

# Structural Dynamics of Ribulose-1,5- Bisphosphate Carboxylase/Oxygenase

Michiel van Lun

*Faculty of Natural Resources and Agricultural Sciences  
Department of Molecular Biology  
Uppsala*

Doctoral Thesis  
Swedish University of Agricultural Sciences  
Uppsala 2013

Acta Universitatis agriculturae Sueciae

2013:40

Cover: CO<sub>2</sub>/O<sub>2</sub> density in Rubisco pop art  
(photo and design: Michiel van Lun)

ISBN (electronic) 978-91-576-7819-5

ISSN 1652-6880

ISBN (print) 978-91-576-7818-8

© 2013 Michiel van Lun, Uppsala

Print: SLU Service/Repro, Uppsala 2013

## Structural dynamics of Ribulose-1,5-bisphosphate Carboxylase/Oxygenase

### Abstract

Ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) assimilates carbon dioxide ( $\text{CO}_2$ ) from air into biomass. Due to its slow turnover, the reaction is a rate-limiting step in photosynthetic carbon fixation. The carboxylation reaction catalyzed by Rubisco is subject to inhibition by oxygen ( $\text{O}_2$ ) in a competing, non-productive reaction that reduces the efficiency of the enzyme by up to 50%. This makes Rubisco a target for engineering to increase crop yield.

The specificity of Rubisco for  $\text{CO}_2$  over  $\text{O}_2$  is a measure how well the enzyme is able to suppress the unwanted oxygenation reaction and varies between organisms. The specificity of Rubisco from several marine algae surpasses that of crop plants. Diatoms with high  $\text{CO}_2$  specificity from the arctic waters around Svalbard have been cultured, the Rubisco protein has been isolated and characterised, and the crystal structure has been determined. The holoenzyme structure is similar to the structure of Rubisco from plants, but the fold of the small subunits differs and has a shorter  $\beta\text{A}-\beta\text{B}$  loop and carboxy-terminal extension that extends into the solvent channel, that appears to provide extra stability to the holoenzyme.

The holoenzyme is a hexadecamer consisting of 8 large, catalytic, and 8 small subunits ( $\text{L}_8\text{S}_8$ ) with a mass of 500 kD. The dynamics of the interaction between the subunits in this large protein will likely influence catalysis and  $\text{CO}_2/\text{O}_2$  specificity. In order to examine the interface communication between subunits, molecular dynamics simulations have been performed on Rubisco enzymes from different organisms and with different holoenzyme structures, showing that the number of contacts and the size of the interaction area differ significantly in the different complexes examined. Single-residue mutations that affect specificity in Rubisco from the unicellular green alga *Chlamydomonas reinhardtii* also influence the protein dynamics and interactions across the subunit interfaces.

The migration of the gaseous substrates,  $\text{CO}_2$  and  $\text{O}_2$  in and around Rubisco, was investigated using molecular dynamics simulations. The results indicate that at equal concentrations of the gas, Rubisco has a preference for binding  $\text{CO}_2$  over  $\text{O}_2$ . Amino acids with small hydrophobic side chains are the most proficient in attracting  $\text{CO}_2$ , indicating a significant contribution of the hydrophobic effect in the interaction. On average, residues in the small subunit bind approximately twice as much  $\text{CO}_2$  as do residues in the large subunit, suggesting the small subunit may function as a reservoir for  $\text{CO}_2$  storage.

**Keywords:** Ribulose-1,5-bisphosphate carboxylase/oxygenase, Rubisco,  $\text{CO}_2/\text{O}_2$ , specificity, diatoms, protein interface, gas diffusion, X-ray crystallography, molecular dynamics simulations.

*Author's address:* Michiel van Lun, SLU, Department of Molecular Biology,  
P.O. Box 590, 751 24 Uppsala, Sweden  
*E-mail:* michiel@xray.bmc.uu.se

## Dedication

For Yuri, who went through doing a PhD twice.

# Contents

<b>List of Publications</b>	<b>7</b>
<b>Abbreviations</b>	<b>11</b>
<b>Foreword. Poor old Noah (ii)</b>	<b>13</b>
<b>1 Introduction</b>	<b>17</b>
1.1 An important but very inefficient enzyme	17
1.2 Molecular forms of Rubisco	18
1.3 Carboxylation/oxygenation reactions	21
1.4 Variation in CO <sub>2</sub> concentration	23
1.5 Similarity of CO <sub>2</sub> and O <sub>2</sub>	24
1.6 Diversity in catalytic rate and CO <sub>2</sub> /O <sub>2</sub> specificity	26
1.7 Rubisco with high specificity	27
1.8 Mutations in large and small subunits	28
1.9 The unicellular green algae <i>Chlamydomonas reinhardtii</i> as a model organism	28
1.10 A dynamical structure	30
<b>2 Aim of this thesis</b>	<b>31</b>
<b>3 Methods</b>	<b>33</b>
<b>4 Results and Discussion</b>	<b>35</b>
4.1 Specificity and Rubisco from marine algae	35
4.2 Composition of interfaces, dynamic and implications for function	37
4.3 Atomic fluctuations on the interface	43
4.4 Interface composition	46
4.5 Studies on gas migration and affinity	47
4.6 CO <sub>2</sub> solubility	48
4.7 Residue affinity for CO <sub>2</sub>	50
4.8 Subunit affinity for CO <sub>2</sub> and O <sub>2</sub>	51
4.9 Route for CO <sub>2</sub> to the active site	52
4.10 Factors important for overall efficiency	53
<b>5 Conclusions</b>	<b>54</b>
<b>References</b>	<b>56</b>



## List of Publications

This thesis is based on the work contained in the following papers, referred to by Roman numerals in the text:

- I Vålegård, K., Pearce, F. G., Haslam, R. P., Madgwick P. J., Andralojc, P. J., Kristoffersen, A.K., van Lun, M., Jewess, P., Taylor, T. C., Klein, U., Eilertsen, H. C., Parry, M. A. J. and Andersson, I. Unusual posttranslational modifications revealed in crystal structures of diatom Rubisco. *Manuscript*.
- II van Lun, M., van der Spoel, D., Andersson, I. (2011). Subunit interface dynamics in hexadecameric Rubisco. *Journal of Molecular Biology* 411, 1083-1098.
- III van Lun, M., Hub, J., van der Spoel, D., Andersson, I. CO<sub>2</sub> and O<sub>2</sub> distribution in Rubisco suggests the small subunit functions as a CO<sub>2</sub> reservoir. *Manuscript*.

Paper II is reproduced with the permission of the publisher.

Additional Publications:

Öster, L. M., Lester, D. M., van Scheltinga, A. T., Svenda, M., van Lun, M., Genereux, C. and Andersson, I. (2006). Insights to cephamycin biosynthesis: The crystal structure of CMcI from *Streptomyces clavuligerus*. *Journal of Molecular Biology*, 358, 546-558

Van Lun, M., Van der Spoel, D. (2006). Computational studies of protein folding. *Handbook of Theoretical and Computational Nanotechnology. Bioinformatics, Nanomedicine and Drug Design*. (Reith, M., Schommers, W., Ed.), 6, American Scientific Publishers, 25650 North Lewis Way, Stevenson Ranch, 91381-1439 California, USA

The contribution of Michiel van Lun to the papers included in this thesis was as follows:

- I Performed sample collection, evaluated some experiments and performed structure refinement.
- II Planned, performed and evaluated all experiments. Responsible for writing the manuscript, including the preparation of all figures.
- III Planned, performed and evaluated majority of experiments. Responsible for majority of writing and preparation of all figures.



# Abbreviations

2PG	2-phosphoglycolate
3PGA	3-phosphoglycerate
ATP	Adenosine triphosphate
CABP	2-carboxyarabinitol-1,5-bisphosphate
CC	Calvin cycle
CPU	Central processing unit
GB	Gigabyte, $10^9$ byte
GROMACS	GRoningen MACHine for Chemical Simulations
$K_{cat}$	Catalytic constant incorporating the rate constants for all the reactions between Enzyme-Substrate complex and Enzyme+Product; a measure of optimal catalytic production
$K_M$	Michaelis constant; 'substrate affinity'
L	Rubisco large subunit
MD	Molecular dynamics
NADPH	Nicotinamide adenine dinucleotide phosphate
OPLS	Optimized potentials for liquid simulations (force field)
pcr	Polymerase chain reaction
PDB	Protein Data Bank
RMSF	Root mean square fluctuation
Rubisco	Ribulose-1,5-bisphosphate carboxylase/oxygenase
RuBP	Ribulose-1,5-bisphosphate
S	Rubisco small subunit
TIPnP	Transferable intermolecular potential $n$ point (water)
V	Reaction Velocity
$\Omega$	Specificity factor



## Foreword. Poor old Noah (ii)

According to one of the oldest historical accounts, grasses were created on the third day of earth's existence. This is acknowledged by at least two world religions (see e.g. Genesis 1:9-12) and with such backup it is not strange that even today, people claim there is an intelligent design to all we see around us. We can derive from this account that grasses were the first known species of plant. On that same day seed plants and fruit bearing trees were created. The day after, daylight was created and this - though not mentioned literally - can be marked as the day photosynthesis started.

As a workhorse for using the energy given by the light to grow, an enzyme was given to plants that we now name Rubisco. This enzyme has the remarkable capability of being able to capture gas from the air and 'eat' it. It is what makes a grass (or indeed any plant) grow. Without Rubisco, no plant can live. The plants were given an enzyme that does this job really well when there is a lot of edible gas around. Plants like a concentration of about 1-5% edible carbon dioxide in the air with no oxygen. On the conditions that the first plants encountered in the aforementioned account, information is sparse, but since humans appeared within a week after this we can only assume that carbon dioxide levels were brought down to the human safe level of less than 1% and that there was enough oxygen to breathe from that day on.

Initially this seemed a solution for the waste problem that plants had: they were breathing out toxic oxygen. But they were stuck with an enzyme that was made to work in much higher carbon dioxide levels. On top of this, oxygen proved to be equally edible, but it would not make the plant grow. From now on plants had to make vast numbers of enzyme to make up for this flaw. Making and recycling Rubisco takes up most of a plant's energy, whereas the solution seems much simpler: creating an efficient enzyme that can easily tell food from fake.

Knowing that the levels of oxygen would go up to such high level that it would impair enzyme efficiency indicated that planning was absent or that the design was unintelligent. Perhaps, instead of maintaining that there was a design, we should consider another possibility. For this we turn for a brief moment to bacteria.

Suppose we have a bacterium that is susceptible for antibiotics. We can let this bacterium grow on a dish if we provide it with food. Within a day there will be countless colonies thriving visible to the human eye. Now we take a few billion bacteria - they are so small that they will fit on the tip of a needle all together - and put them on another dish with more food. This dish contains also an antibiotic mixed in the food that is toxic for the bacterium. What you can see after a day is that there are far less colonies on this dish, because most of the bacteria died within an hour of being put on it. But a few managed to grow. These are bacteria that obtained antibiotic resistance. It can be shown by comparing their genetic makeup with their ancestors, that they have acquired different genes. These bacteria have not been created at the beginning of history nor have they been designed. They got their new skill through a random mutation in their DNA. Now mutations are a rare thing. DNA is very carefully copied with every cell division, but from time to time they happen. What makes this work is the sheer amount of individuals involved. It took billions of dying bacteria and only a few had the right mutation to adapt and survive.

Enzymes behave in a similar way, evolving from simple to complex systems by mutations. If we look at another historical account, the collective plant genes, we can see that Rubisco is reshaped once every few hundred million years, though the basic large subunit is formed only once in history. It got adapted once it was there. At first the enzyme appeared when the earth's carbon dioxide concentrations were high and oxygen was low. When that slowly changed, plants survived and adapted what they could to become efficient enough to live. Residues mutated at a natural rate, taking many generations. Slowly, more efficient mutations outgrew the originals, but the original build was too complex to be completely changed from one generation to the other so the basic layout stayed. Subunits were added that stabilize the molecule. In a further adaptation to local temperature differences, the basic substrate changed from a 3-carbon molecule to a 4-carbon molecule in regions that are constantly warm, allowing stable growth in those areas. In the arctic waters high specificity for carbon dioxide over oxygen became decisive, for in cold waters carbon dioxide availability is low and chemical reactions are slow. These algae cannot compensate Rubisco's inefficiency with high turnover. In a changing world, the most adapted, in other words: the fittest, survived.

Rubisco's inefficiency and survival is a case of evolution in action. It can be explained without the help of a theory of divine creation.



# 1 Introduction

## 1.1 An important but very inefficient enzyme

Ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) is an enzyme crucial to life. Without it no plant would grow, no animal or human would have food. Yet, for all its importance it is remarkably inefficient. It is so inefficient that most plants have to resort to producing it in huge amounts to be able to survive, sometimes up to 50% of plant leaf protein consists only of Rubisco (Ellis, 1979). To give an idea how inefficient it actually is, a comparison with other enzymes is enlightening.

The turnover number of an enzyme is defined as the number of substrate molecules converted to product per unit time by an enzyme that is fully saturated with its substrate. The turnover number of the enzyme catalase is one of the highest known. Catalase splits hydrogen peroxide in water and oxygen at up to 40,000,000 molecules per second. Most biological enzymes are not that fast, but still turn over up to several hundreds or thousands of molecules per second (Table 1). Rubisco enzymes on the other hand have a turnover number as low as 3-10 per second in case of some higher plants.

It is safe to say that Rubisco is an inefficient enzyme and that raises many questions about one of the most basic processes of life. Why is it so slow? Why has it not been replaced with a less costly mechanism? Explanations are hinted by a number of factors, though so far we do not fully understand all the details of it. By now, Rubisco has been the subject of thousands of scientific papers, many of which try to solve a piece of the mystery. The details provided in those papers are intriguing and sometimes raise new questions. Nevertheless, they give us an insight in the fine and complex workings of life and evolution,

to which, hopefully, the papers presented in this thesis add new ideas. In this summary, first the structure of the enzyme and enzymatic reaction is considered, to see the details of the process of carbon fixation. Then the molecular forms of Rubisco will be discussed, emphasizing the important role of Rubisco from marine organisms. Naturally occurring mutations are presented in the chapter thereafter and this is followed by a discussion of the dynamics of Rubisco. The gas substrate and other factors to consider are discussed last.

Table 1. *Enzyme turnover. Source: Biochemistry 5th edition. Berg JM, Tymoczko JL, Stryer L. New York: W.H.Freeman; 2002 and: Biochemistry, Vol 1, AvJ Stenesh, New York 1998*

Enzyme	Turnover (s <sup>-1</sup> )
Catalase	40,000,000
Carbonic anhydrase	600,000
3-Ketosteroid isomerase	280,000
Acetylcholinesterase	25,000
Urase	10,000
Penicillinase	2,000
Lactate dehydrogenase	1,000
Beta-galactosidase	208
Chymotrypsin	100
DNA polymerase I	15
Tyrosyl-tRNA, synthetase	8
Tryptophan synthetase	2
Pepsin	0.5
Lysozyme	0.5

## 1.2 Molecular forms of Rubisco

The most common form of Rubisco is a hexadecamer (L<sub>8</sub>S<sub>8</sub>) of eight large (L) subunits complemented by eight small (S), non-catalytic, subunits. The large subunit contains a carboxy-terminal  $\alpha/\beta$  barrel domain that interacts with the amino-terminal domain of an adjacent L subunit to form an active site. A large-large subunit dimer is therefore required to complete the active site (Andersson et al., 1989). In some prokaryotes, the catalytic subunit of Rubisco forms a homodimer (L<sub>2</sub>) with two catalytic sites (in e.g. *Rhodospirillum rubrum*).

Unlike the hexadecameric form of Rubisco, which is not soluble when stripped of its small subunits, the dimeric Rubisco forms a soluble complex. Dimeric Rubisco shows only about 30% sequence identity with hexadecameric Rubisco (Tabita et al., 2007) nevertheless three-dimensional structures of the catalytic L subunits of both forms are almost identical. Especially the active site is highly conserved (Tabita, 1999, Andersson & Taylor, 2003). There is a third form of Rubisco, so far found in archaea, that comprises 10 large subunits, arranged as five dimers (Kitano et al., 2001).

The hexadecameric form of Rubisco has a 422 symmetry. The large subunits are arranged in four dimers, grouped around the four-fold axis. The small subunits are arranged as two tetramers that cover the top and bottom of the large subunits. Along the four-fold axis is a solvent accessible channel that varies in size between species, depending on the size of loops in the small subunit. The small subunits are not directly connected to the active site and therefore do not take part directly in catalysis. However, the small subunits are important for holoenzyme stability (Esquivel et al., 2002) and thus, indirectly, influence catalysis. Each small subunit interacts with three large subunits and two small subunits. The large subunits also have more than one interaction with neighboring subunits and in total there are 48 subunit interfaces in hexadecameric Rubisco. They can be classified into seven different types of interfaces (Knight et al., 1990). This is illustrated in Figure 1 using the structure of wild-type Rubisco from *Chlamydomonas reinhardtii*. The seven interfaces include three L-L subunit interaction interfaces, three L-S subunit interfaces, and one S-S subunit interface. The LL1 interface is the largest interface and involves two L subunits dimerizing “head to tail” to form a functional dimer. The core of Rubisco is made up of four functional dimers. Within the core, each L subunit is part of two additional interfaces, LL2 and LL3, involving the L subunits of the neighboring dimer. In addition, each L subunit makes contact with three neighboring S subunits in the interfaces LS1, LS2, and LS3. There is only one type of S-S subunit interface, SS1.

An active site in  $L_8S_8$  Rubisco is located in each large subunit at the LL1 interface and has an elongated, or funnel-like shape (Andersson et al., 1989). The N-terminal domain of the neighbouring dimer on the LL1 interface provides residues that are needed for completion of the active site. In all, there are twenty residues involved in substrate binding, most are charged (a detailed description can be found in Andersson et al., 1989 and Knight et al., 1990). The active site is open to solvent when no substrate is present. The C-terminal loop of the large subunit (residues 331-338) is flexible and substrate binding induces this loop relocate, closing the active site from the solvent and lysine

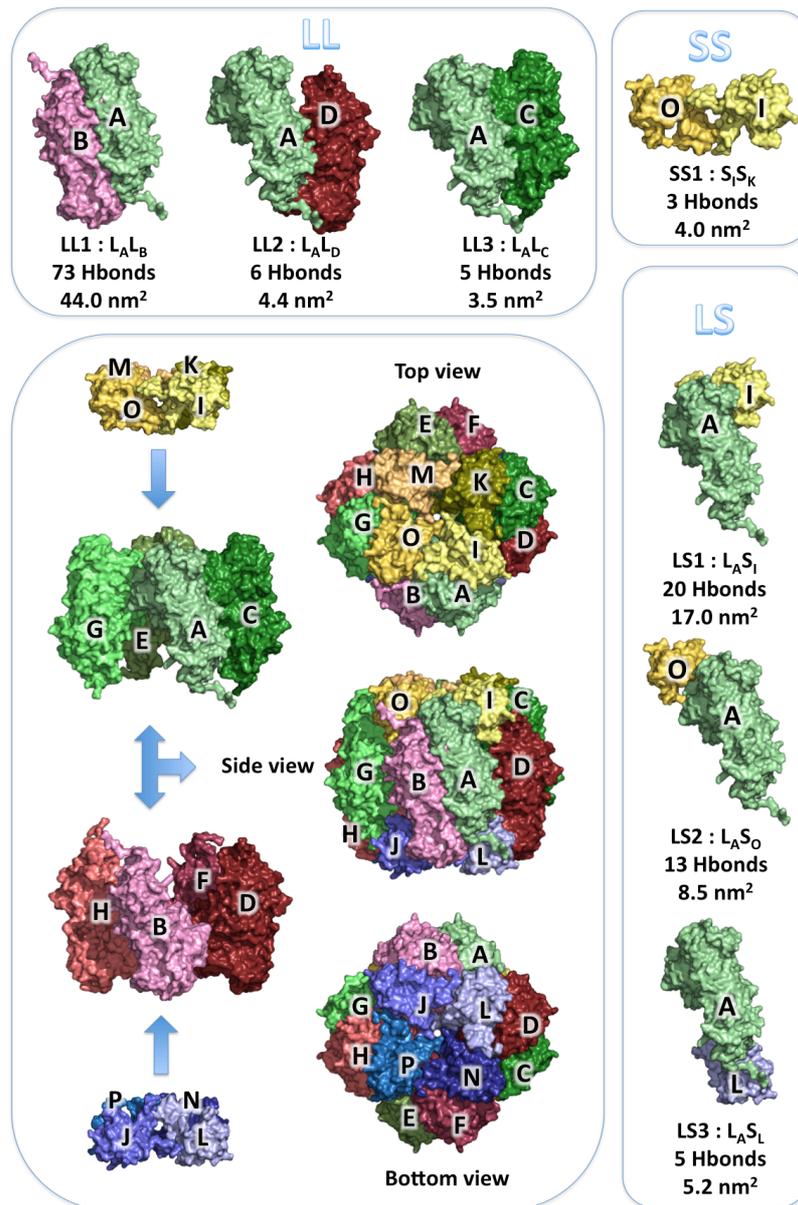


Figure 1. *L<sub>8</sub>S<sub>8</sub>* Rubisco architecture and interfaces

334, located at the tip of the loop, forms direct interactions with both the substrate and the other large subunit (Taylor and Andersson, 1996).

### 1.3 Carboxylation/oxygenation reactions

By incorporating carbon dioxide (CO<sub>2</sub>) into ribulose-1,5-bisphosphate, (RuBP), Rubisco ensures a net carbon uptake that is used for building biomass. Aside from the carboxylation of RuBP, Rubisco catalyzes a number of side reactions. The reaction with oxygen (O<sub>2</sub>) produces a toxic by-product of no real value for the organism.

The transition from dark to light in plant leaves results in a rise in pH. This facilitates deprotonation of the amino group of active site lysine 201. The carbon fixation reaction starts by this lysine binding one CO<sub>2</sub> (Lorimer 1976). The binding is reversible and depends on pH. After this, a magnesium ion (Mg<sup>2+</sup>) stabilises the carbamate group. The Mg<sup>2+</sup> is the center for catalysis. With this in place, a second CO<sub>2</sub> can be coordinated by the magnesium ion. Before or after this, the substrate RuBP can enter Rubisco's active site. Once Lys201 is carbamylated, and Mg<sup>2+</sup> and the substrate RuBP are in place, the reaction can begin (Figure 2). Assisted by groups on the protein, the C3 proton is abstracted, creating an enediol intermediate. The hydrogen electron pair of RuBP moves to C3-C2, creating a negative charge on the C2 oxygen. CO<sub>2</sub> then makes a nucleophilic attack on the 2-3-enediolate, forming the instable 3-keto-2-carboxyarabinitol-1,5-bisphosphate. With the help of a water molecule the C3 gets hydrated, splitting this molecule in the process into one 3-phosphoglycerate (3PGA) directly and a carboxylate anion of 3-phosphoglycerate, which is finally stereospecifically protonated to form the second product molecule as illustrated in Figure 2 (Taylor & Andersson, 1997; Cleland et al., 1998; Kanappan & Gready, 2008). Upon product formation the active site opens and reaction products are released. 3-Phosphoglycerate is then processed further in the cell's metabolism into substances that let its host organism grow, but for virtually all phototrophic organisms, this is the only point where inorganic carbon enters the metabolism.

Oxygenation is a competitive reaction that can happen when O<sub>2</sub> is present in the active site after activation. Whereas the active site is made for linking CO<sub>2</sub> to the enediol, it is possible for O<sub>2</sub> to be bound to the substrate instead. This

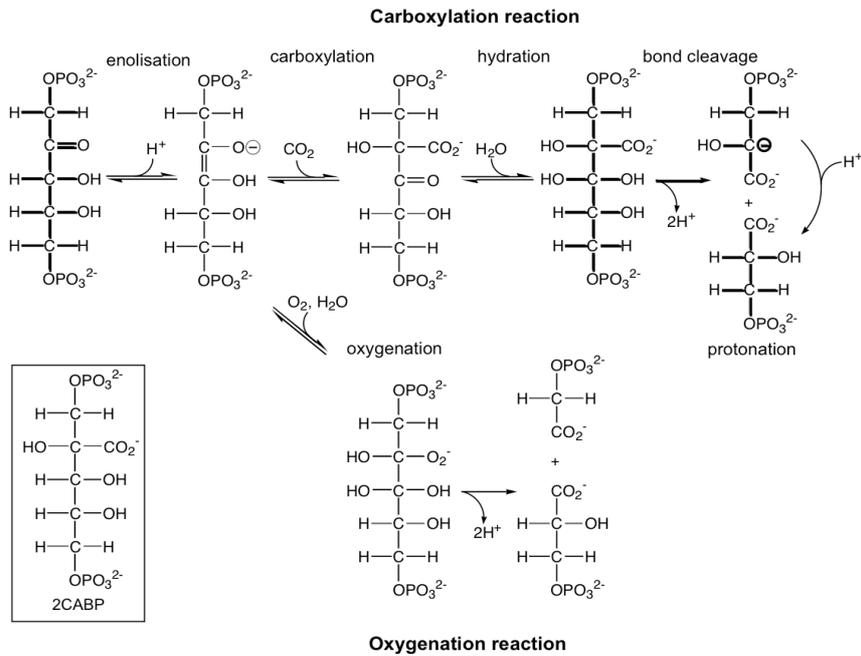


Figure 2. The carboxylation and oxygenation reactions catalyzed by Rubisco

does not lead to the usual end product of two 3-phosphoglycerate (6 carbon atoms), but produces one molecule of 3-phosphoglycerate plus one molecule of 2-phosphoglycolate (5 carbon atoms in total). This means there is no net carbon fixation. In addition, 2-phosphoglycolate is toxic in high concentrations and must ultimately be metabolised, making the process energetically unfavorable. The complex recovery route is illustrated in Figure 3 and will not be discussed in full detail here. Suffice to say that the recovery uses the cell's resources (ATP and NADPH) and in this respect the process is unwanted. Given the high  $O_2$  pressure in the air, this is a serious by-product route to count with. In extreme cases up to two thirds of the product may be phosphoglycolate with one third phosphoglycerate in tobacco leaf if  $CO_2$  availability is limited (to 0.01 mM  $HCO_3^-$ ) and  $O_2$  saturated (Pearce, 2006). Under regular circumstances the loss is around 30%, but may be as high as 50% (Watson et al., 1997). Apart from phosphoglycolate there are a number of other compounds that can be produced in side-reactions involving RuBP, such as xylulose-1,5-bisphosphate, pentodiulose-1,5-bisphosphate and carboxytetrial-1,4-bisphosphate, but they are around 1% or less of the product formation even in  $O_2$  saturation and as such do not play a major role (Pearce, 2006).

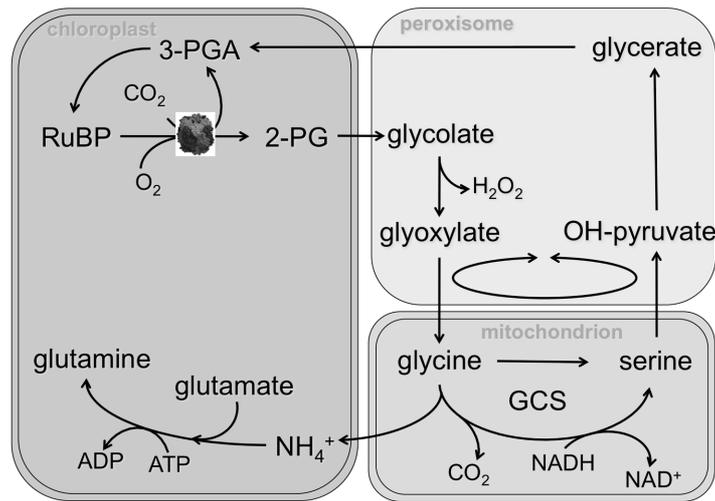


Figure 3. Photorespiration. The main product of oxygenation, 2-phosphoglycolate (2-PG) and its complex recovery route from chloroplast (left) to peroxisome (top right), mitochondria (bottom right) and back to return as phosphoglycerate input for the Calvin Cycle (CC).

#### 1.4 Variation in $\text{CO}_2$ concentration

Around 2.7 billions years ago the first Rubisco appeared that, connected to photoreception, converted  $\text{CO}_2$  into biomass and excreted toxic oxygen as a waste product. It is suggested that it evolved from a methanogen (Tabita et al. 2007), or a methionine salvage pathway (Ashida et al. 2005), though its true ancestor we will probably never know. The environment where it started to contribute to the carbon cycle was a different one than the current, likely with a 1000 times higher level of  $\text{CO}_2$  in the atmosphere and much lower concentrations of  $\text{O}_2$ . In the course of geological history the concentrations of  $\text{CO}_2$  and  $\text{O}_2$  have fluctuated, but the amount of carboxylation-competitive  $\text{O}_2$  has increased and been significantly higher in the atmosphere for most of the time at least since the Cambrian explosion. Interestingly, the current dimeric, decameric and hexadecameric complexes of Rubisco already existed before the great drop in  $\text{CO}_2$  concentration so they should not be taken as a response to

changes in the atmosphere, rather a form of local adaptation that benefitted its host organism (Whitney et al., 2011).

Currently the earth's atmosphere consists of 20.95% oxygen. CO<sub>2</sub> levels are rising steadily for over a century and the most recent CO<sub>2</sub> count was 396.8 ppm (0.04%, February 2013). This formidable difference in concentration does not reflect the local in vivo condition in which Rubisco operates. The concentration of either gas near Rubisco is not directly derived from the atmospheric concentration. CO<sub>2</sub> availability in aquatic environment is lower than in air because of unfavourable solvation. Most CO<sub>2</sub> in the sea will be converted to HCO<sub>3</sub><sup>-</sup> and enter algae in this form. Differences in pH and slow diffusion increase local variation in CO<sub>2</sub> availability. The level of cross membrane diffusion, availability of CO<sub>2</sub> transporters and temperature influence how much CO<sub>2</sub> reaches Rubisco in various species. The CO<sub>2</sub> concentration is further influenced by presence of carbonic anhydrases and carbon concentrating mechanisms. The availability of light alters these variables also. It is for these reasons, if one takes all these factors together, that the local concentration of CO<sub>2</sub> is hard to know exactly. Estimates have been made that put the available CO<sub>2</sub> at 2-10 μM (e.g. Pons et al., 2009). The local O<sub>2</sub> concentration is unknown. Caution has to be taken in extrapolating these numbers across species and habitats.

## 1.5 Similarity of CO<sub>2</sub> and O<sub>2</sub>

The reasons for Rubisco's O<sub>2</sub> susceptibility lie partly in the similarities of O<sub>2</sub> and CO<sub>2</sub>. From an electrostatic point of view they are both uncharged and with only a small dipole. The small dipole of both molecules has a similar pattern with a more positive center and light negative charges at either end of the molecule (for a visualisation see Kannappan & Gready, 2008). In terms of size O<sub>2</sub> is smaller (O-O bond length of 1.21 Å, versus 2.32 Å in O=C=O. The Van der Waals radius of O is 1.52 Å). Both molecules can easily pass a biological membrane through diffusion (Hub and De Groot 2008), nevertheless specific membrane transporters have also evolved.

Table 2. *Kinetic properties of various Rubiscos (under saturation of RuBP and CO<sub>2</sub>). Data from: 1) Morell et al., 1990; 2) Tcherkez and Farquhar, 2006; 3) Pearce, 2006; 4) Jordan, 1983; 5) Genkov and Spreitzer, 2009; A) Whitney et al., 2001. Table on the next page.*

	$\Omega$	$K_m^{CO_2}$ ( $\mu M$ )	$V_{max}$ ( $S^{-1}$ )	$K_{cat}$ ( $S^{-1}$ ) $V/K_m$	$K_m^{O_2}$ ( $\mu M$ )	$K_{cat}^C/K_m^C$ ( $M^{-1} \cdot S^{-1}$ )
<b>Bacteria</b>						
<i>Rhodospirillum rubrum</i> <sup>1)</sup>	12	89		7.3±0.3	406	1.1 x 10 <sup>5</sup>
<i>Riftia pachyptila symbiont</i> <sup>2)</sup>	6.2					
<i>Chromatium vinosum</i> <sup>2)</sup>	41	37		6.7		1.5 x 10 <sup>5</sup>
<b>Cyanobacteria</b>						
<i>Synechococcus PCC6301</i> <sup>3)</sup>	43	280		13.9		0.5 x 10 <sup>5</sup>
<i>Synechococcus PCC7002</i> <sup>2)</sup>	52	246		13.4		0.5 x 10 <sup>5</sup>
<i>Anabaena variabilis</i> <sup>2)</sup>	43					
<i>Aphonacapsa alpicola</i> <sup>4)</sup>	48±1	80±22				
<i>Plectonema boryanum</i> <sup>4)</sup>	54±2	100±38				
<b>Green algae</b>						
<i>Chlamydomonas reinhardtii</i> <sup>5)</sup>	60 ±1	31 ±2			527±20	
<i>Euglena gracilis</i> <sup>2)</sup>	54					
<b>C3 higher plants</b>						
<i>Nicotiana tabacum</i> <sup>1)</sup>	82±2	10.7±0.6	3.4±0.1	10.7 <sup>A)</sup>	295±71	10 x 10 <sup>5</sup>
<i>Triticum aestivum</i> <sup>2)</sup>	90	14		2.5		1.8 x 10 <sup>5</sup>
<i>Oryza sativa</i> <sup>2)</sup>	85					
<i>Spinacia oleracea</i> <sup>2)</sup>	82	14		3.7		2.6 x 10 <sup>5</sup>
<i>Atriplex glabriuscula</i> <sup>2)</sup>	87					
<i>Heliantus maximus</i> <sup>4)</sup>	77	10				
<i>Lyckopersicon escelentum</i> <sup>4)</sup>	82	8.2				
<i>Medicago sativa</i> <sup>4)</sup>	77	15.7				
<i>Petroselinum crispum</i> <sup>4)</sup>	77	11.6				
<b>C4 higher plants</b>						
<i>Amaranthus hybridus</i> <sup>2)</sup>	82	16		3.8		2.4 x 10 <sup>5</sup>
<i>Sorghum bicolor</i> <sup>2)</sup>	70	30		5.4		1.8 x 10 <sup>5</sup>
<i>Zea mays</i> <sup>2)</sup>	79	34		4.4		1.3 x 10 <sup>5</sup>
<i>Setaria italica</i> <sup>4)</sup>	58	31.2				
<i>Portulaca oleracea</i> <sup>4)</sup>	78	13.6				
<i>Echinochloa crus-galli</i> <sup>4)</sup>	83	18.4				
<b>Non-green algae</b>						
<i>Galdiera partita /sulfuraria</i> <sup>1)</sup>	166±6	3.3±0.4	1.2±0.1	3.3 <sup>A)</sup>	374±92	10 x 10 <sup>5</sup>
<i>Phaeodactylum</i> <sup>1)</sup>	113±1	27.9±0.4	3.4±0.1		467±22	
<i>Griffithsia monolis</i> <sup>1)</sup>	167±3	9.3±0.8	2.6±0.1	9.3 <sup>A)</sup>		10 x 10 <sup>5</sup>

## 1.6 Diversity in catalytic rate and CO<sub>2</sub>/O<sub>2</sub> specificity

Rubisco from individual species has unique catalytic properties, owing to variations in amino acids sequence. Along with this, the local conditions in which Rubisco has to operate can vary. Rubisco's preference for CO<sub>2</sub> to O<sub>2</sub> is referred to as the CO<sub>2</sub>/O<sub>2</sub> specificity. The specificity factor is defined as  $V_c K_m_o / V_o K_m_c$ , where  $V_c$  and  $V_o$  are the maximal velocities of carboxylation and oxygenation, respectively, and  $K_m_c$  and  $K_m_o$  are the Michaelis constants for CO<sub>2</sub> and O<sub>2</sub>, respectively (Laing et al., 1974). The relative rates for carboxylation and oxygenation are defined by the product of the specificity factor (often referred to as  $\Omega$  and sometimes as  $\tau$ ) and the ratio of CO<sub>2</sub> to O<sub>2</sub> concentrations at the active site.  $K_m$  is defined as the substrate concentration at which the reaction rate is half  $V_{max}$  and describes the substrate concentration needed for effective catalysis. A high  $K_m$  indicates a high concentration of substrate is needed. Often this number is interpreted as substrate affinity.  $k_{cat}$  represents the number of substrate molecules turned over per second, or, shortly, the turnover number. The ratio  $k_{cat}/K_m$  is an indication for overall enzyme efficiency.

The rates of carboxylation and the specificity are shown for a number of Rubisco from different species in Table 2. This table gives an idea of the differences that Rubisco from various species can exhibit. Bacterial Rubiscos are dimeric and generally have a low specificity, but in an oxygen-poor environment they will perform well. Cyanobacteria have higher substrate specificities (>40) and higher  $k_{cat}$  (>10 per second) than bacteria. They have a larger enzyme-substrate affinity however, so their effective growth (turnover) rate may not be as good as one would expect based on  $\Omega$  and  $k_{cat}$ .

Vascular plants have a higher specificity, but their catalytic rates are lower compared to cyanobacteria. Wheat (*Triticum aestivum*), rice (*Oryza sativa*), spinach (*Spinacia oleracea*) and maize (*Zea mays*) have specificity factors around 80. Their  $k_{cat}$  is less than 5 per second. Interestingly, non-green algae show a high CO<sub>2</sub> specificity, with values well over 100. They make a case for comparing specificity with crop Rubisco.

Substrate specificity is an important factor in effective turnover for Rubisco, together with catalytic rate. Taken together, there seems to be a balance between the two (Jordan et al., 1983, Zhu et al., 2004). It would in principle be possible to overcome Rubisco's inefficiency by either scaling up catalytic rate to make more end product or to make the reaction more specific so that no side reactions occur. Consider two organisms: one with a low specificity for the enzyme's substrate and a high catalytic rate, or another organism with a high specificity and a low catalytic rate. Hereby both

organisms may achieve a net equal growth. If we assume higher specificity as an aspect of evolution, it appears at least roughly to follow the division between higher plants and photosynthetic bacteria, but what good is higher specificity if a higher plant then just scales back its reaction rate? Would not the end result be the same? Yet there is a difference and the importance is in understanding that a higher specificity with lower reaction rate allows better competitive growth, especially in environments with high O<sub>2</sub> competition. If we count growth over millennia or even millions of years, a better growth of only a fraction of a percent is enough to outcompete other organisms. It has been argued that higher plants simply do not have to be more efficient than they actually are (Tcherkez and Farquhar, 2006). As inefficient as Rubisco may seem to us, it is enough to thrive.

Still, that does not satisfy a growing world population and we would be helped to great extent if we could find a way to increase Rubisco activity and thereby crop growth by only, for example, 3%. The reaction catalyzed by Rubisco is the rate limiting reaction in the Calvin cycle. That makes it the obvious target for molecular engineering, therefore understanding the function and efficiency of Rubisco has been a research goal for many decades.

## 1.7 Rubisco with high specificity

Each year, 123 Pg ( $120 \times 10^{15}$  g) of carbon is exchanged through terrestrial organisms, 90 Pg is exchanged between the atmosphere and oceans. Another 6-8 Pg is emitted through burning of fossil fuels. Every year about 10% of atmospheric CO<sub>2</sub> is recycled (0.1% of the total carbon amount in the biosphere) (Field et al., 1998; Beer et al., 2009; Scholes and Noble, 2001).

Seventy percent of the earth is covered by oceans. Marine photosynthesis contributes to about 50% of the global carbon sequestration, which amounts to some 2-4 Pg yr<sup>-1</sup> terrestrial (Le Quere et al., 2009) and 2.2 Pg yr<sup>-1</sup> marine sequestration (Takahashi et al. 2002). At a glance, it looks as if terrestrial Rubisco is more efficient by area. However, one has to take into account other factors for marine photosynthesis such as the difference in number of photosynthetic organisms and that a major part of marine photosynthesis takes place in colder waters. Perhaps a better way of realizing the potential of marine Rubisco is that marine organisms face limitations in light and CO<sub>2</sub> availability that are much more extreme than for terrestrial plants. Rubisco from non-green algae, diatoms and rhodophytes, have a higher CO<sub>2</sub> specificity than plant Rubisco. The importance for the ecosystem and the relatively high CO<sub>2</sub>

specificity of many arctic diatoms motivated us to collect and characterize them in a series of expeditions to the Barentsz sea in May of 2004, 2005 and 2006.

## 1.8 Mutations in large and small subunits

It appears that the active site of Rubisco has remained strongly conserved, and so has overall fold of the enzyme. Directed mutagenesis of large subunit active site residues invariably results in reaction impairment (reviewed in Hartman and Harpel, 1994; Spreitzer and Salvucci 2002). Natural variation in Rubisco sequence occurs outside these active site residues, indicating that the causes for variation in catalytic rate and specificity are to be found there. Mutations in the large subunit far from the active site have been shown to influence specificity (Spreitzer, 2001)

Mutations in the small subunit of Rubisco from prokaryotes and algae, influences  $\Omega$  (Kostov et al., 1997, Spreitzer et al., 2001, Genkov et al., 2006), despite the small subunit not taking part in catalysis. The ten most conserved small subunit residues have been investigated (Genkov and Spreitzer, 2009). Alanine substitution has no apparent effect in three cases (S16A, P19A, E92A) and results in holoenzyme instability (L18A), decreased specificity (Y32A, E43A) or affected catalysis (Y32A, E43A, W73A, L78A, P79A, F81A). Nine of the ten conserved small subunit residues are located at the interface of large and small subunit near large subunit  $\alpha$ -helix 8. Interface interaction apparently plays a role in communicating these structural differences to the active site, but the mechanism is poorly understood.

## 1.9 The unicellular green algae *Chlamydomonas reinhardtii* as a model organism

Studying mutations in plant Rubisco poses a problem since plants cannot thrive without a functional enzyme. There are additional complications if one would try site-directed mutagenesis in plants. The large subunit is encoded in the *rbcL* gene in the chloroplast, whereas the small subunit *rbcS* gene is located in the nucleus (Kawashima, 1972). Transport into the chloroplast and assembly of holoprotein requires signal sequences and species-specific chaperones. Mutagenesis of plant Rubisco is therefore only recently possible, but is time-

consuming and restricted to some favourable cases (Whitney et al., 2011). The unicellular green algae *C. reinhardtii* on the other hand, is an excellent organism to study photosynthesis related mutations in Rubisco, because photosynthesis deficient mutants of *C. reinhardtii* can survive on acetate as a carbon source (Spreitzer and Metz, 1981). Numerous photosynthesis-deficient *C. reinhardtii* mutants, with amino acid substitutions in both L and S subunits, have been identified, and recovery mutants have been selected and characterized (Spreitzer et al., 2005, Genkov et al., 2006, Karkehabadi et al., 2007) (reviewed in Spreitzer and Salvucci, 2002). Structural studies, together with biochemical characterizations, were carried out on several of these mutants (Karkehabadi et al. 2007, Karkehabadi, Peddy et al., 2005, Karkehabadi, Taylor et al., 2005, Garcia-Murria et al., 2008) and have provided detailed information on an atomic/molecular level. By screening of naturally occurring mutants in *C. reinhardtii*, a temperature-conditional photosynthesis-deficient mutant was discovered. This mutant, with a large subunit L290F substitution, has a 13% reduced  $\Omega$ . Further genetic selection for recovery of photosynthetic growth resulted in two large subunit mutations, A222T and V262L, that restored thermal stability and increased  $\Omega$  (Du et al., 2000, Hong and Spreitzer, 1997). Structural studies showed that the L290F substitution resulted in increased temperature factors at specific sites, compared to the wild-type protein (Karkehabadi, Taylor et al., 2005). This hinted to a dynamic explanation for the decreased specificity. Structural analysis further showed that destabilization by the L290F mutation appeared to be transmitted to the  $\beta$ A- $\beta$ B loop of the small subunit. This loop varies considerably in length and in sequence, in diverse species. Replacing the  $\beta$ A- $\beta$ B loop of *C. reinhardtii* Rubisco (28 residues) with the shorter corresponding loops from *Synechococcus* (10 residues) or spinach (22 residues) Rubiscos (Karkehabadi, Peddy et al., 2005) did not result in significant changes in the loop conformation, and the active sites appeared structurally intact. However, both chimeric enzymes had altered catalytic properties, and the introduction of the *Synechococcus* loop caused an 11% decrease in CO<sub>2</sub>/O<sub>2</sub> specificity (Table 3). The results of these structural studies suggested that there is an indirect effect of mutation on structure and function through the dynamics of the protein.

Table 3. Specificity and kinetic constants for Rubisco from *C. reinhardtii* and two artificial small subunit loop replacement mutations (data from Karkehabadi, Peddy et al., 2005)

	Wild type	T46-G64	H47-F53
	A47-R71	( <i>Spinach</i> )	( <i>Synechococcus</i> )
$\Omega$	63 $\pm$ 2	62 $\pm$ 3	56 $\pm$ 1
$K_{cat}$ (CO <sub>2</sub> )	35 $\pm$ 2	26 $\pm$ 3	36 $\pm$ 1
$V_c$ ( $\mu$ mol/h/mg)	111 $\pm$ 6	55 $\pm$ 4	60 $\pm$ 6

### 1.10 A dynamical structure

The crystal structure of a protein is an average of billions of protein copies. One important characteristic of proteins however is that they are dynamic in behaviour, but the information on the dynamics of proteins are missing in crystal structures of proteins. Smaller fluctuations are taken into account in structure refinement in a per-atom assigned B-factor (or temperature factor) that indicates the level of dynamic fluctuation. Using crystal structures it is possible to obtain dynamic information of events at the atomic level through computational analysis. Current computer systems are fast enough to do calculations on atomic level of large protein complexes solvated in water, opening a new level of protein structure investigation.

## 2 Aim of this thesis

The understanding of the relation between enzyme structure and function requires a comprehensive and detailed view of all its amino acids involved. Not only residues directly involved in catalysis, but also key structural amino acids are important. What influence structure, amino acid composition and atomic fluctuation have on the general affinity of Rubisco for its substrate CO<sub>2</sub> is not well understood. The exact location and dynamics of each atom may influence protein function to some degree. Earlier crystallographic work suggested the importance of protein dynamics in communicating the effect of mutation on structure and function (Karkehabadi, Peddy et al., 2005; Karkehabadi, Taylor et al, 2005). The dynamics of Rubisco from *C. reinhardtii* and several mutants with altered specificity and catalytic rate have been examined computationally in atomic detail. To investigate differences in dynamics caused by mutations, a detailed analysis of the interface communication in the protein complex was performed.

Rubisco from marine organisms are structurally homologous with Rubisco from crop plants and *C. reinhardtii*. Rubisco from arctic organisms have a high specificity for CO<sub>2</sub> within their habitat and this raises questions about the differences in their structure and enzymatic properties compared to Rubisco from crops. The structure and enzymatic properties of Rubisco from selected marine diatoms was studied as apart of this thesis.

CO<sub>2</sub> migration to the active site in Rubisco has never been investigated, yet such studies could provide useful knowledge for future work attempting to improve the enzyme. The CO<sub>2</sub> and O<sub>2</sub> affinity of each amino acid in Rubisco from spinach and *C. reinhardtii* was estimated, in order to understand the importance of specific amino acids. The approach here is again dynamical as the atomic fluctuation at ambient temperature influences Rubisco's CO<sub>2</sub>

affinity. The aim of this thesis is to understand Rubisco's structural dynamics and the influence on the performance of Rubisco as a catalyst.

### 3 Methods

The computational work in the papers presented in this thesis is largely performed with molecular dynamics. Calculations on the dynamics of proteins have restraints. Historically the size of a system and limitation in computational power would determine the detail of the method of computation. Proteins have been represented in the past with amino acids as points on a discrete lattice and even as just containing two residue types – H (hydrophobic) and P (polar) (reviewed in Dill 1993, Dill et al., 1995, Karplus and Sali, 1995). All-atomic simulations have gradually become standard, allowing for more accurate results. They do not take into account changes beyond atomic level, however, thereby posing a limitation when it comes to hydrogen bonding and polarization. Ideally, one would like a quantum mechanical representation of proteins, to allow for chemical reactions to be observed and electrostatic interactions to be modeled accurately. Unfortunately, quantum mechanical representation of complete protein systems is still out of reach, as that would require several orders of magnitude more calculation time than atom level calculations. The present method of choice for full protein simulation is all-atomic molecular dynamics.

A set of parameters describing atoms and their physical properties, has to be chosen such that the protein is modeled accurately. Jorgensen et al. developed the OPLS (optimized potentials for liquid simulations) force field during the last half of the 1980s as a means to perform simulations of structures and dynamics of biological systems (Jorgensen and Tirado-Rives, 1988). This provides a reasonable set of parameters for description of atomic interactions in solvated proteins. Water plays an important role in protein interactions and a good description of water is therefore indispensable. The special nature of a water molecule has led to explicit force-field descriptions of water as a three-atom molecule (e.g. TIP3P and SPC, SPC/E, Jorgensen et al., 1983, Berendsen

et al., 1981, Berendsen et al., 1987), a four-point molecule (TIP4P, Jorgensen et al., 1983), a five-point molecule, TIP5P (Mahoney and Jorgensen, 2000) and even a six-point molecule (TIP6P, Nada and Van der Eerden, 2003), where three points describe the atom locations and the other points are needed to set the electrical charges and masses (reviewed e.g in Vega et al., 2009, Cheng and Smith, 2007, Van der Spoel et al., 1998). These contribute to more accurate description (in theory at least), but also increase the computational load. A typical Rubisco system in this study used 55,000 water molecules, while Rubisco ‘only’ consists of ~20,000 atoms. Choosing a more complex water model for calculation inevitably means more computation time. A trade-off has to be found for every system that vindicates accuracy of the results. The water model used in the papers included in this thesis, TIP4P, proved to be the most useful in terms of computational cost and accuracy.

Calculations presented in this thesis have been carried out at super computer centers in Stockholm, Umeå and Linköping. Without them none of the calculations would have been possible. At the start of this thesis work, it took 3 months of cluster computation time to run one 20 ns Rubisco calculation comprising 300,000 particles. On a current desktop PC these calculations would take the best part of a decade each. Fortunately, the MD software used (GROMACS) has been parallelized tremendously in recent years, allowing the use of hundreds of CPUs simultaneously without too significant loss of calculation time through inter-node communication, shortening trajectory calculation times from months to days. One could ask why, with such increased speed, not extend simulation time so that hundreds of ns or even  $\mu$ s of trajectory is available for analysis. The size of the trajectory files is however also something to consider and it has to be determined what simulation length delivers statistically relevant data. A simulation time of Rubisco of 50 ns generates 70 GB of trajectory, which will be backed up, split for analysis, et cetera, and easily multiplies to a few hundred GB to work with. For one paper there were at least 40 of these calculations. Reading this amount of data from disk for analysis takes significant time as well and not every super computer center currently allows temporal storage of such an amount of data. Thus, the current length of computation and amount of data used is a necessary tradeoff.

## 4 Results and Discussion

### 4.1 Specificity and Rubisco from marine algae

For Paper I in this thesis, diatoms were sampled in open arctic waters as well as below the ice, in the coastal areas of northern Norway and in the Barents sea. From these samples, monocultures were grown. In addition, cultures were generated from collected diatom spores. The cultures were first tested on their ability to grow in normal conditions and under limitation of CO<sub>2</sub>, nitrogen, silicate (diatom outer cell walls are made of hydrated silicon dioxide) and phosphorus. This resulted in cultures of *Thalassiosira antarctica*, *Thalassiosira gravida*, *Thalassiosira hyalina*, *Thalassiosira nordenskiöldii*, *Skeletonema costatum*, *Bacteriosira bathyomphala*, and *Chaetoceros socialis*. The mean fastest grower of these is *T. antarctica* at 2-3°C and at 7°C (Table 3).

Table 4. Growth (in doubling per day) for seven arctic diatoms.

Species	Max. growth at 2-3°C	Max. growth at 7°C
<i>Chaetoceros socialis</i>	-0.19	0.09
<i>Thalassiosira nordenskiöldii</i>	0.02	0.11
<i>Thalassiosira hyalina</i>	-0.24	-0.32
<i>Thalassiosira antarctica</i>	0.54	0.12
<i>Thalassiosira gravida</i>	-0.25	-0.08
<i>Skeletonema costatum</i>	-0.22	-0.11
<i>Bacteriosira bathyomphala</i>	0.13	0.12

The seven fastest growing diatom species were further mass cultivated. For these cultures Rubisco specificity factors were finally determined based on  $^{14}\text{C}$  incorporation in 3PGA (carboxylation) and oxygen consumption measurement with an oxygen electrode (oxygenation). Together with the known concentrations of  $\text{CO}_2$  and  $\text{O}_2$  the substrate specificity ( $\Omega$ ) can be calculated (Table 4), This was done for three temperatures between  $15^\circ\text{C}$  and  $35^\circ\text{C}$ . Rubiscos from arctic diatoms have their highest specificity at lower temperatures (around 140 at  $10^\circ\text{C}$ , Haslam et al., 2005). Nevertheless the specificity at room temperature is as high as that of Rubisco from vascular plants. The lowest of these is *S. costatum* ( $\Omega=72$ ), a species that is commonly found further south (Degerlund and Eilertsen, 2009).

Table 5. Relative specificities and kinetic constants of Rubisco from marine diatoms.

Species	$\Omega$ $15^\circ\text{C}$	$\Omega$ $25^\circ\text{C}$	$\Omega$ $35^\circ\text{C}$	$V_c^{\max}$ $25^\circ\text{C}$ ( $\mu\text{mol min}^{-1} \text{mg}^{-1}$ )	$K_M^c$ $25^\circ\text{C}$ ( $\mu\text{M}$ )
<i>Rubisco</i>					
<i>Thalassiosira hyalina</i> <sup>A</sup>	105.5 $\pm$ 4.1	99.3 $\pm$ 3.2	86.7 $\pm$ 1.1	2.67 $\pm$ 0.07	49.8 $\pm$ 2.8
<i>Skeletonema marinoni</i>	96.2 $\pm$ 8.5	95.7 $\pm$ 6.6	84.1 $\pm$ 4.7	2.84 $\pm$ 0.05	47.6 $\pm$ 1.7
<i>Chaetoceros socialis</i>		92 $\pm$ 2.1			
<i>Thalassiosira antarctica</i> <sup>A</sup>		90 $\pm$ 3.2			
<i>Chaetoceros socialis</i>		89 $\pm$ 2.1			
<i>Bacteriosira bathyomphala</i>	93.5 $\pm$ 10.4	86.5 $\pm$ 4.2	76.4 $\pm$ 1.5	2.34 $\pm$ 0.06	47.6 $\pm$ 1.7
<i>Thalassiosira nordenskioldii</i> <sup>A</sup>		81.5 $\pm$ 2.1		2.57 $\pm$ 0.05	122.8 $\pm$ 3.9
<i>Skeletonema costatum</i>		72 $\pm$ 2.2			

A) From Haslam et al., 2005

The crystal structures of Rubisco from five species have been determined (*T. antarctica*, *T. hyalina*, *T. nordenskioldii*, *S. costatum* and *B. bathyomphala*). The structures are  $\text{L}_8\text{S}_8$  Rubisco, but are a form I variation that differs from plant Rubisco. This form I C/D is characterized by a short  $\beta\text{A}-\beta\text{B}$  loop and a carboxy-terminal extension in its small subunit that forms a  $\beta$  hairpin. In a hexadecamer four  $\beta$  hairpins form a  $\beta$  barrel, defining the central solvent channel opening (Figure 1 in Paper I). Form I A/B Rubisco in higher plants and cyanobacteria Rubisco has a less extensive small-small subunit interface, allowing for a more open solvent channel.

There are a number of post-translational modifications in the structures of diatom Rubisco: *T. antarctica*, *T. hyalina* and *B. bathyomphala* have a (most likely) methylated Cys109. Lys150 appears to have hydroxyl groups at  $\text{C}\gamma$  and  $\text{C}\delta$  that can form hydrogen bonds with neighbouring subunit residue Gly147. Lys346 is tri-methylated at  $\text{N}\epsilon$ . Two hydroxy prolines (Pro48, Pro155) have

also been found. Hydroxy proline and methylated cysteine have been detected earlier in Rubisco from *C. reinhardtii* (Taylor et al., 2001). Diatom Pro155 corresponds to hydroxylated Pro151 in *C. reinhardtii*. The reasons for these side chain modifications are unclear.

The amino acids sequences of diatom Rubisco show variation within and between species. The protein sequence of *T. nordenskiöldii* Rubisco deduced from our electron density maps differs in several positions from the published DNA sequences, but corresponds to the gene sequence obtained from our material (Paper I, supplementary Figure 1). In addition, the gene sequence of *T. hyalina* and *B. bathyomphala* differed in several positions from the sequences deduced from the electron density maps and from the previously published sequences. In all cases, the sequences were determined from both strands and were checked multiple times in the case of ambiguities. There are a number of possible explanations. Differences may be due to contaminations, per induced artefacts or to differences in the material used for DNA sequencing and structure determination. The latter may be due to the well-known problems associated with the identification of closely related diatom species (Syvertsen, 1977; Hasle & Syvertsen, 1997; Quillfeldt, 2001). There is also a possibility that the sequence indeed varies within species. Some diatoms may carry multiple genes for each subunit. Plants and green algae have more than one nuclear encoded gene for the small subunit (e.g. 2 in *C. reinhardtii* and 22 in wheat, Spreitzer, 2003). Multiple large and small subunit genes occur in some prokaryotes (Heinhorst et al., 2002; Spreitzer, 2003). It is not unthinkable that diatoms could carry more than one gene for the Rubisco large subunit.

#### 4.2 Composition of interfaces, dynamic and implications for function

There are four types of non-covalent protein interactions in proteins. First there are hydrogen bond interactions, involving hydrogen shared by nitrogen or oxygen. This happens either with an R-N-H group as donor on one side and a double bonded R-C=O acceptor on the other, or an R-O-H donor on one side and a R-C=O on the other, where R is an arbitrary group. Distances and energies for possible hydrogen bonds are listed in Table 6.

Second is the electrostatic interaction between two particles (1 and 2), determined by Coulombs law where the force is  $F=q_1q_2/4\pi\epsilon_0r^2$ , where  $q_1$  and  $q_2$  are charges or interacting particles,  $r$  is their distance and  $\epsilon_0$  the vacuum permittivity. If both particles (or chemical moieties) are charged, this type of interaction is also called ionic bond or salt bridge.

Thirdly there are Van der Waals interactions. Van der Waals interactions occur mainly within 2.5 to 4 Å and find their origin in the charge fluctuations around atoms, creating an attraction at a distance less than ~4 Å and a heavy repulsion if the distance is shorter than the Van der Waals atom radius (e.g. 2 Å for C, 1.4 Å for O, 1.2 Å for H). The Van der Waals attraction is a weaker force than electrostatic or hydrogen bond interactions, but plays an important role in protein interface interactions

Table 6. *Non-covalent bonds in proteins and their typical lengths (after: Biochemistry 5th edition. Berg JM, Tymoczko JL, Stryer L. New York: W.H.Freeman; 2002)*

Interaction	Length (Å)	Strength (kJ/mol)
Coulomb	3	~5.9-12
Van der Waals	2.5-4	~2-4
hydrogen bond		
O-H...O	2.70	21
O-H...N	2.88	29
N-H...O	3.04	8
N-H...N	3.10	13

Finally, the hydrophobic interaction is an effective force that arises from the electrostatic interaction of the surrounding water, which repels apolar atoms.

The large subunit homodimer interface from the first available crystal structure of plant Rubisco has been described in Knight et al., (1990). It was noted that the interface is formed by tight and extensive interactions that have a somewhat polar character. The location of Cys247 of two large subunits indicated the presence of a disulphide bond, that would strengthen the dimer interaction, and the bond formed despite the presence of a reducing agent (2-mercaptoethanol or dithiothreitol) in the buffers used for purification and crystallization. In vivo, under the influence of light, reducing conditions prevail and disulphide bonds may be broken. This may decrease stability of the dimer and could influence catalysis. The 247 cysteine pair is not completely conserved in all Rubisco forms though. The number of protein structures deposited in the Protein Data Bank has grown exponentially since the first Rubisco spinach structure was determined, but the homodimer interface of L<sub>8</sub>S<sub>8</sub> Rubisco's large subunit remains one of the largest protein interface structures determined (with solvent accessible surface >40 nm<sup>2</sup>, Figure 1). This interface is the largest in Rubisco and is present in structures from higher plants (tobacco, spinach and rice), *C. reinhardtii* and *Synechococcus*. Dimeric (L<sub>2</sub>) Rubisco has fewer residues included in its homodimer interface, but is still sizeable compared to a set of PDB representatives (Figure 4), being close to 30

nm<sup>2</sup>. Large-large subunit homodimers in hexadecameric (L<sub>8</sub>S<sub>8</sub>) Rubisco with a substrate analogue bound in the active site have a significantly larger interface area than the unliganded and unactivated enzyme. Upon substrate binding, the interface area increases 10 nm<sup>2</sup> in size. This effect can be seen in spinach Rubisco with and with a substrate analogue bound in the active site (8RUC and 1AUS) and also in the tobacco enzyme (4RUB and 3RUB).

The number of hydrogen bonds on the large homodimer interface of the different forms of Rubisco varies considerably as does the size of the interface. Note: the number of hydrogen bonds in Figure 4 includes salt bridges so it represents all polar interactions. The number of hydrogen bonds in L<sub>8</sub>S<sub>8</sub> Rubisco is more than the average number of hydrogen bonds calculated for a set of representative structures from the Protein Data Bank, illustrating the tight interaction provided by the Rubisco large subunit homodimer interface. For the L<sub>8</sub>S<sub>8</sub> structures with a substrate analogue in the active site, the number of hydrogen bonds in the homodimer interface increases even more, with almost 50%.

Table 7. Rubisco structures labeled in Figure 4. CABP is the substrate analogue 2-carboxyarabinitol-1,5-bisphosphate.

PDB code	Organism	Form	Complex	Resolution (Å)
1GK8	<i>C. reinhardtii</i> wild type	L <sub>8</sub> S <sub>8</sub>	Mg <sup>2+</sup> -CO <sub>2</sub> /CABP	1.40
1UWA	<i>C. reinhardtii</i> L290F mutant	L <sub>8</sub> S <sub>8</sub>	Mg <sup>2+</sup> -CO <sub>2</sub> /CABP	2.30
1UW9	<i>C. reinhardtii</i> L290F/A222T mutant	L <sub>8</sub> S <sub>8</sub>	Mg <sup>2+</sup> -CO <sub>2</sub> /CABP	2.05
1UZH	<i>C. reinhardtii</i> /Synechococcus loop mutant	L <sub>8</sub> S <sub>8</sub>	Mg <sup>2+</sup> -CO <sub>2</sub> /CABP	2.20
1UZD	<i>C. reinhardtii</i> /Spinach loop mutant	L <sub>8</sub> S <sub>8</sub>	Mg <sup>2+</sup> -CO <sub>2</sub> /CABP	2.40
1AUS	Spinach	L <sub>8</sub> S <sub>8</sub>	Mg <sup>2+</sup> -CO <sub>2</sub>	2.20
8RUC	Spinach	L <sub>8</sub> S <sub>8</sub>	Mg <sup>2+</sup> -CO <sub>2</sub> /CABP	1.60
3RUB	Tobacco	L <sub>8</sub> S <sub>8</sub>	SO <sub>4</sub> <sup>2-</sup> x 2	2.00
4RUB	Tobacco	L <sub>8</sub> S <sub>8</sub>	Mg <sup>2+</sup> -CO <sub>2</sub> /CABP	2.70
1EJ7	Tobacco	L <sub>8</sub> S <sub>8</sub>	PO <sub>4</sub> <sup>3-</sup> x 2	2.45
1RBL	<i>Synechococcus</i> PCC6301	L <sub>8</sub> S <sub>8</sub>	Mg <sup>2+</sup> -CO <sub>2</sub> /CABP	2.20
5RUB	<i>Rhodospirillum rubrum</i>	L <sub>2</sub>	-	1.70
9RUB	<i>Rhodospirillum rubrum</i>	L <sub>2</sub>	Mg <sup>2+</sup> -CO <sub>2</sub> /RuBP	2.60
1RUS	<i>Rhodospirillum rubrum</i>	L <sub>2</sub>	3PGA	2.90
2D69	<i>Pyrococcus horikoshii</i>	L <sub>8</sub>	SO <sub>4</sub> <sup>2-</sup>	1.90
1GEH	<i>Thermococcus kodakaraensis</i>	L <sub>10</sub>	SO <sub>4</sub> <sup>2-</sup> x 2	2.80
3A12	<i>Thermococcus kodakaraensis</i>	L <sub>10</sub>	Mg <sup>2+</sup> -CO <sub>2</sub> /CABP	2.30

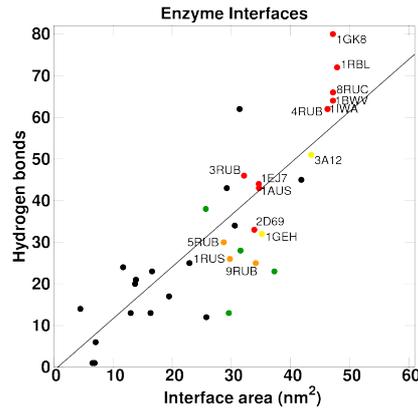


Figure 4. Homodimer interfaces of Rubisco (coloured dots) and a selection of representative structures from the Protein Databank (black dots).  $L_8S_8$  Rubisco in red,  $L_2$  Rubisco in orange,  $L_{10}$  Rubisco in yellow. See table 7 for PDB code abbreviations

As expected there is a correlation between the number of contacts and the interface surface area as shown in Figure 4. The distribution of polar interactions per interface area follows a linear relationship and may serve as a reference for the hydrophilicity/hydrophobicity of the interface. The area below the line is dominated by hydrophobic contacts, whereas the area above the line shows a dominance of hydrophilic interactions. The correlation curve shown was calculated based on 17 representative PDB entries selected according to defined criteria (see Paper II). Details of the selected structures are provided in the supplementary data of Paper II. The  $L_8S_8$  large subunit homodimers from structures with a substrate analogue bound in the active site have a much higher number of hydrogen bonds than expected based on the correlation of area and contacts.

$L_8S_8$  Rubisco contains 48 different subunit interfaces and although the LL1 interface is the largest and most important, the role of the other interfaces has not been investigated in depth previously. Large-large subunit contacts LL2 and LL3 are one tenth of the size of LL1, but still 4.4 and 3.5 nm<sup>2</sup>, corresponding to the size of interaction in a globular small protein complex. The small to large subunit interfaces, LS1, LS2 and LS3, range in size from 17 to 5.2 nm<sup>2</sup>.

The dynamics of the interfaces and nature of interface contacts is explored in more detail here. The size of the interfaces and the number of hydrogen bonds and polar contacts are subject to change at room temperature. MD simulations of Rubisco from *C. reinhardtii* show an interesting transition in the interface structure of the  $L_8S_8$  complex (Figure 5). The LL1 interface has a

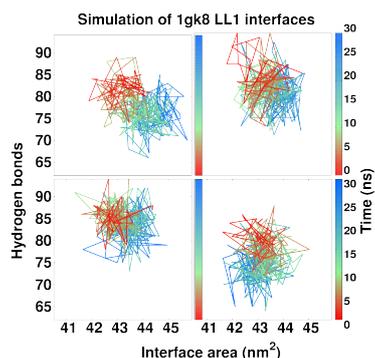


Figure 5. MD Simulation of 4 LL1 interfaces in Rubisco from *C. reinhardtii* (wild-type with substrate analogue CABP bound) using OPLS force field and TIP4P water. The number of hydrogen bonds (here including polar contacts) decreases in the first 5 ns and then stabilizes.

decrease in number of interface contacts in the first 5 ns of simulation. This is followed by stabilization in the remainder of simulation. The interface area also shows variation. The total simulation time was 30 ns and the structure contained the substrate analogue 2-carboxyarabitol-1,5-bisphosphate (CABP).

The number of hydrogen bonds in all interfaces after simulation is shown in Figure 6. Each interface type shows variation in size and number of contacts and deviates somewhat from the size and contacts listed in Figure 1, reflecting the transient nature of hydrogen bonds and the motion of residues at room temperature (summarized in Tables 8 and 9). Wild-type Rubisco (1GK8) and the single/double mutant structures (1UWA, 1UW9) have similar LL1 and LS1 interface sizes and number of hydrogen bonds/polar contacts. The variation in interface size and number of hydrogen bonds does not allow a discrimination of the mutated structures. The replacement of small subunit loop A47-R71 with T46-G64 from spinach does result in a tighter, but not larger LL1 and LS1 interface. Replacement of this loop with the corresponding shorter loop (residues H47-F53) from *Synechococcus* Rubisco has a detrimental effect on the sizes of the LL1 and LS1 interfaces, but not on the number of hydrogen bonds on the LS1 interface.

The ratio of hydrogen bonds per interface area is comparable for all subunits, but the LL1, LL2, LL3, and LS2 interfaces have a more hydrophilic character than the other interfaces. An average of 1.7 hydrogen bonds per square nanometer is found in the LL interfaces (average of LL1, LL2, and

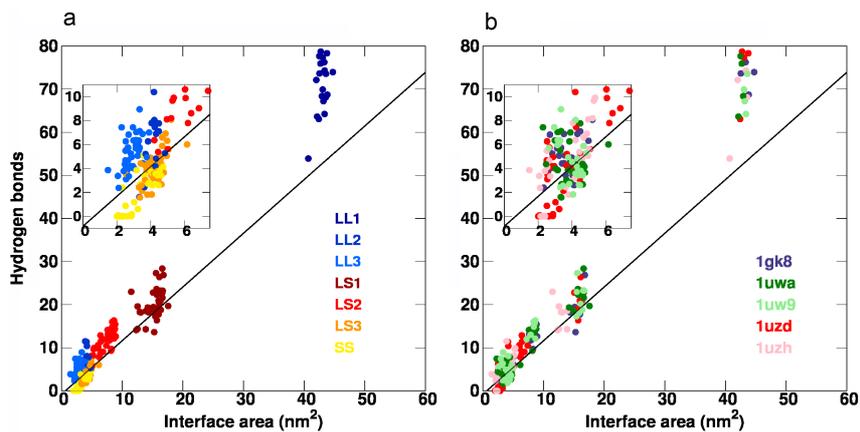


Figure 6. Hydrogen bonds (including polar contacts) and interface areas of 7 types of interfaces (a) in Rubisco from *C. reinhardtii* (1GK8) and four mutants (b, see table 5 for PDB code abbreviations). Data are averaged from the last 2 ns of a 30 ns simulation.

LL3), compared to 1.3 hydrogen bonds per square nanometer in LS interfaces (average of LS1, LS2, and LS3) or 0.9 hydrogen bonds per square nanometer in SS interfaces (average of SS1). The LS2 interface has 1.6 hydrogen bonds per square nanometer and is more hydrophilic than any other LS interface.

It is unusual for a protein to have high polarity on the inside. Protein folding studies show that the inside of proteins tends to be hydrophobic, serving as a core for protein folding. Water as solvent interacts best with outer residues that are more hydrophilic. The unusually high polarity of Rubisco's large subunit homodimer interface may help to bring the two subunits tighter together.

Table 8. Variation in subunit interface sizes and number of hydrogen bonds (including polar contacts) in Rubisco from *C. reinhardtii* and mutants, Data are averaged from the last 2 ns of a 30 ns simulation

Interface	Size (nm <sup>2</sup> )	Hydrogen bonds
LL1	41 – 45	53-79
LL2	3.2 - 4.8	2-11
LL3	1.5 - 3.9	3-9
LS1	12 - 17	14-28
LS2	4.2 - 9	5-16
LS3	2.5 - 6	2-8
SS	2 - 4.7	0-5

Table 9. Variation in LL1 and LS1 interface size and hydrogen bonds (hb) for *C. reinhardtii* Rubisco mutant structures. Data are averaged from the last 2 ns of a 30 ns simulation.

<i>C. reinhardtii</i>	LL1 nm <sup>2</sup>	LL1 hb	LS1 nm <sup>2</sup>	LS1 hb
wild-type (1GK8)	42.5 - 44.7	69 - 76	14-17	14-26
L290F (1UWA)	42.2 - 42.9	64 - 78	14-18	19-28
L290F/A222T (1UW9)	43.1 - 43.6	64 - 74	15-17	17-27
T46-G64 (Spinach) (1UZD)	42.4 - 42.8	63 - 79	15-17	19-26
H47-F53 ( <i>Synechococcus</i> ) (1UZH)	40.7 - 43.6	53 - 74	11-14	14-23

### 4.3 Atomic fluctuations on the interface

The non-covalent contacts on the interfaces mediate the communication between subunits. The communication is influenced by atomic fluctuations. In Figure 7, the averaged fluctuations of C $\alpha$  atoms in *C. reinhardtii* Rubisco during 30 ns of simulation are shown. There are large fluctuations on the N-terminal end of the small chain and the C-terminal end of the large chain, that is involved in active site closure (Taylor and Andersson, 1996). Overall the large subunit is stable with most residues displacing less than 1 Å during the simulation. There are a few residues that fluctuate more than 1Å. The single mutant L290F and double mutant L290F/A222T exhibit virtually the same average fluctuation, with differences limited to less than 0.5 Å movement. Around residue 290 no higher fluctuation is observed in contrast to the earlier observed higher B-factor (Karkehabadi, Taylor et al. 2005). From this, it is hard to conclude the reason for altered  $\Omega$  in the single mutant and restoration in the double mutant and it is likely that mutation A222T is compensating for the loss in specificity in an indirect way. Taking a closer look at fluctuations at the LL1 interface (Figure 8), minor differences in the mutant structures are detected that appear to be due to the presence or absence of CABP in the active site. In the presence of CABP there is a larger interaction area and less fluctuation in most residues in the wild type structure, but also an increased number of isolated high fluctuation points. The pattern of fluctuation in the mutant structures is different, this is also true for the double mutant, which does not reproduce the patterns of the wild type interface. Although double mutant L290F/A222T restores wild type specificity, it does not display

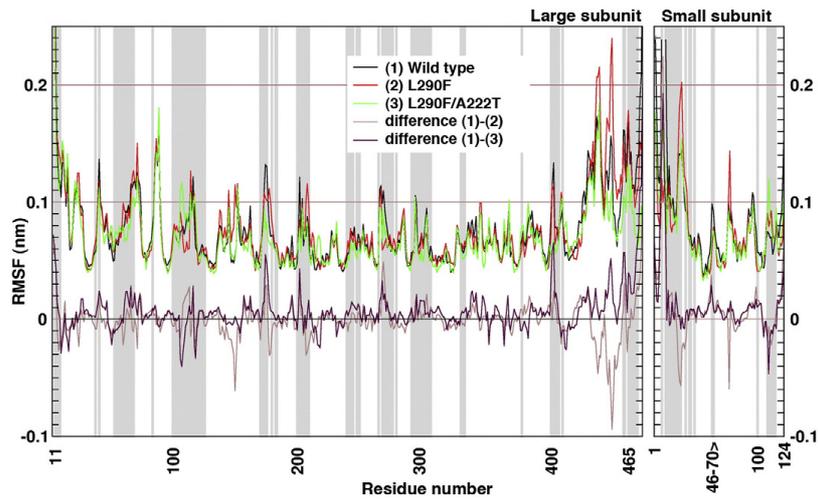


Figure 7. The per-residue dynamics in *C. reinhardtii* wild type Rubisco CABP complex (black), the single mutant L290F (red), and the double mutant L290F/T222A (green). The averaged root mean square fluctuation (rmsf) of C $\alpha$  atoms is displayed on the y-axis. The lines below show the difference in fluctuations between the wild type and L290F (maroon) and between the wild type and L290F/T222A (light brown). Interface residues are indicated by gray shading of the background.

fluctuations observed in the wild type interface. This suggests that one or more other processes may play a role in specificity restoration.

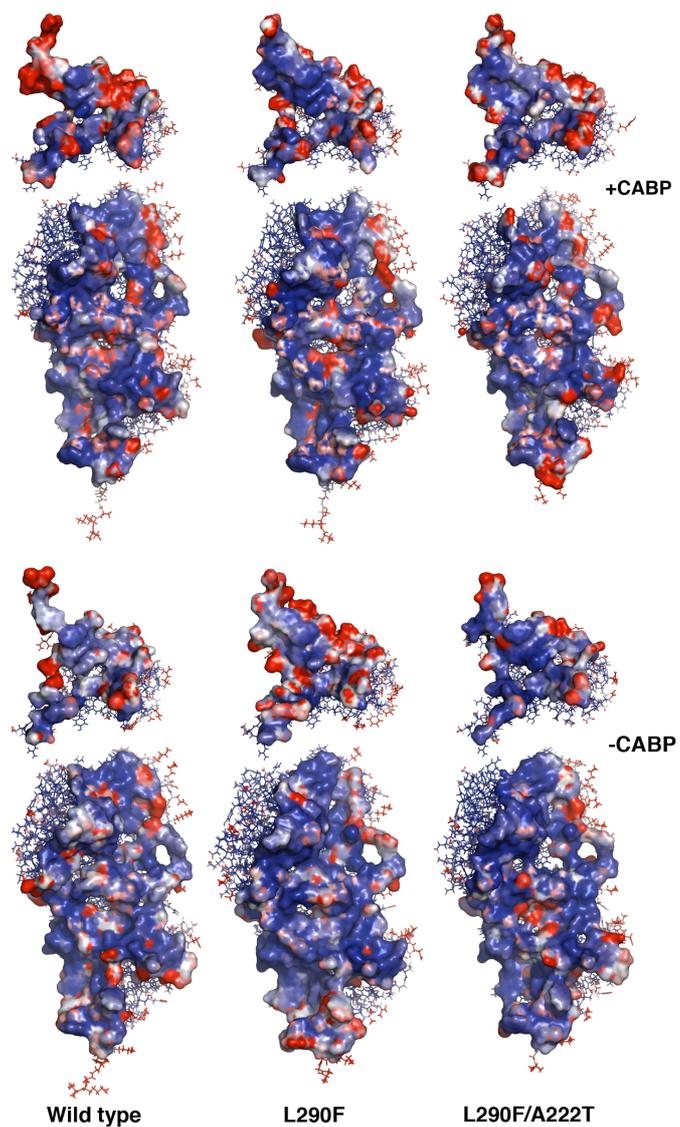


Figure 8. Fluctuations in LL1 and LS1 interfaces in *C. reinhardtii* Rubisco. From left to right is shown wild type, single mutant L290F and double mutant L290F/A222T. The upper half shows structures with substrate analogue CABP bound, the bottom half without CABP. Colouring is from blue (low fluctuation), through white to red (highest fluctuation). The interfaces are shown in surface representation, the rest of the subunit in wireframe.

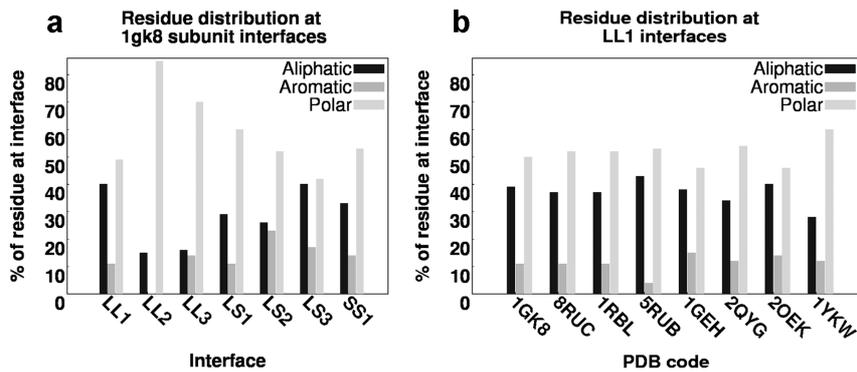


Figure 9. Amino acid residues at the interfaces of Rubisco. (a) Amino acid distribution at the seven different types of interfaces of *C. reinhardtii* wild-type Rubisco. The distribution is similar in the *C. reinhardtii* mutants. (b) Amino acid distribution at the LL1 interfaces in different forms of Rubisco: 1GK8, 8RUC, and 1RBL are  $L_8S_8$  enzymes, 5RUB is a  $L_2$  enzyme, 1GEH is a ( $L_2$ )<sub>3</sub> enzyme, and 2QYG, 2OEK, and 1YKW are Rubisco-like proteins. Amino acids are grouped according to properties: nonpolar (G, A, V, L, I, and M; black), aromatic (F, Y, and W; gray), and polar [S, T, C, P, N, Q, K, R, H, D, E, and other modified amino acids (X); light gray].

#### 4.4 Interface composition

Polar residues require the opposite (partial) charge in order to interact. Hydrophobic residues will prefer to interact with each other rather than with polar water molecules. In this sense polar interactions can be called specific interactions. The composition of Rubisco interfaces shows a large amount of polar residues (Figure 9).

Over 50% of the residues at the seven types of interfaces are polar and there are a number of salt bridges. On the LL1 interface are Glu-Arg salt bridges (E109-R253 and E110-R213), that are conserved in  $L_8S_8$  Rubisco. The LL2 interface has a salt bridge at K146-E110, connecting to the LL1 interface through common residue E110. The small LL3 interface includes two salt bridges (Y165-K183 and K161-D216). The LL2 interface has relatively most polar residues of all interfaces, but forms less hydrogen bonds (1.58 per  $\text{nm}^2$ ) than the LL1 interface.

The LS1 interface also forms less hydrogen bonds per  $\text{nm}^2$  than the LL1 interface, the LS2 interface however is more polar and has as much hydrogen bonds per  $\text{nm}^2$  as the LL1 interface (1.33, 1.58 and 1.65 respectively)

## 4.5 Studies on gas migration and affinity

The migration of gas into proteins has been investigated earlier. Lakowicz and Weber (1973) used the ability of O<sub>2</sub> to quench the fluorescence of tryptophan to show that O<sub>2</sub> diffuses into proteins at a diffusion rate 20-50% of that of O<sub>2</sub> in water. They also established that all buried interior tryptophan residues are reached by O<sub>2</sub> in a ns timescale, implying that proteins determined by X-ray crystallography are not rigid structures, but undergo rapid structural fluctuations. Later studies revealed specific gas channels inside some proteins, for example for H<sub>2</sub> and O<sub>2</sub> in hydrogenases (Montet et al., 1997, Cohen et al., 2006) and O<sub>2</sub> in myoglobin - already proposed in 1966, while the first low resolution structure had been published eight years earlier (Kendrew et al., 1958). It was also shown later e.g by Tilton et al., (1984, 1988) by linking xenon filled cavities and with MD calculations by Elber and Karplus (1990). Other members of the globin family (neuroglobin, hemoglobin) as well as nitrophorins (NO migration) show similar gas conducts (reviewed in Arroyo-Manez et al., 2011).

The migration of CO<sub>2</sub> in proteins has received far less attention and no cavity or tunnel for CO<sub>2</sub> transport in a protein have been described. There are fewer enzymes that utilize CO<sub>2</sub> as a substrate, but the lack of research is nevertheless surprising given the importance of carbon fixation. There is a limited amount of protein structures in the Protein Data Bank that bind CO<sub>2</sub> (reviewed in Cundari et al., 2009) and they often involve a residue with a strong hydrogen bond to the oxygen atoms in CO<sub>2</sub>, indicating that this is the preferred mode of stable binding of a gas molecule in a protein (and in some cases also for enzyme function). The migration of CO<sub>2</sub>, on the other hand, is helped by residue specific affinity, but not by strong binding, as is the case in the structures from the Protein Data Bank. The gas-protein interactions that define the migration route for CO<sub>2</sub> in Rubisco have a different character than the interactions found in the PDB, as they have to be more transient. There has been a computational study investigating the ability of proteins to act as CO<sub>2</sub> storage in the context of CO<sub>2</sub> emission compensation (Drummond et al., 2012), but this did not focus on finding a route from the solvent to the active site. The interaction of CO<sub>2</sub> around the active site in Rubisco has been investigated using quantum chemistry and molecular dynamics (El-Hendawy et al., 2012). Their approach was to start with a CO<sub>2</sub> molecule in the active site and follow it on its way out. They used a partial L subunit to assess the release of CO<sub>2</sub> from the active site and combined that with quantum mechanical calculations, rather than focusing on finding a defined migration route and concluded that CO<sub>2</sub> release and diffusion is a slow process (mainly due to Coulomb interaction)

along a narrow path to the solvent, possibly involving some protein surface contacts.

#### 4.6 CO<sub>2</sub> solubility

The discovery of the existence of tunnels in certain proteins coincided with the growth in number of protein structures in the Protein Data Bank. The mode of gas migration in numerous proteins and especially those that have a deeply buried active site, seems to be through tunnels, however this does not mean that gas is not diffusing in the bulk of the protein. In Rubisco the active site is located near the surface. There is very little guidance for CO<sub>2</sub> if a straight route from solvent to active site is considered. CO<sub>2</sub> is not a very soluble molecule (Wilhelm et al, 1977) therefore is expected to accumulate at the protein, initially at a random place at the surface. The pathway for CO<sub>2</sub> to the active site is described in Paper III. In a series of MD simulations the distribution of CO<sub>2</sub> was determined at various concentrations for four Rubisco structures. The same was done for O<sub>2</sub> in order to rationalize the relation between surface properties and specificity.

The CO<sub>2</sub> model used in Paper III consists of five points of which three have charge (0.94 in the center and -0.47 at either end) and the two extra points contain the masses (atomic weight 22 each) (Merz, 1991). The O<sub>2</sub> model used in Paper III has three points; two negative charges at the outer ends (-0.11), a positive charge at the center (0.22). The masses are located at the negative charge points (atomic weight 16 each) (Hub and De Groot, 2008). The Gibbs energy of hydration has been determined to be 0.7 kJ/mol for CO<sub>2</sub> and 7.75 kJ/mol for O<sub>2</sub>, indicating that CO<sub>2</sub> has higher solubility in water and that neither molecule is a good solute.

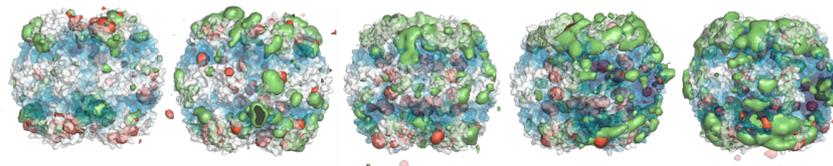


Figure 10. The effect of increases (from left to right) in  $\text{CO}_2$  and  $\text{O}_2$  concentration on their distribution in Rubisco from *C. reinhardtii*.  $\text{CO}_2$  density is in green;  $\text{O}_2$  density is in red; protein is in a gradient from white to blue, where darker blue indicates nearness to the active site. The viewpoint is on 45 degrees off the 2-fold axis, meaning that the center of each image shows an LL dimer with an active site at roughly 1/3 and 2/3 of the image height.

The effect of increases in concentration for the distribution of  $\text{CO}_2$  and  $\text{O}_2$  is visualized in Figure 10 for Rubisco from *C. reinhardtii*. The accumulation of  $\text{CO}_2$  in defined regions of Rubisco is much higher than that of  $\text{O}_2$  at similar concentration. Both gases diffuse into the protein, rather than remaining at the surface.  $\text{CO}_2$  can be found at any location during the course of simulation, with the most visited sites being around the large-small subunit interfaces and around the active site entrance.  $\text{O}_2$  is diffusing from solvent into Rubisco as well, but to less extent and in addition the distribution in Rubisco is in more isolated sites, seemingly more randomly distributed. This result is most interesting when recalling that the solubility of  $\text{CO}_2$  in water is higher than that of  $\text{O}_2$ . The difference in energy for Rubisco binding  $\text{CO}_2$  than  $\text{O}_2$  is more favourable by approximately 10 kJ/mol.

The highest density of  $\text{CO}_2$  in Rubisco is subsurface, just under the outer edges of the hexadecamer complex. A higher subsurface concentration of  $\text{CO}_2$  is also found at the inside of the protein, close to the central solvent channel. The lowest density of  $\text{CO}_2$  is found at the highest center of mass in the protein complex: the interior of the large subunits. This can be attributed to a higher viscosity of protein compared to water as a solvation medium. It is important to note, however, that within a 40 ns simulation virtually every amino acid is visited by at least one  $\text{CO}_2$  molecule. The accessibility of Rubisco for small gas molecules can be explained by side chain and small backbone movements of residues at room temperature. There are no large cavities where gas accumulates. There is also no measurable increase in protein size after taking up  $\text{CO}_2$  or  $\text{O}_2$ , which in itself is not too surprising, because of the relatively small volume of the gas molecules. Likely the gas molecules fill up existing voids that are created by the amino acid flexibility.

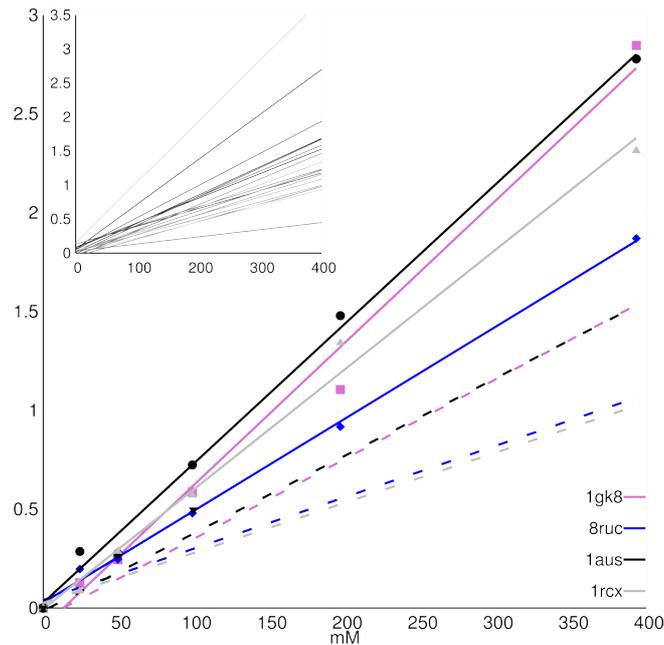


Figure 11. Affinity for CO<sub>2</sub> and O<sub>2</sub> of Rubisco from *C. reinhardtii* (PDB code 1gk8) and spinach (8ruc, 1aus, 1rcx). Shown is the cumulative distribution of gas within 6 Å of any Rubisco residue during 20 ns simulation as a function of the concentration of gas in mM. Dashed lines are for O<sub>2</sub>, solid lines for CO<sub>2</sub>. The inset shows an example CO<sub>2</sub> affinity per residue type from spinach.

Rubisco has a higher affinity for CO<sub>2</sub> than for O<sub>2</sub>, but this affinity is not equal for all residue types. Figure 11 illustrates this by showing the average affinity of Rubisco complexes for both CO<sub>2</sub> and O<sub>2</sub>. The individual affinity per residue type shows a larger spread than the averages (Figure 11 inset). This difference in affinity could in principle be exploited for engineering.

#### 4.7 Residue affinity for CO<sub>2</sub>

The residue specific affinity for CO<sub>2</sub> and O<sub>2</sub> has been investigated in Paper III. The affinity calculation is based on the number of gas molecule visits for each time-step within a 6 Å cutoff of an amino acid during simulation. The resulting value is influenced by simulation time and gas concentration and allows therefore only a relative comparison. The affinity of amino acids in Rubisco for

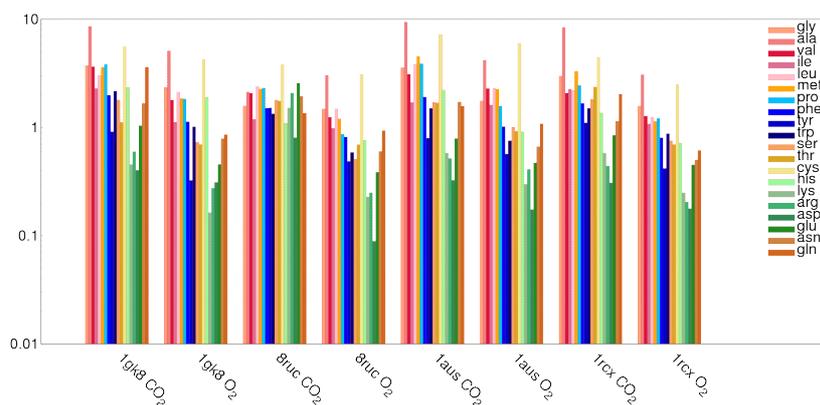


Figure 12. Affinity of amino acids in Rubisco for CO<sub>2</sub> and O<sub>2</sub>. Rubisco complexes are from *C. reinhardtii* (PDB code 1gk8) and spinach (8ruc, laus, 1rcx).

CO<sub>2</sub> and O<sub>2</sub> is shown (on a logarithmic scale) in Figure 12. It appears that small hydrophobic amino acids (alanine, valine, leucine, isoleucine) are most frequently visited by CO<sub>2</sub>. The larger hydrophobic amino acids (phenylalanine, tyrosine and tryptophan) are also visited frequently, but not as often (values are normalized). Side chain charge (arginine, lysine, aspartic acid, glutamic acid) does not benefit CO<sub>2</sub> affinity. Amino acids with a polar, but uncharged side chain (serine, threonine) attract CO<sub>2</sub> better than amino acids with a charged side chain. Compared with CO<sub>2</sub>, O<sub>2</sub> has lower affinity for all residue types, except cysteine, that apparently offers relatively favourable O<sub>2</sub> interaction.

The values are corrected for solvent accessible surface area, though interior residues have inherently less access due to higher viscosity of protein compared to water. This factor remains difficult to quantify, therefore extrapolation of these data to general amino acid affinity for CO<sub>2</sub> or O<sub>2</sub> is not justified.

#### 4.8 Subunit affinity for CO<sub>2</sub> and O<sub>2</sub>

As a consequence of the different affinity for CO<sub>2</sub> and O<sub>2</sub> of distinct amino acids, the large and small subunits will have different affinity for the gases, because they have different amino acid compositions. In *C. reinhardtii* the small subunit binds CO<sub>2</sub> 2.1 times more than the large subunit while O<sub>2</sub> is bound 1.4 times more. Thus both gases visit the small subunit more often than the large subunit, and CO<sub>2</sub> visits even more than O<sub>2</sub>. Amino acid substitutions

in the small subunit of Rubisco have been shown to improve specificity by enhancing the thermal stability of the enzyme (Spreitzer, 2003). Its amino acid composition may also be a factor that increases enzyme specificity.

The origin of the difference in binding affinity per subunit type could be due to selection pressure. Since the small subunits are not in direct contact with the active site, a higher mutational variation may be permitted. The different genomic coding of the small subunits (nuclear genome rather than chloroplast) allows for a higher mutational frequency. On the other hand, contacts required for holoenzyme stability may provide a constraint on the sequence variation.

#### 4.9 Route for CO<sub>2</sub> to the active site

The distribution of CO<sub>2</sub> in the simulations in Paper III provides a possibility to reconstruct a migration route to the active site. The absence of well-defined tunnels and cavities makes direct entry of CO<sub>2</sub> to Rubisco's active site from solvent less likely. A route along the protein surface is a possibility, yet the fact that protein provides a better solvation to CO<sub>2</sub> than water indicates that an interior route is the most favourable. The areas in hexadecameric Rubisco from *C. reinhardtii* and spinach with the highest CO<sub>2</sub> visitations are in the small and large subunits around the large-small subunit interfaces and around the active site entrance. These areas form a continuous area (in terms of similar amount of CO<sub>2</sub> visitations). The low solubility of CO<sub>2</sub> is a driving force towards migration into the protein. The gas will accumulate in the areas that favour contact, while at the same time the binding affinity will not be as high as to prevent further migration. From one of the large-small subunit interface areas migration towards the active site entrance is conceivable. The interconnection of this (large) high-affinity region and the (smaller) active site entrance suggests that the small subunit may act as a funnel for CO<sub>2</sub> migration. The migration of a CO<sub>2</sub> molecule can be along the surface of the protein, but the highest density of CO<sub>2</sub> is found underneath the surface, therefore a migration between the surface residues fits the data. In contrast, the regions with high O<sub>2</sub> affinity in Rubisco are not well connected and despite the fact that the active site entrance also has a high affinity for O<sub>2</sub>, migration of O<sub>2</sub> through high affinity regions seems more unlikely. The ability of Rubisco to guide its substrate with moderate binding affinity appears to enhance its specificity.

#### 4.10 Factors important for overall efficiency

Features such as altered dynamics, holoenzyme stability and CO<sub>2</sub> affinity ultimately depend on amino acid sequence variation. There are also a number of external factors influencing Rubisco activity.

Many organisms have evolved carbon concentrating mechanisms in order to increase local available CO<sub>2</sub> concentration (reviewed in Raven et al., 2012, Meyer and Griffiths, 2012). The ability to raise CO<sub>2</sub> concentration relieves the need for specificity.

Another factor is the ATP dependent protein Rubisco activase. The side reactions catalyzed by Rubisco lead to Rubisco-ligand complexes that are kinetically trapped. To facilitate substrate release, the active site can be opened by Rubisco activase in an ATP dependent manner and the ligand released (Wang and Portis, 1992, reviewed in Portis, 2008).

There are genetic factors such as the different encoding of small subunit (nuclear) and large subunit (chloroplast) that complicates hexadecamer formation. Signal sequences are required for transport of the small subunit through the chloroplast membrane and species-specific folding chaperones are necessary for enzyme folding and assembly (Li and Tabita, 1997, Bracher et al., 2011). These proteins in turn require regulation (which can be an engineering target itself, see e.g. Parry et al., 2013).

## 5 Conclusions

Rubisco has been studied for many decades, yet not everything of its molecular function is known. Understanding the influence of molecular structure and dynamics on its reaction mechanism may assist crop engineering. The increasing number of available crystal structures aids in understanding the consequences of sequence variation between species. Knowledge of the structure of highly specific Rubisco from arctic diatoms adds new information and may deliver clues to specific residues for engineering Rubisco into a more efficient enzyme. Diatom Rubiscos share a specific small subunit structure element that may tighten the interaction of the subunits in the holoenzyme. They show variation in amino acid sequence at specific sites within and between species. The structures are further characterized by the presence of a large number of intriguing residue modifications of unknown function. Rubisco from arctic algae has higher CO<sub>2</sub> specificity than Rubisco from crop plants. The specificity is however higher at lower temperature. This may provide a barrier to crop engineering, but needs to be tested.

The subunit interfaces of hexadecameric Rubisco can be grouped in seven types of which the large-large subunit LL1 is the largest. The large-small subunit interactions comprise a significant part of the large subunit's surface. Large-large subunit interactions are more hydrophilic than large-small subunit interactions. The size of the interaction area is subject to change in a dynamic structure, but the hydrophilic interactions vary even more so. This is illustrated by the tightness of the LL1 subunit in Rubisco structures with or without a substrate analogue in the active site.

The effect of a single mutation is difficult to detect, but is likely to result in changes in protein dynamics. It is difficult to assess which of these changes are responsible for altered specificity. Specificity restoration of double mutant L290F/A222T in *C. reinhardtii* exhibits not a return of wild type dynamics, but

shows another pattern, indicating that one or more other processes may play a role. The consequences of single mutation may be multiple and not straightforward to interpret. Decreasing small-small subunit interaction is however clearly connected to decreased holoenzyme stability.

The migration of CO<sub>2</sub> to the active site is a process that seems to be guided by the specific amino acid composition of Rubisco. Rubisco affinity for CO<sub>2</sub> is higher than for O<sub>2</sub> despite the higher water solubility for CO<sub>2</sub>. The small subunits have an even higher affinity for CO<sub>2</sub> than the large subunits. This leads to the conclusion that the small subunits may have a previously unconsidered property: that of CO<sub>2</sub> reservoir. It also implies that the CO<sub>2</sub> and O<sub>2</sub> affinity of altered amino acids needs to be considered when engineering Rubisco.

## References

- Andersson, I., Knight, S., Schneider, G., Lindqvist, Y., Lundqvist, T., Brändén, C.-I. & Lorimer, G. H. (1989). Crystal structure of the active site of ribulose-bisphosphate carboxylase. *Nature*, **337**, 229-234.
- 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**(1): 160-174.
- Andersson, I. and Taylor, T. C. (2003) Structural framework for catalysis and regulation in ribulose-1,5-bisphosphate carboxylase/oxygenase. *Archives of Biochemistry and Biophysics*. **414**, 130–140
- Arroyo-Manez, P., Bikiel, D. E., Boechi, L., Capece, L., Di Lelia, S., Estrin, D. A., Marti, M. A., Moreno, D. M., Nadra, A. D. and Petruk, A. A. (2011). Protein dynamics and ligand migration interplay as studied by computer simulation. *Biochimica Et Biophysica Acta-Proteins and Proteomics* **1814**, 1054-1064.
- Ashida, H., A. Danchin, et al. (2005). "Was photosynthetic RuBisCO recruited by acquisitive evolution from RuBisCO-like proteins involved in sulfur metabolism?" *Research in Microbiology* **156**(5-6): 611-618.
- Beer, C., M. Reichstein, et al. (2010). "Terrestrial Gross Carbon Dioxide Uptake: Global Distribution and Covariation with Climate." *Science* **329**(5993): 834-838.
- Berendsen, J.C., Postma, J. P. M., van Gunsteren, W. F. and Hermans, J., in *Intermolecular Forces*, edited by B. Pullman (Reidel, Dordrecht, Holland, 1981), p. 331.
- Berendsen, H. J. C., J. R. Grigera, et al. (1987). "The missing term in effective pair potentials." *Journal of Physical Chemistry* **91**(24): 6269-6271.
- Bracher, A., Starling-Windhof, A., Hartl, F. U., Hayer-Hartl, M. (2011). Crystal structure of a chaperone-bound assembly intermediate of form I Rubisco. *Nature Structural & Molecular Biology* **18**, 875-U1500.
- Chan, P. H. and S. G. Wildman (1972). "Chloroplast dna codes for primary structure of large subunit of fraction I protein." *Biochimica Et Biophysica Acta* **277**(3): 677.
- Chen, F. and P. E. Smith (2007). "Simulated surface tensions of common water models." *Journal of Chemical Physics* **126**(22): 3

- Cohen, J., Kim, K., King, P., Seibert, M. and Schulten, K. (2005). Finding gas diffusion pathways in proteins: Application to O<sub>2</sub> and H<sub>2</sub> transport in Cpl FeFe -hydrogenase and the role of packing defects. *Structure* **13**, 1321-1329
- Cundari, T. R., Wilson, A. K., Drummond, M. L., Gonzalez, H. E., Jorgensen, K. R., Payne, S., Braunfeld, J., De Jesus, M. and Johnson, V. M. (2009). CO<sub>2</sub>-Formatics: How Do Proteins Bind Carbon Dioxide? *Journal of Chemical Information and Modeling* **49**, 2111-2115.
- Degerlund, M. and Eilertsen, H. C. (2010). Main Species Characteristics of Phytoplankton Spring Blooms in NE Atlantic and Arctic Waters (68-80A degrees N). *Estuaries and Coasts* **33**, 242-269.
- Dill, K. A. (1993). "Folding proteins - finding a needle in a haystack." *Current Opinion in Structural Biology* **3**(1): 99-103.
- Dill, K. A., S. Bromberg, et al. (1995). "Principles of protein folding-a perspective from simple exact models." *Protein Science*. **4**: 561-602.
- Du, Y. C., S. J. Hong, et al. (2000). "RbcS suppressor mutations improve the thermal stability and CO<sub>2</sub>/O<sub>2</sub> specificity of rbcL-mutant ribulose-1,5-bisphosphate carboxylase/oxygenase." *Proceedings of the National Academy of Sciences of the United States of America* **97**(26): 14206-14211.
- Drummond, M. L., Wilson, A. K. and Cundari, T. R. (2012). Nature of Protein-CO<sub>2</sub> Interactions as Elucidated via Molecular Dynamics. *Journal of Physical Chemistry B* **116**, 11578-11593.
- Dubbs, J. M. and F. R. Tabita (2004). "Regulators of nonsulfur purple phototrophic bacteria and the interactive control of CO<sub>2</sub> assimilation, nitrogen fixation, hydrogen metabolism and energy generation." *Fems Microbiology Reviews* **28**(3): 353-376.
- Eilertsen, R. and Karplus, M. (1990). Enhanced sampling in molecular-dynamics - use of the time-dependent hartree approximation for a simulation of carbon-monoxide diffusion through myoglobin. *Journal of the American Chemical Society* **112**, 9161-9175.
- Ellis, R. J. (1979). "Most abundant protein in the world." *Trends in Biochemical Sciences* **4**(11): 241-244
- Esquivel, M. G., Anwaruzzaman, M. and Spreitzer, R. J. (2002). Deletion of nine carboxy-terminal residues of the Rubisco small subunit decreases thermal stability but does not eliminate function. *Febs Letters* **520**, 73-76.
- Field, C. B., M. J. Behrenfeld, et al. (1998). "Primary production of the biosphere: Integrating terrestrial and oceanic components." *Science* **281**(5374): 237-240.
- Garcia-Murria, M.G., Karkehabadi, S., Marín-Navarro, J., Satagopan, S., Andersson, I., Spreitzer, R.J. & Moreno, J. (2008). "Structural and functional consequences of the replacement of proximal residues Cys(172) and Cys(192) in the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase from *Chlamydomonas reinhardtii*." *Biochemical Journal* **411**: 241-247.
- Genkov, T., Y. C. Du, et al. (2006). "Small-subunit cysteine-65 substitutions can suppress or induce alterations in the large-subunit catalytic efficiency and holoenzyme thermal stability of ribulose-1,5-bisphosphate carboxylase/oxygenase." *Archives of Biochemistry and Biophysics* **451**(2): 167-174.
- Genkov, T. and Spreitzer, R. J. (2009) Highly conserved small subunit residues influence rubisco large subunit catalysis. *Journal of Biological Chemistry* **284**(44):30105–30112.

- Hartman, F. C. and M. R. Harpel (1994). "Structure, function, regulation, and assembly of d-ribulose-1,5-bisphosphate carboxylase oxygenase." *Annual Review of Biochemistry* **63**: 197-234.
- Haslam, R. P., Keys, A. J., Andralojc, P. J., Madgwick, P. J., Andersson, I., Grimsrud, A., Eilertsen, H. C. and Parry, M. A. J. (2005). Specificity of diatom Rubisco. *Plant Responses to Air Pollution and Global Change* (Omasa, K., Nouchi, I. and DeKok, L. J., Eds.), Springer-Verlag Tokyo, Tokyo.
- Hasle, G. R. and Syvertsen (1997). Marine diatoms. *Identifying marine phytoplankton* (ed. Thomas, C. R.). Academic Press, San Diego, pp 5-385.
- Heinhorst, S., Baker, S. H., Johnson, D. R., Davies, P. S., Cannon, G. C. and Shively, J. M. (2002). Two copies of form I RuBisCO genes in *Acidithiobacillus ferrooxidans* ATCC 23270. *Current Microbiology* **45**, 115-117.
- Hong, S. K. and R. J. Spreitzer (1997). "Complementing substitutions at the bottom of the barrel influence catalysis and stability of ribulose-bisphosphate carboxylase/oxygenase." *Journal of Biological Chemistry* **272**(17): 11114-11117.
- Hub, J. S. and De Groot, B. L. (2008). Mechanism of selectivity in aquaporins and aquaglyceroporins. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 1198-1203.
- Jorgensen, W. L., J. Chandrasekhar, et al. (1983). "Comparison of simple potential functions for simulating liquid water." *Journal of Chemical Physics* **79**(2): 926-935.
- Jorgensen, W. L. and J. Tirado-Rives (1988). "The OPLS Potential Functions for Proteins. Energy Minimizations for Crystals of Cyclic Peptides and Crambin." *Journal of the American Chemical Society* **110**: 1657-1666.
- Jordan, D. B. and W. L. Ogren (1983). "Species variation in kinetic-properties of ribulose 1,5-bisphosphate carboxylase oxygenase." *Archives of Biochemistry and Biophysics* **227**(2): 425-433.
- Kannappan, B. and Gready, J. E. (2008) Redefinition of rubisco carboxylase reaction reveals origin of water for hydration and new roles for active- site residues. *Journal of the American Chemical Society*, 130(45):15063-15080.
- Karkehabadi, S., Peddi, S.R., Anwaruzzaman, M., Taylor, T.C., Cederlund, A., Genkov, T., Andersson, I. & Spreitzer, R.J. (2005). "Chimeric Small Subunits Influence Catalysis without Causing Global Conformational Changes in the Crystal Structure of Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase." *Biochemistry* **44**(29): 9851-9861.
- Karkehabadi, S., Taylor, T.C., Spreitzer, R.J. & Andersson, I. (2005). "Altered Intersubunit Interactions in Crystal Structures of Catalytically Compromised Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase." *Biochemistry* **44**(1): 113-120.
- Karkehabadi, S., Taylor, T.C., Spreitzer, R.J. & Andersson, I. (2007). "Structural analysis of altered large-subunit loop-6/carboxy-terminus interactions that influence catalytic efficiency and CO<sub>2</sub>/O<sub>2</sub> specificity of ribulose-1,5-bisphosphate carboxylase/oxygenase." *Biochemistry* **46**(39): 11080-11089.
- Karplus, M. and A. Sali (1995). "Theoretical-studies of protein-folding and unfolding." *Current Opinion in Structural Biology* **5**(1): 58-73.

- Kawashim.N and S. G. Wildman (1972). "Studies on fraction I protein .4. mode of inheritance of primary structure in relation to whether chloroplast or nuclear dna contains code for a chloroplast protein." *Biochimica Et Biophysica Acta* **262**(1): 42-
- Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H. and Phillips, D. C. (1958). 3-dimensional model of the myoglobin molecule obtained by X-ray analysis. *Nature* **181**, 662-666.
- Kitano, K., N. Maeda, et al. (2001). "Crystal structure of a novel-type archaeal Rubisco with pentagonal symmetry." *Structure* **9**(6): 473-481
- Knight, S., Andersson, I. and Brändén, C. I. (1990). Crystallographic analysis of ribulose 1,5-bisphosphate carboxylase from spinach at 2.4 Å resolution. Subunit interactions and active site. *Journal of Molecular Biology*. 215, 113-160.
- Kostov, R. V., C. L. Small, et al. (1997). "Mutations in a sequence near the N-terminus of the small subunit alter the CO<sub>2</sub>/O<sub>2</sub> specificity factor for ribulose bisphosphate carboxylase/oxygenase." *Photosynthesis Research* **54**(2): 127-134.
- Lakowicz, J. R. and Weber, G. (1973). Quenching of protein fluorescence by oxygen - detection of structural fluctuations in proteins on nanosecond time scale. *Biochemistry* **12**, 4171-4179
- Laing, W. A., W. L. Ogren, et al. (1974). "Regulation of Soybean Net Photosynthetic CO<sub>2</sub> Fixation by the Interaction of CO<sub>2</sub>, O<sub>2</sub> and Ribulose 1, 5-Diphosphate Carboxylase1." *Plant Physiology* **54**: 678-685.
- Le Quere, C., M. R. Raupach, et al. (2009). "Trends in the sources and sinks of carbon dioxide." *Nature Geoscience* **2**(12): 831-836.
- Li, L. A. and Tabita, F. R. (1997). Maximum activity of recombinant ribulose 1,5-bisphosphate carboxylase/oxygenase of *Anabaena* sp. strain CA requires the product of the *rbcX* gene. *Journal of Bacteriology* **179**, 3793-3796,
- Lorimer, G. H., Badger, M. R. and Andrews, T. J. (1976). Activation Of Ribulose-1,5-Bisphosphate Carboxylase By Carbon-Dioxide And Magnesium-Ions - Equilibria, Kinetics, A Suggested Mechanism, And Physiological Implications. *Biochemistry* **15**, 529-536
- Mahoney, M. W. and W. L. Jorgensen (2000). "A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions." *Journal of Chemical Physics* **112**(20): 8910-8922.
- Merz, K. M. (1991). CO<sub>2</sub> binding to human carbonic anhydrase-II. *Journal of the American Chemical Society* **113**, 406-411.
- Meyer, M. T., T. Genkov, et al. "Rubisco small-subunit alpha-helices control pyrenoid formation in *Chlamydomonas*." *Proceedings of the National Academy of Sciences of the United States of America* **109**(47): 19474-19479.
- Montet, Y., Amara, P., Volbeda, A., Vernede, X., Hatchikian, E. C., Field, M. J., Frey, M. and FontecillaCamps, J. C. (1997). Gas access to the active site of Ni-Fe hydrogenases probed by X-ray crystallography and molecular dynamics. *Nature Structural Biology* **4**, 523-526.
- Morell, M. K., Kane, H. J. and Andrews, T. J. (1990) Carboxylterminal deletion mutants of ribulosebisphosphate carboxylase from *rhodospirillum- rubrum*. *Febs Letters*, 265(1-2):41-45.
- Nada, H. and J. van der Eerden (2003). "An intermolecular potential model for the simulation of ice and water near the melting point: A six-site model of H<sub>2</sub>O." *Journal of Chemical Physics* **118**(16): 7401-7413.

- Parry, M. A. J., Andralojc, P. J., Scales, J. C., Salvucci, M. E., Carmo-Silva, A. E., Alonso, H. and Whitney, S. M. (2013). Rubisco activity and regulation as targets for crop improvement. *Journal of Experimental Botany* **64**, 717-730.
- Pearce, F. G. (2006) Catalytic by-product formation and ligand binding by ribulose biphosphate carboxylases from different phylogenies. *Biochemical Journal*, 399:525–534.
- Peterson, R. B. (1989). "Partitioning of noncyclic photosynthetic electron-transport to O<sub>2</sub>-dependent dissipative processes as probed by fluorescence and CO<sub>2</sub> exchange." *Plant Physiology*. **90**(4): 1322-1328.
- Pons, T. L., Flexas, J., von Caemmerer, S., Evans, J. R., Genty, B., Ribas-Carbo, M. and Bruynoli, E. (2009). "Estimating mesophyll conductance to CO<sub>2</sub>: methodology, potential errors, and recommendations." *Journal of Experimental Botany*. **60**, 2217-2234.
- Portis, A. R., Li, C. S., Wang, D. F. and Salvucci, M. E. (2008). Regulation of Rubisco activase and its interaction with Rubisco. *Journal of Experimental Botany* **59**, 1597-1604.
- Quillfeldt, C. H., von (2001). Identification of easily confused common diatom species in Arctic spring blooms. *Botanica marina* **44**, 375-389.
- Scholes, R. J. and I. R. Noble (2001). "Climate change - Storing carbon on land." *Science* **294**(5544): 1012-1013.
- Spreitzer, R. J. and L. Metz (1981). "Photosynthesis-deficient Mutants of *Chlamydomonas reinhardtii* with Associated Light-sensitive Phenotypes." *Plant Physiology*. **67**(3): 565-569.
- Spreitzer, R. J., M. G. Esquivel, et al. (2001). "Alanine-scanning mutagenesis of the small-subunit beta A-beta B loop of chloroplast ribulose-1,5-bisphosphate carboxylase/oxygenase: Substitution at Arg-71 affects thermal stability and CO<sub>2</sub>/O<sub>2</sub> specificity." *Biochemistry* **40**(19): 5615-5621.
- Spreitzer, R. J. and M. E. Salvucci (2002). "Rubisco: structure, regulatory interactions, and possibilities for a better enzyme." *Annual Review of Plant Biology*. **53**: 449-75.
- Spreitzer, R. J., S. R. Peddi, et al. (2005). "Phylogenetic engineering at an interface between large and small subunits imparts land-plant kinetic properties to algal Rubisco." *Proceedings of the National Academy of Sciences of the United States of America* **102**(47): 17225-17230.
- Syvrtsen, E. E. (1977). *Thalassiosira rotula* and *T. gravida*: Ecology and morphology. *Nova Hedwigia*, Beih. **54**: 99-112.
- Tabita, F. R. (1999). Microbial ribulose 1,5-bisphosphate carboxylase/oxygenase: a different perspective. *Photosynthesis Research*. **60**, 1–28
- Tabita, F. R., T. E. Hanson, et al. (2007). "Function, structure, and evolution of the RubisCO-like proteins and their RubisCO homologs." *Microbiology and Molecular Biology Reviews* **71**(4): 576-+
- Takahashi, T., S. C. Sutherland, et al. (2002). "Global sea-air CO<sub>2</sub> flux based on climatological surface ocean pCO<sub>2</sub>, and seasonal biological and temperature effects." *Deep-Sea Research Part II-Topical Studies in Oceanography* **49**(9-10): 1601-1622.
- Taylor, T. C. and I. Andersson (1996). "Structural transitions during activation and ligand binding in hexadecameric Rubisco inferred from the crystal structure of the activated unliganded spinach enzyme." *Nature Structural Biology* **3**(1): 95-101.
- Tcherkez, G. G. B., G. D. Farquhar, et al. (2006). "Despite slow catalysis and confused substrate specificity, all ribulose biphosphate carboxylases may be nearly perfectly optimized."

- Proceedings of the National Academy of Sciences of the United States of America* **103**(19): 7246-7251.
- Tilton, R. F., Kuntz, I. D. and Petsko, G. A. (1984). Cavities in proteins - structure of a metmyoglobin-xenon complex solved to 1.9-Å. *Biochemistry* **23**, 2849-2857.
- Tilton, R. F., Singh, U. C., Kuntz, I. D. and Kollman, P. A. (1988). Protein ligand dynamics - a 96 picosecond simulation of a myoglobin xenon complex. *Journal of Molecular Biology* **199**, 195-211.
- Van der Spoel, D., P. J. van Maaren, et al. (1998). "A systematic study of water models for molecular simulation." *Journal of Chemical Physics* **108**: 10220-10230.
- Vega, C., J. L. F. Abascal, et al. (2009). "What ice can teach us about water interactions: a critical comparison of the performance of different water models." *Faraday Discussions* **141**: 251-276.
- Wang, Z. Y. and Portis, A. R. (1992). Dissociation of ribulose-1,5-bisphosphate bound to ribulose-1,5-bisphosphate carboxylase oxygenase and its enhancement by ribulose-1,5-bisphosphate carboxylase oxygenase activase-mediated hydrolysis of ATP. *Plant Physiology* **99**, 1348-1353.
- Watson, G. M. F. and F. R. Tabita (1997). "Microbial ribulose 1,5-bisphosphate carboxylase/oxygenase: A molecule for phylogenetic and enzymological investigation." *Fems Microbiology Letters* **146**(1): 13-22.
- Whitney, S. M., P. Baldett, et al. (2001). "Form I Rubiscos from non-green algae are expressed abundantly but not assembled in tobacco chloroplasts." *Plant Journal* **26**(5): 535-547.
- Whitney, S. M., Houtz, R. L. and Alonso, H. (2011). Advancing Our Understanding and Capacity to Engineer Nature's CO<sub>2</sub>-Sequestering Enzyme, Rubisco. *Plant Physiology* **155**, 27-35.
- Wilhelm, E., Battino, R. and Wilcock, R. J. (1977). Low-pressure solubility of gases in liquid water. *Chemical Reviews* **77**, 219-262.
- Zhu, X. G., A. R. Portis, et al. (2004). "Would transformation of C-3 crop plants with foreign Rubisco increase productivity? A computational analysis extrapolating from kinetic properties to canopy photosynthesis." *Plant Cell and Environment* **27**(2): 155-165.

## Acknowledgements

This PhD work is as much the work of others as it is mine. Without the help and support of numerous people this thesis would not have been lying in front of you. For this I would like to acknowledge a number of colleagues. First and most of all, I would like to thank **Inger Andersson** for taking me on this scientific journey, for continuous support and inspiration. Your scientific knowledge of Rubisco, its biochemistry and biology, was of great importance to me, and helped forming the questions that need to be answered. Your experience in the unwritten workings of science and academia guided me to many interesting observations.

I also would like to thank **David van der Spoel** for co-supervising my work and bringing the computational expertise to help me work out the technical part of the thesis work. Despite being from the same country I think in all these years we did not have any conversation without any English or Swedish word popping in.

I would like to thank **Karin Valegård** for stepping in as acting supervisor. I appreciate the effort you gave to let me smoothly finish off my PhD.

Outside the supervisory trio there have been many colleagues over the years who in one way or the other, contributed to shaping this PhD. Many have already moved on to other location or career and I apologize beforehand if I forget to mention anyone. I thank **Janos Hajdu** for letting me hang around in his group without actually working for him. I thank the people I shared office with: **Filipe** for his resourceful knowledge of anything computer related, **Elsa** for being such pleasant and cheerful companion, **Alexandra** for energy, and organization, **Bianca** for all off work fun, **Min Yang** for being thoughtful, **Johan** for his optimism and **Marvin** for his unique company.

I thank the current and past members of the **Hajdu group/corridor** (in no particular order): **Max, Gijs, Tomas, Dirk, Gunilla, Margareta, Martin S, Jakob, Claudia, Janne, Hemanth, Kerstin, Carl, Jessica, Francisca, Kenta, Asawari, Tony, Sara, Charlotte, Daniel W, Calle, Gösta, Magnus, Rosie, Jorge, Johan R, Diane, Anke, Remco, Jochen, Erik, Tiago, Dusco, Malin** and **Nic**. And this also includes the past and present members of David's MD crowd: **Erik, Daniel L, Gadzikano, Grace, Tang, Per L, Daniel D** and **Özge**. It has been a pleasure working in this strange mixture of expertise, genius and nerdyness.

From the past and present crystallography crowd from SLU/UU: Special thanks to **Saeid** and all the work that he performed during his PhD: it has been a great resource for my work. Thanks to **Tom** for teaching me structure refinement. Thanks to **Mats** for taking well care of the department in a challenging time. **Ellenor** for managing my paper work and making the department run like a clock. I thank **Henrik H. Nisse, Torleiv, Henrik B.** and co for keeping the department strong. I thank **Majid** for being such a nice guy. I thank **Terese** for her interesting lectures and keeping fish. I thank **Christoffer, Nina, Anatoly, Mikael, Alwyn, Sherry, Anna J, Henrik I, Stefan, Nils 2, Avinash, Maria, Sanjee, Jerry, Adrian, Anna SL, Ana Laura** for their nice company at lunches and elsewhere.

I would also like to thank a number of moved-on members: **Evalena** for always wearing fashionable colour, **Glareh** for her great company and humour, **Talal** for helping me solve my first space group and decide to stay in Uppsala in this awkward moment in the swimming pool, **Emma** for sensibility, **Daniel E** for being such a sport, **Åsa** for her enthusiasm, **Lena** for all her words, **Annette, Alina, Lars** for explaining phases and fourier transforms and also for hiring **this Czech guy** that I will mention later on, **Gerard** for being accurate and precise, **Marian** for being part of the Czech subcommunity, **Magnus P** (or **J**) for endless optimism, **Ulla** for organizing the seminars, **Hasse** for introducing me to the secrets of protein structures, **Torsten** for his amazing voice, **Jonas** for always finding the right words to say and fun at teaching, **Gunnar “&” Jenny Berglund** for being such nice people, **Al M-H** for introducing me to English words I did not know exist (or probably only do in Tasmania), **Ola** for guiding me on my first teaching mission, **Fariborz** for his sincerity to others, **Kaspers** for viruses, **Wimal** for looking like the most experienced scientist of us all, **Martin W** and **Louise** for being nice, **Linda** for co-writing a paper, **Anton** for exciting seminars, **Wojciech** for his unrestrained opinions on work and everything else. Thanks to **Patrik, Urszula, Jimmy, Martin A, Martin Hä, Martin Hö** and all the other **Martins**.

Thanks to the social non-coding RNA guys: **Fredrik S, Andrea H, Lotta, Heiko, and Steffi.**

Thanks to the Chinese invasion force: **Miau, Yang, Yafei, Xiaohu, Xiaodi, Zhaolin, Li, Lu, Wangshu;** you have all been a great company.

I thank **ChaSan** for useful discussions and motivating me to writing my first paper.

I thank \*the other\* computational guys, **Samuel, Johan Åqvist** and his lads: **Hugo, Massoud, Henrik K, Yasmin, Lars P, Johan S, Fredrik Ö, Sinisa** and also **Göran.**

I thank **Tyson** and **Marek** for reasons totally unrelated to work.

All these people and me would be lost without **Mark, Christer** and **Erling:** the people with the least papers in the lab are probably the most important.

I thank **Pavel** for coming to Uppsala to do a PhD and introducing me to numerous Czech cultural icons in our flat (one of which still gives me headaches). I thank him for introducing this curious deck of cards that still performs a central role in the life of me and my friends. Most of all I thank him for being there when needed. One former colleague deserves a special mention: I thank **Agata** for help in teaching, for lunches, Polish lessons and becoming one of my best friends in Uppsala.

In the scope of acknowledgements I would like to include two not work related people. I would like to thank my parents for their support and kindness over all the years.