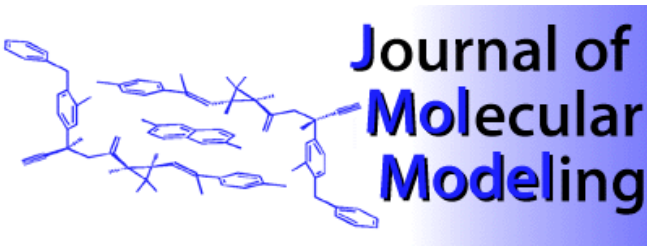
J. Mol. Model. **2000**, *6*, 648 – 653



FULL PAPER

Structural Basis of Inactivation of Thiol Protease by N-Acetyl-p-benzoquinone Imine (NAPQI). A Knowledge-Based Molecular Modeling of the Adduct of NAPQI with Thiol Protease of the Papain Family

Asok K. Pal, Sreya Ghosh, Asim K. Bera, Suparna Bhattacharya, Sibani Chakraborty, Kanai L.Ghatak, and Asok Banerjee

Department of Biophysics, Bose Institute. P1/12 C.I.T. Scheme VIIM, Calcutta-700054. E-mail: ashoke@boseinst.ernet.in

Received: 25 April 2000/ Accepted: 12 July 2000/ Published: 13 December 2000

Abstract Inhibition of hepatic cysteine proteases by non-steroidal anti-inflammatory drug (NSAID) metabolites is implicated in several pathological conditions. It has been reported in the literature that N-acetyl-p-benzoquinone imine (NAPQI), a reactive metabolite of acetaminophen (APAP) can quickly arylate and oxidize thiol (cysteine) protease of the papain family to form an adduct in the pathogenesis of acetaminophen-induced hepatotoxicity. It was also clarified by earlier NMR studies that the 3-position of the aromatic ring (C-3) is the only site of conjugation with cysteinyl thioethers for protein arylation. In a recent study, the adduct of NAPQI has been identified and characterized by LC/MS/MS, LC/NMR and UV spectroscopy, and two possible covalent binding modes corresponding to the 2-position (model-1) and the 3 -position (model-2) of the aromatic ring of NAPQI have been proposed. The work presented here has been initiated to check the structural viability of inhibition for the two proposed adducts at the atomic level. Results of our investigation by computer-assisted molecular modeling structurally demonstrate why model-2 would be more applicable to the static x-ray structure of the complex at physiological pH. This coordinated computational and molecular biology experiment can be used for metabolic screening of NSAIDs. A combinatorial approach of this kind alleviates the doubts in interpreting the results of metabolic function and enhances our insights obtained from either computational or experimental studies alone.

Keywords Knowledge based modeling, Cysteine protease, Inhibitor, Adduct, P450, NAPQI, NSAIDs, Papain, Calotropin DII, Drug design

Introduction

Recently, non-steroidal anti-inflammatory drug (NSAID) metabolites have been drawing special attention for their important role and direct involvement in the inhibition of cysteine (thiol) proteases causing hepatotoxicity. Thiol proteases found in plants, mammalian lysozymes and bacteria form an important group of proteolytic enzymes and participate in a number of physiological processes. Uncontrolled proteolysis of certain cysteine proteases may cause several pathological disorders like inflammation, muscular dystrophy, heart disease, Alzheimer's disease, multiple scle-

Correspondence to: A. Banerjee

rosis, viral diseases, hereditary diseases, tumor malignancy and other health-related disorders. In plants, thiol proteases are involved in the defence mechanism to protect the fruit against fungi, insects and parasites. These enzymes are essential in the life cycle of a number of parasites *e.g. Trypanosoma brucei* (sleeping sickness), *Plasmodium falciparum* (malaria) and *Entamoeba histolytica* (dysentry) and thus become an attractive target for designing specific inhibitors as potent drugs against several such diseases.

Hepatotoxicity caused by large doses of acetaminophen is due to the covalent inhibition of hepatic cysteine proteases under depleted hepatic glutathione levels. N-acetyl-p-benzoquinone Imine (NAPQI), a highly reactive metabolite, is formed by cytochrome P450 oxidation of acetaminophen,[1] a widely used analgesic and antipyretic drug. This metabolite normally reacts with sulfydryl groups in glutathione. However, after large doses of acetaminophen the metabolite is formed in amounts sufficient to deplete hepatic glutathione; under these circumstances, NAPQI quickly oxidizes cysteine protease (Papain family) to form adducts. Imbalances that result in a net decrease in hepatic glutathione level are associated with pathological conditions like hepatic necrosis.[2-4] It was also clarified by earlier NMR studies [5] that the 3-position of the aromatic ring (C-3) of NAPQI is the only site of conjugation with cysteinyl thioethers for protein arylation. In a recent study, the adduct of NAPQI was identified and characterized by LC/MS/MS, LC/NMR and UV spectroscopy, and two possible covalent binding modes corresponding to the 2-position (model-1) and the 3-position (model-2) of the aromatic ring of NAPQI have been proposed.[6, 7]

NAPQI is found to inhibit papain activity markedly and is not only an electrophile that can form covalent adducts with nucleophiles such as glutathione (GSH) and protein thiols, but also a powerful chemical oxidant with a standard reduction potential, E_0 = 0.978± 0.0001V.[8] Based on the chemical reactivity, nucleophilic attack by the thiolate anion of the cysteine protease can take place at either of the two positions, C-1 or C-4, of the aromatic ring of NAPQI leading to the formation of two *ipso* adducts, respectively, the N-acyl-thiohemiaminal adduct and the N-acyl-thiohemiketal adduct (Figure1). Both adducts undergo rapid intramolecular rearrangement by 1,2 shift of the cystein residue leading to the formation of a paracetamol derivative covalently attached at the C-2 and C-3 positions with the cysteine protease, respectively. However, the nucleophilic attack on the aromatic ring of NAPQI must be dictated by the net positive charge developed on the aromatic carbon atoms. It has been shown that the amount of formation of C-2 thiohemiaminal adduct decreases as the basicity increases.[3] The formation of the C-4 thiohemiketal *ipso* adduct as the major product at higher pH values is consistent with the thiolate addition *via* a 'charge controlled' reaction dictated by the high positive net atomic charge (+0.30) at C-4 of NAPQI, with a somewhat lower positive charge (+0.18) at C-1.[9] Since NAPQI reacts with protein thiols to form *ipso* adducts in biological systems, the possible formation and reaction of *ipso* adducts of NAPQI with a typical model cysteine protease, papain, has been investigated structurally by knowledge-based molecular modeling for the two possible modes of inhibition to draw a rational conclusion on the favorable mode of binding. Papain, an endopeptidase, has an active site of length 25 Å consisting of seven subsites (subsites: S1-S4 and S1'-S3'). The active site is relatively accessible to a variety of low molecular weight chemicals.[10] At the optimal pH, which is 6~7, the

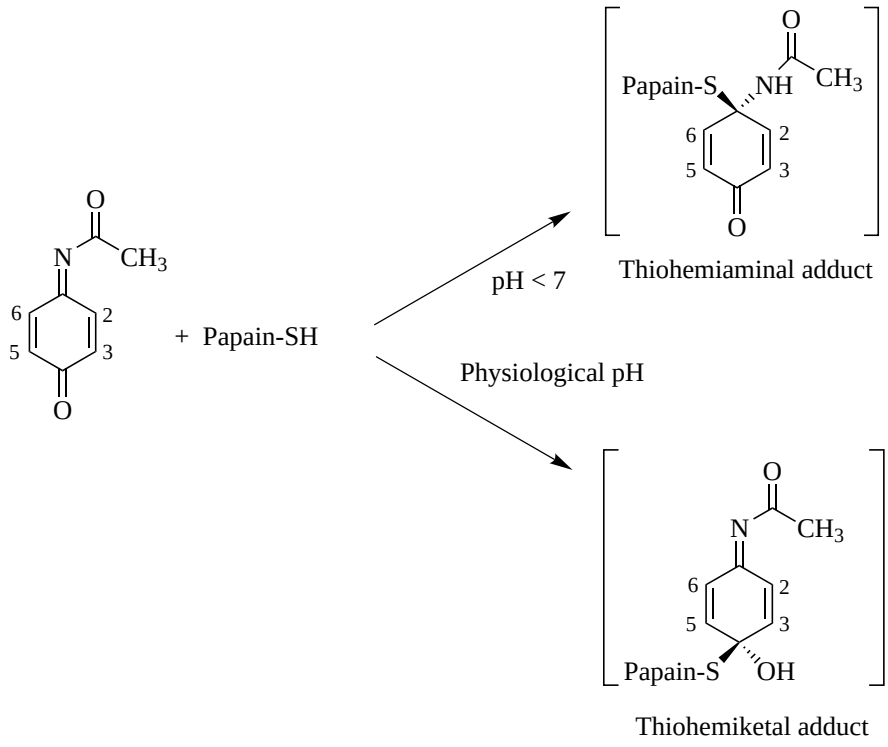


Figure 1 *Schematic diagram of the two possible Papain-NAPQI adducts*

active site of cysteine protease is believed to exist as a zwitterion (Asn···His^{+}···Cys^{-}) with the thiolate anion of the cystein-25 residue functioning as a nucleophile during the initial stage of catalysis.[11] Results of our modeling studies strongly suggest that the adduct, formed by the reaction of NAPQI with the active cysteinyl thiol of this family of enzymes to inactivate them, works in a similar fashion. This is in conformity with the modeling study of the NAPQI adduct with Calotropin DII, another cysteine protease of the papain family, whose x-ray crystallographic structure has been solved recently [12, 13] at 2.1Å resolution in our group.

Although covalent bond formation is a crucial step for revealing the inhibitior activity for both the binding modes, a concrete structural reason for an additional preference for the second binding mode at somewhat higher (physiological) pH values is not clear to date. Our modeling study clearly points out exactly what interactions are responsible for extra stabilization for the second mode of inhibition.

The formation of the thiohemiketal adduct by nucleophilic attack of thiolate anion at the active site of papain has been viewed as a 1,4-conjugate addition reaction (Michael-type addition reaction) and subsequent aromatization leading to the formation of a paracetamol derivative (4-acetyl amino phenol). The bound papain complex (model-2) is shown in Figure 2.

Materials and methods

A full set coordinates of the reported structure of Papain at a resolution of 1.6 Å is available as 1PPP in the Brookhaven protein data bank. The initial atomic coordinates of the two possible models of the paracetamol derivative bound papain complex are made by making a covalent bond between Cys_{25} Sγ with the C-2 aromatic carbon atom for model-1 and the C-3 aromatic carbon atom for model-2 manually. Since the attachment of the inhibitor introduces bad steric interactions into the model, the conformations of the side chains were adjusted manually to relieve bad steric overlap while holding the backbone conformation constant for both the models. To relieve residual strain while maintaining the integrity of the models, the structures were subjected to constrained energy minimization. Molecular modeling was performed with the DISCOVER programs (MSI, San. Diego, InsightII package, version 95.0, CA) using the cvff force field, on a Silicon Graphics IndigoII workstation. First, the total energy was minimized *via* mechanics using a steepest descent algorithm with all main chain atoms fixed. This allows the inhibitor and all the side chains of the protein to relax. The bond length, valence angle and torsional angle terms were taken from the default Discover parameter set. The total empirical energy was derived from the cvff force field. The empirical energy consists of energy terms for bonds, bond angles, dihedral angles, chirality or planarity, hydrogen bonding and non-bonded interaction. The non-bonded interactions included van der Waals' and electrostatic terms, which used the Lennard-Jones and Coulomb functions respectively. The cutoff radius for non bonded interactions was 10 Å. Finally the emergent structures were energy minimized for 300 iterations by the conjugate gradient method to yield the final structures in hand. Both the refined structures were analyzed by PROCHEK.[14] A similar energy minimization was done on Calotropin DII as a cross-verification of the models found and the results are found to be extremely similar. Now the structural basis of two possible binding modes and their relative priorities at different pH levels were explored by careful inspection of the potential interactions and hydrogen bonding networks. Functional implications of the two individual models were compared. Finally, the C-3 carbon atom bound paracetamol-enzyme complex (model-2) was retained as a more favorable model at physiological pH values. The structural guideline on the basis of which model-1 has been given low priority is explained in some detail in the discussion part of this paper. This is in agreement with the conclusion drawn by other workers from kinetic studies.

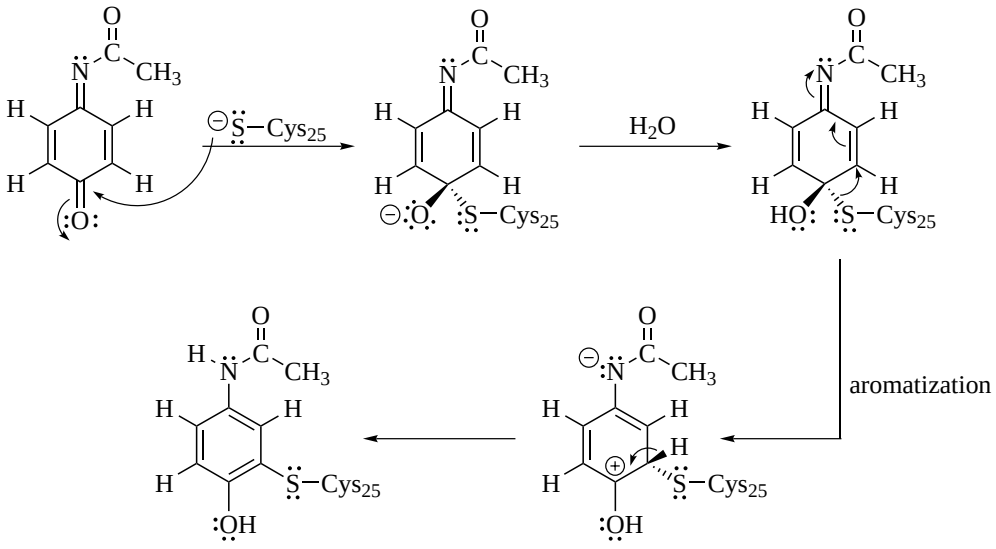


Figure 2 *The proposed reaction mechanism for the formation of covalently bound paracetamol derivative-papain complex (model-2)*

Table 1 *Significant interaction atomic distances (Å) between paracetamol derivative and active site of papain for Model-2*

Type of interaction	Atoms of inhibitor	Residue of Papain	Distance(Å)
Covalent bond	C3	Sg(Cys-25)	1.82
Hydrogen bond	O1	N(Gly-66)	2.91
	O	N(Ala-160)	2.59
Electrostatic short contact	O	Nd^1(His-159)	3.79

Results and discussions

Papain and Calotropin DII have 212 and 215 amino acid residues, respectively, and consist of two domains. In the usual convention followed for papain numbering, one consists of 10-108 and 207-212 residues (L-domain), and another consists of the remaining residues (R-domain). The Cys25 residue of the catalytic center is positioned at the groove created by these two domains. The inhibitor (4-acetyl aminophenol) binds in the active site cavity of papain in the non-primed direction (known as S-subsites), similar to those seen in the crystal structures of papain complexed with several inhibitors like aldehydes, chloromethyl ketones and epoxides. The significant interactions between paracetamol and the active site residues of papain for model-2 are listed in Table 1. Paracetamol is fixed at the active site of papain by covalent bond formation with the Cys_{25} Sγ atom and specific hydrogen bond formations stabilize the molecular conformation in this region. The hydroxyl oxygen of the paracetamol derivative is positioned in the vicinity of the protonated imidazole ring of His_{159} and participates in an electrostatic interaction ($O^{-}\cdots HN^{+}$ = 3.79 Å) and consequently interrupts the participation of the His_{159} residue in the hydrolytic charge-relay system, an essential interaction for inhibitory activity of cysteine proteases of the papain family. This electrostatic interaction requires higher pH values. In model-1, this electrostatic interaction is absent and the corresponding distance is 6.01 Å. This is in agreement with the recent kinetic study, where it is reported that the amount of formation of C-1 thiohemiaminal adduct decreases as the basicity increases, with a concomitant increase in formation of the C-4 thiohemiketal adduct. Formation of the carbon-sulfur bond, which is not easily broken, accounts for the irreversible nature of the adduct (complex). In model-2, the carbonyl oxygen of the paracetamol derivative is hydrogen-bonded (2.91Å) to the backbone nitrogen atom of Gly_{66}, while the hydroxyl oxygen makes a hydrogen bond (2.59Å) to the NH of Ala_{160}. In contrast, in model-1, only the second hydrogen bond is formed with a distance of 2.64Å. A schematic drawing of the significant interactions for both the models is shown in Figure 3 and the interaction distances for model-2 are listed in Table 1. Most of the conformational φ/ψ torsion angles are found within the outer limit boundary of Ramachandran plot [15] for both models. Compared to the structure of papain oxidized at the Sγ atom of Cys_{25}, no significant difference was observed in the secondary structures. A similar result was obtained with the Calotropin DII structure, a member of the papain family. It can therefore be said that the secondary structure of the papain family is stable and is not significantly affected by the binding of the paracetamol derivative. Our modeling study suggests that at somewhat higher (physiological) pH values the nucleophilic attack by the thiolate anion of Cys25 residue takes place at the 4-position of the aromatic ring of NAPQI, forming a N-acyl thiohemiketal adduct, which on molecular rearrangement by 1,2 shift leads to a covalently attached complex at the 3-position. This is probably in order to exploit some extra stability due to electrostatic and hydrogen bonding interaction. Moreover, we have discovered that the presence of these additional interactions

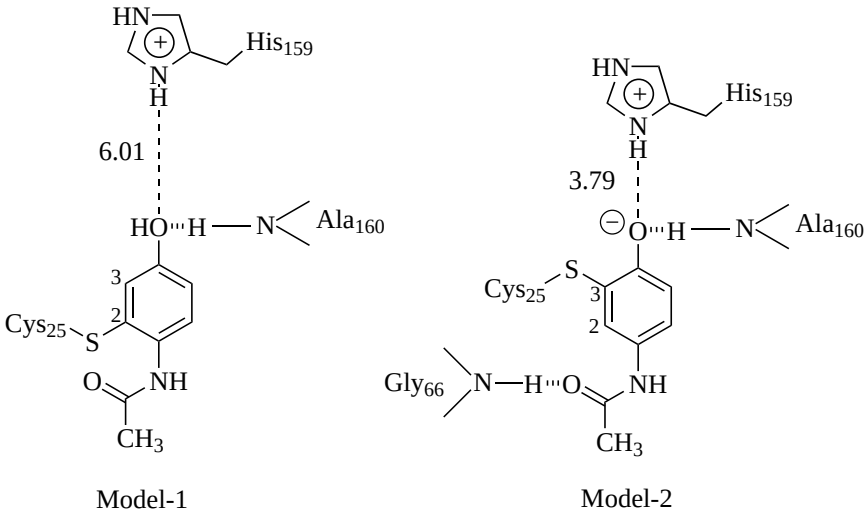


Figure 3 *Schematic drawing showing the interactions between the covalently bound paracetamol derivative and some key residues of papain for both the models. The preferred model-2 has two extra stabilizing interactions over model-1*

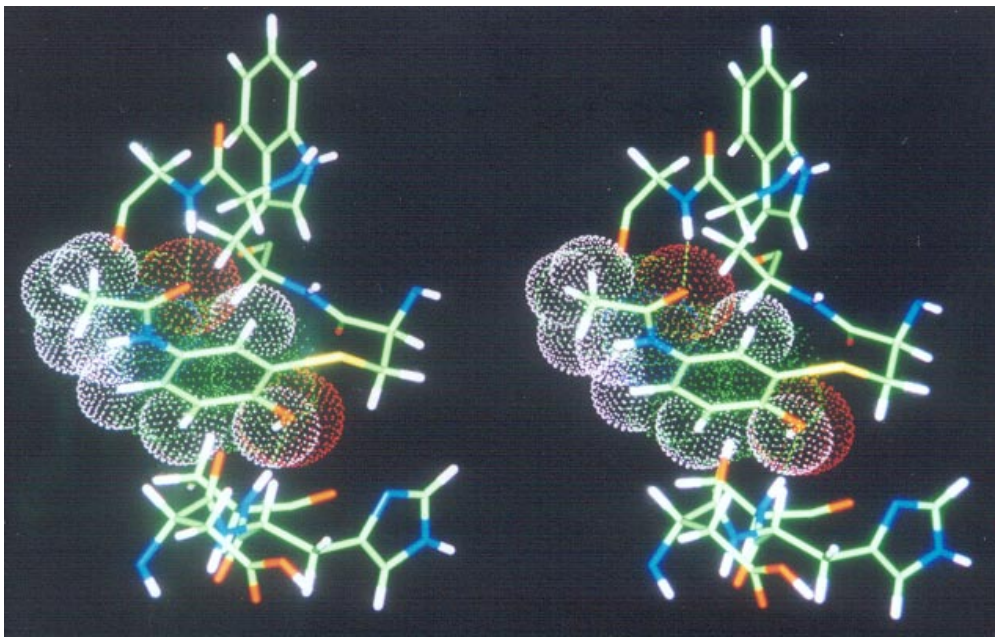


Figure 4 *Stereoscopic view of the interacting zone of the complex (model-2). Paracetamol derivative is shown within dots of van der Waals' radii*

makes model-2 energetically more favorable with a total energy of 3792.42 kcal compared to 5779.13 kcal for model-1. This is consistent with the results obtained from kinetic studies that at higher pH values the thiolate attack at C-4 is due to the development of a high positive net charge (+0.30) at C-4 of NAPQI compared to a somewhat lower net positive charge (+0.18) at C-1.[9] A colored stereo view of the complex (model-2) is shown in Figure 4.

Conclusion

In the absence of a crystal structure, the computer aided molecular modeling of a protein-inhibitor complex has been exploited to shed insight into the molecular aspects of the structure that might result in a high inhibitory potential. The results obtained not only decisively contribute to the establishment of the standard mechanism of inhibition, they also provide several clues for the understanding of protein-inhibitor interactions in general. Knowing the site of covalent linkage, the first step to developing an understanding of the potential non-covalent anticipated interactions is to identify only amenable side chains of amino acid residues with reactive/interacting functional groups while maintaining nearly the same overall conformation of the parent protein in spite of the presence of the inhibitor. Hence in carrying out modeling studies, one needs to conserve the basic structural motif (units) of the protein and ascertain that the alteration introduced by the site-specific events are very limited.

Knowledge-based molecular modeling is an effective method for calculating 3-D atomic coordinates of a protein-inhibitor complex from which structural basis of inhibition can be highlighted. Energy minimization is an essential part of the modeling process and brings the initial model to an optimal, energetically favorable conformation. However, manual intervention and structural chemical intuition are very important and complement the computational approach of molecular modeling.

A literature search showed that the cysteine protease inhibitory activity of N-acetyl-p-benzoquinone imine (NAPQI), a reactive metabolite of acetaminophen, has two probable binding modes of inhibition of cysteine proteases depending on different experimental pH conditions. Precise information about the molecular conformation of two individual modes is not yet available because of difficulties associated with crystallization and unavailability of x-ray structure. But the higher preference for the availability of model-2 at physiological pH is already documented from kinetic studies. This naturally draws special attention to this interesting problem.

The inhibitory mechanism has been investigated extensively at the atomic level by computer-assisted molecular modeling as a complementary approach to kinetic studies to understand the differences in inhibition from the structural point of view for the two proposed binding modes under different experimental conditions. The results of our modelling study seem to be very convincing to favor the second binding mode additionally.

Finally, until the 3-D structure of the papain-paracetamol complex is determined by x-ray/NMR and refined at atomic resolution, the knowledge-based molecular modeling approach utilizes the principle of complementation, and the model so obtained serves as the next-best structural reference for understanding structure-function correlation of the protein and the inhibitor. Furthermore, the x-ray crystallographic structure determination of the co-complex of a papain family member (Calotropin DII) with a paracetamol derivative is underway. If the force field employed is sufficiently accurate, the model obtained can complement the

static x-ray structure and can provide active site geometry differences of papain for the two probable binding modes.

Acknowledgement The authors acknowledge a grant from DBT, Govt. of India, during the research work. A.K.P acknowledges the support of the Principal and his colleagues of R.K.M.V.C.College.

Coordinates Atomic coordinates have been deposited in the Brookhaven Protein Data Bank. The corresponding accession number is **1EFF**.

Abbreviations

APAP: Acetaminophen
NSAID: Non steroidal anti inflammatory drug
NAPQI: N-acetyl-p-benzoquinone Imine
LC/MS/MS: Liquid chromatography/ tandem mass spectrometry.
LC/NMR: Liquid chromatography/ NMR spectroscopy.

References

1. Dahlin, D. C.; Miwa, G. T.; Lu, A. Y. H.; Nelson, S. D. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 1327-1331.
2. Hinson, J. A. *Rev. Biochem. Toxicol.* **1980**, *12*, 103-130.
3. Nelson, S. D. *Semin. Liver Disease* **1990**, *10*, 267-278.
4. Prescott, L. F. *Paracetamol(Acetaminophen): A critical bibliographic Review*; Taylor & Francis Ltd, London, 1996; pp 285-351.
5. Hinson, J. A.; Monks, T. J.; Hong, M.; Highet, R. J.; Pohl, L. R. *Drug Metab. Dispos.* **1982**, *10*, 47-50.
6. Hoffman, K.; Streeter, A. J.; Axworthy, D. B.; Baillie, T. A. *Mol. Pharmacol.* **1985**, *27*, 566-573.
7. Chen, W.; Shockcor, J. P.; Tonge, R.; Hunter, A.; Gartner, C.; Nelson, S. D. *Biochemistry* **1999**, *38*, 8159-8166.
8. Novak, M.; Pelecanou, M.; Pollack, L. *J. Am. Chem. Soc.* **1986**, *108*, 112-120.
9. Loew, G. H.; Goldblum, A. *Mol. Pharmacol*. **1985**, *27*, 375-386.
10. Kamphuis, I. G.; Drenth, J.; Baker, E. N. *J. Mol. Biol.* **1985**, *182*, 317-329.
11. Lewis, S. D.; Johnson, F. A.; Shafer, J. A. *Biochemistry* **1981**, *20*, 48-51.
12. Ghosh, S.; Mukhopadhyay, B. P.; Pal, A. K.; Seal, A.; Bera, A.; Dey, I.; Sinha, N. K.; Banerjee, A. International seminer-cum-school on macromolecular crystallographic data, SINP, Calcutta, Nov 1995.
13. Ghosh, S.; Bera, A.; Pal, A. K.; Bhattacharya, S.; Mukhopadhyay, B. P.; Halder, U.; Chakraborty, S.; Banerjee, A. Discussion meeting on macromolecular crystallography, IISc, Bangalore, Feb 1998.
14. Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. *J. Appl. Cryst.* **1993**, *26*, 283-291.
15. Ramakrishnan, C.; Ramachandran, G. N. *Biophys. J*. **1965**, *5*, 909-933.