

Species Diversity of Thylakoid Proteins for Eudicots and Monocots with Photosynthesis and Chemical Carcinogenesis NADPH

Jeff Cromwell, PhD

The Mathematical Learning Space Research Portfolio

1 Abstract

Photosynthesis uses light energy to synthesize organic compounds from carbon dioxide and water with a light dependent part (light reaction) and the light independent part (dark reaction, carbon fixation). The light reaction takes place in thylakoid, a membrane-bound compartment inside chloroplasts and cyanobacteria. The light energy is used by photosystems I and II to generate proton motive force and reducing power (NADPH or NADH). The proton motive force is used by ATP synthase to generate ATP, as in the mitochondrial respiratory chain where the supplies of ATP and NAD(P)H are then used to fix carbon dioxide. A major function of the thylakoid membrane and its integral photosystems is the establishment of chemiosmotic potential. The carriers in the electron transport chain use some of the electron's energy to actively transport protons from the stroma to the lumen. During photosynthesis, the lumen becomes acidic, pH 4, compared to pH 8 in the stroma where the change is a 10,000 fold concentration gradient for protons across the thylakoid membrane.

2 Introduction

In the Chemical carcinogenesis - reactive oxygen species, cancer is a multi-step process which involves initiation, promotion and progression with chemical carcinogens, (1) directly or after xenobiotic metabolism, (2) induce DNA damage and act in a 'genotoxic' manner and carcinogens can induce cancer via nongenotoxic mechanisms such as oxidative stress. Trace metals and organic xenobiotics are typical classes of environmental pollutants with prooxidant effects that generate reactive oxygen species (ROS) depending on their ability to lose electrons and catalyze Haber Weiss and Fenton reactions. ROS is generated from the induction of various cytochrome P450 isoenzymes by detoxification of chemical carcinogens. Increased ROS generation creates DNA damage with mutations and have an important role in the initiation and progression of multistage carcinogenesis. Besides DNA damage, ROS further induce multiple

intracellular signaling pathways, notably (1) NF-kappa B, (2) JNK/SAPK/p38, as well as (3) Erk/MAPK. These signaling routes can lead to transcriptional induction of target genes to promote proliferation or confer apoptosis resistance to exposed cells. [401]

Figure 1 has the Chemical carcinogenesis - reactive oxygen species map

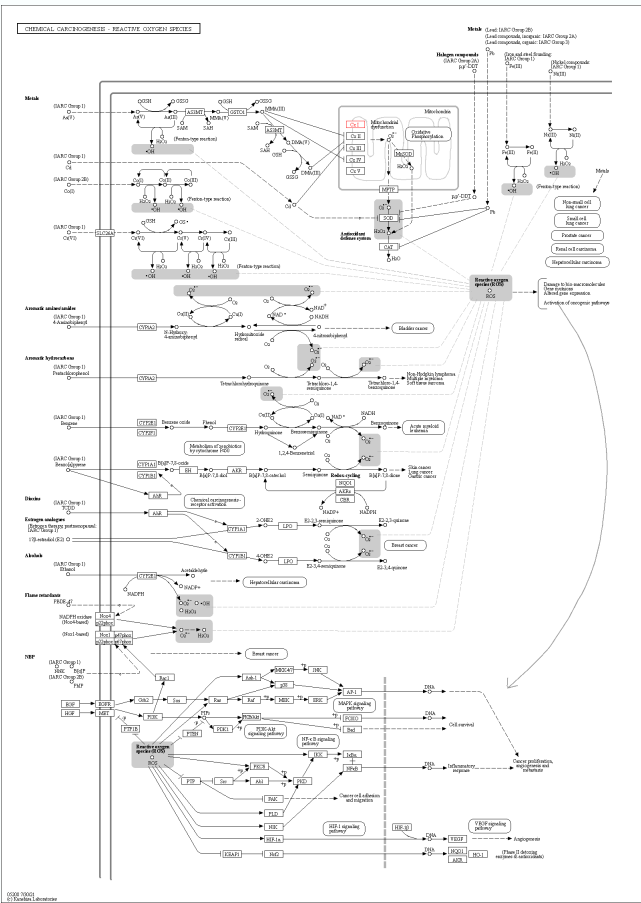


Figure 1: Chemical carcinogenesis - reactive oxygen species [401]

2.1. Water photolysis

The first step in photosynthesis is light dependent water separation to provide the electrons for the photosynthetic electron transport chains as well as protons for the establishment of a proton gradient. The water-separation reaction is on the luminal side of the thylakoid membrane from the light energy captured by the photosystems. This oxidation of water conveniently produces the waste product O₂ needed for cellular respiration with the molecular oxygen formed by the reaction released into the atmosphere. Thylakoids contain many integral and peripheral membrane proteins, as well as luminal proteins.[1]

Figure 2 has the Photosynthesis map

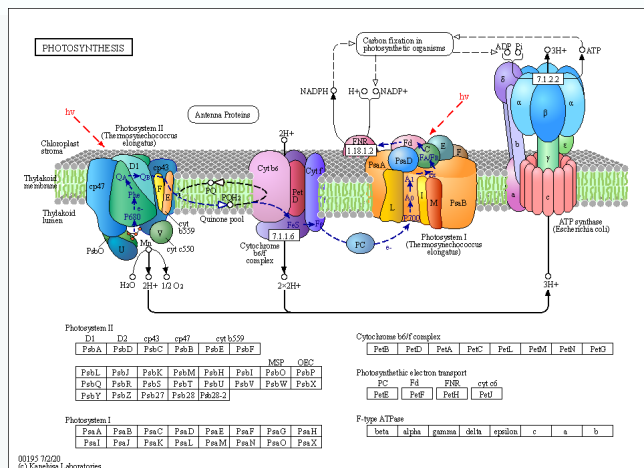


Figure 2: Photosynthesis with C00001 H₂O, C00002 ATP, C00005 NADPH, C00006 NADP⁺, C00007 Oxygen, C00008 ADP, C00009 Orthophosphate, C00034 Manganese, C00080 H⁺, C02061 Plastoquinone and C02185 Plastocyanin-1 [401]

According to these studies, the thylakoid proteome consists of at least 335 different proteins. Out of these, 89 are in the lumen, 116 are integral membrane proteins, 62 are peripheral proteins on the stroma side, and 68 peripheral proteins on the luminal side. Of the thylakoid proteins with known functions, 42 percent are involved in photosynthesis with the next largest functional groups include proteins involved in protein targeting, processing and folding with 11 percent, oxidative stress response (9 percent) and translation (8 percent). [1]

2.2. Integral membrane proteins

Thylakoid membranes contain integral membrane proteins with an important role in light-harvesting and the light-dependent reactions of photosynthesis. There are four major protein complexes in the thylakoid membrane: (a) Photosystems I and II (b) Cytochrome b₆f complex and (c) ATP synthase. Photosystem II is in grana thylakoids and photosystem I and ATP synthase are in stroma thylakoids and the outer layers of grana. The cytochrome b₆f complex is distributed evenly throughout thylakoid membranes. Because of separation the two photosystems in the thylakoid membrane system, mobile electron carriers (plastoquinone and plastocyanin) have to shuttle electrons between them. Plastoquinone shuttles electrons from photosystem II to the cytochrome b₆f complex, whereas plastocyanin carries electrons from the cytochrome b₆f complex to photosystem I. [1]

Together, these proteins use light energy to drive electron transport chains for a chemiosmotic potential across the thylakoid membrane and NADPH, a product of the terminal redox reaction. The ATP synthase uses the chemiosmotic potential to make ATP during photophosphorylation. [1]

2.3. NADPH

Figure 3 has the chemical structure of NADPH, TPNH, Reduced nicotinamide adenine dinucleotide phosphate with formula C₂₁H₃₀N₇O₁₇P₃, EXACT MASS 745.0911 and MOL WEIGHT 745.4209 for KEGG compound C00005 and Pubchem 3307. [800]

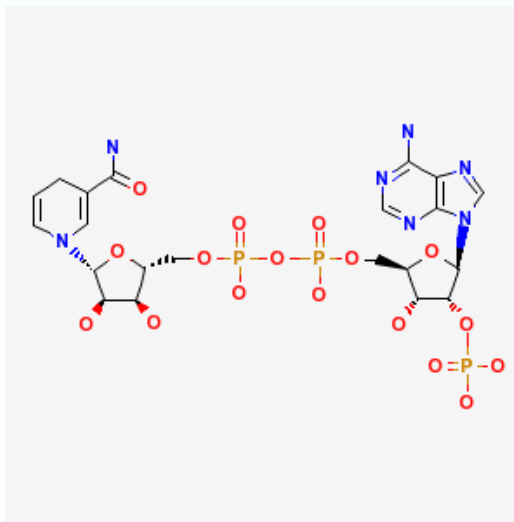


Figure 3: NADPH, TPNH, Reduced nicotinamide adenine dinucleotide phosphate C00005 [401]

There are N1=134 reactions and N2=588 Enzymes with 557 in class 1 Oxidoreductases and rest in 2,3 and 4 Transferases, Hydrolases and Lyases. Categories 1.1.1 Acting on the CH-OH group of donors With NAD⁺ or NADP⁺ as acceptor has 209 and 1.14 Acting on paired donors, with incorporation or reduction of molecular oxygen has 108. Fingerprint Tanimoto-based 2-dimensional similarity search generates N=845 structures in the PubChem database.

Table 1 has the frequency of the top 25 compounds from a subset of pH 7 liquid reactants or products. [401]

	Frequency
+	6654
C00001	2297
<=>	2296
2	674
C00007	608
C00080	531
C00014	230
C00006	220
C00005	219
C00009	212
C00003	201
C00004	197
C03024	129
C03161	129
C00011	116
C00027	101
3	86
C00028	74
C00030	74
C00010	60
4	59
C00033	59
C00031	56
C00022	47
C00138	47

Table 2 has the KEGG Chemical Function (KCF) format. (a) KEGG Chemical Function (KCF) format of NADPH. [401]

	X
1	1 C1y C 45.6720 -22.9148
2	2 N4y N 46.0143 -20.6899
3	3 C1y C 45.2747 -24.1557
4	4 O2x O 44.5777 -22.1203
5	5 C8y C 44.7428 -20.2924
6	6 C8x C 46.7783 -19.6078
7	7 C1y C 43.9237 -24.1557
8	8 O2b O 46.0570 -25.2315
9	9 C1y C 43.5020 -22.8903
10	10 C8y C 44.7366 -18.9598
11	11 N5x N 43.5874 -20.9709
12	12 N5x N 45.9960 -18.5503
13	13 O1a O 43.1595 -25.2377
14	14 P1b P 46.0510 -27.0532
15	15 C1b C 42.2489 -22.4991
16	16 C8y C 43.5691 -18.2934
17	17 C8x C 42.4139 -20.3231
18	18 O1c O 46.0632 -28.6487
19	19 O1c O 47.4202 -27.0594
20	20 O1c O 44.7244 -27.0532
21	21 O2b O 41.1730 -22.4504
22	22 N5x N 42.4139 -18.9721
23	23 N1a N 43.5629 -16.9609
24	24 P1b P 39.7364 -22.4441
25	25 O2c O 37.9454 -22.4441
26	26 O1c O 39.7364 -23.9969
27	27 O1c O 39.7426 -20.9525
28	28 P1b P 36.2825 -22.4441
29	29 O2b O 34.8461 -22.4504
30	30 O1c O 36.2887 -23.9295
31	31 O1c O 36.2825 -20.9525
32	32 C1b C 33.4646 -22.4381
33	33 C1y C 32.2175 -22.8721
34	34 O2x O 31.1172 -22.1385
35	35 C1y C 31.8752 -22.2290
36	36 C1y C 30.0475 -22.9699
37	37 C1y C 30.4937 -24.2536
38	38 O1a O 32.6577 -25.2071
39	39 N1y N 28.7272 -21.6861
40	40 O1a O 29.7479 -25.2989
41	41 C2x C 29.7479 -20.9342
42	42 C2x C 27.4555 -20.9892
43	43 C2y C 29.7541 -19.5956
44	44 C2x C 27.4433 -19.7361
45	45 C1x C 28.5682 -18.9965
46	46 C5a C 30.8667 -18.8620
47	47 N1a N 30.8605 -17.6822
48	48 O5a O 32.0524 -19.4609

Table 1: KEGG Chemical Function (KCF) format. (a) KEGG Chemical Function (KCF) format of NADPH. KCF format has three sections; ENTRY, ATOM and BOND. ENTRY section describes the KEGG ID and the type of the entry. ATOM describes the numbering of the atoms, KEGG Atom Types for the labels on the atoms, atomic species (C for carbon, N for nitrogen, etc), and 2D coordinates of the atoms. [2]

Table 3 has the numbering of the bonds, the numbering of the two atoms in the bond, and the bond order, and steric configuration of the bond. The steric configuration of the bond are steric effects are nonbonding interactions that influence the shape (conformation) and reactivity of ions and molecules. ... Steric repulsive forces between overlapping electron clouds result in structured groupings of molecules stabilized by the way that opposites attract and like charges repel. [401]

	Bond
1	52
2	1 1 2 1 #Up
3	2 1 3 1
4	3 1 4 1
5	4 2 5 1
6	5 2 6 1
7	6 3 7 1
8	7 3 8 1 #Down
9	8 4 9 1
10	9 5 10 2
11	10 5 11 1
12	11 6 12 2
13	12 7 13 1 #Down
14	13 8 14 1
15	14 9 15 1 #Up
16	15 10 16 1
17	16 11 17 2
18	17 14 18 1
19	18 14 19 1
20	19 14 20 2
21	20 15 21 1
22	21 16 22 2
23	22 16 23 1
24	23 21 24 1
25	24 24 25 1
26	25 24 26 1
27	26 24 27 2
28	27 25 28 1
29	28 28 29 1
30	29 28 30 1
31	30 28 31 2
32	31 29 32 1
33	32 33 32 1 #Down
34	33 33 34 1
35	34 33 35 1
36	35 34 36 1
37	36 35 37 1
38	37 35 38 1 #Up
39	38 36 39 1 #Down
40	39 37 40 1 #Up
41	40 39 41 1
42	41 39 42 1
43	42 41 43 2
44	43 42 44 2
45	44 43 45 1
46	45 43 46 1
47	46 46 47 1
48	47 46 48 2
49	48 7 9 1
50	49 10 12 1
51	50 17 22 1
52	51 36 37 1
53	52 44 45 1

Table 2: BOND describes the numbering of the bonds, the numbering of the two atoms in the bond, and the bond order, and steric configuration of the bond. [401] [3]

Figure 3 has the Tanimoto Structural Similarities N=845 with CID5884 with (a) oxygen (b) carbon (c) hydrogen (d) Nitrogen with a location scale [0,1] with 0=start and 1=end for the SMILE sequence.

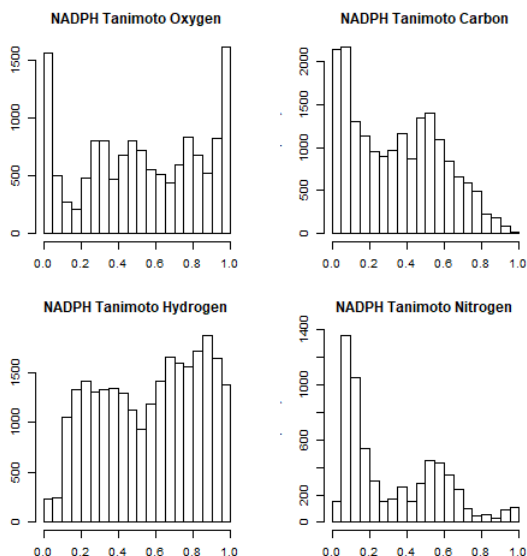


Figure 4: Tanimoto Structural Similarities N=845 with CID5884 with (a) oxygen (b) carbon (c) hydrogen (d) Nitrogen with a location scale [0,1] with 0=start and 1=end for the SMILE sequence. [601] The Pearson distribution with moments are (a) type 2 a 0.9360405 location -0.0176163 scale 1.017616, (b) type 3 shape 1.31062 location 0.003256247 scale 0.2609505, (c) type 2 a 0.9463522 location 0.002016147 scale 0.9966398 and (d) type 3 shape 1.365883 location 0.0120604 scale 0.2251447.

3 Thylakoid Protein Networks

Cyanobacteria are photosynthetic prokaryotes with highly differentiated membrane systems with an internal system of thylakoid membranes of fully functional electron transfer chains of photosynthesis for respiration. The presence of different membrane systems lends these cells a unique complexity among bacteria and reorganization the membranes, synthesize new membrane lipids, and properly target proteins to the correct membrane system is important. The outer membrane, plasma membrane, and thylakoid membranes each have specialized roles in the cyanobacterial cell. [1]

In contrast to the thylakoid network of higher plants, thylakoids in cyanobacteria are organized into multiple concentric shells that split and fuse to parallel layers forming a highly connected network that encloses a single lumen (as in higher-plant chloroplasts) and allows water-soluble and lipid-soluble molecules to diffuse through the entire membrane network. Perforations within the parallel thylakoid sheets allow for the traffic of particles of different sizes throughout the cell: ribosomes, glycogen granules, and lipid bodies. The relatively large distance between the thylakoids provides space for the external light-harvesting antennae, the phycobilisomes. This macrostructure, as in the case of higher plants has flexibility during changes in the physicochemical environment. [1]

In Eudicots where N=93, the distribution is given by Mustard family (9), Caper family (1), Papaya family (1), Rue family (2), Sumac family (2), Mallow family (5), Myrtle family (1), Pea family (15), Rose family (8), Buckthorn family (1), Mulberry family (1), Cucumber family (7), Spurge family (4), Willow family (3), Walnut family (1), Beech family (2), Bittersweet family (1), Grape family (2), Nightshade family (8), Morning-glory family (2), Sesame family (1), Olive family (1), Lopseed family (1), Mint family (1), Daisy family (4), Parsley family (1), Tea family (1), Amaranth family (3), Lotus family (1), Protea family (1), Poppy family (1) and Water-lily family (1). In Monocots, N=18 and the distribution is Grass family (12), Palm family (2), Banana family (1), Orchid family (2) and Asparagus family (1).

3.1. *Oryza sativa ssp. japonica* cultivar *Nipponbare* (Japanese rice)

Monocots with the Grass family includes: (a) osa *Oryza sativa japonica* (Japanese rice) (RefSeq) (b) dosa *Oryza sativa japonica* (Japanese rice) (RAPDB), (c) obr *Oryza brachyantha* (malo sina), (d) bdi *Brachypodium distachyon*, (e) ats *Aegilops tauschii* (wheat D), (f) tdc *Triticum dicoccoides* (wild emmer wheat), (g) taes *Triticum aestivum* (bread wheat), (h) sbi *Sorghum bicolor* (sorghum), (i) zma *Zea mays* (maize), (j) sita *Setaria italica* (foxtail millet), (k) pvir *Panicum virgatum* (switchgrass) and (l) phai *Panicum hallii*

Oryza sativa ssp. japonica cultivar *Nipponbare* (Japanese rice) with T01015 complete Genome Org code osa with the number of protein genes: 28073 and the number of RNA genes: 71.

3.2. *Rosa chinensis* (China rose)

In the Rose family, the following varieties: (a) fve *Fragaria vesca* (woodland strawberry), (b) rcn *Rosa chinensis* (China rose) (c) pper *Prunus persica* (peach), (d) pmum *Prunus mume* (Japanese apricot) (e) pavi *Prunus avium* (sweet cherry), (f) pdul *Prunus dulcis* (almond), (g) mdm *Malus domestica* (apple) and (h) pxb *Pyrus x bretschneideri* (Chinese white pear).

Rosa chinensis (China rose) is from Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliopsida; eudicotyledons; Gunneridae; Pentapetalae; rosids; fabids; Rosales; Rosaceae; Rosoideae; Rosoideae incertae sedis; Rosa. The Number of protein genes is 30924 and the Number of RNA genes: 866.

3.3. Species Eudicots and Monocots Families and Thylakoid Proteins

Consider N1=149 unique protein names and N2=165 unique species in both Table 1 and 2 obtained from the KEGG database. The combination generated N3=1597 sequences to examine.

1	2	3	4
1 Nothobranchius furzeri (turquoise killifish)	Morus notabilis	Spinacia oleracea (spinach)	Synechocystis sp. PCC 6803 GT-S
2 Arabidopsis thaliana (thale cress)	Cucumis sativus (cucumber)	Chenopodium quinoa (quinoa)	Synechocystis sp. PCC 6803 GT-1
3 Arabidopsis lyrata (lyrate rockcress)	Cucumis melo (muskmelon)	Nelumbo nucifera (sacred lotus)	Synechocystis sp. PCC 6803 PCC-N
4 Capsella rubella	Benincasa hispida (wax gourd)	Macadamia integrifolia (macadamia nut)	Synechocystis sp. PCC 6803 PCC-P
5 Camelina sativa (false flax)	Momordica charantia (bitter melon)	Papaver somniferum (opium poppy)	Synechocystis sp. PCC 6714
6 Brassica rapa (field mustard)	Cucurbita maxima (winter squash)	Nymphaea colorata	Synechocystis sp. IPPAS B-1465
7 Brassica napus (rape)	Cucurbita moschata (crookneck pumpkin)	Oryza sativa japonica (Japanese rice) (RefSeq)	Synechococcus elongatus PCC7942
8 Brassica oleracea (wild cabbage)	Cucurbita pepo subsp. pepo (vegetable marrow)	Oryza sativa japonica (Japanese rice) (RAPDB)	Synechococcus sp. CC9902
9 Raphanus sativus (radish)	Ricinus communis (castor bean)	Oryza brachyantha (malo sina)	Synechococcus sp. WH 8103
10 Tarenaya hassleriana (spider flower)	Jatropha curcas	Brachypodium distachyon	Synechococcus sp. PCC 73109
11 Carica papaya (papaya)	Hevea brasiliensis (rubber tree)	Aegilops tauschii (wheat D)	Synechococcus sp. PCC 7003
12 Citrus sinensis (Valencia orange)	Manihot esculenta (cassava)	Triticum diocoides (wild emmer wheat)	Prochlorococcus sp. MIT 0604
13 Pistacia vera (pistachio)	Populus trichocarpa (black cottonwood)	Triticum aestivum (bread wheat)	Prochlorococcus sp. MIT 0801
14 Mangifera indica (mango)	Populus euphratica (Euphrates poplar)	Sorghum bicolor (sorghum)	Synechococcus lividus
15 Theobroma cacao (cacao)	Populus alba (white poplar)	Zea mays (maize)	Thermosynechococcus sp. NK55
16 Gossypium raimondii	Juglans regia (English walnut)	Setaria italica (foxtail millet)	Kovackia sp. CCNU0001
17 Gossypium hirsutum (upland cotton)	Quercus suber (cork oak)	Panicum virgatum (switchgrass)	Halomicronema hongdechloris
18 Gossypium arboreum	Quercus lobata (valley oak)	Panicum hallii	Pseudanabaena sp. ABRG5-3
19 Durio zibethinus (durian)	Tripterygium wilfordii	Phoenix dactylifera (date palm)	Geminocystis sp. NIES-3708
20 Eucalyptus grandis (rose gum)	Vitis vinifera (wine grape)	Elaeis guineensis (African oil palm)	Microcystis panniformis
21 Glycine max (soybean)	Vitis riparia (riverbank grape)	Musa acuminata (wild Malaysian banana)	Microcystis sp. MC19
22 Glycine soja (wild soybean)	Solanum lycopersicum (tomato)	Dendrobium catenatum	Cyanobacterium sp. HL-69
23 Vigna radiata (mung bean)	Solanum pennellii	Phalaenopsis equestris	Trichodesmium erythraeum
24 Vigna angularis (adzuki bean)	Solanum tuberosum (potato)	Asparagus officinalis (garden asparagus)	Oxynema aestuarii
25 Vigna unguiculata (cowpea)	Capsicum annuum	Ambrosia trichopoda	Limnospira fusiformis
26 Cajanus cajan (pigeon pea)	Nicotiana tabacum (common tobacco)	Physcomitrium patens	Moorea producens
27 Abrus precatorius (Indian licorice)	Nicotiana glauca	Chlamydomonas reinhardtii	Phormidium sp. PBR-2020
28 Medicago truncatula (barrel medic)	Nicotiana glauca	Volvox carterii f. nagariensis	Nostoc sp. PCC 7120
29 Cicer arietinum (chickpea)	Nicotiana glauca	Monoraphidium neglectum	Nostoc sp. NIES-3756
30 Arachis duranensis	Ipomoea nil (Japanese morning glory)	Coccomyxa subellipsoidea	Nostoc sp. CENA543
31 Arachis ipaensis	Ipomoea triloba (trilobed morning glory)	Ostreococcus lucimarinus	Nostoc sphaeroides
32 Arachis hypogaea (peanut)	Sesamum indicum (sesame)	Micromonas commoda	Nostoc edaphicum
33 Lupinus angustifolius (narrow-leaved blue lupine)	Olea europaea var. sylvestris (wild olive)	Cyanidioschyzon merolae	Anabaena sp. WA102
34 Fragaria vesca (woodland strawberry)	Erythranthe guttata (spotted monkey flower)	Galdieria sulphuraria	Calothrix sp. 336/3
35 Rosa chinensis (China rose)	Salvia splendens (scarlet sage)	Thalassiosira pseudonana	Nodularia spumigena
36 Prunus persica (peach)	Helianthus annuus (common sunflower)	Acetobacter oryzafermentans	Cylindrospermopsis curvispora
37 Prunus mume (Japanese apricot)	Erigeron canadensis (horseweed)	Acetobacter pomorum	Nostocales cyanobacterium HT-58-2
38 Prunus avium (sweet cherry)	Lactuca sativa (garden lettuce)	Acetobacter sp. JWB	Viruses (Paramecium bursaria Chlorella virus NYs1)
39 Prunus dulcis (almond)	Cynara cardunculus var. scolymus (artichoke)	Acetobacter oryzoeni	Nothobranchius furzeri (turquoise killifish)
40 Malus domestica (apple)	Daucus carota (carrot)	Clostridium pasteurianum DSM 525 = ATCC 6013	Arabidopsis thaliana (thale cress)
41 Pyrus x bretschneideri (Chinese white pear)	Camellia sinensis	Thermoclostridium stercorarium subsp. stercorarium DSM 8532	Arabidopsis lyrata (lyrate rockcress)
42 Ziziphus jujuba (Chinese jujube)	Beta vulgaris (sugar beet)	Syntrophomonas wolfei	

Table 2 has the list of the thylakoid proteins.

1	2	3	4
1 (RefSeq) thylakoid ADP/ATP carrier protein, chloroplastic-like	(RefSeq) thylakoid membrane protein TERC, chloroplastic isoform X4	(GenBank) thylakoid luminal 17.4 kDa protein	(RefSeq) thylakoid luminal 30 kDa protein, probable
2 (RefSeq) cold regulated 314 thylakoid membrane 2	(RefSeq) thylakoid membrane protein slr0575	(GenBank) thylakoid luminal 17.9 kDa protein	(RefSeq) chloroplast thylakoid protein kinase STN7, probable
3 (RefSeq) thylakoid lumen 18.3 kDa protein	(RefSeq) thylakoid membrane protein TERC, chloroplastic isoform X2	(GenBank) thylakoid membrane phosphoprotein 14 kDa protein	(RefSeq) thylakoid luminal 20 kDa protein, putative
4 (RefSeq) thylakoid luminal protein (Mog1/PsbP/DUF1795-like photosystem II reaction center PsbP family protein)	(RefSeq) LOW QUALITY PROTEIN: thylakoid processing peptidase 1, chloroplastic-like	(GenBank) thylakoid luminal protein	(RefSeq) thylakoid luminal protein, chloroplast precursor
5 (RefSeq) thylakoid ascorbate peroxidase	(RefSeq) thylakoid luminal 15.0 kDa protein 2, chloroplastic-like	(GenBank) thylakoid luminal 29 kDa protein	(RefSeq) thylakoid luminal 25.6kDa protein like, chloroplast precursor
6 (RefSeq) thylakoid processing peptide	(RefSeq) thylakoid membrane protein TERC, chloroplastic-like isoform X1	(GenBank) thylakoid membrane slr0575-like protein	(RefSeq) similar to thylakoid luminal 17.4 kD protein, chloroplast precursor
7 (RefSeq) thylakoid luminal P17.1 protein	(RefSeq) thylakoid membrane protein TERC, chloroplastic-like isoform X2	(GenBank) oxygen-evolving complex/thylakoid luminal 25.6 kDa protein	(RefSeq) probable thylakoid lumen rotamase
8 (RefSeq) thylakoid soluble phosphoprotein	(RefSeq) thylakoid luminal 15 kDa protein 1, chloroplastic-like isoform X1	(GenBank) thylakoid rhodanese-like protein	(RefSeq) similar to thylakoid lumen rotamase
9 (RefSeq) thylakoid rhodanese-like protein	(RefSeq) thylakoid luminal protein TL20.3, chloroplastic-like	(GenBank) thylakoid luminal 16.5 kDa protein	(RefSeq) similar to thylakoid sec-independent protein transporter Tha4
10 (RefSeq) chloroplast thylakoid lumen protein	(RefSeq) thylakoid membrane protein TERC, chloroplastic isoform X1	(RefSeq) thylakoid processing peptidase 1, chloroplastic isoform X1	(RefSeq) thylakoid protein
11 (RefSeq) thylakoid luminal 17.9 kDa protein, chloroplast	(RefSeq) thylakoid luminal 29 kDa protein, chloroplastic isoform X1	(RefSeq) thylakoid luminal protein TL20.3, chloroplastic isoform X2	(RefSeq) thylakoid luminal 17.4 kDa protein, chloroplast precursor
12 (RefSeq) thylakoid-associated phosphatase 38	(RefSeq) thylakoid luminal 16.5 kDa protein, chloroplastic-like	(RefSeq) thylakoid luminal 15.0 kDa protein 2, chloroplastic-like isoform X1	(RefSeq) thylakoid processing peptidase
13 (RefSeq) thylakoid ATP/ADP carrier	(RefSeq) thylakoid membrane protein slr0575-like	(RefSeq) LOW QUALITY PROTEIN: thylakoid luminal 17.4 kDa protein, chloroplastic	(RefSeq) thylakoid lumen 15.0-kDa protein
14 (RefSeq) thylakoid lumen 15.0 kDa protein	(RefSeq) thylakoid luminal protein At1g12250, chloroplastic	(RefSeq) thylakoid luminal 15 kDa protein 1, chloroplastic isoform X2	(RefSeq) ATP synthase alpha chain, thylakoid ATPase, plastid protein
15 (RefSeq) thylakoid luminal 17.9 kDa protein, chloroplastic	(RefSeq) thylakoid luminal 15.0 kDa protein 2, chloroplastic isoform X1	(RefSeq) thylakoid luminal 15.0 kDa protein 2, chloroplastic isoform X2	(RefSeq) plastid thylakoid ATPase delta chain, plastid protein
16 (RefSeq) thylakoid luminal 16.5 kDa protein, chloroplastic	(RefSeq) thylakoid processing peptidase 1, chloroplastic-like isoform X1	(RefSeq) probable thylakoid processing peptidase 2, chloroplastic-like	(RefSeq) plastid thylakoid ATPase subunit, plastid protein
17 (RefSeq) thylakoid ADP/ATP carrier protein, chloroplastic	(RefSeq) LOW QUALITY PROTEIN: thylakoid luminal 16.5 kDa protein, chloroplastic	(RefSeq) LOW QUALITY PROTEIN: thylakoid membrane protein slr0575	(GenBank) thylakoid lumen protein
18 (RefSeq) thylakoid luminal 19 kDa protein, chloroplastic	(RefSeq) thylakoid luminal 29 kDa protein, chloroplastic-like isoform X1	(RefSeq) thylakoid-bound ascorbate peroxidase 6	(GenBank) thylakoid-associated protein
19 (RefSeq) thylakoid processing peptidase 1, chloroplastic	(RefSeq) thylakoid membrane protein ssl2009-like	(RefSeq) probable thylakoid lumen protein sl1769	(GenBank) thylakoid-like protein
20 (RefSeq) probable thylakoid processing peptidase 2, chloroplastic	(RefSeq) thylakoid luminal 17.9 kDa protein, chloroplastic isoform X1	(RefSeq) LOW QUALITY PROTEIN: thylakoid luminal protein TL20.3, chloroplastic	(GenBank) thylakoid processing peptidase. Serine peptidase. MEROPS family S26A
21 (RefSeq) thylakoid luminal protein TL20.3, chloroplastic isoform X1	(RefSeq) thylakoid membrane protein slr0575 isoform X1	(RefSeq) thylakoid ADP/ATP carrier protein, chloroplastic-like isoform X1	(GenBank) plasma membrane protein essential for thylakoid formation
22 (RefSeq) thylakoid luminal 15.0 kDa protein 2, chloroplastic	(RefSeq) thylakoid luminal 17.9 kDa protein, chloroplastic isoform X2	(RAP-DB) Similar to thylakoid luminal 17.9 kDa protein, chloroplast.	(GenBank) Putative thylakoid membrane protein, contains 8 pentapeptide repeats
23 (RefSeq) thylakoid luminal 17.4 kDa protein, chloroplastic	(RefSeq) thylakoid luminal 17.4 kDa protein, chloroplastic-like isoform X3	(RAP-DB) Similar to thylakoid-bound ascorbate peroxidase APx8.	(GenBank) thylakoid membrane protein
24 (RefSeq) thylakoid membrane protein TERC, chloroplastic	(RefSeq) thylakoid luminal 15 kDa protein 1, chloroplastic-like	(RefSeq) LOW QUALITY PROTEIN: thylakoid luminal 17.9 kDa protein, chloroplastic	(GenBank) thylakoid-associated single-stranded DNA-binding protein
25 (RefSeq) LOW QUALITY PROTEIN: thylakoid luminal 29 kDa protein, chloroplastic	(RefSeq) thylakoid luminal 17.4 kDa protein, chloroplastic-like isoform X1	(RefSeq) thylakoid luminal 29 kDa protein, chloroplastic isoform X4	(GenBank) thylakoid luminal 15 kDa protein 1, chloroplastic
26 (RefSeq) thylakoid luminal 15 kDa protein 1, chloroplastic isoform X1	(RefSeq) thylakoid membrane protein slr0575-like isoform X1	(RefSeq) thylakoid luminal 17.9 kDa protein, chloroplastic-like isoform X1	(GenBank) Putative thylakoid membrane protein
27 (RefSeq) thylakoid luminal 17.4 kDa protein, chloroplastic isoform X1	(RefSeq) thylakoid luminal 16.5 kDa protein, chloroplastic isoform X1	(RefSeq) thylakoid assembly 8	(GenBank) ring forming protein essential for thylakoid formation Vipp1
28 (RefSeq) thylakoid luminal 29 kDa protein, chloroplastic	(RefSeq) thylakoid luminal 16.5 kDa protein, chloroplastic-like isoform X1	(RefSeq) chloroplast oxygen-evolving complex/thylakoid luminal 25.6kDa protein	(GenBank) putative thylakoid membrane protein
29 (RefSeq) thylakoid luminal 15 kDa protein 1, chloroplastic	(RefSeq) thylakoid ADP/ATP carrier protein, chloroplastic isoform X1	(RefSeq) thylakoid luminal 19 kDa protein	(GenBank) thylakoid signal peptidase I LepB1
30 (RefSeq) thylakoid processing peptidase 1, chloroplastic-like	(RefSeq) LOW QUALITY PROTEIN: thylakoid luminal 19 kDa protein, chloroplastic	(RefSeq) LOW QUALITY PROTEIN: probable thylakoid processing peptidase 2, chloroplastic	(GenBank) high-affinity Na ⁺ (Li ⁺)+H ⁺ thylakoid membrane antiporter NhaS3
31 (RefSeq) thylakoid luminal 17.9 kDa protein, chloroplastic-like	(RefSeq) thylakoid luminal 19 kDa protein, chloroplastic isoform X1	(RefSeq) LOW QUALITY PROTEIN: thylakoid processing peptidase 1, chloroplastic	(GenBank) thylakoid membrane photosystem I accumulation factor
32 (RefSeq) thylakoid luminal 29 kDa protein, chloroplastic-like	(GenBank) thylakoid luminal 15.0 kDa protein	(RefSeq) photosystem II thylakoid luminal 29.8 kDa protein PsbP	(GenBank) putative thylakoid lumen protein
33 (RefSeq) thylakoid luminal protein TL20.3, chloroplastic-like isoform X1	(GenBank) thylakoid lumen 18.3 kDa protein	(RefSeq) thylakoid membrane protein	(GenBank) biogenesis of thylakoid protein
34 (RefSeq) thylakoid luminal 17.4 kDa protein, chloroplastic-like	(GenBank) thylakoid luminal 19 kDa protein	(RefSeq) thylakoid luminal protein	(GenBank) biogenesis of thylakoid protein A
35 (RefSeq) probable thylakoid processing peptidase 2, chloroplastic isoform X1	(GenBank) thylakoid soluble phosphoprotein TSP9 protein	(RefSeq) thylakoid lumen protein	(RefSeq) Paramecium bursaria Chlorella virus NYs1; thylakoid formation protein
36 (RefSeq) thylakoid membrane protein TERC, chloroplastic-like	(GenBank) thylakoid luminal 15 kDa protein	(RefSeq) thylakoid formation protein 1	(RefSeq) thylakoid ADP/ATP carrier protein, chloroplastic-like
37 (RefSeq) thylakoid luminal 19 kDa protein, chloroplastic-like	(GenBank) thylakoid-bound ascorbate peroxidase	(RefSeq) putative thylakoid luminal 29.8 kDa protein	(RefSeq) cold regulated 314 thylakoid membrane 2
38 (RefSeq) thylakoid luminal protein TL20.3, chloroplastic	(GenBank) thylakoid luminal 29.8 kDa protein	(RefSeq) Tat family transporter: pH-dependent thylakoid protein export into thylakoid	(RefSeq) thylakoid lumen 18.3 kDa protein

Molecular properties are abundant in dimensional reduction for collections of sequences. Examples such as stability, binding potential aliphatic and hydrophobicity along with the charge at different pH is a few of available molecular properties to examine based on the gene ontology ids in Table 2. The net charge of a protein sequence based on the Henderson-Hasselbalch equation based on pH 5, 7 and 9. The aliphatic index is the relative volume occupied by aliphatic side chains (Alanine, Valine, Isoleucine, and Leucine) and is a positive factor for the increase of thermostability of globular proteins. The potential protein interaction index proposed by Boman (2003) based in the amino acid sequence of a protein and provides an overall estimate of the potential of a peptide to bind to membranes or other proteins as receptors. A protein have high binding potential if the index value is higher than 2.48. This index predicts the stability of a protein based on its amino acid composition, a protein whose instability index is smaller than 40 is predicted as stable, a value above 40 predicts that the protein may be unstable. Hydrophobicity is an important stabilization force in protein folding; this force changes depending on the solvent in which the protein is found. [1001]

Table 3 has the molecular properties for *Rosa chinensis* (China rose) for each of the thylakoid proteins.

	Species	Protein	Stability Index	Binding Potential	ALiphatc	I.1	CpH5	CpH7	CpH9
498	<i>Rosa chinensis</i> (China rose)	(RefSeq) probable thylakoid processing peptidase 2, chloroplastic	50.229	1.2327	94.0	-0.012772	8.44e+00	1.81158	-4.55458
499	<i>Rosa chinensis</i> (China rose)	(RefSeq) thylakoid luminal 15 kDa protein 1, chloroplastic	19.088	1.3862	84.4	-0.128169	8.62e-01	-2.65045	-7.21622
500	<i>Rosa chinensis</i> (China rose)	(RefSeq) thylakoid luminal 29 kDa protein, chloroplastic	35.773	1.7112	84.1	-0.344092	3.05e+00	-1.08483	-5.24162
501	<i>Rosa chinensis</i> (China rose)	(RefSeq) thylakoid luminal 29 kDa protein, chloroplastic isoform X1	35.506	1.7827	81.4	-0.359574	3.63e+00	0.90610	-2.61588
502	<i>Rosa chinensis</i> (China rose)	(RefSeq) thylakoid luminal protein TL20.3, chloroplastic isoform X1	48.416	1.6834	76.1	-0.234743	1.04e+01	3.19342	-5.46992
503	<i>Rosa chinensis</i> (China rose)	(RefSeq) thylakoid luminal 17.9 kDa protein, chloroplastic	49.541	1.3600	81.1	-0.307627	4.99e+00	0.62118	-3.57338
504	<i>Rosa chinensis</i> (China rose)	(RefSeq) thylakoid luminal 17.4 kDa protein, chloroplastic	34.418	1.4846	86.8	-0.083691	6.19e+00	3.05174	-2.71363
505	<i>Rosa chinensis</i> (China rose)	(RefSeq) thylakoid luminal 16.5 kDa protein, chloroplastic	41.827	1.5464	92.1	-0.100935	5.13e+00	0.65678	-2.49249
506	<i>Rosa chinensis</i> (China rose)	(RefSeq) thylakoid membrane protein slr0575	32.448	1.3082	87.4	-0.128058	1.42e+01	5.82286	0.10253
507	<i>Rosa chinensis</i> (China rose)	(RefSeq) thylakoid membrane protein TERC, chloroplastic	20.010	0.1643	122.7	0.732061	9.69e-01	-0.84109	-3.84147
508	<i>Rosa chinensis</i> (China rose)	(RefSeq) thylakoid luminal 15.0 kDa protein 2, chloroplastic	39.004	1.1985	87.4	-0.082063	9.77e-01	-2.61777	-6.39816
509	<i>Rosa chinensis</i> (China rose)	(RefSeq) LOW QUALITY PROTEIN: thylakoid luminal 19 kDa protein, chloroplastic	29.919	1.5465	68.3	-0.512917	4.18e+00	-1.84206	-5.77130
510	<i>Rosa chinensis</i> (China rose)	(RefSeq) thylakoid membrane protein TERC, chloroplastic isoform X1	32.997	0.7072	104.4	0.390160	5.96e+00	-1.68130	-7.71025
511	<i>Rosa chinensis</i> (China rose)	(RefSeq) LOW QUALITY PROTEIN: thylakoid luminal 29 kDa protein, chloroplastic	31.883	1.6064	80.8	-0.402190	-2.59e+00	-6.09156	-10.20765
512	<i>Rosa chinensis</i> (China rose)	(RefSeq) probable thylakoid processing peptidase 2, chloroplastic	30.410	0.9423	91.5	-0.017391	-2.64e+00	-7.14469	-11.51717

Table 4 has the molecular properties for *Rosa chinensis* (China rose) for each of the thylakoid proteins.

	Species	Protein	Stability Index	Binding Potential	ALipatic	f.1	CpH5	CpH7	CpH9
1220	<i>Oryza sativa japonica</i> (Japanese rice) (RefSeq)	(RefSeq) thylakoid ADP/ATP carrier protein, chloroplastic	39.808	1.0101	90.9	0.000262	2.19e+01	15.07598	9.08661
1221	<i>Oryza sativa japonica</i> (Japanese rice) (RefSeq)	(RefSeq) thylakoid processing peptidase 1, chloroplastic	65.680	1.3368	75.5	-0.143552	1.39e+01	5.52554	-1.70164
1222	<i>Oryza sativa japonica</i> (Japanese rice) (RefSeq)	(RefSeq) thylakoid luminal 15 kDa protein 1, chloroplastic	15.641	1.2523	84.4	-0.075253	3.42e+00	0.40663	-2.62287
1223	<i>Oryza sativa japonica</i> (Japanese rice) (RefSeq)	(RefSeq) thylakoid luminal 17.4 kDa protein, chloroplastic	39.262	1.3700	79.9	0.010169	3.15e+00	0.77958	-5.50364
1224	<i>Oryza sativa japonica</i> (Japanese rice) (RefSeq)	(RefSeq) thylakoid luminal protein TL20.3, chloroplastic isoform X1	38.770	1.6883	71.4	-0.209386	1.01e+01	6.25959	-0.54193
1225	<i>Oryza sativa japonica</i> (Japanese rice) (RefSeq)	(RefSeq) thylakoid luminal 17.9 kDa protein, chloroplastic isoform X1	58.846	1.5643	73.2	-0.208494	9.69e+00	5.49351	-1.75768
1226	<i>Oryza sativa japonica</i> (Japanese rice) (RefSeq)	(RefSeq) thylakoid luminal 29 kDa protein, chloroplastic	44.285	1.6712	84.1	-0.293768	1.08e+01	4.57527	-1.53218
1227	<i>Oryza sativa japonica</i> (Japanese rice) (RefSeq)	(RefSeq) thylakoid membrane protein TERC, chloroplastic	49.331	0.9880	101.1	0.249344	4.69e+00	-0.57573	-4.35237
1228	<i>Oryza sativa japonica</i> (Japanese rice) (RefSeq)	(RefSeq) thylakoid luminal 16.5 kDa protein, chloroplastic	45.977	1.3544	89.6	-0.027615	7.13e+00	3.38410	-0.38791
1229	<i>Oryza sativa japonica</i> (Japanese rice) (RefSeq)	(RefSeq) thylakoid luminal 15.0 kDa protein 2, chloroplastic	36.908	1.0797	86.5	0.099593	4.12e+00	1.02092	-5.41130
1230	<i>Oryza sativa japonica</i> (Japanese rice) (RefSeq)	(RefSeq) thylakoid membrane protein slr0575	54.038	1.1617	87.2	0.003690	9.50e+00	5.38260	1.08299
1231	<i>Oryza sativa japonica</i> (Japanese rice) (RefSeq)	(RefSeq) thylakoid luminal 19 kDa protein, chloroplastic	40.832	1.2830	80.9	-0.140329	5.29e+00	0.68164	-2.43518
1232	<i>Oryza sativa japonica</i> (Japanese rice) (RAPDB)	(RAP-DB) Similar to thylakoid luminal 17.9 kDa protein, chloroplast.	58.477	1.5790	72.7	-0.223346	9.69e+00	5.49351	-1.75768
1233	<i>Oryza sativa japonica</i> (Japanese rice) (RAPDB)	(RAP-DB) Similar to thylakoid-bound ascorbate peroxidase APx8.	47.748	1.5927	68.7	-0.565635	-1.00e+01	-17.29263	-20.83042

For the 1597 proteins for all species the descriptive statistics are in Table 5.

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
Stability Index	1	1597.00	39.32	10.17	39.81	39.74	8.98	0.03	73.37	73.34	-0.39	0.49	0.25
Binding Potential	2	1597.00	1.38	0.33	1.41	1.40	0.28	-0.57	2.45	3.02	-0.62	2.04	0.01
ALipatic	3	1597.00	86.09	10.34	84.56	85.19	8.03	45.00	143.27	98.27	0.92	2.50	0.26
f.1	4	1597.00	-0.12	0.21	-0.13	-0.14	0.16	-0.98	1.11	2.09	0.90	2.60	0.01
CpH5	5	1597.00	6.54	5.32	5.68	6.07	3.61	-13.85	39.44	53.29	1.21	4.11	0.13
CpH7	6	1597.00	1.59	4.69	1.09	1.26	3.22	-22.61	28.43	51.04	0.94	4.20	0.12
CpH9	7	1597.00	-3.63	4.63	-3.79	-3.88	3.28	-27.43	16.63	44.07	0.69	3.75	0.12

[illegible]

4 Conclusions

In this brief mathematical note, the chemical structure of NADPH was examined with respect to the pathways of map00195 Photosynthesis, map00480 Glutathione metabolism, map01100 Metabolic pathways, map04918 Thyroid hormone synthesis, map04925 Aldosterone synthesis and secretion, map04936 Alcoholic liver disease map05208 Chemical carcinogenesis - reactive oxygen species map05415 Diabetic cardiomyopathy with a focus on Photosynthesis, Chemical carcinogenesis - reactive oxygen species and Thylakoid Protein Networks.

5 Bibliography

- [1] Wikipedia contributors. "Thylakoid." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 19 Nov. 2021. Web. 6 Jan. 2022.
- [2] Kotera, M., Tabei, Y., Yamanishi, Y. et al. KCF-S: KEGG Chemical Function and Substructure for improved interpretability and prediction in chemical bioinformatics. BMC Syst Biol 7, S2 (2013). <https://doi.org/10.1186/1752-0509-7-S6-S2>
- [3] Wikipedia contributors. "Steric effects." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 27 Oct. 2021. Web. 6 Jan. 2022.
- [400] Kanehisa, Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K.; KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res. 45. D353-D361 (2017).