

# Teaching Botany Using Bioinformatics Tools

Xiaorong Zhang

Department of Biology, Savannah State University, Savannah, Georgia, USA

Email: [zhangx@savannahstate.edu](mailto:zhangx@savannahstate.edu)

**How to cite this paper:** Zhang, X. R. (2019). Teaching Botany Using Bioinformatics Tools. *Creative Education*, 10, 2137-2146. <https://doi.org/10.4236/ce.2019.1010155>

**Received:** September 16, 2019

**Accepted:** October 11, 2019

**Published:** October 14, 2019

Copyright © 2019 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

Two laboratory activities are designed to reinforce several important concepts in General Botany course, which is a required course for biology majors at Savannah State University (SSU). The first activity requires students to study the relationship between protein structure and function through observing the 3D structure of Rubisco (ribulose-1,5-biphosphate carboxylase and oxygenase)—the enzyme that catalyzes the first step of the Calvin cycle for photosynthesis. This activity also helps students understand the mechanism of enzymatic action through examining the interaction of Rubisco with its co-factor, substrate, competitive inhibitor, and product. The second activity is designed to help students grasp the concept of plant evolution and phylogeny through analyzing the genetic sequences of Rubisco collected from representative species and determining the evolutionary relationships of these species using bioinformatics tools. Through these two laboratory activities, several important topics are linked together, with Rubisco as a common theme, so that students would develop a holistic and coherent view of plant sciences. Furthermore, students would also gain several important bioinformatics skills that they could use and apply in their future studies and careers.

## Keywords

Rubisco, Protein Structure and Function, Cn3D, Enzymatic Action, Molecular Phylogenetics, Plant Evolution

## 1. Introduction

Botany course is commonly required as part of the curriculum for B.S. in biology at four-year colleges or universities. Through the course, biology students are expected to gain a basic understanding of plants—the photosynthetic organisms that sustain other forms of life on the earth—and relate plants to different aspects of their everyday life, such as food, health, and environment.

General Botany is offered to junior and senior biology students as a required

course at SSU. The course covers a broad range of topics of plant sciences at molecular, cellular, and organismic levels; it is accompanied by a laboratory section, which is linked to the lecture topics. In the past, a standard lab manual has been used to conduct the labs by the instructor; they were mostly stand-alone labs and each lab was designed to be completed in 1-hour and 50-minute. These labs tended to be isolated from each other and there was no apparent connection between them.

To foster the inter-connectedness among the topics in biology education and develop students' bioinformatics skills, we designed two lab activities using Rubisco as a common theme. Rubisco is an essential enzyme for the Calvin cycle of photosynthesis, catalyzing the first step, the rate-limiting step, of the cycle. It converts atmospheric carbon dioxide (CO<sub>2</sub>) into organic compounds. As one of the most important and abundant enzymes on the earth, Rubisco has been thoroughly studied and characterized. The enzyme is composed of 16 polypeptide subunits (or chains), including eight copies of a large chain and eight copies of a small chain. The large chain is coded by the gene (*rbcL*) in the chloroplast DNA, and the small chain is coded by the gene (*rbcS*) found in the nuclear DNA (Portis & Parry, 2007).

Rubisco needs two substrates: CO<sub>2</sub> and ribulose 1,5-biphospahte or RuBP (a five-carbon compound). Once the two substrates are gathered in the specific location of the Rubisco's active site, carbon dioxide is linked to RuBP to form two molecules of phosphoglycerate (3PG or PGA), converting inorganic carbon dioxide into organic compounds in the process (Andersson, 2008).

The structure of Rubisco has mostly been determined by X-ray crystallography technique; some of the 3D images are stored in the protein structure databases such as the PDB (Protein Databank) and MMDB (Molecular Modeling Database) and they can be visualized with Cn3D—the software that displays the structure of a biomolecule (Wang et al., 2000). To take advantage of these resources in the databases, we designed a laboratory exercise that would allow students to examine the structure of Rubisco, including its primary, secondary, tertiary, and quaternary structure, and observe the active site of Rubisco and the physical interactions between Rubisco with its cofactor, substrate, competitive inhibitor, and product.

Rubisco is an ancient enzyme that has evolved over two billion years and is found in most of the photosynthetic organisms, including cyanobacteria, algae, and plants (Chase et al., 1993). Therefore, the evolutionary history of the photosynthetic organisms has been recorded in the genetic sequences of Rubisco. As many plant genomes (including their nuclear and chloroplast genomes) have been sequenced, the genetic data of Rubisco are available for phylogenetic and evolutionary analysis of plants. Therefore, we designed another laboratory activity that would require students to retrieve the protein sequences of Rubisco in photosynthetic organisms from the databases and construct a phylogenetic tree of those species.

Through these two activities, several topics of botany course, such as photosynthesis, plant chemistry, plant genetics, plant phylogeny and evolution could be coherently and logically linked together. In addition, students would learn several basic bioinformatics skills, which are definitely important skills for the future biologists.

## 2. Design and Implementation of the Lab Activities

### 2.1. Activity 1—Visualization of the Structure of Rubisco with Cn3D

Students have been previously exposed to the concept of protein structure and function in their freshman biology course (Principles of Biology); they have also gained the basic understanding of enzymes and their catalytic roles in biochemical reactions in the same course. Those concepts are reinforced again in General Botany course. For example, Rubisco is thoroughly discussed in the context of photosynthesis in the lecture. To further enhance students' understanding of this enzyme, we designed a computer-based lab that requires students to visualize the structure of Rubisco and its interaction with its cofactor ( $Mg^{2+}$ ), substrate (RuBP), product (3PG), and competitive inhibitor (D-xylulose-2,2-diol-1,5-bisphosphat or XDP). Students are also encouraged to view the structures of several Rubisco mutants and investigate how a mutation renders the change of the 3D structure of Rubisco and affect its enzymatic function.

In this activity, students are asked to use Cn3D (“see in 3D”) at the NCBI site (<https://www.ncbi.nlm.nih.gov/>) to view the 3D structure of Rubisco at the MMDB or PDB. There are a number of structures of Rubisco stored in these databases, which have mostly been determined by X-ray diffraction technique. We choose several of them for this activity.

#### 2.1.1. The Overall Structure of Rubisco and Its Active Site

First, students are asked to view the overall 3D shape of the entire Rubisco enzyme. Spinach Rubisco in the database (PDB ID: 8RUC) is selected for this purpose (Andersson, 1996). Students are expected to see sixteen chains (or subunits) of Rubisco, including eight large chains and eight small chains, and how these chains are assembled in space. They are also required to examine the secondary structures in Rubisco, including 14  $\alpha$ -helices and 18  $\beta$ -sheets in the large subunit and 2  $\alpha$ -helices and 5  $\beta$ -sheets in the small subunit.

Students are then asked to observe the active site of Rubisco. Eight large subunits of Rubisco are organized into four groups, each of which contains two large chains that are assembled into an antiparallel dimer, with the N-terminal domain of one monomer adjacent to the C-terminal domain of the other. As a result, there are four such dimers in one molecule of Rubisco. The active site is at an interface between monomers within each dimer and arranged around a magnesium ion ( $Mg^{2+}$ ), which is the co-factor of the enzyme. Students are asked to zoom in to see the detail of the active site—the binding of  $Mg^{2+}$  and 2-carboxyar-

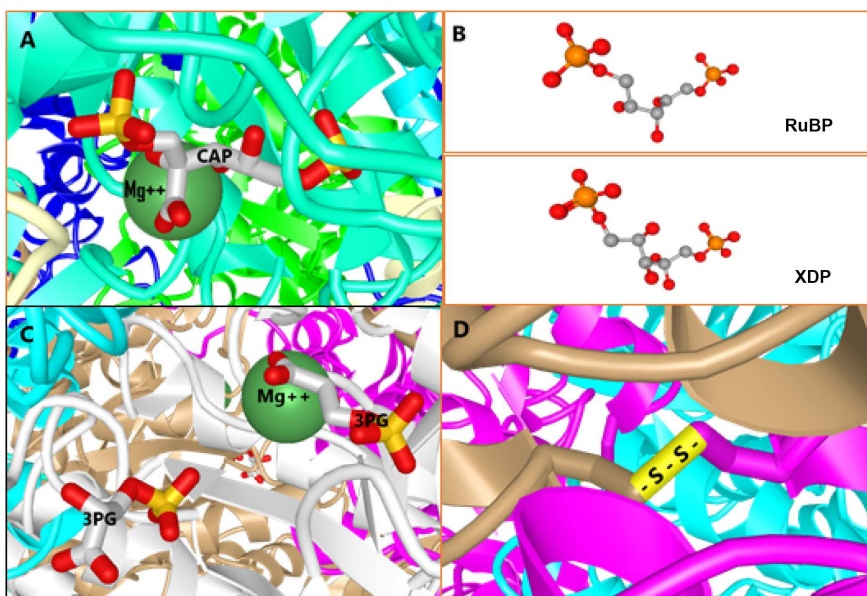
abinitol-1,5-diphosphate (or CAP, an analogue of RuBP) to the active site (**Figure 1(A)**). They are asked to identify amino acids that are part of the active site by selecting the amino acids that are 3 angstroms from the  $Mg^{2+}$ . They are expected to identify three amino acid residues: lysine (K) at 201<sup>th</sup> position, aspartic acid (D) at 203<sup>th</sup> position, and glutamic acid (E) at 204<sup>th</sup> position, all of which are charged amino acids (lysine as a positively charged and aspartic acid and glutamic acid as negatively charged), indicating they interact with other molecules with ionic bonds.

### 2.1.2. The Binding of Substrate, Inhibitor, and Product at the Active Site of Rubisco

Students are also asked to observe the binding of the substrate (RuBP) to the active site of Rubisco through an entry—PDB ID: 1RCX (Taylor & Andersson, 1997a) and the binding of its competitive inhibitor (XDP) to the active site through another entry—PDB ID: 1RCO (Taylor et al., 1996). By inspecting and comparing the inhibitor (XDP) and the natural substrate (RuBP) of the enzyme in their structure and shape (**Figure 1(B)**), students would gain an understanding of how a competitive inhibitor acts: it occupies the active site of Rubisco because of its similar shape to the substrate, preventing the binding of the natural substrate and inhibiting the chemical reaction. In addition, students are asked to observe a Rubisco complex with its product 3-Phosphoglycerate (3PG) through an entry—PDB ID: 1AA1 (Taylor & Andersson, 1997b). Two molecules of 3PG are bound per active site; both of them bind approximately at the same position as its substrate (RuBP) or competitive inhibitor does (**Figure 1(C)**). From these entries, students are also able to see the disulfide bridge that cross-links two large subunits in each dimer—Cys247 (cysteine at 247<sup>th</sup> position) residues of neighboring large chains are involved in the formation of this disulfide bridge (**Figure 1(D)**).

### 2.1.3. Rubisco Mutants

The 3D structures of several Rubisco mutants are also found in the databases. Through a direct observation and comparison of the structures of a wild-type with a mutated Rubisco, students are able to visualize how a single amino acid change at the level of primary structure of Rubisco results in a change of its 3D structure and affects its substrate binding and enzymatic activity. Those mutants are excellent examples of illustrating the close relationships of the amino acid sequence in a protein with its structure and function. For example, a mutant (PDB ID: 1 UWA) is caused by a substitution of leucine by phenylalanine at 290<sup>th</sup> position (L290F) in the large subunit of Rubisco from a green algae (*Chlamydomonas reinhardtii*), as both leucine and phenylalanine are nonpolar amino acids, they are similar in their chemical properties that the substitution does not lead to a global change in the shape of Rubisco; however, it does cause local structural changes and results in a 13% decrease in  $CO_2$  binding, and therefore leads to reduced catalytic activity of the enzyme (Karkehabadi et al., 2005).



**Figure 1.** (A) Binding of  $Mg^{2+}$  cofactor and CAP (an analog of RuBP) to the active site of Rubisco. (B) Molecular structures of RuBP (a substrate of Rubisco) and XDP (a competitive inhibitor of Rubisco). (C) Binding of two molecules of 3PG (the product of the chemical reaction catalyzed by Rubisco) to the active site of Rubisco. (D) Disulfide bridge (-S-S-) formed between two neighboring large subunits of Rubisco. The images are downloaded from NCBI (National Center for Biotechnology Information).

In addition, students are encouraged to explore some of the revertants in the databases, whose phenotypes have reverted to the normal phenotype by a second mutation. Those revertants help reinforce students' understanding of the intricate relationship of a protein's primary structure with its 3D structure and enzymatic function.

## 2.2. Activity II—Phylogenetic Analysis of Rubisco Proteins

Students have been introduced to the taxonomy and evolution of plants in the lecture; as a result, they have gained the basic understanding of the classification and phylogeny of plants. The new area of molecular phylogenetics is also explained to students and principles underlying the phylogenetic analysis of molecular data, such as DNA and protein sequences are also discussed in the lecture.

In this lab activity, students are required to collect protein sequences of the large subunit of Rubisco from at least twenty-five species representing different branches of photosynthetic organisms; those sequences are then compared and analyzed to reveal the evolutionary relationships among those species and generate a phylogenetic tree through the steps described below.

### 2.2.1. Collection of Protein Sequences of the Large Subunit of Rubisco

Students are asked to identify and retrieve the protein sequences of the large subunit of Rubisco by searching the protein databases on the NCBI site (<https://www.ncbi.nlm.nih.gov/>) using the keywords, such as “Rubisco large sub-

unit” and the name of species. Each student needs to retrieve 25 or more protein sequences from different types of photosynthetic organisms, including cyanobacteria, green algae, bryophytes, vascular seedless plants, gymnosperms, and angiosperms (monocots, eudicots, and basal angiosperms). **Table 1** shows a partial list of the organisms from which the protein sequences of the large subunit of Rubisco are derived.

### 2.2.2. Generation of Multiple Sequence Alignment

Once the protein sequences of the Rubisco large subunit are retrieved and formatted, students are asked to generate the multiple sequence alignment of these sequences using Clustal Omega (Thompson et al., 1994), which is available on the European Bioinformatics Institute (EBI) site (<https://www.ebi.ac.uk/Tools/msa/clustalo/>). Students need to examine the alignment visually and understand the mutation events that have led to the mismatches and gaps in the alignment. Mismatches are generally caused by amino acid substitutions, and gaps are usually generated by indels (that is, insertion or deletion mutations). By inspecting the alignment of the protein sequences, students would gain the basic understanding of how the sequences of Rubisco have been diverged by the specific molecular mutations over the course of evolution.

### 2.2.3. Construction of the Phylogenetic Tree

The *Phylogeny Interference Package* (PHYLIP) is downloaded from the website (<http://evolution.genetics.washington.edu/phylip.html>) and used to construct the phylogenetic tree. PHYLIP contains a number of software tools needed for the generation of the tree (Felsenstein, 1989). Specifically, PROTDIST in PHYLIP is used to compute a distance matrix from the alignment of Rubisco protein sequences (obtained from the previous step), which is a table showing the evolutionary distances between all pairs of protein sequences in the dataset; the evolutionary distance is calculated from the number of amino acid differences between a pair of sequences. NEIGHBOR in the same package is then used to generate a neighbor-joining tree (Saito & Nei, 1987) using the distance matrix data generated from PROTDIST. The graphic tree is displayed with Tree View (Page, 1996). Rubisco protein sequence from cyanophyta (or cyanobacteria) is used as an out group to root the tree.

### 2.2.4. Interpretation of the Phylogenetic Tree

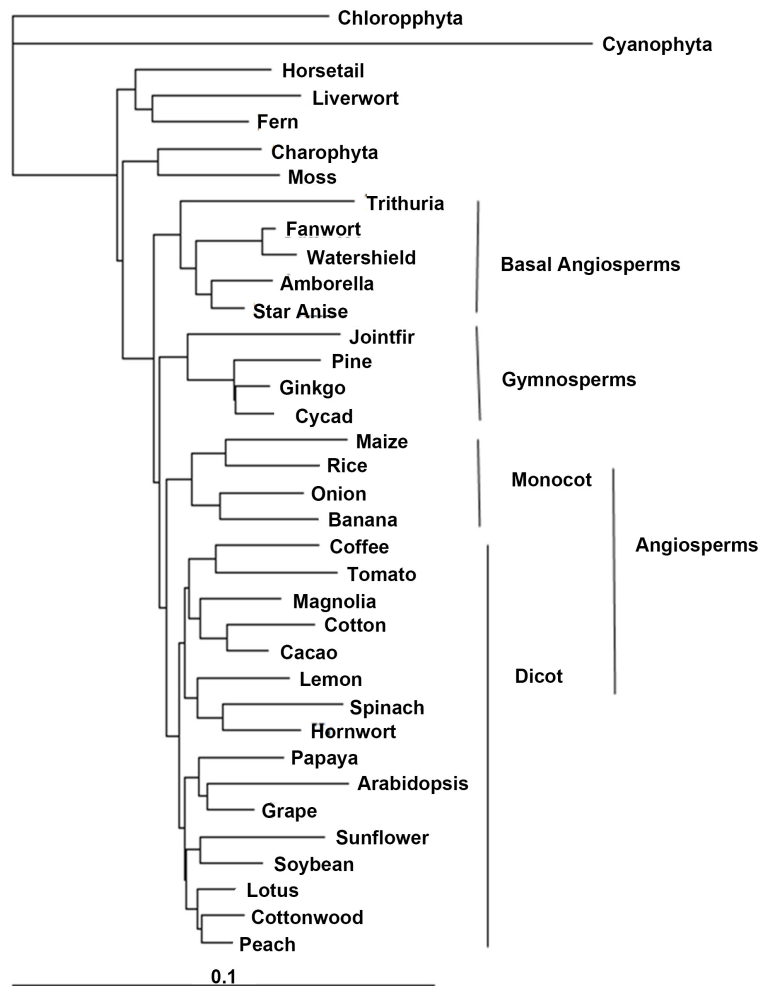
Once a tree is generated, students need to interpret their trees and compare them with the existing trees reconstructed from other data (morphological, anatomical, or molecular) in the textbook or journal articles.

A sample of the phylogenetic trees generated by our students is shown in **Figure 2**. The topology of the tree is in general agreement with the currently accepted view of the organismal phylogeny of photosynthetic organisms. Although the primitive plants—green algae, bryophytes, and ferns—are not clearly grouped, the tree generally displays the evolutionary trend of plants, from bryophytes and ferns to gymnosperms and angiosperms. Within the group of



**Table 1.** A collection of thirty-six photosynthetic organisms from NCBI.

Accession No.	Species	Classification
ABG53398	<i>Trichodesmium erythraeum</i> (cyanobacteria)	Bacteria
YP_764383	<i>Stigeoclonium helveticum</i> (green algae)	Chlorophyta
YP_635747	<i>Chara vulgaris</i> (green algae)	Charophyta
BAC85044	<i>Physcomitrella patens</i> (moss)	Bryophyte
AHN85622	<i>Marchantia polymorpha</i> (liverwort)	Bryophyte
YP_004021777	<i>Equisetum arvense</i> (horsetail)	Vascular Seedless
NP_848068	<i>Adiantum capillus-veneris</i> (fern)	Vascular Seedless
AJT70447	<i>Pinus armandii</i> (pine)	Gymnosperm
BAL72600	<i>Ginkgo biloba</i> (ginkgo)	Gymnosperm
A6BM42	<i>Gnetum parvifolium</i> (jointfir)	Gymnosperm
YP_009158611	<i>Encephalartos lehmannii</i> (cycad)	Gymnosperm
NP_904107	<i>Amborella trichopoda</i> (amborella)	Basal Angiosperm
ARJ63208	<i>Illicium anisatum</i> (Japanese star anise)	Basal Angiosperm
YP_009516884	<i>Cabomba aquatica</i> (fanwort)	Basal Angiosperm
AAF34884	<i>Magnolia grandiflora</i> (magnolia)	Basal Angiosperm
AUD56413	<i>Brasenia schreberi</i> (watershield)	Basal Angiosperm
ACN49416	<i>Nelumbo nucifera</i> (sacred lotus)	Basal Angiosperm
YP_007476359	<i>Trithuria inconspicua</i>	Basal Angiosperm
ABQ81458	<i>Ceratophyllum demersum</i> (hornwort)	Basal Angiosperm
BAA00147	<i>Oryza sativa</i> (rice)	Monocot
NP_043033	<i>Zea mays</i> (maize)	Monocot
CCW72383	<i>Musa acuminata</i> (banana)	Monocot
AEX93806	<i>Allium cepa</i> (onion)	Monocot
AAB68400	<i>Arabidopsis thaliana</i> (thale cress)	Eudicot
NP_054944	<i>Spinacia oleracea</i> (spinach)	Eudicot
AFS41717	<i>Gossypium hirsutum</i> (cotton)	Eudicot
AAA19771	<i>Solanum lycopersicum</i> (tomato)	Eudicot
ABJ89687	<i>Coffea arabica</i> (coffee)	Eudicot
YP_588125	<i>Helianthus annuus</i> (sunflower)	Eudicot
YP_009365832	<i>Citrus lemon</i> (lemon)	Eudicot
YP_538747	<i>Glycine max</i> (soybean)	Eudicot
YP_001109509	<i>Populus trichocarpa</i> (black cottonwood)	Eudicot
ABY86790	<i>Carica papaya</i> (papaya)	Eudicot
AEX58544	<i>Theobroma cacao</i> (cacao)	Eudicot
ADO64983	<i>Prunus persica</i> (peach)	Eudicot
YP_567084	<i>Vitis vinifera</i> (grape)	Eudicot



**Figure 2.** Phylogeny inferred from an analysis of the protein sequences of the Rubisco large subunit identified from thirty-six photosynthetic organisms. The names of these species and the accession numbers of these protein sequences from Genbank are shown in [Table 1](#). The scale bar represents 0.1 substitutions per amino acid site.

angiosperms, three subgroups are well-defined, representing monocot, eudicot, and basal angiosperms (flowering plants which have diverged from the lineage that evolved to monocots and eudicots).

### 3. Discussion

We designed two computer-based activities that were tied to several lecture topics in General Botany, including enzyme chemistry, photosynthesis, and evolution and phylogeny of plants.

We were under the impression that students, in general, showed great interest and enthusiasm for learning during these lab activities, which was demonstrated by their active participation and engagement in these activities. The evaluation of student performance indicated that the implementation of these exercises indeed helped students achieve their learning objectives. For example, most of the



students were able to complete these activities independently and submit a well-written report with valid results and correct interpretations. In addition, most of the students performed satisfactorily on the test, demonstrating their understanding of the concepts and mastering of the skills they learned in the lab. However, we were unable to make an accurate assessment of student learning outcomes this time due to the small size of the class (24 students) and the lack of a control group. We will definitely address this oversight in our future studies.

Although these activities were designed and implemented in General Botany course at SSU, they can be easily modified and adapted to teach the similar topics in other biology courses by selecting different groups of proteins or enzymes relevant to that course.

#### 4. Conclusion

There are several educational implications of this project. First, the traditional labs for General Botany were generally stand-alone labs that were disconnected with each other. These two exercises are developed to foster the interconnectedness of the topics and concepts through a common theme so that students would gain a holistic and coherent view of plant sciences. Second, unlike the traditional recipe-based lab exercises, these two exercises are designed to be inquiry-based and open-ended activities and to promote students' skills in critical thinking, problem-solving, and data analysis and interpretation. Third, students are introduced to several important bioinformatics skills, such as searching databases, retrieving data, and using software to analyze the data, which have become important skills in any areas of biological sciences. Students could apply these valuable skills in their future studies and careers.

#### Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

#### References

- Andersson, I. (1996). Large Structures at High Resolution: The 1.6 Å Crystal Structure of Spinach Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase Complexed with 2-Carboxyarabinitol Bisphosphate. *Journal of Molecular Biology*, 259, 160-174. <https://doi.org/10.1006/jmbi.1996.0310>
- Andersson, I. (2008). Catalysis and Regulation in Rubisco. *Journal of Experimental Botany*, 59, 1555-1568. <https://doi.org/10.1093/jxb/ern091>
- Chase, M. W., Soltis, D. E., Olmstead, R. G., Morgan, D., Les, D. H. et al. (1993). Phylogenetics of Seed Plants: an Analysis of Nucleotide Sequences from the Plastid Gene *rbcL*. *Annals of the Missouri Botanical Garden*, 80, 528-580. <https://doi.org/10.2307/2399846>
- Felsenstein, J. (1989). PHYLIP-Phylogeny Inference Package (Version 3.2). *Cladistics*, 5, 164-166.
- Karkehabadi, S., Taylor, T. C., Spreitzer, R. J., & Andersson, I. (2005). Altered Intersubunit Interactions in Crystal Structures of Catalytically Compromised Ribulose-1,5-Bisphosphate

- hosphate Carboxylase/Oxygenase. *Biochemistry*, *44*, 113-120.  
<https://doi.org/10.1021/bi047928e>
- Page, R. D. M. (1996). Tree View. An Application to Display Phylogenetic Trees on Personal Computer. *Bioinformatics*, *12*, 357-358.  
<https://doi.org/10.1093/bioinformatics/12.4.357>
- Portis, A. R., & Parry, M. A. (2007). Discoveries in Rubisco (Ribulose 1,5-Bisphosphate Carboxylase/Oxygenase): A Historical Perspective. *Photosynthesis Research*, *94*, 121-143.  
<https://doi.org/10.1007/s11120-007-9225-6>
- Saito, N., & Nei, M. (1987). The Neighbor-Joining Method: A New Method for Reconstructing Phylogenetic Trees. *Molecular Biology and Evolution*, *4*, 406-425.
- Taylor, T. C., & Andersson, I. (1997a). The Structure of the Complex between Rubisco and Its Natural Substrate Ribulose 1,5-Bisphosphate. *Journal of Molecular Biology*, *265*, 432-444. <https://doi.org/10.1006/jmbi.1996.0738>
- Taylor, T. C., & Andersson, I. (1997b). Structure of a Product Complex of Spinach Ribulose-1, 5-Bisphosphate Carboxylase/Oxygenase. *Biochemistry*, *36*, 4041-4046.  
<https://doi.org/10.1021/bi962818w>
- Taylor, T. C., Fothergill, M. D., & Andersson, I. (1996) A Common Structural Basis for the Inhibition of Ribulose 1,5-Bisphosphate Carboxylase by 4-Carboxyarabinitol 1,5-Bisphosphate and Xylulose 1,5-Bisphosphate. *Journal of Biological Chemistry*, *271*, 32894-32899. <https://doi.org/10.1074/jbc.271.51.32894>
- Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: Improving the Sensitivity of Progressive Multiple Sequence Alignment through Sequence Weighting, Position-Specific Gap Penalties and Weight Matrix Choice. *Nucleic Acids Research*, *22*, 4673-4680. <https://doi.org/10.1093/nar/22.22.4673>
- Wang, Y., Geer, L. Y., Chappay, C., Kans, J. A., & Bryant, S. H. (2000). Cn3D: Sequence and Structure Views for Entrez. *Trends in Biochemical Sciences*, *25*, 300-302.  
[https://doi.org/10.1016/S0968-0004\(00\)01561-9](https://doi.org/10.1016/S0968-0004(00)01561-9)