JOURNAL OF CHEMICAL EDUCATION

Laboratory Experiment

pubs.acs.org/jchemeduc

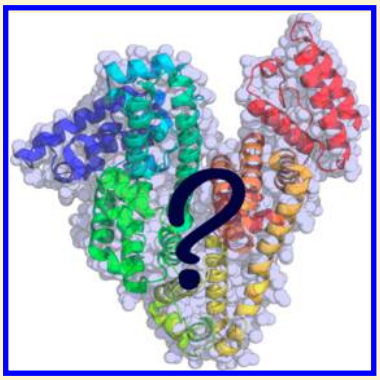
Modeling Human Serum Albumin Tertiary Structure To Teach Upper-Division Chemistry Students Bioinformatics and Homology Modeling Basics

Dušan Petrović and Mario Zlatović*

Faculty of Chemistry, University of Belgrade, Studentski trg 12-16, 11000 Belgrade, Serbia

Supporting Information

ABSTRACT: A homology modeling laboratory experiment has been developed for an introductory molecular modeling course for upper-division undergraduate chemistry students. With this experiment, students gain practical experience in homology model preparation and assessment as well as in protein visualization using the educational version of PyMOL state-of-the-art molecular graphics. Students create a human serum albumin homology model with relatively high resolution at 1.77 Å (heavy atom model) and with reasonable values for bond lengths, angles, and dihedrals. The suggested tasks integrate different fundamental aspects of structural biology and protein modeling. Ramachandran plots and side chain rotamers are discussed. The assignments are shown to be good not only to teach homology modeling basics, but also to introduce several other concepts of structural bioinformatics such as protein sequence data mining, basic local alignment search tool (BLAST), and multiple sequence alignment. Homology modeling is demonstrated to be a cornerstone in molecular docking and drug design projects if the crystal structure of the protein of interest is not yet determined.



KEYWORDS: *Upper-Division Undergraduate, Biochemistry, Chemoinformatics, Computer-Based Learning, Laboratory Instruction, Medicinal Chemistry, Molecular Modeling, Proteins/Peptides*

Throughout the chemistry curriculum at the University of Belgrade in Serbia, students are exposed to various applications of computers in chemistry. A molecular modeling and chemoinformatics course is taught as an elective upper-division undergraduate course for chemistry students, mainly designed for students pursuing a degree in experimental organic chemistry. Since this is the first computational chemistry course focused on applications of computers in the life sciences, the course material covers a broad range of themes in modern molecular modeling. The main goal of this course is to introduce students to various aspects of computational chemistry, with a particular emphasis on computer-aided drug design and computer-aided molecular modeling. In this course, students are confronted with different *ab initio* and semi-empirical quantum mechanical (QM) methods applied to organic chemistry as well as molecular dynamics and docking simulations related to drug design. An important part of the course deals with the structure and activity of proteins with medicinal applications.

Basic protein modeling techniques have been described for undergraduate laboratories in this and other educational journals.[1−9] Although there are numerous reports of protein visualization, molecular docking, and molecular dynamics (MD) simulations as well as quantitative structure−activity relationship (QSAR) studies, there appear to be only few reports of homology modeling[2a−d] and structural bioinformatics[4a−c] lab assignments. Both experimental and computational methods have advanced in the last ten years, which have enabled homology modeling to become a more accurate method. Also, with the development of computers, tasks that required high computational resources ten years ago can be accomplished today on low-priced laptops or even cell phones and tablet devices. There is a need for an experiment for students involving homology modeling with a more contemporary approach.

Although the homology modeling workflow is common among published educational material,[2a−d] a wide variety of different programs was used. Some of the suggested software automates several steps of the protocol, which potentially could lead to not fully understanding the crucial steps in homology modeling. Furthermore, all assignments, including the present one, significantly differ in the amount and type of suggested analysis of the built homology models. Similar experiments and procedures are available via various Internet resources,[10−16] but most of them require additional adaptation for teaching purposes. Some of them deal with aspects of protein modeling that exceed the scope of a modeling course dedicated to students pursuing a degree in experimental organic chemistry and thus demand more time and resources. Therefore, a molecular modeling experiment was designed to cover critical points in homology modeling that can be completed using readily available resources and within a reasonable time, which

DOI: 10.1021/ed500358f
J. Chem. Educ. XXXX, XXX, XXX−XXX

would leave enough time for discussion on specific problems in protein structure and modeling.

HOMOLOGY MODELING

Knowledge of the tertiary structure can be essential for drug design or protein affinity chromatography purification. With the development of experimental methods, a total number of three-dimensional (3D) protein structures determined by X-ray crystallography has shown significant growth: from 1,582 PDB entries in 1993 to over 23,500 in 2003 to over 105,000 in 2015.[17] However, 3D structures of many proteins have still not been determined. Although mechanisms of protein folding are not yet completely known,[18] a structure of a protein can be predicted using molecular modeling. De novo protein structure prediction is generally very complex due to the high number of degrees of freedom and large conformational space; accurate predictions using this method have been obtained only for small proteins.[19] Homology modeling is less dependent on protein size, rather it depends on the percentage of the sequence identity to a template structure.[20]

Evolutionarily related proteins, usually with quite similar sequences, are called homologous. Their sequence and function similarities are usually related to the similarity in their stable tertiary structures.[21] Homology, or comparative, protein modeling predicts the 3D structure of one protein by using the sequence similarities with its homologous proteins.[22] The safe zone for homology modeling is a prediction of the proteins with >150 residues and sequence identity >50%. Structure prediction of small proteins (<100 residues) and with sequence identity <30% is usually considered a twilight zone; hence, it is not recommended for model preparation. For homology models where sequence identity is higher than 90%, accuracy of the model can usually be compared to high resolution experimental structures. If the sequence identity is between 50−90%, root−mean−square deviation of the atomic coordinates (RMSD) can be around 1.5 Å. If the sequence identity drops below 30%, errors are quite large, and a structure is generally unusable.[23]

LABORATORY EXPERIMENT RATIONALE

In the present experiment, students prepare a homology model of human serum albumin (HSA) as if its tertiary structure was not known. The choice of HSA is based on its importance in the human organism, previous students' knowledge of serum proteins and HSA structure (from an introductory biochemistry course), and the possibility to produce good models without alignment adjustment and several other details. This experiment also provides an opportunity to summarize knowledge of protein structure and the specific protein chosen for the modeling. Some additional suggestions on the target for modeling and variations in template are in the instructor notes section of the Supporting Information.

The purpose of this laboratory experiment is to cover three themes: protein structure, homology modeling, and software tools. The key learning goals for each theme are given in Table 1.

Some aspects of protein structure were covered in an introductory biochemistry course, and thus, protein structures related to the problems and procedures in homology modeling are summarized during the work of this experiment. This experiment is intended to aid students' learning process by guiding them through the main phases of the protein modeling.

Table 1. Key Learning Goals Divided in Three Groups

Protein Structure	Homology Modeling	Software Tools
Sources of information on proteins and their structures	Multiple sequence alignments	Databases and data mining
Conformational properties of proteins	Conserved and variable regions	BLAST, Clustal
Types of possible secondary structures	Prediction of secondary structures	Modeling resources
Homologous proteins	Template based protein modeling	PyMOL (or similar protein visualization and handling tool)
	Model refinement, optimization, and validation	
	Scopes and limitations of protein models	

This course is designed primarily for students pursuing a degree in experimental organic chemistry, so a highly theoretical background is omitted. The idea of both the course and this experiment is not to educate students to be molecular modelers, rather it is to provide information on basic techniques in molecular modeling. For this, a simple model is used, without additional alignment adjustment, loop remodeling, structure relaxation, and several other details, although those procedures were discussed. However, even without including these latter procedures in the modeling, relatively good models are prepared.

The models are prepared based on the structure of a similar serum albumin (typically from horse). Students further inspect validity of their models through Ramachandran plots, side chain rotamers, and deviations in bond lengths and angles. Finally, students perform benchmarks against crystal structures of HSA deposited in the PDB to determine the resolutions of their models. Students not only learn how to prepare basic homology models, but also they master several bioinformatics tools as well as protein visualization through PyMOL[24] software. Furthermore, students are encouraged to discuss protein structural features so they could revise and deepen their knowledge of the introductory biochemistry course.

EXPERIMENT

Placement in the Course and Organization

The molecular modeling and chemoinformatics course is taken after biochemistry and organic chemistry courses. It is designed to have introductory theoretical lectures followed by laboratory experiments. Laboratory experiments are held in 4-h sessions. This arrangement allows use of the time available for the lab experiments in the best way without additional elaboration of some aspects of the general subject already covered in the lectures. With a current student-to-instructor ratio of around 5:1, 4 h are more than enough for successful completion of the experiment with time for discussion of key techniques, concepts, and theory behind them.

For the molecular modeling experiment, each student has access to a desktop computer equipped with Internet access, educational-use-only PyMOL version, and Java software. As software requirements are minimal, students can repeat the same or similar tasks from their homes. The experiment can be done in pairs, if needed, and some parts can be given as homework.

DOI: 10.1021/ed500358f
J. Chem. Educ. XXXX, XXX, XXX−XXX

Overview of the Experiment

The experiment consists of seven sections. In the first section, students use the UniProt database[25] to find the HSA protein sequence. After query for the "Human Serum Albumin" at the UniProt database, among all of received answers, students are directed to the sequence of serum albumin from *Homo sapiens*. Retrieved structural and functional properties of HSA are discussed, and the sequence of mature HSA (without signal and propeptide) is downloaded in the FASTA format. The instructor can present a short discussion about the structure of HSA and its role in humans.

Students find proteins homologous to HSA using BLAST[26] in the second section and proceed to multiple sequence alignment using Clustal Omega[27] in the third section. At this point, the instructor and students discuss sequence similarity, role of conserved regions in protein function, quality of selected template for modeling, and alignment problems.

In the fourth section, students prepare a selected PDB file (4F5U[28]) in PyMOL and build a homology model using SWISS-MODEL[29] in the fifth section. Students usually obtain reasonable and uniform models if good template structures are chosen and the procedure in the Supporting Information is followed. In this part, students learn about preparing crystal structures for molecular modeling purposes, structural problems, and ambiguities that can be found in them. Homology modeling and available tools for it are discussed.

In the sixth section, protein models are analyzed using RAMPAGE,[30] VADAR,[31] and MolProbity.[32] The most descriptive tool is visualization of the Ramachandran plot. Allowed regions, exposure of polar and nonpolar residues, protein folding, and rotamers are discussed with students.

Finally, in the seventh section, students perform HSA homology model benchmarking against the experimentally determined HSA structures, available at the PDB database, without (1E78[33]) and with (1GNI[34]) cocrystallized ligands. HSA model and benchmark structures are aligned in PyMOL, and the RMSD values are calculated.

The effects of the cocrystallized ligand present in the tertiary structures and conformational flexibility of proteins are discussed with students. Modeling of holoenzymes and apoenzymes (holoenzyme without a coenzyme) is also discussed.

Further suggestions are given in the Supporting Information, where a step-by-step lab manual for this laboratory experiment is provided.

RESULTS AND DISCUSSION

The experiment was performed for three consecutive years, and 22 students in total completed this particular procedure. This experiment allowed undergraduate students pursuing a degree in experimental organic chemistry to gain practical experience in homology modeling and to supplement knowledge on protein tertiary structures. The suggested tasks were easily completed during one 4-h lab session, and they integrated many fundamental aspects of protein modeling and visualization. The human serum albumin homology model generated in this experiment (Figure 1) differed from the PDB-deposited structure by only 1.77 Å (heavy atom model) and had reasonable values for bond lengths, angles, and dihedrals. Although more accurate models can be prepared, this was only a basic introduction to homology modeling; hence, it clearly and easily depicted all aspects of the method at the desired level. After the experiments were completed, students were able to use different bioinformatics tools as well as several tools for investigating structural characteristics of proteins.

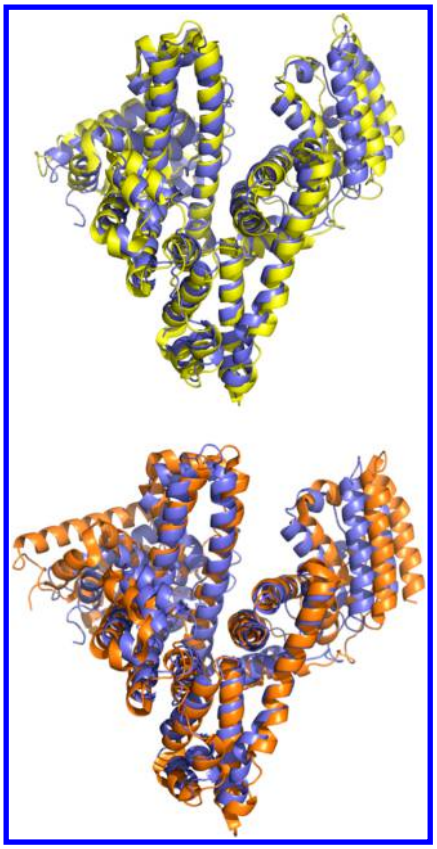


Figure 1. Alignment of the prepared HSA homology model, shown in blue, with (top) 1E78 structure of the HSA without cocrystallized ligand, shown in yellow; and (bottom) 1GNI structure of the HSA, with cocrystallized *cis*-9-octadecenoic acid, shown in orange. Figures were graphically adjusted from the students' data.

An effort was made to gain insight into eventual improvement of student knowledge about homology modeling and protein structure by benchmarking test answers of students having this lab assignment to answers from the first generation students taking this course as a control group (at that time, the bioinformatics and homology modeling assignments were not performed according to this procedure). Improvements were found in answers, mostly on questions concerning protein structure and understanding the principles of comparative modeling. For instance, there were about 20% more correct answers concerning Ramachandran diagrams, 15% more correct answers about amino acid conformations, and 15% more correct answers on structurally conserved and variable regions and their significance. Furthermore, answers were more precise and illustrated a better understanding of the subject.

Homology modeling and molecular docking can be useful for organic chemistry majors as a first step in drug design projects. The feedback from the students after performing this lab assignment was extremely positive. Students not only found the experiment fun, educational, and useful, but also their answers on a final examination and the knowledge shown during the discussions indicated a better understanding of the importance, general principles, possibilities, and limitations of homology modeling and of protein structure as well, which thus fulfilled the main objectives for introducing this experiment.

DOI: 10.1021/ed500358f
J. Chem. Educ. XXXX, XXX, XXX−XXX

■ ASSOCIATED CONTENT

Supporting Information

Step-by-step experiment manual. This material is available via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: mario@chem.bg.ac.rs.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors would like to thank the students of the molecular modeling and chemoinformatics course at the University of Belgrade in Serbia for their valuable feedback and the departmental colleagues for their support and valuable critiques.

■ REFERENCES

(1) Wolfson, A. J.; Hall, M. L.; Branham, T. R. An integrated biochemistry laboratory, including molecular modeling. *J. Chem. Educ.* **1996**, *73* (11), 1026−1029.

(2) (a) León, D.; Uridil, S.; Miranda, J. Structural analysis and modeling of proteins on the Web: An investigation for biochemistry undergraduates. *J. Chem. Educ.* **1998**, *75* (6), 731−734. (b) Centeno, N. B.; Villà-Freixa, J.; Oliva, B. Teaching structural bioinformatics at the undergraduate level. *Biochem. Mol. Biol. Educ.* **2003**, *31* (6), 386−391. (c) Badotti, F.; Barbosa, A. S.; Reis, A. L. M.; do Valle, I. F.; Ambrósio, L.; Bitar, M. Comparative modeling of proteins: A method for engaging students' interest in bioinformatics tools. *Biochem. Mol. Biol. Educ.* **2014**, *42* (1), 68−78. (d) McDougal, O. M.; Cornia, N.; Sambasivarao, S. V.; Remm, A.; Mallory, C.; Oxford, J. T.; Maupin, C. M.; Andersen, T. Homology modeling and molecular docking for the science curriculum. *Biochem. Mol. Biol. Educ.* **2014**, *42* (2), 179−182.

(3) Dabrowiak, J. C.; Hatala, P. J.; McPike, M. A molecular modeling program for teaching structural biochemistry. *J. Chem. Educ.* **2000**, *77* (3), 397−400.

(4) (a) Kossida, S.; Tahri, N.; Daizadeh, I. Bioinformatics by example: From sequence to target. *J. Chem. Educ.* **2002**, *79* (12), 1480−1485. (b) Inlow, J. K.; Miller, P.; Pittman, B. Introductory bioinformatics exercises utilizing hemoglobin and chymotrypsin to reinforce the protein sequence structure−function relationship. *Biochem. Mol. Biol. Educ.* **2007**, *35* (2), 119−124. (c) Likić, V. A. Computer programming and biomolecular structure studies: A step beyond internet bioinformatics. *Biochem. Mol. Biol. Educ.* **2006**, *34* (1), 1−4.

(5) Lamberti, V. E.; Fosdick, L. D.; Jessup, E. R.; Schauble, C. J. C. A hands-on introduction to molecular dynamics. *J. Chem. Educ.* **2002**, *79* (5), 601−606.

(6) Ramos, M. J.; Fernandes, P. A.; Melo, A. Modeling chemical and biological systems: A successful course for undergraduate students. *J. Chem. Educ.* **2004**, *81* (1), 72−75.

(7) Carvalho, I.; Borges, Á. D. L.; Bernardes, L. S. C. Medicinal chemistry and molecular modeling: An integration to teach drug structure−activity relationship and the molecular basis of drug action. *J. Chem. Educ.* **2005**, *82* (4), 588−596.

(8) Ship, N. J.; Zamble, D. B. Analyzing the 3D structure of human carbonic anhydrase II and its mutants using Deep View and the Protein Data Bank. *J. Chem. Educ.* **2005**, *82* (12), 1805−1808.

(9) Sutch, B. T.; Romero, R. M.; Neamati, N.; Haworth, I. S. Integrated teaching of structure-based drug design and biopharmaceutics: A computer-based approach. *J. Chem. Educ.* **2012**, *89* (1), 45−51.

(10) Experiment-10: Homology Modeling. *Sakshat Virtual Labs*; NMEICT of MHRD: India, 2015. http://iitb.vlab.co.in/?sub=41&brch=118&sim=657&cnt=1 (accessed January 2015).

(11) Martz, E. *Comparative ("Homology") Modeling for Beginners with Free Software*, 2001. http://proteinexplorer.org/homolmod.htm (accessed January 2015).

(12) *Homology Modelling*. http://www.pdg.cnb.uam.es/cursos/bcn05/Structures/3D_Practicals/P_homology/index.html (accessed January 2015).

(13) *Swiss-PdbViewer-Tutorial: Homology Modelling*; Swiss Institute of Bioinformatics: Lausanne, Switzerland, 2015. http://spdbv.vital-it.ch/modeling_tut.html (accessed January 2015).

(14) Bordoli, L. *Homology Modeling*. http://edu.isb-sib.ch/file.php/57/HM.htm (accessed January 2015).

(15) *Homology Modeling*. http://www.cs.huji.ac.il/~fora/81855/exercises/ex6.pdf (accessed January 2015).

(16) *Predicting protein structure—Homology modeling exercises*; BITS VIB: Gent, Belgium, 2012. http://wiki.bits.vib.be/index.php/Predicting_protein_structure_-_homology_modeling_exercises (accessed January 2015).

(17) The following search information was used: Search query: "Macromolecule type: contains protein = Yes AND Experimental method: X-ray AND Has Experimental Data = Yes AND Release Date (choose appropriate period)". *RCSB Protein Data Bank. Advanced Search Interface*; RCSB PDB: Washington, DC, 2015. http://www.rcsb.org/pdb/search/advSearch.do?search=new (accessed January 2015).

(18) Dill, K. A.; MacCallum, J. L. The protein-folding problem, 50 years on. *Science* **2012**, *338* (6110), 1042−1046.

(19) Kim, D. E.; Blum, B.; Bradley, P.; Baker, D. Sampling bottlenecks in de novo protein structure prediction. *J. Mol. Biol.* **2009**, *393* (2), 249−260.

(20) Baker, D.; Sali, A. Protein structure prediction and structural genomics. *Science* **2001**, *294* (5540), 93−96.

(21) Kaczanowski, S.; Zielenkiewicz, P. Why similar protein sequences encode similar three-dimensional structures? *Theor. Chem. Acc.* **2010**, *125* (3−6), 643−650.

(22) Saxena, A.; Sangwan, R. S.; Mishra, S. Fundamentals of homology modeling steps and comparison among important bioinformatics tools: An overview. *Sci. Int.* **2013**, *1* (7), 237−252.

(23) Venselaar, H.; Krieger, E.; Vriend, G. Homology Modeling. In *Structural Bioinformatics*; Gu, J., Bourne, P. E., Eds.; Wiley-Blackwell: Hoboken, NJ, 2009; pp 715−732.

(24) *The PyMOL Molecular Graphics System, Educational Version*; Shroedinger: New York, 2015. http://pymol.org/educational/ (accessed January 2015).

(25) The UniProt Consortium. Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Res.* **2014**, *42* (D1), D191−D198.

(26) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215* (3), 403−410. *Basic Local Alignment Search Tool*; NCBI: Bethesda, MD, 2015. http://blast.st-va.ncbi.nlm.nih.gov/ (accessed January 2015).

(27) Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T. J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Söding, J.; Thompson, J. D.; Higgins, D. G. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **2011**, *7* (1), 539. *Clustal Omega*; EMBL-EBI: Cambridgeshire, UK, 2015. http://www.ebi.ac.uk/Tools/msa/clustalo/ (accessed January 2015).

(28) Bujacz, A. Structures of bovine, equine, and leporine serum albumin. *Acta Crystallogr., D: Biol. Crystallogr.* **2012**, *68* (10), 1278−1289.

(29) Biasini, M.; Bienert, S.; Waterhouse, A.; Arnold, K.; Studer, G.; Schmidt, T.; Kiefer, F.; Cassarino, T. G.; Bertoni, M.; Bordoli, L.; Schwede, T. SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* **2014**, *42* (W2), W252−W258. *SWISS-MODEL*; Swiss Institute of Bioinformatics: Lausanne, Switzerland, 2015. http://swissmodel.expasy.org/ (accessed January 2015).

DOI: 10.1021/ed500358f
J. Chem. Educ. XXXX, XXX, XXX−XXX

(30) Lovell, S. C.; Davis, I. W.; Arendall, W. B.; de Bakker, P. I. W.; Word, J. M.; Prisant, M. G.; Richardson, J. S.; Richardson, D. C. Structure validation by Cα geometry: φ, ψ, and Cβ deviation. *Proteins* **2003**, *50* (3), 437−450. *RAMPAGE*. http://mordred.bioc.cam.ac.uk/~rapper/rampage.php (accessed January 2015).
(31) Willard, L. VADAR: A web server for quantitative evaluation of protein structure quality. *Nucleic Acids Res.* **2003**, *3113* (13), 3316−3319. Single (or Multiple) Model Protein Structure Analysis. *VADAR Version 1.8*. http://vadar.wishartlab.com/ (accessed January 2015).
(32) Chen, V. B.; Arendall, W. B.; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C. MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr., D: Biol. Crystallogr.* **2010**, *66* (1), 12−21. *MolProbity*. http://molprobity.biochem.duke.edu/ (accessed January 2015).
(33) Bhattacharya, A. A.; Curry, S.; Franks, N. P. Binding of the general anesthetics propofol and halothane to human serum albumin: High resolution crystal structures. *J. Biol. Chem.* **2000**, *275* (49), 38731−38738.
(34) Petitpas, I.; Grüne, T.; Bhattacharya, A. A.; Curry, S. Crystal structures of human serum albumin complexed with monounsaturated and polyunsaturated fatty acids. *J. Mol. Biol.* **2001**, *314* (5), 955−960.

DOI: 10.1021/ed500358f
J. Chem. Educ. XXXX, XXX, XXX−XXX