JPP JOURNAL OF PROTEINS AND PROTEOMICS
5(1), January-April 2014, pp. 41-48

Research Article

ASSESSMENT OF LOOP RIGIDIFICATION IN ENZYME-INHIBITOR COMPLEXES: A QUANTITATIVE AND PROBABILISTIC STUDY

Sudip Majumder[1,*] and Rangeet Mitra[2]

[1]Amity School of Applied Sciences and [2]Amity School of Engineering and Technology, Amity University, Haryana, India

Abstract: Enzyme-inhibitor interactions are one of the most important protein-protein or protein-ligand interactions in biological systems. During complex formation with its cognate enzyme, many inhibitory loops undergo structural rigidification whereby molecular motion of the atoms around the otherwise flexible loop gets reduced. B-factor or temperature factor is a good indicator of such rigidification. Here we present a quantitative assessment of the extent of rigidification of loop residues in comparison to the rest of the inhibitor molecule comparing B-factor values in free inhibitor structures and in their corresponding enzyme–inhibitor complexes for serine protease inhibitors (SPI). Our study also reveals that crystal packing artefacts might also play a crucial role in determining the B-factor values of individual amino acid residues in a protein.

Keywords: Enzyme-inhibitor complexes; B-factor; loop rigidification.

Introduction

In biological systems, protein-protein interactions have been an intensely investigated area of research for structural biologists and bio-chemists over the past few decades. There have been diverse types of protein-protein interactions investigated ranging from antigen-antibody to ligand-receptor in different signalling cascades. One of the well-investigated interactions in biological systems is the enzyme-inhibitor interaction.

Enzymes are proteins whose roles are important for performing different functionalities in biological systems. However, the mal-regulation of these enzymes is responsible for several diseases and disfunctionalities in biological systems. The most common way by which nature has developed a defence mechanism against this is the use of protease inhibitors. These naturally occurring inhibitors are also proteins. They act against their respective enzymes through various mechanisms, but the most widely used of them is the canonical standard mechanism, where the inhibitors use a surface exposed loop to interact with the active surface of the enzyme (Schecter and Berger, 1967). In the process, these inhibitors bind to the enzymes in a tight, irreversible manner resulting in the regulation of enzyme functionality.

However, the range of binding constants for such interactions varies depending on the nature of the particular enzyme-inhibitor pair (Dasgupta and Sen, 2011). Thus, trypsin inhibitors show dissociation constants in the range of 10^{-9} M^{-1} indicating strong binding (Song and Suh, 1998), mainly due to the strong electrostatic interaction of Asp189 residue in trypsin pocket with the cognate P1 residue (Lys or Arg) of corresponding inhibitor. On the other hand, for chymotrypsin the value for dissociation constant is of the order of 10^{-5}-10^{-7} M^{-1} suggesting much weak interaction (Kortt, 1981).

Corresponding Author: **Sudip Majumder**
E-mail: sudip22m@gmail.com

Received: November 10, 2013
Accepted: January 10, 2014
Published: April 30, 2014

An interesting aspect of these interactions is the drastic reduction of molecular motion and attainment of rigidity in the complexes at the interface area (Qasim *et al.*, 2006; Hanson *et al.*, 2007). Temperature factor (more popularly called the B-factor) provides a measure of such rigidification. It was shown qualitatively that at least for SPI enzyme-inhibitor complexes, the interacting loop region of the inhibitors undergoes stabilization and rigidification in comparison to the free inhibitor structure (Majumder *et al.*, 2012).

In the present work, we have tried to acquire a quantitative estimation of the extent of rigidification that occurs during the formation of enzyme-inhibitor complexes. We have utilized the database prepared by Majumder *et al.* (2012) in their previous work and calculated the average B-factor value for the loop regions in a free inhibitor structure to that of in its complex with corresponding enzyme. Furthermore, we also compared the B-value of the entire molecule (whether it is a free inhibitor or a complex) to the loop regions, which is the hotspot in enzyme-inhibitor interaction. A quantitative estimation of reduction of molecular motion during complex formation can be useful while investigating enzyme-inhibitor interaction. Our study also highlights that crystal artefacts (symmetry in crystal packing) can play a crucial role in determining the extent of rigidification of inhibitory loop during complex formation.

Materials and Methods

The crystal structures of enzyme-inhibitor complexes and those of free inhibitors were downloaded from the Protein data bank (PDB) and B-factor were calculated using the Temperature factor analysis program from CCP4 suit (CCP4 project, 1994). For further statistical analysis Gaussian Mixture Models (GMM) proposed by Reynolds (Reynolds, 1995) was used.

Gaussian mixture models are ubiquitous in parameter estimation. An arbitrary probability density function (pdf represented by p) can be parametrically represented as a sum of Gaussians with means $\{\boldsymbol{\mu}_j\}$ and covariance matrices Σ_j. Mathematically,

$$p(x) = \sum_{j=1}^{M} w_j \frac{1}{\sqrt{(2\pi)^n |\Sigma_j|}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_j)^T \Sigma_j^{-1}(\mathbf{x}-\boldsymbol{\mu}_j)\right).$$

A popular technique for estimating the weights w_j, the fore-mentioned mean and the covariance matrices, is the Expectation Maximization (EM) algorithm. In EM algorithm, iteration is done as follows:

(1) Estimating w_j by maximising the pdf p $(w_j | \mathbf{x}, \{\boldsymbol{\mu}_j\}, \{\Sigma_j\})$.

(2) Estimating $\boldsymbol{\mu}_j$ and Σ_j using Maximum Likelihood (ML) estimation given the data $\mathbf{x}$ and weights w_j from the previous iteration.

These ML estimates of $\boldsymbol{\mu}_j$ and Σ_j are given by,

$$\Sigma_j = \sum_{i=1}^{N} w_j (\mathbf{x}_i - \boldsymbol{\mu}_j)(\mathbf{x}_i - \boldsymbol{\mu}_j)^T$$

$$\boldsymbol{\mu}_j = \sum_{i=1}^{N} w_j \mathbf{x}_i$$

The samples are indexed by i. The weights in turn are estimated by the Maximum Likelihood (ML) rule,

$$w_j = \frac{\sum_i \frac{1}{\sqrt{(2\pi)^n |\Sigma_j|}} \exp\left(-\frac{1}{2}(\mathbf{x}_i-\boldsymbol{\mu}_j)^T \Sigma_j^{-1}(\mathbf{x}_i-\boldsymbol{\mu}_j)\right)}{\sum_i \sum_j \frac{1}{\sqrt{(2\pi)^n |\Sigma_j|}} \exp\left(-\frac{1}{2}(\mathbf{x}_i-\boldsymbol{\mu}_j)^T \Sigma_j^{-1}(\mathbf{x}_i-\boldsymbol{\mu}_j)\right)}$$

It can be proven that these are indeed the estimators for mean covariance matrix and soft-indicator values.

Generally, one of the main determinants in using such model is deciding as to how many Gaussian clusters are needed for the optimal analysis of dataset. These are generally decided by validation. Optimal numbers of clusters are determined on the basis of the value of Log-Likelihood. However, the sample mean ceases to be a sufficient-statistic for non-Gaussian pdfs. Hence for such situations GMM is an appropriate model of choice.

Results and Discussions

As described above, we calculated B-factor values for the inhibitory loop region of 16 enzyme-

inhibitor complexes from PDB. Our study is focussed primarily on the inhibitors that form complexes with trypsin and the chymotrypsin (i.e SPI's). These two categories are the most prevalent and investigated among all the enzyme-inhibitor complexes. To the best of our knowledge, there are about 22 reported structures of trypsin complexes and 10 reported structures of chymotrypsin complexes. However for our purpose, we could use only those inhibitor complexes where the free inhibitor structure was also available. Thus, we constructed a database of 8 such pairs consisting both free-inhibitor and enzyme-inhibitor complexes. For the derivation of B-values we limited the calculation upto 5 residues on each side of the scissile bond (i.e. from P5 to P5'), because the main interaction of the enzyme active site primarily takes places with these residues of the inhibitory loop and the direct manifestation of interaction effect is most likely to be observed amongst these residues.

Table 1 exhibits the value of the B-factor of the inhibitory loops in comparison to that of entire molecule of the inhibitor, both in free state as well as in complexed state with enzyme. It is evident that in free state, there is negligible systematic pattern in the B-values; in some cases the B-values of the inhibitory loop were even higher than that of the entire inhibitor molecule while in others, the situation is exactly reversed. This type of random behaviour can be explained by the fact that in free state the inhibitory loop is just like any other part of the rest of the molecule. That means, in free state the loop does not have any extra stabilization compared to other parts of the molecule as is expected for a loop.

However, the fifth and sixth column of Table 1 presents a situation that is completely different from the above observation. In the complexes, the average B-value of the inhibitory loop is always significantly lower than that of the average B-value of the entire molecule. This observation clearly suggests the drastic reduction of molecular motion and flexibility of inhibitory loop during complex formation. However, considering the different methods of refinement used for solving the above structures, it seems better to express the quantitative rigidification in terms of some normalized or difference parameters (Parthasarathy and Murthy, 1997). Hence in Table 2, we calculated Δ_1 and Δ_2, the differences in B-factors (Table 2), for quantifying the extent of stabilization of the inhibitory loop in comparison to the entire inhibitor molecule, either in free or complexed state. In this table, the pattern of variation is apparent and unanimous showing the consistently higher value of Δ_2 in comparison to its corresponding Δ_1. The physical significance of such result accounts for the rigid orientation acquired by the inhibitory loop during complex formation that leads to tight and irreversible binding of enzyme-inhibitor systems. Interestingly, comparison of B-values of the inhibitory loop of an inhibitor in free and its complex state reveals that generally, the B-value of the loop region in the free inhibitor is always much greater than that of the complex, as expected with rigidification of the loops in complexed state. However, one exception is the case of 3BYB and 3D65 pair where the free inhibitor (3BYB) showed much lower B-values of inhibitory loop in comparison to its complex counterpart (Figure 2).

While investigating this apparent anomaly, we found that in the 3BYB structure, there is a symmetry molecule near the inhibitory loop (Figure 3). Due to this crystal packing symmetry, in the vicinity of the inhibitory loop (at about 7Å), it is much highly stabilized than expected leading to its lower B-values in comparison to its complex counterpart. However, the higher Δ_2 value in this case also strongly supports our observation that in spite of already stabilized inhibitory loop in the free inhibitor structure, the complex molecule shows sufficient rigidification of aforesaid loop compared to the rest of the molecule. This observation also suggests that crystal-packing artefact (presence of symmetry molecule) can sometimes play a significant role in influencing the individual B-values of various amino acid residues in a protein irrespective of its nature. But the rigidification, attained by the inhibitory loop while forming complex with an enzyme, is independent of any crystal artefact and an intrinsic property of enzyme-inhibitor reaction.

We also made a comparison of B-values of the inhibitory loop in free-inhibitor structure as well as that of in complexes (Table 3). The difference

Table 1
Comparison of B-factor of only inhibitory loop with that of the entire molecule in free inhibitor structure and comparison of B–factor of inhibitory loop with B-factor of entire inhibitor molecule in enzyme-inhibitor complexes

Protein Name	*PDB ID*	*B–factor of inhibitory loop in free inhibitor structure*	*B –factor of entire molecule in free inhibitor structure*	*B–factor of inhibitory loop in enzyme-inhibitor complexes*	*B–factor of entire inhibitor molecule in enzyme-inhibitor complexes*
Trypsin inhibitor from soybean (sti)	1AVU	22.22	35.78		
Complex porcine pancreatic trypsin/ soybean trypsin inhibitor	1AVW			16.6	34.45
The crystal structure of the P1 mutant (Leu to Arg) of a winged bean chymotrypsin inhibitor (Kunitz)	1XG6	38.7	45.95		
A binary complex between bovine pancreatic trypsin and an engineered mutant trypsin inhibitor	3VEQ			37.07	50.1
Crystal structure of a Bowman-Birk inhibitor from *Vigna unguiculata* seeds	2R33	41.18	42.64		
Crystal Structure of the Bowman-Birk inhibitor from *Vigna unguiculata* seeds in complex with beta-trypsin	2G81			16.95	29.67
Bovine pancreatic trypsin inhibitor (BPTI) containing only the [5,55] disulfide bond	2ZJX	10.21	12.99		
Crystal structure of trypsin complexed with BPTI	2FTL			9.44	12.67
Crystal structure of omtky3-ch2-asp19i	1DS3	18.65	16.272		
Crystal and molecular structures of the complex of alpha-*chymotrypsin with its inhibitor turkey ovomucoid third domain	1CHO			12.96	17.82
Protease inhibitor ecotin	1ECY	49.08	41.292		
Crystal structure of a complex between bovine chymotrypsin and ecotin	1N8O			44.05	54.8
Structure of the tryptase inhibitor tdpi from a tick	2UUX	16.7	14.726		
Structure of a tick tryptase inhibitor in complex with bovine trypsin	2UUY			9.96	18.61
Crystal structure of textilinin-1, a Kunitz-type serine protease inhibitor from the Australian common brown snake venom	3BYB	15.5	18.6		
Crystal structure of textilinin-1, a Kunitz-type serine protease inhibitor from the Australian common brown snake venom, in complex with trypsin	3D65			27.32	31.68

Table 2
The Differences of B-values of Inhibitory Loop and B-values of Total Molecule Both in Free and Complex Structures

Protein Name	*PDB ID*	*Δ1=B-Factor of entire inhibitor molecule in free state – B-Factor of inhibitory loop in free state*	*Δ2= B-Factor of entire inhibitor molecule in complex – B-Factor of inhibitory loop in complex*
Trypsin inhibitor from soybean (sti)	1AVU	13.56	
Complex porcine pancreatic trypsin/ soybean trypsin inhibitor	1AVW		17.85
The crystal structure of the P1 mutant (Leu to Arg) of a Wwinged bean chymotrypsin inhibitor (Kunitz)	1XG6	7.25	
A binary complex betwwen bovine pancreatic trypsin and a engineered mutant trypsin inhibitor	3VEQ		13.03
Crystal structure of a Bowman-Birk inhibitor from *Vigna unguiculata* seeds	2R33	1.46	
Crystal Structure of the Bowman-Birk inhibitor from *Vigna unguiculata* seeds in complex with beta-trypsin	2G81		12.72
Bovine pancreatic trypsin inhibitor (BPTI) containing only the [5,55] disulfide bond	2ZJX	2.78	
Crystal structure of trypsin complexed with BPTI	2FTL		3.23
Crystal structure of omtky3-ch2-asp19i	1DS3	2.378	
Crystal and molecular structures of the complex of alpha-*chymotrypsin with its inhibitor turkey ovomucoid third domain	1CHO		4.86
Protease inhibitor ecotin	1ECY	7.78	
Crystal structure of a complex between bovine chymotrypsin and ecotin	1N8O		10.75
Structure of the tryptase inhibitor tdpi from a tick	2UUX	1.974	
Structure of a tick tryptase inhibitor in complex with bovine trypsin	2UUY		8.65
Crystal structure of textilinin-1, a Kunitz-type serine protease inhibitor from the Australian common brown snake venom	3BYB	3.1	
Crystal structure of textilinin-1, a Kunitz-type serine protease inhibitor from the Australian common brown snake venom, in complex with trypsin	3D65		4.36

between these two values does show a spread instead of any consistent number. However, that is expected considering the different resolutions in which the data were collected for different crystal structures and also the various methods used in the respective structure refinements.

To validate our data via probabilistic models, we considered the B-values in Table 1 in free and complex-inhibitor structure. The distributions of B-values of all proteins were modelled as one dimensional GMM (Reynolds, 1995) both in free and complex state. For the B-values of inhibitory loop in free inhibitor structures, our simulation in MATLAB suggested detection of the following clusters with means 16.65 and 42.98 with weights 0.62 and 0.38 respectively. The

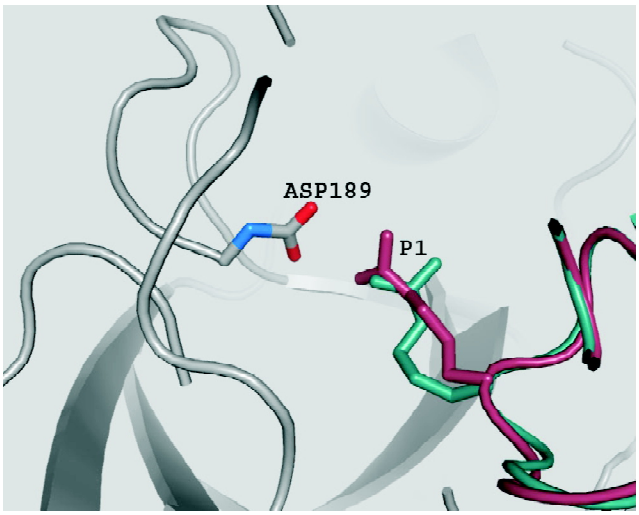

Figure 1: Superposition of soybean trypsin inhibitor (STI) [1AVU] with complex of soybean trypsin inhibitor and porcine pancreatic trypsin [1AVW]. The inhibitory loop of 1AVU is shown in marine blue, whereas that of 1AVW is shown in raspberry red; P1 residue of inhibitory loop (the cognate residue for enzymes recognition) is shown in stick. Trypsin polypeptide in the complex is shown in grey with Asp189 in trypsin pocket shown in stick.

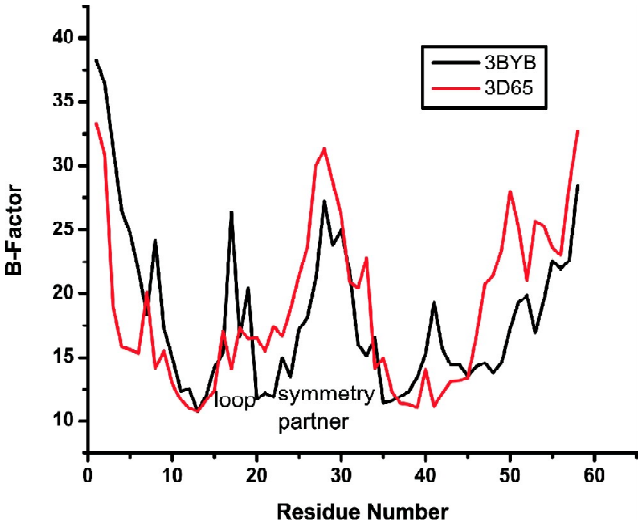

Figure 2: The apparent anomalous B-factor plot of 3BYB and 3D65; 3BYB is shown in black while 3D65 is shown in red. Presence of symmetry molecule just after the loop is pointed in 3BYB B-factor graph, which is responsible for very low B-value near the loop in the free inhibitor structure

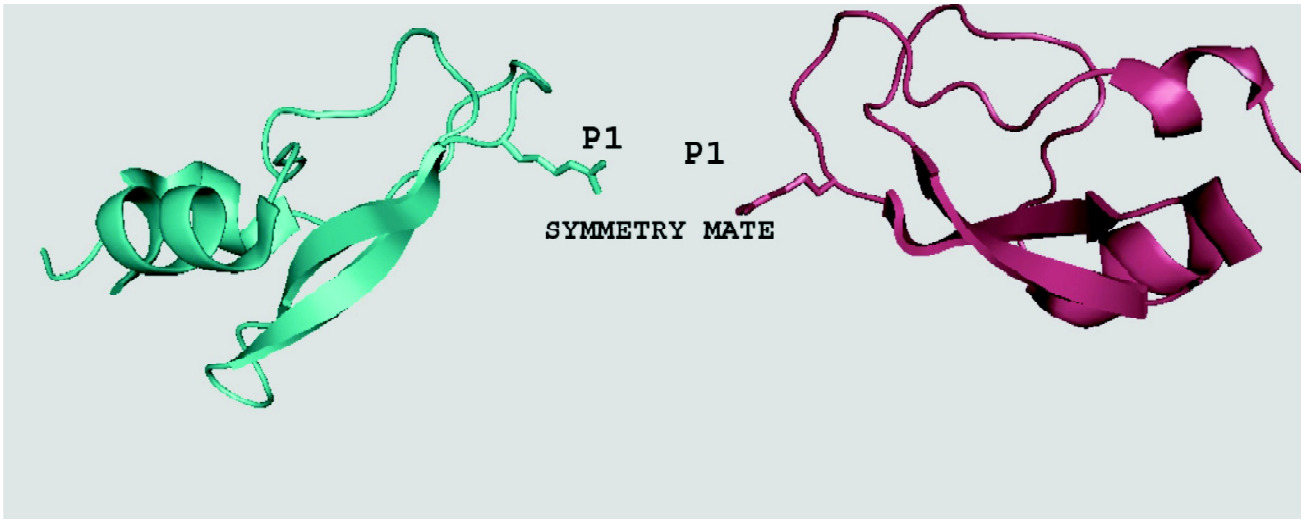

Figure 3: Symmetry molecule pair near the inhibitory loop of 3BYB shown in cyan blue with symmetry pair shown in raspberry red and the P1 residue of the inhibitory loop (the cognate residue for enzyme recognition) is shown in stick.

Table 3
Comparison of B-values of the inhibitory loop in free inhibitor and inhibitor-enzyme complex structures

Protein name	*PDB ID*	*B–factor of inhibitory loop in free inhibitor structure*	*B–factor of inhibitory loop in enzyme-inhibitor complexes*	*Difference between B-values of free inhibitory loop and inhibitory loop in complex*
Trypsin inhibitor from soybean (sti)	1AVU	22.22		
Complex porcine pancreatic trypsin/ soybean trypsin inhibitor	1AVW		16.6	5.6
The crystal structure of the P1 mutant (Leu to Arg) of a Wwinged bean chymotrypsin inhibitor (Kunitz)	1XG6	38.7		
A binary complex betwwen bovine pancreatic trypsin and a engineered mutant trypsin inhibitor	3VEQ		37.07	1.63
Crystal structure of a Bowman-Birk inhibitor from *Vigna unguiculata* seeds	2R33	41.18		
Crystal Structure of the Bowman-Birk inhibitor from *Vigna unguiculata* seeds in complex with beta-trypsin	2G81		16.95	24.23
Bovine pancreatic trypsin inhibitor (BPTI) containing only the [5,55] disulfide bond	2ZJX	10.21		
Crystal structure of trypsin complexed with BPTI	2FTL		9.44	0.77
Crystal structure of omtky3-ch2-asp19i	1DS3	18.65		
Crystal and molecular structures of the complex of alpha-*chymotrypsin with its inhibitor turkey ovomucoid third domain	1CHO		12.96	5.69
Protease inhibitor ecotin	1ECY	49.08		
Crystal structure of a complex between bovine chymotrypsin and ecotin	1N8O		44.05	5.03
Structure of the tryptase inhibitor tdpi from a tick	2UUX	16.7		
Structure of a tick tryptase inhibitor in complex with bovine trypsin	2UUY		9.96	6.74
Crystal structure of textilinin-1, a Kunitz-type serine protease inhibitor from the Australian common brown snake venom	3BYB	15.5		
Crystal structure of textilinin-1, a Kunitz-type serine protease inhibitor from the Australian common brown snake venom, in complex with trypsin	3D65		27.32	11.82

choice of 2 clusters gave the optimal value of Log-Likelihood.

For the inhibitory loop in the inhibitor-complex structures, our simulation of the dataset reported means 35.95 and 13.16 with weights 0.38 and 0.62 respectively. Again choice of two clusters gave the optimal value of Log-Likelihood.

Given this data, the overall mean of the B-value of inhibitory loop in free structures and inhibitory loop in complexes were [0.62*16.65+0.38*42.98]=26.65 and [0.37*35.95 + 0.63*13.16]=21.59 consecutively.

As discussed earlier, the sample mean could/could-not have been a sufficient statistic for the data (depending on whether the data is Gaussian or non-Gaussian). However, as the above GMM based analysis hold true for any arbitrary distribution, it is irrelevant whether the sample mean convey all the information about the unknown pdf; as in the GMM based analysis two Gaussian components with significant weights were found via simulation, indicating a highly non-Gaussian distribution. Results indicate that dearth of data is compensated by the use of an analytical technique optimal for both Gaussian and non-Gaussian distributions.

Hence we found that in complexes, the B-values of the inhibitory loop tend to be low in a probabilistic sense. As GMM can model any distribution, we find that in complex structures, the average B-values of the inhibitory loop tend to be lower than that of inhibitory loop in free structures. Hence, the assumed rigidity of inhibitory loop upon complex formation with enzymes is also supported by probabilistic techniques. It also accounts for the fact that some anomalous lower B-value of inhibitory loop in free inhibitor structures compared to its complex analogue (like the case of 3BYB and 3D65) is solely due to extrinsic artefacts like crystal packing defects due to the presence of symmetry molecules.

Abbreviations

CCP4, Collaborative Crystallography Project 4; SPI, serine protease inhibitor; GMM, Gaussian mixture model; pdf, probability density function; ML, maximum likelihood

References

Collaborative Computational Project, Number 4 (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr Sect D 50, 760–763.

Dasgupta, J. and Sen, U. (2011). Structural aspects of strong inhibition and role of scaffold for serine protease inhibitors. J Proteins Proteomics 2, 125-133.

Hanson, W.M., Domek, G.J., Horvath, M.P. and Goldenberg, D.P. (2007). Rigidification of a high resolution protein structures. J. Mol. Biol. 366, 230–243.

Kortt, A. A. (1981). Specificity and stability of the chymotrypsin inhibitor from winged bean seed (*Psophocarpus tetragonolobus* (L) Dc). Biochim. Biophys. Acta 657, 212-221.

Majumder, S., Khamrui, S., Dasgupta, J., Dattagupta, J.K. and Sen, U. (2012). Role of remote scaffolding residues in the inhibitory loop pre-organization, flexibility, rigidification and enzyme inhibition of serine protease inhibitors. Biochim Biophys Acta 1824, 882-890.

Parthasarathy, S. and Murthy, M.R.N. (1997). Analysis of temperature factor distribution in protein crystallography. Acta Crystallogr Sect D 50, 760–763.

Qasim, M.A., Van Etten, R.L., Yeh, T., Saunders, C., Ganz, P.J., Qasim, S., Wang, L. and Laskowski, Jr.M. (2006). Despite having a common P1 Leu, eglin C inhibits alpha-lytic proteinase a millionfold more strongly than does turkey ovomucoid third domain. Biochemistry 45, 11342-11348.

Reynolds, D.A. (1995). Speaker identification and verification using Gaussian mixture models. Speech Comm. 17, 91-108.

Schecter, I. and Berger, A. (1967). On the size of the active site in proteases. I. papain. Biochem. Biophys. Res. Commun. 27, 157–162.

Song, H.K. and Suh, S.W. (1998). Kunitz-type soybean trypsin inhibitor revisited: refined structure of its complex with porcine trypsin reveals an insight into the interaction between a homologous inhibitor from *Erythrina caffra* and tissue-type plasminogen activator. J. Mol. Biol. 275, 347–363.