Carbohydrate Research 543 (2024) 109201

Contents lists available at ScienceDirect

Carbohydrate Research

journal homepage: www.elsevier.com/locate/carres

In silico approaches for better understanding cysteine cathepsin-glycosaminoglycan interactions

Krzysztof K. Bojarski [a,*,1], Alexis David [b,c], Fabien Lecaille [b,c,1], Sergey A. Samsonov [d,1]

[a] Department of Physical Chemistry, Gdansk University of Technology, Narutowicza 11/12, Gdansk, 80-233, Poland
[b] Université de Tours, Tours, France
[c] INSERM, UMR 1100, Centre d'Etude des Pathologies Respiratoires (CEPR), Team "Mécanismes Protéolytiques dans l'Inflammation, Tours, France
[d] Faculty of Chemistry, University of Gdansk, Wita Stwosza 63, Gdansk, 80-308, Poland

ARTICLE INFO

Keywords:
Molecular modeling
Molecular docking
Molecular dynamics
Proteolytic modulation
Procathepsin processing
Glycosaminoglycans

ABSTRACT

Cysteine cathepsins constitute the largest cathepsin family, with 11 proteases in human that are present primarily within acidic endosomal and lysosomal compartments. They are involved in the turnover of intracellular and extracellular proteins. They are synthesized as inactive procathepsins that are converted to mature active forms. Cathepsins play important roles in physiological and pathological processes and, therefore, receive increasing attention as potential therapeutic targets. Their maturation and activity can be regulated by glycosaminoglycans (GAGs), long linear negatively charged polysaccharides composed of recurring dimeric units. In this review, we summarize recent computational progress in the field of (pro)cathepsin-GAG complexes analyses.

1. Introduction

1.1. Glycosaminoglycans structure and biological role

GAGs represent a class of long linear negatively charged polysaccharides. GAG chains are composed of repeating disaccharide units, in which (with the exception of keratan sulfate) an amino sugar and an uronic acid are present [1]. Each GAG possesses a distinct sulfation pattern, delineating the exact locations of sulfate groups along the carbohydrate chain [2]. Taking into account high number of feasible sulfation patterns within a GAG repeating disaccharide unit, these molecules are chemically extremely heterogenous. There are six main classes of GAGs: heparin (HP), heparan sulfate (HS), chondroitin sulfate (CS), dermatan sulfate (DS), hyaluronic acid (HA), and keratan sulfate (KS) (Fig. 1).

GAGs are mainly present in the extracellular matrix (ECM) and in the acidic lysosomes [4]. With the exception of HA, GAGs are covalently bound to core proteins, forming proteoglycans [5]. GAGs fulfill various biological functions in living organisms, related to their structural properties. HA is composed of N-acetyl-D-glucosamine (GlcNAc) and D-glucuronic acid (GlcA) units with −1 electrostatic charge per disaccharide unit, making it one of the least charged GAGs. HA can form unbranched chains with a wide spectrum of polysaccharide chain length, starting from 100 to even 30 000 disaccharide units per one chain [5], depending on the tissue origin [6]. HA plays a role in tissue differentiation [7], as evidenced by its association with cell migration and proliferation. Furthermore, HA has been found to induce gene expression related to wound healing in macrophages, endothelial cells, eosinophils, and specific epithelial cells [8–10]. The presence of HA degradation products triggers an inflammatory response, serving as an injury signal [11]. Disaccharide unit of CS consisting of N-acetyl-D-galactosamine (GalNAc) and GlcA. Sulfation of hydroxyl groups can occur either at the 4th carbon atom position of GalNAc (chondroitin 4-sulfate; C4–S, CS-A) or at the 6th carbon atom position of the same aminosugar (chondroitin 6-sulfate; C6–S, CS-C), leading to a net charge of −2 per disaccharide unit. In some instances, sulfation can occur at both the 4th and 6th carbon atom positions of GalNAc, producing "highly sulfated" chondroitin (chondroitin 4,6-sulfate, C4,6-S, CS-E) with a net charge of −3 per disaccharide unit. These variants of CS are identified animals other than mammals. CS carbohydrate chains are much shorter compared to HA as they are usually composed of 40–100 recurring disaccharide units [12]. CS primarily participates in bone resorption [13]. DS shares a disaccharide structure similar to CS but with iduronic acid (IdoA) instead of GlcA. DS chains usually consists of 40–100 recurring disaccahride units [12]. In terms of biological role DS exhibits antithrombotic activity [14]. KS lacks uronic acid in its structure and its

* Corresponding author.
E-mail address: krzysztof.bojarski@pg.edu.pl (K.K. Bojarski).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.carres.2024.109201
Received 12 April 2024; Received in revised form 25 June 2024; Accepted 28 June 2024
Available online 6 July 2024

disaccharide unit mainly composed of GlcNAc and D-galactose (Gal), each sulfated at the 6th position. KS carbohydrate chains are made up of 5–34 disaccharide units, making it one of the shortest GAGs [15]. KS can be found in the cornea, contributing to proper tissue hydration necessary for corneal transparency [16]. Both HP and HS share a disaccharide unit consisting of D-glucosamine (GlcN) and IdoA for HP, and GlcA/IdoA for HS with a varying sulfation pattern (GlcNSAc, GlcNSAc(6S), GlcNS(6S), GlcNS(3S6S), IdoA(2S), GlcA(2S)). The difference which allows to distinguish between HP and HS is the percentage of these residues and their distribution in the linear chains [17–20]. It is proposed that there are 20 variants of HS disaccharide unit, making it one of the most heterogenous GAGs [21]. Their polysaccharide chains consist of 10–50 and 25–200 disaccharide units, for HP [22] and HS [23], respectively. Both GAGs play essential roles in angiogenesis, anticoagulation, cell growth and development, and defense against viral invasion [24]. Additionally, they are vital for protease storage and contribute to wingless signaling when they are bound to a protein core [25]. Besides altered sulfation, abnormal GAG expression affects cell biological

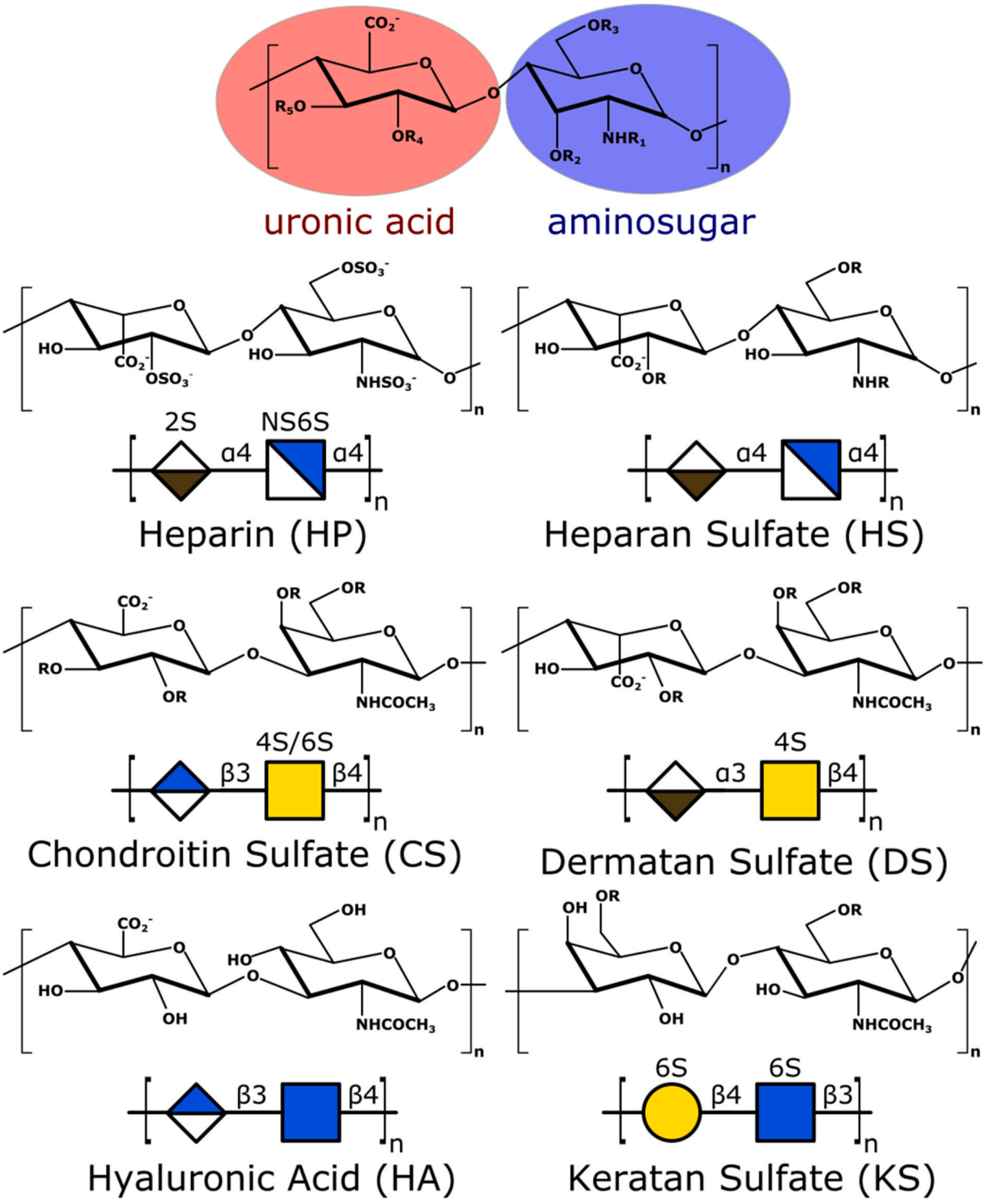

Fig. 1. Dimeric units of different GAG classes employed in molecular modelling studies described in this review along with their symbol nomenclature for glycans (SNFG) representation [3].

behavior by dysregulating enzymatic activity of their protein targets and finally drives to the development of various severe diseases, such as cancers [14], autoimmune diseases [26], Alzheimer's disease [27], Parkinson's disease [28], arthritis [29], and mucopolysaccharidosis [30]. Considering their involvement in development of aforementioned diseases and their biological significance GAGs emerge as pivotal promising targets for regenerative medicine [31,32]. Specifically, GAGs present opportunities for potential modifications aimed at enhancing their essential biological properties. These modifications could entail the addition of extra sulfate groups [33–35] or the substitution of sulfates with alternative negatively charged groups, such as phosphate groups [36]. The objective of these modifications is to modulate the strength and selectivity of protein-GAG interactions, thereby augmenting the inhibition or activation of protein activity.

1.2. Challenges in analysing glycosaminoglycan containing systems

The molecular description of protein-GAG interactions often necessitates the application of both experimental and computational methods for their accurate characterization. While these complementary approaches offer diverse perspectives on protein-GAG interactions, each one faces distinct limitations. Up to date there are two main experimental methods that have been so far successfully applied to determine structure of protein-GAG systems: X-ray crystallography and nuclear magnetic resonance (NMR). Their challenges for protein-GAG complexes are the following.

Application of X-ray crystallography requires primarily the acquisition of a crystal of the studied complex. Due to the specific conditions of crystallization, this process is not always straightforward, while the obtained structure may not properly reflect protein-GAG complex in solvent.

NMR analysis of protein-GAG interactions necessitates sufficient purified concentrated samples for experimentation. Furthermore, proteins, which size exceed a molecular weight of 35 kDa are challenging to analyse with this method due to the complexity of the spectrum assignment.

Considering that analysis of protein-GAG complexes at the atomic level with aforementioned methods is both time consuming and a costly processes, computational approaches are usually employed. The *in silico* analysis of protein-GAG systems also features essential challenges due to the GAG intrinsic properties that often require application of advanced techniques.

GAGs are highly flexible molecules [37]. With the increase of length of the polysaccharide chain, the number of degrees of freedom that should be accounted for increases linearly. This originates from the pyranose rings which may adapt various conformations [38] corresponding to different binding affinities when these molecules are bound [39], which also increases the number of degrees of freedom that should be considered in modeling. Conformational flexibility is not only limited to pyrasone rings as glycosidic linkages of GAGs may also adapt different conformations, which correspond to different binding affinities [40]. Furthermore, it was observed that the degree of HP chain bending increased with the length of the polysaccharide chain, while maintaining approximately linear structure [41]. This is pivotal in molecular docking since many conventional molecular docking approaches are limited to handle ligands with a large number of degrees of freedom. Additionally, the negatively charged chains of GAG interact with the long flexible positively charged side chains of Arg and Lys residues [42,43], which adds more conformational space of the complex, and therefore increases the computational cost and time of the analysis [44].

Considering the density of negative charges of GAGs, electrostatics is one of the major driving force of interactions with protein, conferring an additional challenge in GAG-protein identification [45]. Moreover, the driving force for assembly is also dominated by long-range, ion specific interactions [46] as well by solvent molecules [47]. The role of ions on GAG molecular recognition was discussed in several experimental studies [48–54]. Furthermore, application of *ab initio* and molecular dynamics approach revealed that ions might have direct effect on ring puckering [55] as well as GAG global conformational properties [56].

GAGs from the same class may bind to their protein targets with comparable binding free energies. As a result, despite the simplicity of their carbohydrate backbones, GAGs structures can be exceedingly complex mainly due to their sulfation pattern (position and density), which is a critical determinant of protein-GAG interactions, a key aspect dictating structural attributes, molecular recognition, and biological activity [57].

GAGs can likely interact with protein at the same site but with different poses, yielding similar binding free energies. This phenomenon is known as multipose binding [58,59]. Such low specificity GAG-binding was observed for GAG complexes with Interleukin-8 [60], Hepatocyte Growth Factor [61] and SARS-CoV-2 protein [62]. Furthermore, a GAG might bind in the same pose but with antiparallel orientations, resulting in indistinguishable binding free energies [63].

In recent years, several computational tools for discovering protein interactions have been developed. However, despite these advancements, there remains a lack of consistent and reliable frameworks for studying GAG interactions [64]. Among the known challenges in this area are predicting binding poses, determining the appropriate system size when solvating in explicit solvent, and keeping force field parameters up-to-date [65,66].

The availability of experimental structures of protein-GAG complexes is limited. According to GAG-DB, to date, there are only 109 such structures available [67]. Consequently, the application of machine learning algorithms remains a significant challenge. This stems from the fact that with an insufficient number of available experimental structures, machine learning algorithms lack access to a diverse range of structural information essential for accurate analysis and effective predictions. Moreover, the scarcity of thermodynamic data further complicates the development of effective machine learning methods in this field. Finally, setting up proper scoring functions remains a substantial challenge [68].

1.3. Molecular modelling methods to characterise protein-glycosaminoglycan interactions

Despite the challenges discussed above, a combination of following computational approaches is usually employed to characterise protein-GAG interactions constituting a general pipeline with the following steps (Fig. 2).

Calculation of electrostatic potential isosurfaces provides data on potential GAG binding regions on the protein surface by employing, for example, Poisson-Boltzmann Surface Area (PBSA) method from AmberTools package [69]. In this approach, solvent is treated implicitly by Poisson-Boltzmann equation [70] which allows to perform analysis in short time. Protein is represented as a dielectric body with vacuum permittivity whose shape is defined by atomic coordinates and radii. As a result, electrostatic potential data of the protein surface is obtained, which allows to identify regions that might be electrostatically attractive or repulsive for GAGs. This method has been succesfully applied in different studies, especially in work of Samsonov et al. in which for most of the experimentally available protein-GAG complexes, authors correctly predicted binding regions [71].

Molecular docking aids in prediction of GAG binding poses. In standard molecular docking protocol few steps are performed in order to obtain protein-GAG structures. First, ligand binding poses are generated. This could be performed for the molecules with fixed conformation and with no conformational flexibility allowed (rigid molecular docking). It is also possible to allow for certain conformational changes (mostly for the ligand) and then to perform flexible molecular docking by employing different algorithms such as fast shape matching, incremental construction, simulated annealing, distance geometry-based algorithm or evolutionary programming. Moreover, ligand binding poses could be

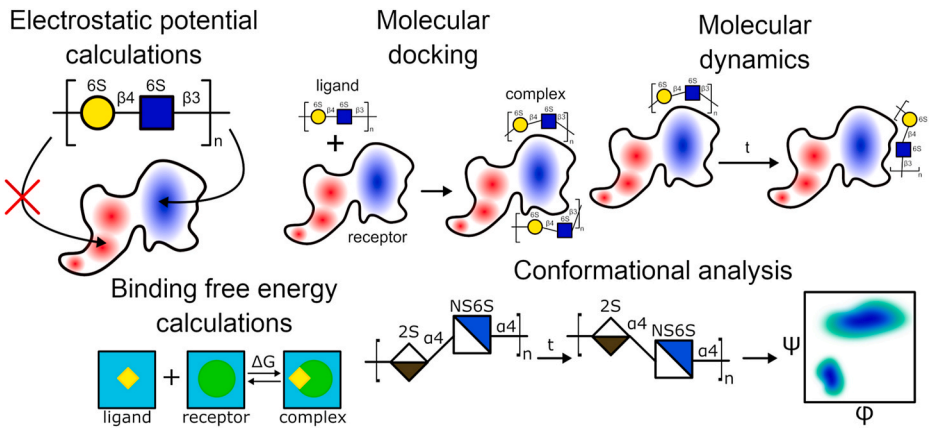

Fig. 2. Schematic representation of the computational pipeline employed to study protein-GAG interactions.

predicted taking into consideration whole surface of the receptor or just part of it, usually based on the results from PBSA calculations. In the next step calculated structures are ranked. The obtained structures are classified by scores reflecting interaction energies that could have very diverse origin and the nature of physical principles considered. In the prediction of protein-GAG complexes Autodock 3 [72] software is oftently employed with parameters adjusted to protein-GAG systems [71], as it was shown to perform well for a wider variety of protein-GAG complexes [73]. Furthermore, in order to properly address solvent-mediated and ion-mediated interactions, advanced molecular docking methods such as steered molecular dynamics [74,75] or repulsive scaling replica exchange molecular dynamics [76] are employed.

Molecular dynamics (MD) allows to study dynamical evolution of a protein-GAG complex over time by numerically solving Newton's motion equations [77,78]. Forces between the particles and their potential energies are calculated by using specific interatomic potentials defined in specific force fields, such as bonded, non-bonded, electrostatic and van der Waals interactions as well as dihedral angles. MD might be performed in all-atom (in which each atom is considered in calculations) or coarse-grained respresentation [79] (a reduced model in which group of atoms are treated as one pseudoatom). MD simulations for GAG-containing systems are oftenly performed with application of ff14SB/GLYCAM06j [69,80,81] and CHARMM36 m [66,82] force fields.

Binding free energy analysis is employed to estimate the binding affinity in a protein-GAG complex and the impact of specific amino acid residues on its stability. In order to do that, Molecular Mechanics Poisson-Boltzmann Surface Area (MM-PBSA) method is employed [83,84], which uses Generalised-Born approximation. By summing up electrostatic component with hydrophobic contribution term it is possible to calculate full solvation free energy. This calculated energy is next computed with binding free energy *in vacuo* in order to produce binding free energy in solvent. In this approach free energy components of binding and per residue individual energy impacts such as bonded (bond, angle, dihedral) and non-bonded (van der Waals, electrostatic, polar solvation, surface area) energy components are post-processed based on the obtained MD trajectories.

Considering the challenges inherent in employing computational methodologies by applying GAG-specific approaches, it was nevertheless feasible to characterise *in silico* protein-GAG interactions in various biological systems. Among the studied protein targets that have been studied successfully are growth factors [40,63], cysteine cathepsins [85,86], SARS-CoV-2 protein [87], N-deacetylase-N-sulfotransferase [88], antithrombin [89], bone morphogenic protein 6 [90], chemokines [91] and collagen [92].

1.4. Cysteine cathepsins: structure, synthesis, expression, and biological roles

Human cysteine cathepsins (Cat) (CatB, C, F, H, K, L, O, S, V, X, and W) belong to the family C1 of clan CA cysteine proteases, sharing the typical papain-like fold, which is composed of two domains (L-left domain, R-right domain), with the V-shaped active site cleft (catalytic residues: Cys25, His159 and Asn175 papain numbering [93]) at the interface (Fig. 3). Cysteine cathepsins are monomeric proteins (20–35 kDa mass range) except for CatC, which forms tetramers (circa. 200 kDa).

Cathepsins are synthesized as inactive precursors (zymogens) termed as procathepsins. Procathepsins undergo post-translational glycosylation and are directed to the lysosomal compartment mainly through cellular mannose-6-phosphate receptors, and alternatively by other receptors depending on cell-type [94,95]. They can be found in acidic endosomes/lysosomes compartments and are responsible for the intracellular endocytosed protein turn-over. Furthermore, they are also released outside the lysosomes (cytosol, mitochondria, nucleus, or extracellular environment) where they exert many proteolytic events, including zymogen activation, cellular protein digestion, and extracellular matrix (ECM) and basement membrane (BM) remodeling [96–100]. However, in various human diseases, their proteolytic activity can be dysregulated, ensuing frequently in their extracellular overexpression, as observed in cancer, muscular dystrophy, hepatitis, rheumatoid arthritis, cardiovascular and bone diseases, lung diseases, immune system-related disorders, and neurodegenerative diseases [95]. Consequently, cysteine cathepsins including CatB, CatC, CatK, CatL, and CatS have been proposed as potential therapeutic targets for several inflammatory disorders. Cysteine cathepsins show highest enzymatic activity at acidic pH (optimal pH 5) and are unstable at neutral pH [96]. However, CatS remains stable and active at neutral pH, contrary to the other related cathepsins [101]. CatB and CatK are found partially active at neutral pH in the presence of GAGs [102,103]. Cathepsin activity is subject to regulation by various physicochemical conditions (e.g., pH, temperature, ionic strength, and redox status), and endogenous protein inhibitors like members of the cystatin family. Cathepsins activity can be also inhibited by their respective propeptides [104–110]. In addition, the proteolytic activity of mature CatB, CatK, CatS, CatV, and CatX is modulated by sulfated GAGs [30,111]. Besides the scope of this review focused on mammalian cysteine cathepsins, a number of studies have also reported that GAGs can also modulate the activity and the stability of other cysteine proteases from other organism such as plant [112] and microorganism pathogens (i.e., Trypanosoma, Leishmania, Schistosoma [113–116]). In the last two decades, GAGs have emerged as regulators of human cysteine cathepsin enzymatic activities in disease (i.e., osteoporosis [117], rheumatoid arthritis [118], Alzheimer's disease [119],

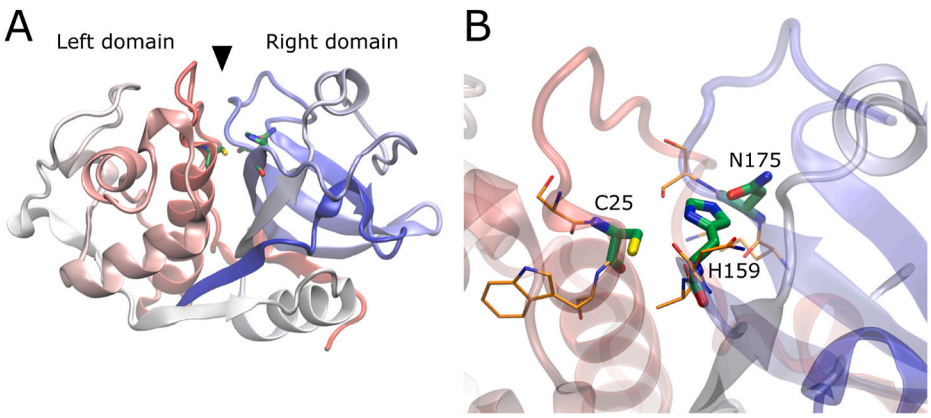

Fig. 3. Structure of CatK (PDB ID: 4X6H, 1.0 Å) in cartoon representation (A) with its active site cleft (B). Catalytic site residues are shown in green sticks.

carcinoma [120] and mucopolysaccharidosis [121]). At the same time, structural information on cathepsin-GAG interactions is limited. Mature human CatK-C4-S and DS hexamer complexes are the only X-ray crystal structures available up to now (PDB IDs: 3C9E, 3H7D, 4N8W, 4N79), and structure prediction of cathepsin-GAG interactions predominantly stems from *in silico* approaches [85,86,122,123].

Therefore, a comprehensive understanding of the molecular mechanisms underlying GAG-cysteine cathepsin interactions to selectively modulate protease activity is one promising avenue for clinical utility. In this review, we summarize recent progress on molecular interactions between GAG and cathepsin (pro- and mature forms) by computational studies and highlight their significance as pillar of investigation, alongside the conventional experimentation's workbench. Specifically, our focus is on elucidating the latest *in silico* studies dedicated to the CatB, CatK, CatL, CatS, and CatV complexes with GAGs. These complexes are not only biologically relevant but have also been extensively explored through diverse experimental approaches in the past.

2. *In silico* methods in analysing (pro)cathepsin-glycosaminoglycan interactions

2.1. Modulation of procathepsins B and S processing by GAGs

Cysteine cathepsins are synthesized as inactive pre-proenzymes. After the removal of the N-terminal signal peptide in the endoplasmic reticulum, the propeptide helps in the proper folding of cathepsins and their delivery to the endosomes/lysosomes. Their inactivity arises from the presence of the propeptide that sterically blocks the active site. In lysosomes, the propeptide is cleaved and removed by other proteases [94], or by autocatalytic processing [124,125]. In previous studies, GAGs (i.e., CS, HP, DS) and dextran sulfate favors the autocatalytic removal of the propeptide of cathepsins B, K, L, and S at both acidic (pH $\leq$ 5.0) and higher (neutral) pH. This suggests that negatively charged GAGs interact with specific basic residues of CatB, K, L, and S propeptides, which have a theoretical isolectric point (pI) above 9.0, except for CatK (pI ~6.3 [126]) and are therefore positively charged. Even though no studies have been carried out on procathepsin V processing, it can be hypothesized that GAGs likely contribute to its maturation in regards to the pI of its prepropeptide (~9.5). This might be supported by the fact that CatV shares similar percentage of basic residues (Lys, Arg, His) present in propeptide with CatB, CatK, CatL and CatS (from 17.7 % to 22.4 %) (Fig. 4). Furthermore, some of these basic residues are highly conserved throughout proenzymes of these cathepsins, suggesting their role as putative GAG binding sites. Similarly, sequential signatures for heparin-binding sites (i.e., Cardin/Weintraub motifs [127]) are present in the prodomain of CatS, K, L, and V. No such motifs are found in the procathepsin B sequence, which is markedly shorter (62 residues) compared to the other forementioned cathepsins. Mutations in one of these motifs (K99 and K104) in the proenzyme of CatL altered the processing of procathepsin L [128]. This suggests that this propeptide consensus sequence is important in terms of its effect on CatL-like zymogen destabilization and activation in the extracellular environment.

In silico methods were employed to investigate the structure, corresponding potential enzymatic activities, and energetic characteristics of the interactions of diverse type and length of GAGs with procathepsins B [85] and S [86]. Molecular docking, MD simulations, and binding free energy calculations were performed to scrutinize the molecular interactions between procathepsins B, S, and GAGs (CS, DS, HP, HS, defined as a desulfated HP throughout the computational studies discussed in this review, and HA) with different degrees of polymerization (dp). Docking results for proCatB indicated that two residues of the propeptide (K50 and R54) were important in the interaction with C4–S (dp4 and dp6), C6–S (dp6), DS (dp6), HP (dp6), HS (dp4 and dp6). In addition, another GAG binding region on proCatB surface (residues R85, K86, K127, K130, K141 and K144) was present for C6–S (dp4), DS (dp4), HP (dp4 and dp6), HS (dp4). Presence of such binding regions suggests the possibility of allosteric regulation mechanism that may affect conformation of the active site of the enzyme and the propeptide. In a computational study Novinec proposed binding of C4–S to CatK may affect conformation of an active site [130]. Similar allosteric regulation of CatB by HS was observed by Costa et al. [131]. Furthermore, in their experimental study, Caglic et al. suggested that GAG binding to mature CatB may induce conformational change of propeptide [132]. It is likely that long GAG chains (>dp10) bind to both binding regions, and thus, regulate conformational flexibility of the propeptide and may allosterically affect the active site. For proCatS, GAGs showed stronger preference in binding to the propeptide as all clustered docking structures were bound at least partially to the propeptide. At the same time, it was possible to identify also favorable docking solutions structures, in which a GAG was not bound to the propeptide. Results of MD-based binding free energy analysis showed that HS formed the most stable complex with the propeptide of cathepsins B and S, favouring the HS tetramer (dp4) and hexamer (dp6), respectively. Stable complexes were also reported with HP, corroborating experimental results on the relevance of GAGs in the maturation *in vitro* of procathepsins. Subsequent microsecond MD simulations were conducted for cathepsins B and S (pro- and mature forms) in the absence and presence of HS (dp4), to study the effect of GAG on the proenzyme conformational flexibility and the geometry of the active site cleft. For proCatB, HS (dp4) binding drives to an increased flexibility of the active site, favoring its more open conformation and, therefore, autoproteolytic activity. In contrast, HS (dp4) binding decreases flexibility of the propeptide and as a result, may stabilize conformation in which the active site is accessible. In the absence of HS (dp4), the propeptide was more flexible, which may impel conformational change that uncovers CatB active site and makes it

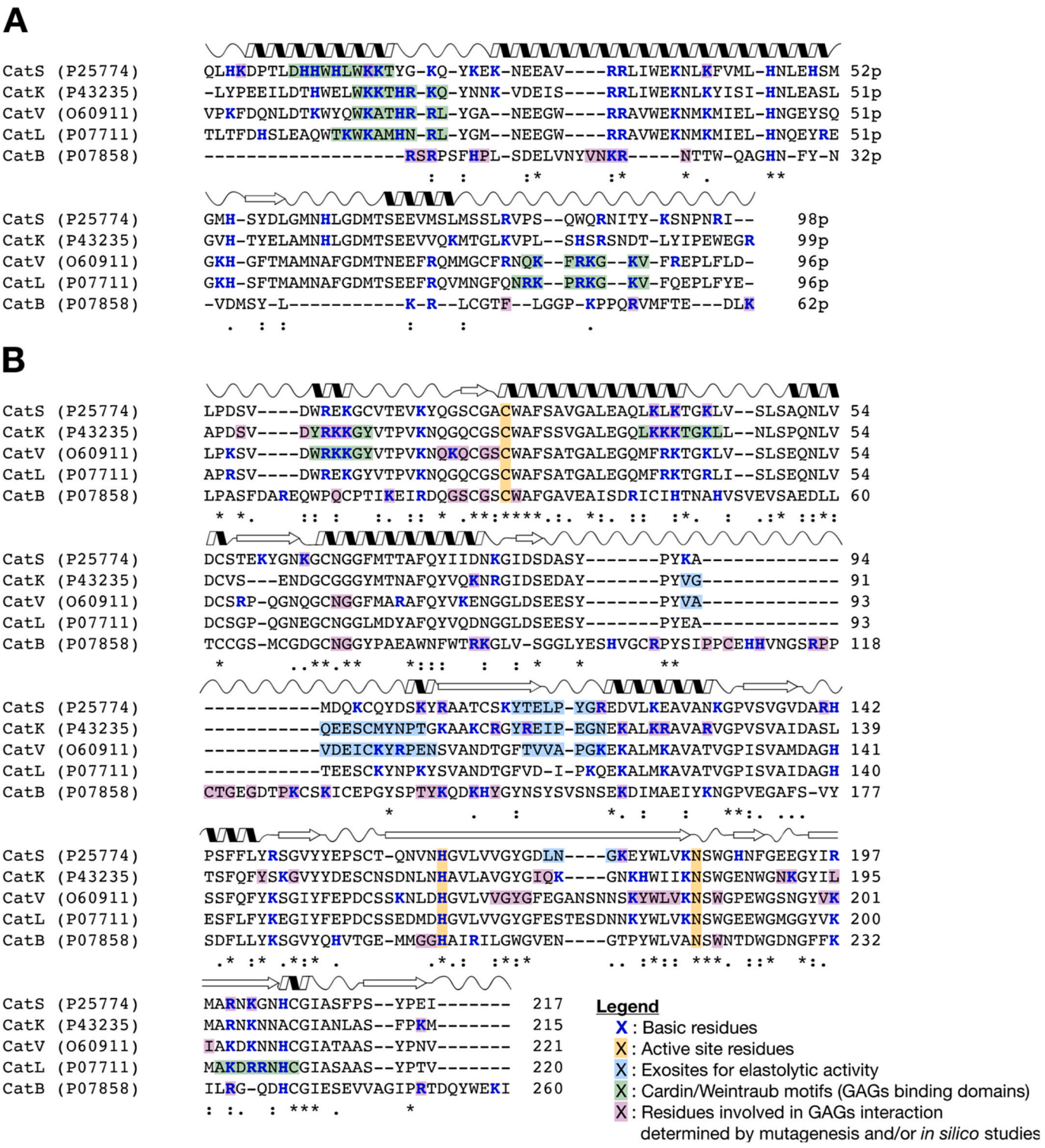

Fig. 4. Multiple sequence alignment of cathepsins S, K, V, L, and B. A) Primary sequences of prodomain and B) mature form of cathepsins. Sequences are from UniProtKB database and are aligned using Clustal Omega 1.2.4 program [129]. Corresponding secondary structure of CatL (proform and mature form) is placed above the sequences. Conservation of each residue can be found below the sequences with symbols signifying a conserved site (*), a conservative substitution (), a semi- (.) or a non-conserved site (). Cardin/Weintraub motifs [127] refer to similar XBBXBX and XBBBXXBX sequences with X and B, corresponding to hydropathic (neutral and hydrophobic) and basic residues, respectively.

accessible for another proCatB molecule. Regarding proCatS, it was suggested that the binding pose of HS has an additional alternative effect on zymogen structural stability. HS could bind elsewhere on the mature domain of CatS, thus promoting conformational changes of the propeptide from a closed inactive form to an opened active form. Altogether, these results supported the hypothesis that sulfated GAG binds to proCatB and proCatS in a conformation corresponding to available crystallographic structures of these unbound proenzymes (PDB ID: 3PBH, 2C0Y) and affect conformational flexibility of the propeptide and active site. What remained unexplored was whether GAG forms a

stable complex with proCatB and proCatS molecule in 'opened active' conformation (where the active site is accessible) as well. Therefore, coarse-grained modeling was employed to model corresponding structures. Subsequent *in silico* analysis for HP and HS, representing the most and the least charged GAGs respectively, suggested that these two GAGs form stable complexes with both proenzymes in their 'opened active' state.

Based on the computational results obtained for proCatS and proCatB, two molecular mechanisms of their GAG-mediated maturation were proposed as summarized in Fig. 5. Similar mechanism for proCatK processing was also suggested, as described [133]. Lemaire and collegues observed that binding of C4–S to proCatK may increase the rate of proenzyme's maturation akin to how HS induced conformational change of proCatS as observed in in silico study (Fig. 5A). Furthermore, they proposed that under acidic pH, proCatK may undergo conformational changes, which render the enzyme into an open active conformation (Fig. 5B). Of note, to facilitate the autocatalytic activation of recombinant papain-like cysteine proteases (C1 family, clan CA), the negatively charged GAG dextran sulfate is commonly used [134].

2.2. Specificity of the interactions between mature cathepsins and GAGs

Albeit mature form of human CatK, L, S, and V share high protein identity (>56 %), the electrostatic potentials of their molecular surface are markedly different (Fig. 6) [30]. The predicted isoelectric point (pI) of CatK, V, and S are 8.9, 8.6, and 7.6, respectively, and are higher than that of CatL (pI: 4.7). Multiple sequence alignement of cathepsins indicated difference in the number and position of basic residues and acidic residues (Fig. 4). Main differences primarily located in the active site (or in its close vicinity) of cathepsins, specifically the substrate binding site, which is weakly positive for CatV, while that of CatL is negative. This characteristic is likely determinant for their substrate preference and GAG interactions.

Recently, a systematic prediction of the interactions between CatB, CatK, CatL, CatS, and CatV and different types and lengths of GAG, followed by MD-based refinement and binding free energy analysis was performed [122]. Based on the results, authors suggested that stability of cathepsin-GAG complex increased with the GAG chain length, except for CatL, which electrostatic potential is predominanlty characterized by negatively charged surface [30], indicating that electrostatic interactions are likely the driving force of cathepsin-GAG interactions. For all analyzed cathepsin-GAG complexes three main binding regions were conserved (Figs. 7 and 8). One binding site is located on the back of the enzyme (R-domain) away from the active site, and partially overlaps with the C4–S binding site reported in CatK-C4-S crystallographic structure (PDB ID: 3C9E). Alternatively, GAGs bind to the L-domain of CatK, particularly to basic residues K77, R111 R127, and K214. Another binding site is located directly the active site of all cathepsins, which may dampen substrate cleavage sterically.

Identification of novel *in silico* binding sites might be essential for characterisation of GAG-mediated allosteric regulation of cathepsins' activity. The work by Costa et al. delves into the allosteric regulation of CatB by HP (dp2) [131]. The authors identified two regions where GAG binding occurs: one was located near the N-terminus and it partially overlaps with experimentally determined 4N8W-like binding site (Fig. 7B) and near the one of the recently proposed *in silico* binding site (*in silico BS1*, Fig. 8A). The other binding site was located on the L-domain of CatB in a vicinity of another recently reported *in silico* binding site (Fig. 8B). Authors also concluded that short HP chains likely bind to R-domain, while long ones might bind to L-domain instead. Next

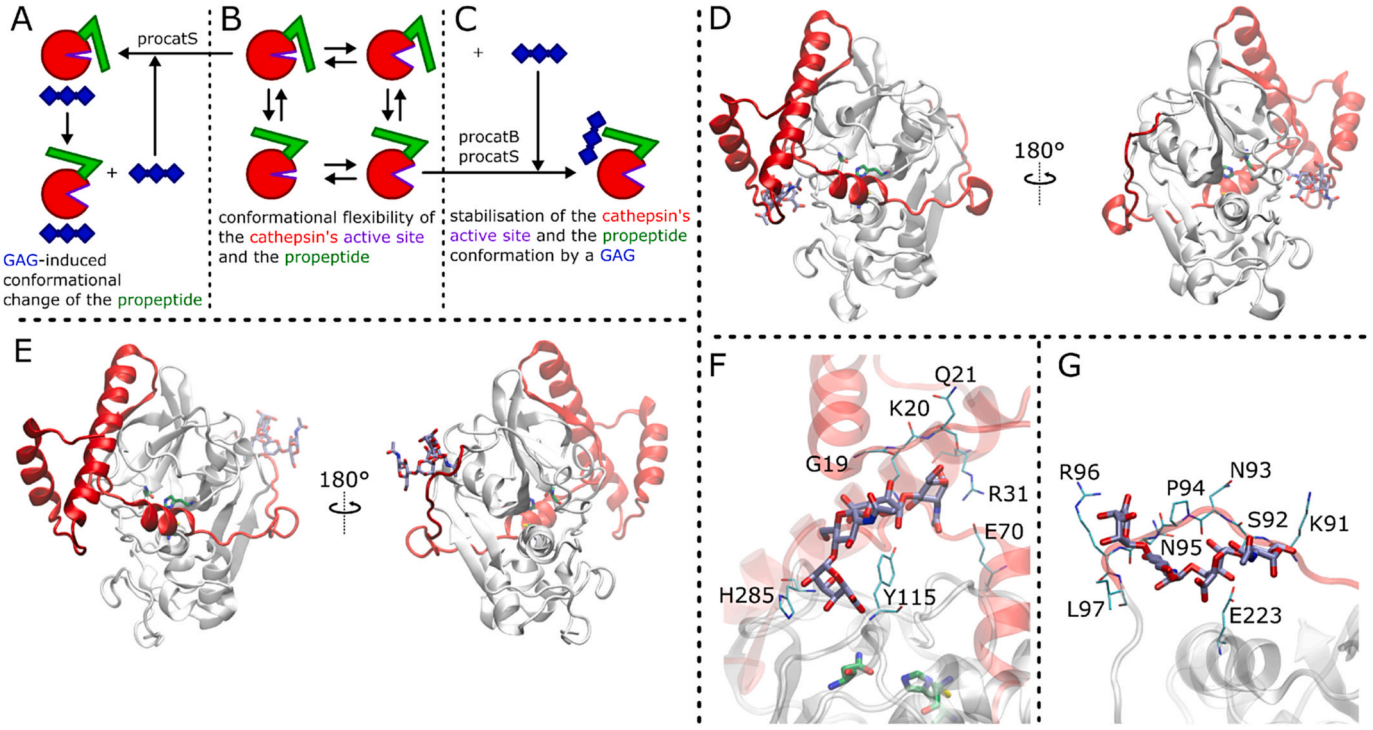

Fig. 5. Proposed mechanisms of GAG-mediated procathepsins B and S maturation (A, B, C). Similarly to one previously suggested for proCatB in the absence of GAG, propeptide of cathepsin S could change its conformation, adapting one that exposes the active site into an open conformation (B). GAG binding to proCatS in this altered conformation stabilizes its structure, including that of the active site, thus facilitating the maturation (C). An alternative processing pathway was also suggested for proCatS, which involves a direct interaction of GAG with the mature cathepsin region of the proenzyme (A). Formation of such a complex induces a conformational change of propeptide, which renders the active site accessible for another proCatS molecule (autoproteolytic process). (D, E) Structures of ProcatS-HS (dp4) with predicted binding regions. The propeptide and mature enzyme are depicted in red and white cartoon, respectively, and HS (dp4) is in cyan sticks. Panels (F, G) provide close-up views of the binding sites shown in (D, E), respectively. GAG binding may either induce a conformational change in the propeptide (E, G) or stabilize its conformation (D, F) by interacting with both propeptide and mature enzyme.

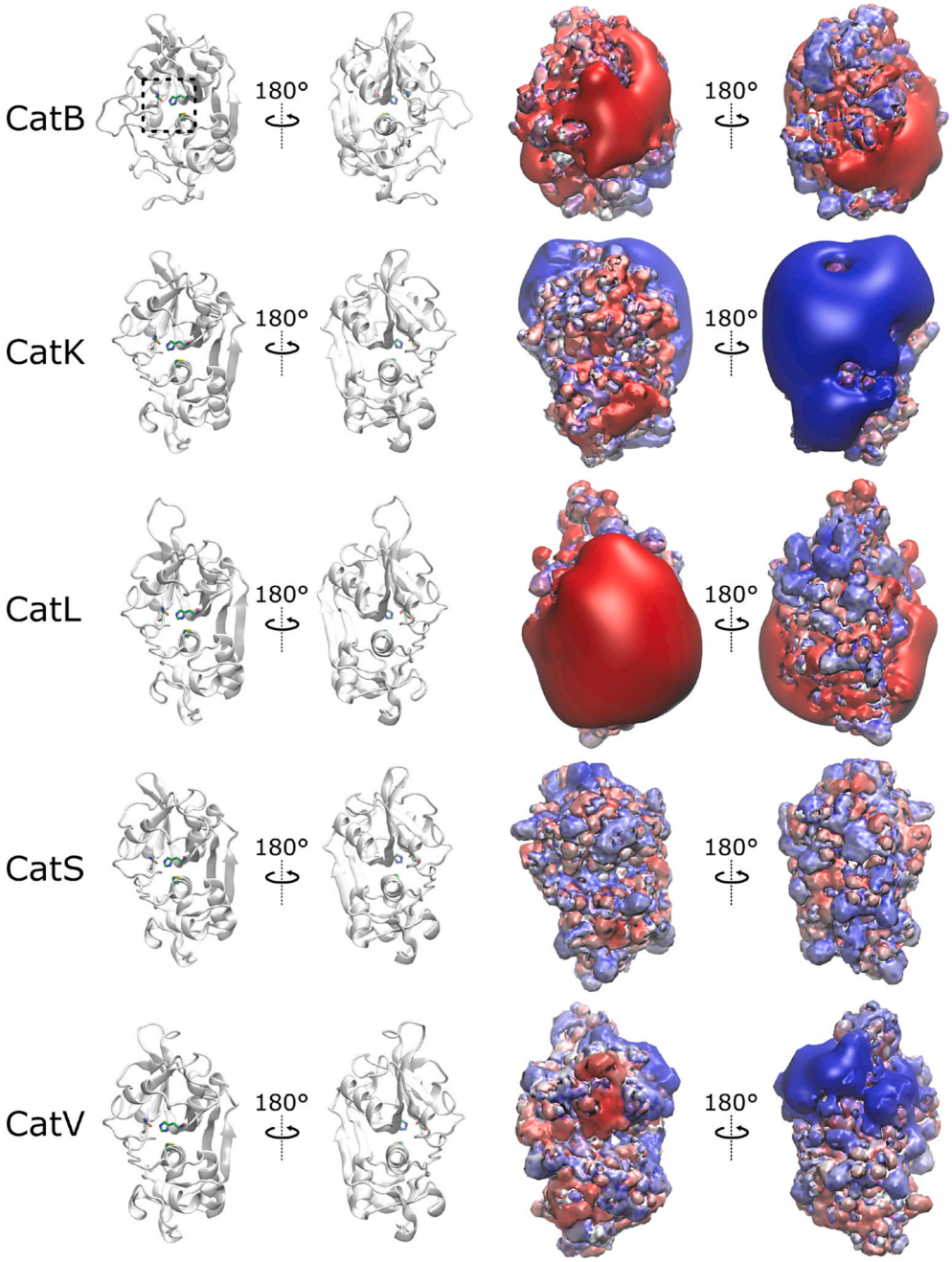

Fig. 6. Crystallogrpahic structures of CatB (PDB ID: 6AY2, 1.6 Å), CatK (PDB ID: 4X6H, 1.0 Å), CatL (PDB ID: 2XU3, 0.9 Å), CatS (PDB ID: 2H7J, 1.5 Å), CatV (PDB ID: 1FHO, 1.6 Å) in white cartoon representation and calculated electrostatic potential isosurfaces in surface representation (red, −3 kcal/(mol·e); blue, +3 kcal/(mol·e), respectively). Position of an active site is depicted with dotted square in reference to CatB. Electrostatic potential isosurfaces were calculated with "pbsa" module of AMBER [69] and visualised with VMD [135].

Costa and colleagues conducted MD simulations of the CatB-HP complex under varying pH conditions. Interestingly, they found that the unbound CatB remained highly stable at acidic pH levels. However, as the pH increased, the enzyme exhibited increased conformational flexibility. Consequently, the interactions between active site residues were disrupted, potentially elucidating the enzyme's inhibition under alkaline conditions. Furthermore, the binding of HP to CatB at this pH induced a rearrangement of contacts between enzyme domains, thereby favoring the maintenance of helical content and active site stability. These findings shed light on the intricate interplay between pH, GAG binding, and the structural dynamics of CatB, offering valuable insights into its regulatory mechanisms. The allosteric mechanism was further explored in the study conducted by Novinec [130]. This investigation focused on elucidating the conformational flexibility of unbound CatF, CatK, CatL, CatS, and CatV, as well as CatK in complex with C4–S (dp6) and small substrates NSC13345 and NSC94914. Notably, the study characterized three binding regions for C4–S: two of these regions aligned with the crystallographic structures available for CatK-C4-S (dp6) complexes (Fig. 7A and B), while the third was situated between the L- and R-domains, a region that had recently been investigated *in silico* as a potential binding site (Fig. 8C). It was hypothesized that the binding region of the allosteric regulator could play a crucial role in modulating enzymatic activity. Additionally, the study suggested that the binding of any of the aforementioned effectors did not significantly impact the overall conformational flexibility of the entire CatK molecule, but rather affected its active site. These findings offer valuable insights into the

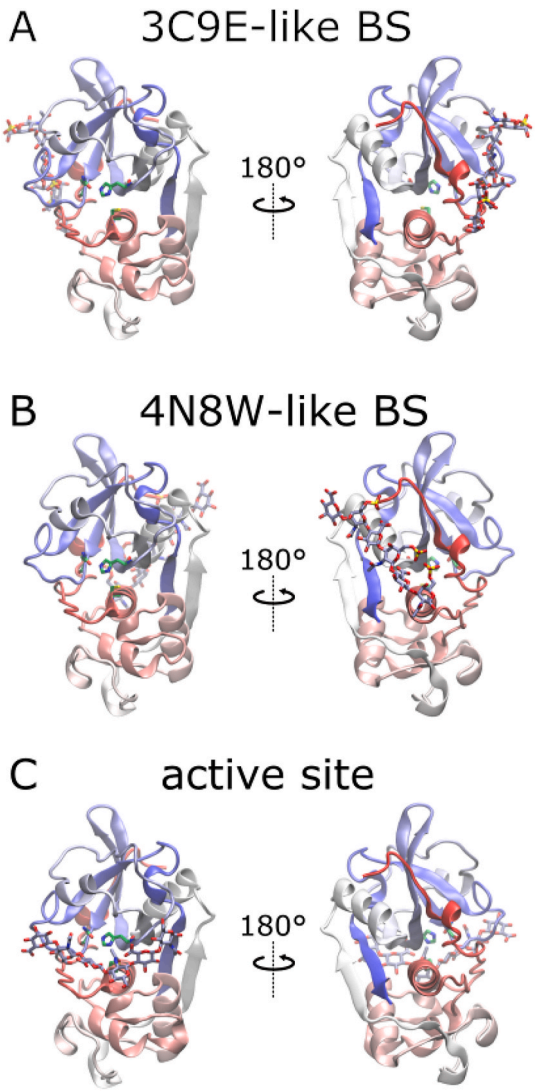

Fig. 7. Structures of Cat-GAG (dp6) complexes which represent experimentally determined binding sites (BS): (A) CatK-DS; (B) CatK-C4-S; (C) CatS-HA.

intricate regulatory mechanisms governing cathepsin function.

Analysis of binding free energies revealed that sulfation pattern is essential in the stability of cathepsin-GAG complex. For cathepsins B, L and S differences in stability were observed for C4–S and C6–S, which have the same net charge. In general, CatB-C6-S, CatL-C4-S and CatS-C4-S complexes were more energetically favorable than CatB-C4-S, CatL-C6-S, and CatS-C6-S, respectively. It is essential to point out that albeit CatL-GAG complexes were reported with favorable binding free energies, suggesting a potential binding, the lifetimes of these complexes were shorter than the other investigated Cat-GAG complexes. These lifetimes were estimated by taking into account the time during which GAG was bound to the cathepsin, using RMSD as a metric. This corroborates previous results, which reported the absence of complex formation for cathepsin L in the presence of C4–S, using gel mobility shift assays [103]. For cathepsins K and V, complexes with C4–S and C6–S were of similar affinity. In summary, these findings suggest that the sulfation pattern of GAG has important biological consequences for cathepsin-GAG interactions. Additionally, it can be concluded that binding free energies might correlate with the lifetimes of Cat-GAG complexes, as shorter lifetimes have been reported for complexes with lower binding affinity.

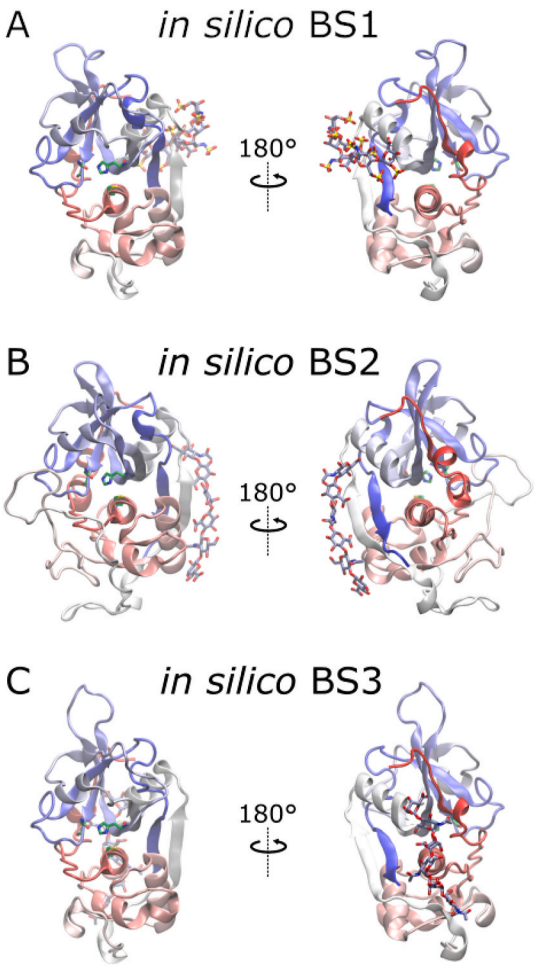

Fig. 8. Structures of Cat-GAG (dp6) complexes which represent binding sites (BS) identified with application of *in silico* approaches. (A) CatV-C6-S; (B) CatB-HA; (C) CatL-HS.

2.3. Chondroitin 4-sulfate enhances collagenolytic activity of human cathepsin K

Human CatK (hCatK) is highly expressed in osteoclasts and primarily responsible for efficiently cleaving the triple helical chains of both type I and II collagens, thus playing a vital role in bone resorption [136]. The potent collagenolytic activity of CatK is induced by the formation of a specific high molecular weight CatK-C4-S complex, extensively discussed in the literature [137,138]. Of note, in the absence of C4–S, the ability of CatK to efficiently cleave the collagen fibrils is hindered. Two experimental structures were obtained for the CatK-C4-S complex, revealing two distinct binding sites for C4–S (PDB ID: 3C9E and 4N8W). These two crystal structures have been associated with low and high collagenolytic activities of CatK, respectively [137,138]. Given its biological significance, dysfunction of CatK could lead to severe disorders such as pycnodysostosis [139], osteoporosis [140], rheumatoid arthritis, bone cancer [141], lung fibrosis [142], and lymphangioleiomyomatosis [143]. Consequently, effective treatment relies on identifying inhibitors that can mitigate the excessive activity of CatK by dicephering its molecular regulatory mechanisms. All synthetic inhibitors that have been clinically evaluated so far exclusively target the active site of CatK, thus blocking both its collagenolytic activity and other biological proteolytic activities such as maturation of thyroid hormones and TGF-β 1 hydrolysis [144]. Therefore, strategies that consist of blocking its active site may trigger unwanted side effects. An alternative approach would consist of targeting these two GAG-binding exosites to selectively

hamper the excessive collagenolytic activity of CatK in disease [145]. In these lines, application of computational methods could potentially allow to investigate such binding at accessible time scales. MD simulation approach has been successfully applied to uncover molecular mechanisms of CatK-substrate [146] and CatK-GAG interactions [122].

Computational endeavors effectively identified two main C4–S (dp6) binding sites on CatK surface with a greater preference of the binding pose corresponding to the 4N8W structure compared to the 3C9E structure in terms of the predicted binding affinity (Fig. 7). Similarly for other tested GAGs (C6–S, DS, HA, HP, HS), the same two binding regions corresponding to crystallographic structures 3C9E and 4N8W were identified. In addition, novel binding regions were identified on different parts of the L-domain or between L- and R-domains of CatK (*in silico* BS1, BS2 and BS3, Fig. 8).

To validate the reliability and robustness of *in silico* GAG-binding sites in hCatK, enzymatic assays were tested on its rat counterpart, which is a common animal model for the evaluation of hCatK inhibitors. These two enzymes share high sequence identity (88 %) and similarity (93 %), initially suggesting a similar C4–S-mediated mechanism. However, experimental evidence reported that albeit C4–S could enhance the collagenolytic activity of hCatK, it had negligeable effect on rCatK [123]. Subsequent computational studies aimed to discern the atomic level origin of this disparity. Analysis of electrostatic potential isosurfaces for rCatK and hCatK revealed similar GAG putative binding regions on their surfaces. Binding free energy analysis of hCatK-C4-S (dp6) complex allowed prediction of the most stable complex that corresponded to the crystallographic structure 4N8W. Preference towards 4N8W binding region was also reported in computational study of Maszota-Zieleniak et al. [147]. By employing Repulsive Scaling Replica Exchange Molecular Dynamics [76] authors reported a C4–S binding to both experimentally determined binding regions with a clear preference for the 4N8W. In contrast to hCatK-C4-S complex, the rCatK-C4-S complex corresponded to the low collagenolytic activity (3C9E-like structure). The results derived from the C4–S (dp6) simulations suggested that for longer polysaccharide chains, differences between the two enzymes were more pronounced, which may explain their distinct collagenolytic activities in the presence of C4–S. In fact, docking solutions for C4–S (dp12) to rCatK and hCatK revealed notable disparities in the obtained binding poses. For rCatK it was observed that C4–S (dp12) binds to the protein surface in position, where it can interact with amino acids responsible for CatK-C4-S interactions in both 3C9E and 4N8W crystallographic structures at the same time. For hCatK C4–S (dp12) could interact with amino acids involved in C4–S binding in either 3C9E or 4N8W experimental structure only. MD simulations and binding free energy analyses suggested stability of the rCatK complex with C4–S (dp12) connecting at both binding regions.

In another study, Marcisz et al. described C4–S (dp12) binding to hCatK with Umbrella Sampling approach, starting from experimental structure 3C9E, in which C4–S (dp6) was elongated [148]. Authors observed preference towards structure in which C4–S (dp12) was located in similar binding region to previously reported rCatK-C4-S (dp12) complex. Marcisz et al. also observed high complexity of the free energy landscape in the proximity of the native pose.

All these computational findings that enabled the rational interpretation of the experimental data, facilitated the proposal of a molecular mechanism for collagen degradation involving rCatK and hCatK in the presence of C4–S (Fig. 9). For rCatK, the binding of C4–S (dp12) to both the 3C9E and 4N8W binding sites leads to the formation of a molecular complex associated with lower collagenolytic activity [137]. In contrast, for hCatK, C4–S (dp12) binds only to 4N8W, forming a complex characterized by higher collagenolytic activity [138].

2.4. GAGs inhibit enzymatic activity of human cathepsins S and V

2.4.1. Cathepsin S

Cathepsin S plays a pivotal role with other proteases in the

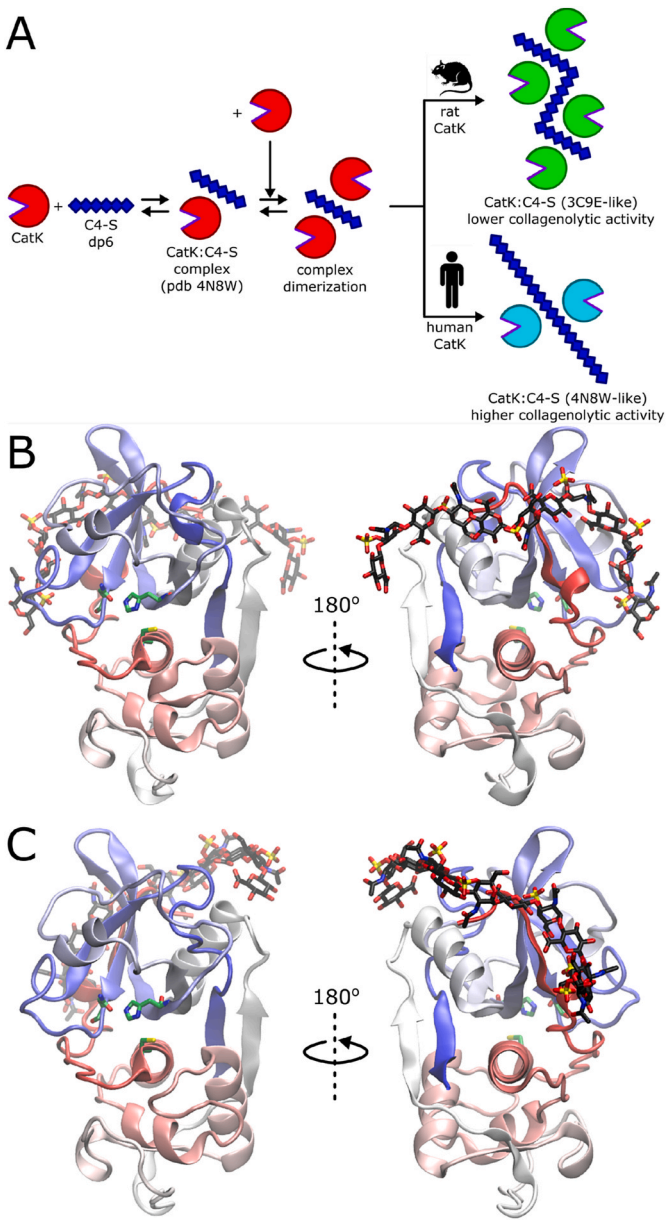

Fig. 9. Mechanism of collagen hydrolysis by rCatK and hCatK in the presence of C4–S (A). Structures of hCatK (B) and rCatK (C) with C4–S (dp12).

degradation of ECM and BM proteins [149–151]. Its dysregulation has been linked to various inflammatory skin conditions, as well as disorders such as arthritis, asthma, psoriasis, and atherosclerosis [152,153]. Consequently, CatS has been proposed as a potential therapeutic target in these human diseases [154–156]. To understand the impact of GAG on the enzymatic activity of CatS, several GAGs, including C4–S, C6–S, HP, HS, and HA, were investigated both computationally and *in vitro* [157]. Among them, C4–S and HS exhibited the strongest inhibitory effect on type IV collagen degradation as well its peptidase activity. This enzymatic activity supression effect was reversed upon the addition of NaCl, suggesting the significance of electrostatic interactions. Additionally, intrinsic fluorescence spectroscopy suggested that C4–S could induce subtle local conformational changes in CatS structure, primarily affecting Trp residues exposed to the hydrophilic environment. Furthermore, C4–S was found to decrease the rate of proCatS maturation in a dose-dependent manner. Conversely, other GAGs tested had no effect on proCatS processing rate, regardless of their concentration. Molecular docking results for C4–S indicated the presence of three potential

binding regions, one located within the active site, which corroborates *in vitro* experiments. Two others GAG-binding sites are positioned away from the catalytic cleft, one exosite located near domains that bind to elastin [158]. Moreover, it was observed that substrate binding prior to GAG binding facilitated the interaction of C4–S with the enzyme's active site. This binding scenario could explain the inhibitory mechanism of C4–S on CatS, wherein GAG binding to the substrate impedes substrate dissociation, consequently decreasing its enzymatic activity. Additionally, it was concluded that the minimum size for specific C4–S binding is a tetramer, and further GAG elongation had no significant effects on binding affinity. Similar findings were obtained for C6–S, although in this case, an alternative binding region was more preferred than one including the active site.

2.4.2. Cathepsin V

In addition to cathepsins K and S, CatV, also known as cathepsin L2, has a highly tissue-restricted expression (e.g., cornea, thymus, heart, brain, lung and skin) [121]. CatV plays a crucial role in the release of antigenic peptides and the maturation of MHC class II molecules, and participates in the degradation and turnover of insoluble elastin fibrils. Likewise, CatV exhibits the highest elastolytic activity yet described among mammalian elastase-like proteases. In addition to its diverse roles in physiological processes, CatV participates in autoimmune diseases, mostly related to its dysregulation or its "out of place" activity, supporting the hypothesis that targeting CatV may have utility as a novel and promising therapeutic target. It was also discussed that dysregulation of CatV might lead to mucopolysaccharidosis. HS, a major GAG that accumulated in mucopolysaccharidosis types I, II, and III (MPS I, –II, and –III) inhibited the enzymatic activity of CatV, in a dose-dependent manner [121] by a non-competitive mechanism. Molecular modeling studies suggested that HS binding sites are predominantly located near the active site of CatV, which may account for its inhibition [159]. Therefore, in order to define key residues involved in HS binding, structure of CatV was compared with CatL. Both enzymes share high primary sequence identity (80 %) but their electrostatic potential surfaces are essentially different [30]. Moreover, in contrast to CatL, CatV formed a complex *in vitro* with C4–S similar to that described for CatK [160,161]. C4–S bound to CatV exosites located near the domain that interacts with elastin. In addition, high concentrations of C4–S strongly inhibited the elastolytic activity of CatV. Molecular docking results indicated that HS bound to the active site of CatV (in a similar manner to what was observed for other cathepsin-GAG complexes, Fig. 7C), while for CatL such HS binding poses within enzymatic site were not identified. Analysis of key CatV amino acid residues involved in the binding of HS suggested that K20 is a critical residue for binding, as confirmed by site-directed mutagenesis study. Findings presented in this study suggest that the inhibition of elastolytic activity of CatV in MPS patients due to the increased HS levels may contribute to MPS development, particularly lung-related disorders, as a result of HS binding to the CatV active site. These insights also offer valuable clues about the potential structure of specific agents that could prevent CatV-HS complex formation and thus restore its biological activities.

3. Conclusions

Numerous reports have shed light on the biological relevance of cysteine cathepsin-GAG interactions. GAGs have been found to impact the enzymatic activity of (pro)cathepsins by binding to their active site. It was discussed that GAGs regulate maturation of proCatB and proCatS and that a particular GAG binding region might determine which mechanism is induced. *In silico* studies revealed that dodecameric C4–S, in opposite to human CatK, binds to rat CatK in a way that bridges two experimentally identified binding sites resulting in different mechanism of collagen degradation. Furthermore, computational studies showed that in many cases an increase of GAG chain length and net sulfation, favors stability of a cathepsin-GAG complex, suggesting the key role of electrostatic interactions in these molecular systems. Molecular modeling also allowed to propose potential C4–S binding regions on the surface of CatS essential for its inhibition and provided putative mechanisms for such process. Computational studies proposed that HS binding to CatV active site might inhibit its elastolytic activity, which might be important in the development of MPS disorders. The following studies demonstrate that *in silico* methods can be effectively utilized to accurately characterise protein-GAG interactions. However, despite recent progress, some questions remain unanswered. One aspect of procathepsin maturation requiring further investigation is the conformation of the proenzyme, particularly regarding the accessibility of the active site. Such investigation could elucidate whether there exists a single specific conformation or multiple conformations, thereby aiding in the identification of potential patterns in the maturation processes of different procathepsins. Another aspect of protein-GAG interactions necessitating further exploration is the role of GAGs in (pro)cathepsin enzymatic activity. In this regard, the application of Quantum Mechanics/Molecular Mechanics approaches could be applied to gain essential insights into GAG-mediated cathepsin enzymatic processes at the quantum level. The application of QM/MM methods could provide a more detailed view of the GAG-mediated procathepsin maturation process. Furthermore, this approach would allow for the study of the kinetics behind cathepsin enzymatic reactions and the effect of GAG binding on these processes. Simultaneously, there is room for the advancement of experimental approaches for characterizing protein-GAG complexes. For instance, the cryo-EM method holds promise in this regard, despite its ongoing development. Despite its successful application in studying various biological complexes, to date, no experimental structure of a protein-GAG complex has been obtained using this approach.

In conclusion, studies performed with theoretical approaches provided data on GAG binding exosites that could be targeted for development of novel and innovative therapeutic option to reduce collagenolytic and elastinolytic activities of specific cathepsins in human disease. At the same time, numerous facets of cathepsin-GAG interactions remain unexplored, leaving this field wide open for further comprehensive investigation. Better understanding of cathepsin-GAG interactions paves the way towards molecular design of novel GAG-based drugs which might be an effective compounds in treatment of diseases caused by dysregulation of cathepsins enzymatic activity.

CRediT authorship contribution statement

Krzysztof K. Bojarski: Writing – review & editing, Writing – original draft, Conceptualization. **Alexis David:** Writing – original draft. **Fabien Lecaille:** Writing – review & editing, Writing – original draft. **Sergey A. Samsonov:** Writing – review & editing, Writing – original draft.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgements

This work was supported by the National Science Center of Poland (Narodowe Centrum Nauki, grant number: UMO-2023/48/C/ST4/00163 (K. K. Bojarski) and UMO-2018/30/E/ST4/00037 (S. A. Samsonov)). This work was also supported by institutional funding from INSERM and by a grant from the French National Research Agency (Modulate, no. ANR-22-CE44-0013). A. David holds a doctoral

fellowship from Région Centre-Val de Loire (CVdL), France.

References

[1] J. D. Esko, K. Kimata, U. Lindahl, Proteoglycans and Sulfated Glycosaminoglycans, Essentials of Glycobiology. second ed.

[2] H. Habuchi, O. Habuchi, K. Kimata, Sulfation pattern in glycosaminoglycan: does it have a code? Glycoconj. J. 21 (2004) 47–52, https://doi.org/10.1023/B:GLYC.0000043747.87325.5e.

[3] A. Varki, R.D. Cummings, J.D. Esko, H.H. Freeze, P. Stanley, J.D. Marth, C. R. Bertozzi, G.W. Hart, M.E. Etzler, Symbol nomenclature for glycan representation, Proteomics 9 (24) (2009) 5398–5399, https://doi.org/10.1002/pmic.200900708.

[4] S.I. Lamberg, A.C. Stoolmiller, Glycosaminoglycans. a biochemical and clinical review, J. Invest. Dermatol. 63 (6) (1974) 433–449, https://doi.org/10.1111/1523-1747.ep12680346.

[5] J. Necas, L. Bartosikova, P. Brauner, J. Kolar, et al., Hyaluronic acid (hyaluronan): a review, Vet. Med. 53 (8) (2008) 397–411.

[6] R.D. Price, M. Berry, H.A. Navsaria, Hyaluronic acid: the scientific and clinical evidence, J. Plast. Reconstr. Aesthetic Surg. 60 (10) (2007) 1110–1119, https://doi.org/10.1016/j.bjps.2007.03.005.

[7] A.S. Maharjan, D. Pilling, R.H. Gomer, High and low molecular weight hyaluronic acid differentially regulate human fibrocyte differentiation, PLoS One 6 (10) (2011) e26078, https://doi.org/10.1371/journal.pone.0026078.

[8] C.M. McKee, M.B. Penno, M. Cowman, M.D. Burdick, R.M. Strieter, C. Bao, P. W. Noble, et al., Hyaluronan (ha) fragments induce chemokine gene expression in alveolar macrophages. the role of ha size and cd44, J. Clin. Invest. 98 (10) (1996) 2403–2413, https://doi.org/10.1172/JCI119054.

[9] B. Oertli, X. Fan, R.P. Wüthrich, Characterization of cd44-mediated hyaluronan binding by renal tubular epithelial cells, Nephrol. Dial. Transplant.: Offic. Pub. Europ. Dialysis Transplant Assoc.-Europ. Renal Assoc. 13 (2) (1998) 271–278, https://doi.org/10.1093/oxfordjournals.ndt.a027818.

[10] Y. Kawano, V. Patrulea, E. Sublet, G. Borchard, T. Iyoda, R. Kageyama, A. Morita, S. Seino, H. Yoshida, O. Jordan, et al., Wound healing promotion by hyaluronic acid: effect of molecular weight on gene expression and in vivo wound closure, Pharmaceuticals 14 (4) (2021) 301, https://doi.org/10.3390/ph14040301.

[11] W.J. Chen, Functions of Hyaluronan in Wound Repair, Hyaluronan, 2002, pp. 147–156, https://doi.org/10.1046/j.1524-475x.1999.00079.x.

[12] K. Sugahara, T. Mikami, T. Uyama, S. Mizuguchi, K. Nomura, H. Kitagawa, Recent advances in the structural biology of chondroitin sulfate and dermatan sulfate, Curr. Opin. Struct. Biol. 13 (5) (2003) 612–620, https://doi.org/10.1016/j.sbi.2003.09.011.

[13] J. Martel-Pelletier, S.K. Tat, J.-P. Pelletier, Effects of chondroitin sulfate in the pathophysiology of the osteoarthritic joint: a narrative review, Osteoarthritis Cartilage 18 (2010) S7–S11, https://doi.org/10.1016/j.joca.2010.01.015.

[14] M. Belting, Glycosaminoglycans in cancer treatment, Thromb. Res. 133 (2014) S95–S101, https://doi.org/10.1016/S0049-3848(14)50016-3.

[15] J.L. Funderburgh, Mini review keratan sulfate: structure, biosynthesis, and function, Glycobiology 10 (10) (2000) 951–958, https://doi.org/10.1093/glycob/10.10.951.

[16] V.H. Pomin, Keratan sulfate: an up-to-date review, Int. J. Biol. Macromol. 72 (2015) 282–289, https://doi.org/10.1016/j.ijbiomac.2014.08.029.

[17] R. Lever, B. Mulloy, C.P. Page, Heparin-a Century of Progress, vol. 207, Springer Science & Business Media, 2012, https://doi.org/10.1007/978-3-642-23056-1.

[18] B. Casu, U. Lindahl, Structure and biological interactions of heparin and heparan sulfate, Adv. Carbohydr. Chem. Biochem. 57 (2001) 159–206, https://doi.org/10.1016/S0065-2318(01)57017-1.

[19] D. A. Lane, L. Ulf, Heparin: Chemical and Biological Properties, clinical applications.

[20] H.G. Garg, R.J. Linhardt, C.A. Hales, Chemistry and Biology of Heparin and Heparan Sulfate, Elsevier, 2011.

[21] S. Ricard-Blum, F. Lisacek, Glycosaminoglycanomics: where we are, Glycoconj. J. 34 (2017) 339–349, https://doi.org/10.1007/s10719-016-9747-2.

[22] R. Saravanan, Isolation of low-molecular-weight heparin/heparan sulfate from marine sources, Adv. Food Nutr. Res. 72 (2014) 45–60, https://doi.org/10.1016/B978-0-12-800269-8.00003-8.

[23] J. Kreuger, D. Spillmann, J.-p. Li, U. Lindahl, Interactions between heparan sulfate and proteins: the concept of specificity, J. Cell Biol. 174 (3) (2006) 323, https://doi.org/10.1083/jcb.200604035.

[24] A. Onishi, K. St Ange, J.S. Dordick, R.J. Linhardt, Heparin and anticoagulation, Front. Biosci.-Landmark 21 (7) (2016) 1372–1392.

[25] M. Tsuda, K. Kamimura, H. Nakato, M. Archer, W. Staatz, B. Fox, M. Humphrey, S. Olson, T. Futch, V. Kaluza, et al., The cell-surface proteoglycan dally regulates wingless signalling in drosophila, Nature 400 (6741) (1999) 276–280, https://doi.org/10.1038/22336.

[26] C. Hansen, E. Otto, K. Kuhlemann, G. Förster, G. Kahaly, Glycosaminoglycans in autoimmunity, Clin. Exp. Rheumatol. 14 (1996) S59–S67.

[27] T. Ariga, T. Miyatake, R.K. Yu, Role of proteoglycans and glycosaminoglycans in the pathogenesis of alzheimer's disease and related disorders: amyloidogenesis and therapeutic strategies—a review, J. Neurosci. Res. 88 (11) (2010) 2303–2315, https://doi.org/10.1002/jnr.22393.

[28] S. Lehri-Boufala, M.-O. Ouidja, V. Barbier-Chassefière, E. Hénault, R. Raisman-Vozari, L. Garrigue-Antar, D. Papy-Garcia, C. Morin, New roles of glycosaminoglycans in α-synuclein aggregation in a cellular model of Parkinson disease, PLoS One 10 (1) (2015) e0116641, https://doi.org/10.1371/journal.pone.0116641.

[29] J.Y. Wang, M.H. Roehrl, Glycosaminoglycans are a potential cause of rheumatoid arthritis, Proc. Natl. Acad. Sci. USA 99 (22) (2002) 14362–14367, https://doi.org/10.1073/pnas.222536599.

[30] A. David, T. Chazeirat, A. Saidi, G. Lalmanach, F. Lecaille, The interplay of glycosaminoglycans and cysteine cathepsins in mucopolysaccharidosis, Biomedicines 11 (3) (2023) 810, https://doi.org/10.3390/biomedicines11030810.

[31] A. Köwitsch, G. Zhou, T. Groth, Medical application of glycosaminoglycans: a review, J. Tissue Eng. Regenerat. Med. 12 (1) (2018) e23–e41, https://doi.org/10.1002/term.2398.

[32] R.A. Scott, A. Panitch, Glycosaminoglycans in biomedicine, Wiley Interdisciplin. Rev.: Nanomed. Nanobiotechnol. 5 (4) (2013) 388–398, https://doi.org/10.1002/wnan.1223.

[33] J.R. Savage, A. Pulsipher, N.V. Rao, T.P. Kennedy, G.D. Prestwich, M.E. Ryan, W. Y. Lee, A modified glycosaminoglycan, gm-0111, inhibits molecular signaling involved in periodontitis, PLoS One 11 (6) (2016) e0157310, https://doi.org/10.1371/journal.pone.0157310.

[34] W.Y. Lee, J.R. Savage, J. Zhang, W. Jia, S. Oottamasathien, G.D. Prestwich, Prevention of anti-microbial peptide ll-37-induced apoptosis and atp release in the urinary bladder by a modified glycosaminoglycan, PLoS One 8 (10) (2013) e77854, https://doi.org/10.1371/journal.pone.0077854.

[35] J. Zhang, X. Xu, N.V. Rao, B. Argyle, L. McCoard, W.J. Rusho, T.P. Kennedy, G. D. Prestwich, G. Krueger, Novel sulfated polysaccharides disrupt cathelicidins, inhibit rage and reduce cutaneous inflammation in a mouse model of rosacea, PLoS One 6 (2) (2011) e16658, https://doi.org/10.1371/journal.pone.0016658.

[36] K.K. Bojarski, J. Becher, T. Riemer, K. Lemmnitzer, S. Möller, J. Schiller, M. Schnabelrauch, S.A. Samsonov, Synthesis and in silico characterization of artificially phosphorylated glycosaminoglycans, J. Mol. Struct. 1197 (2019) 401–416, https://doi.org/10.1016/j.molstruc.2019.07.064.

[37] B. Casu, M. Petitou, M. Provasoli, P. Sinay, Conformational flexibility: a new concept for explaining binding and biological properties of iduronic acid-containing glycosaminoglycans, Trends Biochem. Sci. 13 (6) (1988) 221–225, https://doi.org/10.1016/0968-0004(88)90088-6.

[38] B.M. Sattelle, J. Shakeri, A. Almond, Does microsecond sugar ring flexing encode 3d-shape and bioactivity in the heparanome? Biomacromolecules 14 (4) (2013) 1149–1159, https://doi.org/10.1021/bm400067g.

[39] S.A. Samsonov, M.T. Pisabarro, Importance of idoa and idoa (2s) ring conformations in computational studies of glycosaminoglycan–protein interactions, Carbohydr. Res. 381 (2013) 133–137, https://doi.org/10.1016/j.carres.2013.09.005.

[40] K.K. Bojarski, A.K. Sieradzan, S.A. Samsonov, Molecular dynamics insights into protein-glycosaminoglycan systems from microsecond-scale simulations, Biopolymers 110 (7) (2019) e23252, https://doi.org/10.1002/bip.23252.

[41] S. Khan, J. Gor, B. Mulloy, S.J. Perkins, Semi-rigid solution structures of heparin by constrained x-ray scattering modelling: new insight into heparin–protein complexes, J. Mol. Biol. 395 (3) (2010) 504–521, https://doi.org/10.1016/j.jmb.2009.10.064.

[42] F.S. Alotaibi, M.M. Alsadun, S.A. Alsaiari, K. Ramakrishnan, E.A. Yates, D. G. Fernig, Interactions of proteins with heparan sulfate, Essays Biochem. (2024) EBC20230093, https://doi.org/10.1042/EBC20230093.

[43] M. Parafioriti, S. Elli, J.C. Muñoz-García, J. Ramírez-Cárdenas, E.A. Yates, J. Angulo, M. Guerrini, Differential solvent deep-std nmr and md simulations enable the determinants of the molecular recognition of heparin oligosaccharides by antithrombin to be disentangled, Int. J. Mol. Sci. 25 (9) (2024) 4669, https://doi.org/10.3390/ijms25094669.

[44] A. Imberty, H. Lortat-Jacob, S. Pérez, Structural view of glycosaminoglycan–protein interactions, Carbohydr. Res. 342 (3–4) (2007) 430–439, https://doi.org/10.1016/j.carres.2006.12.019.

[45] B. Mulloy, The specificity of interactions between proteins and sulfated polysaccharides, An. Acad. Bras. Cienc. 77 (2005) 651–664, https://doi.org/10.1590/S0001-37652005000400007.

[46] J. Potthoff, K.K. Bojarski, G. Kohut, A.G. Lipska, A. Liwo, E. Kessler, S. Ricard-Blum, S.A. Samsonov, Analysis of procollagen c-proteinase enhancer-1/glycosaminoglycan binding sites and of the potential role of calcium ions in the interaction, Int. J. Mol. Sci. 20 (20) (2019) 5021, https://doi.org/10.3390/ijms20205021.

[47] P. Kaliannan, M.M. Gromiha, K. Ramamurthi, M. Elanthiraiyan, Solvent accessibility studies on glycosaminoglycans, Biophys. Chem. 74 (1) (1998) 11–22, https://doi.org/10.1016/S0301-4622(98)00153-7.

[48] N.E. Woodhead, W.F. Long, F.B. Williamson, Binding of zinc ions to heparin. analysis by equilibrium dialysis suggests the occurrence of two, entropy-driven, processes, Biochem. J. 237 (1) (1986) 281, https://doi.org/10.1042/bj2370281.

[49] I. Stevic, N. Parmar, N. Paredes, L.R. Berry, A.K. Chan, Binding of heparin to metals, Cell Biochem. Biophys. 59 (2011) 171–178, https://doi.org/10.1007/s12013-010-9129-5.

[50] G. Multhaup, A.I. Bush, P. Pollwein, C.L. Masters, Interaction between the zinc (ii) and the heparin binding site of the alzheimer's disease βa4 amyloid precursor protein (app), FEBS Lett. 355 (2) (1994) 151–154, https://doi.org/10.1016/0014-5793(94)01176-1.

[51] R. Eckert, H. Ragg, Zinc ions promote the interaction between heparin and heparin cofactor ii, FEBS Lett. 541 (1–3) (2003) 121–125, https://doi.org/10.1016/S0014-5793(03)00322-3.

[52] Q. Han, Y. Fu, H. Zhou, Y. He, Y. Luo, Contributions of zn (ii)-binding to the structural stability of endostatin, FEBS Lett. 581 (16) (2007) 3027–3032, https://doi.org/10.1016/j.febslet.2007.05.058.
[53] F. Zhang, X. Liang, J. M. Beaudet, Y. Lee, R. J. Linhardt, The effects of metal ions on heparin/heparin sulfate-protein interactions, J. Biomed. Tech. Res. 1 (1). doi: 10.19104/jbtr.2014.101.
[54] U. Uciechowska-Kaczmarzyk, S. Babik, F. Zsila, K.K. Bojarski, T. Beke-Somfai, S. A. Samsonov, Molecular dynamics-based model of vegf-a and its heparin interactions, J. Mol. Graph. Model. 82 (2018) 157–166, https://doi.org/10.1016/j.jmgm.2018.04.015.
[55] V. Lutsyk, W. Plazinski, Exploring ring conformation in uronate monosaccharides: insights from ab initio calculations and classical molecular dynamics simulations, J. Phys. Chem. B 128 (2) (2024) 472–491, https://doi.org/10.1021/acs.jpcb.3c06556.
[56] M.M. Kogut, A. Danielsson, S. Ricard-Blum, S.A. Samsonov, Impact of calcium ions on the structural and dynamic properties of heparin oligosaccharides by computational analysis, Comput. Biol. Chem. 99 (2022) 107727, https://doi.org/10.1016/j.compbiolchem.2022.107727.
[57] C.I. Gama, S.E. Tully, N. Sotogaku, P.M. Clark, M. Rawat, N. Vaidehi, W. A. Goddard III, A. Nishi, L.C. Hsieh-Wilson, Sulfation patterns of glycosaminoglycans encode molecular recognition and activity, Nat. Chem. Biol. 2 (9) (2006) 467–473, https://doi.org/10.1038/nchembio810.
[58] K. Atkovska, S.A. Samsonov, M. Paszkowski-Rogacz, M.T. Pisabarro, Multipose binding in molecular docking, Int. J. Mol. Sci. 15 (2) (2014) 2622–2645, https://doi.org/10.3390/ijms15022622.
[59] P.R.B. Joseph, P.D. Mosier, U.R. Desai, K. Rajarathnam, Solution nmr characterization of chemokine cxcl8/il-8 monomer and dimer binding to glycosaminoglycans: structural plasticity mediates differential binding interactions, Biochem. J. 472 (1) (2015) 121–133, https://doi.org/10.1042/BJ20150059.
[60] T. Hofmann, S.A. Samsonov, A. Pichert, K. Lemmnitzer, J. Schiller, D. Huster, M. T. Pisabarro, M. von Bergen, S. Kalkhof, Structural analysis of the interleukin-8/glycosaminoglycan interactions by amide hydrogen/deuterium exchange mass spectrometry, Methods 89 (2015) 45–53, https://doi.org/10.1016/j.ymeth.2015.02.011.
[61] J. Gallagher, Fell–muir lecture: heparan sulphate and the art of cell regulation: a polymer chain conducts the protein orchestra, Int. J. Exp. Pathol. 96 (4) (2015) 203–231, https://doi.org/10.1111/iep.12135.
[62] M. Parafioriti, M. Ni, M. Petitou, C.J. Mycroft-West, T.R. Rudd, N.S. Gandhi, V. Ferro, J.E. Turnbull, M.A. Lima, M.A. Skidmore, et al., Evidence for multiple binding modes in the initial contact between sars-cov-2 spike s1 protein and cell surface glycans, Chem.–Eur. J. 29 (1) (2023) e202202599, https://doi.org/10.1002/chem.202202599.
[63] K.K. Bojarski, S.A. Samsonov, Role of oligosaccharide chain polarity in protein–glycosaminoglycan interactions, J. Chem. Inf. Model. 61 (1) (2020) 455–466, https://doi.org/10.1021/acs.jcim.0c01402.
[64] A. Almond, Multiscale modeling of glycosaminoglycan structure and dynamics: current methods and challenges, Curr. Opin. Struct. Biol. 50 (2018) 58–64, https://doi.org/10.1016/j.sbi.2017.11.008.
[65] N.V. Sankaranarayanan, B. Nagarajan, U.R. Desai, So you think computational approaches to understanding glycosaminoglycan–protein interactions are too dry and too rigid? think again, Curr. Opin. Struct. Biol. 50 (2018) 91–100, https://doi.org/10.1016/j.sbi.2017.12.004.
[66] B. Nagarajan, N.V. Sankaranarayanan, U.R. Desai, Molecular dynamics simulations of glycosaminoglycan oligosaccharide using newer force fields, bioRxiv (2019) 561969, https://doi.org/10.1101/561969.
[67] S. Pérez, F. Bonnardel, F. Lisacek, A. Imberty, S. Ricard Blum, O. Makshakova, Gag-db, the new interface of the three-dimensional landscape of glycosaminoglycans, Biomolecules 10 (12) (2020) 1660, https://doi.org/10.3390/biom10121660.
[68] S. Perez, O. Makshakova, J. Angulo, E. Bedini, A. Bisio, J.L. de Paz, E. Fadda, M. Guerrini, M. Hricovini, M. Hricovini, et al., Glycosaminoglycans: what remains to be deciphered? JACS Au 3 (3) (2023) 628–656, https://doi.org/10.1021/jacsau.2c00569.
[69] D. Case, D. Cerutti, T. Cheatham III, T. Darden, R. Duke, T. Giese, H. Gohlke, A. Goetz, D. Greene, N. Homeyer, et al., Amber17. 2017, San Fr. Univ. Calif.
[70] W. Im, D. Beglov, B. Roux, Continuum solvation model: computation of electrostatic forces from numerical solutions to the Poisson-Boltzmann equation, Comput. Phys. Commun. 111 (1–3) (1998) 59–75, https://doi.org/10.1016/S0010-4655(98)00016-2.
[71] S.A. Samsonov, M.T. Pisabarro, Computational analysis of interactions in structurally available protein–glycosaminoglycan complexes, Glycobiology 26 (8) (2016) 850–861, https://doi.org/10.1093/glycob/cww055.
[72] G.M. Morris, D.S. Goodsell, R.S. Halliday, R. Huey, W.E. Hart, R.K. Belew, A. J. Olson, Automated docking using a lamarckian genetic algorithm and an empirical binding free energy function, J. Comput. Chem. 19 (14) (1998) 1639–1662, https://doi.org/10.1002/(SICI)1096-987X(19981115)19:14<1639::AID-JCC10>3.0.CO;2-B.
[73] U. Uciechowska-Kaczmarzyk, I.C. de Beauchene, S.A. Samsonov, Docking software performance in protein-glycosaminoglycan systems, J. Mol. Graph. Model. 90 (2019) 42–50, https://doi.org/10.1016/j.jmgm.2019.04.001.
[74] M. Suan Li, B. Khanh Mai, Steered molecular dynamics-a promising tool for drug design, Curr. Bioinf. 7 (4) (2012) 342–351, https://doi.org/10.2174/157489312803901009.
[75] S.A. Samsonov, J.-P. Gehrcke, M.T. Pisabarro, Flexibility and explicit solvent in molecular-dynamics-based docking of protein–glycosaminoglycan systems, J. Chem. Inf. Model. 54 (2) (2014) 582–592, https://doi.org/10.1021/ci4006047.
[76] T. Siebenmorgen, M. Zacharias, Efficient refinement and free energy scoring of predicted protein–protein complexes using replica exchange with repulsive scaling, J. Chem. Inf. Model. 60 (11) (2020) 5552–5562, https://doi.org/10.1021/acs.jcim.0c00853.
[77] M. Karplus, J.A. McCammon, The dynamics of proteins, Sci. Am. 254 (4) (1986) 42–51.
[78] S.A. Hollingsworth, R.O. Dror, Molecular dynamics simulation for all, Neuron 99 (6) (2018) 1129–1143, https://doi.org/10.1016/j.neuron.2018.08.011.
[79] S.Y. Joshi, S.A. Deshmukh, A review of advancements in coarse-grained molecular dynamics simulations, Mol. Simulat. 47 (10–11) (2021) 786–803, https://doi.org/10.1080/08927022.2020.1828583.
[80] K.N. Kirschner, A.B. Yongye, S.M. Tschampel, J. González-Outeiriño, C. R. Daniels, B.L. Foley, R.J. Woods, Glycam06: a generalizable biomolecular force field. carbohydrates, J. Comput. Chem. 29 (4) (2008) 622–655, https://doi.org/10.1002/jcc.20820.
[81] A. Pichert, S.A. Samsonov, S. Theisgen, L. Thomas, L. Baumann, J. Schiller, A. G. Beck-Sickinger, D. Huster, M.T. Pisabarro, Characterization of the interaction of interleukin-8 with hyaluronan, chondroitin sulfate, dermatan sulfate and their sulfated derivatives by spectroscopy and molecular modeling, Glycobiology 22 (1) (2012) 134–145, https://doi.org/10.1093/glycob/cwr120.
[82] J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B.L. De Groot, H. Grubmüller, A.D. MacKerell Jr., Charmm36m: an improved force field for folded and intrinsically disordered proteins, Nat. Methods 14 (1) (2017) 71–73, https://doi.org/10.1038/nmeth.4067.
[83] B.R. Miller III, T.D. McGee Jr., J.M. Swails, N. Homeyer, H. Gohlke, A.E. Roitberg, Mmpbsa. py: an efficient program for end-state free energy calculations, J. Chem. Theor. Comput. 8 (9) (2012) 3314–3321, https://doi.org/10.1021/ct300418h.
[84] S. Genheden, U. Ryde, The mm/pbsa and mm/gbsa methods to estimate ligand-binding affinities, Expet Opin. Drug Discov. 10 (5) (2015) 449–461, https://doi.org/10.1517/17460441.2015.1032936.
[85] K.K. Bojarski, A.S. Karczyńska, S.A. Samsonov, Role of glycosaminoglycans in procathepsin b maturation: molecular mechanism elucidated by a computational study, J. Chem. Inf. Model. 60 (4) (2020) 2247–2256, https://doi.org/10.1021/acs.jcim.0c00023.
[86] K.K. Bojarski, S.A. Samsonov, In silico insights into procathepsin s maturation mediated by glycosaminoglycans, J. Mol. Graph. Model. (2023) 108406, https://doi.org/10.1016/j.jmgm.2023.108406.
[87] G. Paiardi, S. Richter, P. Oreste, C. Urbinati, M. Rusnati, R. C. Wade, The binding of heparin to spike glycoprotein inhibits sars-cov-2 infection by three mechanisms, J. Biol. Chem. 298 (2). doi:10.1016/j.jbc.2021.101507.
[88] C.J. Mycroft-West, S. Abdelkarim, H.M. Duyvesteyn, N.S. Gandhi, M.A. Skidmore, R.J. Owens, L. Wu, Structural and mechanistic characterization of bifunctional heparan sulfate n-deacetylase-n-sulfotransferase 1, Nat. Commun. 15 (1) (2024) 1326, https://doi.org/10.1038/s41467-024-45419-4.
[89] S. Elli, E. Stancanelli, Z. Wang, M. Petitou, J. Liu, M. Guerrini, Degeneracy of the antithrombin binding sequence in heparin: 2-o-sulfated iduronic acid can replace the critical glucuronic acid, Chem.–Eur. J. 26 (51) (2020) 11814–11818, https://doi.org/10.1002/chem.202001346.
[90] A. Denardo, S. Elli, S. Federici, M. Asperti, M. Gryzik, P. Ruzzenenti, F. Carmona, P. Bergese, A. Naggi, P. Arosio, et al., Bmp6 binding to heparin and heparan sulfate is mediated by n-terminal and c-terminal clustered basic residues, Biochim. Biophys. Acta Gen. Subj. 1865 (2) (2021) 129799, https://doi.org/10.1016/j.bbagen.2020.129799.
[91] A.E. Proudfoot, Chemokines and glycosaminoglycans, Front. Immunol. 6 (2015) 145683, https://doi.org/10.3389/fimmu.2015.00246.
[92] F. Delacoux, A. Fichard, C. Geourjon, R. Garrone, F. Ruggiero, Molecular features of the collagen v heparin binding site, J. Biol. Chem. 273 (24) (1998) 15069–15076, https://doi.org/10.1074/jbc.273.24.15069.
[93] D. Turk, G. Gunčar, M. Podobnik, B. Turk, Revised definition of substrate binding sites of papain-like cysteine proteases, Biol. Chem. 379 (1998) 137–147, https://doi.org/10.1515/bchm.1998.379.2.137.
[94] C. Chwieralski, T. Welte, F. Bühling, Cathepsin-regulated apoptosis, Apoptosis 11 (2006) 143–149, https://doi.org/10.1007/s10495-006-3486-y.
[95] T. Yadati, T. Houben, A. Bitorina, R. Shiri-Sverdlov, The ins and outs of cathepsins: physiological function and role in disease management, Cells 9 (7) (2020) 1679, https://doi.org/10.3390/cells9071679.
[96] V. Turk, V. Stoka, O. Vasiljeva, M. Renko, T. Sun, B. Turk, D. Turk, Cysteine cathepsins: from structure, function and regulation to new frontiers, Biochim. Biophys. Acta Protein Proteonomics 1824 (1) (2012) 68–88, https://doi.org/10.1016/j.bbapap.2011.10.002.
[97] O. Vasiljeva, T. Reinheckel, C. Peters, D. Turk, V. Turk, B. Turk, Emerging roles of cysteine cathepsins in disease and their potential as drug targets, Curr. Pharmaceut. Des. 13 (4) (2007) 387–403, https://doi.org/10.2174/138161207780162962.
[98] M. Fonović, B. Turk, Cysteine cathepsins and extracellular matrix degradation, Biochimica et Biophysica Acta (BBA)-general subjects 1840 (8) (2014) 2560–2570, https://doi.org/10.1016/j.bbagen.2014.03.017.
[99] M. Vizovišek, M. Fonović, B. Turk, Cysteine cathepsins in extracellular matrix remodeling: extracellular matrix degradation and beyond, Matrix Biol. 75 (2019) 141–159, https://doi.org/10.1016/j.matbio.2018.01.024.
[100] E. Vidak, U. Javoršek, M. Vizovišek, B. Turk, Cysteine cathepsins and their extracellular roles: shaping the microenvironment, Cells 8 (3) (2019) 264, https://doi.org/10.3390/cells8030264.

[101] H. Kirschke, B. Wiederanders, D. Brömme, A. Rinne, Cathepsin s from bovine spleen. purification, distribution, intracellular localization and action on proteins, Biochem. J. 264 (2) (1989) 467–473, https://doi.org/10.1042/bj2640467.
[102] P.C. Almeida, I.L. Nantes, J.R. Chagas, C.C. Rizzi, A. Faljoni-Alario, E. Carmona, L. Juliano, H.B. Nader, I.L. Tersariol, Cathepsin b activity regulation: heparin-like glycosaminoglycans protect human cathepsin b from alkaline ph-induced inactivation, J. Biol. Chem. 276 (2) (2001) 944–951, https://doi.org/10.1074/jbc.M003820200.
[103] Z. Li, Y. Yasuda, W. Li, M. Bogyo, N. Katz, R.E. Gordon, G.B. Fields, D. Brömme, Regulation of collagenase activities of human cathepsins by glycosaminoglycans, J. Biol. Chem. 279 (7) (2004) 5470–5479, https://doi.org/10.1074/jbc.M310349200.
[104] E. Carmona, É. Dufour, C. Plouffe, S. Takebe, P. Mason, J.S. Mort, R. Ménard, Potency and selectivity of the cathepsin l propeptide as an inhibitor of cysteine proteases, Biochemistry 35 (25) (1996) 8149–8157, https://doi.org/10.1021/bi952736s.
[105] J. Guay, J.-P. Falgueyret, A. Ducret, M.D. Percival, J.A. Mancini, Potency and selectivity of inhibition of cathepsin k, l and s by their respective propeptides, Eur. J. Biochem. 267 (20) (2000) 6311–6318, https://doi.org/10.1046/j.1432-1327.2000.01730.x.
[106] K. Schilling, A. Körner, S. Sehmisch, A. Kreusch, R. Kleint, Y. Benedix, A. Schlabrakowski, B. Wiederanders, Selectivity of propeptide-enzyme interaction in cathepsin l-like cysteine proteases, Biol. Chem. 390 (2) (2009) 167–174, https://doi.org/10.1515/BC.2009.023.
[107] R.E. Burden, P. Snoddy, C.A. Jefferies, B. Walker, C.J. Scott, Inhibition of cathepsin l-like proteases by cathepsin v propeptide, Biol. Chem. 388 (5) (2007) 541–545, https://doi.org/10.1515/BC.2007.053.
[108] M. Horn, L. Dolečková-Marešová, L. Rulíšek, M. Máša, O. Vasiljeva, B. Turk, T. Gan-Erdene, M. Baudyš, M. Mareš, Activation processing of cathepsin h impairs recognition by its propeptide, Biol. Chem. 386 (9) (2005) 941–947, https://doi.org/10.1515/BC.2005.109.
[109] G. Maubach, K. Schilling, W. Rommerskirch, I. Wenz, J.E. Schultz, E. Weber, B. Wiederanders, The inhibition of cathepsin s by its propeptide—specificity and mechanism of action, Eur. J. Biochem. 250 (3) (1997) 745–750, https://doi.org/10.1111/j.1432-1033.1997.00745.x.
[110] C.J. Billington, P. Mason, M.-C. Magny, J.S. Mort, The slow-binding inhibition of cathepsin k by its propeptide, Biochem. Biophys. Res. Commun. 276 (3) (2000) 924–929, https://doi.org/10.1006/bbrc.2000.3553.
[111] M. Novinec, B. Lenarčič, B. Turk, Cysteine cathepsin activity regulation by glycosaminoglycans, BioMed Res. Int. (2014), https://doi.org/10.1155/2014/309718.
[112] P.C. Almeida, I.L. Nantes, C.C. Rizzi, W.A. Judice, J.R. Chagas, L. Juliano, H. B. Nader, I.L. Tersariol, Cysteine proteinase activity regulation: a possible role of heparin and heparin-like glycosaminoglycans, J. Biol. Chem. 274 (43) (1999) 30433–30438, https://doi.org/10.1074/jbc.274.43.30433.
[113] E. Del Nery, M.A. Juliano, A.P.C. Lima, J. Scharfstein, L. Juliano, Kininogenase activity by the major cysteinyl proteinase (cruzipain) from trypanosoma cruzi, J. Biol. Chem. 272 (41) (1997) 25713–25718, https://doi.org/10.1074/jbc.272.41.25713.
[114] C. Serveau, A. Boulangé, F. Lecaille, F. Gauthier, E. Authié, G. Lalmanach, Procongopain from trypanosoma congolense is processed at basic ph: an unusual feature among cathepsin l-like cysteine proteases 384 (6) (2003) 921–927, https://doi.org/10.1515/BC.2003.103.
[115] W.A. Judice, M.A. Manfredi, G.P. Souza, T.M. Sansevero, P.C. Almeida, C. S. Shida, T.F. Gesteira, L. Juliano, G.D. Westrop, S.J. Sanderson, et al., Heparin modulates the endopeptidase activity of leishmania mexicana cysteine protease cathepsin l-like rcpb2. 8, PLoS One 8 (11) (2013) e80153, https://doi.org/10.1371/journal.pone.0080153.
[116] J.P. Dalton, J. Dvorak, Activating the cathepsin b1 of a parasite: a major route with alternative pathways? Structure 22 (12) (2014) 1696–1698, https://doi.org/10.1016/j.str.2014.11.003.
[117] T.D. Rachner, S. Khosla, L.C. Hofbauer, Osteoporosis: now and the future, Lancet 377 (9773) (2011) 1276–1287, https://doi.org/10.1016/S0140-6736(10)62349-5.
[118] Y. Hashimoto, H. Kakegawa, Y. Narita, Y. Hachiya, T. Hayakawa, J. Kos, V. Turk, N. Katunuma, Significance of cathepsin b accumulation in synovial fluid of rheumatoid arthritis, Biochem. Biophys. Res. Commun. 283 (2) (2001) 334–339, https://doi.org/10.1006/bbrc.2001.4787.
[119] W. Wendt, X.-R. Zhu, H. Lübbert, C.C. Stichel, Differential expression of cathepsin x in aging and pathological central nervous system of mice, Exp. Neurol. 204 (2) (2007) 525–540, https://doi.org/10.1016/j.expneurol.2007.01.007.
[120] I. Santamaría, G. Velasco, M. Cazorla, A. Fueyo, E. Campo, C. López-Otín, Cathepsin l2, a novel human cysteine proteinase produced by breast and colorectal carcinomas, Cancer Res. 58 (8) (1998) 1624–1630.
[121] F. Lecaille, T. Chazeirat, A. Saidi, G. Lalmanach, Cathepsin v: molecular characteristics and significance in health and disease, Mol. Aspect. Med. 88 (2022) 101086, https://doi.org/10.1016/j.mam.2022.101086.
[122] K. Bojarski, J. Sage, G. Lalmanach, F. Lecaille, S. Samsonov, In silico and in vitro mapping of specificity patterns of glycosaminoglycans towards cysteine cathepsins b, l, k, s and v, J. Mol. Graph. Model. 113 (2022) 108153, https://doi.org/10.1016/j.jmgm.2022.108153.
[123] F. Lecaille, T. Chazeirat, K.K. Bojarski, J. Renault, A. Saidi, V.G.N. Prasad, S. Samsonov, G. Lalmanach, Rat cathepsin k: enzymatic specificity and regulation of its collagenolytic activity, Biochim. Biophys. Acta Protein Proteonomics 1868 (2) (2020) 140318, https://doi.org/10.1016/j.bbapap.2019.140318.
[124] V. Turk, B. Turk, D. Turk, Lysosomal cysteine proteases: facts and opportunities, EMBO J. 20 (17) (2001) 4629–4633, https://doi.org/10.1093/emboj/20.23.6570.
[125] K. Ishidoh, E. Kominami, Processing and activation of lysosomal proteinases. Biological Chemistrydoi:10.1515/BC, 2002, p. 206.
[126] F. Lecaille, J. Kaleta, D. Brömme, Human and parasitic papain-like cysteine proteases: their role in physiology and pathology and recent developments in inhibitor design, Chem. Rev. 102 (12) (2002) 4459–4488, https://doi.org/10.1021/cr0101656.
[127] A.D. Cardin, H. Weintraub, Molecular modeling of protein-glycosaminoglycan interactions, Arteriosclerosis: Offic. J. Am. Heart Assoc. Inc. 9 (1) (1989) 21–32, https://doi.org/10.1161/01.ATV.9.1.21.
[128] M. Fairhead, S.M. Kelly, C.F. van der Walle, A heparin binding motif on the pro-domain of human procathepsin l mediates zymogen destabilization and activation, Biochem. Biophys. Res. Commun. 366 (3) (2008) 862–867, https://doi.org/10.1016/j.bbrc.2007.12.062.
[129] F. Sievers, A. Wilm, D. Dineen, T.J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding, et al., Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega, Mol. Syst. Biol. 7 (1) (2011) 539, https://doi.org/10.1038/msb.2011.75.
[130] M. Novinec, Computational investigation of conformational variability and allostery in cathepsin k and other related peptidases, PLoS One 12 (8) (2017) e0182387, https://doi.org/10.1371/journal.pone.0182387.
[131] M.G. Costa, P.R. Batista, C.S. Shida, C.H. Robert, P.M. Bisch, P.G. Pascutti, How does heparin prevent the ph inactivation of cathepsin b? allosteric mechanism elucidated by docking and molecular dynamics, BMC Genom. 11 (2010) 1–15, https://doi.org/10.1186/1471-2164-11-S5-S5. Springer.
[132] D. Caglic, J.R. Pungercar, G. Pejler, V. Turk, B. Turk, Glycosaminoglycans facilitate procathepsin b activation through disruption of propeptide-mature enzyme interactions, J. Biol. Chem. 282 (45) (2007) 33076–33085, https://doi.org/10.1074/jbc.M705761200.
[133] P.A. Lemaire, L. Huang, Y. Zhuo, J. Lu, C. Bahnck, S.J. Stachel, S.S. Carroll, L. T. Duong, Chondroitin sulfate promotes activation of cathepsin k, J. Biol. Chem. 289 (31) (2014) 21562–21572, https://doi.org/10.1074/jbc.M114.559898.
[134] D. Brömme, F.S. Nallaseth, B. Turk, Production and activation of recombinant papain-like cysteine proteases, Methods 32 (2) (2004) 199–206, https://doi.org/10.1016/s1046-2023(03)00212-3.
[135] W. Humphrey, A. Dalke, K. Schulten, Vmd: visual molecular dynamics, J. Mol. Graph. 14 (1) (1996) 33–38, https://doi.org/10.1016/0263-7855(96)00018-5.
[136] B.R. Troen, The role of cathepsin k in normal bone resorption, Drug News Perspect. 17 (1) (2004) 19–28, https://doi.org/10.1358/dnp.2004.17.1.829022.
[137] Z. Li, M. Kienetz, M.M. Cherney, M.N. James, D. Brömme, The crystal and molecular structures of a cathepsin k: chondroitin sulfate complex, J. Mol. Biol. 383 (1) (2008) 78–91, https://doi.org/10.1016/j.jmb.2008.07.038.
[138] A.H. Aguda, P. Panwar, X. Du, N.T. Nguyen, G.D. Brayer, D. Brömme, Structural basis of collagen fiber degradation by cathepsin k, Proc. Natl. Acad. Sci. USA 111 (49) (2014) 17474–17479, https://doi.org/10.1073/pnas.1414126111.
[139] B.D. Gelb, G.-P. Shi, H.A. Chapman, R.J. Desnick, Pycnodysostosis, a lysosomal disease caused by cathepsin k deficiency, Science 273 (5279) (1996) 1236–1238, https://doi.org/10.1126/science.273.5279.1236.
[140] L.T. Duong, A.T. Leung, B. Langdahl, Cathepsin k inhibition: a new mechanism for the treatment of osteoporosis, Calcif. Tissue Int. 98 (2016) 381–397, https://doi.org/10.1007/s00223-015-0051-0.
[141] D. Qian, L. He, Q. Zhang, W. Li, D. Tang, C. Wu, F. Yang, K. Li, H. Zhang, Cathepsin k: a versatile potential biomarker and therapeutic target for various cancers, Curr. Oncol. 29 (8) (2022) 5963–5987, https://doi.org/10.3390/curroncol29080471.
[142] F. Bühling, C. Röcken, F. Brasch, R. Hartig, Y. Yasuda, P. Saftig, D. Brömme, T. Welte, Pivotal role of cathepsin k in lung fibrosis, Am. J. Pathol. 164 (6) (2004) 2203–2216, https://doi.org/10.1016/S0002-9440(10)63777-7.
[143] S. Marchand-Adam, M. Pronost, A. Saidi, F. Lecaille, G. Lalmanach, Cathepsin k: both a likely biomarker and a new therapeutic target in lymphangioleiomyomatosis? Rare Disease and Orphan Drugs J. 2 (2023) 3, https://doi.org/10.20517/rdodj.2022.24.
[144] F. Lecaille, D. Brömme, G. Lalmanach, Biochemical properties and regulation of cathepsin k activity, Biochimie 90 (2) (2008) 208–226, https://doi.org/10.1016/j.biochi.2007.08.011.
[145] P. Panwar, K. Søe, R.V. Guido, R.V. Bueno, J.-M. Delaisse, D. Brömme, A novel approach to inhibit bone resorption: exosite inhibitors against cathepsin k, Br. J. Pharmacol. 173 (2) (2016) 396–410, https://doi.org/10.1111/bph.13383.
[146] S. Ma, L.S. Devi-Kesavan, J. Gao, Molecular dynamics simulations of the catalytic pathway of a cysteine protease: a combined qm/mm study of human cathepsin k, J. Am. Chem. Soc. 129 (44) (2007) 13633–13645, https://doi.org/10.1021/ja074222+.
[147] M. Maszota-Zieleniak, M. Marcisz, M.M. Kogut, T. Siebenmorgen, M. Zacharias, S. A. Samsonov, Evaluation of replica exchange with repulsive scaling approach for docking glycosaminoglycans, J. Comput. Chem. 42 (15) (2021) 1040–1053, https://doi.org/10.1002/jcc.26496.
[148] M. Marcisz, S. Anila, M. Gaardløs, M. Zacharias, S.A. Samsonov, Studying specificity in protein–glycosaminoglycan recognition with umbrella sampling, Beilstein J. Org. Chem. 19 (1) (2023) 1933–1946, https://doi.org/10.3762/bjoc.19.144.
[149] H.A. Chapman, R.J. Riese, G.-P. Shi, Emerging roles for cysteine proteases in human biology, Annu. Rev. Physiol. 59 (1) (1997) 63–88, https://doi.org/10.1146/annurev.physiol.59.1.63.

[150] G.-P. Shi, J.A. Villadangos, G. Dranoff, C. Small, L. Gu, K.J. Haley, R. Riese, H. L. Ploegh, H.A. Chapman, Cathepsin s required for normal mhc class ii peptide loading and germinal center development, Immunity 10 (2) (1999) 197–206, https://doi.org/10.1016/S1074-7613(00)80020-5.

[151] J. Sage, E. Leblanc-Noblesse, C. Nizard, T. Sasaki, S. Schnebert, E. Perrier, R. Kurfurst, D. Brömme, G. Lalmanach, F. Lecaille, Cleavage of nidogen-1 by cathepsin s impairs its binding to basement membrane partners, PLoS Onedoi: 10.1371/journal.pone.0043494.

[152] J.S. Ainscough, T. Macleod, D. McGonagle, R. Brakefield, J.M. Baron, A. Alase, M. Wittmann, M. Stacey, Cathepsin s is the major activator of the psoriasis-associated proinflammatory cytokine il-36γ, Proc. Natl. Acad. Sci. USA 114 (13) (2017) E2748–E2757, https://doi.org/10.1073/pnas.1620954114.

[153] N. Cimerman, P.M. Brguljan, M. Krašovec, S. Šuškovič, J. Kos, Circadian and concentration profile of cathepsin s in sera from healthy subjects and asthmatic patients, Pflügers Archiv 442 (2001) r204–r206, https://doi.org/10.1007/s004240100026.

[154] W.-H. Boehncke, E. Weber, H. Schmid, G. Schwarz, M. Braun, C.J. Schröter, T. Burster, T. Flad, D. Dressel, H. Kalbacher, Cathepsin s activity is detectable in human keratinocytes and is selectively upregulated upon stimulation with interferon-γ, J. Invest. Dermatol. 119 (1) (2002) 44–49, https://doi.org/10.1046/j.1523-1747.2002.01800.x.

[155] A. Schönefuß, W. Wendt, B. Schattling, R. Schulten, K. Hoffmann, M. Stuecker, C. Tigges, H. Lübbert, C. Stichel, Upregulation of cathepsin s in psoriatic keratinocytes, Exp. Dermatol. 19 (8) (2010) e80–e88, https://doi.org/10.1111/j.1600-0625.2009.00990.x.

[156] K. Chung, T. Pitcher, A.D. Grant, E. Hewitt, E. Lindstrom, M. Malcangio, Cathepsin s acts via protease-activated receptor 2 to activate sensory neurons and induce itch-like behaviour, Neurobiolo. Pain 6 (2019) 100032, https://doi.org/10.1016/j.ynpai.2019.100032.

[157] J. Sage, F. Mallévre, F. Barbarin-Costes, S.A. Samsonov, J.-P. Gehrcke, M. T. Pisabarro, E. Perrier, S. Schnebert, A. Roget, T. Livache, et al., Binding of chondroitin 4-sulfate to cathepsin s regulates its enzymatic activity, Biochemistry 52 (37) (2013) 6487–6498, https://doi.org/10.1021/bi400925g.

[158] P.-M. Andrault, P. Panwar, D. Brömme, Characterization of cathepsin s exosites that govern its elastolytic activity, Biochem. J. 477 (1) (2020) 227–242, https://doi.org/10.1042/BCJ20190847.

[159] T. Chazeirat, S. Denamur, K.K. Bojarski, P.-M. Andrault, D. Sizaret, F. Zhang, A. Saidi, M. Tardieu, R.J. Linhardt, F. Labarthe, et al., The abnormal accumulation of heparan sulfate in patients with mucopolysaccharidosis prevents the elastolytic activity of cathepsin v, Carbohydr. Polym. 253 (2021) 117261, https://doi.org/10.1016/j.carbpol.2020.117261.

[160] Z. Li, W.-S. Hou, C.R. Escalante-Torres, B.D. Gelb, D. Brömme, Collagenase activity of cathepsin k depends on complex formation with chondroitin sulfate, J. Biol. Chem. 277 (32) (2002) 28669–28676, https://doi.org/10.1074/jbc.M204004200.

[161] Y. Yasuda, Z. Li, D. Greenbaum, M. Bogyo, E. Weber, D. Brömme, Cathepsin v, a novel and potent elastolytic activity expressed in activated macrophages, J. Biol. Chem. 279 (35) (2004) 36761–36770, https://doi.org/10.1074/jbc.M403986200.