JBC JOURNAL OF BIOLOGICAL CHEMISTRY

FOR PEER REVIEW - CONFIDENTIAL

Unusual posttranslational modifications revealed in crystal structures of diatom Rubisco

Tracking no: JBC/2018/003518R1

Author Affiliation:
Karin Valegård (Uppsala University,), P. Andralojc (Rothamsted Research, United Kingdom), Richard Haslam (Rothamsted Research, United Kingdom), F. Grant Pearce (University of Canterbury, New Zealand), Gunilla Eriksen (Arctic University of Norway, Norway), Pippa Madgwick (Rothamsted Research, United Kingdom), Anne Kristoffersen (University of Oslo, Norway), Michiel van Lun (Uppsala University, Sweden), Uwe Klein (University of Oslo, Norway), Hans Eilertsen (Arctic University of Norway, Norway), Martin Parry (Lancaster University, United Kingdom), Inger Andersson (Uppsala University, Sweden)

Abstract:

Competing interests:

Author contributions:
Karin Valegård: Formal analysis; Investigation; Writing-original draft; Writing-review and editing P. Andralojc: Data curation; Formal analysis; Investigation; Writing-original draft; Writing-review and editing Richard Haslam: Formal analysis; Investigation; Writing-review and editing F. Grant Pearce: Formal analysis; Investigation; Writing-review and editing Gunilla Eriksen: Formal analysis; Investigation; Writing-review and editing Pippa Madgwick: Formal analysis; Investigation; Writing-review and editing Anne Kristoffersen: Formal analysis; Investigation; Writing-review and editing Michiel van Lun: Formal analysis; Investigation; Writing-review and editing Uwe Klein: Conceptualization; Resources; Data curation; Formal analysis; Supervision; Investigation; Writing-original draft; Writing-review and editing Hans Eilertsen: Conceptualization; Resources; Formal analysis; Supervision; Funding acquisition; Investigation; Writing-original draft; Writing-review and editing Martin Parry: Conceptualization; Resources; Formal analysis; Funding acquisition; Investigation; Writing-review and editing Inger Andersson: Conceptualization; Resources; Data curation; Formal analysis; Supervision; Funding acquisition; Investigation; Visualization; Writing-original draft; Project administration; Writing-review and editing

Funding:

Reviewer Instructions

Structural and functional analyses of Rubisco from arctic diatom species reveal unusual posttranslational modifications

Karin Valegård[1], P. John Andralojc[2], Richard P. Haslam[2], F. Grant Pearce[1, 5], Gunilla K. Eriksen[3], Pippa J. Madgwick[2], Anne K. Kristoffersen[4,6], Michiel van Lun[1], Uwe Klein[4], Hans C. Eilertsen[3], Martin A. J. Parry[2, 7], and Inger Andersson[1]*

From the [1]*Department of Cell and Molecular Biology, Uppsala University, Box 596, S-751 24 Uppsala, Sweden;* [2]*Plant Biology and Crop Science, Rothamsted Research, Harpenden, Herts, AL5 2JQ, UK;* [3]*Norwegian College of Fisheries Science, the Arctic University of Norway, N-9037 Tromsø, Norway and* [4]*Department of Molecular Biosciences, University of Oslo, P.O. Box 1041, Blindern, N-0316 Oslo, Norway*

Running title: Rubisco from Arctic Diatoms

Present Address: [5]*School of Biological Sciences, University of Canterbury, Private Bag 4800, Christchurch, New Zealand;* [6]*Department of Oral Biology, Faculty of Dentistry, University of Oslo, Norway*; [7]*Lancaster Environment Centre, Lancaster University, LA1 4YQ Lancaster, UK.*

*To whom correspondence should be addressed. Inger Andersson, Department of Cell and Molecular Biology, Uppsala University, Box 596, S-751 24 Uppsala, Sweden; inger.andersson@icm.uu.se; Tel. +46-18-4714288. Fax. +46-18-511755.

Keywords: carbon fixation, Rubisco, diatoms, CO_2/O_2 specificity, crystal structure, post-translational modifications.

ABSTRACT

The catalytic performance of the major CO_2-assimilating enzyme, ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco), restricts photosynthetic productivity. Natural diversity in the catalytic properties of Rubisco indicates possibilities for improvement. Oceanic phytoplankton contain some of the most efficient Rubisco enzymes, and diatoms in particular are responsible for a significant proportion of total marine primary production as well as being a major source of CO_2 sequestration in polar cold waters. Until now, the biochemical properties and three-dimensional structures of Rubisco from diatoms were unknown. Here, diatoms from Arctic waters were collected, cultivated and analyzed for their CO_2 fixing capability. We characterized the kinetic properties of five, and determined the crystal structures of four Rubiscos selected for their high CO_2-fixing efficiency. The DNA sequences of the *rbc*L and *rbc*S genes of the selected diatoms were similar, reflecting their close phylogenetic relationship. The V_{max} and K_M for the oxygenase and carboxylase activities at 25°C and the specificity factors ($S_{c/o}$) at 15, 25 and 35°C, were determined. The $S_{c/o}$ values were high, approaching those of mono- and dicot plants, thus exhibiting good selectivity for CO_2 relative to O_2. Structurally, diatom Rubiscos belong to Form I C/D, containing small subunits characterised by a short βA-βB loop and a carboxy-terminal extension that forms a β-hairpin structure (βE-βF loop). Of note, the diatom Rubiscos featured a number of posttranslational modifications of the large subunit, including 4-hydroxy-proline, beta-hydroxyleucine, hydroxylated, and nitrosylated cysteine, mono-, and di-hydroxylated lysine, and tri-methylated lysine. Our studies suggest adaptation toward achieving efficient CO_2-fixation in Arctic diatom Rubiscos.

Ribulose-1,5-bisphosphate carboxylase/oxygenase (EC 4.1.1.39, Rubisco) is the principal enzyme responsible for the assimilation of CO_2 into the biosphere. It catalyses the primary photosynthetic CO_2 reduction reaction, the carboxylation of ribulose-1,5-bisphosphate (RuBP) by CO_2. The product of this reaction, 3-phosphoglycerate (3PGA), is subsequently converted into biomass. Rubisco is found in all photoautotrophic organisms including photosynthetic bacteria, archaea, algae and plants. The oceans are of particular

importance for the global carbon cycle and are estimated to have a net uptake flux of circa 2 petagram (Pg) of carbon per year (1). This takes place as pressure difference driven diffusion influenced by wind and enhanced by sequestration by marine phytoplankton. Oceanic contribution to the global net primary production (NPP) has been estimated to 45-50 Pg carbon per year (2), which amounts to circa 45% of the total NPP. This is remarkable, considering the low availability of free dissolved CO_2 and its slow diffusion in sea water, but is connected to the overall fast growth rates of phytoplankton often leading to >> 50% increase in biomass per day (3).

Carbon fixation resulting from Rubisco's activity amounts to more than 10^{11} tons of atmospheric CO_2 annually (2). However, Rubisco is an inefficient catalyst because of its low turnover rate and its tendency to catalyse a reaction with O_2 rather than CO_2, leading to the oxygenation of its sugar phosphate substrate to yield 2-phosphoglycolate (2PG). This inherent oxygenase activity leads to a significant loss of carbon to the atmosphere, and a decrease in the carboxylation efficiency (for reviews, see e.g. refs. 4-6). Because of the significance of Rubisco to crop production, plant nitrogen and water usage, and the global carbon cycle, there is considerable interest in investigations aimed at reducing the oxygenase activity. Rubisco has been the subject of intense research, including structural, mechanistic, and mutagenesis studies. However, despite the vast amount of data available, the molecular basis for its CO_2/O_2 discrimination is not fully understood.

The ratio of carboxylation and oxygenation, measured as the CO_2/O_2 specificity factor (see further below), is not fixed, and there is substantial variation among phototrophs (7). For instance, the specificity factor is very low (around 20) in anoxygenic nonsulphur purple bacteria, but considerably higher (ca 80-100) in Rubisco from green plants. Form I Rubisco, to which plant Rubiscos belong, can be further divided into two subgroups: green-like, containing higher plants, green algae and cyanobacteria, and red-like, containing among others eukaryotic non-green algae, i.e. diatoms and rhodophytes (8-10). The latter group contain some of the most CO_2 efficient forms of Rubisco.

The genetic, phylogenetic, kinetic, and structural characteristics of red-like Rubiscos from marine organisms are to a large extent unknown. For example, little is known about Rubisco from psychrophilic organisms that live in Arctic waters. The objective of the present work was therefore to study Arctic/coldwater microalgae to provide new information on Rubisco function at the molecular level. We have studied the natural variation in Rubisco from northern diatoms, which thrive at the light limited low temperature environments within and below the ice and make up the main part of primary production in fish rich areas (11).

Results

Growth experiments

The growth data were standardized in order to compare measurements obtained by different methods (see Experimental procedures), and hence only serves the purpose to compare the species in a relative manner. The mean growth rate obtained from the different methods was 0.47 doublings day^{-1}, with minimum ca. 0.05 and maximum ca. 1.2 doublings day^{-1}. The expected maximum doubling rate at 3-4 ^{o}C is ca. 1.0 doublings per day (12); considering that we used a L:D photoperiod of 14:10, we conclude that the growth achieved for our seven chosen species was in the maximum range. The results from the growth rate experiments showed that the overall fastest grower at 2-3 ^{o}C was *Thalassiosira antarctica* (interpreted from the average growth data). In this temperature range it was followed by *Bacterosira bathyomphala* and *T. nordenskioeldii* (Table 1). The slowest growers in this temperature range were *T. gravida* and *T. hyalina*. At 7 ^{o}C the fastest growers were *T. antarctica*, *B. bathyomphala* and *T. nordenskioeldii*. When both temperatures were considered the fastest growers were *T. antarctica*, *B. bathyomphala* and *Chaetoceros socialis,* thus *T. antarctica* performed best at both temperatures. The overall mean increase in growth rates from low to high temperature regimes were ca. 0.03 (in standardized relative units), and only *C. socialis*, *T. nordenskioeldii*, *T. gravida* and *Skeletonema marinoi* responded with increased growth rates when the temperature increased (Table 1).

Determination of kinetic constants of Rubisco enzymes from Arctic diatoms

The CO_2-fixation efficiency of Rubisco shows considerable species-specific variation (13). Our objective was to identify the most efficient Rubiscos among diatoms, a group of microalgae that are prime candidates for finding

new highly efficient Rubisco enzymes. The partitioning between the carboxylation and oxygenation reactions (ν_C/ν_O) is dependent on the relative concentrations of the gaseous substrates and the relative catalytic efficiencies (V_{max}/K_M) of the two activities in accordance with the following relationship: $\nu_C/\nu_O = (V_cK_o/V_oK_c)$ ($[CO_2]/[O_2]$) where ν_C and ν_O are the velocities of carboxylation and oxygenation, respectively, V_c and V_o the maximal velocities of the two reactions, and K_c and K_o the Michaelis constants for CO_2 and O_2, respectively. The composite of constants in the equation is referred to as the specificity factor, and often referred to as $S_{c/o}$ τ, or Ω (14). The specificity factor is usually determined from the product of the measured 3PGA/2PG concentrations and the known $[CO_2]/[O_2]$ ratio.

Optimisation of a Rubisco purification procedure for use with marine diatoms was undertaken, and a suitable protocol was developed that resulted in over 80% pure Rubisco. Diatom Rubisco content was generally much lower than in plants, confirming earlier observations (15).

HPLC analysis of ^{14}C-ribulose bisphosphate (RuBP) oxygenation and carboxylation was first evaluated to examine diatom Rubisco CO_2/O_2 specificity. This method is labour intensive, highly sensitive to relatively small changes in 3PGA and 2PG concentrations, and requires tightly controlled reaction conditions. Therefore a method based on the oxygen electrode was employed giving real time data collection. Wheat Rubisco was used as an internal standard. In addition K_M and V_{max} were determined for Rubisco from diatom species with wheat as a control.

Using assays and screening protocols especially developed for diatom Rubisco enzymes, specificity factors were determined for five diatom species at a range of temperatures from 15 - 35 ^oC (Table 2, ref. 16). The specificity factors of diatom Rubiscos were close to that for wheat Rubisco. In all of the Arctic species examined the specificity factor increased at decreasing temperatures. None of the Arctic species examined had a higher specificity factor than wheat, even when values were extrapolated to 0 ^oC. However, unlike wheat Rubisco, diatom Rubiscos were not deactivated when exposed for prolonged periods (~ 24 h) to temperatures of 4 ^oC (data not shown). These observations suggest structural adaptations to the low temperatures in the extreme environment these diatoms occupy.

Crystal structures of Rubisco enzymes from Arctic diatoms

Crystals of diatom Rubisco species were obtained and the corresponding structures were determined. Details of data collection and refinement are summarised in Table 3. Overall the quality and resolution of the data were very good, with the best crystals diffracting to better than 2 Å resolution. However, because some crystals were thin in at least one dimension, the corresponding data were anisotropic. The quality of the structures was significantly improved by the use of TLS refinement implemented in the refinement programs REFMAC5 and PHENIX, but the quality of the *T. nordenskioeldii* and *B. bathyomphala* structures remained sub-standard and were not included in the final set of structures. The RbcL sequence from *T. nordenskioeldii* (O98947) was used for an initial fit to the electron density maps and subsequently modified to fit the density as refinement progressed. In parallel to this, genomic DNA was extracted from the cell cultures and partial sequences of *rbc*L and *rbc*S were determined to aid model building (Supporting Fig. 1). The sequences of Rubisco from diatoms in this study were highly similar – as would be expected in view of their close relationship. GUG is the translation start codon of all *rbc*S genes sequenced. This codon normally codes for valine but the protein structures show that, as expected, methionine was inserted in this position.

Diatom Rubisco (Fig. 1A) is a hexadecamer of eight large (L, 490 residues) and eight small (S, 139 residues) subunits and belong to form I C/D (reviewed in ref. 5). This form includes a small subunit that is distinct from the small subunits of form I A/B enzymes (in e.g. cyanobacteria and higher plants) and is characterised by a short βA-βB loop and a carboxy-terminal extension (βE-βF loop) that forms a β hairpin structure. The β hairpins from four small subunits together form a β-barrel that lines the entrance to the central solvent channel at each end of the holoenzyme (Fig. 1B). Form I C/D structures have previously been observed in Rubisco from the betaproteobacterium *Cupriavidus necator* (formerly *Ralstonia eutropha*), and the red algae *Galdieria partita* and *G. sulphuraria* (17-19). The diatom structures are highly similar: structures can be

superimposed with root mean square deviations (r.m.s.d.s) of 0.15-0.32 Å for all Cα atoms.

Posttranslational modifications

The structures of diatom Rubisco feature a number of posttranslational modifications in the large subunit (Table 4, Figs. 2, 3). Rubisco is activated by carbamoylation of the ε-amino group of an active-site lysine residue and subsequent coordination to Mg^{2+} (20, 21). Thus, as expected for the fully-activated enzyme examined in the present study, electron density corresponding to a carbamoyl group is observed at the ε-amino group of Lys205 (corresponding to Lys201 of spinach Rubisco). 4-Hydroxy-Pro residues are present at positions 48 and 155. Hydroxy-Pro-155 (Fig. 2A) corresponds to hydroxy-Pro151 of Rubisco from the green algae *Chlamydomonas reinhardtii* (22), whereas hydroxy-Pro48 appears to be unique to diatom Rubiscos. Both residues are relatively buried in the interior of the protein. Electron density corresponding to a modification of the sulphur atom of Cys109 was detected in some (*T. antarctica*, *T. hyalina*, *B. bacterosira*), but not all diatoms. Analysis of this extra density shows it is most consistent with hydroxylation. A large extra density at Sγ of Cys457 suggested a different modification; we first considered carbamylation or methylation (methylcysteine was detected in Rubisco from *C. reinhardtii*, ref. 22), but analysis of side-chain atom temperature factors and difference density maps after refinement indicated such modifications to be unlikely. Instead, nitrosylation of the Cys-sulphur was found to best fit the extra density (Fig. 2B). The S-nitroso group of Cys457 is accessible to solvent, suggesting that it may be involved in interactions with binding partners. Nitrosylation at Cys457 was most prominent in Rubisco from *C. socialis*, but the modification could also be detected at lower occupancy in the enzymes from *T. antarctica* and and *T. hyalina* (Table 4). It was only faintly detected in the enzyme from *S. marinoi*; this is likely because of the low resolution of the data. Lys150 features additional density at Cγ and Cδ most consistent with hydroxylation (Fig. 3). This is a relatively unusual modification that has not been reported previously. Lys150 is located on the dimer interface of the large subunits, and forms several hydrogen bonds with its 3- and 4-hydroxyl groups to Ser147 of the adjacent subunit (Fig. 3B). Such contacts have been shown to influence stability, catalysis and specificity in Rubisco (23-25). It is therefore likely that these interactions, which would not be present in the unmodified enzyme, will confer extra stability to the holoenzyme. Additional density at Cδ of Lys198 was interpreted as mono-hydroxylated lysine. Lys346 shows extra density at its Nε corresponding to trimethylation (Fig. 2C). Trimethyl-Lys346 is located at the exterior of the hexadecamer and is accessible to the solvent. Trimethyl-lysine has been detected at position 14 of some plants (26), although it has not yet been observed in a crystal structure, presumably due to disorder of the N-terminus. Trimethylation of residue 346 appears to be unique to the present structures. Leu174 is hydroxylated at Cß (Fig. 2D); the modification introduces an additional hydrogen bond contact to the main chain nitrogen of Asp202. As mentioned above, *rbc*S sequences deduced from the crystal structure differ from the DNA sequence at residue 1. All of the modifications are unambiguous for each of the four-eight copies in the asymmetric unit.

Discussion

Finding a Rubisco enzyme that has its carboxylation reaction enhanced relative to its oxygenase reaction and engineering this trait into the Rubisco enzymes of economically important crop plants has potential implications both with regard to agricultural and environmental considerations. Besides increasing yield, it would potentially allow growing of crop plants in areas with a short growing season, i.e. short summers, and – in areas with longer growing seasons – to obtain more than one harvest per season. Current concerns regarding global warming and the greenhouse effect point to the need for a better understanding of global carbon fluxes, in particular in the oceans and between the ocean and the atmosphere.

Little is known about the biochemical properties of Rubisco from marine microorganisms, which are estimated to be responsible for about half of the total NPP. Initial findings indicate that Rubisco enzymes from marine microalgae carry a number of unusual features, which make them prime candidates for further studies. Young *et al.* (27) reported the kinetic constants of Rubisco enzymes from a set of diatoms, which were all from southern origin and had a relatively large geographic spread. Much less has been reported about the properties of Rubisco from northern diatoms, and no

structures of diatom Rubisco have been described so far.

Specificity factors measured from diatom Rubisco are high relative to those of cyanobacteria. Although the specificity factor serves as an important first diagnostic parameter to indicate changes in efficiency of engineered Rubisco enzymes, it is but one parameter that determines the net efficiency of Rubisco enzymes. It is becoming increasingly obvious that environmental factors, such as the temperature and the aridity of the environment from which the organism evolved are important factors that influence Rubisco's carboxylation capacity (28). In the case of marine phytoplankton, CO_2 and light limitations are important factors to consider. Phytoplankton have adopted carbon concentrating mechanisms (CCM) to offset the problems of CO_2 limitation and use the high levels of bicarbonate in sea water (29, 30). Evidence for a CCM in diatoms to date is mainly from model diatoms (31), whereas direct evidence for a CCM in northern diatoms is currently lacking. Common with previously analysed red-type Rubisco enzymes, the northern diatoms show a reduced affinity for O_2 (Table 2), but lack the very high affinity for CO_2 observed for non-green algae such as *Griffithsia monilis* (32, 33). This, together with the low concentration of free dissolved CO_2 in sea water, would point to the need for a CCM. The operation of a CCM may increase photosynthetic light requirements. Thus, it seems possible that the evolution of high specificity factors in diatoms (compared to e.g. ocean-living cyanobacteria) may contribute to their ability to grow well in the light-limited environment typical of the early bloom or under the Arctic ice or during periods, when the maximum solar elevation is low for longer periods (34). As the catalytic efficiency of Rubisco increases one would expect that less nitrogen (as the constituent amino acids of Rubisco) would be required to maintain a given photosynthetic rate. From our results, the species that had the highest specificity factor relative to the others, *T. hyalina*/*T. antarctica*, also had the highest overall growth rate and the highest growth rate at the lowest (2 ^{o}C) temperature. In addition, the diatom with the lowest specificities, *S. marinoi*, is considered a more southern species that does not normally enter the true Arctic growth regime (11). This, in our opinion, indicates that high specificity Rubisco may be a cold water/Arctic evolutionary adaptation connected to competitive advantages. Hobson *et al*. (35) have reported high specific activities and low cellular concentrations of Rubisco in diatoms relative to green algae, illustrating the coupling between carbon assimilation and nitrogen metabolism. Although speculative, improvements in Rubisco specificity would be ecologically significant if they affect the competitive ability of a species.

The Rubisco *rbc*L/S sequences obtained from genomic DNA extracted from the cultured diatom material were generally confirmed by the amino acid sequences deduced from the electron density maps. Most of the differences may not be of significant consequence for the function of the enzyme; for instance, the commonly observed Ile/Val substitution (or Ile/Leu) belong to the same class of apolar amino acids with similar physico-chemical properties.

T. nordenskioeldii, *T. gravida* and *T. antarctica* are common in the northern cold water to temperate regions (11), whereas *T. hyalina* is described as an Arctic species (36). It is well known that it is difficult to distinguish morphologically between *Thalassiosira* species, e.g. the morphologically determined identity of *T. gravida* may be confused by the fact that it may change morphology when the temperature is lowered, whereby it resembles *T. rotula* (37). *T. gravida* may also easily be misidentified with *T. antarctica* (38). There are also different "types" of *T. antarctica;* the one cultivated here probably most resembles *T. antarctica* var. borealis (39). In addition, the genetic information for the group is largely incomplete, and, as a consequence, the available species concepts may be incomplete, and phenotypic (and physiological) adaptation may well occur over short intervals of time. Considering the more southern origin of the diatoms that have been studied to date, the differences that we observe in this study may well be due to true variation occurring in the species collected in Arctic/north Atlantic waters.

There is also the possibility that certain species may carry several copies of *rbc*L and/or *rbc*S genes and that these may be expressed differentially. Plants and green algae are known to have multiple nuclear-encoded *rbc*S genes, for instance wheat carries over 20 *rbc*S genes whereas *C. reinhardtii* has 2 copies (reviewed in ref. 40). Some prokaryotes even have multiple copies of both *rbc*S and *rbc*L genes (41). Multiple copies are assumed to be the result of lateral gene transfer (8), but this has not been addressed specifically for diatoms. In addition, some chloroplasts have been found to exhibit maternal, paternal, and biparental modes of

inheritance even within the same species, the latter has been shown in e.g. *Pseudo-nitzschia* (42). While it is presently not possible to distinguish between these alternatives, it is not unlikely that the conditions in a mass culture may give rise to sequence variations (multiple sequences) in a manner observed here.

While the carbamylation of an active-site lysine residue has been established as essential for activity, the roles of other posttranslational modifications of Rubisco have not been elucidated with regard to functional significance (reviewed in refs. 43, 44). Hydroxyproline is a major component of collagen, where the absence of the hydroxyl group on proline (caused by a deficiency in vitamin C) results in the disease scurvy. Hydroxyproline is also found in diverse proteins localised to the plant cell wall (45, 46), but this residue has never been observed in Rubisco from vascular plants. While S-hydroxycysteine is detected for the first time in Rubisco, 4-hydroxyproline and S-methylcysteine have been observed earlier in Rubisco from the unicellular green alga *C. reinhardtii* (22), but there is yet no known function for these modifications in algae. Redox regulation of Rubisco activity via cysteine residues has been extensively studied in *C. reinhardtii* (47). Nitric oxide (NO) signaling regulates various physiological processes in animals, plants and algae. In the diatom *S. costatum*, a link has been found between nitric oxide concentrations and programmed cell death (48), suggesting a role for nitric oxide in the massive cell loss occurring at the end of a diatom bloom. If nitrosylation of Rubisco at Cys457 is part of this mechanism remains to be elucidated, but we note that S-nitrosocysteine has also been detected at the corresponding position (Cys460) in Rubisco from the red algae *G. sulphuraria* (19). The presence of mono-, di-hydroxylated or tri-methylated lysines in the diatom Rubisco enzymes investigated in this study is also enigmatic. Hydroxylysines have been detected in the hydrolysates of peptides and proteins exposed to $HO•/O_2$, and subsequently treated with $NaBH_4$ (49), but such oxidising conditions are difficult to imagine in the live diatom cell. Hydroxylysine is a component of collagen and has also been reported to become incorporated instead of lysine in the bacterial cell wall (50). The diatom cell is characterised by its silica-based cell wall. Silica-precipitating peptides from the diatom *Cylindrotheca fusiformis* have been shown to contain post-translationally modified lysines (51) that are necessary for their silica-precipitating activity. These lysine residues are ε-dimethylated, ε-trimethylated or δ-hydroxylated or contain a combination of these modifications. It is not clear why lysine residues of diatom Rubiscos should be modified in the same way. It may be that Rubisco has evolved to utilise the enzymes responsible for these modifications, and that these confer some sort of advantage, for instance insensitivity to tryptic proteolysis (44, 52), or stability. Occurrence of these modifications in all of the diatoms used in this study confirm their close relationship. Trimethylation of Lys14 of the large subunit of Rubisco occurs in some plants, e.g those belonging to the *Solanaceae* or *Cucurbitaceae* families, but not all (26, 43). It is possible that the interaction with other proteins, e.g. chaperones or Rubisco activase may be regulated by trimethylation, but at present there is no experimental evidence to substantiate this assumption. Similar to the S-nitroso-group of Cys457, the trimethyl group of Lys346 is located on the surface of the protein, thereby enabling potential contacts with other binding partners.

Conclusion

Oceanic primary production is dominated by phytoplankton and diatoms account for a significant proportion of the NPP (2, 3). Here we present structural and functional data on a large set of diatoms from Arctic cold waters. Our results indicate adaptation of diatom Rubiscos in response to the environment in which they live, including low contents of Rubisco protein, high specificity factors approaching those of the most efficient crop plants coupled with low oxygen sensitivities, and a number of post-translational modifications.

Experimental procedures

Collection of algae in the Arctic east ice area and selection of species for cultivation

To establish diatom monocultures, samples of algae were collected with 20 μm phytoplankton nets during three spring cruises to the Atlantic and Arctic Barents Sea (2004 - 2006). As an initial guideline, when species were selected, it was assumed that the quantitatively most important species recorded during field investigations were the fastest and most successful growers (for compilation of

abundances, see ref. 11). The chosen potential candidates were representatives from the genera *Chaetoceros*, *Thalassiosira*, *Bacterosira* and *Skeletonema*.

Growth rate measurements

Small-scale cultivation experiments were performed aimed at assessing the fastest growers at nutrient replete conditions, i.e. CO_2 and autoclaved natural sea water with added nitrogen, phosphorous, and silicate to f/10 concentrations. These experiments were performed in irradiance and temperature controlled/logged rooms at two irradiances and temperatures (fluorescent daylight tubes, L:D=14:10, scalar irradiance 25 and 125 μmol quanta m^{-2} s^{-1}, 2-3 and 7 °C) using 25 and 1500 ml non toxic Erlenmeyer plastic flasks. Since monitoring growth from a single measure of biomass [e.g. chlorophyll a (Chl*a*) that may vary with species and light level] may not be sufficient to detect "true" increase in overall biomass, several methods were applied to detect the fastest growers. The methods were increase in (i) cell numbers (inverted microscope counting), (ii) organic bound carbon and nitrogen (Carlo Erba Elemental analyzer) and (iii) *in vitro* Chl*a* and pheophytin content (53). We computed growth as doublings day^{-1} from the formula

$$\mu = \mathrm{Log}_2 C2 \;\; \mathrm{Log}_2 C1/D$$

where μ=doublings day^{-1}, $C2$ and $C1$ are cell numbers and D number of days.

In addition, ^{14}C radioactive tracer photosynthesis (carbon assimilation) measurements were performed applying 5 μCurie aqueous sodium-bicarbonate/100 ml^{-1} culture (for method see ref. 54). The scalar irradiance exposure gradients were 330, 172, 102, 53, 13, and 0 μmol quanta m^{-2} s^{-1}. Further we calculated both α-slope photosynthesis curve (mgC mgChl*a* l^{-1} h^{-1} μmol quanta m^{-2} s^{-1}) and P_{max} –max. photosynthesis (mgC mgChl*a* l^{-1} h^{-1}). The above experiments were repeated several times in order to achieve robust data sets for statistical analysis (n=2680). In the end, analysis of the data for the highest growth rates, maximum photosynthesis (P_{max}), and slope photosynthesis (α) for each species, condition and experiment were standardized using the formula

$$x_{\mathrm{new}} = x - \mu/\sigma$$

where x is measured growth rate, μ is population mean doubling and σ is population standard deviation. The standardized results were then pooled for each species, the results ranked, and the following seven diatom species were considered fast growers and were chosen for further investigation: *B. bathyomphala, T. antarctica, T. hyalina, T. nordenskioeldii, T. gravida, C. socialis, and S. marinoi*.

Mass cultivation

The selected species were mass cultured in specially constructed 300 l plexiglass cylinders in temperature and irradiance controlled rooms. Cultivation took place at ~ 4 °C under a 14:10 h L:D regime and at optimal (I_{max}) scalar irradiances determined during the small scale ^{14}C experiments. When the desired culture densities had been reached (150-500 μg Chl*a* l^{-1}) the cultures were harvested onto specially designed 20 μm mesh plankton net devices. The samples were stored at –80 °C prior to further analysis.

Purification of Rubisco for determination of specificity factors

Twenty-five ml of extraction buffer [100 mM Bicine, pH 8.0, 6% PEG 4000, 5 mM DTT, 1 mM each of benzamidine, phenylmethyl sulphonylfluoride (PMSF), ε-amino-n-caproic acid (ε-ACA) and EDTA, 1% (v/v) Tween-80, 0.2 mM EGTA, 0.5% (w/v) PVPP, 0.5% (v/v) protease inhibitor cocktail (Sigma P5955), and 1% (w/v) washed sand] was ground to a frozen powder in liquid nitrogen (N_2). To this was added 20-40 ml of a harvested mass culture from above, then ground to a frozen powder. A further 175 ml of extraction buffer was added, 25-50 ml at a time, with frequent grinding until thawing was complete. On thawing, polysaccharide hydrolases were added (200,000 U lysozyme, 40 U pectinase, 8 U cellulase, all supplied by Sigma UK) and the ice-cold homogenate was sonicated (6-8 μm amplitude) for 30 s followed by a 60 s interval. This was repeated until the total sonication time was 2 min. The extract was clarified by centrifugation (22,000 x g, 20 min, 4 °C), brought to 20% (w/v) PEG 4000 and 20 mM $MgCl_2$, then stirred for 30 min at 4 °C. The resulting protein precipitate was sedimented by centrifugation (22,000 x g, 20 min, 4 °C) and redissolved in 8 ml of ice cold Gradient Buffer [10 mM TRIS, pH 8.0 (HCl), 10 mM $MgCl_2$, 10 mM $NaHCO_3$, 5 mM DTT, 1 mM EDTA, 1 mM KH_2PO_4, 1 mM benzamidine, 1 mM ε-ACA,] using a pre-cooled homogeniser to achieve a lump-free suspension. The suspension was clarified by centrifugation (235,000 x g, 20 min, 4 °C), applied to a previously prepared sucrose gradient (0.3 to 1.2 M sucrose in Gradient

Buffer) at a rate of 4 ml suspension per 35 ml of sucrose gradient, centrifuged for 190 min at 370,000 x g at 4 ^{o}C, fractionated into 1 ml aliquots then snap frozen in liquid N_2. A small sample previously taken from each fraction was assayed for protein content and Rubisco activity (55). Fractions containing the Rubisco activity peak (between fraction 9 and 14 from the bottom) were pooled and passed through PD-10 columns (2 ml of sample per column) pre-equilibrated in Column Buffer (100 mM Bicine, pH 8.1, 10 mM $MgCl_2$, 10 mM $NaHCO_3$, 5 mM DTT, 0.5 mM EDTA, 1 mM ε-ACA, 1 mM benzamidine, 1 mM KH_2PO_4). The resulting protein eluates were combined and passed through 0.45 µm regenerated cellulose filters before sample concentration using Centriplus concentrators (Millipore Amicon, MWCO < 150 kDa). The final volume of the resulting Rubisco was approximately 0.50 ml, which was snap frozen in liquid N_2 prior to short term storage at -80 ^{o}C.

Preparation of Rubisco used for determination of kinetic constants used a simplified procedure, omitting the sonication, sucrose gradient and ultrafiltration treatments, simply consisting of homogenisation, sedimentation, PEG precipitation, clarification of the redissolved protein, passage through PD-10 columns pre-equilibrated with Column Buffer supplemented with 2% (w/v) PEG 4000, followed by freezing in liquid N_2, prior to short term storage at -80 ^{o}C.

Determination of Specificity Factors

Specificity factors for diatom Rubisco were determined by real time data collection based on rates of carboxylation and oxygenation measured by ^{14}C incorporation and an oxygen electrode, respectively. Wheat Rubisco was used as an internal standard, and prior to use, a freeze dried stock of wheat Rubisco was dissolved in CO_2-free 0.1 M Bicine, pH 8.2, containing 20 mM $MgCl_2$. The purified Rubisco samples were then desalted by centrifugation through G25 Sephadex columns previously equilibrated with CO_2-free 0.1 M Bicine, pH 8.2, containing 20 mM $MgCl_2$. Potassium phosphate (400 mM, pH 8.2) was then added to give a final concentration of 4 mM. $NaH^{14}CO_3$ (37 GBq mol^{-1}) was then added to a final concentration of 10 mM and the wheat Rubisco activated by incubation at 37 ^{o}C for 40 min. Diatom Rubisco showed no increase in activity in response to warming but maintained activity for 24 h when kept at 4 ^{o}C (data not shown). Reaction mixtures were prepared in an oxygen electrode (Model DW1, Hansatech, Kings Lynn. UK) by first adding 0.95 ml 100 mM Bicine, pH 8.1, containing 10 mM $MgCl_2$ and 20 µg (50 WA units) of carbonic anhydrase, pre-equilibrated with CO_2-free air at 25ºC, and 0.02 ml of 0.1 M $NaH^{14}CO_3$ 18.5 GBq/mol. A sufficient amount of activated Rubisco was then added in 25 µl to complete the reaction in 5 min. The reaction was started by the addition of 10 µl of 18.5 mM RuBP. RuBP oxygenation was calculated from the oxygen consumption and carboxylation from the amount of ^{14}C incorporated into 3PGA when all the RuBP was consumed (56). A number of reaction mixtures containing pure wheat Rubisco were interspersed with those containing Rubisco from diatoms. In addition, measurements of specificity at 15 and 35 ºC were made. The procedure followed was similar to that at 25 ºC. Mean initial concentrations of O_2 in solution in equilibrium with air were 305, 254 and 227 µM at 15, 25 and 35 ºC respectively, as determined by the integrated Hansatech software. Initial concentrations of CO_2 in solution were calculated from amounts of $NaHCO_3$ added, using pKa values for H_2CO_3 of 6.19, 6.11 and 6.06 at 15, 25 and 35 °C, respectively. The specificity values were normalised to the average value for wheat Rubisco, of 94 [± 4 (SD), n = 4] at 25 ^{o}C. The determinations were repeated 3 – 5 times at each temperature, using material pooled from 2 or 3 biological replicates.

Determination of Catalytic Parameters

These were measured essentially as previously described (57). Carboxylation activity was measured at 8, 16, 24, 36, 68 and 100 µM CO_2 (aq) in equilibrium with a gas phase of N_2 containing 2%, 21%, 56% or 92% (v/v) O_2, at 25 °C. K_M and V_{max} for carboxylation (K_c and V_c, respectively) were calculated at each O_2 concentration using a Michaelis-Menten kinetic model. K_M and V_{max} for oxygenation (K_o and V_o, respectively) were calculated as follows: $K_o = [O_2] / [(K_{M,app} / K_c) – 1]$ and $V_o = (V_c \times K_o) / (K_c \times S_{c/o})$ where K_c is the Michaelis-Menten constant for CO_2 in the absence of O_2, and $K_{M,app}$ is the apparent Michaelis-Menten constant for CO_2 as measured in the reactions equilibrated with 21%, 56%, or 92% O_2. Specific mixtures of N_2 and O_2 were prepared using a gas divider (Signal Group, UK) and concentrations of O_2 in solution were calculated at 100% relative humidity and

standard atmospheric pressure (101.3 kPa). The solubility of O_2 was taken as 257.5 μM. The concentration of CO_2 in solution (in equilibrium with HCO_3^-) was calculated assuming a pKa of 6.11 for the first ionization of carbonic acid, taking into consideration the pH of each buffer solution (measured on the day of assay). Carbonic anhydrase (≥ 50 WA units per 1 ml reaction; Sigma, UK); was present in the reaction solution to maintain equilibrium between $NaHCO_3$ and CO_2. The Rubisco samples used in these assays had all been equilibrated in $NaHCO_3$ and $MgCl_2$ containing buffers during the purification procedures (above), and were found not to require any further activation prior to assay. Control reactions were performed by measuring CO_2 fixation (acid stable ^{14}C) in reaction solutions lacking RuBP or $NaHCO_3$, and following substitution of RuBP for 3PGA, or following total inhibition of Rubisco by prior treatment with an excess of the tight-binding Rubisco inhibitor, 2'-carboxyarabinitol-1,5-bisphosphate (CABP). These controls confirmed that the activity measured (i.e. all acid stable ^{14}C detected) was entirely due to Rubisco.

Radioactive content of ^{14}C-labelled compounds was measured in 0.40 ml aqueous solutions, following the addition of 3.6 ml Ultima Gold Scintillation cocktail (Perkin-Elmer, UK), using a Tri-Carb 2910 TR Liquid Scintillation Analyser (Perkin-Elmer, USA).

Values of Michaelis-Menten constants and maximum velocities were estimated using EnzFitter (Biosoft: Software for Science, UK). Turnover number (k_{cat}; mol product, mol active site^{-1} s^{-1}) was calculated from the corresponding V_{max} values (V_c and V_o; μmol acid-stable ^{14}C mg Rubisco^{-1} min^{-1}) after determination of Rubisco concentration in the samples. This was accomplished using the [^{14}C]CABP binding assay described in ref. 58.

Sequencing of Rubisco genes from marine diatom species

Total genomic DNA was isolated and the DNAs were used as templates in PCR reactions to amplify the *rbc*L/S genes. Internal PCR primers were designed according to marine algal *rbc*L/S sequences that are already deposited in databases. Sequences of the 5' and 3' ends of the genes were amplified using the internal and a set of external primers designed according to genes flanking the *rbc*L/S gene cluster. These genes were found in a preliminary assembly of the genome of the diatom *T. pseudonana* on the web site of the Joint Genome Institute of the US Department of Energy (http://www.jgi.doe.gov/).

Genomic DNA from *C. socialis*, *T. antarctica, T. hyalina*, *T. nordenskioeldii, S. marinoi,* and *B. bathyomphala* was extracted by standard methods. Oligonucleotides were designed to amplify a region of the diatom genome including the Rubisco large and small subunit genes, *rbc*L and *rbc*S. In most cases, a faithful DNA polymerase (PicoMaxx from Statagene) was used to amplify this region and the sequences of *rbc*L and *rbc*S from each species were determined. For each species, each base has been covered by at least 2 sequencing reactions from independently generated PCR products. If there was any difference between the first two sequences, a third independently generated PCR fragment was sequenced. Two species initially gave more than one DNA sequence. In these cases sequencing was repeated with DNA isolated from a new culture.

Isolation and purification of Rubisco for structure determination

To yield crystallisation-grade purity, frozen algae in glycerol (ca. 20 g) were thawed and suspended in 10 ml of extraction buffer [50 mM Bicine, pH 8.0, 10 mM $MgCl_2$, 10 mM $NaHCO_3$, 1 mM EDTA, 5 mM 2-mercaptoethanol, 1 Complete protease-inhibitor tablet (Roche Molecular Biochemicals), 5 μl Benzonase nuclease (Novagen)]. The algal suspension was disrupted in a One-shot cell disrupter (Constant System LTD; www.constantsystem.com). The extract was centrifuged (15 minutes, 20 000 rpm, Sorvall SS34). The supernatant was passed through a 0.45 μm syringe filter and applied to a Superdex 200 column (120 ml) equilibrated with purification buffer. Fractions (2 ml) were collected and analysed by SDS-PAGE. Fractions containing Rubisco were pooled, diluted with an equal volume of 0.1 M NaCl in purification buffer and further purified on a MonoQ ion exchange column (8 ml). The sample was loaded onto the column and equilibrated with low salt (0.1 M NaCl in purification buffer). The protein was eluted with a linear 0.1-0.5 M NaCl gradient in 120 ml of purification buffer. Fractions (2 ml) were collected and analysed by SDS-PAGE. Pooled fractions containing Rubisco yielded 2-5 mg of pure protein from 20 g of algae.

Crystallisation, data collection, structure determination and refinement

Prior to crystallization, the activated enzyme was concentrated to 20 mg ml^{-1} using Vivaspin 6 (Vivascience) and incubated with 0.001 M CABP. Crystals were grown using the hanging-drop vapour diffusion method at 20 °C. The drop contained equal amounts of the protein sample in crystallisation buffer (0.05 M HEPES, pH 7.5, 0.05 M NaCl, 0.01 M $NaHCO_3$, and 0.005 M $MgCl_2$) with 0.001 M CABP added, and a well solution consisting of the crystallisation buffer with 7-13% PEG 4000 as a precipitating agent. The crystals were flash-cooled in liquid N_2 using a mother liquor with 30% ethylene glycol added as a cryo-protectant and maintained at 100 K for data collection. Diffraction data were collected at Max-lab, Lund, Sweden and at the European Synchrotron Radiation Source (ESRF) Grenoble, France (Table 3). The data were processed using DENZO/SCALEPACK (59) and XDS (60). The crystal structures were solved by molecular replacement using the program MOLREP (61). The initial search model consisted of a set of one large and one small subunit of *G. partita* Rubisco (PDB code 1bwv). Using the data for Rubisco from *T. antarctica*, eight solutions corresponding to eight different orientations of the search model in the hexadecamer of the asymmetric unit were found. The RbcL sequence from *T. nordenskioeldii* (O98947) was used for an initial fit to the electron density maps; this crude fit was subsequently improved using results obtained from sequencing of the gene and by inspection of electron density maps. Subsequently, the refined model of *T. antarctica* Rubisco was used as a search model to solve the remaining structures (Table 3). Modifications of the sequence were made as above.

Refinement was performed using REFMAC5 (62), and PHENIX (63). For cross-validation, 5% of the data was excluded from the refinement for R_{free} calculations. Refinement consisted of one round of rigid body refinement using data to 3 Å, followed by refinement using a maximum likelihood target function with non-crystallographic symmetry (NCS) restraints. NCS restraints were released towards the end of refinement of the structures to the highest resolution. TLS refinement (64) was used in the final stages with each subunit as a TLS group. Solvent molecules were added using ARP/wARP (65) and were manually inspected in O (66). Throughout the refinement, the $2mF_o$-DF_c and mF_o-DF_c sigma A weighted maps (67) were inspected and the models were manually adjusted using O (66).

Coordinates and structure factors have been deposited in the Protein Data Bank with accession codes 5MZ2, 5N9Z, 5OYA, and 6FTL.

Acknowledgements: We wish to acknowledge MAX-lab, Lund, Sweden and the European Synchrotron Radiation Facility, Grenoble for providing beam time and data-collection facilities and thank the beam line staff for excellent support. We thank Philip Jewess and Thomas C. Taylor for assistance at the beginning of the project. This work was supported by grants from the European Union, (QLK3-CT-2002-01945) and the Swedish Research Council for Environment, Agricultural Sciences, and Spatial Planning (FORMAS). PJA, PJM and MAJP acknowledge support from the Biotechnology and Biological Sciences Research Council, UK, through the 20:20 Wheat® Institute Strategic Program (BBSRC BB/J/00426X/1).

Conflict of interest: The authors declare that they have no conflicts of interest with the contents of this article.

References

1. Takahashi, T., Sutherland, S. C., Sweeney, C., Poisson, A., Metzl, N., Tilbrook, B., Bates, N., Wanninkhof, R., Feely, R. A., Sabine, C., Olafsson, J., and Nojiri, Y. (2002) Global sea-air CO_2 flux based on climatological surface ocean pCO_2, and seasonal biological and temperature effects. *Deep-Sea Research Part II-Topical Studies in Oceanography* **49**, 1601-1622
2. Field C. B., Behrenfeld M. J., Randerson J. T., and Falkowski P. (1998) Primary production of the biosphere: Integrating terrestrial and oceanic components. *Science* **281**, 237-240
3. Falkowski, P. G. (2012) Ocean Science: The power of plankton. *Nature* **483**, S17-S20
4. Parry, M. A. J., Andralojc, P. J., Mitchell, R. A. C., Madgwick, P. J., and Keys, A. J. (2003) Manipulation of Rubisco: the amount, activity, function and regulation. *J. Exp. Bot*. **54**, 1321 - 1333
5. Andersson, I., and Backlund, A. (2008) Structure and function of Rubisco. *Plant Physiol. Biochem*. **46**, 275-291
6. Carmo-Silva, E., Scales, J. C., Madgwick, P. J., and Parry, M. A. J. (2016) Optimizing Rubisco and its regulation for greater resource use efficiency. *Plant Cell Env*. **38**, 1817-1832
7. Jordan, D. B., and Ogren, W. L. (1981) Species variation in the specificity of ribulose bisphosphate carboxylase/oxygenase. *Nature* **291**, 513-515
8. Delwiche, C. F., and Palmer, J. D. (1996) Rampant horizontal transfer and duplication of rubisco genes in eubacteria and plastids. *Mol. Biol. Evol*. **13**, 873-882
9. Tabita, F. R. (1999) Microbial ribulose-1,5-bisphosphate carboxylase/oxygenase: a different perspective. *Photosynth. Res*. **60**, 1-28
10. Watson, G. M. F., and Tabita, F. R. (1997) Microbial ribulose 1,5-bisphosphate carboxylase/oxygenase: A molecule for phylogenetic and enzymological investigation. *FEMS Microbiol. Lett*. **146**, 13-22
11. Degerlund, M., and Eilertsen, H. C. (2009) Main species characteristics of phytoplankton spring blooms in NE Atlantic and Arctic waters (68-80°N). *Estuaries and Coasts* **33**, 242-269
12. Eppley, R. W. (1972) Temperature and phytoplankton growth in sea. *Fishery Bull.* **70**, 1063-1085
13. Galmés, J., Kapralov, M. V., Andralojc, P. J., Conesa, M. À., Keys, A. J., Parry, M. A. J., and Flexas, J. (2014) Expanding knowledge of the Rubisco kinetics variability in plant species: environmental and evolutionary trends. *Plant Cell Environ.* **37**, 1989-2001
14. Laing, W. A., Ogren, W. L., and Hageman, R. H. (1974) Regulation of soybean net photosynthetic CO_2 fixation by the interaction of CO_2, O_2, and ribulose 1,5-diphosphate carboxylase. *Plant Physiol,* **54**, 678-685
15. Losh, J. L., Young, J. N., and Morel, F. M. M. (2013) Rubisco is a small fraction of total protein in marine phytoplankton. *New Phytologist* **198**, 52-58
16. 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. *In Plant Responses to Air Pollution and Global Change* (eds Omasa, K., Nouchi, I., and De Kok, L. J.). Springer-Verlag Tokyo, pp 157-164
17. Hansen, S., Vollan, V. B., Hough, E., and Andersen, K. (1999) The crystal structure of Rubisco from *Alcaligenes eutrophus* reveals a novel central eight-stranded β-barrel formed by β-strands from four subunits. *J. Mol. Biol*. **288**, 609-621
18. Sugawara, H., Yamamoto, H., Shibata, N., Inoue, T., Okada, S., Miyake, C., Yokota, A., and Kai, Y. (1999) Crystal structure of carboxylase reaction-oriented ribulose-1,5-bisphosphate carboxylase/oxygenase from a thermophilic red alga, *Galdieria partita*, *J. Biol. Chem*. **274**, 15655-15661
19. Stec, B. (2012) Structural mechanism of RuBisCO activation by carbamylation of the active site lysine. *Proc. Natl. Acad. Sci.* **109**, 18785-18790
20. Lorimer, G. H., Badger, M. R., and Andrews, T. J. (1976) The 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
21. Lorimer, G. H., and Miziorko, H. M. (1980) Carbamate formation on the epsilon-amino group of a lysyl residue as the basis for the activation of ribulosebisphosphate carboxylase by CO_2 and Mg^{2+}. *Biochemistry* **19**, 5321-5328

22. Taylor, T. C., Backlund, A., Spreitzer, R. J., Björhall, K., and Andersson, I. (2001) First crystal structure of Rubisco from a green alga - *Chlamydomonas reinhardtii*. *J. Biol. Chem*. **276**, 48159-48164
23. Chen, Z., Chastain, C. J., Al-Abed, S. R., Chollet, R., and Spreitzer, R. J. (1988) Reduced CO_2/O_2 specificity of ribulose-bisphosphate carboxylase/oxygenase in a temperature-sensitive mutant of *Chlamydomonas*. *Proc. Natl. Acad. Sci.* **85**, 4696-4699
24. Karkehabadi, S., Taylor, T. C., Spreitzer, R. J., and Andersson, I. (2005) Altered intersubunit interactions in crystal structures of catalytically compromised ribulosebisphosphate carboxylase/oxygenase. *Biochemistry* **44**, 113-120
25. van Lun, M., van der Spoel, D., and Andersson, I. (2011) Subunit interface dynamics in hexadecameric Rubisco. *J. Mol. Biol.* **411**, 1083-1098
26. Houtz, R. L., Poneleit, L., Jones, S. B., Royer, M., and Stults, J. T., (1992) Posttranslational Modifications in the Amino-terminal region of the large subunit of ribulose- 1,5-bisphosphate carboxylase/oxygenase from several plant species. *Plant Physiol*. **98**, 1170-1174
27. Young, J. N., Heureux, A. M. C., Sharwood, R. E., Rickaby, R. E. M., Morel, F. M. M., and Whitney, S. M. (2016) Large variation in the Rubisco kinetics of diatoms reveals diversity among their carbon-concentration mechanisms. *J. Exp. Bot.* **67**, 3445-3456
28. Galmés, J., Flexas, J., Keys, A. J., Cifre, J., Mitchell, R. A. C., Madgwick, P. J., Haslam, R. P., Medrano, H., and Parry, M. A. J., (2005) Rubisco specificity factor tends to be larger in plant species from drier habitats and in species with persistent leaves. *Plant Cell Environ*. **28** 571-579
29. Giordano, M., Beardall, J., and Raven, J. A. (2005) CO_2 concentrating mechanisms in algae: mechanisms, environmental modulation, and evolution. *Annu. Rev. Plant Biol.* **56**, 99-131
30. Reinfelder, J. R. (2010) Carbon concentrating mechanisms in eukaryotic marine phytoplankton. *Annu. Rev. Mar. Sci.* **3**, 291–315
31. Hopkinson, B. M., Dupont, C. L., and Matsuda, Y. (2016) The physiology and genetics of CO_2 concentrating mechanisms in model diatoms. *Curr. Op. Plant Biol.* **31**, 51-57
32. Badger, M. R., Andrews, T. J., Whitney, S. M., Ludwig, M., Yellowlees, D. C., Leggat, W., and Price, G. D. (1998) The diversity and coevolution of Rubisco, plastids, pyrenoids, and chloroplast-based CO2-concentrating mechanisms in algae. *Can. J. Bot.* **76**, 1052-1071
33. Savir, Y., Noor, E., Milo, R., and Tlusty, T. (2010) Cross-species analyses traces adaption of Rubisco toward optimality in a low-dimensional landscape. *Proc. Natl. Acad. Sci.* **17**, 3475-3480
34. Eilertsen, H. C., and Degerlund, M. (2010) Phytoplankton and light during the northern high-latitude winter. *J. Plankton Res.* **32**, 899-912
35. Hobson, L. A., Morris, W. J., and Guest, K. P. (1985) Varying photoperiod, ribulose 1,5-bisphosphate carboxylase/oxygenase and CO_2 uptake in *Thalassiosira fluviatilis* (Bacillariophyceae). *Plant Physiol.* 79, 833-837
36. Smayda, T. (1958) Biogeographical studies of marine phytoplankton. *Oikos* 9, 158- 191
37. Syvertsen, E. E. (1977) *Thalassiosira rotula* and *T. gravida*: Ecology and morphology. *Nova Hedwigia*, Beih. 54: 99-11
38. Quillfeldt, C. H., von (2001) Identification of easily confused common diatom species in Arctic sproing blooms. *Botanica marina* **44**, 375-389
39. Hasle, G. R., and Syvertsen E. E. (1997) Marine diatoms. *In Identifying marine phytoplankton* (ed. Thomas, C. R.). Academic Press, San Diego, pp 5-385
40. Spreitzer, R. J. (2003) Role of the Rubisco small subunit. *Arch. Biochem. Biophys*. **414**, 141-149
41. 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. *Curr. Microbiol.* **45**, 115-117
42. Ghiron, J. H. L., Amato, A., Montresor, M., and Kooistra, W. H. C. F. (2008) Plastid inheritance in the planktonic raphid pennate diatom *Pseudo-nitzschia delicatissima* (Bacillariophyceae). *Protist* **159**, 91-98
43. Houtz, R. L., and Portis, A. R. (2003) The life of ribulose 1,5-bisphosphate carboxylase/oxygenase-posttranslational facts and mysteries. *Arch. Biochem. Biophys*. **414**, 150-158

44. Houtz, R. L., Magnani, R., Nayak, N. R., and Dirk, L. M. A. (2008) Co- and post-translational modifications in Rubisco: unanswered questions. *J. Exp. Bot.* **59**, 1635-1645
45. Lamport, D. T. A., and Northcote, D. H. (1960) Hydroxyproline in primary cell walls of higher plants. *Nature* **188**, 665–666
46. Cassab, G. I. (1998) Plant cell wall proteins. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **49**, 281-309
47. Moreno, J., Garcia-Murria, M. J., and Marin-Navarro, J. (2008) Redox modulation of Rubisco conformation and activity through its cysteine residues. *J. Exp. Bot.* **59**, 1605-1614
48. Chung, C.-C., Hwang, S.-P. L., and Chang, J. (2008) Nitric oxide as a signaling factor to upregulate the death-specific protein in a marine diatom, *Skeletonema costatum*, during blockage of electron flow in photosynthesis. *Appl. Environ. Microbiol.* **74**, 6521-6527
49. Morin, B., Bubb, W. A., Davies, M. J., Dean, R. T., and Fu, S. (1998) 3-Hydroxylysine, a potential marker for studying radical-induced protein oxidation. *Chem. Res. Toxicol.* 11 (11), 1265–1273.
50. Smith, W. G., Gilboe, D. P., and Henderson, L. M. (1965) Incorporation of hydroxylysine into the cell wall and a cell-wall precursor in *Staphylococcus aureus*. *J. Bacteriol.* **89**, 136-140
51. Kröger, N., Deutzmann, R., and Sumper, M. (2001) Silica-precipitating peptides from diatoms. *J. Biol. Chem*. **276**, 26066-26070.
52. Houtz, R. L., and Mulligan, R. M. (1991) Protection of tryptic-sensitive sites in the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase by catalysis. *Plant Physiol.* **96**, 335-339
53. Holm-Hansen, O., Lorentzen, C. J., Holmes, R. W., and Strickland, J. D. H. (1965) Fluorometric determination of chlorophyll. *Journal Conseil International Pour l'Exploration de la Mer* 30, 3-15
54. Sargent, J. R., Eilertsen, H. C., Falk-Petersen, S., and Taasen, J. P. (1985) Carbon assimilation and lipid production in phytoplankton in northern Norwegian fjords. *Mar. Biol.* **85**, 109-116
55. Parry, M. A. J., Delgado, E., Vadell, J., Keys, A. J., Lawlor, D. W., and Medrano, H. (1993) Water stress and the diurnal activity of ribulose-1,5-bisphosphate carboxylase in field grown *Nicotiana tabacum* genotypes selected for survival at low CO_2 concentrations. *Plant Physiol. Biochem*. **31**, 113-120
56. Parry, M. A. J., Keys, A. J., and Gutteridge, S. (1989) Variation in the specificity factor of C_3 higher plant Rubiscos determined by the total consumption of ribulose-P_2. *J. Exp. Bot.* **40**, 317-320
57. Carmo-Silva, A. E., Keys, A. J., Andralojc, P. J., Powers, S. J., Celeste Arrabaca, M., and Parry, M. A. J. (2010) Rubisco activities, properties, and regulation in three different C4 grasses under drought. *J. Exp. Bot.* **61**, 2355-2366
58. Yokota, A., and Canvin, D. T. (1985) Ribulose bisphosphate carboxylase/oxygenase content determined with [^{14}C]carboxypentitol bisphosphate in plants and algae. *Plant Physiol.* **77,** 735-9
59. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307-326
60. Kabsch, W. (2010) XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125-132
61. Vagin, A., and Teplyakov, A. (1997) MOLREP: an Automated Program for Molecular Replacement. *J. Appl. Crystallogr.* **30**, 1022-1025
62. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr*. *D Biol*. *Crystallogr*. **53**, 240-55
63. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213-221
64. Winn, M. D., Isupov, M. N., and Murshudov, G. N. (2001) Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr. D Biol. Crystallogr.* **57**, 122-33

65. Lamzin, V. S., Perrakis, A., and Wilson, K. S. (2001) The ARP/WARP suite for automated construction and refinement of protein models. In *International Tables for Crystallography. Vol F, Crystallography of Biological Macromolecules* (Rossman, M. G., and Arnold, E. , eds), pp. 720-722, Kluwer Academic Publishers, Dordrecht
66. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110-9
67. Pannu, N. S., and Read, R. J. (1996) Improved structure refinement through maximum likelihood, *Acta Crystallogr. A* **52***,* 659-668

FOOTNOTES
This work was supported by the European Union (No. QLK3-CT-2002-01945) and the Swedish Research Council for Environment, Agricultural Sciences, and Spatial Planning (FORMAS).

Abbreviations*:* Rubisco, ribulose-1,5-bisphosphate carboxylase/oxygenase; RuBP, D-ribulose-1,5-bisphosphate; 3PGA, 3-phospho-D-glycerate; Pg, petagram; 2PG, 2-phosphoglycolate; NPP, net primary production; L, large subunit; S, small subunit; *rbc*L, Rubisco large subunit gene; *rbc*S, Rubisco small subunit gene; CABP, 2'-carboxyarabinitol-1,5-bisphosphate; r.m.s.d., root mean square deviation.

Table 1. **Overall standardized maximum growth rates of seven diatom species chosen for further analysis**. The results are means of repeated experiments with species specific n>48. See Experimental procedures for explanation of standardisation method.

Species	Max. growth at 2-3 °C	Max. growth at 7 °C
Chaetoceros socialis	-0.19	0.09
Thalassiosira nordenskioeldii	0.02*	0.11
Thalassiosira hyalina	-0.24	-0.32*
Thalassiosira antarctica	0.54*	0.12
Thalassiosira gravida	-0.25	-0.08
Skeletonema marinoi	-0.22	-0.11*
Bacterosira bathyomphala	0.13*	0.12*

*Species that had standardized growth rates statistically significantly different from the other species at the same temperature (Anova $p<0.05$).

Table 2. Relative specificities and kinetic constants of Rubisco from marine diatoms.

Species	Temperature (°C)	Normalised Specificity factors			Kinetic Constants n=6			
		mean	sd	n	V_cmax (μmol min^{-1} mg Rubisco^{-1})	V_omax (μmol min^{-1} mg Rubisco^{-1})	K_MC (μM)	K_MO (mM)
Thalassiosira hyalina	15	106	4	4	3.5±0.1	1.2 ±0.2	50 ±3	1.64±0.20
	25	99	3	4				
	35	87	1	3				
Bacterosira bathyomphala	15	94	10	3	3.9 ±0.1	0.8 ±0.1	81 ±4	1.30 ±0.18
	25	87	4	3				
	35	76	2	3				
Skeletonema marinoi	15	96	9	4	4.0 ±0.1	1.6 ±0.2	48 ±2	1.81 ±0.19
	25	96	7	4				
	35	84	5	3				
Thalassiosira nordenskioeldii	25	82	2	5	4.0 ±0.1	0.6 ±0.1	122 ±4	1.42 ±0.53
Thalassiosira antarctica	25	90[a]	3.2[a]	5[a]	3.2 ±0.2	0.7 ±0.3	93 ±10	1.74 ±0.75
Triticum aestivum	15	113	8	4	2.5 ±0.1[b]	0.8 ±0.03[b]	10.9±0.9[b]	0.34 ±0.03[b]
	25	100	3	4				
	35	90	3	3				

[a]From ref. 16.
[b]From ref. 57.

Table 3. Data collection and refinement statistics

Values in parentheses are for the outer resolution shell.

Species	*Thalassiosira antarctica* var borealis	*Thalassiosira hyalina*	*Skeletonema marinoi*	*Chaetoceros socialis*
PDB id	5MZ2	5N9Z	6FTL	5OYA
Data collection				
X-ray source	ESRF ID14:2	ESRF ID14:3	Lund X711	ESRF ID29
Wavelength (Å)	0.933	0.931	1.087	0.969
Resolution (Å)	1.9	1.9	2.6	1.8
Space group	$P2_1$	$P2_1$	$P4_12_12$	$C2$
Unit cell a, b, c (Å) β (°)	118.3, 220.1, 124.4, β=118.4	118.0, 220.0, 124.3, β=118.3	a=b=111.0, c=396.4	118.2, 219.1, 220.2, β=90.2
No. reflections	3 067 613	2 653 566	1 535 835	1 195289
No. unique reflections	437 345	432 731	75 003	495 177
Completeness (%)	99.9 (99.7)	99.0 (98.5)	96.9 (75.8)	96.1 (94.9)
R_{merge}[a]	0.117 (0.388)	0.066 (0.251)	0.178 (0.956)	0.092 (0.668)
Refinement				
Residues in model	L: 3/4-483/484 S: 1-139	L: 4-484 S: 1-139	L: 3-484 S: 1-139	L: 15/16-483 S: 1-139
No water molecules	3 659	3 107	354	2946
No ethylene glycol molecules	20	15	18	24
R_{cryst}[b]	0.143 (0.166)	0.151 (0.179)	0.171 (0.296)	0.167 (0.317)
R_{free}[c]	0.176 (0.218)	0.182 (0.223)	0.238 (0.364)	0.192 (0.331)
R.m.s.d. from ideal values bond lengths (Å) bond angles (°)	 0.006 0.922	 0.007 0.973	 0.011 1.57	 0.012 1.36
Ramachandran analysis Favoured (%) Allowed (%) Outliers (%)	 98 2 0	 98 2 0	 96 4 0	 97 3 0

[a] $R_{merge} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $\langle I(hkl) \rangle$ is the average intensity of symmetry-equivalent reflections.

[b] $R_{work} = \sum_{hkl} ||F_o| - |F_c|| / \sum_{hkl} |F_o|$ where F_o and F_c are the observed and calculated structure factor amplitudes, respectively.

[c] R_{free} calculated from a randomly chosen 5% of all unique reflections.

Table 4. **Posttranslational modifications in diatom Rubiscos**.
HYP, 4-hydroxyproline; CSO, S-hydroxycysteine; LYO, 4-hydroxylysine; LOH, 3,4-dihydroxylysine; 8RE, 3,4-dihydroxylysine; HLU, beta-hydroxyleucine; HL2, beta-hydroxyleucine; KCX, lysine-NZ-carboxylic acid; M3L, N-trimethyllysine; SNC, S-nitrosocysteine. LOH and 8RE are stereoisomers at the Cß position of dihydroxylysine. HLU and HL2 are stereoisomers at the Cß position of beta-hydroxyleucine.

Species	Pro48	Cys109	Lys150	Pro155	Leu174	Lys198	Lys205	Lys346	Lys457
T. antarctica	HYP	CSO	LYO	HYP	HLU	LYO	KCX	M3L	n.m.[a]
T. hyalina	HYP	CSO	8RE	HYP	HLU	LYO	KCX	M3L	n.m.[a]
S. marinoi	-	n.m.[a]	LOH	HYP	HLU	LYO	KCX	M3L	n.m.[a]
C. socialis	HYP	CSO	LOH	HYP	HL2	-	KCX	M3L	SNC

[a]Not modelled because of low occupancy.

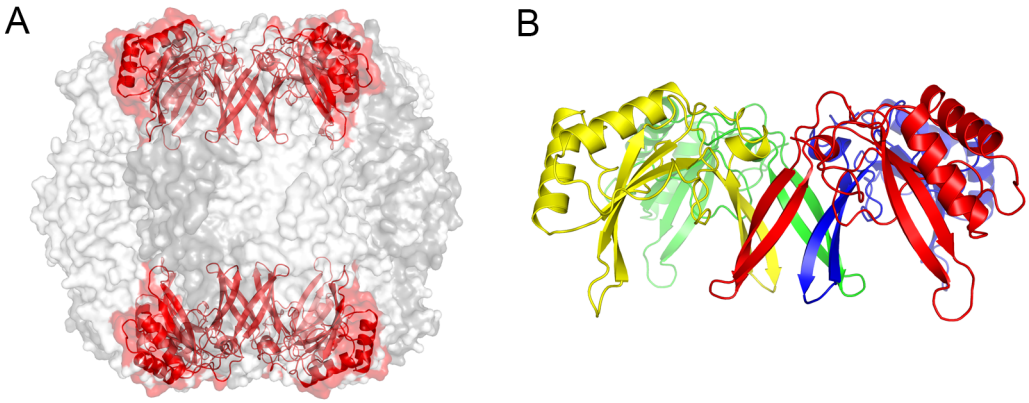

Figure 1. Structure of form I D diatom Rubisco. a) Overall structure, showing large subunits in two shades of grey and small subunits in red and b) structure of the capping cluster of S subunits, coloured yellow, red, blue and green, respectively. The C-terminal strands from four different S subunits form a ß-barrel that lines the central solvent channel.

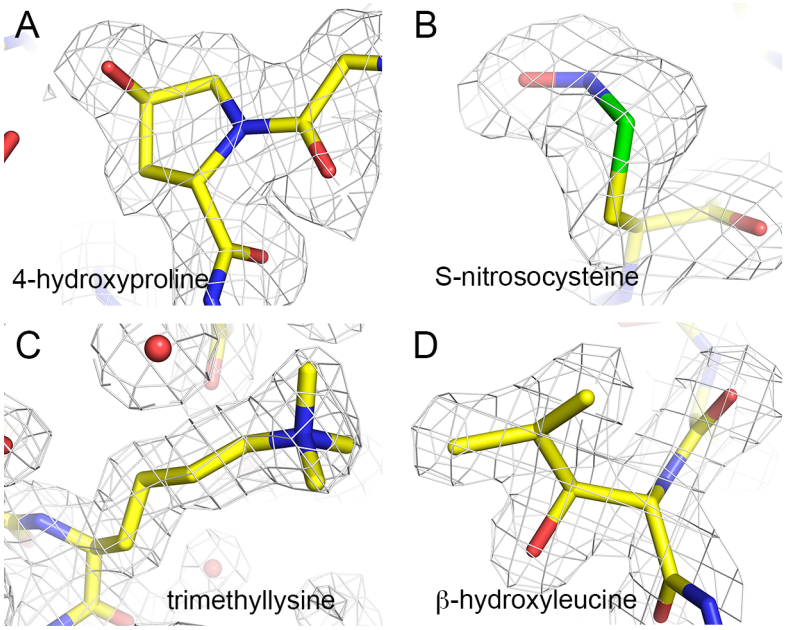

Figure 2. Representative electron density for post-translationally modified residues: a) 4-hydroxyproline, b) S-nitrosocysteine, c) ε-*N*,*N*,*N*-trimethyllysine, and d) ß-hydroxyleucine.

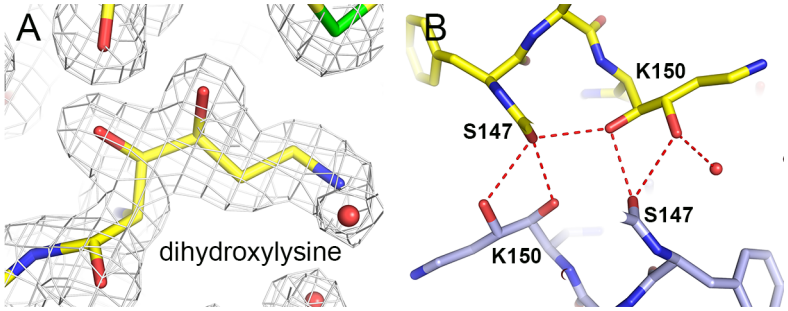

Figure 3. Posttranslationally modified Lys150. a) Electron density for γ,δ-dihydroxylysine, and b) contacts of the modified residue at the L_2 dimer interface. One L subunit in yellow and the second L subunit in blue.