

RESEARCH ARTICLE

Structural Dynamics Investigation of Human Family 1 & 2 Cystatin-Cathepsin L1 Interaction: A Comparison of Binding Modes

Suman Kumar Nandy, Alpana Seal*

Department of Biochemistry & Biophysics, University of Kalyani, Kalyani, West Bengal, India

* aseal@klyuniv.ac.in



OPEN ACCESS

Citation: Nandy SK, Seal A (2016) Structural Dynamics Investigation of Human Family 1 & 2 Cystatin-Cathepsin L1 Interaction: A Comparison of Binding Modes. PLoS ONE 11(10): e0164970. doi:10.1371/journal.pone.0164970

Editor: Claudio M Soares, Universidade Nova de Lisboa Instituto de Tecnologia Quimica e Biologica, PORTUGAL

Received: July 8, 2016

Accepted: October 4, 2016

Published: October 20, 2016

Copyright: © 2016 Nandy, Seal. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: The authors received no specific funding for this work.

Competing Interests: The authors have declared that no competing interests exist.

Abstract

Cystatin superfamily is a large group of evolutionarily related proteins involved in numerous physiological activities through their inhibitory activity towards cysteine proteases. Despite sharing the same cystatin fold, and inhibiting cysteine proteases through the same tripartite edge involving highly conserved N-terminal region, L1 and L2 loop; cystatins differ widely in their inhibitory affinity towards C1 family of cysteine proteases and molecular details of these interactions are still elusive. In this study, inhibitory interactions of human family 1 & 2 cystatins with cathepsin L1 are predicted and their stability and viability are verified through protein docking & comparative molecular dynamics. An overall stabilization effect is observed in all cystatins on complex formation. Complexes are mostly dominated by van der Waals interaction but the relative participation of the conserved regions varied extensively. While van der Waals contacts prevail in L1 and L2 loop, N-terminal segment chiefly acts as electrostatic interaction site. In fact the comparative dynamics study points towards the instrumental role of L1 loop in directing the total interaction profile of the complex either towards electrostatic or van der Waals contacts. The key amino acid residues surfaced via interaction energy, hydrogen bonding and solvent accessible surface area analysis for each cystatin-cathepsin L1 complex influence the mode of binding and thus control the diverse inhibitory affinity of cystatins towards cysteine proteases.

Introduction

Cystatins are typical emergency inhibitors of C1 and C13 family of cysteine proteases (CPs) [1], classified into four groups, namely family 1 or stefins, family 2 or cystatins, family 3 or kininogens and family 4 or cystatins devoid of CP activity [2, 3]. In spite of substantial difference in protein sequence, all members of cystatin superfamily shares the characteristic cystatin fold formed by five-stranded anti-parallel β -sheets (β 1- β 5) wrapped around a core of a five-turn α -helix (α 1) lying almost perpendicular to the sheets (S1 Fig). The connectivity within the cystatin fold is: (N)- β 1- α 1- β 2-L1- β 3-(AS)- α 2- β 4-L2- β 5-(C); where AS is a broad “appending structure” positioned at the opposite side relative to the N-terminus and β hairpin loops L1 and L2. Stefins lack AS loop and the second α -helix (α 2). Structural studies have recognized three conserved regions in cystatin fold, the exposed L1 loop containing highly conserved

(Q-X-V-X-G) region flanked between projecting N-terminal segment and C-terminal L2 loop comprising equally conserved PW segment (with an exception of human stefins), altogether form a tripartite, largely hydrophobic, wedge-shaped edge complementary to the active site of papain-like CPs [4, 5].

But the conservation of tertiary structure or similar mechanism of inhibition [4] does not tally well with the inhibition profile of cystatins. The inhibitory affinities of the human family 1 & 2 cystatins, even towards a particular CP, differ notably [6]. In case of human stefins and cystatins, cystatin C emerges as the best overall inhibitor of C1 family of CPs; whereas SD type of cystatins are the poorest of the lot. Human stefin B, cystatin F appears to be 100 fold poorer inhibitor of cathepsin L in comparison to cystatin C; while stefin A, cystatin M/E are found to be 1000 fold, cystatin SN shows 10^5 fold and cystatin S & SA exhibits 10^6 fold lower inhibitory activity [6, 7]. Even in case for human S-type cystatins, where more than 90% sequence similarity is observed, cystatin S comes out as a significantly poorer inhibitor in comparison to cystatin SA & SN [6]. Furthermore, the relative contributions of three conserved regions, viz. N-terminal, L1 & L2 β -hairpin loops, also vary appreciably. In contrast to the previous observations [8], N-terminal segment of human stefin A and B is reported to contribute about 40% of the total free energy of binding for papain, cathepsin B & L [9, 10]; although the first β hairpin loop fails to account for stefin A in CP binding [11] but credited for stefin B-papain interaction [8]; and the second loop of stefin A shows variable affinity depending on the CP [12], responsible for 20–30% of total binding energy in bovine cystatin B-papain, -cathepsin H and -cathepsin B interaction. The L1 and L2 loop of cystatin C is recognized to hold the major share in total free energy of binding of papain, actinidin, cathepsins B and H [13]. For SD-type cystatins, the QXVXG region exhibits main effect on papain inhibition, whereas all three conserved regions participate for cathepsin CPs [6]. Apart from three conserved regions, C-terminal residues also takes part significantly in papain, cathepsin B & H interaction of stefin B [14].

In vivo studies further elucidate the considerable correlation of cathepsin B & L activities with the inhibitory activities of stefin A & B in breast carcinoma [15] and head & neck carcinoma [16] patients, respectively. Substantial reduction in stefin B inhibitory activity and concomitant increase in cathepsins S and L activity is observed in EPM1 pathogenesis [17]. Various reports suggests the levels of cystatin C [18], cystatin M/E [19] and stefin A [20] in tissue and extracellular fluids can serve as relatively reliable markers for diagnosis and prognosis of a variety of diseases [1, 21] while cathepsin B & L levels may well be used as a potential indicator of tumor aggressiveness and metastasis [22, 23].

In this paper the detailed molecular mechanisms behind the diverse inhibitory interaction of human family 1 & 2 cystatins, namely stefin A, B & cystatin C, D, F, M/E, S, SA, SN, with cathepsin L1 (CL1) were investigated. Firstly, the CL1-cystatin/stefin complexes were built through molecular docking and then atomistic molecular dynamics studies of the binding mode of all nine complexes were performed. The integrity of the complexes were verified, important residues involved in interaction were identified by interaction energy (IE), solvent accessible surface area (SASA), and hydrogen bond (HB) analysis and in turn correlated with the varied contributions of the conserved regions of cystatins involved in binding. Altogether our results explained the nature of interaction and variable affinity of cystatins towards CL1.

Methods

Prediction of intrinsic disordered region

The amino acid sequences of human CL1 and all nine cystatins were retrieved from UniProt and were subjected to DisEMBL 1.5 [24], Globplot 2.3 [25], regional order neural network

(RONN) [26] and protein disorder prediction system (PrDOS) [27] to specify regions of higher flexibility and complementarity.

PDB file preparation

Initial atomic coordinates of six proteins, to wit CL1, stefin A, stefin B, cystatin C, cystatin D, and cystatin F were selected on basis of active/inhibitory site mutants, incomplete chains, missing residues, domain swapping etc. from the Research Collaboratory for Structural Bioinformatics Protein Data Bank (www.pdb.org) and downloaded accordingly [28]. Crystal waters were removed, atoms were reordered in certain residues, terminals were modified and conformers (if any) were deleted on basis of occupancies and further the occupancies were corrected. Then crystal structures were subjected to Swiss PDB Viewer [29], mainly to clear the nomenclature issues. The specific information on file preparation is given in [S1 Table](#). For the structures of cystatin M/E, cystatin S, cystatin SA and cystatin SN comparative models prepared previously [30] were used.

Complex formation—molecular docking

CASp (Computed Atlas of Surface Topology of proteins) [31], cons-PPISP (consensus-Protein-Protein Interaction Site Predictor) [32] and the InterProSurf server [33] were utilized to locate the binding-sites in prepared PDB files. To find out the binding modes of cathepsin-cystatin/stefin complexes, crystallographic and modeled structures of stefin A, B, cystatin C, D, E, M/E, S, SA & SN were docked with CL1 using protein-protein rigid body docking program ZDOCK [34] of Discovery Studio (DS) 2.5. The active site of CL1 (CYS 25, HIS 163, ASN 187) [UniProt ID: P07711] and the possible inhibitory sites of cystatins [6, 35] were defined to screen the potential docking configurations. A total of 2000 docked poses were generated for each of the cystatin/stefin-cathepsin complexes and ranked based on the default ZDOCK scoring function [36, 37] combining shape complementarities with electrostatics and desolvation energy (DE), and subsequently re-ranked with a more detailed weighted energy function ZRANK [38]. On basis of known cathepsin-cystatin/stefin conformations, binding site root mean square deviation (RMSD) were calculated and all docked poses within 7 Å deviation were subjected to RDOCK [39] for further refinement. The best docking pose with the lowest RDOCK energy was considered.

Molecular dynamics simulations

In total, 27 systems were prepared for simulation. For each of the nine CL1-cystatin/stefin complexes, we simulated (a) the complex of the CL1 and its binding partner, (b) the CL1 alone, and (c) the binding partner alone. The initial coordinates of simulation groups (b) and (c) were taken directly from group (a). All molecular dynamics simulations (MDS) were performed using GROMACS (Groningen Machine for Chemical Simulations) 4.5.4 [40] simulation packages employing Charmm27 (Chemistry at Harvard Macromolecular Mechanics) force field [41]. The systems were solvated by explicit SPC/E (extended simple point charge) water model [42] in cubic boxes maintaining a minimum 10 Å edge distance. Counter ions (Cl^- and Na^+) were added by randomly replacing water molecules to achieve a neutral simulation cell of physiological ionic strength (0.10 molar). Composition of each system was given in [S2 Table](#). All systems were then minimized using a steepest descents (SD) algorithm [43] in order to eliminate steric clashes or inappropriate geometry.

A 100 ps NVT equilibration was performed at 300 K with position restraints applied to all of the backbone atoms to ease any bad contacts at the side chain solvent interface. The velocity rescale thermostat [44] was used with a temperature coupling time constant (τ_t) of 0.1 ps.

All bond lengths were constrained using the linear constraint solver (LINCS) [45] algorithm allowing for a 2 fs time step. Long-range electrostatic interactions were approximated using the particle mesh Ewald (PME) method [46] with a fourth-order spline interpolation and a 0.15 nm Fourier grid spacing. The short range non-bonded interactions were defined as van der Waals (VDW) and electrostatic interactions for particles within 10 Å. Then all position restraints were withdrawn and a 100 ps NPT simulation was conducted; an isotropic Parrinello-Rahman barostat [47] was set to 1.0 bar of pressure in all directions with a pressure coupling time constant (τ_p) of 2.0 ps. Afterwards 150 ns production run was followed using velocity rescale thermostat and Parrinello-Rahman barostat; LINCS and PME treatments were also implemented as described. Snapshots of the trajectory were taken in every 2 ps.

The GROMACS suite of tools along with a secondary structure recognition algorithm DSSP (Define Secondary Structure of Proteins) [48] and VMD [49] was used for all types of MDS analysis. Every complex trajectory was subjected to gromos clustering and the middle most structure of the largest cluster was taken as representative structure (RS) and compared with the final docking output of same complex (DC) to assess the effect of refinement. Principal component analysis (PCA) was used to track the collective motions of bound and unbound inhibitors. Further, the change in solvent-accessible surface area (Δ SASA), HB pattern and IEs of each residue in the binding interface were determined. Microsoft Excel program was used for preparation of the graph and DS visualizer was employed for 3-D figure generation.

Results and Discussion

All the intrinsic disorder predicting servers detected the N- and C-terminal region of all ten proteins as disordered regions. Additionally the regions near chain break (residues 172–179) and various loop regions (residues 11–25, 54–69, 80–109, 154–161) of CL1 were recognized as disordered region. Further, AS loop of cystatin C, D & M/E was noticed as disordered region in RONN; Globplot & RONN spotted part of β_4 of cystatin C and D; in all S-type cystatins, β_4 , α_2 and β_2 was noted as disordered regions by RONN and DisEMBL respectively; in addition DisEMBL marked β_3 of cystatin SN as highly disordered loops.

In the following sub-sections, we first checked the consistency among nine simulations of unbound CL1, then stability of the all nine complexes were assessed, the effect of refinement was quantified and binding modes of complexes were compared in view of diverse participation of conserved regions in VDW and electrostatic interactions. The detailed analyses of all nine individual CL1-stefin/cystatin complexes were placed in [S1 Text](#). The unbound CL1s were named after the corresponding CL1-stefin/cystatin complex (say, CL1_A denotes the unbound CL1 taken from CL1-Stefin A complex) from which its initial coordinates were taken as stated in MDS section under Methods. The amino acid residues of stefins/cystatins were referred according to their positions in the alignment ([S2 Fig](#)) and for CL1 the crystal structure (PDB ID: 1ICF) numbering was followed.

Consistency of simulations: unbound cathepsins

All nine simulations of unbound CL1 were appeared to be in stable state in terms of intra-protein, protein-solvent electrostatic and VDW energy ([Fig 1A](#)). Electrostatic energy profile exhibited more fluctuation than the VDW one, protein-water interface recorded higher variation compared to intra-protein electrostatics.

Although average structures might not be physically meaningful or energetically accessible but can be used as a tool for characterizing sampling. Here a true representative average structure of every trajectory was computed by an iterative procedure; average structure was calculated using the first frame and then repeated with the resulting average structure until the

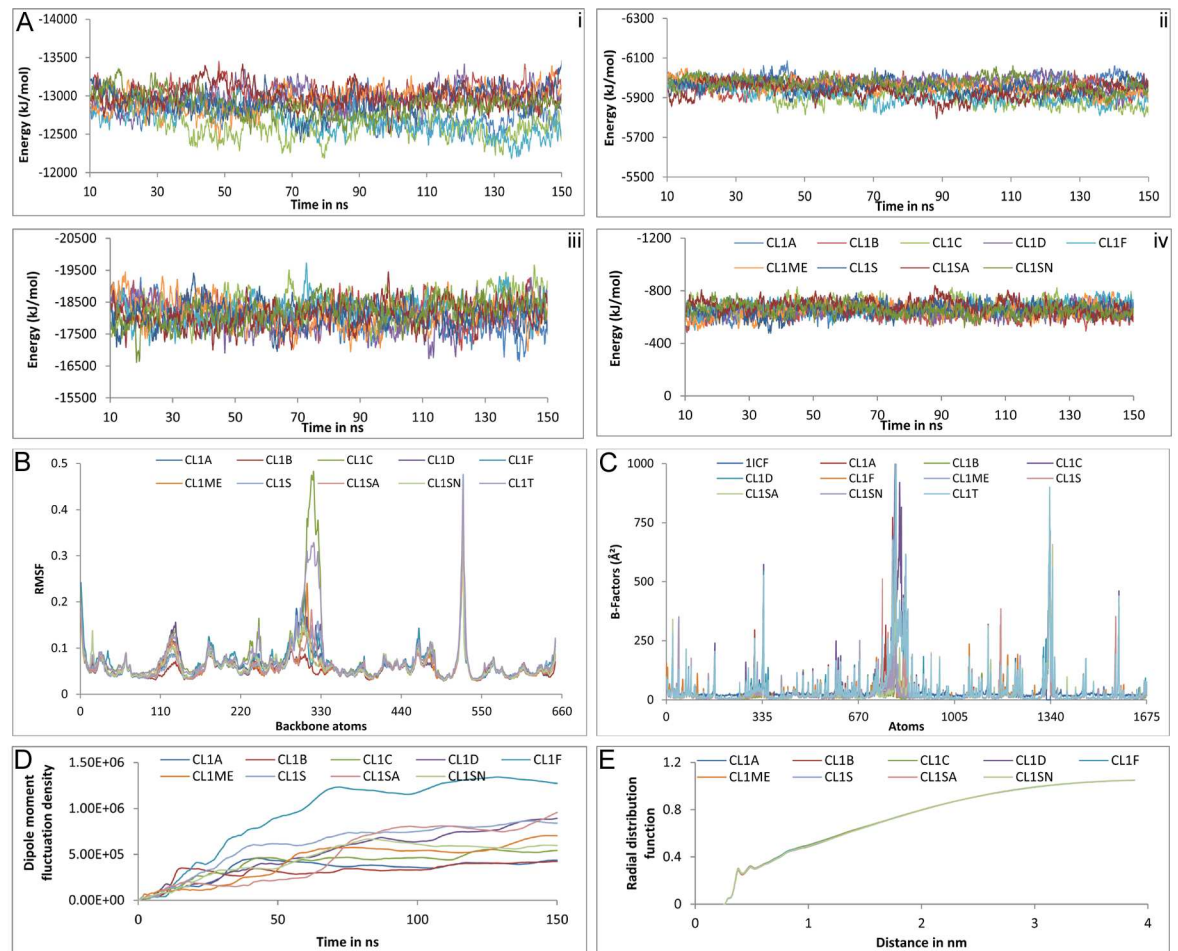


Fig 1. Convergence of the properties of un-bound cathepsin L1s in solution. A. Non-bonded electrostatic (i, iii) and VDW (ii, iv) energy profile in intra-protein (i, ii) and protein-water (iii, iv) interfaces. Color code is same for i-iv. B. RMSF analysis. C. Crystallographic and simulated B-factors. D. Protein dipole moment fluctuation densities. E. Radial distribution functions between backbone atoms of protein and water oxygen atoms.

doi:10.1371/journal.pone.0164970.g001

RMSD between successive average structure becomes $<0.001 \text{ \AA}$ [50]. The individual trajectory-average structures of CL1 differed from the crystal structure by $0.873\text{--}1.033 \text{ \AA}$, compared to a $0.272\text{--}0.419 \text{ \AA}$ difference with respect to the global trajectory-average; while they differed from each other by $0.399\text{--}0.639 \text{ \AA}$. S3 Table showed, the individual trajectory-averages of unbound CL1s (CL1_A—CL1_{SN}) were closer to the global trajectory average (CL1_{GT}) than the crystal structure (PDB ID: 1ICF). Little variation among these average structures pointed towards convergence of simulation and indicated toward possible overlap of conformational spaces between the trajectories [51]. In fact PCA depicts an overlap of $0.327\text{--}0.614$ among individual trajectories and $0.449\text{--}0.643$ between individual and global trajectories (S4 Table). Stable RMSD output and atomic motions of unbound receptors also seconded the same by exhibiting similar nature of fluctuation (S1 Text and Fig 1B). Discrepancies observed in the residues 70–80 and 103–108 of unbound cathepsin of cystatin C complex, which also contributed at global trajectory (CL1_{GT}), were already predicted as disordered region.

Structural properties of the unbound cathepsins illustrated little variation in individual trajectories and resembled more with the global trajectory rather the crystal structure (S5 Table). The molecular volume, density and radius of gyration (Rg) recorded minute variation of 0.3--

0.6%, 0.4–0.5% & 0.4–1% with relative to the crystal structure and 0.4%, 0.4% & 0.2–0.4% in case of global trajectory, respectively. Only noticeable change was observed in case of SASA, 7.4–10.6% in comparison to crystal structure and 0.8–2.1% for global trajectory. This might be due to difference in environment between solvent simulation and crystallographic structure determination. In contrast, to all these structural characteristics crystallographic and simulated B-factors hardly correlated with each other, simulated ones reported markedly higher values (Fig 1C). This might be due to huge difference in determination techniques, time-scales and environment [52].

Dielectric relaxation of unbound CL1s was considered through dipole moments fluctuation density, which were stabilized around 90 ns although CL1F described distinctly higher values (Fig 1D). Hydration of CL1 surface defined by radial distribution functions were seemed to be indistinguishable from each other (Fig 1E). Thus we can say the non-bonded energies, conformational differences, dielectric relaxation, hydration of protein and most of the structural properties except B-factor looked to be converged within the limit of the simulation.

Stability of the complexes

DSSP profile demonstrated overall stability of secondary structural elements in receptor and inhibitor in both bound and unbound state over the period of simulation; recording only a maximum change of 3% in secondary structural elements between bound and unbound proteins (S1 Text). RMSD analysis also depicted by and large stability of all complexes—stefin A, B and cystatin C, D & SA complexes got stabilized readily within 10–20 ns whereas cystatin F, M/E, S & SN complexes illustrated high relaxation time of ≈ 60 ns (Fig 2A). Δ RMSF study (Δ RMSF = RMSF_{bound} – RMSF_{unbound}) recorded mostly higher fluctuation of the bound proteins with comparatively lower fluctuation at the sites of direct interaction (Fig 2B–2D). Apart from the global observations some local attributes were also noticed, such as marked stabilization of β -hairpin loops in bound state especially in case of stefin A, B and cystatin C complexes (Fig 2B and 2C and S3A, S3B, S7A, S7B, S11A and S11B Figs); split of $\beta 5$ into β -sheet & β -bridge in bound cystatin D leads to higher RMSF at C-terminal (Fig 2C and S15A and S16B Figs) and presence of 2nd α -helix in all cystatins but cystatin M/E were observed both in bound and unbound state (S1 Text). The four residue insertions in the AS loop of cystatin M/E (S2 Fig), also recognized as disordered region, might be the reason behind disappearance of $\alpha 2$, as well of its unique appearance in DSSP analysis, and higher fluctuation in RMSF analysis (Fig 2C, S23A, S23B and S24B Figs). CL1 illustrated little difference between bound and unbound

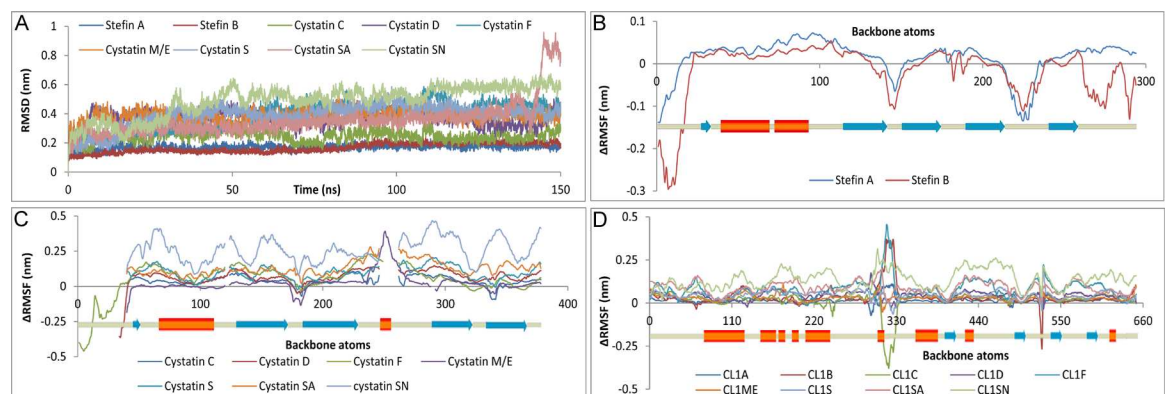


Fig 2. Stability of the complexes. RMSD of cathepsin L1-stefin/cystatin complexes (A) and Δ RMSF (RMSF_{bound} – RMSF_{unbound}) in stefins (B), cystatins (C) and cathepsin L1s (D).

doi:10.1371/journal.pone.0164970.g002

states except near the region of chain break (residue 175) and the loop region (residues 95–113), both noted as disordered segment (Fig 2D). These two regions were also registered high RMSF values and contributed in RMSD plot of CL1s associated with stefin B, cystatin C, F, S, SN complexes and unbound receptors (Fig 2A and 2D and S8A, S8C, S12A, S12C, S20A, S20C, S28A, S28C, S36A and S36C Figs). Cystatin C, D, M/E and SN complexes exhibited more variation in RMSD profile, CL1-Stefin B complex reported an increase in RMSD near 75ns and again stabilized ≈ 100 ns; cystatin SA complex also illustrated an increase in RMSD in post 130ns simulation; in the final 35 ns cystatin SN reported almost a plateau (Fig 2A). The peculiarities of the complex RMSD were gathered either from the receptor or the inhibitor, such as bound receptor of stefin B complex showed a similar increase in RMSD after 75 ns due to enhanced fluctuation of disordered loop residues 99–112 (Fig 2A and S8A and S8C Fig); markedly higher fluctuation in AS loop was responsible for late increase in RMSD in cystatin SA complex (Fig 2A and S32B Fig); while higher RMSF of the bound inhibitors were accountable for the RMSD fluctuations of cystatin C (S12A and S12B Fig), M/E (S24A and S24B Fig—AS loop) and higher RMSF of both bound receptor and inhibitor were liable for RMSD fluctuations of cystatin D (S16 Fig), SN (S36 Fig). The bound CL1s and stefins/cystatins portrayed either lower or comparable RMSD with the unbound one. In case the bound ligand reported higher RMSD values compared to the unbound ligand, the bound receptor registered lower RMSD compared to the unbound CL1 and vice-versa—thus illustrating the stability of the complexes (S1 Text). Further the Rg and total SASA served as additional measure of stability of the complexes recording only $\approx 2\%$ variation (S6 Table).

Quantification of refinement by MDS

Each of the CL1-stefin/cystatin complex trajectories was subjected to gromos clustering [53] and the middle most structure of the largest cluster was taken as representative structure (RS) and compared with the final docking output of same complex (DC) to quantify the effects of refinement through MDS.

Comparison of secondary structure of RSs with corresponding DCs revealed certain regions of the receptor were transformed into helix (residues 58–60), while in other regions helices (residues 64–66, 102–104) and sheets (83–84, 105–107) were altered into turn and coils as an effect of refinement. In case of inhibitors, after refinement, various secondary structure modifications were observed in 1st β -sheet (sheet structure shifted towards N-terminal in Cystatin M/E & SA, reappeared in cystatin SN, shortened in Cystatin C & D); 2nd and 3rd β -sheet were shortened in cystatin F, M/E, SN and cystatin M/E, S respectively; elongation of 5th β -sheet was seen in stefin complexes, while the same was broken into two parts in cystatin C & D complexes and got shortened in cystatin M/E and SA (S7 Table).

Superposition of receptors of RS with equivalent DC illustrated differences in binding conformations of the inhibitors after MDS; shift of inhibitor centroids, relative motion between secondary structures were observed on refinement, Cystatin F and M/E recorded the highest RMSD (Fig 3 and S8 Table). Depending upon the degree of changes implicated on refinement, the binding interface of the DC & RS differs accordingly; cystatin D, F and SN showed the most number of alterations (S9 Table). The RS of stefin A moved a bit far from the receptor along horizontal axis, shift of centroids by 2.64 Å, in comparison to the docked complex (Fig 3A and S8 Table); which enables the refined complex to identify the importance of Pro3 [9] of stefin A and Asn66 of CL1 in molecular interaction and HB forming that remained unrecognized in the docking studies (Table 1, S1 Text and S6 Fig). Refined Stefin B exhibited least deviation and centroid motion (Fig 3B, S8 Table) but good enough to reveal the crucial roles of Met1, Ala49, His75 of the inhibitor [10, 14] and Gly68 of CL1 in hydrogen bonding,

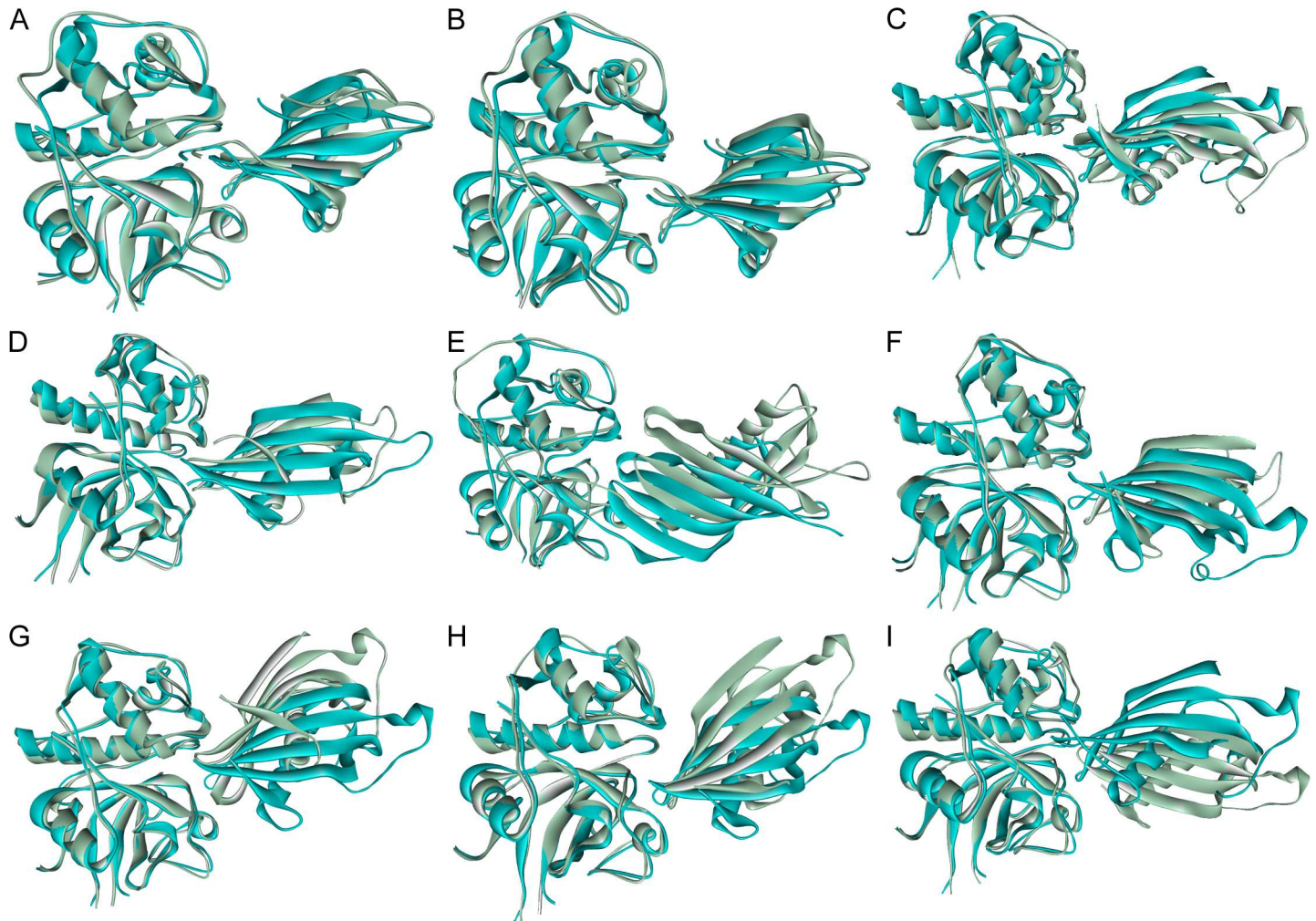


Fig 3. Change in conformation of complexes after refinement. Superimposed structures of refined complexes and corresponding docking outputs of cathepsin L1 complexes with (A) stefin A, (B) stefin B, (C) cystatin C, (D) cystatin D, (E) cystatin F, (F) cystatin M/E, (G) cystatin S, (H) cystatin SA, (I) cystatin SN. Refined structure was shown in green and docked output in cyan.

doi:10.1371/journal.pone.0164970.g003

electrostatic and VDW interaction (Table 1, S1 Text and S10 Fig). In DCs of stefin A & B, catalytic site residue Cys25 of CL1 was predicted to form HB with stefins, but was lost on refinement—in agreement with the exosite nature of stefins. Cystatin C was tilted and centroid was repositioned in its refined form—the N-terminal and L1 loop came up from the docking plane whereas the L2 loop goes down due to clockwise (β_2 , β_4 , β_5) and anti-clockwise (β_1 , β_3) motions of β -sheets (Fig 3C and S8 Table). As a result the HB between Gly67 and Asp162 was the only HB of DC that retained after refinement and Gly68 of CL1, N-terminal Gly19, entire L1 loop residues of cystatin C specially Ile64 except Lys62 emerges as new interaction centre (Table 1, S1 Text and S14 Fig). RS of cystatin D also took an inclined conformation in comparison to the DC, undergone a large diagonal shift in YZ plane, while β -hairpin loops of both structures remained almost fixed, the N-terminal & far end of RS (C-terminal, AS loop) moved away with respect to docked conformation (Fig 3D and S8 Table). The electrostatic, VDW interaction and hydrogen bonding pattern remained almost same except the N-terminal participation got reduced and contribution of two β -hairpin loops amplified after refinement (Table 1, S1 Text and S18 Fig). In cystatin F DC, mainly the N-terminal along with two

Table 1. Key residues of interaction.

| Cathepsin L1 complex with | Receptor | Inhibitor |
|---------------------------|---|--|
| Stefin A | <u>Gly20</u> , Gln21 , <u>Gly23</u> , <u>Cys65</u> , Asn66 , <u>Gly68</u> , <u>Glu141</u> , <u>Leu144</u> , <u>Phe145</u> , <u>Met161</u> , Asp162 , Trp189 | Met1 , Ile2 , Pro3 , Gly4 , <u>Gln46</u> , Val47 , Val48 , Ala49 , <u>Gly50</u> , <u>Phe70</u> , Lys71 , <u>Leu73</u> , <u>Pro74</u> |
| Stefin B | <u>Gly20</u> , Gln21 , <u>Gly23</u> , Cys65 , Asn66 , Gly68 , <u>Leu69</u> , <u>Met70</u> , Asp71 , <u>Leu144</u> , <u>Phe145</u> , <u>Tyr146</u> , <u>Met161</u> , Asp162 , Trp189 , <u>Glu192</u> | Met1 , Met2 , Cys3 , Gly4 , <u>Pro6</u> , <u>Gln46</u> , Val48 , Ala49 , <u>Gly50</u> , <u>Asn52</u> , <u>Phe70</u> , <u>Gln71</u> , <u>Ser72</u> , <u>Leu73</u> , His75 , <u>Lys78</u> , <u>Ser83</u> , Tyr97 |
| Cystatin C | <u>Gln60</u> , <u>Glu63</u> , <u>Asn66</u> , Gly68 , <u>Leu69</u> , <u>Gly139</u> , <u>His140</u> , <u>Leu144</u> , Asp162 | Gly19 , <u>Pro20</u> , Gln63 , Ile64 , Val65 , Ala66 , Gly67 , <u>Tyr115</u> , <u>Pro118</u> , <u>Trp119</u> |
| Cystatin D | <u>Glu63</u> , Asn66 , Gly68 , <u>Leu69</u> , <u>Gly139</u> , <u>Asp160</u> , Asp162 | <u>Gly18</u> , <u>Gln63</u> , Ile64 , Val65 , Gly66 , <u>Asn69</u> , <u>Gln113</u> , <u>Asn115</u> , Trp119 |
| Cystatin F | <u>Gln21</u> , Glu63 , <u>Cys65</u> , Asn66 , <u>Leu69</u> , Asp71 , <u>Tyr72</u> , <u>Gln78</u> , Asp114 , <u>Gly139</u> , <u>His140</u> , <u>Glu159</u> , Asp160 , Met161 , <u>Asp162</u> , Ala214 , Ser216 | Thr6 , <u>Cys7</u> , <u>Ser13</u> , Arg14 , Lys16 , Lys66 , Trp119 |
| Cystatin M/E | <u>Gln60</u> , Glu63 , Asn66 , Gly68 , <u>Leu69</u> , <u>Gly139</u> , <u>Leu144</u> , <u>Glu159</u> , Asp160 , Met161 , Asp162 | Arg21 , <u>Gln29</u> , <u>Leu64</u> , Val65 , Ala66 , Gly67 , <u>Ile68</u> , Lys69 , <u>Pro118</u> , Trp119 |
| Cystatin S | <u>Glu63</u> , Asn66 , <u>Gly68</u> , <u>Leu69</u> , <u>Gly139</u> , <u>Leu144</u> , <u>Asp162</u> , <u>Trp189</u> | <u>Gln63</u> , Phe65 , Gly66 , Asn69 , <u>Phe71</u> , <u>Glu113</u> , <u>Pro118</u> , Trp119 , <u>Asn127</u> |
| Cystatin SA | Gln21 , Glu63 , Asn66 , <u>Gly139</u> , <u>Leu144</u> , Asp160 , Asp162 , <u>Glu192</u> | Gly19 , <u>Gln63</u> , <u>Ile64</u> , Val65 , <u>Gly66</u> , <u>Gln113</u> , <u>Tyr115</u> , Trp119 , <u>Arg122</u> , <u>Ser124</u> , <u>Val126</u> , <u>Asn127</u> , Arg129 |
| Cystatin SN | <u>Gln21</u> , <u>Asn66</u> , <u>Tyr72</u> , Asp137 , <u>Gly139</u> , Asp160 , Asp162 | Gly19 , Tyr21 , <u>Val65</u> , <u>Gly66</u> , <u>Tyr115</u> , Trp119 |

Residues that reports high IE (>30KJ/mol) are in bold, hydrogen bonded residues are underlined and residues marked in SASA analysis are italicized.

doi:10.1371/journal.pone.0164970.t001

β -hairpin loops blocked the catalytic site of the receptor; while in the RS, the role of N-terminal remained almost same but L1 & L2 loops moved away from the catalytic core (Fig 3E and S8 Table). Different conformations of N-terminal and counter rotations of β -hairpin loop forming sheets of RS have contributed to the repositioning of cystatin centroids; different conformations of terminal regions and AS loop as well recorded the highest deviation among refined cystatins (S8 Table). Correspondingly the interactions with L1 and L2 loop got diminished; Thr6 of cystatin F and Glu63, Asn66 of receptor came into the interaction sphere (Table 1, S9 Table, S1 Text and S22 Fig). All three conserved sites of cystatin M/E stayed in almost same position before and after refinement, although the refined inhibitor undergoes an anti-clockwise rotation for which the far end of the refined inhibitor and L2 loop moved away from the docked position (Fig 3F). A shift of centroid was observed, L1 loop of RS popped up the docking plane due to counter rotations of β 2 & β 3 sheets (S8 Table). Two HBs between L2 loop and receptor residues were lost and rest of the interaction pattern remain roughly the same (Table 1 and S9 Table, S1 Text and S26 Fig). Cystatin S got pulled towards the receptor on refinement—documented second highest centroid movement; rotated anti-clockwise by keeping the point of interaction of L1 loop almost same, while L2 loop dipped below the docking plane due to counter rotations of β 4 & β 5 and N-terminal moved off from the receptor (Fig 3G and S8 Table). Interacting residues remained more or less identical before and after MDS (S9 Table) but their involvement in interaction changed a lot; after refinement, HB pattern got altered totally, L1 loop residues dominated the interaction (Table 1, S1 Text and S30 Fig). Position of three conserved site remained almost stationary for cystatin SA before and after refinement but the far end of the inhibitor (C-terminal, AS loop) came closer the receptor as in case of cystatin S but reported a lower centroid shift (Fig 3H and S8 Table). Residue association in electrostatic & VDW interaction stayed nearly same but lot of short lived HBs formed with L2 loop of the inhibitor on refinement owing to its sinking below the docking plane due to counter

rotations of $\beta 4$ & $\beta 5$ (Table 1, S8 and S9 Table, S1 Text and S34 Fig). RS of cystatin SN moved away from the receptor rotated in clockwise direction, accounted for the highest centroid shift, portraying the reverse scenario of cystatin S & SA (Fig 3I and S8 Table). Consequently N-terminal region acted as the principal interaction hub and number of interacting residues reduced heavily from the β -turns (Table 1, S9 Table, S1 Text and S38 Fig).

Comparative study of the interaction profile of complexes

Cathepsin L1. Our observations goes well with the current knowledge that human family 1 and 2 cystatins does not directly bind to the catalytic site residues of CL1 (Cys25, His163, Asn187), rather blocks the access to these residues by binding in nearby positions (Table 1). The CL1 residues were portrayed in Fig 4 according to their frequency of participation in CL1-stefin/cystatin interaction. We found four distinct regions of CL1 that contributed heavily in the interaction, to wit

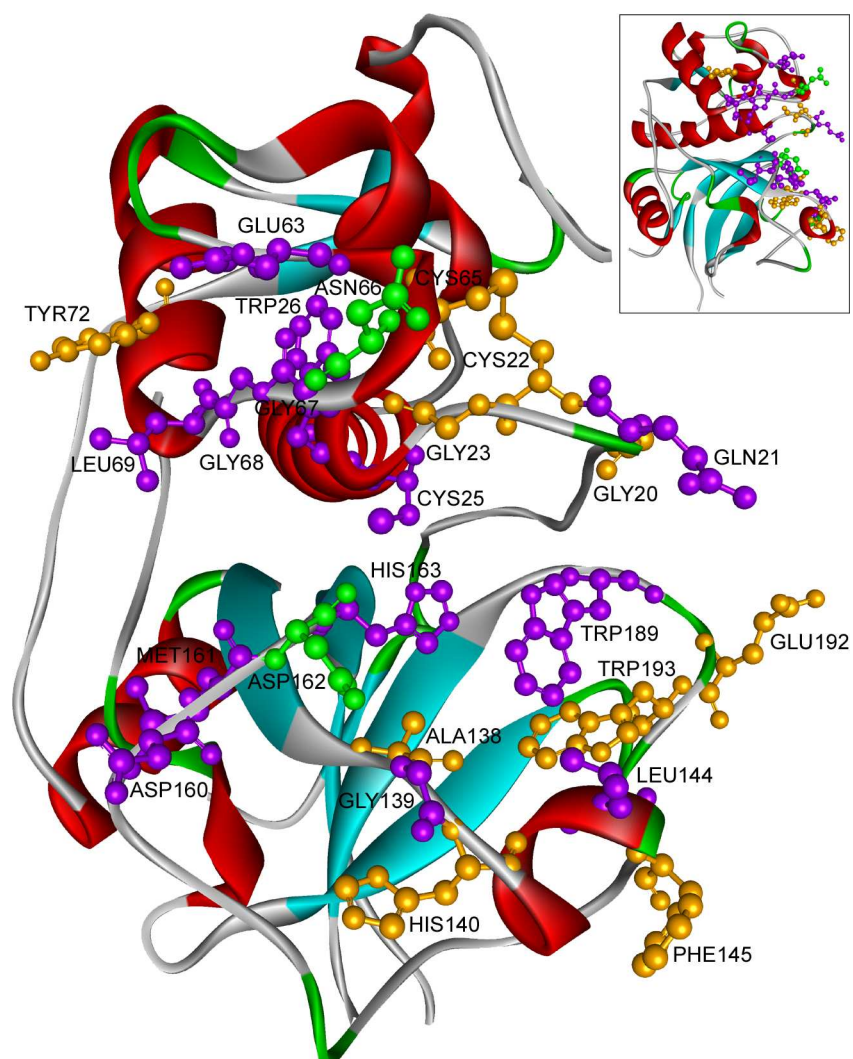


Fig 4. Participation of cathepsin L1 residues in cathepsin L1-stefin/cystatin interaction. Residues always involve in interaction were colored in green, mostly took part were in purple and occurred seldom were in orange. cathepsin L1 (inset) and its 90° rotated were shown in picture. Residues with IE >10KJ/mol were considered for the representation.

doi:10.1371/journal.pone.0164970.g004

1. residues Gln21;
2. residues Glu63, Asn66, Gly68, Leu69;
3. residues Gly139, Leu144; and
4. residues Asp160, Asp162 (Table 1).

The first group was more important for stefins; the third group of residues mainly participated in VDW contacts; the residues of second and fourth group contributed to both electrostatic and VDW interaction. In stefin interaction few more residues took part, such as Gly20, Gly23, Cys65, Phe145, Met161 and Trp189. Among all these residues Asn66 and Asp162 came common in all the complexes, participated in all kind of interaction with high IE as well (Table 1).

Cystatins. Among all nine CL1-cystatin complexes, VDW interaction dominated the IE profile in five complexes (Stefin A, B, Cystatin C, D and S), electrostatic interaction did so in three (Cystatin F, M/E and SN) and in cystatin SA-CL1 complex, electrostatic and VDW contacts were of equal importance (Table 2). Correlated motions always remained largely associated with the conserved region that dominated the overall interaction of any particular complex. Out of three conserved regions, the conserved L2 loop served as a VDW interaction centre; generally provided 10–30% of total potential energy of interaction to cystatins and 20–25% for stefins with two marked exceptions of cystatin S and SA, where L2 loop hold a larger share, 42% and 53% of IE, respectively. The highly conserved PW motif (S2 Fig) provided 50–90% of the total L2 loop contribution, while Trp alone contributed 40–75%. Cystatin SA-CL1 complex came up as an exception, where PW motif provided only 20% of the L2 loop IE and the rest amount was contributed by the adjacent β 4, β 5 sheet residues (S34A Fig). In stefins, as

Table 2. Relative contribution of conserved regions and their participation in electrostatic and van der Waals interaction.

| Cathepsin L1 complex with | Contribution in Interaction Energy (IE) in % | | | | IE (E:V) |
|---------------------------|--|-----------|---------|------------|-----------|
| | N-terminal | L1 Loop | L2 Loop | C-terminal | |
| Stefin A | 39 | 35 | 24 | 2 | 28:72 |
| E:V | 26:74 | 23:77 | 37:63 | 49:51 | |
| Stefin B | 45 | 23 | 21 | 11 | 40:60 |
| E:V | 48:52 | 25:75 | 33:67 | 52:48 | |
| Cystatin C | 19 | 68 | 13 | - | 32:68 |
| E:V | 42:58 | 33:67 | 16:84 | | |
| Cystatin D | 6 | 61 | 33 | - | 32:68 |
| E:V | 33:67 | 34:66 | 28:72 | | |
| Cystatin F | 73 | 17 | 10 | - | 73:27 |
| E:V | 78:22 | 87:13 | 13:87 | | |
| Cystatin M/E | 27 | 55 | 18 | - | 61:39 |
| E:V | 100:0 | 61.5:38.5 | 4:96 | | |
| Cystatin S | 6 | 52 | 42 | - | 21.5:78.5 |
| E:V | 60:40 | 34:66 | 1:99 | | |
| Cystatin SA | 26 | 21 | 53 | - | 53:47 |
| E:V | 99:1 | 25:75 | 41:59 | | |
| Cystatin SN | 67 | 19 | 14 | - | 69:31 |
| E:V | 93.5:6.5 | 17:83 | 20:80 | | |

E- Electrostatic Interaction & V–van der Waals Interaction

doi:10.1371/journal.pone.0164970.t002

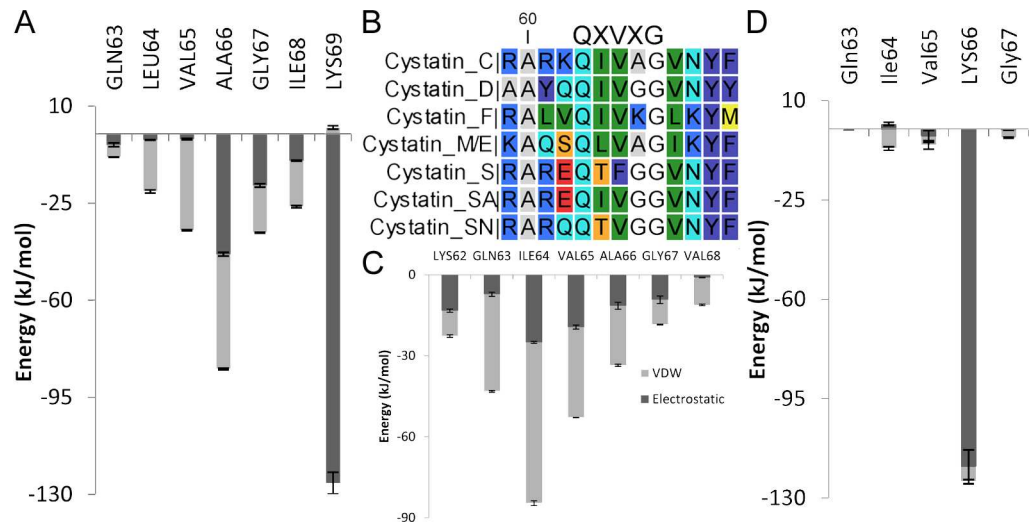


Fig 5. Contribution of L1 loop in cathepsin L1-cystatin interaction. Interaction energy profile of L1 loop of cystatin M/E (A), cystatin C (C), cystatin F (D) and alignment of L1 loop residues of cystatins (B). Error bars represent the estimated error in GROMACS calculation. Legends were same in A, C, D.

doi:10.1371/journal.pone.0164970.g005

the conserved Trp remained absent Pro alone contributed 12–15% of IE, comparable to the cystatin complexes.

L1 loop seemed to be of major importance for Cystatin C, D, M/E & S, where it provided 52–68% of total IE, for the rest the share varied between 17–35% (Table 2). L1 loop region in general acted as VDW interaction centre for stefin/cystatin complexes, 65–83% of L1 loop contribution was provided by VDW contacts, while cystatin F and M/E complexes emerged as exceptions. The third and fourth residue of the QXVXG motif looked to be of more importance, provided 50–60% of the L1 loop contribution; for cystatin F the contribution became highest (93% of L1 loop IE) and reached the low for cystatin C & M/E (30.7% & 35.4% respectively). All hydrophobic residues in QIVAG contributed in cystatin C and the most hydrophobic second Ile residue played the central role (Fig 5C)—nullifies the usual dominance of 3rd & 4th residue. CL1-cystatin C complex also recorded the highest L1 loop contribution (68% of IE) with the help of strongest hydrophobic triplet IVA among the cystatins (Table 2 and Fig 5B). The presence of Lys residue within the QXVXG motif (Fig 5D) of cystatin F disrupted the hydrophobic core and as a result VDW contribution, merely turned into 13% of L1 loop's share of IE. Only one residue, Lys66, recorded high IE in the IE profile of cystatin F (Fig 5D), in contrast to regular presence of 2–3 residues in IE profile from L1 loop with a high score and noted change in SASA. In case of cystatin M/E the residue Lys69, showed high electrostatic IE, formed multiple HBs, and influenced the interaction of L1 loop in a great extent (Fig 5A and S26B Fig). Although the hydrophobic core and the regular involvement of QXVXG persisted, due to the presence of more hydrophobic residues Ala66 & Ile68 in the region (Fig 5B), but the overall interaction profile of L1 loop (Fig 5A) moved into the electrostatic way (61.5% of L1 loop contribution).

Role of N-terminal also varied widely among the complexes, while it provided ≈40% of IE to the stefins, the contribution of this region became as low as 6% for cystatin D & S complexes, turned moderate (19–27% of total IE) for cystatin C, M/E & SA and recorded highest for cystatin F and SN complexes, 73% & 67% of total IE, respectively (Table 2). Mostly N-terminal acted as an electrostatic interaction centre, only in cystatin C, D and stefin A, VDW interaction prevailed; for stefin B electrostatic and VDW contacts were equally important. The C-terminal

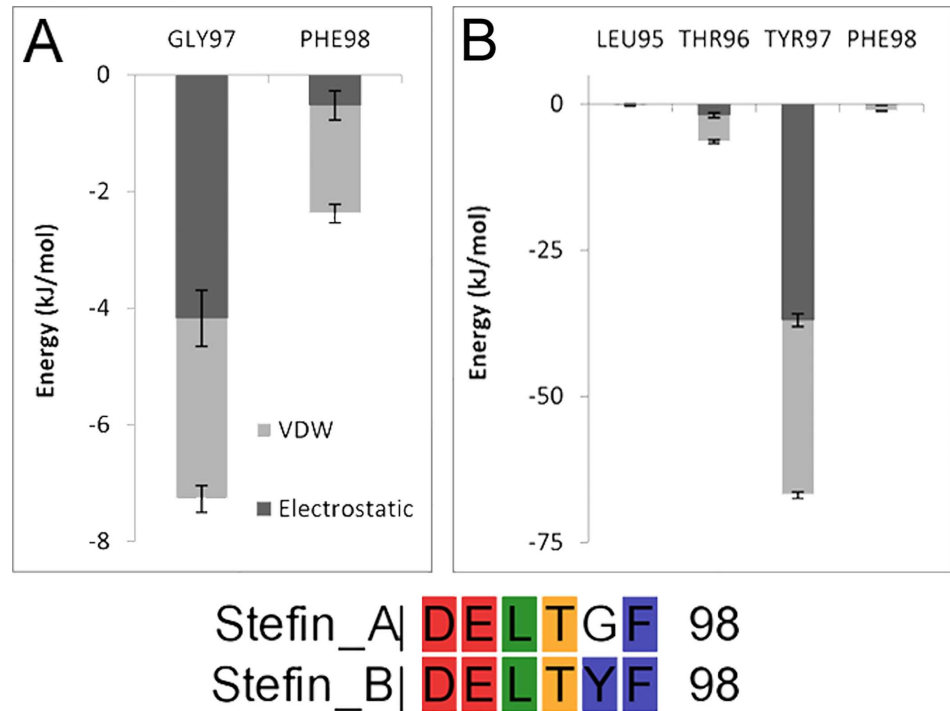


Fig 6. C-terminal interaction profile of stefin-cathepsin L1 complexes. Potential energy of interaction of C-terminal region of Stefin A (A) and Stefin B (B) along with their C-terminal alignment.

doi:10.1371/journal.pone.0164970.g006

also shared equal contribution of electrostatic and VDW interaction but appeared to be of significance for stefin-CL1 complexes only.

Further similarities were noticed between the interaction profiles of the complexes, which also portrayed diverse approaches of cystatins towards CL1. Depending on parallel roles of inhibitor and receptor residues, similarities in IE profile and diverse contributions of conserved regions, stefin/cystatin-CL1 complexes might be grouped into stefin A, B; cystatin C, M/E; cystatin D, S and cystatin F, SN, while cystatin SA stood alone. This differential interaction patterns most likely had been the reason of varying affinity of cystatins towards cathepsins along with the dissimilar amino acid compositions at three interacting sites. For example, presence of Gly instead Tyr (residue 97) was the obvious reason behind minor contribution of C-terminal of stefin A compared to stefin B-CL1 complex (Table 2 and Fig 6); similarly occurrence of more hydrophobic Ala at 66th position helped cystatin C and M/E to establish hydrophobic contacts more efficiently than others (Fig 5A-5C), which also contributed to the L1 loop participation (Table 2) & CL1 affinity as well; consecutive presence of less hydrophobic Thr and Phe at L1 loop (Fig 5B, 64th & 65th position) of cystatin S might be the reason of its inactive nature; differential affinity of S-type of cystatins in spite of their huge sequence similarity might be attributed to their extremely diverse interaction profile (S30A, S34A and S38A Figs), varying contribution of conserved regions (Table 2).

Conclusions

In this work, inhibitory complexes of CL1 and human family 1 and 2 cystatins were modeled; built complexes were verified with available experimental data [5, 9, 10, 12, 13, 14, 54] and used for further study of molecular interactions between receptor and inhibitor. Cystatins were exosite binding inhibitors of CL1, interacted with the receptor through their three conserved

regions in more than one way as documented by their interaction profiles. While the CL1-cystatin F, M/E, SN complexes were dominated by electrostatic interaction, the interaction profile of cystatin SA was equiposed between electrostatic and VDW contacts and in the rest, VDW interaction prevailed. L1 and L2 loop mainly acted as VDW interaction centre, whereas electrostatic interaction prevailed in N-terminal. In fact, L1-loop dictated the balance of the overall interaction profile, whether it will be dominated by VDW contacts or electrostatic interaction will play the key role. The little differences in amino acid compositions in the respective conserved sites controlled the diverse interaction patterns of cystatins, which in turn determined the wide-ranging affinity of cystatins towards cathepsins. These results were in tune with the available information and explained the nature of interactions & variable affinities of cystatins as well. Our findings will be helpful in understanding the details of cystatin inhibition and will serve as a guideline for mutation & protein designing studies.

Supporting Information

S1 Fig. Structure and connectivity of cystatin fold. The prepared human cystatin C (S1 Table) was used for visualization.
(TIF)

S2 Fig. Alignment of human family 1 and 2 cystatins. Disulfide bonds and conserved regions involved in CP inhibition were marked. Alignment was performed in Clustal X2 [55] with default parameters and visualized by CLC sequence viewer 7 (<http://www.clcbio.com/products/clc-sequence-viewer/>).
(TIF)

S3 Fig. DSSP analysis of Stefin A–Cathepsin L1 complex. Secondary structure content of Stefin A in bound (A) and unbound (B) state and that of Cathepsin L1 in bound (C) and unbound (D) form.
(TIF)

S4 Fig. RMSD & RMSF analyses for Stefin A–Cathepsin L1 complex. (A) Average backbone RMSD of complex, inhibitor (I) and receptor (R) in bound and unbound state. RMSF of Stefin A (B) and Cathepsin L1 (C) in complex and in absence of their binding partner.
(TIF)

S5 Fig. PCA of Stefin A–Cathepsin L1 complex. (A) Covariance matrix illustrating correlated and anticorrelated motions of bound (top left) and unbound (bottom right) steffin A. The secondary structure of steffin A backbone is represented along the axes (from left to right and from bottom to top). (b) Motion of the largest eigenvector of steffin A in absence (left) and presence (right) of cathepsin L1.
(TIF)

S6 Fig. Binding interface characterization of Stefin A–Cathepsin L1 complex. (A) Potential energy of interaction between binding interface residues of steffin A (I) & cathepsin L1 (R). Error bars represent the estimated error in GROMACS calculation. (B) Average number of HBs formed among interface residues; error bars designate standard deviation. (C) Appreciable changes in SASA on complex formation among binding interface residues.
(TIF)

S7 Fig. DSSP analysis of Stefin B–Cathepsin L1 complex. Secondary structure content of Stefin B in bound (A) and unbound (B) state and that of Cathepsin L1 in bound (C) and unbound

(D) form.
(TIF)

S8 Fig. RMSD & RMSF analyses for Stefin B-Cathepsin L1 complex. (A) Average backbone RMSD of the complex, inhibitor (I) and receptor (R) in bound and unbound state. RMSF of Stefin B (B) and Cathepsin L1 (C) in complexed form and in free state in solution.
(TIF)

S9 Fig. PCA of Stefin B-Cathepsin L1 complex. (A) Covariance matrix illustrating correlated and anticorrelated motions of bound (top left) and unbound (bottom right) steфин B. The secondary structure of steфин B backbone is represented along the axes (from left to right and from bottom to top). (b) Motion of the largest eigenvector of steфин B in absence (left) and presence (right) of cathepsin L1.
(TIF)

S10 Fig. Binding interface characterization of Stefin B-Cathepsin L1 complex. (A) Potential energy of interaction between binding interface residues of steфин B (I) & cathepsin L1 (R). Error bars represent the estimated error in GROMACS calculation. (B) Average number of HBs formed among interface residues. Error bars designate standard deviation. (C) Appreciable changes in SASA on complex formation among binding interface residues.
(TIF)

S11 Fig. DSSP analysis of Cystatin C-Cathepsin L1 complex. Secondary structure content of cystatin C in bound (A) and unbound (B) state and that of cathepsin L1 in bound (C) and unbound (D) form.
(TIF)

S12 Fig. RMSD & RMSF analyses for Cystatin C-Cathepsin L1 complex. (A) Average backbone RMSD of the complex, inhibitor (I) and receptor (R) in bound and unbound state. RMSF of cystatin C (B) and cathepsin L1 (C) in complexed form and in free state in solution.
(TIF)

S13 Fig. PCA of Cystatin C-Cathepsin L1 complex. (A) Covariance matrix illustrating correlated and anticorrelated motions of bound (top left) and unbound (bottom right) cystatin C. The secondary structure of cystatin C backbone is represented along the axes (from left to right and from bottom to top). (b) Motion of the largest eigenvector of cystatin C in absence (left) and presence (right) of CL1.
(TIF)

S14 Fig. Binding interface characterization of Cystatin C-Cathepsin L1 complex. (A) Potential energy of interaction between binding interface residues of cystatin C (I) & cathepsin L1 (R). Error bars represent the estimated error in GROMACS calculation. (B) Average number of HBs formed among interface residues. Error bars designate standard deviation. (C) Appreciable changes in SASA on complex formation among binding interface residues.
(TIF)

S15 Fig. DSSP analysis of Cystatin D-Cathepsin L1 complex. Secondary structure content of cystatin D in bound (A) and unbound (B) state & that of cathepsin L1 in bound (C) and unbound (D) form.
(TIF)

S16 Fig. RMSD & RMSF analyses for Cystatin D-Cathepsin L1 complex. (A) Average backbone RMSD of the complex, inhibitor (I) and receptor (R) in bound and unbound state. RMSF

of cystatin D (B) and cathepsin L1 (C) in complexed form and in free state in solution. (TIF)

S17 Fig. PCA of Cystatin D-Cathepsin L1 complex. (A) Covariance matrix illustrating correlated and anticorrelated motions of bound (top left) and unbound (bottom right) cystatin D. The secondary structure of cystatin D backbone is represented along the axes (from left to right and from bottom to top). (b) Motion of the largest eigenvector of cystatin D in absence (left) and presence (right) of cathepsin L1. (TIF)

S18 Fig. Binding interface characterization of Cystatin D-Cathepsin L1 complex. (A) Potential energy of interaction between binding interface residues of cystatin D (I) & cathepsin L1 (R). Error bars represent the estimated error in GROMACS calculation. (B) Average number of HBs formed among interface residues. Error bars designate standard deviation. (C) Appreciable changes in SASA on complex formation among binding interface residues. (TIF)

S19 Fig. DSSP analysis of Cystatin F-Cathepsin L1 complex. Secondary structure content of cystatin F in bound (A) and unbound (B) state & that of cathepsin L1 in bound (C) and unbound (D) form. (TIF)

S20 Fig. RMSD & RMSF analyses for Cystatin F-Cathepsin L1 complex. (A) Average backbone RMSD of the complex, inhibitor (I) and receptor (R) in bound and unbound state. RMSF of cystatin F (B) and cathepsin L1 (C) in complexed form and in free state in solution. (TIF)

S21 Fig. PCA of Cystatin F-Cathepsin L1 complex. (A) Covariance matrix illustrating correlated and anticorrelated motions of bound (top left) and unbound (bottom right) cystatin F. The secondary structure of cystatin F backbone is represented along the axes (from left to right and from bottom to top). (b) Motion of the largest eigenvector of cystatin F in absence (left) and presence (right) of cathepsin L1. (TIF)

S22 Fig. Binding interface characterization of Cystatin F-Cathepsin L1 complex. (A) Potential energy of interaction between binding interface residues of cystatin F (I) & cathepsin L1 (R). Error bars represent the estimated error in GROMACS calculation. (B) Average number of HBs formed among interface residues. Error bars designate standard deviation. (C) Appreciable changes in SASA on complex formation among binding interface residues. (TIF)

S23 Fig. DSSP analysis of Cystatin M/E-Cathepsin L1 complex. Secondary structure content of cystatin M/E in bound (A) and unbound (B) state and that of cathepsin L1 in bound (C) and unbound (D) form. (TIF)

S24 Fig. RMSD & RMSF analyses for Cystatin M/E-Cathepsin L1 complex. (A) Average backbone RMSD of the complex, inhibitor (I) and receptor (R) in bound and unbound state. RMSF of cystatin M/E (B) and cathepsin L1 (C) in complexed form and in free state in solution. (TIF)

S25 Fig. PCA of Cystatin M/E-Cathepsin L1 complex. (A) Covariance matrix illustrating correlated and anticorrelated motions of bound (top left) and unbound (bottom right) cystatin M/

E. The secondary structure of cystatin M/E backbone is represented along the axes (from left to right and from bottom to top). (b) Motion of the largest eigenvector of cystatin M/E in absence (left) and presence (right) of cathepsin L1.

(TIF)

S26 Fig. Binding interface characterization of Cystatin M/E-Cathepsin L1 complex. (A) Potential energy of interaction between binding interface residues of cystatin M/E (I) & cathepsin L1 (R). Error bars represent the estimated error in GROMACS calculation. (B) Average number of HBs formed among interface residues. Error bars designate standard deviation. (C) Appreciable changes in SASA on complex formation among binding interface residues.

(TIF)

S27 Fig. DSSP analysis of Cystatin S-Cathepsin L1 complex. Secondary structure content of cystatin S in bound (A) and unbound (B) state & that of cathepsin L1 in bound (C) and unbound (D) form.

(TIF)

S28 Fig. RMSD & RMSF analyses for Cystatin S-Cathepsin L1 complex. (A) Average backbone RMSD of the complex, inhibitor (I) and receptor (R) in bound and unbound state. RMSF of cystatin S (B) and cathepsin L1 (C) in complexed form and in free state in solution.

(TIF)

S29 Fig. PCA of Cystatin S-Cathepsin L1 complex. (A) Covariance matrix illustrating correlated and anticorrelated motions of bound (top left) and unbound (bottom right) cystatin S. The secondary structure of cystatin S backbone is represented along the axes (from left to right and from bottom to top). (b) Motion of the largest eigenvector of cystatin S in absence (left) and presence (right) of cathepsin L1.

(TIF)

S30 Fig. Binding interface characterization of Cystatin S-Cathepsin L1 complex. (A) Potential energy of interaction between binding interface residues of cystatin S (I) & cathepsin L1 (R). Error bars represent the estimated error in GROMACS calculation. (B) Average number of HBs formed among interface residues. Error bars designate standard deviation. (C) Appreciable changes in SASA on complex formation among binding interface residues.

(TIF)

S31 Fig. DSSP analysis of Cystatin SA-Cathepsin L1 complex. Secondary structure content of cystatin SA in bound (A) and unbound (B) state and that of cathepsin L1 in bound (C) and unbound (D) form.

(TIF)

S32 Fig. RMSD & RMSF analyses for Cystatin SA-Cathepsin L1 complex. (A) Average backbone RMSD of the complex, inhibitor (I) and receptor (R) in bound and unbound state. RMSF of cystatin SA (B) and cathepsin L1 (C) in complexed form and in free state in solution.

(TIF)

S33 Fig. PCA of Cystatin SA-Cathepsin L1 complex. (A) Covariance matrix illustrating correlated and anticorrelated motions of bound (top left) and unbound (bottom right) cystatin SA. The secondary structure of cystatin SA backbone is represented along the axes (from left to right and from bottom to top). (b) Motion of the largest eigenvector of cystatin SA in absence (left) and presence (right) of cathepsin L1.

(TIF)

S34 Fig. Binding interface characterization of Cystatin SA-Cathepsin L1 complex. (A) Potential energy of interaction between binding interface residues of cystatin SA (I) & cathepsin L1 (R). Error bars represent the estimated error in GROMACS calculation. (B) Average number of HBs formed among interface residues. Error bars designate standard deviation. (C) Appreciable changes in SASA on complex formation among binding interface residues. (TIF)

S35 Fig. DSSP analysis of Cystatin SN-Cathepsin L1 complex. Secondary structure content of cystatin SN in bound (A) and unbound (B) state & that of cathepsin L1 in bound (C) and unbound (D) form. (TIF)

S36 Fig. RMSD & RMSF analyses for Cystatin SN-Cathepsin L1 complex. (A) Average backbone RMSD of the complex, inhibitor (I) and receptor (R) in bound and unbound state. RMSF of cystatin SN (B) and cathepsin L1 (C) in complexed form and in free state in solution. (TIF)

S37 Fig. PCA of Cystatin SN-Cathepsin L1 complex. (A) Covariance matrix illustrating correlated and anticorrelated motions of bound (top left) and unbound (bottom right) cystatin SN. The secondary structure of cystatin SN backbone is represented along the axes (from left to right and from bottom to top). (b) Motion of the largest eigenvector of cystatin SN in absence (left) and presence (right) of cathepsin L1. (TIF)

S38 Fig. Binding interface characterization of Cystatin SN-Cathepsin L1 complex. (A) Potential energy of interaction between binding interface residues of cystatin SN (I) & cathepsin L1 (R). Error bars represent the estimated error in GROMACS calculation. (B) Average number of HBs formed among interface residues. Error bars designate standard deviation. (C) Appreciable changes in SASA on complex formation among binding interface residues. (TIF)

S1 Table. Structure files preparation.
(DOCX)

S2 Table. Composition of simulated systems.
(DOCX)

S3 Table. Pairwise RMSD (Å) of the crystal structure (IICF), individual trajectory-averages (CL1A—CL1SN) and global trajectory-average structure (GT) using mainchain atoms.
(DOCX)

S4 Table. Conformational space normalized overlaps between cathepsin L1 (CL1).
(DOCX)

S5 Table. Structural Properties of unbound CL1s.
(DOCX)

S6 Table. Structural properties of complexes as a measure of stability.
(DOCX)

S7 Table. Changes in secondary structure after refinement through MDS.
(DOCX)

S8 Table. Reciprocal arrangements of cystatins in Docked & Refined structures.
(DOCX)

S9 Table. Changes in Binding Interface (BI).
(DOCX)

S10 Table. Flexibility of bound and unbound cystatins.
(DOCX)

S1 Text. Simulation results of Cathepsin L1-Stefin/Cystatin complexes.
(DOCX)

Acknowledgments

We convey our sincere thanks to DBT sponsored Bioinformatics Infrastructure Facility (BIF), University of Kalyani, where the whole study took place. We also acknowledge Shri Rajabrta Bhuyan for his help.

Author Contributions

Conceptualization: SKN AS.

Data curation: SKN.

Formal analysis: SKN.

Investigation: SKN.

Methodology: SKN.

Project administration: SKN.

Resources: AS.

Software: SKN.

Supervision: SKN AS.

Validation: SKN.

Visualization: SKN.

Writing – original draft: SKN.

Writing – review & editing: SKN AS.

References

1. Turk V, Stoka V, Turk D. Cystatins: biochemical and structural properties, and medical relevance. *Front Biosci.* 2008; 13: 5406–5420. PMID: [18508595](#)
2. Barrett J, Fritz H, Grubb A, Isemura S, Jarvinen M, Katunuma N, et al. Nomenclature and classification of the proteins homologous with the cysteine-proteinase inhibitor chicken cystatin. *Biochem J.* 1986; 236: 312. PMID: [3491603](#)
3. Turk V, Stoka V, Vasiljeva O, Renko M, Sun T, Turk B, et al. Cysteine cathepsins: from structure, function and regulation to new frontiers. *Biochim Biophys Acta.* 2012; 1824: 68–88. doi: [10.1016/j.bbapap.2011.10.002](#) PMID: [22024571](#)
4. Bode W, Engh R, Musil D, Thiele U, Huber R, Karshikov A, et al. The 2.0 Å x-ray crystal structure of chicken egg white cystatin and its possible mode of interaction with cysteine proteinases. *EMBO J.* 1988; 7: 2593–2599. PMID: [3191914](#)
5. Stubbs MT, Laber B, Bode W, Huber R, Jerala R, Lenarcic B, et al. The refined 2.4 Å X-ray crystal structure of recombinant human stefin B in complex with the cysteine proteinase papain: a novel type of proteinase inhibitor interaction. *EMBO J.* 1990; 9: 1939–1947. PMID: [2347312](#)

6. Dickinson DP. Salivary (sd-type) cystatins: Over one billion years in the making—but to what purpose?. *Crit Rev Oral Biol Med.* 2002; 13: 485–508. PMID: [12499242](#)
7. Cheng T, Hitomi K, van Vlijmen-Willems IM, de Jongh GJ, Yamamoto K, Nishi K, et al. Cystatin M/E is a high affinity inhibitor of cathepsin V and cathepsin L by a reactive site that is distinct from the legumain-binding site. A novel clue for the role of cystatin M/E in epidermal cornification. *J Biol Chem.* 2006; 281: 15893–15899. doi: [10.1074/jbc.M600694200](#) PMID: [16565075](#)
8. Machleidt W, Thiele U, Assfalg-Machleidt I, Förger D, Auerswald EA. Molecular mechanism of inhibition of cysteine proteinases by their protein inhibitors: kinetic studies with natural and recombinant variants of cystatins and stefins. *Biomed Biochim Acta.* 1991; 50: 613–620. PMID: [1801731](#)
9. Estrada S, Pavlova A, Björk I. The contribution of N-terminal region residues of cystatin A (stefin A) to the affinity and kinetics of inhibition of papain, cathepsin B, and cathepsin L. *Biochemistry.* 1999; 38: 7339–7345. doi: [10.1021/bi990003s](#) PMID: [10353845](#)
10. Pol E, Björk I. Contributions of individual residues in the N-terminal region of cystatin B (stefin B) to inhibition of cysteine proteinases. *Biochim Biophys Acta.* 2003; 1645: 105–112. PMID: [12535617](#)
11. Nikawa T, Towatari T, Ike Y, Katunuma N. Studies on the reactive site of the cystatin superfamily using recombinant cystatin A mutants. Evidence that the QVVAG region is not essential for cysteine proteinase inhibitory activities. *FEBS Lett.* 1989; 255: 309–314. PMID: [2676604](#)
12. Pavlova A, Björk I. The role of the second binding loop of the cysteine protease inhibitor, cystatin A (stefin A), in stabilizing complexes with target proteases is exerted predominantly by Leu73. *Eur J Biochem.* 2002; 269: 5649–5658. PMID: [12423365](#)
13. Björk I, Brieditis I, Raub-Segall E, Pol E, Hakansson K, Abrahamson M. The importance of the second hairpin loop of cystatin C for proteinase binding. Characterization of the interaction of Trp-106 variants of the inhibitor with cysteine proteinases. *Biochemistry.* 1996; 35: 10720–10726. doi: [10.1021/bi960420u](#) PMID: [8718861](#)
14. Pol E, Björk I. Importance of the second binding loop and the C-terminal end of cystatin B (stefin B) for inhibition of cysteine proteinases. *Biochemistry.* 1999; 38: 10519–10526. doi: [10.1021/bi990488k](#) PMID: [10441148](#)
15. Gabrijelcic D, Svetic B, Spaić D, Skrk J, Budihna M, Dolenc I, et al. Cathepsins B, H and L in human breast carcinoma. *Eur J Clin Chem Clin Biochem.* 1992; 30: 69–74. PMID: [1316176](#)
16. Kos J, Smid A, Krasovec M, Svetic B, Lenarcic B, Vrhovec I, et al. Lysosomal proteases cathepsins D, B, H, L and their inhibitors stefins A and B in head and neck cancer. *Biol Chem Hoppe Seyler.* 1995; 376: 401–405. PMID: [7576236](#)
17. Rinne R, Saukko P, Järvinen M, Lehesjoki AE. Reduced cystatin B activity correlates with enhanced cathepsin activity in progressive myoclonus epilepsy. *Ann Med.* 2002; 34: 380–385. PMID: [12452481](#)
18. Sokol JP, Schiemann WP. Cystatin C antagonizes transforming growth factor beta signaling in normal and cancer cells. *Mol Cancer Res.* 2004; 2: 183–195. PMID: [15037657](#)
19. Zhang J, Shridhar R, Dai Q, Song J, Barlow SC, Yin L, et al. Cystatin m: a novel candidate tumor suppressor gene for breast cancer. *Cancer Res.* 2004; 64: 6957–6964. doi: [10.1158/0008-5472.CAN-04-0819](#) PMID: [15466187](#)
20. Li W, Ding F, Zhang L, Liu Z, Wu Y, Luo A, et al. Overexpression of stefin A in human esophageal squamous cell carcinoma cells inhibits tumor cell growth, angiogenesis, invasion, and metastasis. *Clin Cancer Res.* 2005; 11: 8753–8762. doi: [10.1158/1078-0432.CCR-05-0597](#) PMID: [16361563](#)
21. Kos J, Krasovec M, Cimerman N, Nielsen HJ, Christensen IJ, Brünner N. Cysteine proteinase inhibitors stefin A, stefin B, and cystatin C in sera from patients with colorectal cancer: relation to prognosis. *Clin Cancer Res.* 2000; 6: 505–511. PMID: [10690531](#)
22. Park IC, Lee SY, Jeon DG, Lee JS, Hwang CS, Lee SH, et al. Enhanced expression of cathepsin L in metastatic bone tumors. *J Korean Med Sci.* 1996; 11: 144–148. doi: [10.3346/jkms.1996.11.2.144](#) PMID: [8835762](#)
23. Thomssen C, Schmitt M, Goretzki L, Oppelt P, Pache L, Dettmar P, et al. Prognostic value of the cysteine proteases cathepsins B and cathepsin L in human breast cancer. *Clin Cancer Res.* 1995; 1: 741–746. PMID: [9816040](#)
24. Linding R, Jensen LJ, Diella F, Bork P, Gibson TJ, Russell RB. Protein disorder prediction: implications for structural proteomics. *Structure.* 2003; 11: 1453–1459. PMID: [14604535](#)
25. Linding R, Russell RB, Neduva V, Gibson TJ. GlobPlot: Exploring protein sequences for globularity and disorder. *Nucleic Acids Res.* 2003; 31: 3701–3708. PMID: [12824398](#)
26. Yang ZR, Thomson R, McNeil P, Esnouf RM. RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics.* 2005; 21: 3369–3376. doi: [10.1093/bioinformatics/bti534](#) PMID: [15947016](#)

27. Ishida T, Kinoshita K. PrDOS: prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Res.* 2007; 35: W460–W464. doi: [10.1093/nar/gkm363](https://doi.org/10.1093/nar/gkm363) PMID: [17567614](https://pubmed.ncbi.nlm.nih.gov/17567614/)
28. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. *Nucleic Acids Res.* 2000; 28: 235–242. PMID: [10592235](https://pubmed.ncbi.nlm.nih.gov/10592235/)
29. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis.* 1997; 18: 2714–2723. doi: [10.1002/elps.1150181505](https://doi.org/10.1002/elps.1150181505) PMID: [9504803](https://pubmed.ncbi.nlm.nih.gov/9504803/)
30. Nandy SK, Bhuyan R, Seal A. Modelling family 2 cystatins and their interaction with papain. *J Biomol Struct Dyn.* 2013; 31: 649–664. doi: [10.1080/07391102.2012.706403](https://doi.org/10.1080/07391102.2012.706403) PMID: [22881286](https://pubmed.ncbi.nlm.nih.gov/22881286/)
31. Dundas J, Ouyang Z, Tseng J, Binkowski A, Turpaz Y, Liang J. CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Res.* 2006; 34: W116–W118. doi: [10.1093/nar/gkl282](https://doi.org/10.1093/nar/gkl282) PMID: [16844972](https://pubmed.ncbi.nlm.nih.gov/16844972/)
32. Chen H, Zhou HX. Prediction of interface residues in protein-protein complexes by a consensus neural network method: test against NMR data. *Proteins.* 2005; 61: 21–35. doi: [10.1002/prot.20514](https://doi.org/10.1002/prot.20514) PMID: [16080151](https://pubmed.ncbi.nlm.nih.gov/16080151/)
33. Negi SS, Schein CH, Oezguen N, Power TD, Braun W. InterProSurf: a web server for predicting interacting sites on protein surfaces. *Bioinformatics.* 2007; 23: 3397–3399. doi: [10.1093/bioinformatics/btm474](https://doi.org/10.1093/bioinformatics/btm474) PMID: [17933856](https://pubmed.ncbi.nlm.nih.gov/17933856/)
34. Chen R, Li L, Weng Z. ZDOCK: an initial-stage protein-docking algorithm. *Proteins.* 2003; 52: 80–87. doi: [10.1002/prot.10389](https://doi.org/10.1002/prot.10389) PMID: [12784371](https://pubmed.ncbi.nlm.nih.gov/12784371/)
35. Kordiš D, Turk V. Phylogenomic analysis of the cystatin superfamily in eukaryotes and prokaryotes. *BMC Evol Biol.* 2009; 9: 266. doi: [10.1186/1471-2148-9-266](https://doi.org/10.1186/1471-2148-9-266) PMID: [19919722](https://pubmed.ncbi.nlm.nih.gov/19919722/)
36. Chen R, Weng Z. Docking unbound proteins using shape complementarity, desolvation, and electrostatics. *Proteins.* 2002; 47: 281–294. PMID: [11948782](https://pubmed.ncbi.nlm.nih.gov/11948782/)
37. Chen R, Weng Z. A novel shape complementarity scoring function for protein-protein docking. *Proteins.* 2003; 51: 397–408. doi: [10.1002/prot.10334](https://doi.org/10.1002/prot.10334) PMID: [12696051](https://pubmed.ncbi.nlm.nih.gov/12696051/)
38. Pierce B, Weng Z. ZRANK: reranking protein docking predictions with an optimized energy function. *Proteins.* 2007; 67: 1078–1086. doi: [10.1002/prot.21373](https://doi.org/10.1002/prot.21373) PMID: [17373710](https://pubmed.ncbi.nlm.nih.gov/17373710/)
39. Li L, Chen R, Weng Z. RDOCK: refinement of rigid-body protein docking predictions. *Proteins.* 2003; 53: 693–707. doi: [10.1002/prot.10460](https://doi.org/10.1002/prot.10460) PMID: [14579360](https://pubmed.ncbi.nlm.nih.gov/14579360/)
40. Hess B, Kutzner C, Van Der Spoel D, Lindahl E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J chem theo comput.* 2008; 4: 435–447.
41. MacKerell AD Jr, Bashford D, Bellott M, Dunbrack RL Jr, Evanseck JD, Field MJ, et al. An all-atom empirical potential for molecular modeling and dynamics of proteins. *J Phys Chem B.* 1998; 102: 3586–3616. doi: [10.1021/jp973084f](https://doi.org/10.1021/jp973084f) PMID: [24889800](https://pubmed.ncbi.nlm.nih.gov/24889800/)
42. Berendsen HJC, Grigera JR, Straatsma TP. The missing term in effective pair potentials. *J Phys Chem.* 1987; 91: 6269–6271.
43. Leach AR. Energy Minimisation and Related Methods for Exploring the Energy Surface. In: *Molecular Modelling Principles and Applications.* Harlow: Prentice Hall; 2001. pp. 253–302.
44. Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. *J chem phys.* 2007; 126: 14101.
45. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: A linear constraint solver for molecular simulations. *J Comp Chem.* 1997; 18: 1463–1472.
46. Darden T, York D, Pedersen L. Particle mesh Ewald: An N · log (N) method for Ewald sums in large systems. *J Chem Phys.* 1993; 98: 10089–10092.
47. Parrinello M, Rahman A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J Appl phys.* 1981; 52: 7182–7190.
48. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers.* 1983; 22: 2577–2637. doi: [10.1002/bip.360221211](https://doi.org/10.1002/bip.360221211) PMID: [6667333](https://pubmed.ncbi.nlm.nih.gov/6667333/)
49. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Mol Graph.* 1996; 14: 33–38. PMID: [8744570](https://pubmed.ncbi.nlm.nih.gov/8744570/)
50. Caves LS, Evanseck JD, Karplus M. Locally accessible conformations of proteins: multiple molecular dynamics simulations of crambin. *Protein Sci.* 1998; 7: 649–666. doi: [10.1002/pro.5560070314](https://doi.org/10.1002/pro.5560070314) PMID: [9541397](https://pubmed.ncbi.nlm.nih.gov/9541397/)
51. Grossfield A, Feller SE, Pitman MC. Convergence of molecular dynamics simulations of membrane proteins. *Proteins.* 2007; 67: 31–40. doi: [10.1002/prot.21308](https://doi.org/10.1002/prot.21308) PMID: [17243153](https://pubmed.ncbi.nlm.nih.gov/17243153/)

52. Hünenberger PH, Mark AE, Van Gunsteren WF. Fluctuation and cross-correlation analysis of protein motions observed in nanosecond molecular dynamics simulations. *J Mol Biol.* 1995; 252:492–503. doi: [10.1006/jmbi.1995.0514](https://doi.org/10.1006/jmbi.1995.0514) PMID: [7563068](https://pubmed.ncbi.nlm.nih.gov/7563068/)
53. Daura X, Gademann K, Jaun B, Seebach D, van Gunsteren WF, Mark AE. Peptide folding: when simulation meets experiment. *Angewandte Chemie International Edition.* 1999; 38: 236–240.
54. Pol E, Björk I. Role of the single cysteine residue, Cys 3, of human and bovine cystatin B (stefin B) in the inhibition of cysteine proteinases. *Protein Sci.* 2001; 10: 1729–1738. doi: [10.1110/ps.11901](https://doi.org/10.1110/ps.11901) PMID: [11514663](https://pubmed.ncbi.nlm.nih.gov/11514663/)
55. Higgins DG, Sharp PM. CLUSTAL: a package for performing multiple sequence alignment on a micro-computer. *Gene.* 1988; 73: 237–244. PMID: [3243435](https://pubmed.ncbi.nlm.nih.gov/3243435/)