Extension of the Molecular Structure of the HIV-1 Protease via Quantum Mechanical Methods

Robert M Mullowney
Bioinformatics and Bioengineering Summer Institute,
Virginia Commonwealth University: VCU BBSI
Summer 2004

Introduction:

The Human Immunodeficiency Virus (HIV), the retrovirus that causes Acquired Immune Defiency Syndrome (AIDS), is currently being studied all over the world because of its wide-spread impact on the world's population. One scientific goal is to find a "cure" for AIDS. There are currently nineteen (19) drugs approved by the Federal Drug Administration (FDA) for the treatment of AIDS; however, HIV often quickly becomes resistant to these treatments. Accordingly, it is important to understand the process by which HIV becomes drug resistant.

In this study we will focus on the HIV-1 protease. The HIV protease is an enzymatic protein that cleaves large polyprotein (Gag and Gag-Pol) into usable proteins that are essential for HIV reproduction and viability. This homodimeric (comprised of two identical subunits) protein contains 99 amino acids in each subunit (for a total of 198) and is an aspartic protease. As such, the HIV-1 protease contains the conversed triad, Asp-Thr-Gly, at the active site (residues 25-27 with the same residues from the other chain). The HIV-1 protease is the target of seven (7) FDA approved dugs (protease inhibitors, PIs) for the treatment of HIV/AIDS. Resistance to protease inhibitors often arises from a single mutation or polymorphism. These mutations are a result of random mutations that arise from HIV's highly error prone replication. Due to HIV's rapid rate of replication, it is estimated that every possible mutation occurs over a hundred (100) times per day in an untreated HIV-infected individual. Therefore, in the presence of a PI, the strains containing resistance-conferring mutations will proliferate and be replicated in subsequent generations. Currently, the best known method to combat HIV is to give patients an assortment of drugs, often referred to as a cocktail, so that in the event that HIV becomes resistant to a specific drug, the others will still effectively inhibit protease activity. This strategy is complicated because some mutations offer resistance to several, or even all of the inhibitors while still allowing the natural substrate to enter the enzyme's active site. As drug resistance becomes an increasing problem, it is more important to discover new and better drugs. An ideal situation would be to find a drug that could always inhibit the protease, even after it undergoes mutation.

There are currently 179 crystal structures of the HIV-1 protease in the RCSB Protein Data Bank (http://www.rcsb.org/pdb/) (PDB). These structures represent numerous different mutations as well as different inhibitors, substrates, and inhibitor and substrate analogues bound to the protease. One major drawback to current crystallographic techniques is that at current resolutions, the location of protons (hydrogen atoms) is not very clear. The location of protons plays a key role in the shape and chemical nature of

the enzyme's active site, hydrogen bonds, and water molecules.

In this project we hope to use quantum mechanics to perform molecular orbital and molecular dynamic calculations to gain a better understanding of the true chemical nature of the HIV-1 protease mutants, especially the active site. Previously, these types of calculations have only been preformed on simple models of the active site and substrates. We, on the other hand, hope to examine the entire molecule. We hope to be able to determine bond orders, SCF energies, and the Universal Force field (UFF).

Methods:

We will obtain a crystal structure from the PDB and use ChemSite (a molecular viewing application for Windows based PCs) to add hydrogen atoms to the molecule and optimize the structure (ChemSite uses the AMBER parameters for optimizations). We will then uses this file as input into our calculation program. We will be using a Fortran program to do CNDO/2 or INDO (Complete/Incomplete Neglect of Deorthoginalized Orbitals) calculations. This code will be taken from ENZY4, CORN, CNDV, or NACH (these are all various versions of the same program, reference 1). The code to do the molecular dynamics calculations will be taken from another program. These programs were written and/or modified by Don Shillady and his students.

I must also write a script to convert the output of ChemSite into a format suitable for input into our program. The Cartesian format (*.cta) in ChemSite most closely resembles the format required by our program. Our input files must contain right justified fields as well as some simple changes to the data present in the file. This will be done using regular expressions and string/array manipulation in perl. The script will also have to count the total number of atoms and molecular orbitals present in the molecule. Another script may be required to obtain the required information to do the molecular dynamics calculations.

In addition to the advanced chemistry knowledge required for this project, I need to learn to be comfortable in a Unix computing environment (including the Vi text editor), improve my perl programming skills, and obtain a low level familiarity with Fortran77.

Possible results and implications:

The biggest problem that this project faces is that, in its current form, the family of CNDO programs cannot handle molecules as large as the HIV-1 protease. Supposedly, Charles Castevens, a recent Ph.D. student at VCU, enlarged ENZY to accommodate 4000 atoms (just barely larger than the HIV-1 protease); however, initial searches left us unable to find a final copy of his updated program. Size limitations are imposed by the need to address very large matrices in order to complete the molecular orbital calculations. INDO cannot be used without augmenting the program because it has insufficient parameters for dealing with sulfur atoms (present in nearly all proteins—in Met and Cys residues). Another strategy would be to add the required parameters to the INDO subroutines and then attempt to expand its capabilities. Another problem is that Fortran77 is not well supported on many of VCU's computing clusters. We also hope that we can complete enough of these calculations so that comparisons may be made between similar structures.

References:

(1) Pople, John and David Beveridge; *Approximate Molecular Orbital Theory*. McGraw-Hill, Inc, New York: **1970**.

(2) Shafer, Robert W. Genotypic Testing for HIV-1 Drug Resistance. HIV InSite Knowledge Base Chapter, April **2004**. http://hivinsite.ucsf.edu/InSite?page=kb-03-02-07

(3) Rappe, AK, CJ Casewit, KS Colwell, A Goddard III, and WM Skiff. *UFF, a Full Periodic Table Force Field for Molecular Mechanics and Molecular Dynamics Simulations*. Journal of the American Chemical Society; **1992**, 114, 10024-35.

(4) Lee, Hsing, Thomas A Darden, and Lee G Pedersen. *An ab Initio Quantum Mechanical Model for the Catalytic Mechanism of HIV-1 Protease*. Journal of the American Chemical Society; **1996**, 118, 3946-50.

(5) Brik, Ashraf and Chi-Huey Wong. *HIV-1 protease:mechanism and drug discovery*. Organic & Biomolecular Chemistry; **2003**, 1, 5-14

(6) Liu, Haiyan, Florian Muller-Plathe, and Wilfred F van Gunsteren. *A Combined Quantum/Classical Molecular Dynamics Study of the Catalytic Mechanism of HIV Protease*. Journal of Molecular Biology; **1996**, 261, 454-69.

(7) Piana, Stefano and Paolo Carloni. *Conformational Flexibility of the Catalytic Asp Dyad in HIV-1 Protease: an Ab Initio Study on the Free Enzyme*. Proteins; **2000**, 39, 26-36.

(8) Mahalingam, Bhuvaneshwari, John M Louis, Jason Hung, Robert W Harrison, and Irene T Weber. *Structural Implications of Drug-Resistant Mutants of HIV-1 Protease: High Resolution Crystal Structures of the Mutant Protease/Substrate Analogue Complexes*. Proteins; **2001**, 43, 455-64.