



HAL
open science

ChloroKB, a cell metabolism reconstruction of the model plant *Arabidopsis thaliana*

Pauline Gloaguen, Yves Vandenbrouck, Jacques Joyard, Gilles Curien

► To cite this version:

Pauline Gloaguen, Yves Vandenbrouck, Jacques Joyard, Gilles Curien. ChloroKB, a cell metabolism reconstruction of the model plant *Arabidopsis thaliana*. *Comptes Rendus Biologies*, 2021, 344 (2), pp.157-163. 10.5802/crbio1.49 . hal-03282179

HAL Id: hal-03282179

<https://hal.science/hal-03282179v1>

Submitted on 19 Jul 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **News and views / Knowledge base**

2 **ChloroKB, a cell metabolism reconstruction of the model plant *Arabidopsis thaliana***

3 **ChloroKB, une reconstruction du métabolisme cellulaire de la plante modèle *Arabidopsis thaliana***

4 **Authors :** ¹Gloagen Pauline, ²Vandenbrouck Yves, ¹Joyard Jacques and ¹Curien Gilles

5 ¹ Université Grenoble Alpes, CNRS, CEA, INRAe, IRIG-LPCV, UMR5168, 38000 Grenoble, France.

6 ² Université Grenoble Alpes, Inserm, CEA, IRIG-BGE, U1038, 38000, Grenoble, France.

7 Pauline Gloagen ; Email : pauline.gloagen@cea.fr

8 Yves Vandenbrouck ; Email : yves.vandenbrouck@cea.fr

9 Jacques Joyard ; Email : jacques.joyard@univ-grenoble-alpes.fr ; Phone : 33 6 20 64 28 21

10 Gilles Curien ; Email : gilles.curien@cea.fr ; Phone : 33 4 38 78 35 26 or 33 4 38 78 25 09

11

12 *English version*

13 Will we one day be able to model plant metabolism as we can model the flight of an aircraft? Beyond
14 the obvious advantage of large-scale modelling of plant metabolism for the optimised production of
15 molecules of interest, for example, having an integrated predictive model at our disposal would allow
16 us to totally change the way we can validate our knowledge. A validated model would mean that we
17 really understood how plant cell metabolism works. It would then be possible to analyse the impact
18 of disturbances in environmental conditions on cellular functioning, an essential preliminary step for
19 a detailed understanding of how plants live in their environment. The task seems immense because
20 plant cell metabolism is certainly the most complex eukaryotic metabolism to understand given the
21 nature of the energy source used (light) and the existence of an additional compartment that uses
22 and controls this energy (the chloroplast). Endocytosis of a photosynthetic microorganism a few
23 billion years ago by a heterotrophic eukaryotic host not only added a subcellular compartment but
24 also completely reorganised the metabolic flows within the host cell and the transport reactions
25 between this compartment and the other cell compartments [1]. It is therefore not possible to use
26 bacterial or yeast models for plant metabolism modelling. The situation is made even more complex
27 by the fact that higher plants genome encodes the metabolism of about ten different cell types with
28 quite different metabolisms in leaves, roots, meristems, phloem, pollen, fruits, seeds, etc. Upstream
29 of predictive modelling, it is therefore necessary to collect experimental data, reference them,
30 compare them and transform this knowledge into an "entity" that can be analysed by both humans
31 and computers.

32 This work has been carried out for about ten years now for the model plant *Arabidopsis* and is
33 available online (<http://chloroKB.fr>). Initially ChloroKB (for Chloroplast Knowledge Base) focused on
34 chloroplast metabolism [2]: it is thus possible to explore the biosynthesis pathways for amino acids
35 (see [L branched Chain Aminoacids](#)), lipids (see [Galactolipid synthesis](#); [3]), or pigments
36 (chlorophylls, carotenoids [4], etc.; e.g. see [Chlorophyll synthesis](#)), their subcellular location (within
37 the thylakoids, stroma and envelope membranes) and their interactions with other cell
38 compartments (Figure 1) [5].

39 With recent updates, the reconstructed network now covers a very large part of *Arabidopsis*
40 mesophyll cell metabolism with nearly 1600 proteins and more than 1000 metabolites. This network
41 is presented in the form of interconnected metabolic maps (currently 189) showing subcellular
42 compartments, metabolites, proteins, complexes, reactions and transport. Each "object" (protein,
43 transporter, complex) is clickable, allowing the user to access the data available online but scattered
44 throughout the literature. Currently this reconstructed network is a tool for exploring metabolism
45 which very quickly provides access to integrated knowledge of a particular metabolism (e.g., see
46 [Ascorbate glutathion pathway](#)) (Figure 2).

47 ChloroKB differs from other metabolic databases (BioCYC, Uniprot, KEGG) by its high level of data
48 integration, the consideration of subcellular compartmentalisation and thus metabolite transport,
49 the representation of "fine" controls such as complex formation, activity modification by allosteric
50 control or phosphorylation (e.g. see [PEP carboxylase regulation](#)). Sub-cellular localisation has been
51 improved compared to the version initially published with a coding which now considers the
52 different cellular membranes and inter-membrane spaces (plasma membrane, endoplasmic
53 reticulum, inner and outer membranes of mitochondria and chloroplast). It will thus eventually be
54 possible to model membrane lipid metabolism and intermembrane traffic. The use of CellDesigner
55 [6] as a standard for the representation of biochemical networks facilitates the reading of the
56 network and communication, especially with young researchers, who must in a short period of time

57 acquire knowledge accumulated over more than fifty years. In addition to this visual aspect, which
58 allows ChloroKB to be used as a memory aid, as a source of expert information or as a "google map"
59 type exploration tool, ChloroKB's data have been structured in such a way as to allow for
60 mathematical modelling. Flux balance analysis modelling for example allows prediction of steady-
61 state fluxes distribution at the system level under some assumptions (for example optimization of
62 biomass production or of a specific flux) [7]. This kind of modelling requires a stoichiometry matrix of
63 the network. A stoichiometry matrix captures mathematically the structure of a network and consists
64 in a table with reactions in columns and metabolites in rows containing +1 or -1 or 0 or other values
65 depending on the amount of each metabolites used in each reaction or transport steps of the
66 network. A complete stoichiometry matrix of the network represented in ChloroKB will soon be
67 available online, which can be used as a reference for modelling work dedicated to a particular issue
68 [8], [9], [10]. Finally, a large amount of quantitative data (metabolite concentrations) is available
69 which will be useful to constrain metabolic models of the Flux balance analysis type.

70 *French version*

71 Pourrons-nous un jour modéliser le métabolisme d'une plante comme nous sommes désormais
72 capable de modéliser le vol d'un avion ? Au-delà des avantages évidents d'une modélisation à grande
73 échelle du métabolisme de la plante pour la production optimisée de molécules d'intérêt, par
74 exemple, disposer d'un modèle prédictif intégré permettrait de changer totalement notre manière
75 de valider nos connaissances. Un modèle validé signifierait que nous avons réellement compris
76 comment ce métabolisme fonctionne. Il serait alors envisageable d'analyser la réponse cellulaire aux
77 modifications des conditions environnementales, étape préliminaire essentielle à la compréhension
78 fine de la vie de la plante dans son environnement. La tâche semble immense car le métabolisme de
79 la cellule végétale est certainement le métabolisme eucaryote le plus complexe à appréhender étant
80 donné la nature de la source d'énergie utilisée (la lumière) et l'existence d'un compartiment
81 supplémentaire qui utilise et contrôle cette énergie (le chloroplaste). L'endocytose d'un
82 microorganisme photosynthétique il y a quelques milliards d'années par un hôte eucaryote
83 hétérotrophe a non seulement ajouté un compartiment subcellulaire mais a totalement réorganisé
84 les flux métaboliques au sein de la cellule hôte et les réactions de transports entre ce compartiment
85 et les autres compartiments [1]. Il n'est donc pas possible d'utiliser des modèles de bactéries ou de
86 levures pour modéliser le métabolisme de la plante. La situation est rendue encore plus complexe
87 par le fait que le génome des plantes supérieures code le métabolisme d'une dizaine de types
88 cellulaires différents aux métabolismes très différents (cellules chlorophylliennes, cellules racinaires,
89 cellule de phloème, pollen, cellules méristématiques, etc.). En amont de la modélisation prédictive il
90 est donc nécessaire d'effectuer un travail de collation des données expérimentales, de les référencer,
91 de les confronter et de transformer cette connaissance en une « entité » lisible à la fois par l'être
92 humain et par l'ordinateur.

93 C'est ce travail qui a été effectué depuis maintenant une dizaine d'années pour la plante modèle
94 *Arabidopsis thaliana* et qui est disponible en ligne (<http://chloroKB.fr>). Initialement ChloroKB (pour
95 *Chloroplast Knowledge Base*) était centré sur le métabolisme du chloroplaste [2] : il est ainsi possible
96 (i) d'explorer les fonctions chloroplastiques et en particulier les voies de biosynthèse des acides
97 aminés (voir [L branched Chain Aminoacids](#)), des lipides (voir [Galactolipid synthesis](#) ; [3]), ou des
98 pigments (chlorophylles, caroténoïdes...; e.g. voir [Chlorophyll synthesis](#), [4]) ainsi que (ii) leur
99 localisation subcellulaire (au sein des thylakoïdes, du stroma et de l'enveloppe) et (iii) leurs
100 interactions avec l'ensemble des compartiments cellulaires (Figure 1) [5].

101 Avec ses mises à jour récentes, le réseau reconstruit de la cellule chlorophyllienne d'*Arabidopsis*
102 *thaliana* couvre désormais une très grande partie du métabolisme cellulaire. Il se présente sous la
103 forme de cartes métaboliques interconnectées (actuellement au nombre de 189) sur lesquels
104 figurent les compartiments subcellulaires, les métabolites, les protéines, les complexes, les réactions
105 et les transports. Chaque « objet » (protéine, transporteur, complexe) est cliquable, permettant à
106 l'utilisateur d'accéder aux données certes disponibles en ligne mais dispersées dans la littérature.
107 Actuellement ce réseau reconstruit est un outil d'exploration du métabolisme qui permet très
108 rapidement d'avoir accès à une connaissance intégré d'un métabolisme particulier (e.g. voir
109 [Ascorbate glutathion pathway](#)) (Figure 2).

110 ChloroKB se distingue des autres bases de données métaboliques (BioCYC, Uniprot, KEGG) par son
111 niveau élevé d'intégration des données, la prise en compte de la compartimentation subcellulaire et
112 donc du transport de métabolites, la représentation des contrôles « fins » comme la formation de
113 complexes, de modification d'activité par contrôle allostérique ou par phosphorylation (e.g. voir [PEP](#)
114 [carboxylase regulation](#)). Le niveau de précision des localisations a été amélioré par rapport à la
115 version initialement publiée avec un codage qui prend désormais en compte les différentes
116 membranes et les espaces intermembranaires (membrane plasmique, réticulum endoplasmique,
117 membranes interne et externes de la mitochondrie et du chloroplaste). Il sera ainsi à terme possible
118 de modéliser le métabolisme lipidique membranaire et le trafic intermembranaire. L'utilisation de
119 CellDesigner [6] comme standard de représentation des réseaux biochimiques facilite la lecture du
120 réseau ainsi que la communication, en particulier avec les jeunes chercheurs, qui doivent en peu de
121 temps acquérir une connaissance accumulée pendant plus de cinquante ans. Au-delà de cet aspect
122 visuel qui permet d'utiliser ChloroKB comme aide-mémoire, comme source d'informations
123 expertisées ou comme outil d'exploration de type « google map », les données de ChloroKB ont été
124 structurées de manière à permettre une modélisation mathématique. Une méthode de modélisation
125 appelée Flux balance analysis permet par des méthodes de programmation linéaire de calculer la
126 distribution des flux à l'état stationnaire dans le système métabolique sous certaines hypothèses
127 d'optimisation (de la biomasse par exemple ou d'un flux de métabolite particulier) [7]. Ce type de
128 modélisation requière une abstraction du réseau qui se présente sous la forme d'une matrice de
129 stoichiométrie (un tableau contenant les réactions en colonnes et les métabolites en ligne).
130 L'occurrence et la stoichiométrie des réactions et des transports est représenté par des 0, +1 or -1,
131 +2 etc. Une matrice de stoichiométrie complète du réseau reconstruit sera d'ici peu disponible en
132 ligne sur le site de ChloroKB et pourra servir de référence pour des travaux de modélisation dédiés à
133 une question particulière [8],[9],[10]. Enfin de très nombreuses données expérimentales
134 quantitatives (concentrations de métabolites) sont accessibles et seront utiles pour contraindre les
135 modèles métaboliques de type « Flux balance analysis ».

136 Références

- 137 • [1]. Selosse, M.-A. & Joyard, J. (2021), Symbiose et évolution : à l'origine de la cellule
138 eucaryote, Encyclopédie de l'Environnement, [en ligne ISSN 2555-0950] url :
139 [https://www.encyclopedie-environnement.org/vivant/symbiose-evolution-lorigine-de-](https://www.encyclopedie-environnement.org/vivant/symbiose-evolution-lorigine-de-cellule-eucaryote/)
140 [cellule-eucaryote/](https://www.encyclopedie-environnement.org/vivant/symbiose-evolution-lorigine-de-cellule-eucaryote/).
- 141 • [2]. Gloaguen, P., Bournais, S., Alban, C., Ravanel, S., Seigneurin-Berny, D., Matringe, M.,
142 Tardif, M., Kuntz, M., Ferro, M., Bruley, C., Rolland, N., Vandenbrouck, Y. & Curien, G. (2017)
143 ChloroKB: A Web application for the integration of knowledge related to chloroplast
144 metabolic network, *Plant Physiol.* 174, 922-934.
- 145 • [3]. Joyard J., Ferro M., Masselon C., Seigneurin-Berny D., Salvi D., Garin J. & Rolland N.
146 (2009) Chloroplast proteomics and the compartmentation of plastidial isoprenoid
147 biosynthetic pathways. *Mol. Plant.* 2, 1154-1180

- 148 • [4]. Joyard J., Ferro M., Masselon C., Seigneurin-Berny D., Salvi D., Garin J. & Rolland N.
149 (2010) Chloroplast proteomics highlights the subcellular compartmentation of lipid
150 metabolism. *Prog. Lipid Res.* 49, 128-1385.
- 151 • [5]. Rolland N., Curien G., Finazzi G., Kuntz M., Maréchal E., Matringe M., Ravanel S.,
152 Seigneurin-Berny D. (2012) The biosynthetic capacities of the plastids and integration
153 between cytoplasmic and chloroplast processes. *Ann. Rev.Genetics* 46, 233-264
- 154 • [6]. Funahashi, A., Matsuoka, Y., Jouraku, A., Morohashi, M., Kikuchi, N. & Kitano, H. (2008)
155 CellDesigner 3.5: A versatile modeling tool for biochemical networks, *Proc IEEE.* 96, 1254-
156 1265.
- 157 • [7]. Sweetlove, L. J., Ratcliffe, R. G. (2011) Flux-balance modeling of plant metabolism,
158 *Frontiers in Plant Science*, 2, article 38
- 159 • [8]. Arnold, A. & Nikoloski, Z. (2014) Bottom-up Metabolic Reconstruction of Arabidopsis and
160 Its application to determining the metabolic costs of enzyme production, *Plant Physiol.* 165,
161 1380-1391.
- 162 • [9]. Shameer, S., Ratcliffe, R. G. & Sweetlove, L. J. (2019) Leaf energy balance requires
163 mitochondrial respiration and export of chloroplast NADPH in the light, *Plant Physiol.* 180,
164 1947-1961.
- 165 • [10]. Blatke, M. A. & Brautigam, A. (2019) Evolution of C4 photosynthesis predicted by
166 constraint-based modelling, *Elife.* 8.

167 **Figures captions**

168 **English**

169 **Figure 1.** Exploration of carotenoid metabolism in chloroplasts. The different colours correspond to
170 experimental evidence for protein localisation in the envelope (yellow), or the thylakoid (green) or
171 absence of evidence (red). Demonstrated reactions are coded in black and hypothetical reactions in
172 grey. The violin icons allow navigation from map to map. Zooming option in the online version
173 facilitates the understanding.

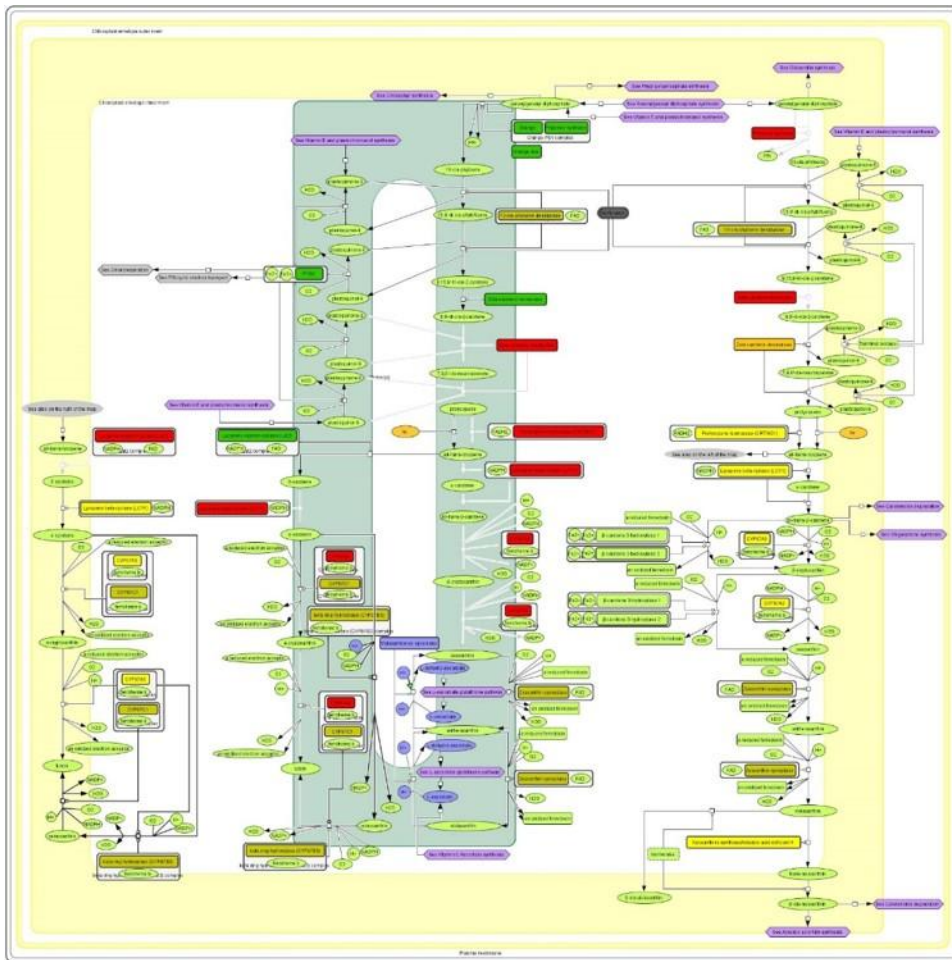
174 **Figure 2.** Exploration of Ascorbate glutathion pathway in photosynthetic *Arabidopsis* cells. A
175 complete overview of this metabolism is provided across 6 subcellular compartments, namely
176 thylakoid (green) plastid (yellow), peroxisome (blue), mitochondrion intermembrane space and
177 matrix (red) and cytosol (white). Apoplasic L-ascorbate metabolism is represented in another map
178 (top violin icon).

179 **Français**

180 **Figure 1.** Exploration du métabolisme des caroténoïdes dans les chloroplastes. Les différentes
181 couleurs correspondent aux preuves expérimentales de la localisation des protéines dans l'enveloppe
182 (jaune), ou le thylakoïde (vert) ou à l'absence de preuves (rouge). Les réactions démontrées sont
183 codées en noir et les réactions hypothétiques en gris. Les icônes violettes permettent de naviguer de
184 carte en carte. L'option de zoom dans la version en ligne facilite la compréhension.

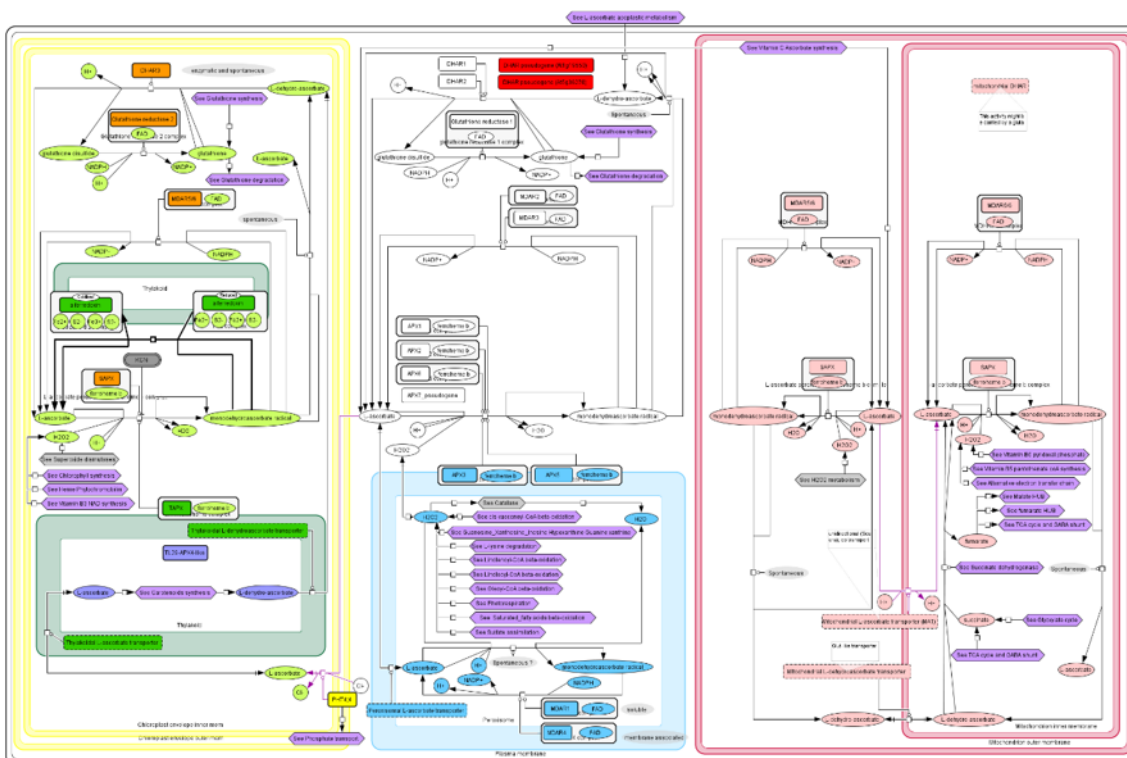
185 **Figure 2.** Exploration de la voie Ascorbate glutathion dans les cellules photosynthétiques
186 d'*Arabidopsis*. Un aperçu complet de ce métabolisme est fourni à travers 6 compartiments
187 subcellulaires, à savoir les thylakoïdes (vert), le plaste (jaune), le peroxysome (bleu), l'espace
188 intermembranaire et la matrice de la mitochondrie (rouge) et le cytosol (blanc). Le métabolisme
189 apoplastique du L-ascorbate est représenté sur une autre carte (icône du violon en haut).

191 Figure 1



192

193 Figure 2



194