Informatics in Medicine Unlocked 11 (2018) 28–35

Contents lists available at ScienceDirect

Informatics in Medicine Unlocked

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

Computational analysis of sequential and structural variations in stromelysins as an insight towards matrix metalloproteinase research

Beutline Malgija [a], Host Antony David Rajendran [a], Uma Maheswari [a], Nivetha Sarah Ebenezer [a], Joyce Priyakumari [a,*], Shanmughavel Piramanayagam [b]

[a] Bioinformatics Infrastructure Facility for BTISnet, Madras Christian College, Chennai, Tamil Nadu, 600059, India
[b] Computational Biology Lab, Department of Bioinformatics, Bharathiar University, Coimbatore, Tamil Nadu, 641046, India

ARTICLE INFO

Keywords:
Matrix metalloproteinases
Zinc-binding motif
Structure prediction
Ramachandran plot
Stromelysin
Superfamily

ABSTRACT

Matrix metalloproteinases are zinc-dependent protein and peptide hydrolases. They are broadly involved in metabolic regulation through both extensive protein degradation and selective peptide-bond hydrolysis. Stromelysins belong to this group of proteinases and involved in various physiological and pathological functioning of the cell. This study aims at assessing the sequential and structural aspects of stromelysins based on *in silico* approaches. Deduced stromelysin sequences were predicted to possess regulatory domain, protease domain, and proline-rich hinge regions. Sequential analysis revealed MMP-3 and 10 are more similar than MMP-11 regarding stability and aminoacid distribution. Secondary structure prediction showed that beta-sheets dominated other secondary structural elements (alpha helices, coils, and turns) in stromelysins. Validation of predicted models with different approaches confirms the accuracy and best quality of models. The binding mode of zinc atom provides information regarding their interaction with stromelysins. The predicted models showed little variation in binding mode with their natural inhibitor, TIMP-1. The predicted models will be used in an extensive range of studies for functional analysis and improvement activity of stromelysins.

1. Introduction

Matrix metalloproteinases (MMPs), also called matrixins are zinc-dependent endopeptidases capable of degrading most of the extracellular matrix (ECM) components [1] such as proteoglycans, insoluble collagen fibers and soluble ECM proteins (fibronectin, laminin etc), involved in various physiological and pathological conditions. About 23 types of MMPs have been identified in humans [2,3], and many are reported to have their role in cancer [4,5]. MMPs also act in cardiovascular remodeling and have a distinct spatial and temporal role that ensures normal physiology of the heart [6] and vasculature [7]. They can be categorized into collagenases, gelatinases, stromelysins, matrilysins, metallo-elastases and membrane type MMPs (MT-MMPs) based on the structure and substrate specificity. MMPs in various categories are shown in Supplementary Fig. S1.

Though all the MMPs have their role in various physiological and pathological conditions, several studies have reported the involvement of stromelysins in various conditions like cancer [8,9] and cardiovascular diseases [10,11]. Stromelysin is composed of MMP-3, MMP-10, and MMP-11 corresponding to stromelysin 1, 2 and 3 respectively. They have relatively broad substrate specificity. The majority of stromelysins cleave non-collagenous ECM proteins like proteoglycans, glycoproteins, fibronectin, and laminin [12]. MMP-3 and MMP-10 have an identical spectrum of activity, but MMP-3 is more potent [13].

MMPs contribute to hepatocellular carcinoma development, participating in tumor angiogenesis, growth, and dissemination [14]. Noel et al. (2000) [15] reported the expression of stromelysin-3 (MMP-11) in aggressive carcinomas promoting tumor development. It is originally identified in stromal cells of primary breast cancer [16]. Stromelysins also play an important role in weakening the ECM components (collagen, elastin, fibronectin, and proteoglycans), thus harming the aortic wall. MMP-10 is associated with disease severity and mortality in patients with peripheral arterial disease [17]. MMPs are produced as zymogens requiring activation by other proteases [18,19] and once activated, proteolytic activity is further regulated by a family of endogenous inhibitors, the tissue inhibitors of metalloproteinases (TIMPs) [20,21]. The deposited three dimensional structures of MMP-3 [22,23], MMP-10 [24], MMP-11 [25] only focused on the domain regions. No complete 3D

* Corresponding author. Bioinformatics Infrastructure Facility for BTISnet, Madras Christian College, Chennai, Tamil Nadu, 600059, India.
E-mail address: joycebenezer@gmail.com (J. Priyakumari).

https://doi.org/10.1016/j.imu.2017.12.003
Received 23 August 2017; Received in revised form 11 November 2017; Accepted 2 December 2017
Available online 19 March 2018

structures were reported yet. Due to their importance in various pathological and normal functioning and the lack of perfect three-dimensional structures, this study aims at predicting their structural models and analyzing the sequential, structural variation and their binding modes with TIMP.

2. Materials and methods

2.1. Primary sequence analysis

The amino acid sequences of three different stromelysins MMP-3, MMP-10, and MMP-11 (UniProt IDs: P08254, P09238, and P24347) were retrieved from UniProt database [26]. The aminoacid composition was computed using PEPSTATS analysis tool [27]. Physico-chemical parameters such as molecular weight, aminoacid composition, aliphatic index, Extinction coefficient, Iso-electric point etc. were calculated using Protparam server [28]. The calculated PI aids in determining isoelectric point of the protein whereas GRAVY (Grand Average of Hydropathicity) score infers the interaction pattern of the protein with the water molecules. Putative functional motifs of the target protein were predicted using NCBI conserved domains and MOTIF Finder [29]. Multiple sequence alignment was performed by Clustal omega [30].

2.2. Homology modeling

The signal peptide and pro-peptides were removed from target sequences and only the chains were subjected for homology modeling. BLAST was used to search the homologous crystal structure available in Protein Data Bank (PDB) [31]. Multiple template structures were utilized for building the three-dimensional structures of targets MMP-3 (Template PDB IDs: 1FBL, 1SU3, 1UMS, 2CLT, 3BA0 and 4FU4), MMP-10 (1FBL, 1SU3, 1UMS, 2CLT, 3BA0, 3V96 and 4FU4) and MMP-11 (1FBL, 1SU3, 1UMS, 1HV5, 2CLT, 3BA0, 4FU4 and 966C). The targets were aligned separately with the templates along with zinc atoms and top five models were predicted for each using MODELLER 9.18 [32].

2.3. Structure validation and comparison

Validation of the best models was performed using RAMPAGE [33], PROCHECK [34], ResProx [35], PROSESS [36] and ProSA analysis [37]. Based on stereochemical properties, the best model was taken for further refinement using Swiss PDB Viewer [38]. The generated model was visualized, inspected and analyzed using PyMOL [39]. The superimposition of the three predicted models was generated using PyMOL.

2.3.1. Molecular docking

The binding affinities of the modeled proteins with their natural inhibitor TIMP1 were checked using Patchdock Server [40]. The docked complexes were visualized and analyzed for their interactions using Discovery Studio Visualizer.

3. Results

3.1. Sequence analysis

Important physicochemical properties like isoelectric point, molecular weight, aminoacid composition, aliphatic index, Extinction coefficient and GRAVY score were calculated for the three stromelysins as depicted in Table 1. Stromelysins showed little variation in their physicochemical properties like molecular weight, pI, aliphatic index and the number of residues in different groups (aliphatic, aromatic, polar, nonpolar, acidic and basic). Moreover, a change in stability was observed for MMP-11, which is unstable whereas the other two stromelysins were found stable. A theoretical isoelectric point of 5.34, 5.18, and 5.53 respectively for MMP-3, MMP-10, and MMP-11 indicate that they are negatively charged. Also, they contain more negatively charged residues

Table 1
Physicochemical parameters of stromelysin.

S. No	Parameters	Value		
		MMP-3	MMP-10	MMP-11
1	Number of aminoacids	378	378	391
2	Molecular weight	42837.2	43009.2	44255.9
3	Theoretical pI	5.34	5.18	5.53
4	Instability index	30.43	35.33	41.40
5	Aliphatic index	71.75	66.38	71.89
6	GRAVY	−0.419	−0.381	−0.310
7	Negatively charged residues (No.)	53	51	49
8	positively charged residues (No.)	39	34	37
9	Aliphatic residues (No.)	93	91	106
10	Aromatic residues (No.)	63	72	71
11	Non-polar residues	206	211	237
12	Polar residues	172	167	154
13	Basic residues	51	47	50
14	Acidic residues	53	51	49

pI- Isoelectric point. Negative GRAVY score values indicates their hydrophilic property. Instability index indicates the nature of the protein (>=40 is instable and >40 is stable).

(Asparagine and Glutamine) than positively charged (Arginine and Lysine). The instability index ensures the stable and unstable nature of the protein, with value 40 or more than 40.0 indicates the unstable nature whereas the value less than 40.0 indicates its stable nature [41]. Predicted instability index of three stromelysin proteins suggests the stable nature of MMP-3 (30.43) and MMP-10 (35.33), whereas MMP-11 (41.40) is unstable. Negative GRAVY score values (−0.419, −0.381, −0.310 respectively) of the targets indicate their hydrophilic nature. Isoelectric point of the target proteins showed the acidic nature of stromelysin family with pI (Isoelectric point) values less than 7.

Conserved domain search revealed that stromelysins contain three functional domains. Peptidase M10 superfamily (Martixin) and HX (Hemopexin-like repeats) superfamily are common to all. MMP-11 contains DNA polymerase III gamma 3 superfamily domain instead of peptidoglycan binding domain in MMP-3 and MMP-10 (Fig. 1). Multiple sequence alignment showed various domain regions including regulatory domain or cysteine switch, protease domain or zinc-binding motif and proline-rich hinge region as shown in Fig. S2 (supplementary file).

Fig. 2 shows the different secondary structural features and the aminoacids occupying the regions based on multiple sequence alignment of the predicted models. The figure shows the protease domain and the hinge region. Regulatory domain, which resides in the propeptide is absent in the predicted structures as we modeled only the protein chains of stromelysins. Apart from these domains, several other conserved regions are visible by this alignment. These conserved regions may also be responsible for their activity.

3.2. Homology modeling

A high level of sequence identity assurances a more reliable alignment between the target and the template. Different Stromelysin proteins (MMP-3, MMP-10, and MMP-11) from *Homo sapiens* were selected for homology-based searching of template structures using BLASTP against the structural database of PDB. Structures that showed maximum identity with a high score and least E-value were used to build the 3D models of different targets. Multiple templates were aligned with each targets using modeller 9v18.

The Modeller software constructed five different structure models for each target, with three different types of scores namely mol pdf, Discrete Optimized Protein Energy (DOPE) and GA341 score, based on physiochemical properties such as Van der Waals interaction, hydrophilicity, hydrophobicity, atomic charges and atomic energy. Various scores of the predicted models are shown in Supplementary Table S1. Models with least DOPE score for each target were considered further for validation. Model5, model1, and model5 respectively for MMP-3, MMP-10, and