



**Molecular
Omics**

**Crosslinking mass spectrometry unveils novel interactions
and structural distinctions in the model green alga
*Chlamydomonas reinhardtii***

Journal:	<i>Molecular Omics</i>
Manuscript ID	MO-RES-06-2021-000197.R1
Article Type:	Research Article
Date Submitted by the Author:	22-Aug-2021
Complete List of Authors:	Smythers, Amanda; University of North Carolina Chapel Hill, Chemistry Iannetta, Anthony; University of North Carolina Chapel Hill, Chemistry Hicks, Leslie; University of North Carolina Chapel Hill, Chemistry

SCHOLARONE™
Manuscripts

1 **Cover page**

2

3 **Title**

4

5 Crosslinking mass spectrometry unveils novel interactions and structural distinctions in the model green
6 alga *Chlamydomonas reinhardtii*

7

8 **Author names**

9

10 Amanda L. Smythers^{1#}, Anthony A. Iannetta^{1#}, and Leslie M. Hicks^{1*}

11

12 **Affiliations**

13

14 ¹Department of Chemistry, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
15 # denotes equal contributions

16

17 **Correspondence**

18

19 Dr. Leslie M. Hicks, Department of Chemistry, University of North Carolina at Chapel Hill,
20 Kenan Laboratories, 125 South Road, CB#3290, Chapel Hill, NC 27599-3290, United States

21

22 E-mail: lmhicks@unc.edu

23

24 Phone/Fax: (919) 843-6903 / (919) 962-2388

25

26 **Keywords**

27

28 Crosslinking mass spectrometry, interactome, photosynthesis, *Chlamydomonas*, LC-MS/MS

29

30 **Running title**

31

32 XL-MS unveils protein interactome of *Chlamydomonas reinhardtii*

1 Abstract

2 Interactomics is an emerging field that seeks to identify both transient and complex-bound
3 protein interactions that are essential for metabolic functions. Crosslinking mass spectrometry
4 (XL-MS) has enabled untargeted global analysis of these protein networks, permitting largescale
5 simultaneous analysis of protein structure and interactions. Increased commercial availability of
6 highly specific, cell permeable crosslinkers has propelled the study of these critical interactions
7 forward, with the development of MS-cleavable crosslinkers further increasing confidence in
8 identifications. Herein, the global interactome of the unicellular alga *Chlamydomonas reinhardtii*
9 was analyzed via XL-MS by implementing the MS-cleavable disuccinimidyl sulfoxide (DSSO)
10 crosslinker and enriching for crosslinks using strong cation exchange chromatography. Gentle
11 lysis via repeated freeze-thaw cycles facilitated *in vitro* analysis of 157 protein-protein crosslinks
12 (interlinks) and 612 peptides linked to peptides of the same protein (intralinks) at 1% FDR
13 throughout the *C. reinhardtii* proteome. The interlinks confirmed known protein relationships
14 across the cytosol and chloroplast, including coverage on 42% and 38% of the small and large
15 ribosomal subunits, respectively. Of the 157 identified interlinks, 92% represent the first empirical
16 evidence of interaction observed in *C. reinhardtii*. Several of these crosslinks point to novel
17 associations between proteins, including the identification of a previously uncharacterized Mg-
18 chelatase associated protein (Cre11.g477733.t1.2) bound to five distinct lysines on Mg-chelatase
19 (Cre06.g306300.t1.2). Additionally, the observed intralinks facilitated characterization of novel
20 protein structures across the *C. reinhardtii* proteome. Together, these data establish a framework
21 of protein-protein interactions that can be further explored to facilitate understanding of the
22 dynamic protein landscape in *C. reinhardtii*.

1 Introduction

2 Crosslinking mass spectrometry (XL-MS) is a powerful approach that couples
3 biochemistry with molecular and structural biology through simultaneous analysis of protein-
4 protein interactions (PPIs), conformations, and structure.(1-5) While alternative methods to
5 investigate PPIs rely on genetic transformation (*e.g.*, the yeast two-hybrid assay) and/or the use of
6 highly specific antibodies (*e.g.*, affinity-purification mass spectrometry), XL-MS is limited only
7 by the reactivity and specificity of the chemical crosslinker, enabling global unbiased delineation
8 of protein networks.(6-10) Further, while other structural approaches often require homogenous
9 protein samples, large sample abundance, and/or crystallization, XL-MS can capture dynamic
10 protein conformations in their native environment, without isolation or purification.(11, 12)

11 Chemical crosslinking involves the covalent linkage of two protein residues that are in
12 close proximity. Crosslinker reagents vary in reactive groups (*e.g.*, NHS ester for amine reactivity,
13 maleimide for cysteine reactivity) and spacer arm length (*e.g.*, 10.3 Å for DSSO). The maximum
14 distance between crosslinked residues is limited by the spacer arm length, but it is accepted that
15 this constraint can be exceeded due to dynamic protein conformational changes (*e.g.*, 30 Å limit
16 for DSSO despite a 10 Å spacer arm).(13) Crosslinking can be coupled with conventional bottom-
17 up workflows (*i.e.*, proteolytic cleavage) for the identification of crosslinked peptides. These
18 detected crosslinks produce three-dimensional information for proteins; crosslinks between two
19 tryptic peptides from two different proteins (interlinks) inform PPIs while crosslinks between two
20 tryptic peptides within the same protein (intralinks) enhance structural knowledge for that protein.

21 XL-MS is an increasingly powerful approach to examine PPIs, yet the requisite data
22 processing is immensely challenging, as linking two peptides increases the proteome search space
23 by n^2 and creates challenging MS/MS fragment spectra that can be difficult to deconvolute. This
24 has been partly addressed via a portfolio of search algorithms and bioinformatics platforms, yet
25 still poses several ongoing challenges including: 1) increased missed cleavages due to crosslinks
26 blocking potential cleavage sites, 2) altered ionization, 3) more complex MS/MS fragmentation,
27 and 4) low abundance of crosslinked peptides compared to linear peptides.(14-16) MS-cleavable
28 crosslinker development has greatly increased the applicability of XL-MS to systems-wide studies,
29 as observed mass shifts from the short or long end of the cleaved crosslinker (*i.e.*, crosslink reporter
30 ions), more efficient peptide fragmentation, and MS/MS spectra of increased quality simplify
31 crosslink identification in database searching and lend confidence to site assignment.(17)

32 *Chlamydomonas reinhardtii* is a unicellular green alga and one of the most widely studied
33 models for photosynthesis, attributed in part to its rapid growth rate, large singular chloroplast,
34 and well annotated genome.(18, 19) *C. reinhardtii* is a beneficial model organism for studying
35 fundamental biochemical processes, including autophagy, signal transduction, and nitrogen flux,
36 among others.(20-25) Despite its long-term interest among plant scientists and cell biologists, its
37 complex interactome has yet to be thoroughly investigated. In the STRING database, there are
38 1,278 experimentally confirmed (medium confidence score >400) unique protein-protein
39 interactions in *C. reinhardtii*, compared to 31,283 in *Arabidopsis thaliana* and 51,599 in
40 *Saccharomyces cerevisiae*.(26) Herein, the amine-reactive, MS-cleavable crosslinker
41 disuccinimidyl sulfoxide (DSSO) was leveraged to analyze the *C. reinhardtii* interactome via XL-

1 MS (Figure 1). Extraction via freeze-thaw enabled global detection of interacting proteins
2 throughout the cell, including 612 intralinks and 157 interlinks at 1% FDR.
3

4 **Experimental**

5 *Cell Growth*

6
7 Wild-type *Chlamydomonas reinhardtii* strain CC-2895 6145c mt- was purchased from the
8 Chlamydomonas Resource Center (St. Paul, MN, USA) and batch cultures were maintained
9 photoheterotrophically on Tris-acetate-phosphate (TAP) agar plates. *C. reinhardtii* was inoculated
10 into 100 mL of TAP medium using a 1 mL inoculum in a foil-covered 250 mL Erlenmeyer flask
11 and grown photoheterotrophically.(27) Cultures were maintained at 22 °C on an Innova 2000
12 platform shaker (New Brunswick Scientific, Enfield, CT, USA) at 120 rpm under constant 100
13 $\mu\text{mol m}^{-2} \text{s}^{-1}$ illumination. Cells were grown to mid-log phase (OD_{750} 0.4-0.5), harvested by
14 centrifuging for 5 min at 3220 g and discarding the supernatant, and flash-frozen using liquid
15 nitrogen. Cell pellets were stored at -80 °C until use.
16

17 *Protein Extraction*

18
19 Three frozen cell pellets (~0.6 g each) were thawed on ice before combining and adding 10 mL of
20 20 mM HEPES buffer, pH 7.8 containing 30 mM sodium chloride, 1.5 mM magnesium chloride,
21 0.5 mM dithiothreitol, and 1x cOmplete EDTA-free protease inhibitor cocktail (Roche, Basal,
22 Switzerland). Soft lysis via five rounds of freeze-thaw at -80 °C for 60 min was performed to
23 ensure the extraction of intact protein interactions. Cellular debris was cleared by centrifugation at
24 3,200 g for 20 min at 4 °C before protein concentrations were estimated using the CB-X assay (G-
25 Biosciences, St. Louis, MO, USA) according to the manufacturer's protocol.
26

27 *Protein Crosslinking*

28
29 A 50 mM stock solution of disuccinimidyl sulfoxide (DSSO, Thermo Fisher Scientific, Waltham,
30 MA, USA) in DMSO was prepared and added to 3 mg of protein lysate at a working concentration
31 of 2 mM. Samples were incubated with end-over-end rotation for 30 min at RT, before quenching
32 the crosslinking reaction with 20 mM Tris, pH 8 for 30 min at RT.
33

34 *Protein Digestion*

35
36 Crosslinked proteins (3 mg) were precipitated with 5x cold acetone and centrifuged at 3200 g and
37 4 °C for 20 min. The supernatant was removed and proteins were resuspended in 200 μL of 100
38 mM Tris-HCl, pH 8 containing 8 M urea. Samples were reduced and alkylated simultaneously
39 using 10 mM tris(2-carboxyethyl)phosphine and 40 mM chloroacetamide for 1 h at 37 °C. Samples

1 were diluted four-fold to 2 M urea with additional 100 mM Tris-HCl, pH 8 before digestion was
2 performed with Trypsin Gold (Promega, Madison, WI, USA) at 37 °C for 16 h using a
3 protease:protein ratio of 1:50 (w/w).

4

5 *Reversed-Phase Solid-Phase Extraction*

6

7 Samples were desalted with 100 mg/1.0 mL Sep-Pak C18 cartridges (Waters, Milford, MA, USA)
8 using a 24-position vacuum manifold (Phenomenex, Torrance, CA, USA) at a flow rate of 1 drop/s.
9 Resin was first pre-eluted using 1 mL of 80% acetonitrile/0.1% trifluoroacetic acid (TFA) before
10 equilibration with 2 mL of 0.1% TFA. Samples were acidified to pH < 3 using 10% TFA, loaded
11 onto the cartridges in two passes, and then washed using 2 mL of 0.1% TFA. Peptides were eluted
12 using 1 mL of 80% acetonitrile/0.1% TFA and concentrated by vacuum centrifugation. Peptides
13 were resuspended in 200 µL of 10 mM potassium phosphate monobasic, 20% acetonitrile, pH 2.7.

14

15 *Strong Cation Exchange Fractionation*

16

17 The crosslinked peptides were fractionated with strong cation exchange (SCX) chromatography
18 using a Shimadzu Prominence HPLC equipped with a UV-vis detector (220 nm) (Shimadzu,
19 Kyoto, Japan) following the method described in Makepeace *et al.*(28) Mobile phase A consisted
20 of 10 mM potassium phosphate monobasic, 20% acetonitrile, pH 2.7, mobile phase B was 10 mM
21 potassium phosphate monobasic, 250 mM potassium chloride, 20% acetonitrile, pH 2.7, and
22 mobile phase C was 10 mM potassium phosphate monobasic, 600 mM potassium chloride, 20%
23 acetonitrile, pH 2.7. Peptides were fractionated on a PolySulfoethyl A column (100 mm x 4.6 mm,
24 3 µm particles; PolyLC) using a linear gradient of increasing mobile phases B and C at a flow rate
25 of 0.5 mL/min. After 10 min of 100% mobile phase A, mobile phase B increased from 0% to 15%
26 in 9.3 min, where it was held for 8.7 min before ramping to 30% in 8 min, where it was held for
27 11 min and then ramping to 100% in 5 min, where it was held for 5 min. After, mobile phase C
28 increased from 0% to 100% in 5 min before returning to 100% mobile phase A in 5 min and re-
29 equilibrating for 25 min. After 10 min into the gradient, fractions were collected every 1 min.
30 These were desalted using reversed-phase solid phase extraction as described above, concentrated
31 under vacuum centrifugation, and resuspended in 15 µL of 0.1% TFA.

32

33 *LC-MS/MS Analysis*

34

35 Fractions were analyzed using an Acquity UPLC M-Class System (Waters) coupled to a Q
36 Exactive HF-X mass spectrometer (Thermo Fisher Scientific). Mobile phase A consisted of water
37 with 0.1% formic acid (Thermo Fisher Scientific) and mobile phase B was acetonitrile with 0.1%
38 formic acid. Injections (4 µL) were made to a Symmetry C18 trap column (100 Å, 5µm, 180µm x
39 20 mm; Waters) with a flow rate of 5 µL/min for 3 min using 99% A and 1% B. Peptides were
40 then separated on a HSS T3 C18 column (100 Å, 1.8µm, 75µm x 250 mm; Waters) using a linear

1 gradient of increasing mobile phase B at a flow rate of 300 nL/min. Mobile phase B increased
2 from 5% to 35% in 90 min before ramping to 85% in 5 min, where it was held for 10 min before
3 returning to 5% in 2 min and re-equilibrating for 13 min. The mass spectrometer was operated in
4 positive polarity and the Nanospray Flex source had spray voltage floating at 2.1 kV, capillary
5 temperature at 320 °C, and funnel RF level at 40. MS survey scans were collected with a scan
6 range of 350 – 2000 m/z at a resolving power of 120,000 and an AGC target of 3×10^6 with a
7 maximum injection time of 50 ms. A top 20 data-dependent acquisition was used where HCD
8 fragmentation of precursor ions having +2 to +7 charge state was performed using a normalized
9 collision energy setting of 28. MS/MS scans were performed at a resolving power of 30,000 and
10 an AGC target of 1×10^5 with a maximum injection time of 100 ms. Dynamic exclusion for
11 precursor m/z was set to a 10 s window.

12

13 *Data Analysis*

14

15 Acquired spectral files (*.raw) from the 61 fractions were analyzed using Proteome Discoverer
16 v.2.5 with the incorporated XLinkX nodes and searched against the Joint Genome Institute's v.5.6
17 database (https://phytozome-next.jgi.doe.gov/info/Creinhardtii_v5_6, 19,523 entries, accessed
18 02/2020) appended with the NCBI chloroplast and mitochondrial databases (chloroplastic-NCBI:
19 BK000554; mitochondrial-NCBI: NC_001638.1; 77 entries, accessed 02/2020) and sequences for
20 common laboratory contaminants (<https://www.thegpm.org/cRAP/>, 116 entries, accessed
21 02/2020). For crosslinked peptides, target-decoy searches of MS/MS data used a trypsin protease
22 specificity with the possibility of two missed cleavages, peptide/fragment mass tolerances of 15
23 ppm/0.02 Da, fixed modification of cysteine carbamidomethylation, and variable modifications of
24 N-terminus acetylation and methionine oxidation. Identified crosslinks are reported at 1% FDR,
25 controlled at the CSM level using XlinkX and its target-decoy database searching strategy.
26 Detailed settings for Proteome Discoverer nodes are found in Supplemental Information 1.

27

28 Initial analysis of protein interlinks was achieved using the STRING database as well as KEGG
29 mapper.(29-32) In the STRING database, active interaction sources were limited to experiments
30 and co-expression with a minimum required interaction score of 0.400. Following STRING and
31 KEGG annotation, the proteins were manually annotated for location and function using a
32 combination of UniProt, PANTHER functional analysis, and the PlaPPiSite
33 (<http://zzdlab.com/plappisite/index.php>).(33-35)

34

35 To generate protein-protein interaction models, an established protocol(36) was applied, where
36 protein structure modeling was achieved through I-TASSER(37) and protein complex docking was
37 performed using HADDOCK.(38) These models were visualized and Euclidean distances of
38 mapped crosslinks were measured using ChimeraX.(39-42) Mapping of detected interlinks was
39 performed using xiNET.(43) XlinkDB 3.0 was used to automatically calculate Euclidean distances
40 of intralinks that were mapped onto known structures or homology models generated by the
41 Integrative Modeling Platform.(44-46) Most intralinks that were mapped were visualized using

1 NGL Viewer.(47) For eukaryotic translation initiation factor 1 alpha, proteins were visualized by
2 mapping to the homology model in SWISS-MODEL, downloading the pdb file, visualizing in
3 Chimera, and manually annotating for detected crosslinks.(39, 48) Detected crosslinks from this
4 study were made private on the XLinkDB 3.0 database
5 (<http://xlinkdb.gs.washington.edu/xlinkdb/>). These data can be accessed by using the filename
6 “Crcrosslinking_Lesliehicks” in any of the sections labeled “Network Name”.

7 8 *Data Availability*

9
10 The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium
11 via the PRIDE partner repository and can be accessed with the dataset identifier PXD026433.(49)

12
13 Username: reviewer_pxd026433@ebi.ac.uk
14 Password: UVg4z1uC

15 16 **Results and Discussion**

17 Crosslinked Proteome Coverage

18
19 Although XL-MS is a powerful tool for the global analysis of PPIs, its success is highly
20 dependent on the permeability and reactivity of the crosslinker. In preliminary trials with DSSO,
21 the crosslinker did not permeate the Chlamydomonas cell wall effectively and did not enable
22 global analysis of the interactome (data not shown). Therefore, gentle, detergent-free lysis via
23 freeze-thaw was used to release proteins from the cell in near-native conformations. Intact protein-
24 protein interactions were crosslinked *in vitro* with DSSO and crosslinked peptides were enriched
25 using strong cation exchange (SCX) fractionation prior to LC-MS/MS analysis (Figure 1C). SCX
26 leverages the highly positively charged crosslinked peptides to separate them from the less charged
27 linear, non-crosslinked peptides. This is essential as non-crosslinked peptides greatly outnumber
28 and suppress the ionization of the low abundant crosslinked peptides. SCX fractionation
29 simultaneously enriches for crosslinked peptides and decreases sample complexity, thus increasing
30 the depth of coverage for crosslinked peptides.

31 The analysis of 61 SCX fractions resulted in the identification of 56,595 crosslink reporter
32 ion peaks, yielding 1,705 crosslink-spectrum matches (CSMs, representative CSM can be found
33 in Figure S1) at 1% FDR grouped into 769 unique crosslinks (Table S1). Also, from these SCX
34 fractions a total of 116,086 non-crosslinked peptides were identified (Table S2), corresponding to
35 the identification of 7,482 proteins (≥ 2 unique peptides, master proteins, 1% FDR, Table S3). The
36 full suite of identified crosslinks is depicted as a circos plot in Figure 2A, while the identified *C.*
37 *reinhardtii* interactome is displayed in Figure 2B (interactive versions of both can be found [here](#)).
38 The majority of the CSMs are contained in later SCX fractions (49-65), supporting the use of SCX
39 as an enrichment technique for more positively charged crosslinked peptides (Figure 2C). Among
40 the 769 detected crosslinks, 157 are between two tryptic peptides derived from two different

1 proteins (interlinks) and 612 are between two tryptic peptides within the same protein (intralinks)
2 (Figure 2D). A majority of the detected proteins with at least one identified crosslink contain one
3 crosslink (63%), while 13% contain two crosslinks and 24% contain three or more (Figure 2E).
4 Additionally, of the 157 identified interlinks, only 12 are currently reported in *C. reinhardtii* in the
5 plant PPI database (Table S4); therefore 92% of the crosslinks observed herein represent the first
6 empirical evidence of interactions between the proteins.(35)

7 8 Intralinks

9
10 Overlaying detected intralinks onto known protein structures and comparing the Euclidean
11 distance between the crosslinked residues to the DSSO maximum crosslinking distance of 30 Å
12 can be used to evaluate the intralink dataset.(13, 50-52) Few proteins with identified intralinks
13 have existing structures for *C. reinhardtii* in the Protein Data Bank (PDB), so the Integrative
14 Modeling Platform in XLinkDB 3.0 was used for homology modeling for the other proteins.(44,
15 46) In this platform, a structural homology model is identified by multiple sequence alignment and
16 used to predict protein structure. Out of all detected intralinks, 44% were mapped onto known
17 structures or homologous proteins to obtain structural information (Table S5). Of these, 96%
18 featured residues within the theoretical distance of DSSO (Figure 3A); this is not surprising as the
19 majority of the observed proteins are highly conserved across taxons and thus have similar
20 structural features in distinct species.

21 The crosslinking data for Photosystem II Oxygen Evolution Enhancer protein 3
22 (Cre08.g372450.t1.2) confirms the structure established via cryogenic electron microscopy.(53)
23 Of the 10 intralinks observed across this protein, none exceeded the 30 Å cutoff, with an average
24 predicted link distance of 12 Å. However, there are relatively few proteins from *C. reinhardtii*
25 with structural data in PDB; therefore, many of the observed intralinks were mapped to proteins
26 in other organisms. For example, 6 intralinks derived from phosphoglycerate kinase
27 (Cre11.g467770.t1.1) were mapped onto phosphoglycerate kinase from *Thermotoga maritima*
28 (PDB ID: 1VPE), a thermophilic bacterium, and the predicted Euclidean distances were compared
29 with the maximum distance enabled by DSSO. All of the crosslinks on *T. maritima*
30 phosphoglycerate kinase contained residues < 30 Å apart, thus indicating that the DSSO
31 crosslinked peptides from *C. reinhardtii* identified in this study conform to the structural
32 predictions of the known protein structure from *T. maritima* (Figure 3B).

33 The lack of structural data for *C. reinhardtii* proteins creates challenges when using
34 crosslinking to confirm tertiary structure. For example, eukaryotic translation elongation factor 1
35 alpha (Cre12.g498600.t1.2) had the second highest abundance of intralinks with 18 total, 14 of
36 which were mapped within the distance restraints for DSSO (Figure 4). However, there are no
37 structures from *C. reinhardtii* in PDB and the crosslinks from *C. reinhardtii* could not be mapped
38 to a singular homologous protein; rather, the *C. reinhardtii* sequence was mapped to proteins with
39 high sequence similarity in *Homo sapiens*, *Oryctolagus cuniculus*, *Pyrococcus horikoshii*, and
40 *Aeropyrum pernix* (Figure 4A). Additionally, three intralinks (K358-K293, K387-K413, K293-

1 K230) were not able to be mapped to proteins with known structures, likely due to differences in
2 the primary sequences in the protein models when compared to that of *C. reinhardtii* (Figure 4B).
3 Therefore, while the average predicted crosslink distance for eukaryotic translation elongation
4 factor 1 alpha was 17 Å (Figure 4C), well within the distance restraint of DSSO, complementary
5 experiments would need to be conducted to fully understand how unique the protein structure is in
6 *C. reinhardtii*.

7 Intralinks that include residues spaced above 30 Å could indicate differences from the
8 homologous structure and/or false positive identifications. The high percentage of intralinks that
9 are within the 30 Å distance limit provides high confidence in the dataset, therefore suggesting
10 that those outside the limit are potential distinctions from the homologous protein structures. One
11 notable example is Ribosomal protein S5/Elongation Factor G/III/V family protein
12 (Cre12.g516200.t1.2), in which 10 intralinks were observed. Of those intralinks, seven had
13 predicted distances less than 30 Å. However, one intralink was not mapped to a known structure
14 and two were outside the DSSO distance restraints; K498-K429 had a predicted distance of 40 Å
15 while K498-K484 had a predicted distance of 36 Å, thus suggesting deviations from the mapped
16 protein structures. Like eukaryotic translation elongation factor 1 alpha, this protein required
17 several organisms' known protein structures to map the crosslinks, with both intralinks that
18 surpassed the DSSO restraint being mapped to proteins from *H. sapiens*. However, further work
19 will need to be conducted to determine the extent to which the structure in *C. reinhardtii* differs
20 from that of other known models.

21 Experimentally determined intralinks can be used as constraints to guide computational
22 modeling for novel protein structures. This approach was implemented to predict the structure of
23 Elongation Factor 3 and ABC transporter (Cre04.g222700.t1.2) by incorporating 11 identified
24 intralinks (5 were not mapped onto a known structure or homology model) to computational model
25 its structure using the iterative threading assembly refinement (I-TASSER) server.(37) From this,
26 the top five structure models were output from I-TASSER (Figure S2). The top template for these
27 models derived from Elongation Factor 3A in *Saccharomyces cerevisiae* (PDB: 2IWH),
28 suggesting that Cre04.g222700.t1.2 folds similarly to this protein. To test the agreement between
29 these structures and the crosslink data, the detected intralinks were mapped onto the models and
30 the distances between the crosslinked residues were calculated. In each case, all crosslinks had
31 measured distances within the maximum restraint for DSSO, providing confidence in the
32 modeling. These data demonstrate the value of these identified intralinks in generating novel
33 protein structures.

34

35 Interlinks

36

37 *Cytosolic interlinks*

38 Overall, 76 cytosolic interlinked protein pairs were identified, including 83 unique proteins
39 (Figure 2B, Table S1). Of these interlinked protein pairs, 63 included at least one protein from the
40 cytosolic ribosome, reflecting the known high abundance of this large protein complex.

1 KeggMapper revealed that 42% of the small ribosomal subunit and 38% of the large ribosomal
2 subunit were identified in this study (Figure 5).(30) This includes 34 proteins with at least one
3 detected interlink and 29 proteins with at least one detected intralink, thus showing remarkable
4 coverage of the ribosome from both a protein-protein relationship perspective as well as a
5 structural perspective, without any attempt at ribosomal enrichment. There were 84 intralinks
6 identified on cytosolic ribosomal subunits, of which 67 were successfully mapped to a homology
7 model and analyzed for Euclidian distance. Of these measurements, only three intralinks were
8 determined to be greater than the maximum distance allowed by DSSO. Crosslinks between K498-
9 K429 and K498-K484 from Ribosomal protein S5 (Cre12.g516200.t1.2) had distances of 40 and
10 36 Å, respectively, and the crosslink between K209-K183 from Cytosolic 80S ribosomal protein
11 L8 (Cre12.g535851.t1.1) had a distance of 38 Å. This likely represents small changes in the
12 structure of the subunits compared to the homology models, though further analysis is needed to
13 fully characterize any deviations.

14 The extensive coverage of the ribosome indicates that crosslinking could be used to profile
15 ribosomal changes in *C. reinhardtii*, which would be advantageous when combined with
16 quantitative analysis for mapping changes to post-translational modifications and/or determining
17 structural dynamics under stress, particularly as these dynamics are known to regulate substrate
18 interaction and biological activity of protein translation.(54-57) Previous work leveraged MS to
19 identify the order of subunit assembly along the ribosomal stalk proteins in *E. coli*.(58) but cannot
20 reveal structural changes occurring of well documented large-scale movements following substrate
21 binding.(59, 60) In contrast, solution-state NMR has been leveraged to understand the structure
22 and motion of the ribosome in *E. coli*, but lacks the ability to identify the subunits responsible for
23 this flexibility.(61) Recent work paired heat treatment with MS and bioinformatic analysis to
24 identify 17 intrinsically disordered proteins within the cytosolic ribosome structure in *C.*
25 *reinhardtii*; however, their contribution(s) to the dynamic ribosome structure and overall flexibility
26 is currently unknown.(62) XL-MS enables simultaneous analysis of both the identification of
27 subunits as well as their structural proximity, which could prove critical in understanding
28 ribosomal regulation in *C. reinhardtii*. Further enrichment of ribosomes paired with quantitative
29 XL-MS may unveil how these proteins' flexibility contribute to ribosomal regulation.

30 Although the majority of detected interlinks related to the ribosome, several other identified
31 interactions point to the existence of novel regulatory points in essential metabolic processes.
32 Additionally, their interlink partners can help confirm or dismantle previous localization
33 predictions. For example, glutaredoxin 6 (Cre01.g047800.t1.1) (Table S1, row 705), which
34 enzymatically deglutathionylates cysteine residues, has been predicted to be localized to the
35 chloroplast despite being encoded in the nuclear genome.(63-65) However, here it was observed
36 in an interlink with a cytosolic nucleotide-binding protein (Cre16.g685200.t1.1), thus
37 contradicting the prediction that glutaredoxin 6 is localized to the chloroplast. Although validation
38 is required, localization of glutaredoxin 6 in the cytosol could indicate a more substantial substrate
39 list than previously predicted. Understanding the spectrum of glutaredoxin substrates is essential

1 to understanding the impact of intracellular oxidative signaling, and how this can be used as a
2 regulatory mechanism in diverse cell processes.

3 Another intriguing result is the crosslink identified between starch phosphorylase
4 (Cre07.g3336950.t1.1) and a previously unannotated protein (Cre02.g142206.t1.1) (Table S1, row
5 742) unique to *C. reinhardtii*. Conducting a protein BLAST search of this unidentified protein
6 against the proteome of *Arabidopsis thaliana* did not reveal any sequence similarity, but did result
7 in sequence similarities of >50% to uncharacterized proteins in five green algae: *Volvox carteri f.*
8 *nagariensis*, *Gonium pectoral*, *Haematococcus lacustris*, *Scenedesmus sp. NREL 46B-D3*, and
9 *Polytomella parva* (Figure 6). Additionally, searching the primary sequence against the NCBI
10 conserved domain database (CDD) distinguished amino acids 66 – 173 as an oxidoreductase
11 containing a GGXGXXG cofactor binding motif; this protein domain is present in many proteins
12 related to the metabolism of steroids, cofactors, carbohydrates, lipids, aromatic compounds, and
13 amino acids, as well as function in redox sensing, though it is not present in the five aligned,
14 uncharacterized proteins with >50% sequence similarity.(66) Together, this data suggests this
15 unannotated protein may be a unique regulatory protein along the starch biosynthesis pathway in
16 *C. reinhardtii*, though its function needs to be further analyzed.

17 Similarly, we identified a novel connection between inositol hexakisphosphate kinase
18 (Cre01.g052650.t1.1) and sulfatase-domain containing protein (Cre01.g012126.t1.2) (Table S1,
19 row 768). Hexakisphosphate kinase catalyzes the conversion of hexakisphosphate (InsP₆) to
20 disphosphoinositol pentakisphosphate (InsP₇). It is also an essential regulatory component of the
21 target of rapamycin pathway in *C. reinhardtii*, through which it engages in crosstalk with the
22 phosphorylation network to modulate carbon metabolism and photosynthesis.(67) In *C.*
23 *reinhardtii*, hexakisphosphate kinase is impacted by phosphorous deprivation, thus establishing
24 the enzyme as a regulatory point in intracellular nutritional sensing. However, its interaction with
25 a sulfatase-domain containing protein could suggest that it is also involved in sulfur recycling
26 and/or sensing— not surprising due to TOR's longstanding association with many forms of
27 nutritional deprivation.(22, 24, 68, 69) Further experiments should explore the role of InsPs during
28 sulfur deficiency to better understand this regulatory mechanism.

29

30 *Chloroplast interlinks*

31 As the chloroplast typically comprises >40% of the cell's volume, an abundance of
32 crosslinked peptides were identified from chloroplast proteins.(18) In total, we identified 43
33 unique proteins involved in 35 interlinked peptide pairs, which were analyzed for known
34 relationships using the STRING database (Figure 7).(29) Several interlinks aligned with well-
35 established protein relationships, such as between glyceraldehyde-3-phosphate dehydrogenase
36 (GAPDH, Cre01.g010900.t1.2) and phosphoribulokinase (PRK, Cre12.g554800.t1.2) (Table S1,
37 row 738), which together form a complex that controls substrate availability for RuBisCO.(70)
38 Previous work has used fluorescence correlation spectroscopy to distinguish the existence of the
39 PRK-GAPDH complex, here confirmed via XL-MS.(71) GAPDH has also been shown to complex
40 with CP12, an intrinsically disordered protein, to successfully recruit and complex with PRK, an

1 essential regulatory step in the operation of the Calvin-Benson cycle.(72, 73) However, CP12 was
2 not identified among crosslinked peptides in this study. This could indicate that CP12's lysines are
3 solvent inaccessible, or more likely that the protein was too low in abundance to identify in this
4 dataset. The latter is suggested by recent X-ray diffraction work in *Arabidopsis thaliana* that
5 analyzed the GAPDH-CP12-PRK complex where the observed lysines were solvent accessible
6 therein. This underscores the need for further targeted analysis via XL-MS and/or high resolution
7 microscopy to delineate the complex structure in *C. reinhardtii*.(74)

8 XL-MS also revealed a known relationship between photosystem II repair protein
9 (Cre10.g430150.t1.2, REP27) and Chloroplast ribosomal protein L11 (Cre10.g423650.t1.2,
10 RPL11), detecting multiple crosslinks between REP27 at position K351 and RPL11 at positions
11 K126, K129, and K146 (Table S1, rows 473, 474, and 598). REP27 is a tetratricopeptide repeat
12 protein that is encoded by the nuclear genome but localized to the chloroplast and is essential for
13 the selective removal and replacement of photodamaged D1 proteins of photosystem II.(75)
14 Although the structure of REP27 has not been elucidated via biochemical techniques, identification
15 of two transmembrane domains facilitated the generation of a folding model of REP27, which
16 predicts a single loop localized to the lumen due to transmembrane helices while both the N-
17 terminus and the C-terminus are localized to the chloroplast stroma.(76) The highly charged C-
18 terminus is essential for mRNA translation initiation and assembly of D1, with *rep27* knockouts
19 possessing few intact D1 proteins when analyzed using western blotting. The data herein adds
20 additional evidence to this mechanism, with the first empirical evidence of a direct linkage between
21 REP27 and RPL11 that likely represents a docking point for the ribosome to the photosystem II
22 repair complex during D1 repair and insertion.(77)

23 While known protein interactions lend confidence to the acquired dataset, unique
24 crosslinks presenting previously undescribed interactions are of particular interest for further
25 examination. For example, a novel interaction was observed between GAPDH and an FKBP-type
26 cis-trans isomerase (Cre10.g466850.t1.1) that facilitates protein disulfide bonds and is essential
27 for the regulation/proliferation of oxidative signaling (Table S1, row 613).(78, 79) GAPDH is
28 highly redox regulated and we have previously observed reversible oxidation on C190 that remains
29 unchanged following nitrogen deprivation.(24) While GAPDH has several documented regulatory
30 redox partners, including thioredoxins,(80-82) it has not previously been connected to FKBP-type
31 cis-trans isomerase. This interaction could therefore facilitate the formation of regulatory disulfide
32 bonds on GAPDH.

33 XL-MS also revealed Mg-chelatase subunit 1 (Cre06.g306300.t1.2), the first-committed
34 enzyme of chlorophyll biosynthesis, to be a hub for protein interactions within the chloroplast.
35 Mg-chelatase was shown to interact with flagellar associated protein 165 (Cre03.g211521.t1.1), a
36 transcription initiation factor component of the TAF4 family (Cre02.g095800.t1.2), DNA
37 polymerase theta (Cre08.g384390.t1.1), an IgA-specific serine endopeptidase
38 (Cre09.g408350.t1.2), and two previously unannotated (predicted) *C. reinhardtii* proteins
39 (Cre11.g477733.t1.2, Cre06.g306300.t1.2). Multiple interlinks were identified between one of the
40 predicted proteins (Cre11.g477733.t1.2) at position K19 and Mg-chelatase at positions K70, K116,
41 K157, K167, K198, K207, and K308 (Figure 8A, 8B). This small 14 kDa protein, referred herein

1 as Mg-chelatase associated protein (MCAP), does not share sequence identity with any proteins in
2 *Arabidopsis thaliana*, suggesting that it may be a novel protein for the regulation of chlorophyll
3 biosynthesis in green algae. While it shares 62.7% sequence overlap with a protein from the green
4 alga *Gonium pectoral* and 59% overlap with a protein from the green alga *Chlamydomonas*
5 *incerta*, these proteins also lack annotations.

6 To generate a protein-protein interaction model for Mg-chelatase and MCAP, an
7 established protocol was applied, which employs I-TASSER for protein structure modeling and
8 High Ambiguity Driven protein-protein Docking (HADDOCK) for protein complex docking.(36-
9 38) The 9 detected intralinks for Mg-chelatase were incorporated in its I-TASSER structure
10 modelling (Figure S3) and the top structures for Mg-chelatase and MCAP were exported for
11 interaction docking. All detected intralinks from Mg-chelatase were mapped onto its top structure
12 and had measured distances within the maximum restraint for DSSO, thereby indicating that model
13 refinement was unnecessary. In contrast, modeling MCAP was more complicated; all the modeled
14 structures had low C-scores (-5 to -4, scored on a range of -5-2, where 2 is best), where the top
15 templates included a putative uncharacterized metacyclic invariant surface protein from
16 *Trypanosoma brucei* (PDB: 5VTL) and *Mycobacterium tuberculosis* quinolinate phosphoribosyl
17 transferase (PDB: 1QPN). This is likely the result of low sequence similarity between MCAP and
18 proteins with known structures. The identified interlinks between Mg-chelatase and MCAP were
19 incorporated into the preliminary complex modeling, top structures for each cluster were exported,
20 and detected interlinks were mapped onto these structures to determine the alignment between the
21 models and the experimental crosslink data (Figure S4). The best overlap between one of the
22 models and the identified interlinks resulted in 3 of 7 crosslinks (K167-K19, K207-K19, K308-
23 K19) having measured distances within the maximum restraint for DSSO. This poor overlap
24 between the detected interlinks and the protein complex docking could be attributed to the poor
25 protein structure modeling obtained for MCAP. A refined complex model was created by repeating
26 the protein complex docking and only incorporating the compatible interlinks (Figure 8C, S5), and
27 the same 3 of 7 crosslinks had measured distances within the maximum restraint for DSSO.

28 MCAP is likely low in abundance as it has not been detected in our previous global
29 proteomics work in *C. reinhardtii*, whereas Mg-chelatase was detected in all studies.(24, 83, 84)
30 This suggests that MCAP and Mg-chelatase may be present in different stoichiometric ratios.
31 Chlorophyll synthesis is essential for ensuring efficient light capture and energy transfer, and
32 investigating the function of MCAP should be a priority in order to validate the reported interaction
33 results, and, if this interaction is confirmed, understand how this biosynthetic pathway differs in
34 comparison to other phototrophs.

1 **Conflict of interest**

2 The authors have no conflict of interest related to the work described in this manuscript.

3

4 **Author Contributions**

5 Investigation, A.L.S.; bioinformatics, A.A.I.; formal analysis, A.L.S. and A.A.I.; writing, A.L.S.,
6 A.A.I., and L.M.H.; funding acquisition, L.M.H. All authors have read and agreed to the published
7 version of the manuscript.

8

9 **Acknowledgements**

10 This research was supported by a National Science Foundation CAREER award (MCB-1552522)
11 awarded to L.M.H. and an NSF Major Research Instrumentation award (CHE-1726291) for the
12 HFX-MS.

13

14 **Figure captions**

15 **Figure 1.** General workflow for *C. reinhardtii* interactomics. **(A)** The MS-cleavable crosslinker
16 DSSO was used in this study. **(B)** DSSO generates reporter ions that enable delineation between
17 crosslinks. These reporter ions are essential for minimizing false discovery rate in identifications.
18 **(C)** Native protein structures and interactions were preserved and extracted using a gentle lysis via
19 freeze-thaw. Proteins were crosslinked with DSSO and proteolyzed using trypsin. Crosslinked
20 peptides were enriched and fractionated to increase depth of coverage for crosslink identifications.
21 After LC-MS/MS analysis, the XLinkX nodes in Proteome Discoverer were used to recognize
22 crosslinks by characteristic crosslink reporter ions (α_S , α_L , β_S , and β_L) in the MS2 spectra and
23 identify crosslink-spectrum matches with database searching using the fragment ions from the
24 MS2 spectra.

25 **Figure 2.** Description of identified *C. reinhardtii* crosslinked peptides following DSSO
26 crosslinking of protein extracts. **(A)** Circos plot of all identified crosslinks. Inner, green curves
27 represent interlinks between two different proteins, outer, orange curves represent interlinks within
28 a homoligomer, and outer, purple curves represent intralinks within the same protein. **(B)** Protein
29 interactome map featuring identified interlinks. Proteins are grouped by subcellular localization.
30 **(C)** Histogram of the number of identified CSMs across each SCX fraction. **(D)** Types of detected
31 crosslinks. **(E)** Distribution of identified unique proteins by number of detected crosslinks.

32 **Figure 3.** Intralinks were mapped onto known protein structures or homology models using the
33 Integrative Modeling Platform in XLinkDB 3.0. **(A)** Histogram showing Euclidean distances
34 between residues of detected intralinks. The red, dashed line represents the DSSO maximum
35 crosslinking distance of 30 Å. Of intralinks that were able to be mapped, 96% featured residues
36 within the theoretical distance of DSSO. **(B)** The 6 intralinks (in green) identified from *C.*
37 *reinhardtii* phosphoglycerate kinase that mapped onto the structure of phosphoglycerate kinase
38 from *T. maritima* (PDB ID: 1VPE). The Euclidean distances between the residues in each
39 intralinks (in red) are less than 30 Å, the maximum crosslinking distance for DSSO.

40

1 **Figure 4.** Intralinks of eukaryotic translation elongation factor 1 alpha (Cre12.g498600.t1.2) were
2 mapped onto homology models using the Integrative Modeling Platform in XLinkDB 3.0 to
3 measure theoretical distances between identified intralinks. **(A)** Visualization of structure maps
4 used for distance determination, conducted via SWISS-MODEL. Lysine residues are shown in
5 pink and crosslinks are shown via green lines. **(B)** Sequence alignment (conducted via Clustal
6 Omega) of the five proteins used to determine the distances between crosslinked residues in
7 eukaryotic translation elongation factor 1 alpha. Green highlight indicates residues that were
8 successfully mapped to homology models, while pink, blue, and purple highlights indicate
9 crosslinks (shown in C) that were not mapped to homology models. **(C)** Table of the intralinks
10 observed on eukaryotic translation elongation factor 1 alpha, the species and protein accession the
11 primary sequence was mapped to, and the distance between bound lysines according to each
12 model.

13 **Figure 5.** Visualization of all crosslinked subunits of the cytosolic ribosome. Intralinks are
14 visualized in pink while interlinks within the large or small ribosomal subunits are visualized in
15 blue. The orange interlinks represent interactions between proteins found in opposite subunits.

16 **Figure 6.** Sequence alignment between Cre02.g142206.t1.1, an uncharacterized protein observed
17 to be crosslinked to starch phosphorylase (Cre07.g336950.t1.1), and the five proteins with
18 sequence similarity >50%. Organisms include *Volvox carteri f. nagariensis*, *Gonium pectoral*,
19 *Haematococcus lacustris*, *Scenedesmus sp. NREL 46B-D3*, and *Polytomella parva*. The
20 oxidoreductase domain is highlighted in pink.

21 **Figure 7.** Protein interactome map featuring identified interlinks localized to the chloroplast.
22 Proteins are color coded by molecular function and clustered biological processes are shaded and
23 labeled. Purple lines denote observed interlinks in this study while grey lines indicate empirical
24 evidence of interactions as compiled by the STRING database, where increasing line thickness
25 indicates increased confidence in interaction.

26 **Figure 8.** Crosslinking unveiled a novel complex between Mg-
27 chelatase (Cre06.g306300.t1.2) and uncharacterized protein Cre11.g477733.t1.2, herein referred
28 to as Mg-chelatase associated protein, or MCAP. **(A)** Sequence of MCAP, which is
29 uncharacterized on UniProt. The crosslinked residue is shown in pink. **(B)** Table of observed
30 crosslinks between Mg-chelatase and MCAP. Crosslinked residues are shown in brackets.
31 Euclidean distance refers to that interlink mapped onto the refined protein complex shown in
32 Figure 8C. **(C)** I-TASSER and HADDOCK were used to generate a protein-protein interaction
33 model between Mg-chelatase (teal) and MCAP (orange). The structure with the most mapped
34 intralinks within the maximum restraint for DSSO (in green) from the refined complex modeling
35 is shown. The boxes labeled with roman numerals show the crosslinked lysine residues from the
36 perspective of the arrows on the model complex structure.

37

38 References

39 1. Leitner A, Faini M, Stengel F, Aebersold R. Crosslinking and mass spectrometry: an
40 integrated technology to understand the structure and function of molecular machines. Trends in
41 biochemical sciences. 2016;41(1):20-32.

- 1 2. Leitner A, Walzthoeni T, Kahraman A, Herzog F, Rinner O, Beck M, et al. Probing
2 native protein structures by chemical cross-linking, mass spectrometry, and bioinformatics.
3 *Molecular & Cellular Proteomics*. 2010;9(8):1634-49.
- 4 3. Sinz A. The advancement of chemical cross-linking and mass spectrometry for structural
5 proteomics: from single proteins to protein interaction networks. *Expert review of proteomics*.
6 2014;11(6):733-43.
- 7 4. Iacobucci C, Götze M, Sinz A. Cross-linking/mass spectrometry to get a closer view on
8 protein interaction networks. *Current opinion in biotechnology*. 2020;63:48-53.
- 9 5. Liu F, Heck AJ. Interrogating the architecture of protein assemblies and protein
10 interaction networks by cross-linking mass spectrometry. *Current opinion in structural biology*.
11 2015;35:100-8.
- 12 6. Smythers AL, Hicks LM. Mapping the plant proteome: tools for surveying coordinating
13 pathways. *Emerging Topics in Life Sciences*. 2021.
- 14 7. Kaur U, Meng H, Lui F, Ma R, Ogburn RN, Johnson JHR, et al. Proteome-Wide
15 Structural Biology: An Emerging Field for the Structural Analysis of Proteins on the Proteomic
16 Scale. *J Proteome Res*. 2018;17(11):3614-27.
- 17 8. Paiano A, Margiotta A, De Luca M, Bucci C. Yeast Two - Hybrid Assay to Identify
18 Interacting Proteins. *Curr Protoc Protein Sci*. 2019;95(1):e70.
- 19 9. Peden EA, Boehm M, Mulder DW, Davis R, Old WM, King PW, et al. Identification of
20 Global Ferredoxin Interaction Networks in *Chlamydomonas reinhardtii*. *Journal of Biological*
21 *Chemistry*. 2013;288(49):35192-209.
- 22 10. Mackinder LC. The *Chlamydomonas* CO₂ - concentrating mechanism and its potential
23 for engineering photosynthesis in plants. *New Phytologist*. 2018;217(1):54-61.
- 24 11. Schneider M, Belsom A, Rappsilber J. Protein tertiary structure by crosslinking/mass
25 spectrometry. *Trends in biochemical sciences*. 2018;43(3):157-69.
- 26 12. O'Reilly FJ, Rappsilber J. Cross-linking mass spectrometry: methods and applications in
27 structural, molecular and systems biology. *Nature structural & molecular biology*.
28 2018;25(11):1000-8.
- 29 13. Merkley ED, Rysavy S, Kahraman A, Hafen RP, Daggett V, Adkins JN. Distance
30 restraints from crosslinking mass spectrometry: Mining a molecular dynamics simulation
31 database to evaluate lysine–lysine distances. *Protein Science*. 2014;23(6):747-59.
- 32 14. Matzinger M, Mechtler K. Cleavable Cross-Linkers and Mass Spectrometry for the
33 Ultimate Task of Profiling Protein–Protein Interaction Networks in Vivo. *Journal of Proteome*
34 *Research*. 2021;20(1):78-93.
- 35 15. Mendes ML, Fischer L, Chen ZA, Barbon M, O'reilly FJ, Giese SH, et al. An integrated
36 workflow for crosslinking mass spectrometry. *Molecular systems biology*. 2019;15(9):e8994.
- 37 16. Götze M, Iacobucci C, Ihling CH, Sinz A. A simple cross-linking/mass spectrometry
38 workflow for studying system-wide protein interactions. *Analytical chemistry*.
39 2019;91(15):10236-44.
- 40 17. Sinz A. Divide and conquer: cleavable cross-linkers to study protein conformation and
41 protein–protein interactions. *Anal Bioanal Chem*. 2017;409(1):33-44.
- 42 18. Harris EH. *The Chlamydomonas Sourcebook: Introduction to Chlamydomonas and Its*
43 *Laboratory Use: Volume 1: Academic press; 2009.*
- 44 19. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, et al. Phytozome: a
45 comparative platform for green plant genomics. *Nucleic acids research*. 2012;40(D1):D1178-
46 D86.

- 1 20. Pérez-Pérez ME, Couso I, Crespo JL. The TOR Signaling Network in the Model
2 Unicellular Green Alga *Chlamydomonas reinhardtii*. *Biomolecules*. 2017;7(3):54.
- 3 21. Pérez-Pérez ME, Couso I, Heredia-Martínez LG, Crespo JL. Monitoring autophagy in the
4 model green microalga *Chlamydomonas reinhardtii*. *Cells*. 2017;6(4):36.
- 5 22. Park JJ, Wang H, Gargouri M, Deshpande RR, Skepper JN, Holguin FO, et al. The
6 response of *Chlamydomonas reinhardtii* to nitrogen deprivation: a systems biology analysis. *The*
7 *Plant journal : for cell and molecular biology*. 2015;81(4):611-24.
- 8 23. Juergens MT, Deshpande RR, Lucker BF, Park J-J, Wang H, Gargouri M, et al. The
9 regulation of photosynthetic structure and function during nitrogen deprivation in
10 *Chlamydomonas reinhardtii*. *Plant physiology*. 2015;167(2):558-73.
- 11 24. Smythers AL, McConnell EW, Lewis HC, Mubarek SN, Hicks LM. Photosynthetic
12 Metabolism and Nitrogen Reshuffling Are Regulated by Reversible Cysteine Thiol Oxidation
13 Following Nitrogen Deprivation in *Chlamydomonas*. *Plants (Basel, Switzerland)*. 2020;9(6).
- 14 25. Takeuchi T, Benning C. Nitrogen-dependent coordination of cell cycle, quiescence and
15 TAG accumulation in *Chlamydomonas*. *Biotechnology for Biofuels*. 2019;12(1):292.
- 16 26. Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, et al. The STRING
17 database in 2021: customizable protein–protein networks, and functional characterization of
18 user-uploaded gene/measurement sets. *Nucleic acids research*. 2021;49(D1):D605-D12.
- 19 27. Ramanan R, Kim B-H, Cho D-H, Ko S-R, Oh H-M, Kim H-S. Lipid droplet synthesis is
20 limited by acetate availability in starchless mutant of *Chlamydomonas reinhardtii*. *FEBS letters*.
21 2013;587(4):370-7.
- 22 28. Makepeace KAT, Mohammed Y, Rudashevskaya EL, Petrotchenko EV, Vögtle FN,
23 Meisinger C, et al. Improving Identification of In-organello Protein-Protein Interactions Using an
24 Affinity-enrichable, Isotopically Coded, and Mass Spectrometry-cleavable Chemical
25 Crosslinker. *Mol Cell Proteomics*. 2020;19(4):624-39.
- 26 29. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING
27 v11: protein-protein association networks with increased coverage, supporting functional
28 discovery in genome-wide experimental datasets. *Nucleic acids research*. 2019;47(D1):D607-
29 d13.
- 30 30. Kanehisa M, Sato Y. KEGG Mapper for inferring cellular functions from protein
31 sequences. *Protein Sci*. 2020;29(1):28-35.
- 32 31. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and
33 interpretation of large-scale molecular data sets. *Nucleic acids research*. 2012;40(D1):D109-D14.
- 34 32. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic acids*
35 *research*. 2000;28(1):27-30.
- 36 33. Consortium TU. UniProt: a worldwide hub of protein knowledge. *Nucleic acids research*.
37 2018;47(D1):D506-D15.
- 38 34. Mi H, Muruganujan A, Huang X, Ebert D, Mills C, Guo X, et al. Protocol Update for
39 large-scale genome and gene function analysis with the PANTHER classification system
40 (v.14.0). *Nature protocols*. 2019;14(3):703-21.
- 41 35. Yang X, Yang S, Qi H, Wang T, Li H, Zhang Z. PlaPPISite: a comprehensive resource
42 for plant protein-protein interaction sites. *BMC Plant Biology*. 2020;20(1):61.
- 43 36. Orbán-Németh Z, Beveridge R, Hollenstein DM, Rampler E, Stranzl T, Hudecz O, et al.
44 Structural prediction of protein models using distance restraints derived from cross-linking mass
45 spectrometry data. *Nature protocols*. 2018;13(3):478-94.

- 1 37. Yang J, Zhang Y. I-TASSER server: new development for protein structure and function
2 predictions. *Nucleic acids research*. 2015;43(W1):W174-W81.
- 3 38. van Zundert GCP, Rodrigues JPGLM, Trellet M, Schmitz C, Kastiris PL, Karaca E, et
4 al. The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular
5 Complexes. *Journal of molecular biology*. 2016;428(4):720-5.
- 6 39. Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, et al. UCSF
7 ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Science*.
8 2021;30(1):70-82.
- 9 40. Goddard TD, Huang CC, Meng EC, Pettersen EF, Couch GS, Morris JH, et al. UCSF
10 ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Science*.
11 2018;27(1):14-25.
- 12 41. Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ. PatchDock and SymmDock:
13 servers for rigid and symmetric docking. *Nucleic acids research*. 2005;33(suppl_2):W363-W7.
- 14 42. Wang J, Youkharibache P, Zhang D, Lanczycki CJ, Geer RC, Madej T, et al. iCn3D, a
15 web-based 3D viewer for sharing 1D/2D/3D representations of biomolecular structures.
16 *Bioinformatics*. 2019;36(1):131-5.
- 17 43. Combe CW, Fischer L, Rappsilber J. xiNET: Cross-link Network Maps With Residue
18 Resolution *. *Molecular & Cellular Proteomics*. 2015;14(4):1137-47.
- 19 44. Schweppe DK, Zheng C, Chavez JD, Navare AT, Wu X, Eng JK, et al. XLinkDB 2.0:
20 integrated, large-scale structural analysis of protein crosslinking data. *Bioinformatics*.
21 2016;32(17):2716-8.
- 22 45. Zheng C, Weisbrod CR, Chavez JD, Eng JK, Sharma V, Wu X, et al. XLink-DB:
23 Database and Software Tools for Storing and Visualizing Protein Interaction Topology Data.
24 *Journal of Proteome Research*. 2013;12(4):1989-95.
- 25 46. Russel D, Lasker K, Webb B, Velázquez-Muriel J, Tjioe E, Schneidman-Duhovny D, et
26 al. Putting the Pieces Together: Integrative Modeling Platform Software for Structure
27 Determination of Macromolecular Assemblies. *PLoS biology*. 2012;10(1):e1001244.
- 28 47. Rose AS, Hildebrand PW. NGL Viewer: a web application for molecular visualization.
29 *Nucleic acids research*. 2015;43(W1):W576-W9.
- 30 48. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-
31 MODEL: homology modelling of protein structures and complexes. *Nucleic acids research*.
32 2018;46(W1):W296-W303.
- 33 49. Perez-Riverol Y, Csordas A, Bai J, Bernal-Llinares M, Hewapathirana S, Kundu DJ, et
34 al. The PRIDE database and related tools and resources in 2019: improving support for
35 quantification data. *Nucleic acids research*. 2019;47(D1):D442-D50.
- 36 50. Banks CAS, Zhang Y, Miah S, Hao Y, Adams MK, Wen Z, et al. Integrative Modeling of
37 a Sin3/HDAC Complex Sub-structure. *Cell Reports*. 2020;31(2).
- 38 51. Bartolec TK, Smith D-L, Pang CNI, Xu YD, Hamey JJ, Wilkins MR. Cross-linking Mass
39 Spectrometry Analysis of the Yeast Nucleus Reveals Extensive Protein-Protein Interactions Not
40 Detected by Systematic Two-Hybrid or Affinity Purification-Mass Spectrometry. *Analytical
41 Chemistry*. 2020;92(2):1874-82.
- 42 52. Klykov O, Steigenberger B, Pektaş S, Fasci D, Heck AJR, Scheltema RA. Efficient and
43 robust proteome-wide approaches for cross-linking mass spectrometry. *Nature protocols*.
44 2018;13(12):2964-90.
- 45 53. Sheng X, Watanabe A, Li A, Kim E, Song C, Murata K, et al. Structural insight into light
46 harvesting for photosystem II in green algae. *Nature Plants*. 2019;5(12):1320-30.

- 1 54. Caldwell P, Luk DC, Weissbach H, Brot N. Oxidation of the methionine residues of
2 Escherichia coli ribosomal protein L12 decreases the protein's biological activity. Proceedings of
3 the National Academy of Sciences of the United States of America. 1978;75(11):5349-52.
- 4 55. Shcherbik N, Pestov DG. The Impact of Oxidative Stress on Ribosomes: From Injury to
5 Regulation. Cells. 2019;8(11).
- 6 56. Simsek D, Barna M. An emerging role for the ribosome as a nexus for post-translational
7 modifications. Curr Opin Cell Biol. 2017;45:92-101.
- 8 57. Dalla Venezia N, Vincent A, Marcel V, Catez F, Diaz JJ. Emerging Role of Eukaryote
9 Ribosomes in Translational Control. Int J Mol Sci. 2019;20(5).
- 10 58. Deroo S, Hyung S-J, Marcoux J, Gordiyenko Y, Koripella RK, Sanyal S, et al.
11 Mechanism and rates of exchange of L7/L12 between ribosomes and the effects of binding EF-
12 G. ACS chemical biology. 2012;7(6):1120-7.
- 13 59. Gao YG, Selmer M, Dunham CM, Weixlbaumer A, Kelley AC, Ramakrishnan V. The
14 structure of the ribosome with elongation factor G trapped in the posttranslocational state.
15 Science. 2009;326(5953):694-9.
- 16 60. Datta PP, Sharma MR, Qi L, Frank J, Agrawal RK. Interaction of the G' domain of
17 elongation factor G and the C-terminal domain of ribosomal protein L7/L12 during translocation
18 as revealed by cryo-EM. Molecular cell. 2005;20(5):723-31.
- 19 61. Wang X, Kirkpatrick JP, Launay HMM, de Simone A, Häussinger D, Dobson CM, et al.
20 Probing the dynamic stalk region of the ribosome using solution NMR. Scientific reports.
21 2019;9(1):13528-.
- 22 62. Zhang Y, Launay H, Schramm A, Lebrun R, Gontero B. Exploring intrinsically
23 disordered proteins in Chlamydomonas reinhardtii. Scientific Reports. 2018;8(1):6805.
- 24 63. Merchant SS, Prochnik SE, Vallon O, Harris EH, Karpowicz SJ, Witman GB, et al. The
25 Chlamydomonas genome reveals the evolution of key animal and plant functions. Science.
26 2007;318(5848):245-50.
- 27 64. Zaffagnini M, Michelet L, Massot V, Trost P, Lemaire SD. Biochemical characterization
28 of glutaredoxins from Chlamydomonas reinhardtii reveals the unique properties of a
29 chloroplastic CGFS-type glutaredoxin. The Journal of biological chemistry. 2008;283(14):8868-
30 76.
- 31 65. Gao X-H, Zaffagnini M, Bedhomme M, Michelet L, Cassier-Chauvat C, Decottignies P,
32 et al. Biochemical characterization of glutaredoxins from Chlamydomonas reinhardtii: Kinetics
33 and specificity in deglutathionylation reactions. FEBS Letters. 2010;584(11):2242-8.
- 34 66. Marchler-Bauer A, Zheng C, Chitsaz F, Derbyshire MK, Geer LY, Geer RC, et al. CDD:
35 conserved domains and protein three-dimensional structure. Nucleic acids research.
36 2013;41(Database issue):D348-52.
- 37 67. Couso I, Evans BS, Li J, Liu Y, Ma F, Diamond S, et al. Synergism between Inositol
38 Polyphosphates and TOR Kinase Signaling in Nutrient Sensing, Growth Control, and Lipid
39 Metabolism in Chlamydomonas. Plant Cell. 2016;28(9):2026-42.
- 40 68. Couso I, Pérez-Pérez ME, Ford MM, Martínez-Force E, Hicks LM, Umen JG, et al.
41 Phosphorus Availability Regulates TORC1 Signaling via LST8 in Chlamydomonas. Plant Cell.
42 2020;32(1):69-80.
- 43 69. Ingargiola C, Turqueto Duarte G, Robaglia C, Leprince A-S, Meyer C. The Plant Target
44 of Rapamycin: A Conduc TOR of Nutrition and Metabolism in Photosynthetic Organisms.
45 Genes. 2020;11(11):1285.

- 1 70. Graciet E, Lebreton S, Gontero B. Emergence of new regulatory mechanisms in the
2 Benson-Calvin pathway via protein-protein interactions: a glyceraldehyde - 3 - phosphate
3 dehydrogenase/CP12/phosphoribulokinase complex. *Journal of Experimental Botany*.
4 2004;55(400):1245-54.
- 5 71. Moparthy SB, Thieulin-Pardo G, Mansuelle P, Rigneault H, Gontero B, Wenger J.
6 Conformational modulation and hydrodynamic radii of CP12 protein and its complexes probed
7 by fluorescence correlation spectroscopy. *The FEBS Journal*. 2014;281(14):3206-17.
- 8 72. Matsumura H, Kai A, Maeda T, Tamoi M, Satoh A, Tamura H, et al. Structure Basis for
9 the Regulation of Glyceraldehyde-3-Phosphate Dehydrogenase Activity via the Intrinsically
10 Disordered Protein CP12. *Structure*. 2011;19(12):1846-54.
- 11 73. Mileo E, Lorenzi M, Eroles J, Lignon S, Puppo C, Le Breton N, et al. Dynamics of the
12 intrinsically disordered protein CP12 in its association with GAPDH in the green alga
13 *Chlamydomonas reinhardtii*: a fuzzy complex. *Mol Biosyst*. 2013;9(11):2869-76.
- 14 74. Yu A, Xie Y, Pan X, Zhang H, Cao P, Su X, et al. Photosynthetic Phosphoribulokinase
15 Structures: Enzymatic Mechanisms and the Redox Regulation of the Calvin-Benson-Bassham
16 Cycle. *Plant Cell*. 2020;32(5):1556-73.
- 17 75. Park S, Khamai P, Garcia-Cerdan JG, Melis A. REP27, a tetratricopeptide repeat nuclear-
18 encoded and chloroplast-localized protein, functions in D1/32-kD reaction center protein
19 turnover and photosystem II repair from photodamage. *Plant Physiol*. 2007;143(4):1547-60.
- 20 76. Dewez D, Park S, García-Cerdán JG, Lindberg P, Melis A. Mechanism of REP27 protein
21 action in the D1 protein turnover and photosystem II repair from photodamage. *Plant physiology*.
22 2009;151(1):88-99.
- 23 77. Theis J, Schroda M. Revisiting the photosystem II repair cycle. *Plant signaling &*
24 *behavior*. 2016;11(9):e1218587.
- 25 78. Jo GA, Lee JM, No G, Kang DS, Kim SH, Ahn SH, et al. Isolation and characterization
26 of a 17-kDa FKBP-type peptidyl-prolyl cis/trans isomerase from *Vibrio anguillarum*. *Protein*
27 *expression and purification*. 2015;110:130-7.
- 28 79. Vallon O. *Chlamydomonas* immunophilins and parvulins: survey and critical assessment
29 of gene models. *Eukaryotic cell*. 2005;4(2):230.
- 30 80. Perez-Perez ME, Mauries A, Maes A, Tourasse NJ, Hamon M, Lemaire SD, et al. The
31 Deep Thioredoxome in *Chlamydomonas reinhardtii*: New Insights into Redox Regulation. *Mol*
32 *Plant*. 2017;10(8):1107-25.
- 33 81. Eroles J, Lorenzi M, Lebrun R, Fournel A, Etienne E, Courcelle C, et al. A New Function
34 of GAPDH from *Chlamydomonas reinhardtii*: A Thiol- Disulfide Exchange Reaction with
35 CP12. *Biochemistry*. 2009;48(25):6034-40.
- 36 82. Eroles J, Mekhalfi M, Woudstra M, Gontero B. Molecular mechanism of NADPH-
37 glyceraldehyde-3-phosphate dehydrogenase regulation through the C-terminus of CP12 in
38 *Chlamydomonas reinhardtii*. *Biochemistry*. 2011;50(14):2881-8.
- 39 83. Ford MM, Smythers AL, McConnell EW, Lowery SC, Kolling DRJ, Hicks LM.
40 Inhibition of TOR in *Chlamydomonas reinhardtii* Leads to Rapid Cysteine Oxidation Reflecting
41 Sustained Physiological Changes. *Cells*. 2019;8(10).
- 42 84. Werth EG, McConnell EW, Couso Lianez I, Perrine Z, Crespo JL, Umen JG, et al.
43 Investigating the effect of target of rapamycin kinase inhibition on the *Chlamydomonas*
44 *reinhardtii* phosphoproteome: from known homologs to new targets. *New Phytologist*.
45 2019;221(1):247-60.
- 46

1

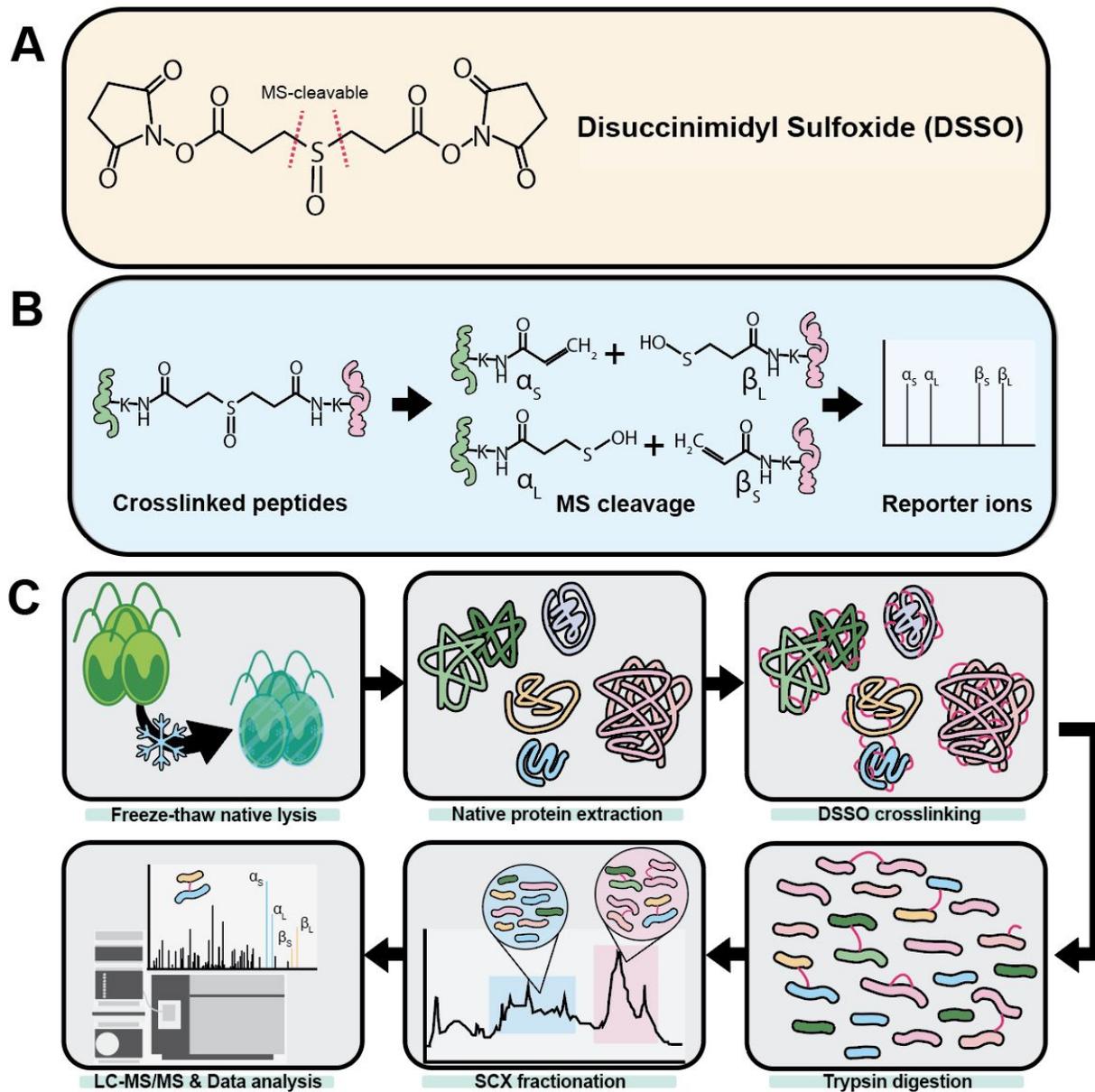


Figure 1. General workflow for *C. reinhardtii* interactomics. **(A)** The MS-cleavable crosslinker DSSO was used in this study. **(B)** DSSO generates reporter ions that enable delineation between crosslinks. These reporter ions are essential for minimizing false discovery rate in identifications. **(C)** Native protein structures and interactions were preserved and extracted using a gentle lysis via freeze-thaw. Proteins were crosslinked with DSSO and proteolyzed using trypsin. Crosslinked peptides were enriched and fractionated to increase depth of coverage for crosslink identifications. After LC-MS/MS analysis, the XLinkX nodes in Proteome Discoverer were used to recognize crosslinks by characteristic crosslink reporter ions (α_S , α_L , β_S , and β_L) in the MS2 spectra and identify crosslink-spectrum matches with database searching using the fragment ions from the MS2 spectra.

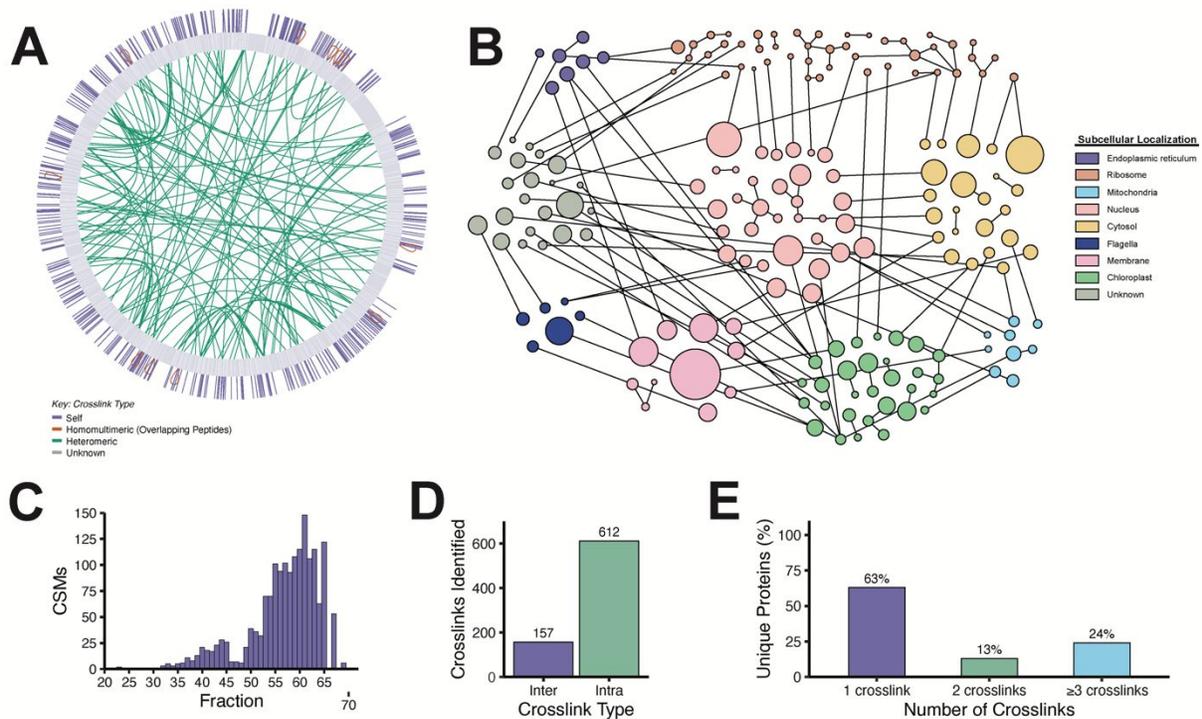


Figure 2. Description of identified *C. reinhardtii* crosslinked peptides following DSSO crosslinking of protein extracts. **(A)** Circos plot of all identified crosslinks. Inner, green curves represent interlinks between two different proteins, outer, orange curves represent interlinks within a homoligomer, and outer, purple curves represent intralinks within the same protein. **(B)** Protein interactome map featuring identified interlinks. Proteins are grouped by subcellular localization. **(C)** Histogram of the number of identified CSMs across each SCX fraction. **(D)** Types of detected crosslinks. **(E)** Distribution of identified unique proteins by number of detected crosslinks.

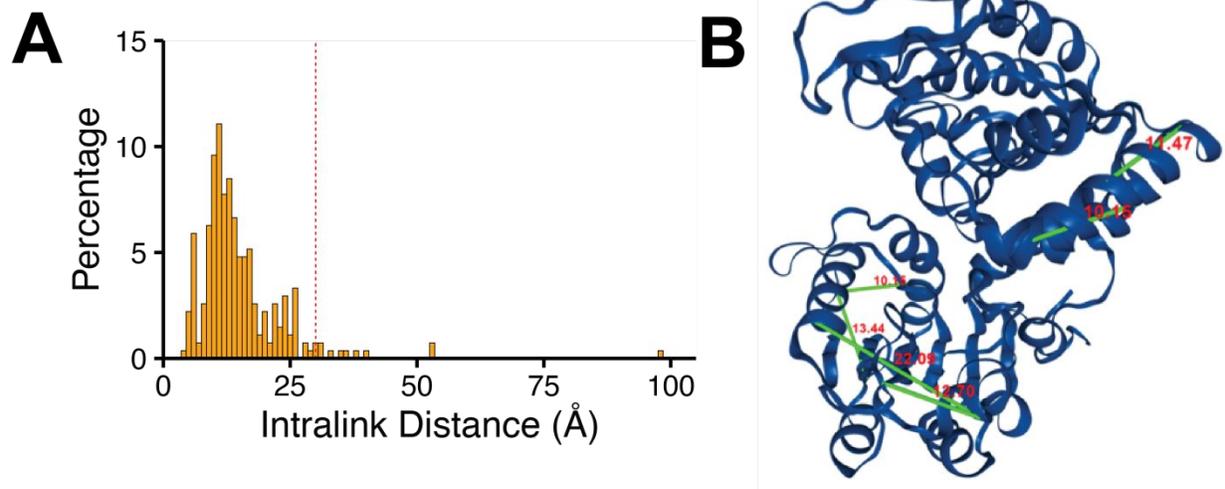


Figure 3. Intralinks were mapped onto known protein structures or homology models using the Integrative Modeling Platform in XLinkDB 3.0. **(A)** Histogram showing Euclidean distances between residues of detected intralinks. The red, dashed line represents the DSSO maximum crosslinking distance of 30 Å. Of intralinks that were able to be mapped, 96% featured residues within the theoretical distance of DSSO. **(B)** The 6 intralinks (in green) identified from *C. reinhardtii* phosphoglycerate kinase that mapped onto the structure of phosphoglycerate kinase from *T. maritima* (PDB ID: 1VPE). The Euclidean distances between the residues in each intralinks (in red) are less than 30 Å, the maximum crosslinking distance for DSSO.

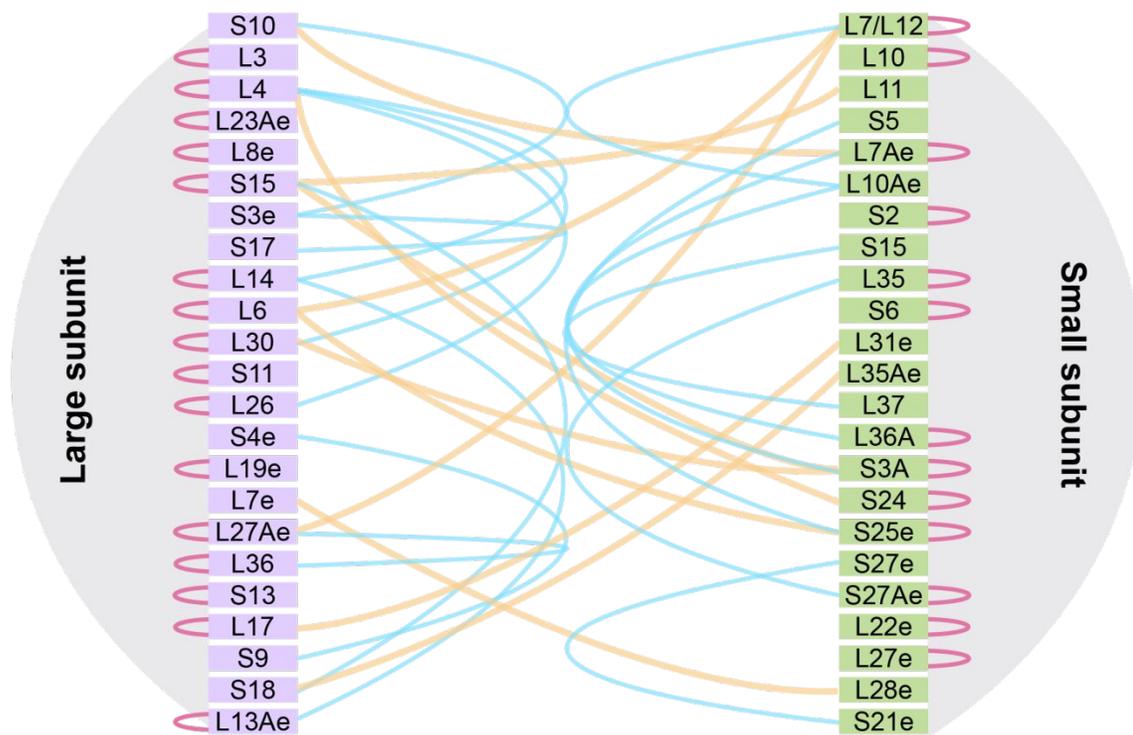


Figure 5. Visualization of all crosslinked subunits of the cytosolic ribosome. Intralinks are visualized in pink while interlinks within the large or small ribosomal subunits are visualized in blue. The orange interlinks represent interactions between proteins found in opposite subunits.

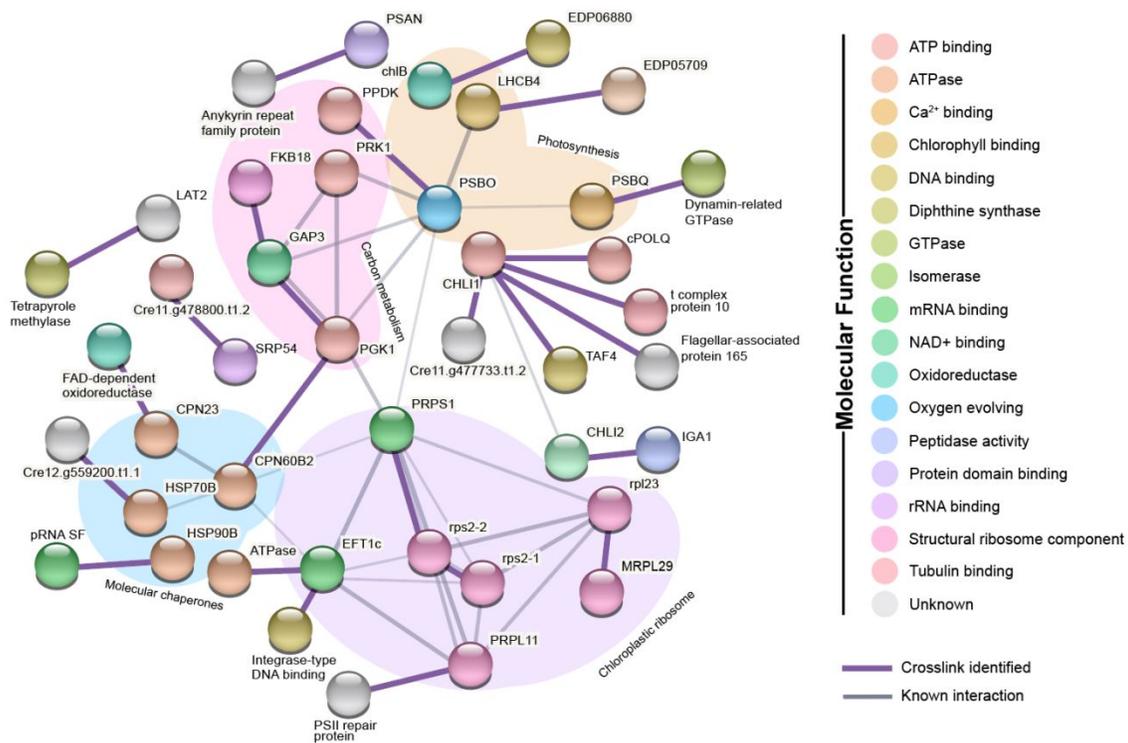


Figure 7. Protein interactome map featuring identified interlinks localized to the chloroplast. Proteins are color coded by molecular function and clustered biological processes are shaded and labeled. Purple lines denote observed interlinks in this study while grey lines indicate empirical evidence of interactions as compiled by the STRING database, where increasing line thickness indicates increased confidence in interaction.

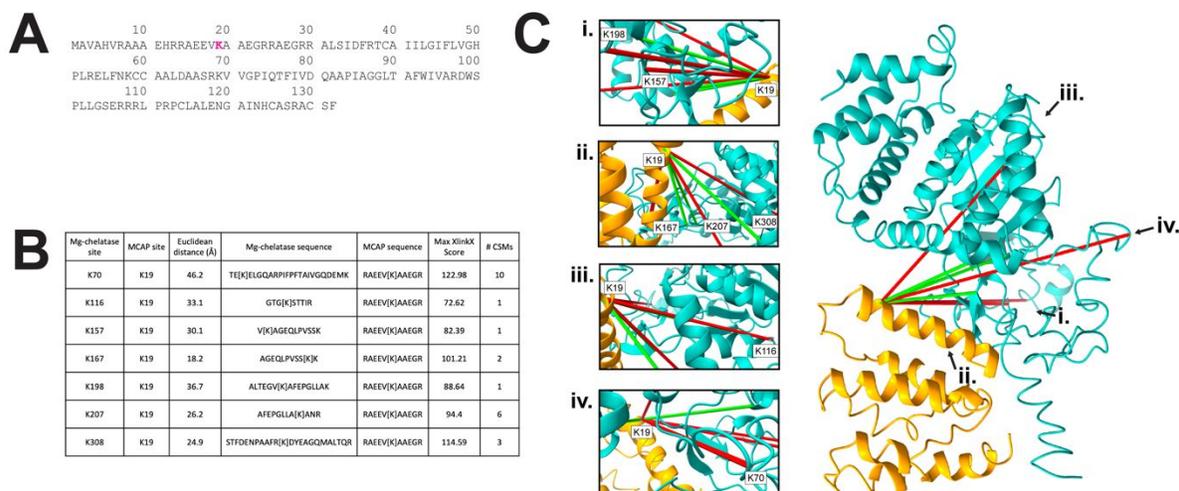


Figure 8. Crosslinking unveiled a novel complex between Mg-chelatase (Cre06.g306300.t1.2) and uncharacterized protein Cre11.g477733.t1.2, herein referred to as Mg-chelatase associated protein, or MCAP. **(A)** Sequence of MCAP, which is uncharacterized on UniProt. The crosslinked residue is shown in pink. **(B)** Table of observed crosslinks between Mg-chelatase and MCAP. Crosslinked residues are shown in brackets. Euclidean distance refers to that interlink mapped onto the refined protein complex shown in Figure 8C. **(C)** I-TASSER and HADDOCK were used to generate a protein-protein interaction model between Mg-chelatase (teal) and MCAP (orange). The structure with the most mapped intralinks within the maximum restraint for DSSO (in green) from the refined complex modeling is shown. The boxes labeled with roman numerals show the crosslinked lysine residues from the perspective of the arrows on the model complex structure.