SAXS Model but not for Model 3. Similarly to what was observed for HP dp6, Model 1 was the strongest in terms of HP dp11 binding. Ca^{2+} ions did not affect the binding sites of HP dp6 on the surface of Models 1, 2 or the SAXS Model but changed binding to Model 3, as described for HP dp6 (Figure 7, Supplementary Figure S7). In all the cases, docked HP dp6 structures overlapped very well with those of HP dp11. It could be concluded that docking HP dp6 defines the core binding unit of a HP chain. The increase in binding affinity with the increase of HP length agrees with the experimental trend observed in this study (Figure 1).

Table 10. Fragment-based molecular docking MD analysis summary for PCPE-1 SAXS Model/Ca^{2+}/HP dp11 interaction.

[1] #	[2] Ca^{2+}	[3] ΔG, kcal/mol	[4] Top$_{MM-GBSA}$ 10 Residues for GAG Binding
1	–	-65.6 ± 12.3	K299, R288, R435, K436, K295, K293, K305, K287, K365, P298
2	–	-64.2 ± 11.2	K436, K434, R275, K279, K295, R435, K365, R288, K299, K287
3	–	-94.7 ± 12.4	K295, K436, R435, K365, K434, K293, K299, R288, K305, V294
1	+	-93.3 ± 12.3	R435, K279, K295, K436, K305, K299, K434, N331, K271, R324
2	+	-73.8 ± 11.9	R435, K434, K436, K299, K295, K279, P298, K293, K287, K305
3	+	-102.5 ± 14.7	K436, K299, K295, R435, R275, K293, K279, K434, K305, P441

[1] Pose number; [2] Ca^{2+} presence; [3] free energy of binding obtained by MM-GBSA; [4] residues identified in the top 10 for binding according to MM-GBSA calculations per cluster ordered by the impact (starting from the most favorable one).
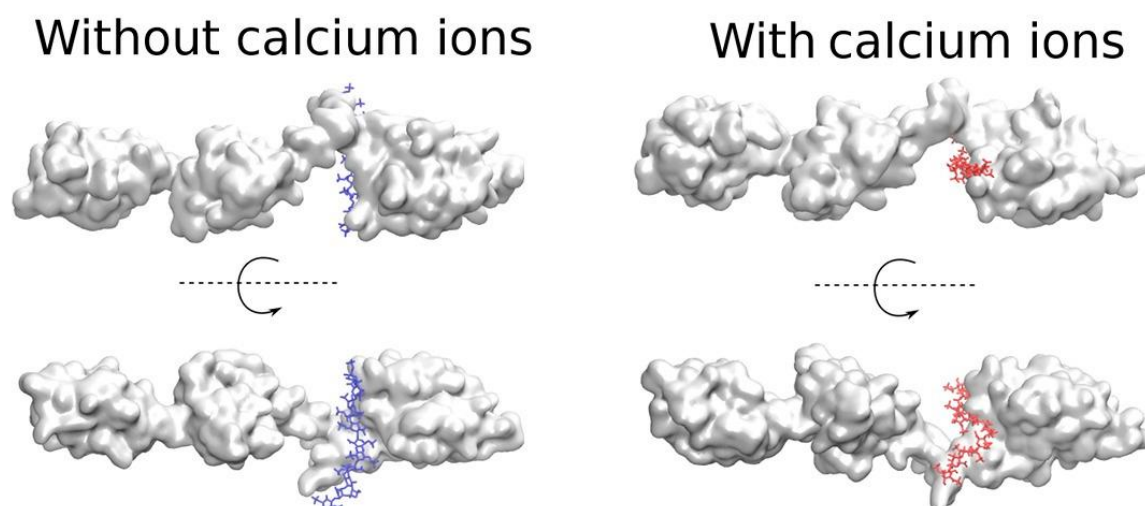
Figure 7. Molecular docking results for the models of the full PCPE-1 SAXS Model in the absence (in blue) and the presence (in red) of Ca^{2+} ions and HP dp11 corresponding to the most favorable free binding energies.

3. Materials and Methods

3.1. Surface Plasmon Resonance (SPR) Binding Assays

The SPR measurements were performed on a BIAcore 3000 instruments (GE Healthcare, Uppsala, Sweden), and the data were analtzed with the BIAevaluation 3.1 Software (GE Healthcare, Uppsala, Sweden) as previously described [28]. Inhibition assays of PCPE-1 binding to HP and HS by HP oligosaccharides (from dp2 up to dp8, generous gift of Rabia Sadir and Hugues Lortat-Jacob, Institut de Biologie Structurale, Grenoble, France) were carried out as previously described [28]. Briefly, HP (Sigma, St Quentin Fallavier, France) and HS (Celsus Laboratories Inc, Cincinnati, OH, USA) from porcine intestinal mucosa were biotinylated and captured on streptavidin previously immobilized on a CM4 sensor chip (GE Healthcare, Uppsala, Sweden). Human recombinant PCPE-1 (1 μM) [28] was incubated with HP oligosaccharides (5 μg/mL) in 10 mM Hepes pH 7.5 + NaCl 0.15 M + P20 0.005% (HBS) + 5 mM $CaCl_2$ for one hour before injection over immobilized HP (39 RUs) and HS (113 RUs) at a flow rate of 30 μL/min for 4 min. The percentages of inhibition were calculated relative to PCPE-1 binding level incubated in the same conditions with HBS + 5 mM $CaCl_2$. The running buffer was HBS and the temperature was set at 25 °C.

3.2. Structures

3.2.1. Protein Structures

The structure of the N-terminal CUB1-CUB2 domains was obtained from the X-ray structure of CUB1-CUB2 fragment of PCPE-1 bound to the C-propeptide trimer of procollagen III (PDB ID: 6FVZ, 2.7 Å). In this structure, the residues 33–275 are resolved (here and further, the numeration of the sequence corresponds to the UniProtKB ID: Q15113). However, the structure of the linker (151–157) between two CUB domains was not determined due to its flexibility [22]. The structure of the NTR domain (313–442) was also obtained from the PDB (PDB ID: 1UAP, 1st NMR model) [23]. The structures of the linker between two CUB domains (151–157) as well as the interdomain linker (276–312) were built in xLeap module of AMBER16, refined using an MD approach and then used for modeling the full-length structure of PCPE-1 with a CG MD approach as described in the Section 3.4.

3.2.2. Glycosaminoglycan Structures

The following GAG structures were used for molecular docking: CS6 dp4, CS6 dp6, DS dp6, HP dp2, dp4, dp3, dp6 with (IdoA(2S) ring in $^1C_4$ conformation, as this conformation is clearly predominant in HP [9,48]). These structures were built in our previous work [49].

3.3. Electrostatic Potential Calculations

The PBSA approach as implemented in AmberTools within AMBER16 package [50] was used to calculate electrostatic potential isosurfaces corresponding to the analyzed proteins. This method proved to be successful for GAG binding site prediction on an extensive protein-GAG dataset from the PDB [31]. The default value for the grid spacing of 1.0 Å and ff99SB force field parameters were used. The electrostatic potential isosurfaces were analyzed in VMD [51]. For visualization, such values of positive and negative electrostatic potential were chosen for each molecular system so that the data were as informative as possible for GAG binding site propensities.

3.4. Coarse-Grained MD Simulations

In order to calculate the conformations of the PCPE-1 linkers (151–157 and 276–312 sequences) to obtain a model of the full-length protein, we applied a CG multiplexed replica exchange molecular dynamics (MREMD) [52,53] approach as implemented in UNRES (United Residues) [54,55], as previously described. The protocol was similar to that used in our previous work [12]. Distance restraints were imposed on the domains during the MREMD simulations. Additionally, in one of the simulations, SAXS-derived restraints [22] were imposed [56]. Each MREMD simulation consisted of 20 trajectories run at temperatures from 265 K to 370 K. Each trajectory consisted of 3.5×10^7 MD steps with 4.89 fs length for simulations with the restraints on the domains only and 7.1×10^4 steps for simulations with additional SAXS-derived restraints. The lower number of steps for simulations with information from the SAXS experiment was used because the radius of gyration of maintained structures was obtained already after that time. Only conformations from the last quarter of the simulation were taken into further analysis with the use of the weighted histogram analysis method (WHAM) [56]. The next step was minimum variance cluster analysis [57] of the conformational ensemble at T = 300 K, which enabled us to obtain five clusters, ranked according to summary probabilities of the ensembles and containing the most probable structures to the cluster with the least probable structures. For each cluster, one representative structure, closest to the cluster centroid, was selected as the representative conformation. The last step was the conversion of CG structures into all-atom ones using the PULCHRA [58] and SCWRL [59] algorithms.

3.5. Molecular Docking

3.5.1. Autodock 3

The docking simulations of GAG ligands to the PCPE-1 were performed with Autodock 3 (AD3) [60], which was previously shown to yield the best performance among docking programs for GAG ligands [10,31]. The blind docking procedure was used: the whole protein surface was available for the ligand when sampling a potential binding site. For all proteins, we used $127 \times 127 \times 127$ grid points for AD3 runs. However, because of the differences in protein size, the grid step was different for different proteins to contain the whole protein molecule within a single grid box: the default value of 0.375 Å was used for the NTR domain, while 0.5 Å was used for CUB1-CUB2 domains, and the grid step of 0.6–1.0 Å was used for different models of the full-length PCPE-1 protein due to their essentially bigger sizes. All GAG ligands were docked to the protein with the use of the Lamarckian genetic algorithm. The initial population size was 300, 10^5 generations, $9995 \cdot 10^5$ energy evaluations and 1000 independent runs with up to 33 torsional angle degrees of freedom were carried out. 1000 docked structures for each molecular system (protein-GAG pair) were obtained and further analyzed; the 50 top-scored ones (according to AD3 scoring function)

were chosen for further clustering with the DBSCAN algorithm [33]. The clustering parameters, neighborhood search radii and minimal numbers of cluster members were manually selected for each system individually in order to yield 2-4 representative clusters. The distance metric used for clustering, which is defined as the root-mean-square of atomic distances for the nearest atoms of the same type, takes into account the periodic nature of GAGs, which is more appropriate for those ligands than the classical root-mean-square deviation (RMSD) [61]. Within each cluster, those poses which were different from one another to be categorized into subgroups were selected for further analysis. Such a procedure was used to account for the multiple pose binding previously observed for GAG ligands [11,62]. The GAG glycosidic linkages from the obtained docking poses were visually filtered in order to avoid incorrect geometries that could be produced by AD3 [63].

3.5.2. Fragment-Based Approach

In order to dock longer HP to the full-length PCPE-1 protein, which is unfeasible for the AD3 protocol described above due to the limitation of the available number of the degrees of freedom and, in general, because of the computationally very expensive conformational sampling required for simulations with such ligands, a fragment-based docking approach we recently developed was applied [64]. In brief, first, HP dp3 of both types, IdoA(2S)-GlcNS(6S)-IdoA(2S) and GlcNS(6S)-IdoA(2S)-GlcNS(6S), were docked using the AD3 docking procedure described in Section 3.5.1. Then, all 1000 solutions for each HP dp3 fragments were used to assemble $\sim 10^5$ dp11 GAG chains applying the approach standard parameters [64]. This was followed by the refinement and all-atom conversion of the chains with a slight modification in the original scripts to avoid the RMSD-based selection of the best fitting structures compared to the experimental ones due to the lack of proper atomistic experimental PCPE-1/GAG complex structures. In particular, instead of the previously described way of selecting structures, a simple RMSD-based clustering (cutoff 4 Å) was performed to filter out the duplicates and find the most relevant structures. From the resulted ~ 5-40 atomistic structures the ones with significantly different docking poses were selected and refined together with the full protein using MD simulations applying the same procedure as described in Section 3.6, except for the minimization procedure where at the first step of 10^4–10^5 steepest descent minimization cycles were applied before the conjugate gradient minimization step.

3.6. All-Atom MD Simulations and MM-GBSA Free Energy Calculations

All-atom molecular dynamics (MD) simulations of the PCPE-1, PCPE-1/Ca^{2+} and PCPE-1/Ca^{2+}/GAG complexes obtained by molecular docking were performed with the use of the AMBER16 MD package [50]. Periodic boundary conditions in a truncated octahedron TIP3P water box with at least 8 Å distance from the solute to the periodic box border were used. Na$^+$ and Cl$^-$ monovalent counterions were used to neutralize the system. ff99SB force field parameters for protein [65] and the GLYCAM06 [66] for GAGs were used, respectively. Prior to MD production runs, two energy-minimization steps were performed: first, 500 steepest descent cycles and 1000 conjugate-gradient cycles with harmonic force restraints on solute (10 kcal/mol/Å^2), then, 3000 steepest-descent cycles and 3000 conjugate-gradient cycles without restraints. After the minimization, the system was heated up to 300 K for 10 ps with harmonic force restraints on solute (10 kcal/mol/Å^2), equilibrated for 100 ps at 300 K and 10^5 Pa in isothermal isobaric ensemble (NTP). This was followed by a 100 ns MD production run in the same NTP ensemble. The SHAKE algorithm, 2 fs time integration, 8 Å cutoff for non-bonded interactions, and the particle mesh Ewald method were used. The trajectories were analyzed using the cpptraj module of AMBER Tools [46]. Free-energy calculations and per-residue energy decomposition were done using molecular mechanics-generalized born surface area (MM-GBSA) model igb = 2 [67] for protein-GAG and protein-Ca^{2+} complexes for the parts of the trajectory where convergence in terms of RMSD was obtained. The obtained energy values account explicitly for the enthalpy and implicitly for the solvent entropy. For this reason, the reported energies should not be strictly interpreted as full free energy of binding: the entropic contribution to binding

was not taken into account explicitly. Entropy calculations were shown to dramatically increase the overall noise in the free binding energies when used within MM-GBSA free energy calculation schemes in general [68] and for GAG containing systems particularly [69].

3.7. Ca^{2+} Ion Position Prediction

We applied and compared several approaches to predict the Ca^{2+} binding sites on the surface of PCPE-1 and its domains. Prior to applying these approaches to PCPE-1, we analyzed their performance for the complex between an annexin V and HP, where Ca^{2+} ions are known to be stable and to contribute to GAG binding (PDB ID: 1G5N, 2.7 Å) [35]. In this structure, 9 Ca^{2+} were resolved.

3.7.1. Molecular Dynamics Approach

We used the MD approach with the protocols described in Section 3.6 to predict the binding sites of Ca^{2+} on the surface of protein. The length of each MD simulation run was 100 ns. In these calculations, Ca^{2+} ions were placed randomly in the periodic box. For annexin V, 9 Ca^{2+} were used corresponding to the number of Ca^{2+} ions observed in the experimental structure. For CUB1-CUB2, NTR domains, three ions were used. The number of ions was chosen in order to sample effectively the protein surface within a reasonable simulation time. For the CUB1-CUB2 and the NTR domains, the simulations were repeated 5 times. The trajectories were analyzed, and the for the frames where the Ca^{2+} ions were stably bound in terms of RMSD convergence for the coordination complex, MM-GBSA free energy calculations were performed. The obtained values for the predicted Ca^{2+} binding sites were compared with the corresponding energies obtained from the simulations of annexin V-Ca^{2+} and CUB1-CUB2-Ca^{2+} crystal structures. The latter was extracted from the structure of the complex of CUB1-CUB2 with procollagen peptide (PDB ID: 6FVZ, 2.7 Å).

3.7.2. FoldX and IonCom

We used the scripts of FoldX available at http://foldx.embl.de and online ion ligand binding site prediction tool IonCom at https://zhanglab.ccmb.med.umich.edu/IonCom for Ca^{2+} binding site predictions. FoldX represents a tool with an implemented empirical force field developed for effective evaluation of the contribution of mutations on the stability, folding and dynamics of proteins and nucleic acids [70,71]. As an output FoldX yields the positions of predicted ions on the surface of the protein. In contrast, IonCom utilizes ab initio training and template-based information to output a list of protein residues potentially involved in ion binding [72]. Both programs were used with the default parameters.

3.8. Visualization and Data Analysis

VMD [51], Chimera [73] and Pymol [74] were used for structural analysis visualization, MD trajectory analysis, as well as for the graphics production. R package was used for data analysis [75].

4. Conclusions

We report here the computational analysis of the interactions of full-length PCPE-1 and its domains with GAGs of various lengths and sequences. This model of full-length PCPE-1 based on SAXS restraints is in agreement with the experimental values of its radius of gyration and length. The full-length protein binds GAGs through the NTR domain and the interdomain linker, while the binding to CUB1-CUB2 domains is weaker, likely non-specific, and less energetically favorable. GAG preferential binding to the NTR domain is mostly electrostatically-driven. CS6 is predicted to bind to a different site of the NTR domain than the other GAGs, which may account for the experimental differences previously observed between CS6 and DS/HP [28]. Fragment-based docking of longer GAG oligosaccharides results in overlap with the docking poses obtained for shorter GAGs and corresponds to more favorable interactions than those established by shorter oligosaccharides in agreement with

the strongest inhibition of PCPE-1-HP interactions by longer HP oligosaccharides reported in this study. Our results suggest that calcium ions may bind to GAGs before they interact with PCPE-1 or may stabilize the structure and conformation of full-length PCPE-1. Although we have used several computational approaches to predict Ca^{2+} binding sites on the protein surface, considering calcium ions as a part of the protein receptor for docking is not an approach applicable to all systems.

From a methodological point of view, we have shown that the size of the clusters identified by molecular docking is not correlated with their free binding energies obtained in the MD simulation. We have successfully applied, for the first time, fragment-based docking to dp11 oligosaccharides, which will be useful for the computational characterization of protein interactions with long GAGs, which are challenging to study using conventional docking approaches.

Supplementary Materials: Supplementary Materials can be found at http://www.mdpi.com/1422-0067/20/20/5021/s1.

Author Contributions: J.P., K.K.B. obtained and analyzed the data. G.K., A.G.L, A.L. adapted the software to the system, obtained and analyzed the data and wrote the manuscript. S.R.-B. and S.A.S. are responsible for the conceptualization of the work and wrote the manuscript. S.R.-B. and E.K. performed and interpreted SPR binding assays. S.A.S supervised J.P., G.K., K.K.B., A.G.L. and acquired the funding.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

AD3	Autodock 3
BMP-1	Bone morphogenetic protein-1
CG	Coarse-grained
CS	Chondroitin sulfate
CUB	Complement, sea urchin protein Uegf, BMP-1
DS	Dermatan sulfate
GAG	Glycosaminoglycan
HP	Heparin
HS	Heparan sulfate
MD	Molecular dynamics
MM-GBSA	Molecular mechanics-generalized born surface area
mTLD	Mammalian tolloid
NTR	Netrin-like domain
PBSA	Poisson-Boltzmann surface area
PCPE-1	Procollagen C-proteinase enhancer-1
SAXS	Small angle X-ray scattering
UNRES	United residue force field

References

1. Esko, J.D.; Kimata, K.; Lindahl, U. Proteoglycans and Sulfated Glycosaminoglycans. In *Essentials of Glycobiology*, 2nd ed.; Varki, A., Cummings, R.D., Esko, J.D., Freeze, H.H., Stanley, P., Bertozzi, C.R., Hart, G.W., Etzler, M.E., Eds.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, USA, 2009; pp. 1–784.
2. Pomin, V.H.; Mulloy, B. Glycosaminoglycans and Proteoglycans. *Pharmaceuticals* **2018**, *11*, 27. [CrossRef]
3. Proudfoot, A.E. Chemokines and Glycosaminoglycans. *Front. Immunol.* **2015**, *6*, 246. [CrossRef]
4. Shute, J. Glycosaminoglycan and chemokine/growth factor interactions. *Handb. Exp. Pharmacol.* **2012**, *207*, 307–324.

5. Iozzo, R.V.; Zoellerm, J.J.; Nyströmm, A. Basement membrane proteoglycans: Modulators Par Excellence of cancer growth and angiogenesis. *Mol. Cells* **2009**, *27*, 503–513. [CrossRef] [PubMed]

6. Almond, A. Multiscale modeling of glycosaminoglycan structure and dynamics: Current methods and challenges. *Curr. Opin. Struct. Biol.* **2018**, *50*, 58–64. [CrossRef] [PubMed]

7. Pichert, A.; Samsonov, S.A.; Theisgen, S.; Thomas, L.; Baumann, L.; Schiller, J.; Beck-Sickinger, A.G.; Huster, D.; Pisabarro, M.T. Characterization of the interaction of interleukin-8 with hyaluronan, chondroitin sulfate, dermatan sulfate and their sulfated derivatives by spectroscopy and molecular modeling. *Glycobiology* **2012**, *22*, 134–145. [CrossRef] [PubMed]

8. Penk, A.; Baumann, L.; Huster, D.; Samsonov, S.A. NMR and Molecular Modeling Reveal Specificity of the Interactions between CXCL14 and Glycosaminoglycans. *Glycobiology* **2019**. [CrossRef] [PubMed]

9. Bojarski, K.K.; Sieradzan, A.K.; Samsonov, S.A. Molecular Dynamics Insights into Protein-Glycosaminoglycan Systems from Microsecond-Scale Simulations. *Biopolymers* **2019**. [CrossRef]

10. Uciechowska-Kaczmarzyk, U.; Chauvot de Beauchene, I.; Samsonov, S.A. Docking software performance in protein-glycosaminoglycan systems. *J. Mol. Graph. Mod.* **2019**, *90*, 42–50. [CrossRef]

11. Rother, S.; Samsonov, S.A.; Hofmann, T.; Blaszkiewicz, J.; Köhling, S.; Schnabelrauch, M.; Möller, S.; Rademann, J.; Kalkhof, S.; von Bergen, M.; et al. Structural and functional insights into the interaction of sulfated glycosaminoglycans with tissue inhibitor of metalloproteinase-3—A possible regulatory role on extracellular matrix homeostasis. *Acta Biomater.* **2016**, *45*, 143–154. [CrossRef]

12. Vallet, S.D.; Miele, A.E.; Uciechowska-Kaczmarzyk, U.; Liwo, A.; Duclos, D.; Samsonov, S.A.; Ricard-Blum, S. Insights into the structure and dynamics of lysyl oxidase propeptide, a flexible protein with numerous partners. *Sci. Rep.* **2018**, *8*, 11768. [CrossRef] [PubMed]

13. Ricard-Blum, S.; Ruggiero, F.; van der Rest, M. The collagen superfamily. *Top. Curr. Chem.* **2005**, *247*, 35–84.

14. Vadon-Le Goff, S.; Hulmes, D.J.S.; Moali, C. BMP-1/tolloid-like proteinases synchronize matrix assembly with growth factor activation to promote morphogenesis and tissue remodeling. *Matrix Biol.* **2015**, *44–46*, 14–23. [CrossRef] [PubMed]

15. Takahara, K.; Lyons, G.E.; Greenspan, D.S. Bone morphogenetic protein-1 and a mammalian tolloid homologue (mTld) are encoded by alternatively spliced transcripts which are differentially expressed in some tissues. *J. Biol. Chem.* **1994**, *269*, 32572–32578.

16. Kessler, E.; Takahara, K.; Biniaminov, L.; Brusel, M.; Greenspan, D.S. Bone Morphogenetic Protein-1: The Type I Procollagen C-Proteinase. *Science* **1996**, *271*, 360–362. [CrossRef]

17. Li, S.W.; Sieron, A.L.; Fertala, A.; Hojima, Y.; Arnold, W.V.; Prockop, D.J. The C-proteinase that processes procollagens to fibrillar collagens is identical to the protein previously identified as bone morphogenic protein-1. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 5127–5130. [CrossRef]

18. Kessler, E.; Hassoun, E. Procollagen C-Proteinase Enhancer 1 (PCPE-1) in Liver Fibrosis. *Methods Mol. Biol.* **2019**, *1944*, 189–201.

19. Hassoun, E.; Safrin, M.; Ziv, H.; Pri-Chen, S.; Kessler, E. Procollagen C-proteinase enhancer 1 (PCPE-1) as a plasma marker of muscle and liverf in mice. *PLoS ONE* **2016**, *11*, e0159606.

20. Takahara, K.; Kessler, E.; Biniaminov, L.; Brusel, M.; Eddy, R.L.; Jani-Saitfl, S.; Shows, T.B.; Greenspan, D.S. Type I procollagen COOH-terminal proteinase enhancer protein: Identification, primary structure, and chromosomal localization of the cognate human gene (PCOLCE). *J. Biol. Chem.* **1994**, *269*, 26280–26285.

21. Bányai, L.; Patthy, L. The NTR module: Domains of netrins, secreted frizzled related proteins, and type I procollagen C-proteinase enhancer protein are homologous with tissue inhibitors of metalloproteases. *Protein Sci.* **1999**, *8*, 1636–1642. [CrossRef]

22. Liepinsh, E.; Banyai, L.; Pintacuda, G.; Trexler, M.; Patthy, L.; Otting, G. NMR structure of the netrin-like domain (NTR) of human type I procollagen C-proteinase enhancer defines structural consensus of NTR domains and assesses potential proteinase inhibitory activity and ligand binding. *J. Biol. Chem.* **2003**, *278*, 25982–25989. [CrossRef] [PubMed]

23. Pulido, D.; Sharma, U.; Vadon-Le Goff, S.; Hussain, S.A.; Cordes, S.; Mariano, N.; Bettler, E.; Moali, C.; Aghajari, N.; Hohenester, E.; et al. Structural Basis for the Acceleration of Procollagen Processing by Procollagen C-Proteinase Enhancer-1. *Structure* **2018**, *26*, 1384–1392. [CrossRef] [PubMed]

24. Gaboriaud, C.; Gregory-Pauron, L.; Teillet, F.; Thielens, N.M.; Bally, I.; Arlaud, G.J. Structure and properties of the Ca^{2+}-binding CUB domain, a widespread ligand-recognition unit involved in major biological functions. *Biochem. J.* **2011**, *439*, 185–193. [CrossRef] [PubMed]

25. Blanc, G.; Font, B.; Eichenberger, D.; Moreau, C.; Ricard-Blum, S.; Hulmes, D.J.S.; Moali, C. Insights into how CUB domains can exert specific functions while sharing a common fold. Conserved and specific features of the CUB1 domain contribute to the molecular basis of procollagen C-proteinase enhancer-1 activity. *J. Biol. Chem.* **2007**, *282*, 16924–16933. [CrossRef]

26. Bernocco, S.; Steiglitz, B.M.; Svergun, D.I.; Petoukhov, M.V.; Ruggiero, F.; Ricard-Blum, S.; Ebel, C.; Geourjon, C.; Deleage, G.; Font, B.; et al. Low resolution structure determination shows procollagen C-proteinase enhancer to be an elongated multidomain glycoprotein. *J. Biol. Chem.* **2003**, *278*, 7199–7205. [CrossRef]

27. Moschcovich, L.; Bernocco, S.; Font, B.; Rivkin, H.; Eichenberger, D.; Chejanovsky, N.; Hulmes, D.J.; Kessler, E. Folding and activity of recombinant human procollagen C-proteinase enhancer. *Eur. J. Biochem.* **2001**, *268*, 2991–2996. [CrossRef]

28. Weiss, T.; Ricard-Blum, S.; Moschcovich, L.; Wineman, E.; Mesilaty, S.; Kessler, E. Binding of procollagen C-proteinase enhancer-1 (PCPE-1) to heparin/heparan sulfate: Properties and role in PCPE-1 interaction with cells. *J. Biol. Chem.* **2010**, *285*, 33867–33874. [CrossRef]

29. Bekhouche, M.; Kronenberg, D.; Vadon-Le Goff, S.; Bijakowski, C.; Lim, N.H.; Font, B.; Kessler, E.; Colige, A.; Nagase, H.; Murphy, G.; et al. Role of the netrin-like domain of procollagen C-proteinase enhancer-1 in the control of metalloproteinase activity. *J. Biol. Chem.* **2010**, *285*, 15950–15959. [CrossRef]

30. Sankaranarayanan, N.V.; Nagarajan, B.; Desai, U.R. So you think computational approaches to understanding glycosaminoglycan-protein interactions are too dry and too rigid? Think again! *Curr. Opin. Struct. Biol.* **2018**, *50*, 91–100. [CrossRef]

31. Samsonov, S.A.; Pisabarro, M.T. Computational analysis of interactions in structurally available protein-glycosaminoglycan complexes. *Glycobiology* **2016**, *26*, 850–861. [CrossRef]

32. Möbius, K.; Nordsieck, K.; Pichert, A.; Samsonov, S.A.; Thomas, L.; Schiller, J.; Kalkhof, S.; Pisabarro, M.T.; Beck-Sickinger, A.G.; Huster, D. Investigation of lysine side chain interactions of Interleukin-8 with Heparin and other glycosaminoglycans studied by a methylation-NMR approach. *Glycobiology* **2013**, *23*, 1260–1269. [CrossRef] [PubMed]

33. Ester, M.; Kriegel, H.; Sander, J.; Xu, X. A density-based algorithm for discovering clusters in large spatial databases with noise. In Proceedings of the International Conference on Knowledge Discovery and Data Mining, Portland, OR, USA, 2–4 August 1996; pp. 226–231.

34. Imberty, A.; Lortat-Jacob, H.; Pérez, S. Structural view of glycosaminoglycan-protein interactions. *Carbohydr. Res.* **2007**, *342*, 430–439. [CrossRef] [PubMed]

35. Capila, I.; Hernáiz, M.J.; Mo, Y.D.; Mealy, T.R.; Campos, B.; Dedman, J.R.; Linhardt, R.J.; Seaton, B.A. Annexin V—Heparin oligosaccharide complex suggests heparan sulfate—Mediated assembly on cell surfaces. *Structure* **2001**, *9*, 57–64. [CrossRef]

36. Woodhead, N.E.; Long, W.F.; Willamson, F.B. Binding of zinc ions to heparin. Analysis by equilibrium dialysis suggests the occurrence of two, entropy-driven, processes. *Biochem J.* **1986**, *237*, 281–284. [CrossRef] [PubMed]

37. Stevic, I.; Parmar, N.; Paredes, N.; Berry, L.R.; Chan, A.K.C. Binding of heparin to metals. *Cell Biochem. Biophys.* **2011**, *59*, 171–178. [CrossRef] [PubMed]

38. Multhaup, G.; Bush, A.I.; Pollwein, P.; Masters, C.L. Binding of heparin to metals. *FEBS Lett.* **1994**, *335*, 151–154. [CrossRef]

39. Eckert, R.; Ragg, H. Zinc ions promote the interaction between heparin and heparin cofactor II. *FEBS Lett.* **2003**, *541*, 121–125. [CrossRef]

40. Ricard-Blum, S.; Féraud, O.; Lortat-Jacob, H.; Rencurosi, A.; Fukai, N.; Dkhissi, F.; Vittet, D.; Imberty, A.; Olsen, B.R.; van der Rest, M. Characterization of endostatin binding to heparin and heparan sulfate by surface plasmon resonance and molecular modeling: Role of divalent cations. *J. Biol. Chem.* **2004**, *279*, 2927–2936. [CrossRef]

41. Han, Q.; Fu, Y.; Zhou, H.; He, Y.; Luo, Y. Contributions of Zn(II)-binding to the structural stability of endostatin. *FEBS Lett.* **2007**, *581*, 3027–3032. [CrossRef]

42. Zhang, F.; Liang, X.; Beaudet, J.M.; Lee, Y.; Linhardt, R.J. The Effects of Metal Ions on Heparin/Heparin Sulfate-Protein Interactions. *J. Biomed. Technol. Res.* **2014**, *1*, 10. [CrossRef]

43. Seo, Y.; Schenauer, M.R.; Leary, J.A. Biologically Relevant Metal-Cation Binding Induces Conformational Changes in Heparin Oligosaccharides as Measured by Ion Mobility Mass Spectrometry. *Int. J. Mass Spectrom.* **2011**, *303*, 191–198. [CrossRef] [PubMed]

44. Lerner, L.; Torchia, D.A. A multinuclear NMR study of the interactions of cations with proteoglycans, heparin, and Ficoll. *J. Biol. Chem.* **1986**, *261*, 12706–12714. [PubMed]

45. Parrish, R.F.; Fair, W.R. Selective binding of zinc ions to heparin rather than to other glycosaminoglycans. *Biochem. J.* **1981**, *193*, 407–410. [CrossRef] [PubMed]

46. Grushka, E.; Cohen, A.S. The Binding of Cu(II) and Zn(II) Ions by Heparin. *Anal. Lett.* **1982**, *15*, 1277–1288. [CrossRef]

47. Grant, D.; Long, W.F.; Williamson, F.B. Infrared spectroscopy of heparin-cation complexes. *Biochem. J.* **1987**, *244*, 143–149. [CrossRef]

48. Sattelle, B.; Shakeri, J.; Almond, A. Does microsecond sugar ring flexing encode 3D-shape and bioactivity in the heparanome? *Biomacromolecules* **2013**, *14*, 1149–1159. [CrossRef]

49. Samsonov, S.; Bichmann, L.; Pisabarro, M.T. Coarse-grained model of glycosaminglycans. *J. Chem. Inf. Mod.* **2015**, *55*, 114–124. [CrossRef]

50. Case, D.A.; Ben-Shalom, I.Y.; Brozell, S.R.; Cerutti, D.S.; Cheatham, T.E., III; Cruzeiro, V.W.D.; Darden, T.A.; Duke, R.E.; Ghoreishi, D.; Gilson, M.K.; et al. *AMBER16*; University of California: San Francisco, CA, USA, 2018.

51. Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38. [CrossRef]

52. Rhee, Y.M.; Pande, V.S. Multiplexed-replica exchange molecular dynamics method for protein folding simulation. *Biophys. J.* **2003**, *84*, 775–786. [CrossRef]

53. Hansmann, U.H.E.; Okamoto, Y. Comparative study of multicanonical and simulated annealing algorithms in the protein folding problem. *Physica A* **1994**, *212*, 415–437. [CrossRef]

54. Liwo, A.; Baranowski, M.; Czaplewski, C.; Gołaś, E.; He, Y.; Jagieła, D.; Krupa, P.; Maciejczyk, M.; Makowski, M.; Mozolewska, M.A.; et al. A unified coarse-grained model of biological macromolecules based on mean-field multipole-multipole interactions. *J. Mol. Model.* **2014**, *20*, 2306. [CrossRef] [PubMed]

55. Liwo, A.; Czaplewski, C.; Ołdziej, S.; Scheraga, H.A. Computational techniques for efficient conformational sampling of proteins. *Curr. Opin. Struct. Biol.* **2008**, *18*, 134–139. [CrossRef] [PubMed]

56. Karczyńska, A.S.; Mozolewska, M.A.; Krupa, P.; Giełdoń, A.; Liwo, A.; Czaplewski, C. Prediction of protein structure with the coarse-grained UNRES force field assisted by small X-ray scattering data and knowledge-based information. *Proteins* **2018**, *86*, 228–239. [CrossRef] [PubMed]

57. Kumar, S.; Bouzida, D.; Swendsen, R.H.; Kollman, P.A.; Rosenberg, J.M. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **1992**, *13*, 1011–1021. [CrossRef]

58. Murtagh, F.; Heck, A. Multivariate Data Analysis. In *Astrophysics and Space Science Library*; Kluwer Academic Publishers: Strasbourg, France, 1987; pp. 55–109.

59. Rotkiewicz, P.; Skolnick, J. Fast procedure for reconstruction of full-atom protein models from reduced representations. *J. Comput. Chem.* **2008**, *29*, 1460–1465. [CrossRef]

60. Morris, G.M.; Goodsell, D.S.; Halliday, R.S.; Huey, R.; Hart, W.E.; Belew, R.K.; Olson, A.J. Automated docking using a Lamarcklan algorithm an empirical binding free energy function. *J. Comput. Chem.* **1998**, *19*, 1639–1662. [CrossRef]

61. Samsonov, S.; Gehrcke, J.-P.; Pisabarro, M.T. Flexibility and explicit solvent in molecular dynamics-based docking of protein-glycosaminoglycan systems. *J. Chem. Inf. Mod.* **2014**, *54*, 582–592. [CrossRef]

62. Joseph, R.R.; Mosier, P.D.; Desai, U.R.; Rajarathnam, K. Solution NMR characterization of chemokine CXCL8/IL-8 monomer and dimer binding to glycosaminoglycans: Structural plasticity mediates differential binding interactions. *Biochem. J.* **2015**, *472*, 121–133. [CrossRef]

63. Nivedha, A.K.; Makeneni, S.; Foley, B.L.; Tessier, M.B.; Woods, R.J. Importance of ligand conformational energies in carbohydrate docking: Sorting the wheat from the chaff. *J. Comput. Chem.* **2014**, *35*, 526–539. [CrossRef]

64. Samsonov, S.A.; Zacharias, M.; Chauvot de Beauchene, I. Modeling large protein-glycosaminoglycan complexes using a fragment-based approach. *J. Comput. Chem.* **2019**, *40*, 1429–1439. [CrossRef]

65. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* **2006**, *65*, 712–725. [CrossRef] [PubMed]

66. Kirschner, K.N.; Yongye, A.B.; Tschampel, S.M.; González-Outeiriño, J.; Daniels, C.R.; Foley, B.L.; Woods, R.J. GLYCAM06: A generalizable biomolecular force field. Carbohydrates. *J. Comput. Chem.* **2008**, *4*, 622–655. [CrossRef] [PubMed]

67. Onufriev, A.; Case, D.A.; Bashford, D. Effective Born radii in the generalized Born approximation: The importance of being perfect. *J. Comput. Chem.* **2002**, *23*, 1297–1304. [CrossRef] [PubMed]

68. Homeyer, N.; Gohlke, H. Free Energy Calculations by the Molecular Mechanics Poisson-Boltzmann Surface Area Method. *Mol. Inf.* **2012**, *31*, 114–122. [CrossRef] [PubMed]

69. Gandhi, N.S.; Mancera, R.L. Free energy calculations of glycosaminoglycan-protein interactions. *Glycobiology* **2009**, *19*, 1103–1115. [CrossRef]

70. Schymkowitz, J.; Borg, J.; Stricher, F.; Nys, R.; Rousseau, F.; Serrano, L. The FoldX web server: An online force field. *Nucleic Acids Res.* **2005**, *33*, W382–W388. [CrossRef]

71. Schymkowitz, J.W.; Rousseau, F.; Martins, I.C.; Ferkinghoff-Borg, J.; Stricher, F.; Serrano, L. Prediction of water and metal binding sites and their affinities by using the Fold-X force field. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 10147–10152. [CrossRef]

72. Hu, X.; Dong, Q.; Yang, J.; Zhang, Y. Recognizing metal and acid radical ion binding sites by integrating ab initio modeling with template-based transferals. *Bioinformatics* **2016**, *32*, 3260–3269. [CrossRef]

73. Pettersen, E.F.; Goddard, T.D.; Huang, C.C.; Couch, G.S.; Greenblatt, D.M.; Meng, E.C.; Ferrin, T.E. UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612. [CrossRef]

74. *The PyMOL Molecular Graphics System*, Version 1.2r3pre; Schrödinger, LLC: New York, NY, USA, 2002.

75. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2013; ISBN 3-900051-07-0. Available online: http://www.R-project.org/ (accessed on 1 May 2017).