Lehrstuhl für Organische Chemie und Biochemie

Technischen Universität München

Computational Approaches for Metabolic Flux Analysis in ^{13}C Perturbation Experiments

Christian Ettenhuber

Vollständiger Abdruck der von der Fakultät für Chemie der Technischen
Universität München zur Erlangung des akademischen Grades eines
Doktors der Naturwissenschaften genehmigten Dissertation.

Vorsitzender:
Univ.-Prof. Dr. Steffen J. Glaser

Prüfer der Dissertation:
1. Univ.-Prof. Dr. Dr. Adelbert Bacher
2. Univ.-Prof. Dr. Alfons Gierl
3. Univ.-Prof. Dr. Sevil Weinkauf

Die Dissertation wurde am 04.01.2005 bei der Technischen Universität München
eingereicht und durch die Fakultät für Chemie am 03.02.2005 angenommen.

Technology and results presented in this thesis are published in excerpts:

Ettenhuber, C., Radykewicz, T., Kofer, W., Koop, H. U., Bacher A., Eisenreich W., *Metabolic flux analysis in complex isotopolog space. Recycling of glucose in tobacco plants*, Phytochemistry, 2005, **66**, 323-335

Eisenreich, W., Ettenhuber, C., Laupitz, R., Theus, C., Bacher, A., *Isotopolog perturbation techniques for metabolic networks: Metabolic recycling of nutritional glucose in Drosophila melanogaster*, Proc. Nat. Acad. Sci. USA, 2004, **101**, 6764-6769

Mühlberger, R., Robelek, R., Eisenreich, W., Ettenhuber, C., Sinner, E. K., Kessler, H., Bacher, A., Richter, G., *RNA DNA discrimination by the antitermination protein NusB*, J. Mol. Biol., 2003, **327**, 973-83

Ihr kennt nur des Geistes Funken:
aber ihr seht den Ambos nicht, der er ist,
und nicht die Grausamkeit seines Hammers![1]

(F. Nietzsche, Von den berühmten Weisen, in:
Also sprach Zarathustra, 1896)

---

[1] English translation: Ye know only the sparks of the spirit: but ye do not see the anvil which it is, and the cruelty of its hammer!

Abstract

Post-genomic experimental and computational analysis of living systems centers on metabolism in trying to create complete models of cellular systems *in silico*. The complex interplay of metabolic networks can be elucidated *in vivo* by retrobiosynthetic techniques using stable-isotope tracers, NMR spectroscopy, and mass spectroscopy. The perturbation of the ^{13}C quasi-equilibrium in biochemical networks of cellular systems affords an artificial ^{13}C distribution shaped by distinct metabolic processes. Formal description of metabolism as scale-free network has initiated the development of novel computational techniques for modeling and simulating biological systems.

Retrobiosynthetic analysis of secondary metabolic products and amino acids by high-resolution ^{13}C NMR spectroscopy is introduced by specification of a general procedure to reconstruct and analyze primary and secondary pathways in plant cell cultures. This experimental approach is extended by utilization of glucose as universal analyte affording detailed information on the *in vivo* carbohydrate metabolism in biological systems. [U-$^{13}C_6$]glucose serves as tracer in either experimental setup. High dilution of the tracer with glucose at natural ^{13}C abundance affords a 22-dimensional glucose isotopolog space. In this case, local information on distinct regions of glucose molecules obtained by ^{13}C NMR spectroscopy is computationally deconvoluted to unequivocal glucose isotopolog signatures. The resulting abundances of glucose isotopologs enable the assessment of catabolic carbon fluxes and to formulate a description of the spatial and temporal metabolic integral of the intermediary network. The glucose isotopologs are simulated *in silico* by a rule-based approach to reconstruct the metabolic events in the experimental system. The topology of the simulated network includes the complete set of dominant catabolic processes in metabolic-active biological systems (glycolysis/glucogenesis, transketolase and transaldolase reaction of the pentose phosphate pathway, citric acid cycle). Inference of metabolic rules elucidates modulations in the global regulation of carbohydrate fluxes.

The deconvolution of the complete 64-dimensional glucose isotopolog space requires integration of several biochemical and analytical techniques. The glucose isotopologs can be populated using $^{13}CO_2$ as tracer in photosynthetically active plants. Their abundance represents a metabolic integral of anabolic and catabolic carbon fluxes in the plant tissues during the experimental time period. A sufficiently determined equation system for the deconvolution procedure is obtained by combining high-resolution ^{13}C NMR data of glucose and 6,7-dimethyl-8-ribityllumazine generated by enzymatical conversion of this glucose, GC/MS and IRMS.

The basic retrobiosynthetic techniques are exemplified by elucidation of the metabolic origin of the benzofuran derivates isoeuparin and 4-hydroxytremetone in *Tagetes patula* root cell culture as well as by reconstruction of the intermediary metabolic network in *Taxus chinensis* cell culture producing the anti-cancer drug taxol. Both experiments were evaluated on the basis of ^{13}C-labeling patterns in amino acids.

Glucose as universal analyte for assessing carbohydrate metabolism is used for elucidation of the intermediary metabolic networks of animals and plants. The following examples are provided in this thesis: (I). steady-state fluxes and the influence of endosymbionts in *Drosophila melanogaster*, (II). steady-state fluxes and the photosynthetic refixation process of $^{13}CO_2$ in *Nicotiana tabacum*, (III). metabolic changes during starch accumulation in developing maize kernels, (III). insights on the metabolic origin of phytoglycogen in maize and deduction of a new spatial model which separates of the biosynthesis of starch and phytoglycogen, (IV). metabolic changes in maize due to hybrid vigor in F_1 generation plants, and (V). the steady-state metabolism of mutant forms of maize.

The consistent integration of the independent and partially redundant experimental data records in the deconvolution procedure for the complete 64-dimensional glucose isotopolog set is exemplified using glucose isolated *Ruta graveolens* as biological model system. The complete set of glucose isotopologs can be attributed to distinct, interlocked carbon fluxes in the plant.

Acknowledgements

The research projects in the scope of this thesis were supported and financed by the Chair of Organic Chemistry and Biochemistry of the Technical University of Munich.

I thank Dr. Wolfgang Eisenreich and Prof. Dr. Dr. Adelbert Bacher for scientific advice. Heidi Hofner, Fritz Wendling, Ralf Laupitz, Tanja Radykewicz, Lilla Margl, Gertrud Spielbauer, Ulrich Genschel, Alexandra Bosshammer, Christopher W. Helsen, Prof. Dr. Sevil Weinkauf, Prof. Dr. Wolfgang Hiller, Prof. Dr. Uwe Baumgarten, Prof. Dr. Ernst W. Mayr, Prof. Dr. Gerald Richter, Prof. Dr. Hans-Ulrich Koop, Prof. Dr. Yehuda Ben-Shaul, and Prof. Dr. Alfons Gierl for their cooperation, experimental expertise and helpful discussions.

Furthermore, I thank Margot Dieden for her affection and continuous support during the time of the doctoral examination procedure and the writing of this thesis.

Abbreviations

ATP	Adenosine triphosphate
COSY	Correlation spectroscopy
DMAPP	Dimethylallyl pyrophosphate
DNA	Deoxyribonucleotide acid
DOXP	1-deoxyxylulose 5-phosphate
DPP	Days post pollination
DRML	6,7-dimethyl-8-ribityllumazine
EMP	Embden-Meyerhoff-Parnas pathway
FID	Free induction decay
GA	Genetic algorithm
GC	Gas chromatography
HMQC	Heteronuclear multiple quantum coherence
HMBC	Heteronuclear multiple bond coherence
HPLC	High pressure liquid chromatography
HSQC	Heteronuclear single-quantum coherence
INADEQUATE	Incredible natural abundance double quantum transition experiment
IPP	Isopentenyl pyrophosphate
IRMS	Isotope ratio mass spectrometry
MS	Mass spectrometry
NAD+/NADH	Nicotineamid adenine dinucleotide (hydrogen)
NADP+/NADPH	Nicotineamid adenine dinucleotide phosphate (hydrogen)
NJ	Neighbour joining
NMR	Nuclear magnetic resonance
NN	Neural network
PPP	Pentose phosphate pathway
RNA	Ribonucleotide acid
RMSD	Root mean square deviation
SP	Shikimate pathway
TCA	Tricarboxylic acid cycle
UPGMA	Unweighted pair group with arithmetic mean
QSAR	Quantitative structure activity relationship
WSP	Water-soluble polyglucan

Content

1. Introduction

Metabolism encloses all chemical processes within a living cell, tissue or organism and can be divided into processes for the synthesis of organic compounds (anabolism) and their breakdown (catabolism). The use of isotopes in order to study metabolism *in vivo* and *ex vivo* has a long tradition, starting in the second quarter of the 20th century (e. g. Nier, 1947 and Urey et al., 1948). Today, these techniques became an essential tool in the fields of biochemistry and biotechnology.

Experimental studies *in vivo* use the propagation of isotope-labeled tracer molecules in the metabolic network to qualitatively and quantitatively assess metabolic pathways. Current techniques to analyze directed flows focus on biotechnological tuning of metabolic pathways to increase yields of commercially interesting compounds (e. g. Peterson et al., 2000; Drysch et al., 2003), the elucidation of biosynthetic pathways (e. g. Schwender et al., 1996; Mahmud et al., 2002), and analyzing the interplay of known metabolic pathways (e. g. Laule et al., 2003; Rodriguez-Concepcion et al., 2004).

The sequencing of whole genomes led to new challenges. Genome sequences of bacteria, protists and higher model organism of the animal (*Drosophila melanogaster*, Adams et al., 2000; *Caenorhabditis elegans*, Ainscough et al., 1998; *Rattus norvegicus*, Gibbs et al., 2004; *Gallus gallus*, Hillier et al., 2004; *Homo sapiens*, Lander et al., 2001 and Venter et al., 2001) and plant kingdom (*Arabidopsis thaliana*, Yamada et al., 2003; *Oryza sativa*, Goff et al., 2002) let us sense the high degree of complexity of metabolic networks existing in single cells, tissues and organisms. Beyond genomics, novel techniques and interdisciplinary approaches between biology, chemistry, physics, computer science and mathematics are required to complete the understanding of processes occurring within an organism. Metabolomics tries to elucidate the interaction of metabolic pathways in a whole organism and therefore to extend the picture shaped by data obtained from the genome and transcriptome of a system.

Stable-isotope labeling techniques and analysis by NMR spectroscopy or mass spectroscopy are among the best choices of analytical approaches for metabolomic studies (Kikuchi et al., 2004; Grivet et a., 2003). Single directed propagation of a tracer in hitherto experimental setups can be analyzed without computational aid, but requiring both, big amounts of time and expert knowledge. Key nodes and sink metabolites in the metabolic network must be set into focus of the analysis. If transformations and modulations of labeling signatures in metabolic key nodes are known, computer simulations of unidirectional, bidirectional and circular flows can help to interpret the interplay and regulation of the metabolic network in a cell, tissue or organism. The creation of software for these simulations is accompanied by a progress of formal description and formal analysis, which is the prerequisite for integrating metabolomic data with existing data from the genome and transcriptome of a living system.

1.1 *In vivo* Metabolic Analysis using NMR spectroscopy

Elements or their isotopes in organic matter (hydrogen, carbon, nitrogen, oxygen, phosphor and sulphur) have paramagnetic nuclei and are intuitively accessible to nuclear magnetic resonance (NMR) spectroscopy. Studying the metabolome of a cellular system, analytic techniques for carbon compounds are required. Stable-isotope perturbation/relaxation techniques shift the quasi-equilibrium of the natural distribution of carbon atoms to an artificial distribution shaped by the biochemical network. Accessing the distributions of ^{13}C atoms in a compound is the most direct and exact method for evaluation the carbon fluxes of a system *in vivo*. Former methods, like ^{14}C-radiolabeling techniques, were only able to show the existence of a tracer qualitatively within a defined compound. NMR spectroscopy not only provides qualitative and quantitative information on paramagnetic nuclei, it also contains spatial information of the environment of a specific nuclei within the molecule (for review, see Eisenreich and Bacher, 2000; Lopaschuk, 1997).

1.1.1 Carbon isotope distributions in organic matter

Carbon exists as ^{12}C, ^{13}C and ^{14}C in naturally occurring matter. The isotopes ^{12}C and ^{13}C are stable with natural abundances of 98.89 % for ^{12}C and 1.11 % for ^{13}C, while ^{14}C has a half-life of 5730 y before β-decay (Paul and Frey, 1991).

The natural abundance of ^{13}C carbon varies less than 0.1 % depending on the origin of the organic compound and is shaped through biophysical and geochemical processes (Schmidt, 2003; Eisenreich and Bacher, 2000; Rossmann et al., 1991). This variance can be neglected for ^{13}C perturbation experiments and the distribution of ^{13}C in an organic molecule can be viewed as being in a state of quasi-equilibrium. Therefore a natural ^{13}C abundance of 1.1 % can be assumed for each carbon position of a molecule (Eisenreich and Bacher, 2000). In a first approximation, multiple-labeled compounds consisting of n ^{13}C atoms occur with an exponentially decreasing abundance equaling $(0.011)^n$, e. g. for $^{13}C_2$ 0.0121 % and for $^{13}C_3$ 0.0001331 % (for details on statistical carbon distributions, see chapter 2.3).

Introducing artificially ^{13}C-enriched compounds in the experimental system, the natural quasi-equilibrium is perturbed and the perturbation is propagated by metabolization of the tracer molecule. Isolation of key compounds of the metabolic network and assessment of the ^{13}C perturbation enables the elucidation of carbon fluxes within the experimental system.

1.1.2 Isotopomers and isotopologs

The isotope-enriched compound can be either an isotopomer or an isotopolog of the original compound. The terms are defined according to the IUPAC nomenclature (IUPAC, 1997):

Isotopologs (also: isotopologues) are molecular entities that differ only in isotopic composition (IUPAC definition I03351; Figure 1.1 A).

Isotopomers have the same number of each isotopic atoms, but differing in their positions. Isotopomers can be either constitutional isotopomers or isotopic stereoisomers (IUPAC definition I03352; Figure 1.1 B).

The following examples demonstrate the definitions: CH_4, CH_3D, CH_2D_2 and CH_1D_3 are different methane isotopologs. CH_2DCH_2OH and CH_3CHDOH are constitutional isotopomers, since they both contain only one deuterium atom. (R)- and (S)-CH_3CHDOH are examples for isotopic stereoisomers. CHD_2CH_2OH and CH_3CHDOH are isotopologs but not isotopomers.

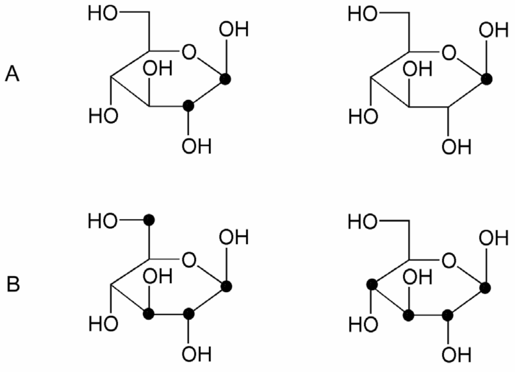

Figure 1.1: Examples of isotopologs and isotopomers of glucose (black circles indicate ^{13}C atoms). A. [1,2-$^{13}C_2$]glucose and [1-$^{13}C_1$]glucose are isotopologs. B. [1,2,3,6-$^{13}C_4$]glucose and [1,2,3,4-$^{13}C_4$]glucose are isotopomers.

1.1.3 Assessment of absolute and relative ^{13}C distributions

For the correct assessment of ^{13}C distributions, an antecedent, correct signal assignment is required to evaluate the ^{13}C NMR experiment. If the complete ^{13}C signal assignment is not available in the literature or in databases (e. g. Bovey, 1967; Saito, 2004), it can be obtained by a comprehensive analysis of ^{1}H^{1}H, ^{13}C^{1}H and ^{13}C^{13}C spin systems. Two-dimensional NMR spectroscopy can provide homocorrelation (COSY, INADEQUATE) and heterocorrelation (HMQC, HMBC) data about the unknown carbon compound (Brevard and Granger, 1981).

The absolute ^{13}C enrichment of a carbon atom *A* in a given position of a compound is obtained by ^{1}H NMR spectroscopy by the ratio of the integral of the ^{1}H^{13}C heteronuclear coupling satellites *S* and the total integral of the hydrogen signal *H*, thus expressed by $A = \frac{S}{S+H}$. Prequesite for this formula is a high relaxation time to avoid the nuclear Overhauser effects. The hydrogen signal should have a non-complex signature, preferably a singlet, and should be located in a non-crowded region of the spectrum. By obtaining the absolute ^{13}C enrichment for one position, all other positions can be specified through cross-calibration (for details, see Eisenreich and Bacher, 2000).

In a ^{13}C NMR spectrum each position in a compound can occur as singlet or as multiplet in case of multiple labels. A signal set can define a specific isotopolog or a group of possible isotopolog species of a compound. If no information from a remote nuclei can be obtained due to limited signal resolution, ^{12}C or ^{13}C have to be assumed for the distant carbon position. This superposition of carbon isotopes generates intrinsically a set of possible isotopologs. In a multiplet, the relative abundances *R* of different isotopolog or isotopolog group signatures of adjacent ^{13}C-labeled carbon atoms can be obtained by calculating the relative ratio of the integrals of a single isotopolog or isotopolog group *L* to the total integrals associated with the signal integral *T* of a specific carbon position. The absolute ^{13}C abundance *B* of a single isotopolog or isotopolog group associated to a specific carbon position can therefore expressed by $B = A \cdot \frac{L}{T}$.

1.1.4 Metabolism forms a scale-free network

Cellular metabolism is best described as graph with a highly dense and complex character. Edges of this graph can be described in terms of function (catalysis) or regulation (substrate-ligand-interaction, e. g. metabolite inhibition or enzyme activation). Nodes of this graph correspond to metabolite concentrations, being either in a steady state between in-going fluxes and out-going fluxes (e. g. pyruvate), or being in a sink state which leads to the accumulation of a distinct metabolite (e. g. chitin).

The extent of the network ranges from hundreds of interactions in the most simple known bacterial species, *Mycoplasma* spp. (Fraser et al., 1995), to thousands of interactions in *Eukarya* (e. g. Waterston et al., 2002, Fujiyama et al., 2002). A single tracer can be distributed *via* this network in a multitude of different routes. Despite the complexity of these networks, the connections between different nodes tend to be short (Clifton and Roe, 1998; Bray, 2003). Each biochemical pathway in the network forms a tightly packed cluster of associated metabolites and reactions. Thus the connectivity in a cluster is higher than between different clusters: the network is cliquish and easily traversable. Networks of this type are scale-free networks (Bray, 2003; Figure 1.2). Scale-free networks can be hierarchically regulated. Environmental modulations are expressed by small changes in the major fluxes inducing metabolic shifts which are amplified in downstream pathways (Almaas et al., 2004).

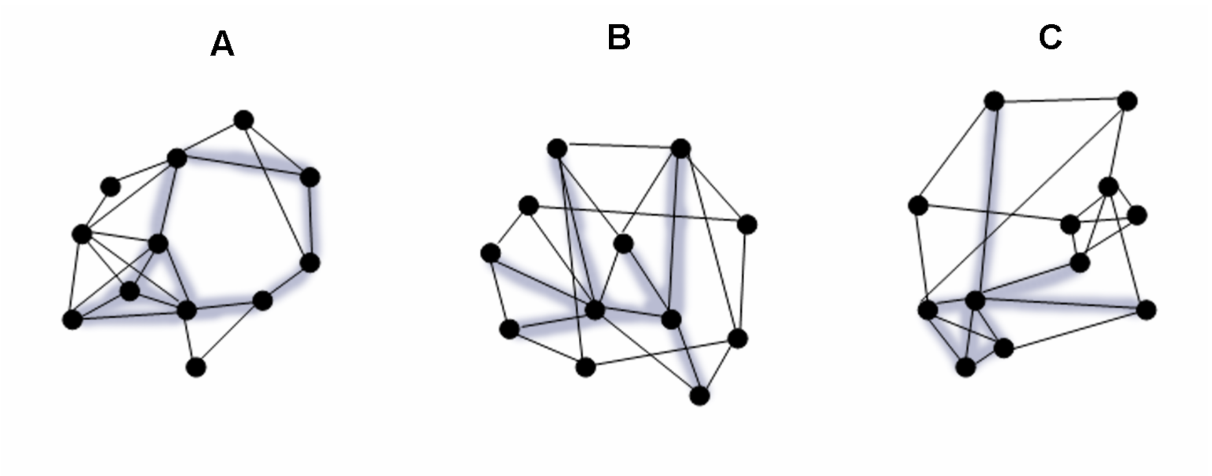

Figure 1.2: Network nodes can be connected in different ways (modified from Bray, 2003). A. Regular networks with short connections between nodes form dense clusters, but the number of steps between clusters tends to be high. B. Random networks form no clusters, but can be traversed on short routes. C. Scale-free networks form dense clusters, but many nodes also possess long-distance connections.

1.2 *In silico* modeling of complex metabolic networks

Living systems are characterized by autopoiesis and emergence (Meng, 2004). Autopoiesis describes the cumulative properties of a system needed for autocatalytic reproduction and maintaining an internal homeostasis. Emergence describes additional abilities of a system, which cannot be reconstructed by cumulating single abilities and functions of the systems components. Both properties make it difficult to apply traditional modeling techniques assuming an exclusively hierarchical control and finite regulation patterns.

Metabolic networks represent a continuous, probabilistic state engine within these systems. The networks are inherently of stochastic character and generate intrinsic noise while operating. Since each natural system is open, additional extrinsic noise is generated from the environment. The theory of the transition probability p between metabolic states X within a time interval Δt was first formulated by Gillespie, 1977, who describes a general metabolic network with M reactions the by relative rates of forward and reverse metabolite fluxes (parameters α and β).

$$p(X, t + \Delta t) = p(X, t) \cdot \left(1 - \sum_{j=1}^{M} \alpha_j \Delta t \right) + p(X, t) \cdot \sum_{j=1}^{M} \beta_j \Delta t$$

Gillespies description of metabolic system was the basis for a whole generation of *in silico* simulations and accompanying biological experiments (cf. Kastner et al., 2002) focusing on single trajectories of metabolites. Due to its low computational performance, the algorithm was subsequently modified and improved. Examples of these adapted algorithms are the Next Reaction Method (Gibson and Bruck, 2000), the Tau-Leap Method (Gillespie, 2001) and Implicit Schemes (Gillespie and Petzold, 2004).

Compartmentation and non-equilibrium conditions cannot be implemented in Gillespies algorithm in a simple way. The Stochsim techniques (Morton-Firth, 1998) overcame these limitations by modeling metabolic networks as graph of interacting nodes connected with probabilistic edges (i. a., Bayesian networks). Nodes with a high number of intrinsic states can be reduced to a single interacting object without performing calculations for all possible states (for experimental applications, cf. Morton-Firth and Bray, 1998; Shimizu, 2002).

Algorithm	**Computational cost**	**Modeling knowledge**	**Speed**	**Accuracy**
Gillespie algorithm	+++	+	-	+++
Gibson algorithm	++	+	++	++
Stochsim	++	+	++	++
Tau-Leap	-	++	+++	+
Spatial-temporal algorithms	+++	+++	- - -	++
Parallel algorithms	-	+	+++	-
Hybrid algorithms	-	++	+++	+

Table 1.1: Comparison of functional-based computational algorithms for modeling complex metabolic networks (modified from Meng, 2004). Evaluation and benchmarking depends on modeling of the system, implementation and computation platform.

Advances in available computer platforms allowed improving the algorithms by exploiting parallel computation for the simulation of multiple nodes (Schwehm, 1996). Furthermore, high-performance computers and computer clusters were able to simulate the spatial and temporal decomposition of metabolic networks *in silico* by modeling biological systems as cellular automata (Stundzia and Lumsden, 1996; Elf et al., 2003).

Despite the advancements in available algorithms and the processor performance, metabolic simulation using stochastic models based on differential equations still requires huge amounts of computing power. Also in many cases the metabolic network becomes oversimplified through applying modeling techniques which deconvolute the graph in plain, linear trajectories (Schmidt et al., 1997; Schmidt et al., 1998; Follstadt and Stephanopoulos, 1998; van Winden et al., 2002).

1.3 Central metabolic pathways

Analysis and modeling of metabolic networks using ^{13}C perturbation studies require detailed understanding of the biochemical interactions and skeletal rearrangements occurring within this network. In the following chapters, the major pathways within the scope of the presented experiments are summarized. The rearrangement of the carbon skeleton of key metabolites and the spatial organization of metabolic pathways is explained.

1.3.1 Glycolysis and glucogenesis

Glycolysis converts glucose to pyruvate. Glucogenesis describes the reverse pathway from pyruvate to glucose. Both pathways are located in the cytosol of cells.

Glycolysis consists of a sequence of ten reactions (for details, see Granner and Pilkis, 1990; Silverman, 1991; evolutional aspects are available in Dandekar et al., 1999): in the first part of the pathway glucose is phosphorylated to glucose 6-phosphate by hexokinase using ATP and isomerized to fructose 6-phosphate by phosphoglucose isomerase. Fructose 6-phosphate is converted to fructose 1,6-bisphosphate by phosphofructokinase using ATP. Aldolase cleaves this compound to dihydroxyacetone phosphate and glyceraldehyde 3-phosphate. Dihydroxyacetone phosphate is isomerized to glyceraldehyde 3-phosphate by triose phosphate isomerase. Glyceraldehyde 3-phosphate becomes oxidized and phosphorylated through glyceraldehyde 3-phosphate dehydrogenase using NAD^+ inorganic phosphate. The resulting 1,3-bisphosphoglycerate transfers phosphate to ADP and becomes 3-phosphoglycerate due to a phosphoryl transfer by the enzyme phosphoglycerate kinase. By action of phosphoglycerate mutase rearrangement 3-phosphoglycerate is converted into 2-phosphoglycerate. Enolase catalyzes of the remaining phosphate and cleavage of water yield phosphoenolpyruvate. Transforming phosphoenolpyruvate to pyruvate catalyzed through pyruvate kinase yields the transfer of another phosphate to ADP yielding ATP.

Glucogenesis requires four additional steps to revert the glycolysis pathway (cf. Granner and Pilkis, 1992). Pyruvate is converted to oxaloacetate by the pyruvate carboxylase by the use of ATP and CO_2. Oxaloacetate is converted to phosphoenolpyruvate by the phosphoenolypyruvate carboxykinase using GTP. The remaining reactions are reversible including to regeneration of fructose 1,6-bisphosphate. Fructose bisphosphatase converts it to fructose 6-phosphate. The reaction of phosphoglucose isomerase is reversible. Glucose 6-phosphate is converted to glucose by glucose 6-phosphatase.

Glycolysis and glucogenesis are reciprocally controlled (cf. Depre et al., 1998). Glycolysis contains three control-points: the enzymes hexokinase, phosphofructokinase and pyruvate kinase. Hexokinase is inhibited by glucose 6-phosphate, which accumulates if the phosphofructokinase is down-regulated. Phosphofructokinase is inhibited by high energy levels (ATP, citrate) and is activated by low energy levels (AMP, frucose 2,6-bisphosphate). The pyruvate kinase is activated by fructose 1,6-bisphosphate and allosterically inhibited by ATP and alanine.

1.3.2 Citrate acid cycle and glyoxylate cycle

The citric acid cycle performs the final oxidation of nutritional glucose and provides building blocks for anabolic biosynthesis. It is located in the mitochondria of eucaryots and in the cytosol of aerobic bacteria. Citric acid cycle and glycolysis are interconnected downstream *via* oxidative decarboxylation of pyruvate leading to acetyl-CoA (Domingo et al., 1999; Izard et al., 1999).

The cycle starts with conjunction of oxaloacetate with acetyl-CoA to citrate catalyzed by the enzyme citrate synthase. The formed citrate is isomerized to isocitrate by the enzyme aconitase. Isocitrate dehydrogenase performs an oxidative decarboxylation of isocitrate to α-ketoglutarate using NAD^+. Again by oxidative decarboxylation, α-ketoglutarate is transformed to succinyl-CoA by α-ketoglutarate dehydrogenase generating again NADH from NAD^+. Coenzyme A is cleaved by introducing phosphate by action of the succinyl-CoA synthetase, which is subsequently used to generate GTP. The resulting succinate is oxidized to fumarate by the enzyme succinate dehydrogenase converting a FAD to $FADH_2$. Succinate is hydratized to malate by fumarase. Malate is finally oxidized to oxaloacetate by malate dehydrogenase generating another NADH (details can be obtained from Velot et al., 1997; Melendez-Hevia et al., 1996; Barnes and Weitzman, 1986).

The citric acid cycle is regulated at three points (cf. Jitrpakdee and Wallace, 1999). Generation of acetyl-CoA by the pyruvate dehydrogenase complex is controlled by reversible phosphorylation. The enzymes isocitrate dehydrogenase and α-ketoglutarate dehydrogenase are inhibited by high levels of ATP. Biosynthetic precursor compounds generated in the cycle form moieties of nucleotides, amino acids, and heme groups.

The glyoxylate cycle is located in the glyoxysomes of cells (for review, see Eastmond and Graham, 2001). It starts by the conversion of oxaloacetate to aspartate and transport of aspartate to the glyoxysomes. Aspartate is reconverted to oxaloacetate and condensed with acetyl-CoA. The subsequent steps as well as further intermediate steps resemble the citric acid cycle. After formation of isocitrate, the enzyme isocitrate lyase cleaves isocitrate to succinate and glyoxylate. Succinate is transported to the mitochondria and oxaloacetate is regenerated. Glyoxylate is condensed with acetyl-CoA to form malate. Malate is exported from the glyoxysomes to the cytosol and is oxidized to oxaloacetate yielding NADH.

1.3.3 Cori cycle and proline cycle

The Cori cycle occurs in higher animals to deliver energy to muscles (for review, see Katz and Tayek, 1999). Glycolysis/glucogenesis would be a futile cycle if limited to one cell, but intercellular cooperation can increase the amount of energy available for rapid muscular movements as required in fast moving animals.

For obtaining ATP fast, the muscle converts glucose to pyruvate (cf. chapter 1.3.1). Pyruvate is converted to lactate to regenerate NAD^+ for glycolysis. Lactate is transported *via* the bloodstream to the liver and is reconverted to pyruvate. Glucogenesis yields new glucose units which are released to the bloodstream.

The proline cycle has an equivalent function as the Cori cycle and occurs in flying insects (Scaraffia and Wells, 2003). It is required to understand the metabolome of *Drosophila melanogaster*, which serves as model organism for intermediary metabolic networks (cf. chapter 3.7.1 and 3.7.7).

Pyruvate is converted to alanine by transamination through glutamate. Alanine is transported *via* the hemolymph to the fat body and is reconverted to pyruvate by reductive transamination using α-ketoglutarate. Pyruvate is converted to oxaloacetate and condenses with acetyl-CoA by catalytic action of the citrate synthase. The first half of the citric acid cycle to α-ketoglutarate is passed and α-ketoglutarate is converted to glutamate as described in the first-step of the reaction sequence in the fat body. Glutamate is converted to proline (cf. chapter 1.3.5). Proline is transported *via* the hemolymph to the flight muscle and is degraded *via* glutamate-γ-semialdehyde to α-ketoglutarate. In the muscle the remaining reaction sequence of the citric acid cycle proceeds. The proline cycle is closed by the regeneration of pyruvate from malate.

1.3.4 Pentose phosphate pathway and Calvin cycle

The pentose phosphate pathway consists of an oxidative and non-oxidative branch (for details on reaction mechanism and evolution, cf. Thorell et al., 2000; Notaro et al., 2000; Martin and Schnarrenberger, 1997). The oxidative branch starts with the conversion of glucose 6-phosphate to 6-phosphoglucono-δ-lactone by the enzyme glucose 6-phosphate dehydrogenase. 6-phosphogluconolactonase catalyses the addition of water yielding 6-phosphogluconate. 6-phosphogluconate is decarboxylated by the enzyme 6-phosphogluconate dehydrogenase using $NADP^+$. The resulting ribulose 5-phosphate can be converted to ribose 5-phosphate or xylulose 5-phosphate by an isomerase and epimerase, respectively. Transketolase reaction between these two C_5-sugars results in sedoheptulose 7-phosphate and glyceraldehyde 3-phosphate. From these compounds, transaldolase yields fructose 6-phosphate and erythrose 4-phosphate. Transketolase action can furthermore generate fructose 6-phosphate and glyceraldehyde 3-phosphate from erythrose 4-phosphate and xylulose 5-phosphate. Fructose 6-phosphate can be reconverted to glucose 6-phosphate and glyceraldehyde 3-phosphate can be feed in the glycolysis pathway.

The Calvin cycle uses ATP and NADPH synthesized in the photosynthesis light reactions to create trioses and hexoses from atmospheric CO_2 (cf. Barber and Andersson, 1994). CO_2 is fixed using ribulose 1,5-bisphosphate through the enzyme ribulose 1,5-bisphosphate carboxylase resulting in two 3-phosphoglycerate units. 3-phosphoglycerate is converted analog to glucogenesis to fructose 6-phosphate and glucose 6-phosphate. Ribulose 1,5-bisphosphate is subsequently regenerated from fructose 6-phosphate or glycerinaldehyde-3-phosphate and dihydroxyacetone phosphate in a series of reactions analog to the pentose phosphate pathway.

1.3.5 Amino acid metabolism

Amino acid anabolism and catabolism can involve complex series of reactions (detailed information can be obtained from Bender, 1985). The following description focuses on the rearrangement of carbon skeletons of amino acids to illustrate the propagation of stable-isotope labels within metabolic networks (Figure 1.3). Many higher organisms are unable to synthesize amino acids resulting from complex reaction sequences involving more than five reactions, i. a., aromatic amino acids, branched amino acids, methionine, and lysine (Bender, 1985).

Glycine and C_3 amino acids are derived from 3-phosphoglycerate and pyruvate. Serine is directly biosynthesized from 3-phosphoglycerate. Cysteine is derived from serine. Glycine can be derived from serine or from threonine. The degradation of glycine leads back to serine. Alanine is biosynthesized from pyruvate. Catabolism of all C_3 amino acids yields pyruvate.

C_4 amino acids (aspartate, asparagine) are biosynthesized from oxaloacetate and are also degraded to oxaloacetate. In most organisms, lysine is also derived from aspartate and acetyl-CoA. Lysine is degraded to acetoacetate and acetyl-CoA. Aspartate can also be converted to homocysteine, which can serve as precursor compound for methionine and threonine. Methionine is degraded *via* S-adenosyl-methionine back to homocysteine. Homocysteine can be degraded to succinyl-CoA, a citric acid cycle intermediate. Degradation of threonine results in glycine and acetyl-CoA.

C_5 amino acids (glutamate, glutamine, proline and arginine) are biosynthesized from α-ketoglutarate. Glutamate serves as precursor for the biosynthesis of glutamine, proline and arginine. Catabolic processes convert glutamine, proline and arginine on different routes back to glutamate and glutamate to α-ketoglutarate.

Pyruvate serves as precursor for branched amino acids. For the biosynthesis of leucine and valine, two pyruvate units are converted to α-ketovalerate (for details, see Schauder, 1992). In this reaction sequence, C-1 of one pyruvate unit is decarboxylated. Leucine and valine are derived from α-ketovalerate. Isoleucine is biosynthesized from α-aceto-α-hydroxybutyrate, which is generated by pyruvate and α-ketobutyrate. Isoleucine, valine and leucine are degraded to ketone bodies, i. a. acetoacetate.

Aromatic amino acid side chains (in phenylalanine and tyrosine, histidine, tryptophane) are generated *via* the shikimate pathway. Erythrose 4-phosphate generated in the pentose phosphate pathway and phosphoenolpyruvate form shikimate and subsequently chorismate in a complex series of reactions. Pathways leading to phenylalanine and tyrosine, histidine, and tryptophane branch from chorismate (for details, see Crawford, 1989). Tyrosine is formed by hydroxylation of phenylalanine. This pathway is also the catabolic pathway of phenylalanine. Tyrosine is metabolized to fumarate and acetoacetate. Histidine and tryptophane also yield acetoacetate, which is inserted in the intermediary metabolism in form of acetyl-CoA. Tryptophane moieties are also degraded to alanine, even this is only a minor pathway for the generation of alanine because of the low abundance of tryptophane in peptides.

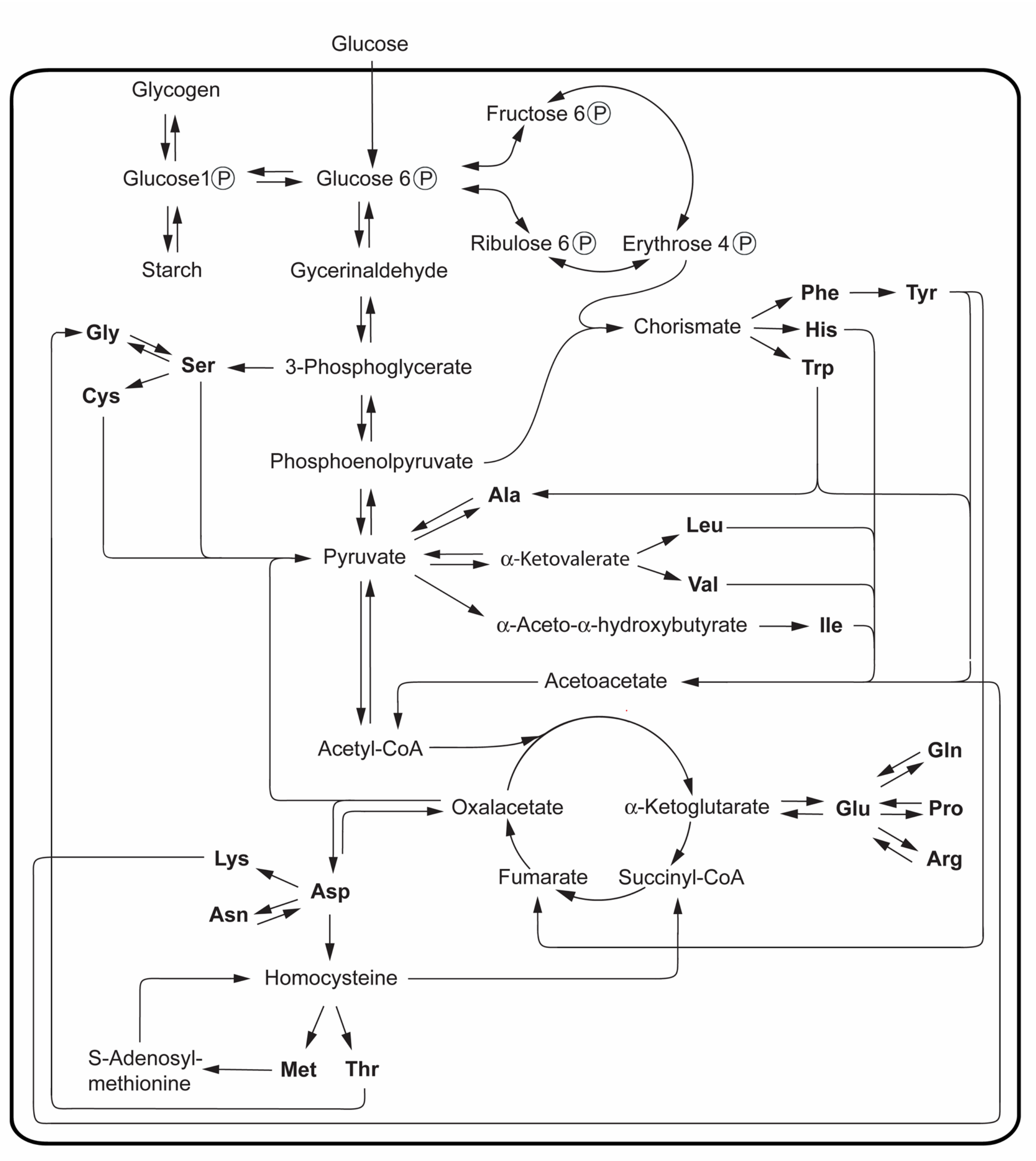

Figure 1.3: Intermediary metabolism and pathways for biosynthesis and degradation of amino acids (for details, see chapter 1.3.5; amino acid are abbreviated by their three letter symbols). The connections in the network illustrate the most common pathways including glycolysis/glucogenesis, citric acid cycle, pentose phosphate pathway and shikimate pathway. Metabolic pathways can be altered or degenerated in specific organisms: many pathways for the biosynthesis of aromatic precursor compounds and complex amino acids (i. a., aromatic amino acids, branched amino acids, methionine, and lysine) vanished in higher organisms of the animal kingdom.

2. Retrobiosynthetic analysis of secondary plant compounds

Retrobiosynthetic analysis (for reviews, see Scott, 1985; Eisenreich and Bacher, 2000) affords information on primary and secondary metabolic pathways. Quantitative evaluation of ^{13}C-labeled amino acid isotopologs determines carbon transfer rates and flux probabilities of intermediary metabolic processes (e. g. Sriram et al., 2004 and Roessner-Tunali et al., 2004). The analysis of unique ^{13}C isotopolog signatures in secondary plant compounds generated from defined ^{13}C-labeled precursor molecules, i. a. isoprenoid moieties or amino acids, yields qualitative and quantitative information on metabolic fluxes and their metabolic history in a biological system (for review, see Eisenreich and Bacher, 2000).

Amino acids offer several advantages compared to other analytes existing in living systems. The reconstruction of primary and secondary metabolic pathways based on data of ^{13}C NMR spectroscopy experiments is possible in every organism (contrary to use for the reconstruction, i. a. glucose, which is not available in all organisms in sufficient amounts for a NMR-based spectroscopic analysis). Despite the capabilities of using this experimental approach, it has to be considered that the metabolic information obtained from amino acids contains different, non-equivalent aspects of metabolic network. Furthermore, different effort is required for isolation and purification of various amino acids and has to be included in the cost-benefit analysis in planning of an experimental setup (cf. chapter 5.3.3).

This chapter describes information which can be obtained from amino acid moieties, the technique to reconstruct metabolic pathways from amino acids and a formal description for the comparison of natural and artificial tracer distributions in order to design stable-isotope perturbation experiments rationally.

2.1 Information content of amino acids

Within the scope of this thesis, the information content of amino acids is defined as magnitude of possible tracer states resulting from different metabolic pathways in the biosynthesis of single amino acids. The maximal information of a tracer signature in amino acids is limited by the size of its biosynthetic precursor with the highest number of carbon atoms and its metabolic history (for details on the biosynthesis of the different amino acid families, see chapter 1.3.5). C_3 and C_4 precursors contain most information, e. g. phosphoenolpyruvate and erythrose 4-phosphate. C_5 blocks like α-ketoglutamate or α-isovalerate contain less information than erythrose 4-phosphate, since they were created from smaller carbon building blocks which already passaged other metabolic processes. For example, α-ketoglutamate is generated from acetyl-CoA and oxaloacetate or α-isovalerate is generated from two units pyruvate, respectively. The relative abundance of ^{13}C-labeled blocks within a compound adds quantitative information on the carbon fluxes in the biosynthetic network.

Amino acids composition of an organism can be used as tool to estimate the cost of isolating single amino acids, thus being a measurement unit for the cost for obtaining a set of amino acid isotopolog signature information. Absolute amino acid compositions are available by mass spectrometry experiments (chapter 5.6), or by analyzing an organism with public available protein databases (i. e., SWISSPROT; Gasteiger et al., 2001). Amino acids count through all organisms available in SWISSPROT, lead to the following averaged amino acid composition: leucine 9.8 %, alanine 7.9 %, serine, 7.4 %, glycine, 7.1 %, valine, 6.8 %, lysine 6.1 %, isoleucine 6.0 %, threonine 5.8 %, arginine, 5.3 %, proline, 5.1 %, asparagine, 4.6 %, glutamic acid, 4.5 %, phenylalanine, 4.2 %, glutamine, 4.1 %, aspartic acid, 4.0 %, tyrosine, 3.3 %, methionine, 2.5 %, histidine, 2.3 %, cysteine, 1.7 % and tryptophane, 1.3 %. Comparing the information content of each amino acid with its assigned cost for retrieving and its anabolic history in a biological system, it is obvious that phenylalanine/tyrosine, leucine, and alanine/serine are the most suited amino acids for intermediary metabolic studies.

2.2 Reconstruction of metabolic pathways

For reconstruction and *in silico* simulation of pathways, the process of formally analyzing carbon moieties can be split up in the interpretation of C_1, C_2, C_3 and C_4 blocks. The results on different carbon precursor molecules are redundant, and can be used as control for the correctness of a conclusion. Interpretation of C_5 blocks, e. g. ribose or ribulose, is not performed within the scope of this thesis. The following interpretation is valid for a biological system grown on medium supplemented with [U-$^{13}C_6$]glucose highly diluted with glucose with natural ^{13}C abundance. It is presumed further, the experimental system performs no refixation of respired $^{13}CO_2$ *via* the Calvin cycle. The interpretation of C_6 units (glucose, fructose) is explained in chapters 3 and 4.

C_1 building blocks can only yield binary information on a metabolic pathway. The pathway is functional, when it generates a $^{13}C_1$ isotopolog in a defined position of the analyte, or is not functional when it affords a $^{12}C_1$ isotopolog at the same position. Prerequesite is a known route from tracer to tracee with suitable pre-labeling of the tracer molecule. For example, the addition of $^{13}CO_2$ to ribulose 1,5-bisphosphate leads to [1-$^{13}C_1$]phosphoglycerate.

C_2 building blocks represented by acetyl-CoA can only yield a minimum of information for quantitative pathway analysis, since possible isotopolog signatures cannot unequivocally attributed to a single pathway. C-1 and C-2 of acetyl-CoA are derived from C-2 and C-3 of pyruvate. The $^{13}C_2$ block is predominantly obtained from passage through glycolysis, the transaldolase reaction of the pentose phosphate pathway, or from glucogenesis *via* oxaloacetate. $^{13}C_1$ at position C-1 of acetyl-CoA is primarily generated by passage of labeled U-$^{13}C_4$ units through the transketolase reaction. [2-$^{13}C_1$]acetyl-CoA can be generated by repetitive passage of $^{13}C_2$ acetyl-CoA units *via* the transketolase reaction subsequent to glucogenesis *via* oxaloacetate.

C_3 building blocks in the form of triose moieties or pyruvate can contain information on all intermediary pathways: glycolysis/glucogenesis, citric acid cycle and transketolase reaction in the pentose phosphate pathway. Isotopologs defined by positions C-1 and C-2 of C_3 blocks were described for C_2 building blocks. Additionally, [3-$^{13}C_1$]triose can be generated by a passage of acetyl-CoA through the citric acid cycle and regeneration of triose moieties in glucogenesis. Universally labeled $^{13}C_3$ units are derived by glycolysis or the transaldolase reaction in the pentose phosphate cycle. Detailed information on transaldolase reaction is lost, since asymmetries in triose moieties resulting from different pathways become averaged by catalytic action of the triose phosphate isomerase.

C_4 building blocks, generally in the form of erythrose 4-phosophate moieties, are generated in the pentose phosphate pathway. Signature descriptions for positions C-2, C-3 and C-4 can be obtained from the description of C_3 building blocks. ^{13}C excess at positions C-1 of C_4 units is primarily generated by secondary cycling and shuffling of a $^{13}C_1$ isotopolog labeled at C-3 through glycolysis/glucogenesis and the pentose phosphate pathway. Universally labeled $^{13}C_4$ units are generated exclusively by single passage of [U-$^{13}C_6$]glucose through the non-oxidative branch of the pentose phosphate pathway by catalytic action of the transketolase enzyme.

Through the isotopolog generation rules for amino acid fragments formally defined in chapter 6.4.3, the intermediary metabolic network of the experimental system can be reconstructed. Each rule encodes the transfer of a defined amount of isotopolog *A* to isotopolog *B* by defining distinct changes in the ^{13}C isotopolog patterns. For example, a rule for the triose moieties passaged through the citric acid cycle can be formulated as $\{XXX\} \rightarrow \{X00\} + \{0XX\}$. The educt in this rule is an arbitrary triose isotopolog {XXX}, which is converted to the product encoded by either {X00} or {0XX} with the same probability. Positions occupied with zero illustrate the conversion of a ^{13}C-label to a ^{12}C-label in a defined carbon position by recombination of the analyte molecule with unlabeled molecules in the experimental system (in catabolic experimental setups; for details, see chapter 3.1). If the zero position is already occupied with ^{12}C no change will occur. X conserves a ^{12}C or ^{13}C label at the given position of the analyte. Inference of these rules is performed according to the 4F algorithm (for details, see chapter 3.6 and chapter 6.4.2 and 6.4.3; a general inference automata is shown in Figure 2.1). It can be summarized that amino acids yield information on glycolysis/glucogenesis, the transketolase reaction of the pentose phosphate pathway, the citric acid cycle and the initial reactions of the shikimate pathway leading to the formation of chorismate, the central precursor for aromatic compounds. The statistical significance of the computer simulations can be tested by a t-test (chapter 6.2.4). Experimental series including a set of metabolic simulations can be analyzed by similarity analysis using clustering algorithms, e. g. UPGMA and NJ (chapter 6.5.3 or 6.5.4), or neural networks (chapter 6.5.5).

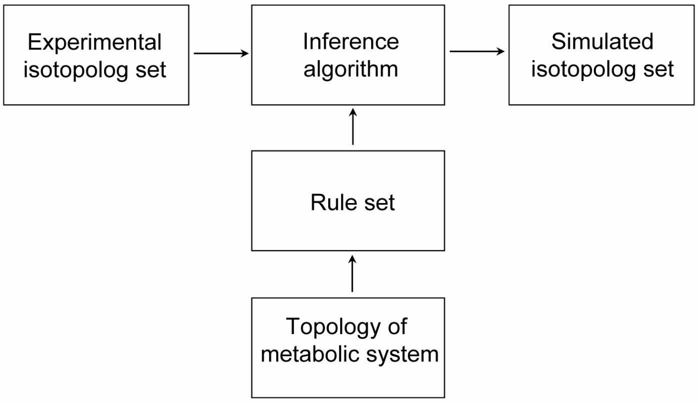

Figure 2.1: The architecture of a general inference automaton which deconvolutes an isotopolog set to metabolic processes. The topology of a metabolic system defines a rule set encoding biochemical reactions which lead to unequivocal isotopolog signatures. The inference algorithm applies these rules to prove the validity of the experimental isotopolog set with respect to the defined metabolic topology. The resulting simulated isotopolog set can be compared to the experimental isotopolog set. High degrees of similarity indicate a well-modeled system.

2.3 Tracer distributions in stable-isotope labeling experiments

Experimental results of ^{13}C perturbation experiments obtained by ^{13}C NMR spectroscopy or mass spectroscopy can be compared with either the natural ^{13}C quasi-equilibrium or with the statistical abundance correlated to the artificial ^{13}C enrichment in the analyte. The similarity between experimental and statistical distributions of isotopomer groups indicates the magnitude of the perturbation in the system.

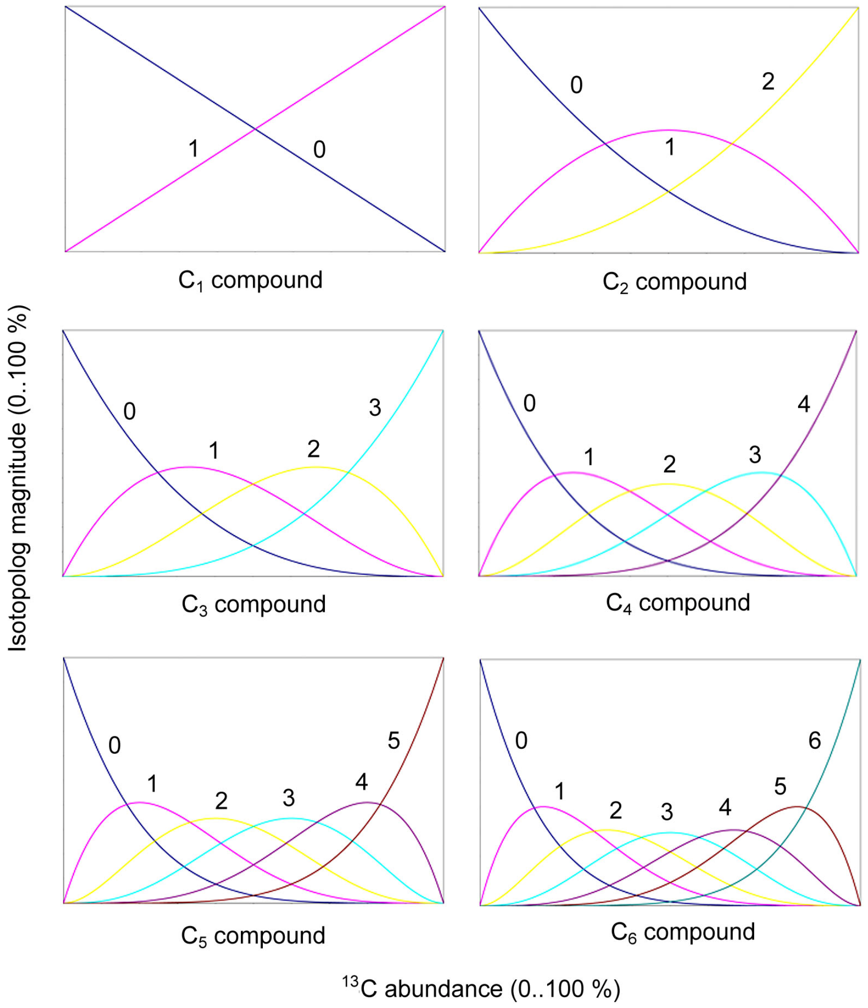

Figure 2.2: Statistical ^{13}C distributions for C_1 to C_6 compounds. Arabic numbers indicate the number of ^{13}C atoms in a compound. The abscissa defines the ratio of ^{12}C to ^{13}C in the experimental system. The ordinate defines the relative fraction of an isotopomer set. Biological ^{13}C perturbation experiments shift the statistical distribution to an artificial distribution generated by the biochemical network in the system.

Assuming a concentration X for ^{12}C and Y for ^{13}C, with the constraints $X + Y = 1$ and $X \in [0;1]$, $Y \in [0;1]$, the abundance L of an isotopolog group with k different isotopologs over n ^{13}C atoms out of m ^{12}C or ^{13}C carbon atoms can be expressed by $L = k \cdot (X)^{m-n} \cdot (Y)^{n}$.

For universally labeled isotopologs with either ^{12}C or ^{13}C, this equation reduces to $L = (X)^{m}$ or $L = (Y)^{m}$, respectively. For natural carbon-based systems, $X = 0.011$ and $Y = 0.989$ applies. Statistical ^{13}C distributions for C_1 to C_6 compounds are illustrated graphically in Figure 2.2. Each function set is symmetric relative to $X = Y = 0.5$, even while single distribution functions can be linear (e. g. for functions 0 and 1 in C_1 compounds), exponential symmetric (e. g. for function 1 in C_2 compounds), or exponential asymmetric (e. g. for functions 1 and 2 in C_3 compounds). The abscissa of Figure 2.2 shows the total ^{13}C abundance and the ordinate the abundance of a single isotopomer group. The total ^{13}C abundance of a compound or a mixture is unequivocally available by IRMS (chapter 5.5). Trivially, the sum of all isotopomer groups adds up to 1.0. For example, for a C_2 compound with 20.0 % total ^{13}C abundance, the $^{12}C_2$ isotopomer group contributes 64.0 %, the $^{13}C_1/^{12}C_1$ isotopomer group contributes 32.0 % and the $^{13}C_2$ isotopomer group contributes 4.0 %. A C_6 compound with 50.0 % abundance for ^{13}C contains 1.6 % $^{12}C_6$, 9.4 % $^{13}C_1/^{12}C_5$, 21.8 % $^{13}C_2/^{12}C_4$, 29.7 % $^{13}C_3/^{12}C_3$, 21.8 % $^{13}C_4/^{12}C_2$, 9.4 % $^{13}C_5/^{12}C_1$ and 1.6 % $^{13}C_6$. For the illustrated example, the symmetry of ^{13}C distributions can be illustrated by the equivalent abundances for the isotopomer groups $^{12}C_6$ *versus* $^{13}C_6$, $^{13}C_1/^{12}C_5$ *versus* $^{13}C_5/^{12}C_1$, and $^{13}C_2/^{12}C_4$ *versus* $^{13}C_4/^{12}C_2$.

With the knowledge of statistical stable-isotope isotopomer distributions, experiments can be designed which favor the generation of distinct isotopomer species. Since biochemical networks will not generate a truly statistical abundance of ^{12}C to ^{13}C, the perturbation of the network can be quantitatively estimated. The ^{13}C perturbation Z can be defined by the sum of m deviations between the experimental isotopomer abundances E and the statistical isotopomer abundances S. Z must be normalized to the total ^{13}C abundance Y in the system, $Z = \sum_{i=0}^{m} \frac{|E_i - S_i|}{Y}$. The experimental perturbation can only be greater than zero and corresponds to the information entropy of a biochemical system (see chapter 3.5).

2.4 Application examples

The following two examples illustrate the retrobiosynthetic approach using amino acid and secondary metabolite isotopolog information obtained by ^{13}C perturbation experiments using [1-$^{13}C_1$]glucose and [U-$^{13}C_6$]glucose as tracers. Firstly, the metabolic history of benzofurans in *Tagetes patula* root cell culture is analyzed and secondly, a quantitative reconstruction of the metabolic network in *Taxus chinensis* cell culture producing the cytostatic drug taxol is performed.

2.4.1 Benzofuran derivates are synthesized *via* phenylalanine and 1-deoxy-D-xylulose 5-phosphate in *Tagetes patula*

Compounds containing benzofuran moieties have been isolated from several species of higher plants including *Asteraceae*, *Rutaceae*, *Liliaceae*, and *Cyperaceae*. In *Tagetes patula* benzofurans were extensively studied and at least seven benzofuran derivates (Figure 2.3) were isolated from leafs, roots and rhizosphere (Bohlmann and Zdero, 1979; Menelaou *et al.*, 1993; Parodi *et al.*, 1988; Sütfeld *et al.*, 1985; Tang *et al.*, 1987). *Tagetes* species are used as feed additives and cosmetics, because their metabolic profile enclose many carotinoids, essential oils, flavonoids and thiophenes (Garg et al., 2000; Flachmann et al., 2004; Gupta 2004). Benzofurans are causative agents for trembling in higher animals. Furthermore, human milk sickness has been associated with benzofurans from certain plants contained in cattle forage (i. e., *Eupatorium rugosum*, *Aplopappus heterophyllus* and *Aplopappus fruticosus*; for further information, see Sharma et al., 1998; Christiensen, 1965).

Two routes for the biosynthesis of benzofurans have been proposed by ^{14}C-radioisotope-labeling studies: route A reports synthesis from acetyl-CoA units *via* polyketide-type intermediates (Lin et al., 1974; Lin and Heinstein 1974), and route B is based on a phenylpropanoid precursor derived from phenylalanine (Siebertz et al., 1988; Monir and Proksch, 1989; Bohlmann and Grenz, 1970). Both, DMAPP (Bohlmann and Grenz, 1970) or IPP (Lin and Heinstein, 1974) were thought to serve as electrophile in this reaction. The source of IPP units can be either mevalonate or 1-deoxyxylulose 5-phosphate. Both pathways exist in higher plants. For example, shikonin produced in *Lithospermum erythrorhizon* cell culture is dependet on mevalonate (Lange et al., 1998), while taxol produced in *Taxus chinensis* cell culture is dependet on 1-deoxyxylulose 5-phosphate (see chapter 2.4.2 and Eisenreich et al., 1996). The mevalonate and DOXP pathways can operate simultaneously and are located in two different compartments. The mevalonate pathway is located in the cytoplasm, while the DOXP pathway occurs in the plastids (Fellermeier et al., 2003). The compartmentation of the pathways is not absolute, as shown by crosstalk analysis of the mevalonate pathway *versus* the DOXP pathway by Hemmerlin et al., 2003.

It will be shown, that isoeuparin and 4-hydroxytremetone in root cultures of *Tagetes patula* are derived from phenylalanine and DMAPP *via* the DOXP pathway.

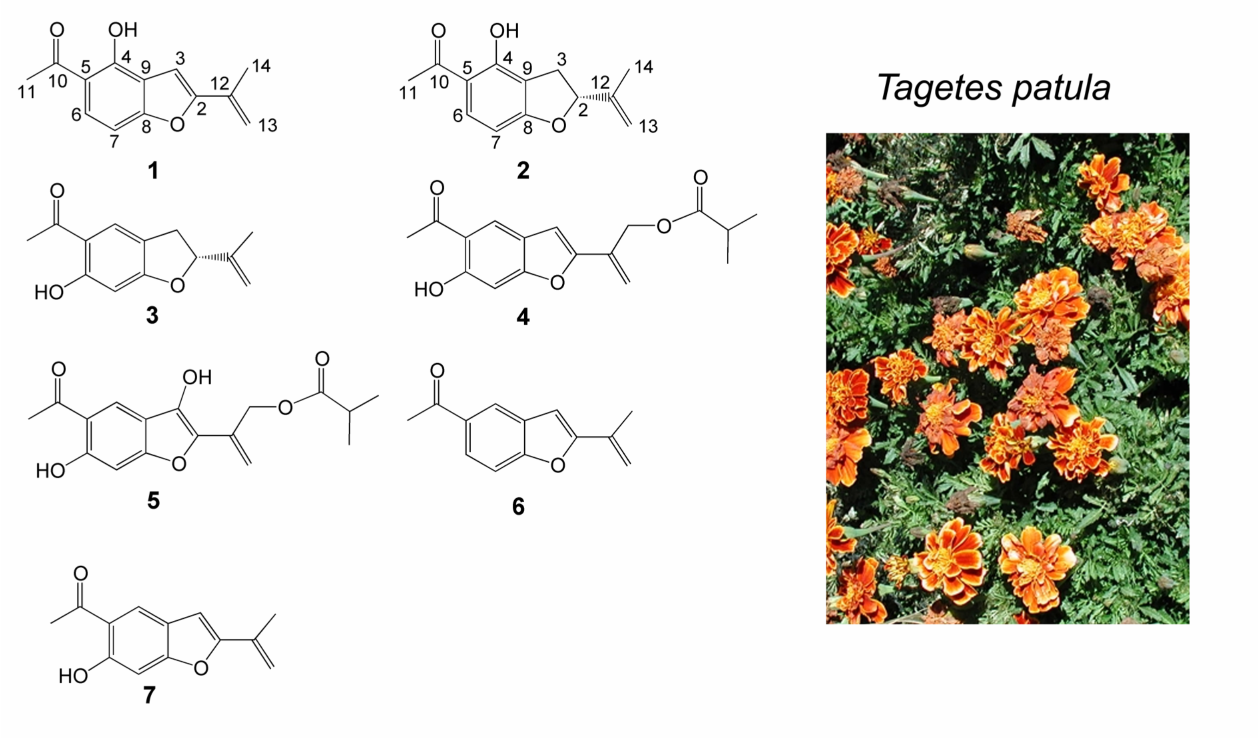

Figure 2.3: Chemical structure of benzofuran compounds isolated from *Tagetes patula* (shown right): (1) isoeuparin, (2) 4-hydroxytremetone, (3) 6-hydroxytremetone, (4) 6-hydroxy-isoeuparin-14-O-isopropylat, (5) 6-hydroxy-tremetone-14-O-isopropylat, (6) tremetone, (7) 6-hydroxy-isoeuparin.

Cell cultures of root tissue of *Tagetes patula* were shown to produce isoeuparin (Sütfeld et al, 1985) and 4-hydroxytremetone (Bigi et al., 1983) as major benzofuran derivates, among 6-hydroxytremetone, 6-hydroxy-isoeuparin-14-O-isopropylat, 6-hydroxy-tremetone-14-O-isopropylat, tremetone and 6-hydroxyisoeuparin (Figure 2.3). Samples of isoeuparin and 4-hydroxytremetone were provided by Lilla Margl, Chair of Organic Chemistry and Biochemistry, Technical University of Munich. The compounds were identified by ^{1}H and ^{13}C NMR spectroscopy and on the basis of $^{1}H^{1}H$ homocorrelation (COSY), $^{1}H^{13}C$ heterocorrelation (HMQC and HMBC) and INADEQUATE NMR experiments. INADEQUATE spectra identified 4 pairs of connected ^{13}C atoms in isoeuparin and 4-hydroxytremetone (C-6/C-7, C-6/C-5, C-12/C-13 and C-2/C-3; Figure 2.4). The optical rotation of 4-hydroxytremetone was determined at $[a]_D^{20} = -107.8°$. The (R)-configuration of (-)-tremetone and (-)-6-hydroxytremetone was shown by Bonner et al., 1964 and Zalkow et al., 1979, therefore an (R)-configuration of 4-hydroxytremetone can be assumed.

Biosynthetic fluxes were reconstructed by supplementing the medium of the *Tagetes patula* root culture with [U-^{13}C$_6$]glucose or [1-^{13}C$_1$]glucose at ratios of 10:1 and 2:1 (w/w), respectively. After a growth period of 13 days, the cells were harvested and isoeuparin and 4-hydroxytremetone were isolated by extraction. The compounds were purified chromatographically and analyzed quantitatively by ^{13}C NMR spectroscopy. Acidic hydrolysis of the biomass afforded the amino acids glycine, alanine, serine, valine, leucine, phenylalanine, and tyrosine.

Quantitative evaluation of ^{13}C NMR spectra of isoeuparin and 4-hydroxytremetone was performed as described (see chapter 1.1.3). Relative abundances due to the relative intensities of ^{13}C^{13}C couplings were cross-calibrated *via* the natural abundance intensity of the signals relative to C-14 to 12.4 mol %. H-14 show heteronuclear coupling to ^{13}C-14 in the ^{1}H spectrum ($^1J_{CH}$=128 Hz) as well as long-range ^{1}H^{13}C coupling between H-14 and ^{13}C-14 ($^2J_{CH}$=5.8 Hz).

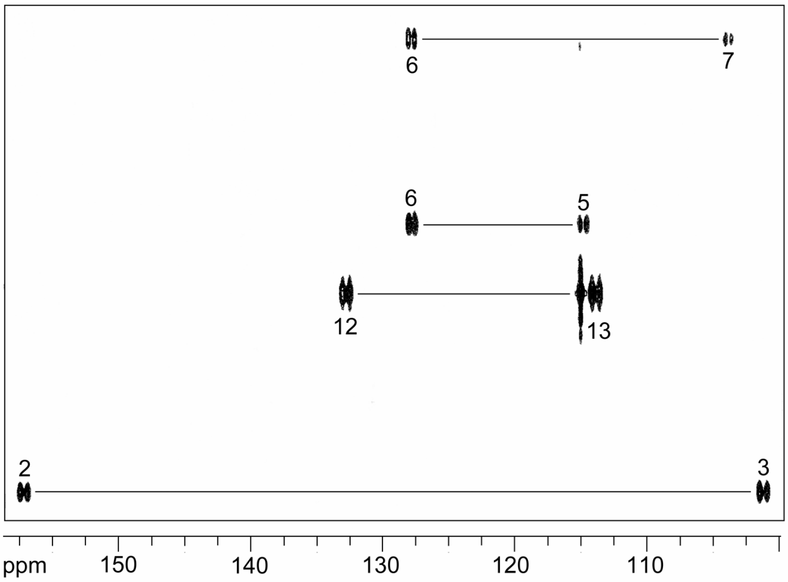

Figure 2.4: INADEQUATE spectra of von isoeuparin from root cultures of *Tagetes patula* supplemented with [U-^{13}C$_6$]glucose. The spectrum shows four adjacent ^{13}C atoms, which can be assigned to the pairs C-2/C-3, C-5/C-6, C-6/C-7, and C-12/C-13.

Labeling patterns of a *Tagetes patula* root culture with [1-$^{13}C_1$]glucose as tracer result in 4-hydroxytremetone and isoeuparin labeled at C-4, C-6, C-10 of the benzyl ring, C-13 of the acetoxy side chain as well as C-3 of the furan or the dihydrofuran ring system (data summarized in Figure 2.5 and 2.7). The excess of about 1.5 ± 0.7 mol % over natural abundance of the remaining carbon atoms arises by reshuffling of the labeled glucose precursor within the metabolic network of the cells.

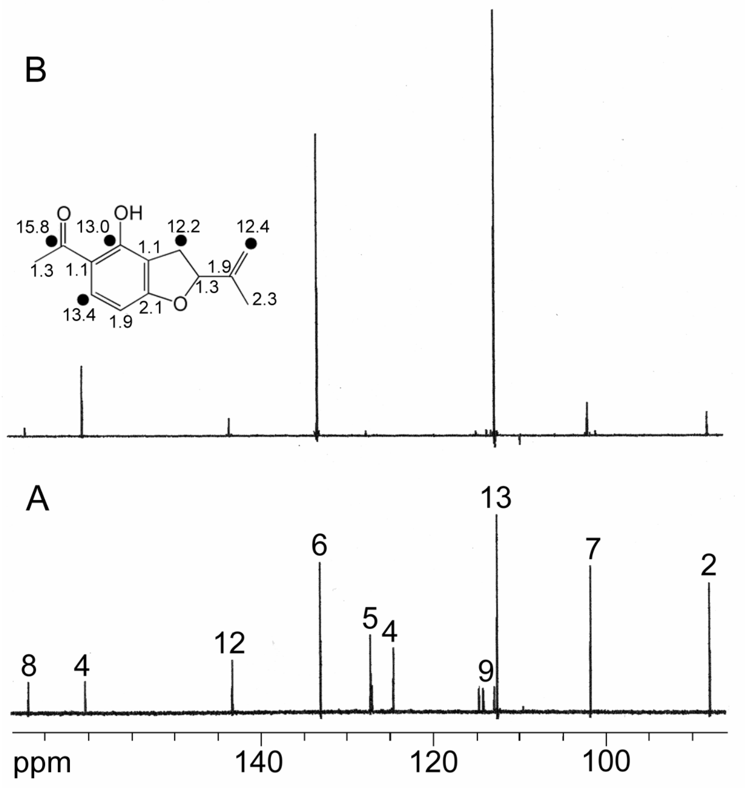

Figure 2.5: *Tagetes patula* grown on [1-$^{13}C_1$]glucose yields labeled isoeuparin with defined ^{13}C NMR signatures (B), which can be analyzed by comparison with natural abundance ^{13}C NMR spectra (A).

The experiment with [U-$^{13}C_6$]glucose as precursor affords complex ^{13}C NMR spectra due to extensive $^{13}C^{13}C$ coupling of the signal generated by the presence of multiple ^{13}C-labeled 4-hydroxytremetone (see Figure 2.6) and isoeuparin isotopolog species. Satellite signals were analyzed by coupling constants deduced from the one-dimensional ^{13}C NMR spectrum and by homonuclear correlation in the INADEQUATE NMR experiment. For example, the high-resolution ^{13}C NMR spectrum of 4-hydroxytremetone resulting from the [U-$^{13}C_6$]glucose

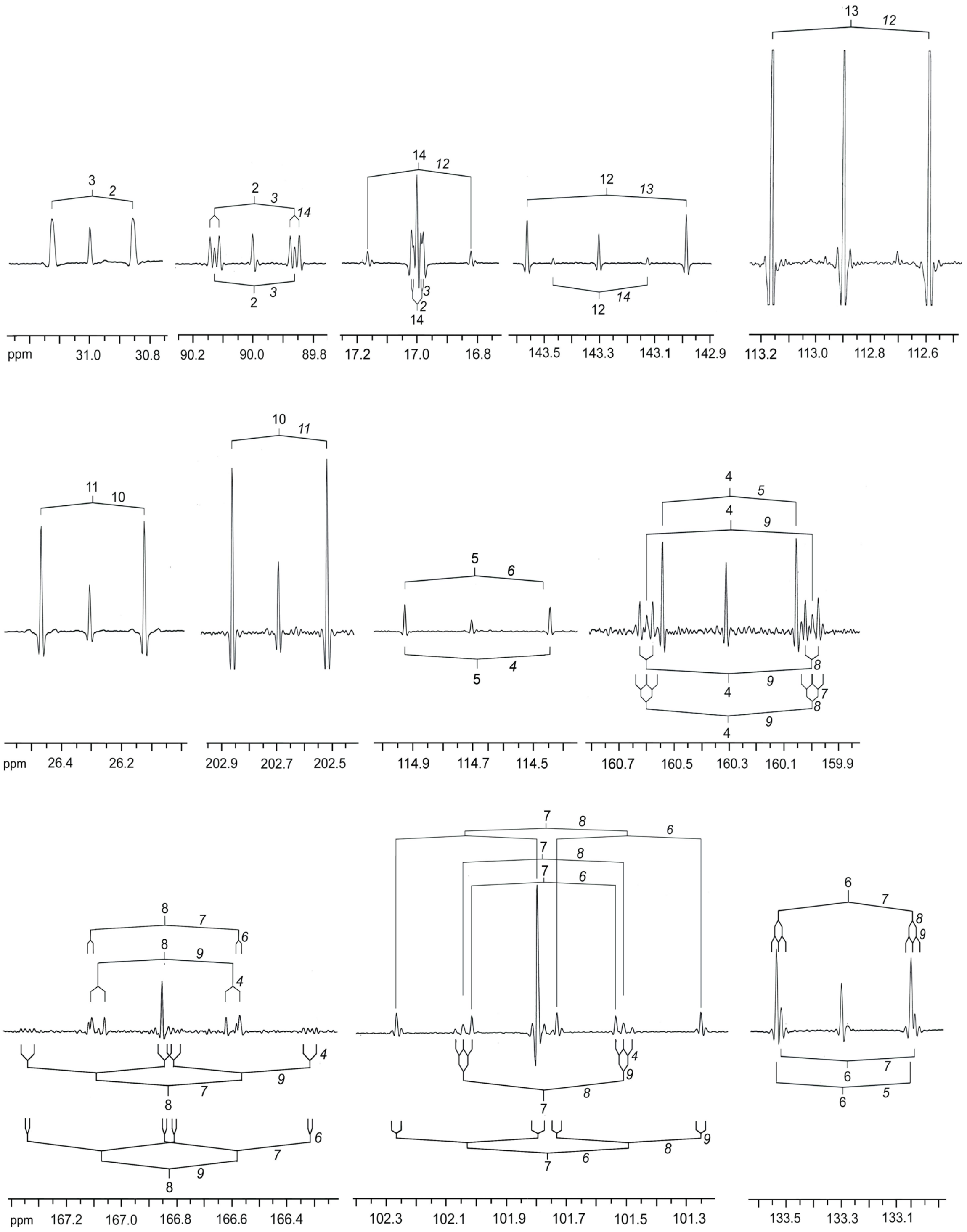

Figure 2.6: High-resolution ^{13}C NMR spectra of 4-hydroxytremetone isolated from *Tagetes patula* root cell culture grown on medium supplemented with [U-$^{13}C_6$]glucose. The processing of the FID with a strong Gaussian function resolves long-range $^{13}C^{13}C$ couplings in the benzofuran skeleton.

experiment show well-resolved $^{13}C^{13}C$ signal coupling satellites. In detail, the ^{13}C spectrum contains 16 dinuclear ^{13}C coupling signatures, 6 trinuclear ^{13}C coupling signatures, and 6 tetranuclear ^{13}C coupling signatures (Figure 2.6). Signals of C-4, C-6, C-7 and C-8 within the benzyl ring system of 4-hydroxytremetone showed the most complex multiplets including $^{13}C_3$ signatures and $^{13}C_4$ signatures. The $^{13}C_3$ carbon moieties were found at C-4/C-8/C-9 or C-6/C-7/C-8 and the $^{13}C_4$ carbon moieties are found at C-4/C-7/C-8/C-9 or C-6/C-7/C-8/C-9. Furthermore, $^{13}C_2$ units with high signal intensity were found at positions C-5/C6 and C-4/ C-5. The spectral $^{13}C^{13}C$ coupling signatures already show qualitatively a symmetrical character of the benzyl ring system in benzofurans. Carbon atom C-14 of 4-hydroxytremetone and isoeuparin shows only statistical coupling to C-12, but shows long-range coupling to C-2 and C-3. The [2,3,14-$^{13}C_3$]benzofuran signature is also seen in the long-range coupling between ^{13}C-14 and $^{13}C_2$ in signal C-2. The C-2/C-3/C-14 triple-labeled $^{13}C_3$ block indicates a common biosynthetic precursor derived from [U-$^{13}C_6$]glucose. The interruption of this block at C-12 already hints DMAPP as putative precursor in benzofuran biosynthesis (for review, see Eisenreich et al., 2004).

The results of the quantitative evaluation of the ^{13}C NMR spectra are summarized in Figure 2.7. The benzyl system of 4-hydroxytremetone and isoeuparin shows $^{13}C_3$ blocks with an abundance of 3.3 ± 0.3 mol %. $^{13}C_4$ blocks can be estimated to 1.1 ± 0.1 mol % in 4-hydroxytremetone. The $^{13}C_4$ signature cannot be determined from ^{13}C NMR spectra of isoeuparin, caused by a disadvantageous signal-to-noise ratio. The distal 2,3,14-$^{13}C_3$ moiety is ^{13}C-labeled with 5.8 mol %. The acetyl side chain at C-10/C-11 is ^{13}C-labeled with 8.2 mol %. The $^{13}C_2$ signature of C-12/C-13 is in the same order of magnitude as the acetyl side chain, showing an abundance of 7.4 mol %.

The existence of isotopologs with continuous $^{13}C_3$ and $^{13}C_4$ carbon signatures excludes the biosynthesis of benzofurans *via* the polyketide pathway, which transfers only $^{13}C_2$ units in the form of acetyl-CoA to the biosynthetic precursor molecule. $^{13}C_3$ and $^{13}C_4$ isotopolog signatures in the benzyl ring system suggest the benzofuran skeleton of 4-hydroxytremetone and isoeuparin is obtained from a shikimate-derived precursor molecule. Central metabolites in the metabolic network can be reconstructed as described (cf. chapter 2.2) from labeled amino acids isolated from *Tagetes patula* root cell culture.

Figure 2.7 A and 2.7 B show quantitatively the resulting ^{13}C perturbations generated through the biochemical network of the plant cell. In the pathways a and b, acetyl-CoA is reconstructed from C-1/C-2 of leucine. Isotopologs of glyceraldehyde 3-phosphate and hydroxyethyl TPP were deduced from C-1 to C-3 of phenylalanine or valine, respectively.

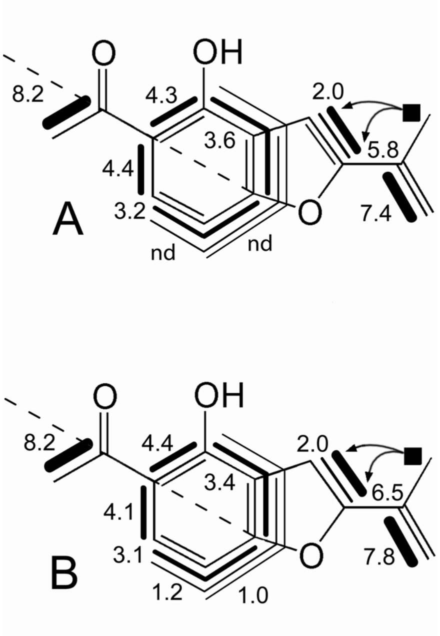

Figure 2.7: ^{13}C-labeling signatures of isoeuparin (A) and 4-hydroxytremetone (B) isotopologs. The signatures of the benzyl ring system are symmetric (dashed line) and indicate of an aromatic precursor compound in the biosynthesis of benzofurans, e. g. acetophenone or 4-hydroxylacetophenone.

The isotopolog composition of an acetophenone type precursor was deduced from the aromatic ring system of phenylalanine in pathway c. From glyceraldehyde 3-phosphate and hydroxyethyl TPP, valid isotopolog allocations of 1-deoxyxylulose 5-phosphate were deduced for pathway d. The arrangement of the DOXP pathway cleaves the $^{13}C_3$ unit introduced by glyceraldehyde 3-phosphate.

For *Tagetes patula* root cell culture grown in medium supplemented with [1-$^{13}C_1$]glucose, the retrobiosynthetic analysis for pathways a-d results in benzofuran isotopologs labeled as follows: Pathway a, the acetophenone unit biosynthesized *via* a polyketide-type pathway would be labeled at C-1, C-3, C-5 and C-7. Pathway b, the DMAPP unit from the mevalonate pathway would be labeled at C-2, C-4 and C-5. Pathway c, the acetophenone unit derived from phenylalanine would be labeled at C-2, C-4 and C-8. Pathway d, the DMAPP unit from the DOXP pathway would be labeled at C-1 and C-4.

For the same experiment with [U-$^{13}C_6$]glucose as tracer, the retrobiosynthetic analysis for pathways a-d affords benzofuran isotopologs labeled as follows: Pathway a, the acetophenone unit biosynthesized *via* a polyketide-type pathway would show four $^{13}C_2$ signatures with the same signal intensity derived from acetyl-CoA. Pathway b, the DMAPP

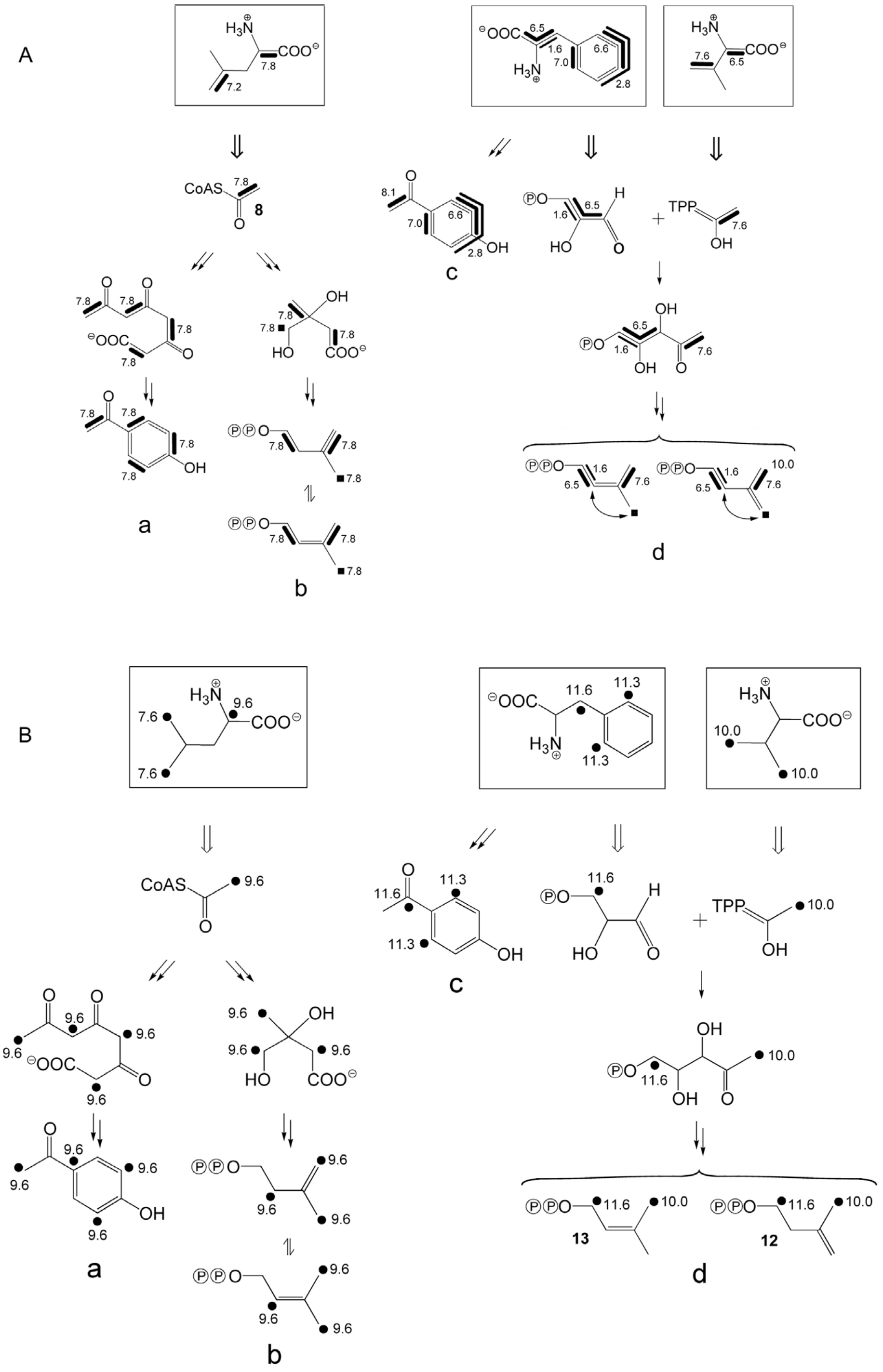

Figure 2.8: *Tagetes patula* root cell culture grown media supplemented with [U-$^{13}C_6$]glucose (A) or [1-$^{13}C_1$]glucose (B), respectively, yields ^{13}C-labeled amino acids. ^{13}C-labeling patterns of acetyl-CoA, acetophenone, glyceraldehyde 3-phosphate, and hydroxyethyl-TTP were deduced from the ^{13}C-labeling patterns of leucine, phenylalanine and valine.

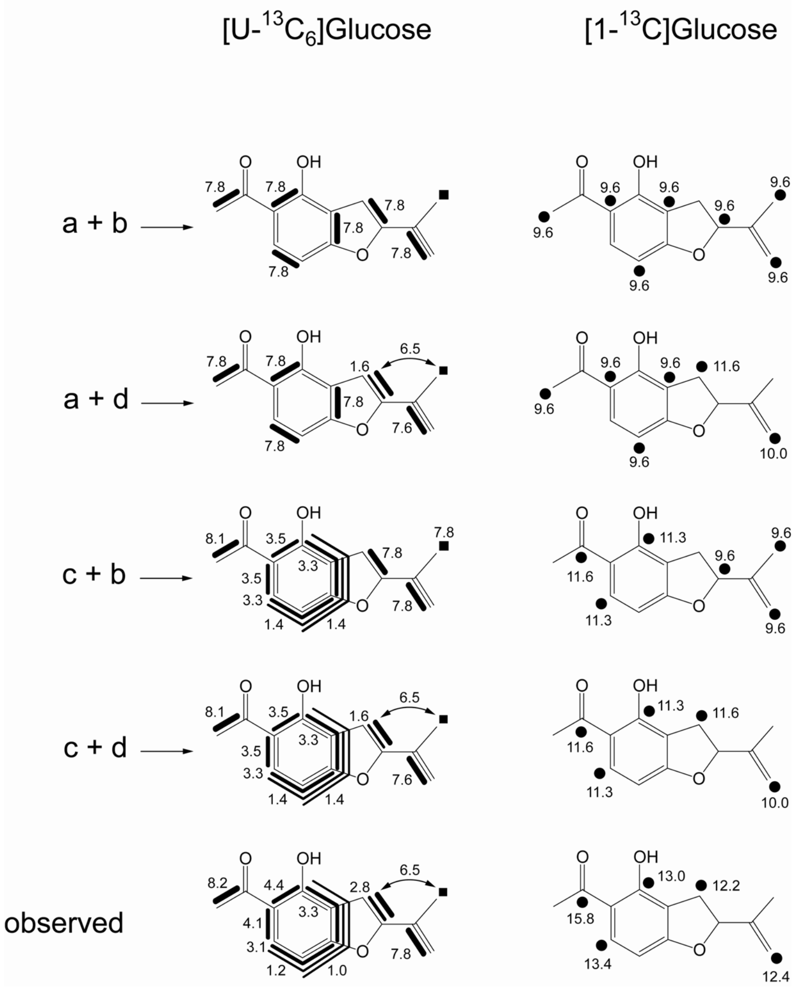

Figure 2.9: Comparison of hypothetical ^{13}C-labeling patterns of benzofurans generated from pathways a to d in Figure 2.8. The observed labeling patterns from benzofuran compounds in root cell cultures of *Tagetes patula* grown on medium supplemented [U-$^{13}C_6$]glucose or [1-$^{13}C_1$]glucose correspond to pathways c and d (cf. Figure 2.8).

unit from the mevalonate pathway would be ^{13}C-labeled with two $^{13}C_2$ and one $^{13}C_1$ block corresponding to C-1/C-2, C-3/C-5 and C-4 of mevalonate. Pathway c, the acetophenone unit derived from phenylalanine would be ^{13}C-labeled with two $^{13}C_2$ units at C-1/C-2 and C-3/C-4. $^{13}C_4$ and $^{13}C_3$ blocks should occur in the moiety of the molecule derived from erythrose 4-phosphate (C-5, C-6, C-7, C-8). Pathway d, the DMAPP unit from the DOXP pathway would show a $^{13}C_2$ and $^{13}C_3$ signature. The $^{13}C_2$ unit is derived from hydroxyethyl TPP, whereas the non-continuous $^{13}C_3$ signature is generated C-1, C-2 and C-3 of 1-deoxyxylulose 5-phosphate.

The possible biosynthetic pathways a to d to precursor molecules (Figure 2.8) can now be used to predict possible ^{13}C-labeling patterns in the benzofuran ring system. As shown in Figure 2.9, the ^{13}C-labeling patterns in benzofurans correspond to the ^{13}C-labeling patterns of precursor moieties derived from the pathways c and d. The metabolic origin of benzofurans can now be unequivocally assigned: benzofurans are biosynthesized from phenylalanine and DMAPP generated in the DOXP pathway.

On the basis of this information, a putative pathway for the biosynthesis of benzofurans can be formulated (Figure 2.10). It is assumed that 4-hydroxy-acetophenone is derived from phenylalanine. 4-hydroxy-acetophenone and DMAPP form 4-hydroxy-5-prenyl-acetophenone by electrophilic substitution. Formation of an intermediary epoxide and nucleophile attack from the 4-hydroxy group may result in the benzofuran ring system. Cleavage of water yields isoeuparin. Hydroxylation and reduction at C-2 would subsequently generate tremetone from isoeuparin.

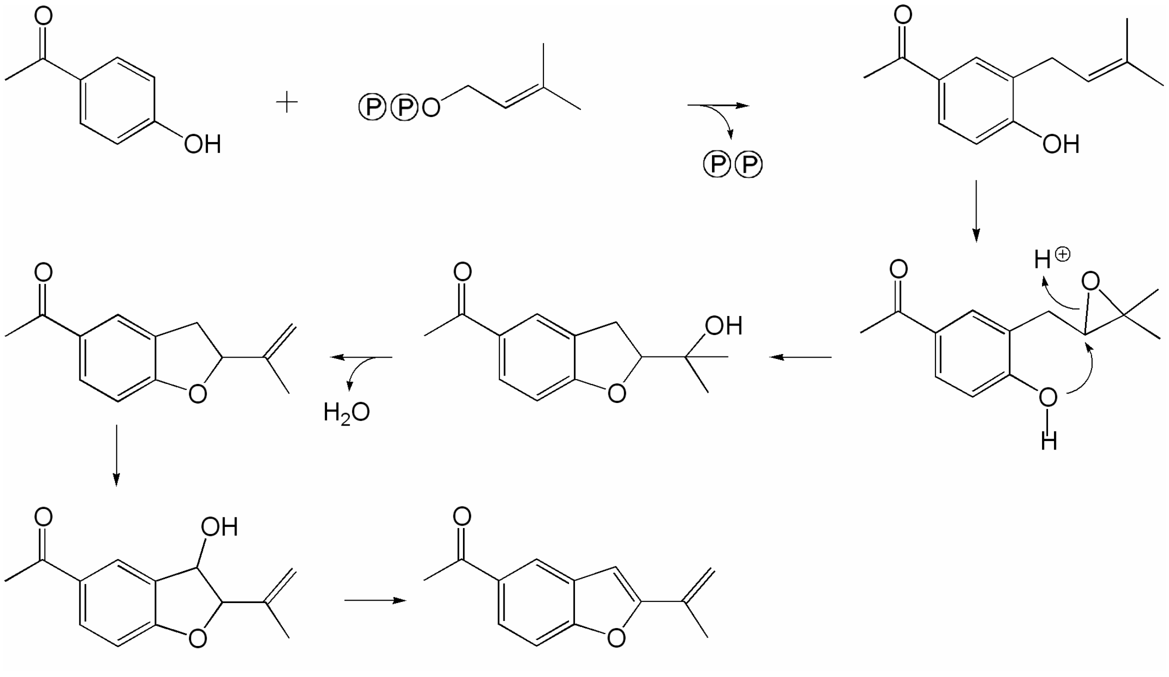

Figure 2.10: Hypothetical biosynthesis of benzofuran compounds in *Tagetes patula* starting from acetophenone and DMAPP. Isoeuparin subsequently serves as precursor for further benzofuran derivates, e. g. 4-hydroxytremetone.

2.4.2 Biosynthesis of taxol and intermediary metabolism of *Taxus chinensis* cells

Taxol and its semisynthetic derivates are among the most important first and second line drugs for the chemotherapy of certain solid tumors including mammary and ovarian carcinoma. Taxol inhibits disassembly of microtubules by binding to the microtubule (+)-end (for review, see Amos and Löwe, 1999).

Introduction of semisynthetic analogs, e. g. Taxotere, has helped to overcome temporary shortage of taxoids from natural sources for tumor therapy. Cell culture and fermentation technology still yields only low amounts of the compound (Lin et al., 1996; Besumbes et al., 2004). Detailed analysis of the biosynthesis of taxol and the intermediary metabolism of *Taxus chinensis* cells may help to optimize biotechnological processes and biotechnological engineering. Taxol is biosynthesized from isoprenoid building blocks *via* the DOXP pathway (Eisenreich et al., 1996). Phenylalanine is incorporated in the form of the two 3-benzoyl moieties and the 2-hydroxy-3-amino-3-phenylpropionyl side chains in taxol as shown by ^{14}C-labeling studies (Fleming et al., 1993 and Fleming et al., 1994).

In the following sections, quantitative assessments of the metabolic origins of the isoprenoid building blocks as well as the benzyl and 2-hydroxy-3-amino-3-phenylpropionyl side chains of taxol are described. The intermediary metabolism of *Taxus chinensis* cells is reconstructed by the 4F algorithm described in chapter 2.2 and chapter 3.6.

Cell culture of *Taxus chinensis* was grown on B5 medium supplemented with [1-$^{13}C_6$]glucose and [U-$^{13}C_6$]glucose as described (Eisenreich et al., 1996). After a growth period of 47 days cells were harvested and taxol was extracted and purified chromatographically. Amino acids were obtained by acid hydrolysis as described (Eisenreich et al., 1996; chapter 5.3.3). The samples were provided by Joo Sun Son, Samyang Genex Biotech Research Institut, Daejeon, Korea. From *Taxus chinensis* cells labeled with [1-$^{13}C_1$]glucose, phenylalanine, tyrosine, valine, alanine, isoleucine, leucine, serine, threonine and lysine were purified and analyzed by ^{1}H and ^{13}C NMR spectroscopy. All isolated amino acids afforded standard labeling patterns resulting from [1-$^{13}C_6$]glucose as described (chapter 2.2 and 2.41). From *Taxus chinensis* cells labeled with [U-$^{13}C_6$]glucose, phenylalanine, valine, serine, isoleucine, leucine, glutamate, aspartate and lysine were purified and analyzed by ^{13}C NMR spectroscopy. From phenylalanine, glutamate, aspartate, valine, and lysine high-resolution ^{13}C NMR spectra were obtained. Signal multiplets generated by ^{13}C^{13}C coupling were evaluated to obtain the amino acid isotopolog compositions and to reconstruct metabolic fluxes in the experimental system.

The ^{13}C NMR spectrum of taxol from *Taxus chinensis* cell culture grown on [1-$^{13}C_1$]glucose was evaluated using available signal assignments (Eisenreich et al., 1996; Shi et al., 2003; Figure 2.11). In the ^{1}H spectrum the signal of H-10 is well separated and appears as multiplet because of heteronuclear ^{1}H^{13}C coupling (Figure 2.11). The ^{13}C abundance of C-10 of taxol was determined at 14.0 mol %. The remaining signals were cross-calibrated to C-10 as described in chapter 1.1.3. The evaluated spectra elucidated the following ^{13}C-labeling patterns: the four isopentenyl units are labeled in their C-1 and C-5 position, all acetyl groups are labeled in their C-2 position and all benzyl groups are labeled in their C-1, C-3 and C-7 positions. The ^{13}C enrichment of the remaining carbon atoms are well in the range of the natural ^{13}C abundance. The labeling pattern of the taxane skeleton as well as the labeling pattern of acetyl and benzyl groups is confirmed by other sources (Eisenreich et al., 2004; Fleming et al., 1993; Fleming et al., 1994).

^{13}C NMR spectra from taxol from *Taxus chinensis* cells grown on [U-$^{13}C_6$]glucose showed a high degree of multiplet coupling and were evaluated as described (chapter 1.1.3). Signals of the isopentenyl moieties were clearly resolved (Figure 2.11 A). C-10 of taxol shows long-range coupling to C-11 and C-13. The same long-range ^{13}C^{13}C coupling is also visible from the ^{13}C signal of C-11 of taxol (cf. Figure 2.11 B). Signals of the 2-hydroxy-3-amino-3-phenylpropionyl side chain were resolved (cf. Figure 2.11 C) and show a C_3 signature grafted by glycolysis (1',2',3'-$^{13}C_3$) and pentose phosphate pathway (2',3'-$^{13}C_2$) as described (see chapter 2.2). While the C-1 signal of the benzyl groups is well resolved, the signals of the aromatic rings A, B and C show extensive overlapping and are located in a crowded region of the spectrum (126.5 – 136.0 ppm; Figure 2.12). Signals of benzyl groups were only partially assigned in the ^{13}C spectrum. Computer simulation of the ^{13}C NMR spectra of distinct benzyl spin systems was unable to assist in evaluation of the remaining benzyl isotopolog signatures (chapter 6.6). Signals of the benzyl ring C were best accessible for evaluation compared to signals of the aromatic rings A and B. The signals C-1_C, C-2_C and C-4_C and C-1_A, C-2_A, C-2_B and C-4_B are well resolved in the ^{13}C NMR spectrum. C-3_C overlaps with the assigned signals of C-3_A, C-3_B, C-1_C and C-4_A in the region of 128.0 to 129.5 ppm. The signals of C-3_A, C-4_A, and C-3_B are located in the region of 128.0 to 129.5 ppm, but cannot be assigned at the available experimental resolution of the ^{13}C NMR spectrum.

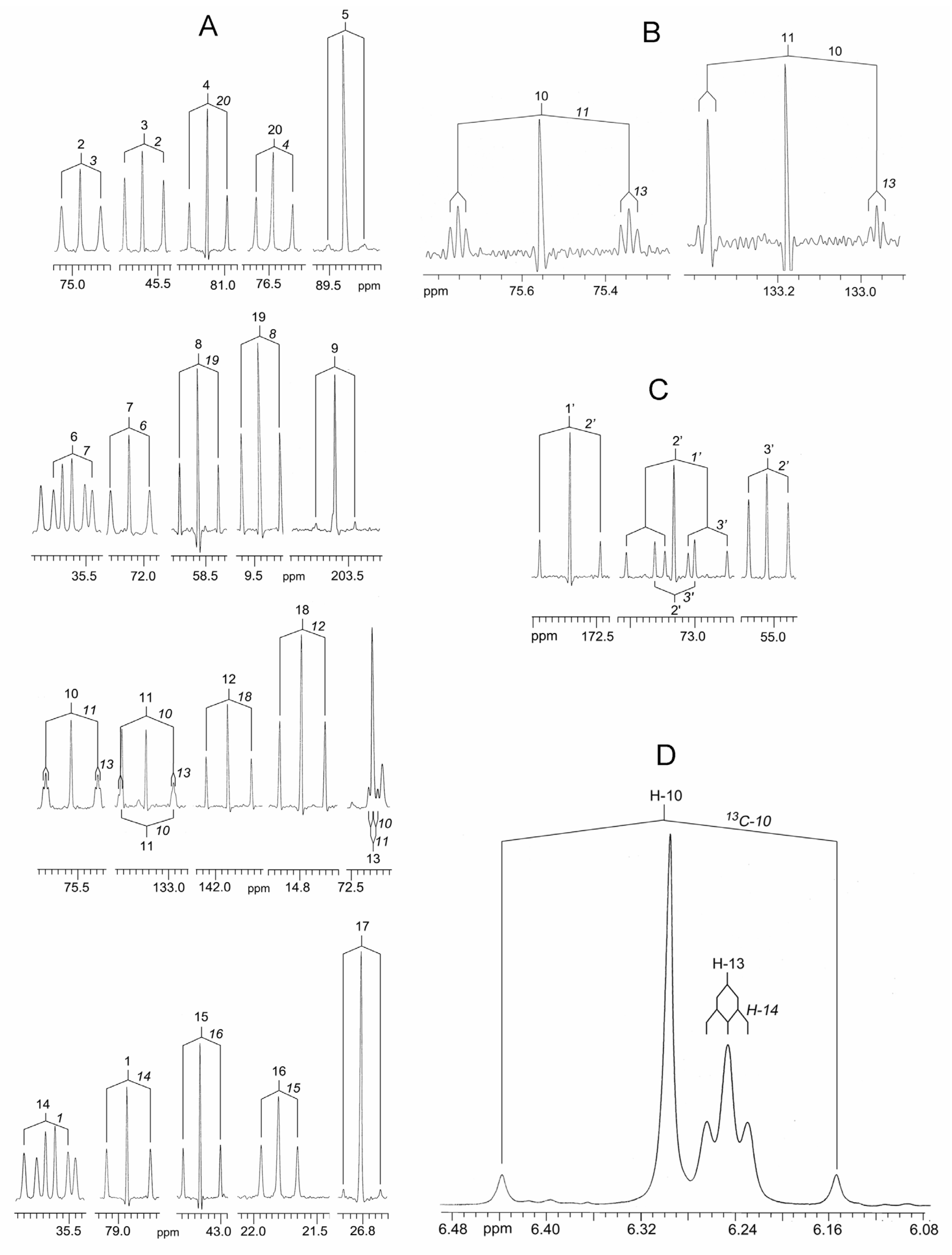

Figure 2.11: ^{13}C NMR signals of taxol obtained by the experiment with *Taxus chinensis* cell culture in B5 medium supplemented with [U-$^{13}C_6$]glucose. Signals of the taxane skeleton (A) and the 1’-3’ propane moiety (C) can be assigned. Signals 10 and 11 in panel A show remote $^{13}C^{13}C$ coupling to C-13 (expansion in panel B). ^{13}C abundances of the carbon positions in taxol were calibrated *via* $^{1}H^{13}C$ coupling of signal H-10 in the ^{1}H NMR spectrum (D).

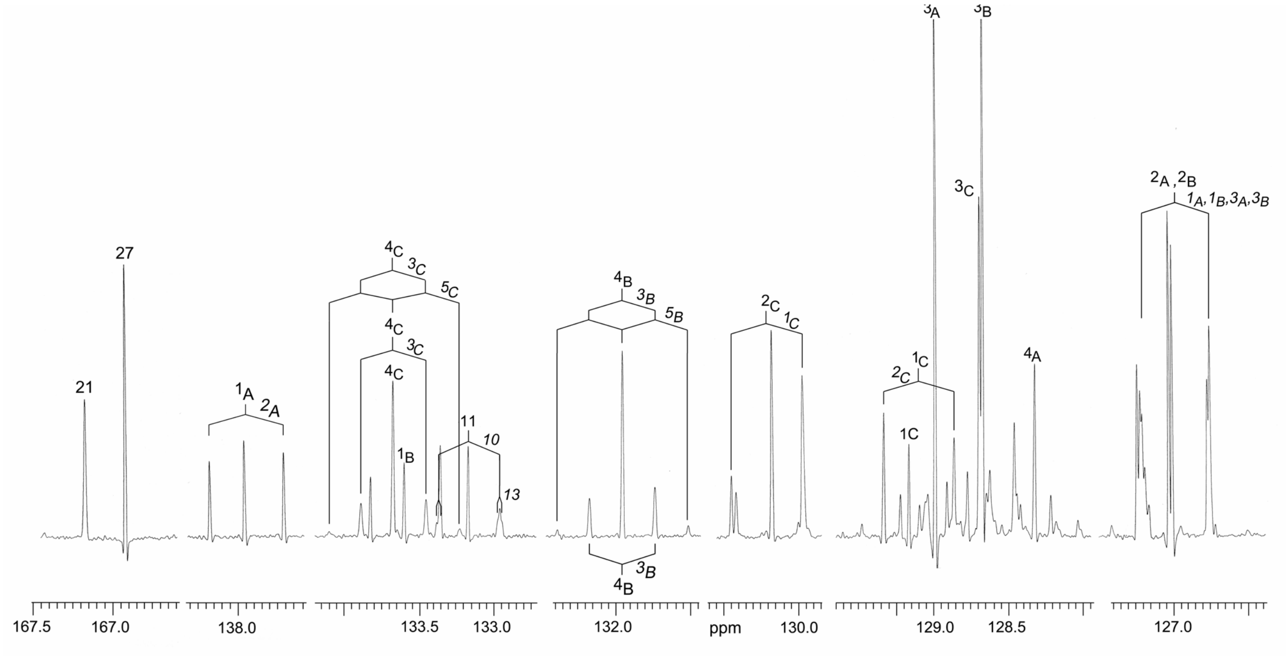

Figure 2.12: High resolution ^{13}C NMR spectra and partial assignment of the benzyl side chains A, B and C of taxol.

The taxane skeleton is formed from geranylgeranyl pyrophosphate (Lin et al., 1996; Eisenreich et al., 1996). Geranylgeranyl pyrophosphate is assembled from IPP and DMAPP units. The experiment with *Taxus chinensis* cell culture in B5 medium supplemented with [1-$^{13}C_1$]glucose affords quantitatively similar labels at the C-1 and C-5 position of all isoprenoid moieties. Biosynthesis of IPP and DMAPP *via* the mevalonate pathway would result in labels at positions C-2, C-4 and C-5 (Figure 2.14 B) of the molecules.

Quantitative predictions of ^{13}C-labeling patterns of key metabolites were derived from the labeling patterns of alanine, leucine and valine as described (see chapter 2.2 and 2.4.1, cf. Figure 2.14). Acetyl-CoA is deduced from alanine, hydroxyethyl-TPP from leucine and valine and erythrose 4-phosphate from phenylalanine and tyrosine. Quantitative estimations of labeling patterns for IPP and DMAPP *via* the mevalonate pathway (Figure 2.14 A) or the DOXP pathway (Figure 2.14 B) can be established and compared to the observed labeling patterns in the taxane moiety. For the experiment with [1-$^{13}C_1$]glucose, labels at C-1 (13.2 ± 0.2 mol %) and C-5 (13.6 ± 0.6 mol %) of IPP or DMAPP are in good agreement with the experimental findings showing an abundance of 14.0 ± 0.6 mol % at C-1 of IPP and 13.4 ± 0.8 mol % at C-5 of IPP as well as 14.0 mol % for C-1 of DMAPP and 14.4 mol % for C-5 of DMAPP (Figure 2.14).

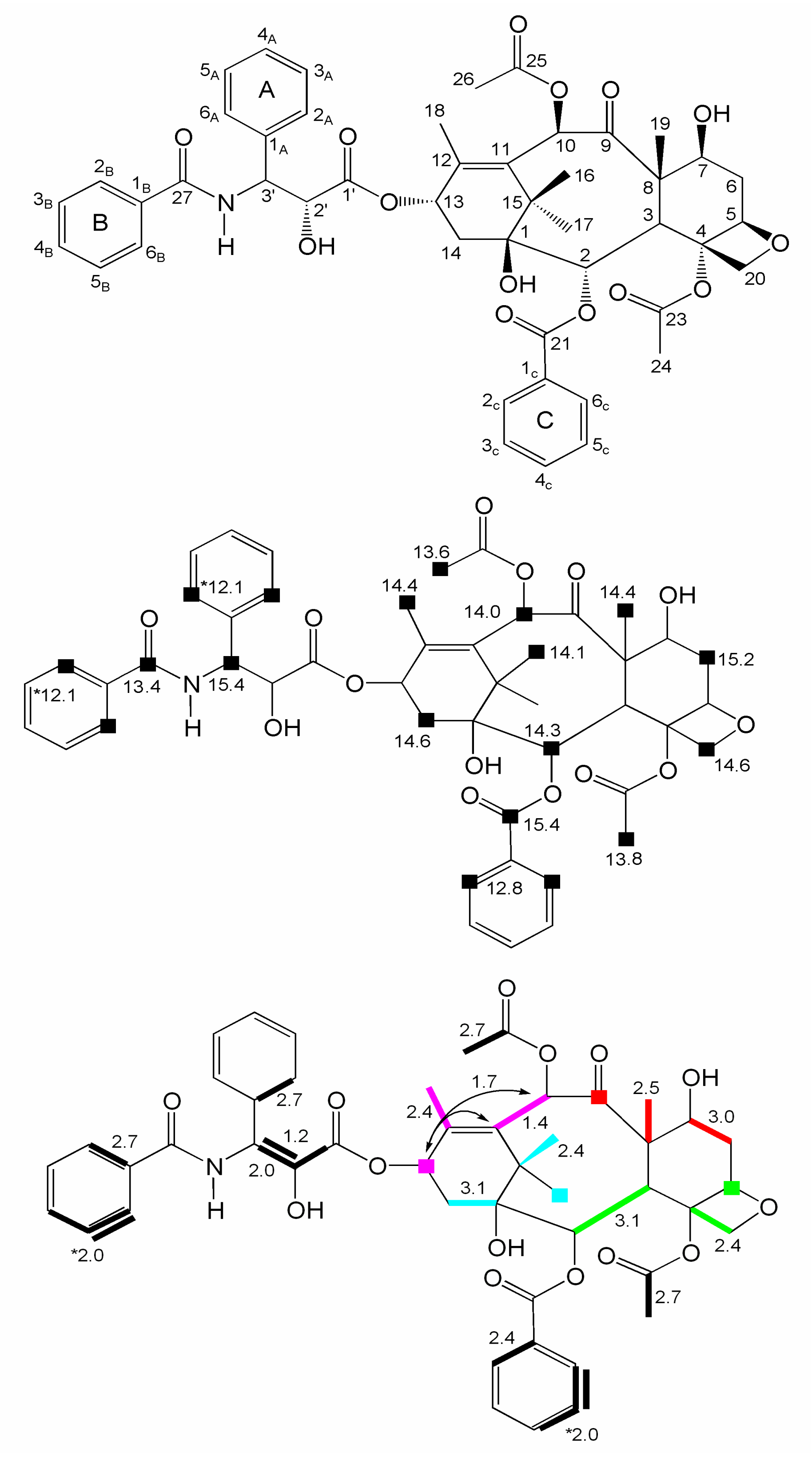

Figure 2.13: Carbon numbering and ^{13}C-labeling patterns of taxol from *Taxus chinensis* grown in B5 medium supplemented with [1-$^{13}C_1$]glucose or [U-$^{13}C_6$]glucose, respectively. NMR signatures with ^{13}C abundances (mol %) are drawn bold. The isoprenoid units forming the taxane skeleton are highlighted.

The experiment with [U-$^{13}C_6$]glucose confirms this result. Acetyl-CoA shows a $^{13}C_2$-signature with 2.3 ± 0.1 mol %. Predictions for 1-deoxy-xylulose 5-phosphate biosynthesized *via* the DOXP pathway would result in a $^{13}C_2$ unit at C-3/C-5 with 2.3 ± 0.1 mol %. C-1, C-2, and C-4 would show a $^{13}C_3$ signature of 2.3 ± 0.1 mol % and a $^{13}C_2$ signature of 1.3 ± 0.2 mol %. The experimental ^{13}C labels of IPP and DMAPP of the taxane skeleton correspond qualitatively and quantitatively to the ^{13}C-labeling patterns calculated for 1-deoxy-xylulose 5-phosphate using the retrobiosynthetic analysis approach on the basis of amino acid data. For IPP units, the $^{13}C_2$ unit at C-3/C-5 shows an abundance of 2.4 ± 0.1 mol %. The ^{13}C labels at C-1, C-2 and C-4 are not unequivocally resolved in the ^{13}C NMR spectrum, but the sum of the $^{13}C_2$ and $^{13}C_3$ signatures can be specified to 3.1 ± 0.1 mol %. In the DMAPP unit of taxol, the ^{13}C abundances correspond to the values calculated for IPP. ^{13}C labels of DMAPP at positions C-1, C-2 and C-4 are resolved in the ^{13}C NMR spectrum (Figure 2.14 B). The $^{13}C_2$ isotopolog of DMAPP shows an abundance of 1.4 mol % and the $^{13}C_3$ isotopolog of DMAPP shows an abundance of 1.7 mol %. All deviations are in the range of the exactness of the applied analytical technique. Therefore, it can be assumed that all isoprenoid moieties are generated from the DOXP pathway (Figure 2.13 and 2.14 A, for additional information on this conclusion, see chapter 2.4.1 and Eisenreich et al., 1996). The mevalonate pathway can be excluded from further analysis.

Acetyl moieties at positions O-4 and O-10 of the taxane skeleton show a $^{13}C_2$ with an abundance of 2.7 mol % (Figure 2.13). It can be concluded, that they were derived directly from acetyl-CoA.

The benzyl ring systems A, B, and C cannot be evaluated completely due to limitations in the resolution of the ^{13}C NMR experiment (Figure 2.13 and 2.14). In the shikimate pathway, the $^{13}C_2$ signature of the phosphoenolpyruvate precursor can be estimated to 2.5 ± 0.2 mol %, which is in good agreement with the acetyl moieties at O-4 and O-10 of taxol. From the erythrose 4-phosphate precursor only an overlapping $^{13}C_2$ and $^{13}C_3$ labeling pattern of the A and C ring system of taxol can be assigned (Figure 2.13) and estimated to 2.0 mol %.

From the amino acids alanine, phenylalanine and aspartate, obtained from the experiment with [U-$^{13}C_6$]glucose, the intermediary carbon fluxes in *Taxus chinensis* cell culture can be partially reconstructed by applying the rules illustrated in chapter 2.2 to a virtual triose phosphate moiety. Only circular flows can be estimated, since no balancing of respiration processes is included in the experiment. Briefly, the flux *via* glycolysis/glucogenesis *versus* pentose phosphate pathway can be reconstructed from resolving the ^{13}C-labeling patterns of the [1,2,3-$^{13}C_3$]triose phosphate unit and [2,3-$^{13}C_3$]triose phosphate unit, which derive directly from the incorporated [U-$^{13}C_6$]glucose proffered to the plant cells.

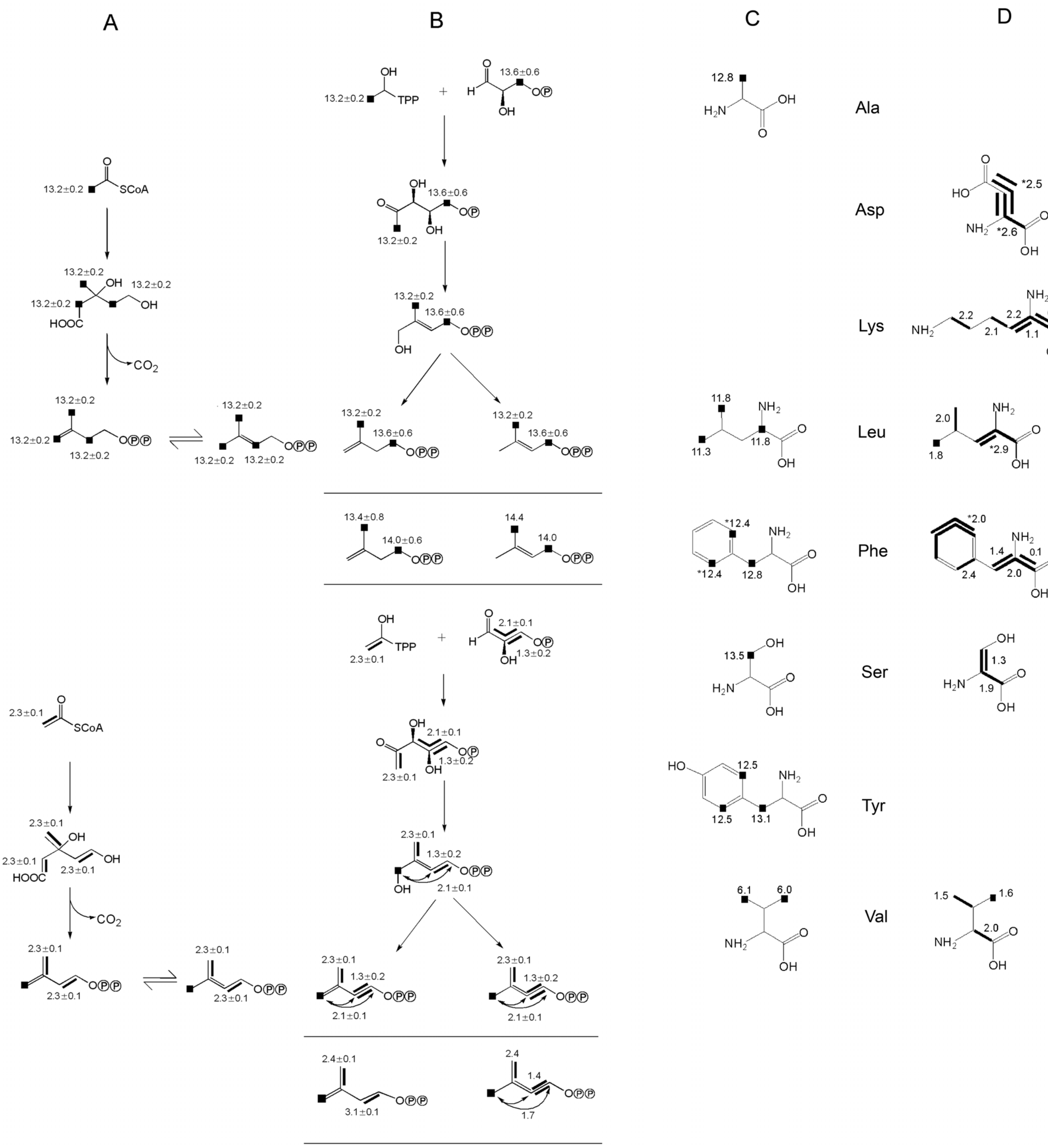

Figure 2.14: Intermediates of the mevalonate pathway (A) and the DOXP pathway (B) leading to IPP and DMAPP were reconstructed from isolated amino acids obtained from the experiment with [1-$^{13}C_1$]glucose (C) or [U-$^{13}C_6$]glucose as tracer. Comparison with isoprenoid moieties in the taxane skeleton shows qualitatively and quantitatively similar labeling patterns.

The abundance of [1,2-$^{13}C_3$] show the relative flux contributions from glucogenesis *via* oxaloacetate. The abundances of single-labeled isotopolog species can be computationally resolved using the 4F algorithm (cf. chapter 2.2 and 3.6), and attributed to specific fluxes. Flux rates *via* the shikimate pathway cannot be estimated because of the non-resolved $^{13}C_4$ signature of the erythrose 4-phosphate precursor in the aromatic rings of taxol, or phenylalanine, or tyrosine, respectively.

Exogenous glucose taken up by *Taxus chinensis* cells grown on B5 medium is primarily metabolized *via* the pentose phosphate pathway (64 ± 7 %, Figure 2.15). Direct flux *via* glycolysis occurs only to 36 ± 7 % of the glucose molecules (Figure 2.15). The backward flux of glucose metabolized through the citric acid cycle can be estimated to be 13 ± 2 %.

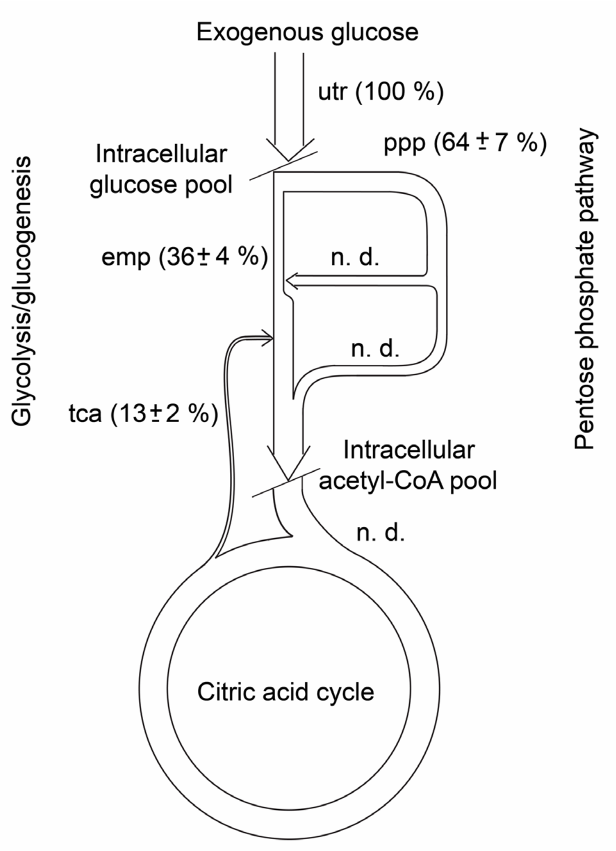

Figure 2.15: Carbon fluxes in *Taxus chinensis* cell culture on B5 medium. The uptake rate of exogenous glucose (utr) is set to 100 %. The intracellular glucose pool is divided by about a third be consumed *via* glycolysis (emp) and two third being consumed *via* the pentose phosphate pathway (ppp). The reflux from the citric acid cycle to the triose pool can be estimated to 13 %.

3. Glucose catabolism elucidates central pathways

Glucose is one of the most central metabolites in biological systems. It is generated by photosynthetic activity of plants, algae and some procaryots and be derived from formaldehyde in abiotic systems (Miller, 1987). Glucose serves as primary energy source for the generation of ATP in most known organisms. Polymers of glucose and glucose derivates are all abundant nature: cellulose, glucosan, chitin, starch and glycogen are the predominant examples (for reviews, see Hills, 2004; Doblin et al., 2003; Merzendorfer and Zimoch, 2003). The key position of glucose in metabolic networks enables a perturbation created by ^{13}C-enriched tracer molecules to spread through all intermediary and secondary metabolic processes. Stable-isotope perturbation experiments for secondary pathways have been illustrated in chapter 2. This chapter focuses on ^{13}C perturbation experiments in catabolic intermediary metabolism using glucose as tracer and analyte.

In order to simplify the following discussion and to highlight the special character of the glucose molecule in experimental setups, a special notation for glucose is introduced. The topology of the glucose molecule is designated in a six digit code. The symbol set of this code contains a binary element {0 for ^{12}C and 1 ^{13}C} and a wildcard symbol {X for ^{12}C or ^{13}C}. The six digits represent the carbon atoms of glucose in descending order relative to the magnitude of the number position: the first digit represents the carbon atom at position 1, the second digit represents the carbon atom at position 2, etc. For example, glucose with ^{13}C in the positions 1 and 2, with ^{12}C in the position 3, and with ^{12}C or ^{13}C in the positions 4, 5 and 6 will be designated {110XXX}. Groups of isotopologs in this notation will subsequently be called X groups. Since each position can be occupied by 0, 1, or X, the total number of glucose X groups is 3^6 eq. 729.

3.1 Isotopolog space of glucose

The dimensionality of an isotopolog space *S* of a compound with *n* carbon atoms can be expressed by $S = 2^n$. In case of ^{13}C-labeled glucose, the state space encloses 64 dimensions. The information density and complexity of an isotopolog space can be estimated by its intrinsic information entropy (cf. chapter 3.5).

Isotopologs with a ^{13}C abundance lower than the noise level of a system can help to condense the information contained in a system. The isotopolog space can be reduced to

$S = 2^n - y$, with *y* being the magnitude of non-abundant isotopologs. The compressed space increases the information density by compression ratio *C*. The Huffman theorem (for additional information, see Huffman, 1952; Gilbert, 1995) expresses *C* by equation $C = \frac{\ell(n^2)}{\ell(H/S)}$, with ℓ as optimal encoding alphabet. ℓ must provide uniqueness to the words generated from the alphabet (i. a., glucose isotopologs) and can be spawned by the Huffman tree algorithm (Huffman, 1952). The X group notation represents a ternary alphabet and serves as optimal code set for an arbitrary isotopolog space (cf. Appendix B).

Each information of the isotopolog space is unique in terms of analytical interpretation. Redundances can occur in a metabolic interpretation of isotopolog spaces, since generally more than one isotopolog can be assigned to a single metabolic process. Therefore, analysis of metabolic networks on the basis of isotopolog data compacts the information content further. If the number of metabolic pathways shaping the analyte is smaller than the magnitude of its isotopolog space, an overdetermined equation system can be created for estimating metabolic fluxes in the experimental system.

The isotopolog state of glucose in 64 dimensions has to be reconstructed using constraints in the form of X groups obtained from analytical methods, namely ^{13}C NMR spectroscopy and mass spectroscopy. ^{13}C NMR spectroscopy generates a maximum of 57 constraints (cf. Appendix A.3.2). In experimental setups with a low abundance of proffered [U-$^{13}C_6$]glucose, this number can reduce to only 26 to 37 constraints (see examples in chapter 3.7). Furthermore, about a third of these constraints are redundant: many general constraints can be reconstructed from more basic constraints. For example, {1XXXXX} can be reconstructed from {10XXXX} and {11XXXX}, also {1000XX} can be reconstructed from {1000X0} and {1000X1}. Mass spectrometry can provide seven X groups containing all possible isotopomers groups of glucose, including [U-$^{12}C_6$]glucose (Previs and Brunengraber, 1998). It has to be noted that this number may increase, if defined fragmentation patterns of the glucose molecule can be obtained and evaluated (for review, see Kelleher, 1999).

It can be concluded, that the reconstruction of 64 dimensions of the glucose isotopolog space is not possible in a direct approach. The equation system is not sufficiently determined for deriving a unique solution. Using only low abundances of [U-$^{13}C_6$]glucose *versus* glucose with natural abundance, the glucose space can be restricted to 22 dimensions assuming the following scenario: bonds between adjacent ^{13}C atoms are broken down and the resulting fragments can be applied as building blocks in anabolic counterprocesses. If the share of the tracer compound is small relative to its unlabeled counterpart the assembly of metabolites emerging from fragments of the tracer compound is also small. Thus the breaking of ^{13}C^{13}C bonds is dominant over the formation of new ^{13}C^{13}C bonds. This dominance of catabolic processes in experimental systems with the illustrated preconditions will be referred as 'dissociative conjecture'.

The dissociative conjecture compacts the isotopolog space from 64 to 22 dimensions enclosing six isotopolog species with one ^{13}C atom, five isotopolog species with two ^{13}C atoms, four isotopolog species with three ^{13}C atoms, three isotopolog species with four ^{13}C atoms, two isotopolog species with five ^{13}C atoms, one isotopolog species with six ^{13}C atoms and one isotopolog species with six ^{12}C atoms (Figure 3.1). For an arbitrary compound, the dissociative conjecture restricts the isotopolog space from $S = 2^n$ to $S = \sum_{i=0}^{n} i$.

Applying the dissociative conjecture to glucose, a number of isotopologs can be determined without computational deconvolution of X groups accessible by ^{13}C NMR spectroscopy. All single ^{13}C-labeled species can be obtained directly: {100000} results from {10XXXX}, {010000} results from {010XXX}, {001000} results from {X010XX}, {000100} results from {XX010X}, {000010} results from {XXX010} and {000001} results from {XXXX01}. The multiple ^{13}C-labeled glucose species {110000} can be derived from {110XXX} and the quantity of [U-^{13}C$_6$]glucose can be derived from {1X1X11}, {X111XX}, {XX111X} or {111XX1}. Experimental variations and redundancy in X groups can be used to determine the exactness of the calculated glucose isotopolog abundances and to define upper and lower error restrains.

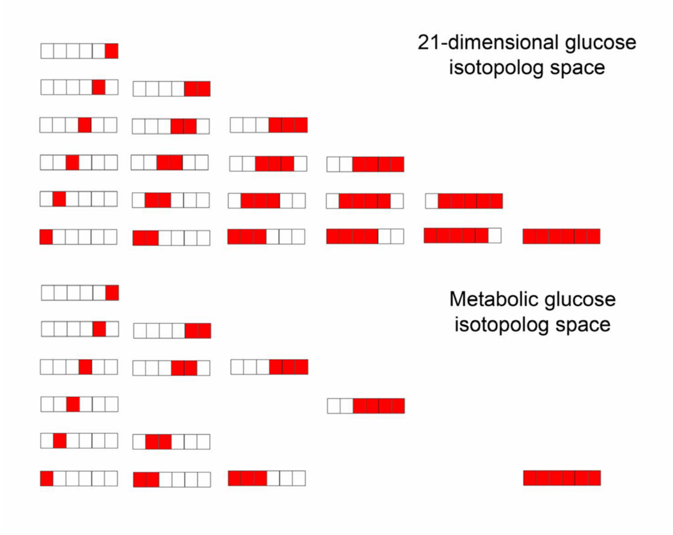

Figure 3.1: The dissociative conjecture compacts the glucose isotopolog space of glucose isotopologs containing at least one ^{13}C atom from 63 dimensions to 21 dimensions. Intermediary metabolic pathways generate only 15 of these 21 glucose isotopologs.

The dimensionality of the glucose isotopolog space resulting from the dissociative conjecture decreases further, if glycolysis/glucogenesis, pentose phosphate pathway and citric acid cycle (for additional information on the single pathways, see chapter 1.3) are assumed as dominant processes in the metabolic network. Only 15 of the 22 possible isotopologs are able to be generated by these processes (Figure 3.1). However, multiple cycling in the metabolic network must be comprised in the interpretation. It can be deduced that intermediary catabolic networks can be completely elucidated using ^{13}C NMR spectroscopy, even if {000000} is not accessible by this technique. Mass spectroscopy yields only seven isotopomer constraints (without information from defined fragmentation events), and cannot provide a sufficient number of constraints to deconvolute even the smallest glucose isotopolog space required for metabolic interpretation.

3.2 General experimental setup

The organism under investigation is cultured on media containing [U-^{13}C$_6$]glucose and glucose with natural ^{13}C abundance in excess. The organism is harvested after an incorporation period. In this time, the proffered [U-^{13}C$_6$]glucose perturbates the natural equilibrium of ^{12}C to ^{13}C in the experimental system. The analyte glucose is isolated by water/methanol extraction. Glucose from the crude extract is purified by ultracentrifugation, filtration and reverse phase HPLC (for details, see chapter 5.3.1). Water is removed under reduced pressure. The glucose is dissolved in D_2O. The final glucose concentration for high-resolution ^{13}C NMR should be more than 5 mg/ml (Figure 3.2) to allow quantitative evaluation. ^{1}H NMR spectroscopy is applied for verifying the sample purity and to calculate the total ^{13}C abundance of C-1β of glucose using heteronuclear ^{1}H^{13}C couplings of the H-1β signal. The ^{13}C abundances of all other carbon atoms of the glucose molecules are cross-calibrated to this value. The use of heteronuclear ^{1}H^{13}C couplings of the H-1α signal is also possible, but is afflicted with a lower signal-to-noise-ratio. Alternatively, the total ^{13}C abundance of glucose can be obtained with high accuracy by isotope ratio mass spectrometry (chapter 5.6). IRMS data can be used for cross-calibration of the total ^{13}C abundance instead of the evaluation of heteronuclear ^{1}H^{13}C signal satellites in the ^{1}H NMR spectrum.

On the basis of the ^{13}C NMR spectrum, showing 12 signal sets disparted in multiplets for the α- and β-anomers of glucose, X groups are calculated by the ratio of the relative integral of each respective satellite set relative to the integral of the total signal (cf. chapter 1.1.3; for examples, see chapter 3.7). Signal assignments of α- and β-glucose were taken from Glawischnig et al., 2002. Coupling constants and isotope shifts were determined from

^{13}C NMR spectra of commercially available ^{13}C-labeled glucoses (Glawischnig et al., 2002; Eisenreich et al., 2004; cf. Appendix A.3.2). Appendix A.3.2 assigns each ^{13}C NMR signal of glucose to its respective X group.

The X groups are computationally deconvoluted to their single isotopologs by a genetic algorithm. The isotopolog data can be applied to simulate the intermediary metabolic pathways and to determine the distribution of glucose cycling events in the metabolic network. The *in silico* deconvolution of the glucose isotopolog space with X groups and the metabolic flux simulation will be further described in chapters 3.3 and 3.6.

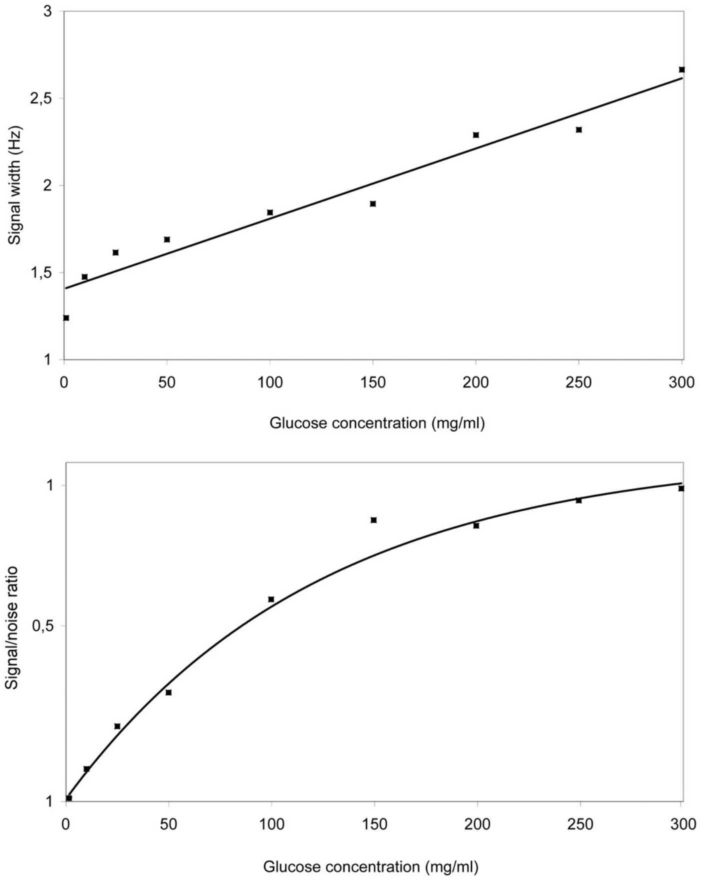

Figure 3.2: Quality of ^{13}C NMR spectra depends on the amount of isolated glucose. The more glucose available from the sample, the less error occurs in X group determination and deconvolution. The functions were determined using by analysis the C-2 signal of [1-^{13}C$_1$]glucose in a ^{13}C NMR spectrum (100 pulses). Signal line width should be minimal to obtain the highest possible resolution in the ^{13}C NMR experiment. The signal/noise ratio should be maximal. The curves can be approximated by $\mathrm{Y} = \alpha \cdot c + \beta$ and $Q = 1 - e^{-c \cdot \delta}$. *Y* is the signal width, *Q* the signal/noise ratio, *c* the glucose concentration in mg/ml and α, β and σ are experimental parameters. In the given setup α eq. 0.004, β eq. 1.406 and σ eq. 0.008.

3.3 Deconvolution of isotopolog groups

X groups form local constraints of the ^{13}C isotopolog space of glucose from a specific experimental setup. The complete isotopolog space can be obtained by resolving these local constraints and reconstruct the ^{13}C abundances for all possible isotopolog species of glucose. For an exact analytical technique with no background noise, the single isotopologs would result from the resolution of a linear equation system by the Gaussian algorithm or a QR-deconvolution (Press et al., 1992). All experimental techniques add background noise to the real signal intensity and reshape the problem fundamentally. The deconvolution has to be formulated as search problem, to find the best fit between a glucose isotopolog space allocation and the X group constraints derived from an experimental setup. The complexity of this deconvolution is dependent on the background noise level and the degree of inconsistence between X groups. Problems occurring in the course of a deconvolution algorithm include performance, reliability, error tolerance and the ability of finding a global minimum along local minima (Press et al., 1992; for examples, see Figure 3.3).

For catabolic processes, several deconvolution algorithms were tested (Table 3.1). Analytical algorithms were found insufficient to find a fit between glucose X group constraints and a set of simulated ^{13}C isotopolog abundances. Steepest descent, Newton algorithm and line search techniques found unstable solutions in repetative iteration processes, probably through multiple minima in the target functions (Figure 3.3). The number of local minima can grow exponentially to the magnitude of perturbations created by X group inconsistencies and background noise. Conjugated gradient methods were more suitable. Oscillation and a varying identification of local minima could be reduced (Figure 3.3). Genetic algorithms were identified as best computational technique for the X group deconvolution process. The advantages and disadvantages of the various methods are summarized in Table 3.1.

Genetic algorithms for computational optimization problems of arbitrary target functions were introduced by Holland, 1975 and Rechenberg, 1973. The formal implementation of the deconvolution process using a genetic algorithm is described in chapter 6.3.4.

The algorithm can be divided into six steps (Figure 3.4): (I). Isotopologs define a binary occupation matrix D. D_{ij} is set to 1 if an isotopolog S_i is contained in X group X_j and is set to 0 if isotopolog S_i is not contained in X group X_j. (II). Vector T_i encoding the ^{13}C abundances (mol %) for a given isotopolog $S_{i,}$ is initialized with random values. (III). Using formula $H_{ij} = D_{ij} \cdot T_i$, a state matrix H is generated. This matrix yields the ^{13}C abundances for X groups associated with the current vector T_i. (IV). The fitness function $F(T_i)$ evaluates the difference of the X groups resulting from H_j and the X groups resulting from the ^{13}C NMR experiment, $F(T_i) = \sum_j \left| X_j - \sum_i H(T_i)_{ij} \right|$. (V). T_i is changed by applying a genetic algorithm (for the detailed parameter set, see chapter 6.3.4). (VI). The steps (III) to (V) are repeated until *F* converges to a minimum after n iterations.

Vector T_n contains the final ^{13}C abundances of each glucose isotopolog carrying at least a single ^{13}C atom. The isotopolog {00000} can be trivially obtained through calculating the remaining share with respect to the sum of all glucose isotopolog species containing one or more ^{13}C atoms.

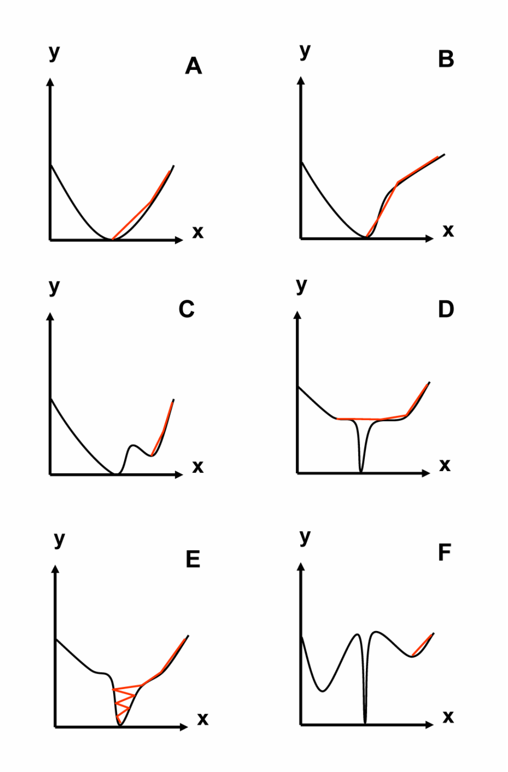

Figure 3.3: Examples for specific problems of finding a global minimum by computational methods (course of optimization drawn in red) in arbitrary two-dimensional functions (drawn in black): A. Normal descent in a convex function. B. Descent in a non-convex but non-pathological function. C. Non-convex function with two minima. A simple descent along the steepest path results in finding only local minima. D. Function with tight minimum. If the step-width of descent is fixed, the minima will be ignored by the algorithm. E. Wrong adaptation of the step-width in the course of descent can lead to oscillations in the deconvolution process. F. Pathological function with multiple minima and a tight global minimum surrounded by a barrier. Most gradient-based optimization algorithms will result in finding one of the local minimum.

Optimization algorithm	**Advantage**	**Disadvantage**
Steepest descent	➢ Performance ➢ Meaningful information can be obtained from the deconvolution course using a distinct target function	➢ Requires convex functions (only one minima; maxima are at the border of the function domain) ➢ Steepest descent may not be the most direct pathway to the minima ➢ Requires external parameter α for stepwidth ➢ Slow on regional plateaus ➢ Descent tends to oscillate easily
Newton algorithm	➢ Descent goes directly in the direction of the minimum	➢ Complex calculation of the Hesse matrix ➢ Requires matrix inversion ➢ Unstable at transitions to non-convex target functions
Line search	➢ Parameter α is implicitly set by the algorithm	➢ Descents in zigzag course, even for plain functions
Conjugated gradient descent	➢ Performance ➢ Stable against oscillation	➢ Unreliable in non-convex environments
Genetic algorithm	➢ Performance ➢ Local and global minima can be found in functions with multiple minima ➢ Stable in complex target functions ➢ Parallelization possible	➢ Heuristic approach ➢ Quality of a minimum can only be assessed after repetitive search process

Table 3.1: Optimization algorithm evaluated for deconvolution of isotopolog space restricted by the dissociative conjecture (see chapter 3.1). Steepest descent, Newton algorithm, line search and conjugated gradient descent are analytical approaches (chapter 6.3). Genetic algorithms are intrinsically probabilistic, but have been found suitable to optimize ^{13}C abundances of glucose isotopologs to fit experimentally obtained X groups constraints.

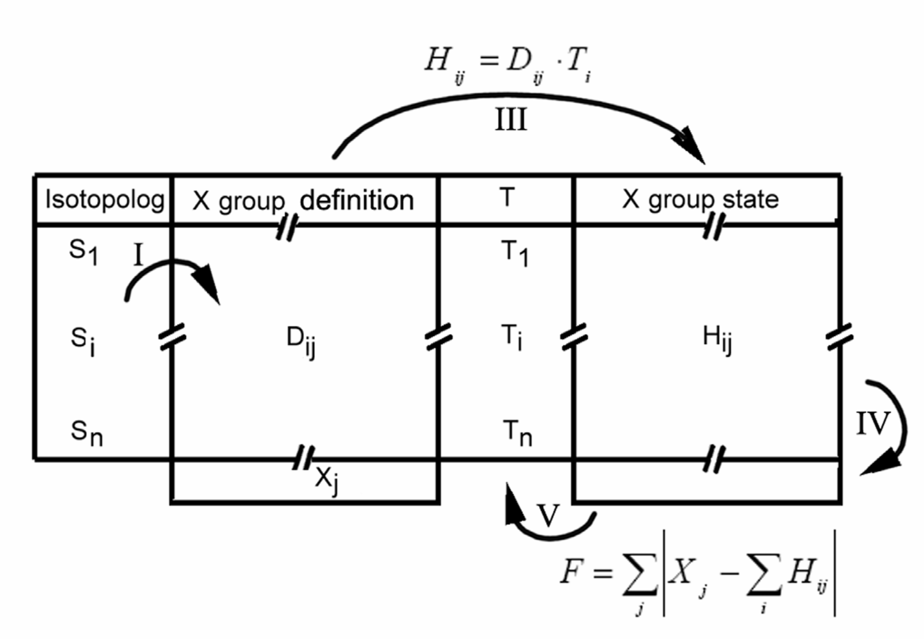

Figure 3.4: Graphical presentation of the interaction scheme implemented for the deconvolution of X groups *via* a genetic algorithm (for details on process phases and symbols, see text).

3.4 Isotopolog signatures arising in catabolic networks

The glucose isotopolog space restricted by the dissociative conjecture (chapter 3.1) results in 22 isotopologs. Isotopolog {000000} is not accessible by ^{13}C NMR spectroscopy, but can be calculated from the sum *T* of glucose isotopolog species carrying one ore more ^{13}C atoms, $\{000000\} = 1 - T$. Glycolysis/glucogenesis, pentose phosphate pathway and citric acid cycle can yield 14 possible ^{13}C-labeling patterns of glucose. These 14 isotopologs can be divided into 6 single-labeled glucose isotopologs and 8 multiple-labeled glucose isotopologs. Most single ^{13}C-labeled isotopologs as well as several multiple labeled isotopolog are generated by multiple cycling events. Futile metabolic cycles cannot be estimated, since they will not transform the ^{13}C signature of the glucose isotopologs. The metabolic signatures arising from catabolic processes can be interpreted for each of these 14 isotopolog species.

{111111} is created by direct incorporation of glucose molecules into tissues or storage forms (e. g. glycogen in plants and animals or starch in plants).

{111000} and {000111} are created by glucogenesis of [U-$^{13}C_3$]triose phosphate units, which result from the aldolase reaction in glycolysis/glucogenesis or in the non-oxidative branch of the pentose phosphate pathway.

{110000} and {001111} are created in the non-oxidative branch of the pentose phosphate pathway by the transketolase reaction. Entering the pathway with [U-$^{13}C_6$]glucose 6-phosphate, it creates a universally labeled erythrose 4-phosphate moiety. Combination with an unlabeled C_2 unit yields {001111} (Figure 3.5 A). Accordingly {110000} can arise by combination of a ^{13}C-labeled C_2 unit with an unlabeled C_4 unit (Figure 3.5 A). Cycling *via* the transaldolase reaction continually transforms {001111} to {001000} and {000111}, but leaves {110000} intact. {001000} can be rearranged to {000100} during glycolysis/glucogenesis by catalytic action of the enzymes aldolase and triose phosphate isomerase.

{000011} can arise from {110000} by glycolysis and glucogenesis, which leads to a C-3/C-4 symmetrical mirroring of the ^{13}C occupied isotopolog positions. Since {000011} and {110000} are generally not abundant in the same ratio, the interplay of these pathways is not in equilibrium *in vivo*.

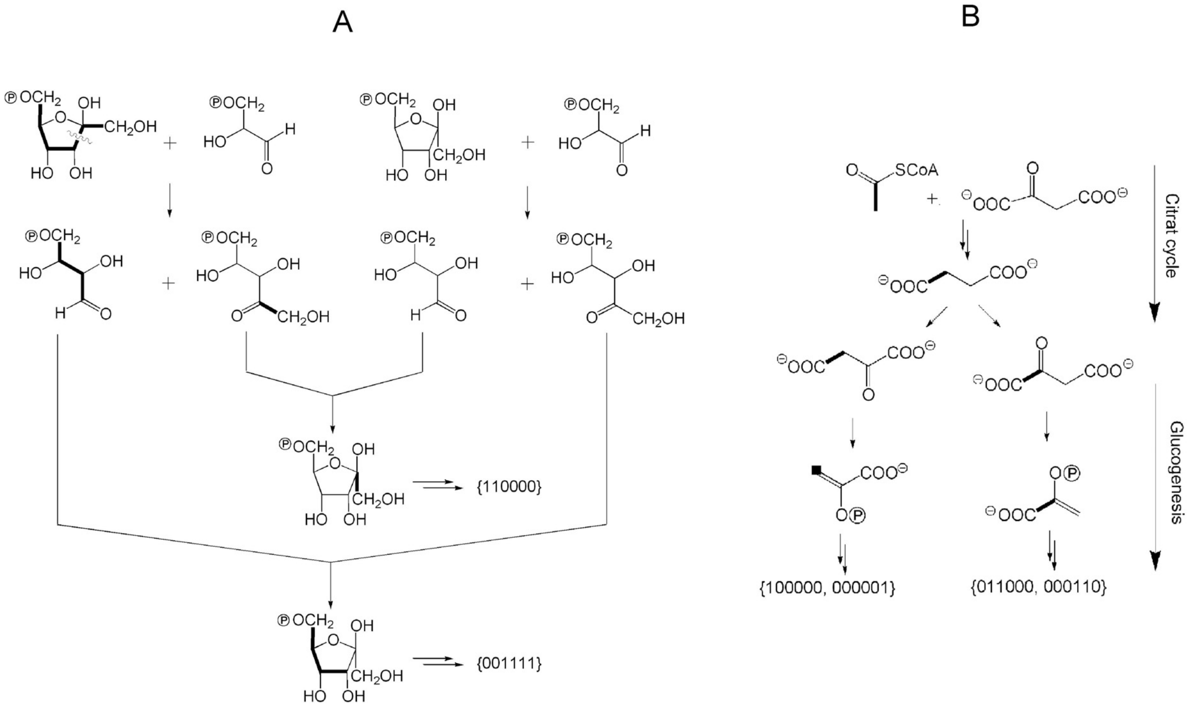

Figure 3.5: A. Formation of {110000} and {001111} glucose phosphate by catalytic action of the transketolase enzyme in the non-oxidative branch of the pentose phosphate cycle. Bonds in bold type connect ^{13}C atoms in multiply ^{13}C-labeled isotopologs derived from [U-$^{13}C_6$]glucose. B. Formation of {100000}, {000001}, {011000}, {000110} through the citric acid cycle, or the Cori cycle, or the proline cycle, respectively.

{011000} and {000110} can be created by glucogenesis of [1,2-$^{13}C_2$]glyceraldehyde 3-phosphate and [1,2-$^{13}C_2$]dihydroxyacetone 3-phosphate. The ^{13}C-labeled triose 3-phosphate compound can be generated from the citric acid cycle, or the Cori cycle, or the proline cycle. The isotopologs {111111}, {111000}, {000111}, {110000} and {000011} are precursors in the formation of [1,2-$^{13}C_2$]- and [3,4-$^{13}C_2$]oxaloacetate *via* the citric acid cycle. The posphoenolpyruvate carboxykinase may subsequently convert [1,2-$^{13}C_2$]oxaloacetate to [1,2-$^{13}C_2$]phosphoenolpyruvate, which can form {011000} and {000110} through glucogenesis (Figure 3.5 B). The same series of reactions starting from [3,4-$^{13}C_2$]oxaloacetate yields [3-$^{13}C_1$]phosphoenolpyruvate which can generate {100000} and {000001} (Figure 3.5 B).

{010000} and {000010} can be generated by multiple cycling of glucose molecules in the metabolic network. Passage of [U-$^{13}C_6$]glucose or [1,2,3-$^{13}C_3$]triose moieties through the citric acid cycle with subsequent passage of the fragments through the transketolase reaction in the pentose phosphate cycle generates these isotopologs.

{001100}, {011100}, {001110}, {111100}, {011110}, {111110}, {011111} are not generated by any central metabolic pathway, if the isotopolog transformations of the glucose molecule dominated by dissociative processes.

3.5 Entropy of isotopolog spaces

The entropy *H* of isotopolog spaces is defined using probability p_i of the occurrence of a distinct isotopolog E_i of a compound within the experimental system and the magnitude of metabolic processes $\Omega(E_i)$ which the isotopolog must have passed (for additional information on the theory of information entropy, see Shannon, 1948). $\Omega(E_i)$ is assessed by exhaustive mapping of metabolic reactions leading to single and multiple ^{13}C-labeled isotopolog species. $\Omega(E_i)$ must be normalized by the total number of distinct metabolic circuits, which define the network topology of the experimental system. The specific probability $p_i, p_i \in [0;1]$ for a single isotopolog can be estimated by knowledge of the complete isotopolog space and calculating the probability of appearance of a distinct isotopolog with respect to the isotopolog set.

$$H = \left(-\sum p_i \cdot \log_2 \Omega\right)^{-1}$$

H can be applied to evaluate the complexity of a metabolic network (cf. chapter 3.4) and to cluster information obtained from experimental isotopolog spaces (possible cluster algorithms are UPGMA, chapter 6.5.4, and NJ, chapter 6.5.4).

H is equivalent to the average information density of the isotopolog space (cf. Peters, 1967). This can be illustrated by the following example: if p_i becomes 1, it means that an isotopolog species will occur with 100 % probability. This isotopolog would describe the complete metabolic system and no additional information can be obtained from the remaining isotopolog space. Probability p_i and $\Omega(E_i)$ will maximize H in this case. In contrast, a small value of *H* (i. a., small information entropy of the system) can be interpreted as a small probability of occurrence for a distinct isotopolog distribution with a short metabolic history.

3.6 Metabolic flux simulation

Metabolic flux analysis describes mathematical models aiming at the simulation and analysis of catabolic and anabolic reaction networks. The resulting flux maps show the interplay of metabolites in organisms in quantitative terms and allow conclusions for biotechnological applications and genetic modifications.

For the simulation of circular flows, a rule-based approach was chosen. Existing functional-based solutions, which encode the inference algorithm in the form of differential equations have been found unable to resolve intrinsically circular problems (for details on evaluated algorithms and software packages, see chapter 6.4.1; for additional information on functional approaches, see Schmidt et al., 1997; Wiechert et al., 1997; Wiechert and de Graaf, 1997; Follstadt and Stephanopoulos, 1998; Schmidt et al., 1998; Fiaux et al., 1999; Park et al., 1999; Wiechert et al., 1999; Wiechert and Murzel, 2001; van Winden et al., 2002; information on design and modeling of circular problems can be found in Yourdon, 1989 and Powell-Douglass, 2004).

For the described experimental setup, using [U-$^{13}C_6$]glucose as tracer and glucose isolated from the biological system as analyte, the algorithm 4F has been designed and implemented to analyze metabolic cycling of [U-$^{13}C_6$]glucose in catabolic networks (chapter 3.1) enclosing the Emden-Meyerhoff-Parnas pathway, the pentose phosphate pathway and the citric acid cycle. The results for each pathway represent a spatial and temporal integration of the metabolic events transforming the glucose isotopolog space in the experimental system. The supported pathways can be discriminated by symmetric or asymmetric shuffling of isotopolog patterns, e. g. the 3:3 rearrangements in glycolysis/glucogenesis and the 2:4 rearrangements by the transketolase reaction. The software can estimate relative exchange fluxes between symmetric moieties in the pathways as well as forward and backward fluxes from asymmetric metabolic transformations.

Input variables are the ^{13}C excess E_i of the experimentally obtained isotopologs S_i. Furthermore the abundance of the ^{13}C-labeled glucose pool P available to the experimental system must be specified. E_i becomes normalized by P to avoid zero-flux effects (for details, see Roscher et al., 2000). The size of the pool of unlabeled glucose equals $1-P$. Output variables of the algorithms is the simulated ^{13}C excess O_i of each isotopolog. The isotopolog pattern generated by carbon fluxes through the defined metabolic pathways must minimize the cumulated deviation L between the input and output sets, $L=\sum_i P\cdot\left|E_i-O_i\right|$.

The 4F algorithm is implemented as .NET software service and available at http://xsystem.org.chemie.tu-muenchen.de. It simulates the molar ^{13}C excess of glucose isotopologs by inference of rules defined in chapter 6.4.2. The rules transform E_i by shuffling small amounts of glucose (iterator size φ) from pools of [U-^{13}C$_6$]glucose and glucose with natural ^{13}C abundance *via* the defined metabolic pathways. The pool of glucose with natural ^{13}C abundance should exceed the pool of [U-^{13}C$_6$]glucose by at least one order of magnitude. 4F rules are not commutative and are executed according to their priority: (I). Rules yielding asymmetric isotopologs have a higher priority than rules yielding symmetric isotopologs. (II). The execution priority of a rule increases with the number of distinct isotopologs, which can be generated by applying the rule on an isotopolog, e. g. [U-^{13}C$_6$]glucose.

For the simulation, a bottom-up approach is used. Glucose isotopologs derived from multiple metabolic cycling are generated by a forward feeding phase, while isotopologs derived from single transformations are generated by the subsequent slipstreams in the simulated network. According to this convention, ^{13}C excess of isotopologs with maximal metabolic pathway lengths is generated first by forward feeding of glucose to the simulated network. Slipstreams from P are subsequently refilling E_i from isotopologs with shorter metabolic pathways lengths. Symmetric isotopologs derived from single metabolic operations become balanced by asymmetric transfer reactions. Balancing operations of symmetric isotopologs are not included in the calculation of the total flux distribution, since they cannot be distinguished in the glucose isotopolog space (Figure 3.6).

The following examples illustrates the procedure: assume an arbitrary symmetric transformation T_1 generating E_{T1A} and E_{T1B} from P. Units of an iterator size φ are removed from P and are added to E_{T1A} and E_{T1B}. Repeating the shuffling n times leads to educt pool $P_{n+1}=P_0-n\cdot\varphi$ and product pool $E_{T1A,n+1}=E_{T1A,0}+n\cdot\varphi$ and $E_{T1B,n+1}=E_{T1B,0}+n\cdot\varphi$. The shuffling stops after the smaller component, E_{T1A} or E_{T1B}, is flooded. A second asymmetric transformation T_2 now generates E_{T2A} and E_{T2B} from E_{T1A}. The forward flux will iteratively remove $k\cdot\varphi$ from E_{T1A} until either E_{T1A} becomes zero or E_{T2A} and E_{T2B} are filled. E_{T2A} and

E_{T2B} will be the final possible state of the example (i. a., single-labeled glucose species). If E_{T1A} becomes zero before E_{T2A} and E_{T2B} are equilibrated, E_{T1A} or E_{T1B} become rebalanced and refilled by slipstream from P. Subsequently the forward flux to E_{T2A} and E_{T2B} is repeated until similarity with the target values is reached. If E_{T1A} is non-zero, a slipstream from P is triggered until E_{T1A} is balanced. If P becomes zero or exceeds defined deviation borders the simulation is terminated. The shuffling events n and k are now related to the total number of shuffling events S in the system. S is increased in parallel to n and k in the iteration phase, but is not increased by balancing operations of symmetric processes. Balancing of asymmetric processes *via* symmetric operations increases S. Thus the ratios $\frac{n}{S}$ and $\frac{k}{S}$ express the transition probability of glucose molecules for a distinct metabolic transformation.

The quality of the generated flux system can be calculated by the relative deviation between v isotopologs $\varepsilon = \sum_{j=1}^{v} \frac{E_j - I_j}{I_j}$. Furthermore the deviation of the pool size of [U-$^{13}C_6$]glucose proffered to the experimental system and the level of consumed glucose by the *in silico* simulation can be used for verification of the results.

Statistical analysis is required for estimating the error propagation of isotopolog data through the network. Isotopolog data from NMR experiments is intrinsically noisy and the statistical significance of the associated modulations of the metabolic networks has to be tested. For this test the sensitivity matrix is calculated, showing the strongest impact of single data records X on the numerically generated network. Calculating the covariance matrix V (chapter 6.2.1) from the sensitivity matrix Y with magnitude N, confidence values can be attributed to the computed flux values. Significance of the modulations can be tested by a χ^2 test for interconnected sets (chapter 6.2.3) or a t-test for non-interconnected sets (chapter 6.2.4).

The complexity of the system grows by square relative to the number of metabolic transformations encoded in the rule set of the simulation, since all rules are interconnected in the metabolic topology definition of the simulated biological system.

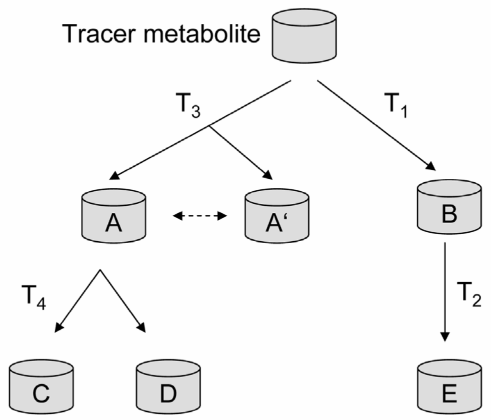

Figure 3.6: Example for the process of the feed forward flux estimation algorithm. A pool of a tracer metabolite is transformed by the metabolic operations T_1, T_2, T_3 and T_4. T_3 is symmetric. In the forward flux phase, T_3 and T_1 fill *A*, *A'* and *B*. T_4 generates from *A* the isotopologs *C* and *D*. *C* and *D* are fill until there are equilibrated or *A* is depleted. In either case a slipstream from the tracer pool and balancing between *A* and *A'* must occur in the counterflux phase. The same reckons for *B* and *E*. Filling *E* depletes *B*. Thus a slipstream from the tracer pool must occur until *B* and *E* are balanced with the experimentally obtained constraints.

3.7 Application examples

The following examples illustrate the analysis of intermediary metabolic fluxes in *Drosophila melanogaster*, *Nicotiana tabacum* and *Zea mais*. Steady state fluxes in *Drosophila melanogaster* and *Nicotiana tabacum* representing important biotechnological model organisms are evaluated. The analysis of the biosynthesis of starch in the cereal endosperm of *Zea mais* is one of the most suited systems for the illustrated experimental and computational techniques. Examples analyzing the metabolic modulations during starch synthesis in developing maize kernels, the origin of phytoglycogen, and hybrid vigor will be presented.

3.7.1 Steady state fluxes in *Drosophila melanogaster*

Drosophila melanogaster serves as one of the most important model organism in genetics and developmental biology (Lewis, 1978; Nüsslein-Volhard and Wieschaus, 1980). Its genome is completely sequenced and annotated, but its metabolic network is barely characterized (Adams et al., 2000).

The 13,600 annotated genes of *Drosophila melanogaster* (Adams et al., 2000) enclose only approximately 320 genes directly associated with primary anabolic and catabolic functions. In detail, 0.37 % of the genes are associated with carbohydrate metabolism, 0.49 % with amino acid metabolism, 0.48 % with energy pathways and electron transport and 0.79 % with lipid metabolism (Adams et al., 2000; Lee and Sonnhammer, 2003). However, this share of the genome can already form a metabolic graph containing 1.0 x 10^5 possible interactions. The small-world character of this graph (cf. chapter 1.1.4; Bray, 2003) reduces the intraconnectivity by at least three orders of magnitude.

Establishing the ^{13}C perturbation/relaxation technique using glucose as tracer and analyte for *Drosophila melanogaster* allows a huge range of experiments, e. g. assessing the impact of environmental changes, nutritional changes, host-parasite interaction (chapter 3.7.7) and drug interaction.

Drosophila melanogaster strain Oregon-R was grown on medium containing 5.6 g/L [U-$^{13}C_6$]glucose. The fly was kept at room temperature. Between one and two days after hatching, adult flies were frozen using liquid nitrogen and ground. The powder was extracted with water and methanol. The mixture was purified by centrifugation, filtration and reverse phase HPLC (chapter 5.3.1). Glucose was prepared by Ralf Laupitz, Chair of Organic Chemistry and Biochemistry, Technical University of Munich.

^{13}C NMR confirmed the presence of glucose with an α/β anomer ratio of 0.7. The purity of the glucose sample was checked using ^{1}H NMR spectroscopy. The ^{1}H NMR spectrum of glucose shows the signals for H-1α (5.26 ppm) and H-1β (4.51 ppm) well separated in a non-crowded region of the spectrum. H-1β shows heteronuclear $^1H^{13}C$ coupling ($^1J_{CH}$ = 170 Hz) which is used for calculating the total ^{13}C abundance of C-1β (cf. chapter 1.1.3). The ^{13}C abundance of C-1β of glucose was determined at 2.2 mol %. The ^{13}C abundance of the remaining carbon atoms were cross-calibrated using signal C-1β.

The ^{13}C NMR spectrum showed 12 signal sets divided in multiplets for the α- and β-anomers of glucose. The signal sets are clustered in 8 regions. The populated areas of the ^{13}C NMR spectrum contain signal sets with information on single carbon positions of glucose (1β, 1α, 2β, 3α), carbon positions in the α- and β-anomer forms of glucose (4α/β, 6α/β), and complex signal clusters resulting from different carbon positions of glucose (3α/5α, 2β/5β). As a consequence of heavy isotope shift effects (Appendix A.3.1) and of non-first order coupling, the satellite patterns of the multiplets are not strictly symmetrical with respect to the central lines attributed to the respective [$^{13}C_1$]isotopologs.

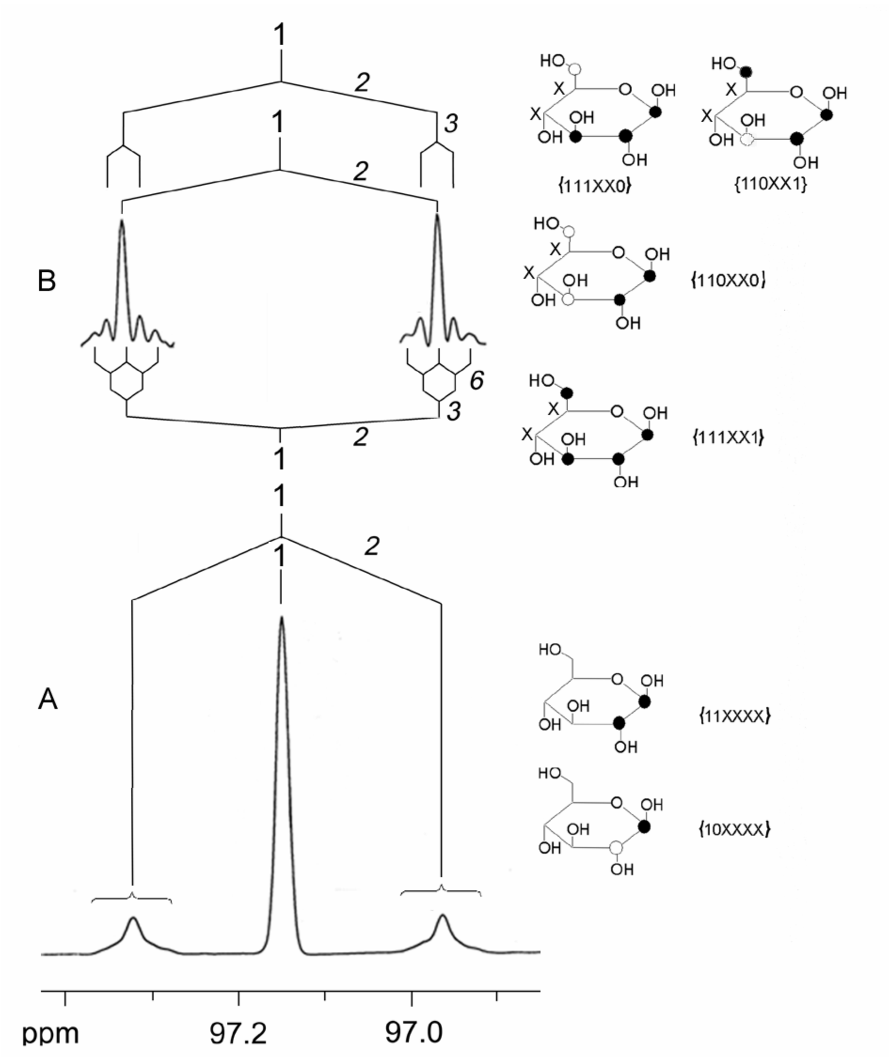

Fig. 3.7: ^{13}C NMR signal of C-1β of glucose isolated from *Drosophila melanogaster* cultivated on a medium containing [U-$^{13}C_6$]glucose. On the basis of $^{13}C^{13}C$ coupling constants summarized in Appendix A.3 all satellite signals were assigned (coupling atom in italic letters). Glucose isotopologs show filled circles indicating ^{13}C, open circles indicating ^{12}C, and X indicating ^{12}C or ^{13}C, and illustrating the translation of ^{13}C NMR spectral information to spatial and semantic information. Next to the structures, the corresponding X groups are encoded by 1 eq. ^{13}C, 0 eq. ^{12}C and X eq. ^{13}C or ^{12}C. A. Spectrum calculated with a mild Gaussian function. B. spectrum calculated with a strong Gaussian function (only satellite signals are shown).

The X group notation of glucose (described in the beginning of chapter 3) will be exemplified using the ^{13}C NMR signal of C-1β of glucose (Figure 3.7). The total signal set of C-1β can be summarized to {1XXXXX}, because only the ^{13}C atom at position C-1 of glucose is determined. Homonuclear $^{13}C^{13}C$ between C-1β and C-2β of glucose results in a single satellite pair (Figure 3.7 A) and resolves the X at position C-2 of {1XXXXX}. Thus, the X group {1XXXXX} can be divided in its share resulting from the central signal of C-1β corresponding to X group {10XXXX} and its share resulting from $^{13}C^{13}C$ coupling satellites of C-1β and C-2β corresponding to X group {11XXXX}. The symbols 0 and 1 of an X group can add up to X using arithmetic rules as they apply for a ternary number system. Therefore, {10XXXX} and {11XXXX} add up to {1XXXXX}.

As shown in Figure 3.7 B, the X group {11XXXX} can be divided further by processing the FID signal of the ^{13}C NMR spectrum with a stronger Gaussian function. The resolved long-range $^{13}C^{13}C$ coupling between C-3β and C-6β of glucose reveals four additional X groups in the C-1β signal set. The ^{13}C satellite signals of C-3β divide {11XXXX} in {110XXX} and {111XXX}, with respect to the signal integral ratios. Furthermore, long-range $^{13}C^{13}C$ coupling to C-6β allows extending the X group set derived from C-1β signal of glucose further (Figure 3.7 B) by {111XX0} and {110XX1}. On the basis of the dissociative conjecture (chapter 3.1), the X group {110XX1} can be excluded for the subsequent analysis. Translation of the remaining ^{13}C NMR signals of α- and β-glucose was performed similar to the evaluation of glucose signal C-1β. The complete translation table for all ^{13}C NMR signals to their corresponding X groups is found in Appendix A.3.2.

The X groups derived from the 12 signal sets in the ^{13}C NMR spectrum are not equivalent. The exactness of an X group corresponds to the signal quality and signal-to-noise ratio associated with its related ^{13}C NMR signals. The signals of C-1β (central peak at 97.20 ppm), C-1α (central peak at 93.28 ppm), C-2β (central peak at 75.41 ppm) and C-3α (central peak at 74.05 ppm) are well separated from signals of the remaining carbon signals of α-glucose and β-glucose. These signals contain explicitly the related single-labeled glucose isotopologs and encode {110000} and {111111} in their associated X group set ({110XX0}, {111XX1}, {X111XX}, cf. chapter 3.1).

The isolated signal sets for 4α/β, 6α/β are also available with high exactness. Signal C-4α/β is well resolved in the spectrum (central peaks at 70.90 ppm and 70.83 ppm) and the X groups of this signal set can be unequivocally determined. Signals C-6α and C-6β (central peaks at 61.85 ppm and 61.97 ppm) overlap, i. a. the X group $\{XXXX11\}_{\beta}$ is contained in the central signal of C-6α. The X groups can be extracted using the corresponding satellite signal of $\{XXXX11\}_\beta$ (62.13 ppm).

Signals for C-3α/C-5α and C-2β/C-5β overlap in the spectrum and are resolved through recalculation of the signal integral values by exploiting the redundant information contained in $^{13}C^{13}C$ coupling satellite signals. For example, $\{XXX010\}_\alpha$ (central peak at 72.55 ppm) overlaps with {XXX111} and {111XXX}. By averaging integrals of isolated satellite signals from X groups $\{XXX111\}_\alpha$ and $\{111XXX\}_\alpha$ and knowledge of the specific coupling pattern (obtained from commercially available glucose species, see Appendix A.3.2), $\{XXX010\}_\alpha$ can be determined by calculation of the difference of signal integrals.

The ^{13}C NMR spectrum in Figure 3.8 affords 30 X groups. 22 X groups cannot be generated using their subsets. For example, {11XXXX} can be reconstructed from {110XX0}, {111XX1} and {111XX0, 110XX1}. In contrast, {111XX1} cannot be reconstructed from more specific X groups due to experimental limits in the resolution of the ^{13}C NMR experiment.

The X groups were used to deconvolute the isotopolog space of glucose for catabolic metabolic processes with a genetic algorithm (chapter 3.3). The deconvolution process afforded an average error of 0.04 mol % per X group with a standard deviation of 0.06. The algorithm determines 6 single-labeled and 7 multiple-labeled glucose isotopologs. All single-labeled isotopologs were well above the natural ^{13}C abundance (1.33 to 1.63 mol % *versus* 1.1 mol %). Multiple-labeled isotopologs were determined with a ^{13}C excess concentration range of 0.15 mol % to 0.32 mol %.

Figure 3.9 compares the resulting glucose isotopolog abundances of the ^{13}C perturbation experiment of *Drosophila melanogaster* with natural ^{13}C abundance glucose isotopologs. The impact of the perturbation is highest for {111111} with an experimental abundance of 0.24 mol %. In contrast, the natural abundance of {111111} can be calculated to 1.8×10^{-10} mol %. The ^{13}C excess of single-labeled species is higher than in multiple-labeled species. This fact already hints a high share of multiple cycling events for glucose molecules in the metabolic network. All glucose isotopologs found were predicted to be generated by catabolic processes (cf. chapter 3.1). Notably, {001111} is below the noise level of the isotopolog space (< 0.01 mol %) and can be excluded from the further analysis.

^{13}C excess of glucose isotopologs can be used for metabolic simulation with the 4F software package (chapter 3.6). Notably, glucose isotopologs as well as simulated carbon fluxes represent integral values of the metabolic events in different tissues and organs in the flies. The simulation afforded 94.6 % similarity with the experimentally obtained concentration values. More than 44 % of the glucose molecules have been transformed more than one

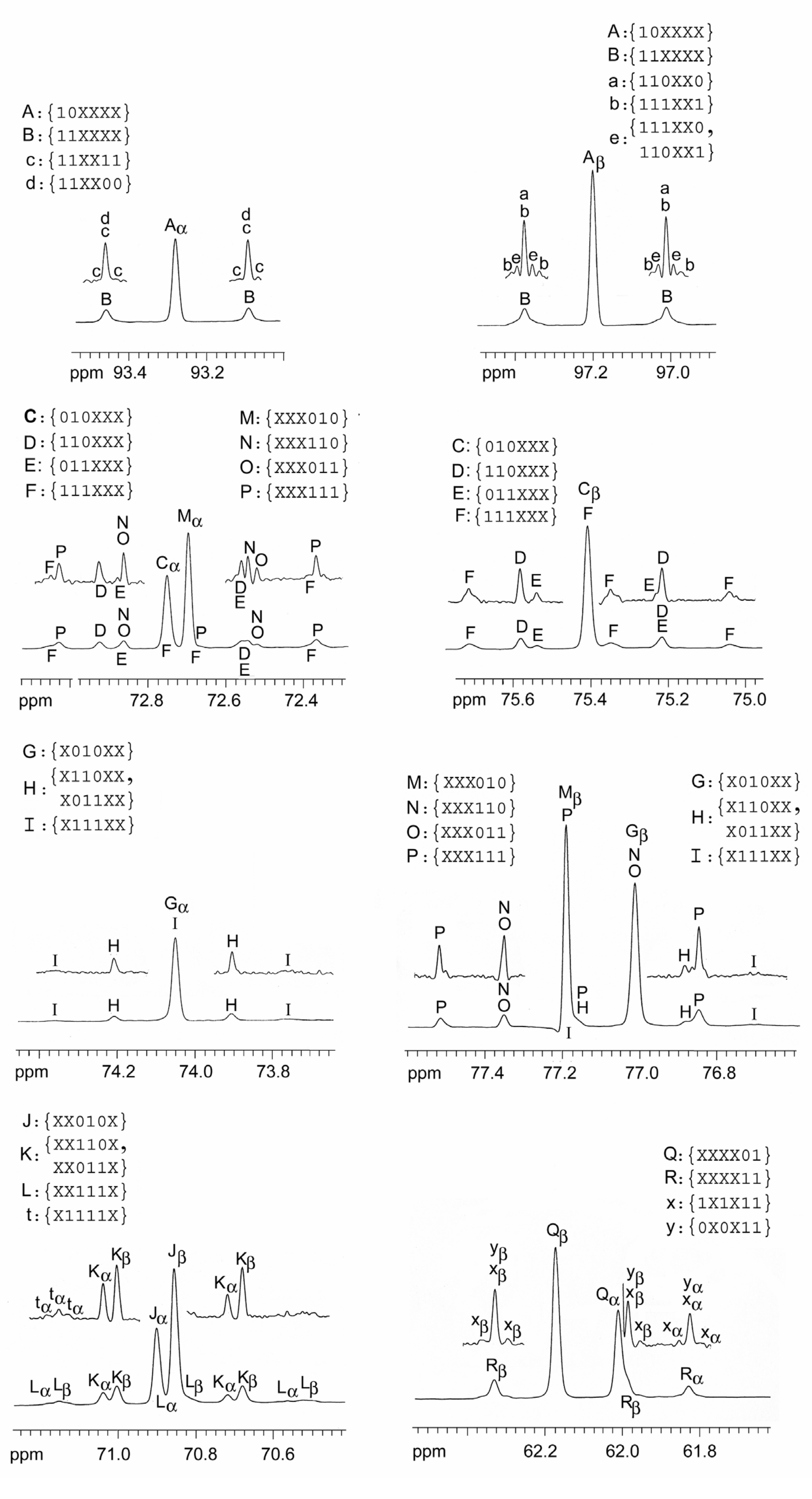

Figure 3.8: ^{13}C NMR signals of glucose isolated from *Drosophila melanogaster* cultivated on a medium containing [U-$^{13}C_6$]glucose. The signals are assigned to corresponding X groups (Appendix A.3.2).

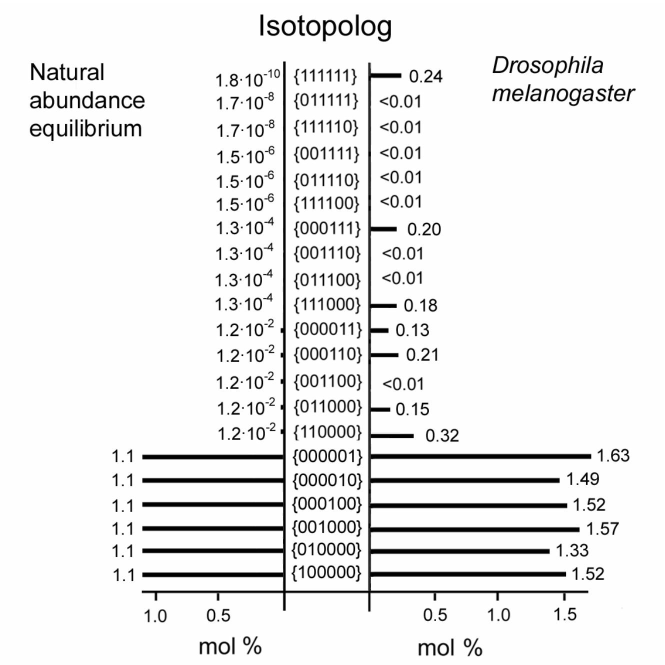

Figure 3.9: Natural ^{13}C abundance of glucose isotopologs *versus* their occurrence in the *Drosophila melanogaster* experiment. The natural ^{13}C equilibrium was perturbed by supplementing the medium with 5.6 g/L [U-$^{13}C_6$]glucose.

time by different metabolic processes (Figure 3.10) and only 56 % of the glucose molecules passed through a single metabolic transformation (Figure 3.10) prior to the termination of metabolic cycling. The simulation shows the pentose phosphate cycle and the citric acid cycle or proline cycle as dominant processes. The later processes cannot be distinguished in the metabolic simulation, since they result in the same glucose isotopolog pattern. Only 16 % of the glucose molecules were metabolized through glycolysis. 8 % of the glucose molecules were directly incorporated in the tissue of the flies.

The determined metabolic distributions and cycling distributions show the first analysis of intermediary metabolic processes in *Drosophila melanogaster* using glucose as tracer and analyte. The illustrated technique is suitable for *in vivo* metabolic studies with the insect. Influences of environment, nutrition and endosymbionts of *Drosophila melanogaster* will be assessed in chapter 3.7.7.

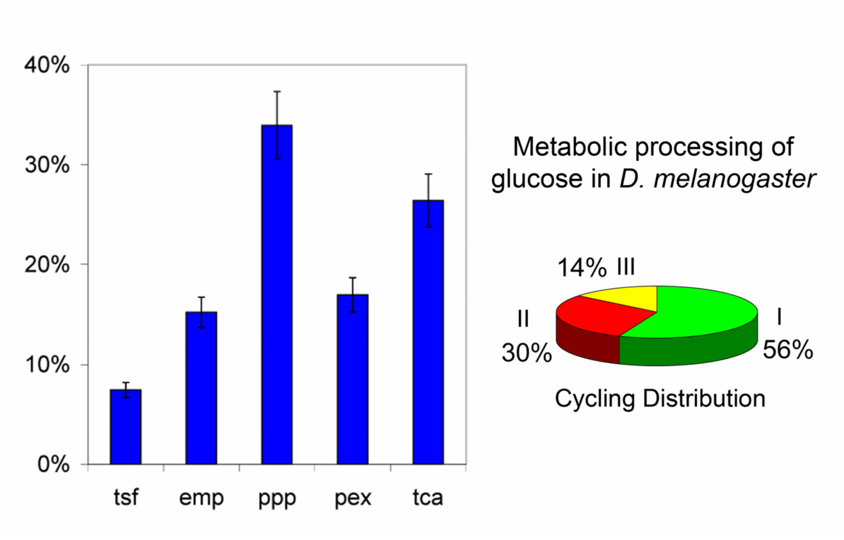

Figure 3.10: Results of the *in silico* simulation of the metabolic network of *Drosophila melanogaster* using the 4F algorithm. Pathways include glycolysis/glucogenesis (emp), the transketolase reaction of the pentose phosphate pathway (ppp), the transaldolase of the pentose phosphate pathway (pex), the citric acid cycle (tca) and the direct incorporation of [U-$^{13}C_6$]glucose in the tissue of the fly (tsf). The cycling distribution shows the relative abundances of single-, double, and triple glucose cycling events reshaping the isotopolog signature.

3.7.2 Steady state fluxes in *Nicotiana tabacum*

The metabolic complexity of plants may significantly exceed the complexity of regulatory and functional networks of animals. This is reflected in the number of genes available in the genomes of plants and animals. For example, the genome sequence of *Arabidopsis thaliana* elucidated 25,000 genes (Arabidopsis Genome Initiative, 2000) and the draft sequence of *Oryza sativa* suggests even 60,000 genes (Sasaki et al., 2002). In contrast, the genome of *Drosophila melanogaster* contains only 13,000 genes (Adams et al., 2000) and the genome of *Homo sapiens* contains less than 25,000 genes (Lander et al., 2001; Venter et al., 2001; International Human Genome Sequencing Consortium, 2004).

The analytical technique for studying intermediary metabolism established for *Drosophila melanogaster* in chapter 3.7.1 is subsequently transferred into an experimental system with *Nicotiana tabacum*. The *Nicotiana* genus divides in more than 100 species and belongs to the nightshade family (*Solanaceae*), which includes important crop plants such as potato or tomato. Only two species of the *Nicotiana* genus are used in agricultural industries:

Nicotiana tabacum and *Nicotiana rustica* (Bogani et al., 1997). All variations of Virginian tobacco are obtained exclusively from strains of *Nicotiana tabacum*. Furthermore, this plant is used as model plant in biotechnological research: techniques for analyzing host/parasite interactions (Ryang et al., 2004), plant physiology (for review, see Katoh and Hashimoto, 2004), and metabolic engineering (Tekoah et al., 2004) have been successfully conducted. ^{13}C perturbation studies on *Nicotiana tabacum* extend the complexity of the experiment with *Drosophila melanogaster* illustrated in chapter 3.7.1. The computer simulations describing the metabolic network of tobacco must include the Calvin cycle. It will be shown that the intermediary metabolic network of tobacco can be elucidated by computational evaluation of an experiment which uses [U-^{13}C$_6$]glucose as tracer and glucose as analyte.

The tobacco plant was grown on B5 medium (Manandhar and Gresshoff, 1980) containing 4 g/L [U-^{13}C$_6$]glucose diluted with 19.8 g/L unlabeled sucrose under aseptic conditions and reduced illumination compared to normal daylight (see chapter 5.2.2). Glucose was isolated after 20 days growth in ^{13}C-labeled medium from the leafs of the plant by water/methanol extraction. Purification was performed as described (for details, cf. chapter 3.2). Glucose were prepared by Tanja Radykewicz, Chair of Organic Chemistry and Biochemistry, Technical University of Munich.

In the ^{1}H NMR spectrum, heteronuclear ^{1}H^{13}C coupling of the signal H-1α (5.14 ppm, $^{1}J_{CH}$ = 170 Hz) assesses a ^{13}C abundance of 2.4 mol % for C-1α of glucose. Interpretation of the ^{13}C NMR spectrum and calculation of the glucose X groups was performed as described (cf. 1.3 and 3.1). The high-resolution ^{13}C NMR spectrum of glucose afforded 29 X groups, which were deconvoluted to isotopologs using a genetic algorithm (chapter 3.3). The deconvolution process resulted in 14 isotopologs with an abundance > 0.01 mol %. The average error was 0.02 mol % per X group with a standard deviation of 0.03. The resulting isotopologs represent a spatial and temporal integral over the metabolic processes occurring in root, phylum and leaf tissues and show extensive remodeling of the proffered universally labeled glucose contained in the medium. Processes in different cell compartments or plant tissues cannot be distinguished in the chosen experimental approach.

Single-labeled glucose isotopologs showed a ^{13}C abundance between 1.62 mol % and 1.83 mol %, which corresponds to a ^{13}C excess of 0.52 mol % and 0.73 mol % over natural ^{13}C abundance (Figure 3.11). Multiple-labeled isotopologs show an excess between 0.01 and 0.39 mol % over natural abundance (Figure 3.11). The dominant isotopolog is {111111} with 0.39 mol %. In this case, the [U-^{13}C$_6$]glucose was taken up *via* the root system, transported through the phylum, and stored in the leafs without metabolic rearrangement of its isotopolog signature. Other multiple ^{13}C-labeled glucoses are the {111000} isotopolog (0.16 mol %), the {000111} isotopolog (0.23 mol %), the {000011} isotopolog (0.10 mol %)

and the {110000} isotopolog (0.24 mol %). At lower ^{13}C abundance, but still above the detection limit, {011000}, {000110} and {001111} were observed with 0.03 mol %, 0.02 mol % and 0.01 mol %. The excess over natural ^{13}C abundance in the total system can be quantified to 4.85 mol %.

In Figure 3.11, the isotopolog space of glucose isolated from tobacco leaves is compared to the isotopolog space of glucose isolated from starch of maize kernels obtained in an experiment by Glawischnig et al., 2002. The experiment with maize was the first study showing the extensive remodeling of proffered [U-$^{13}C_6$]glucose through the metabolic network in plant tissues. The qualitative composition of the glucose isotopolog space of the maize experiment shows no isotopolog, which is not abundant in the experiment with tobacco. However, the deconvolution techniques used in Glawischnig et al., 2002 did not cover the entire isotopolog space.

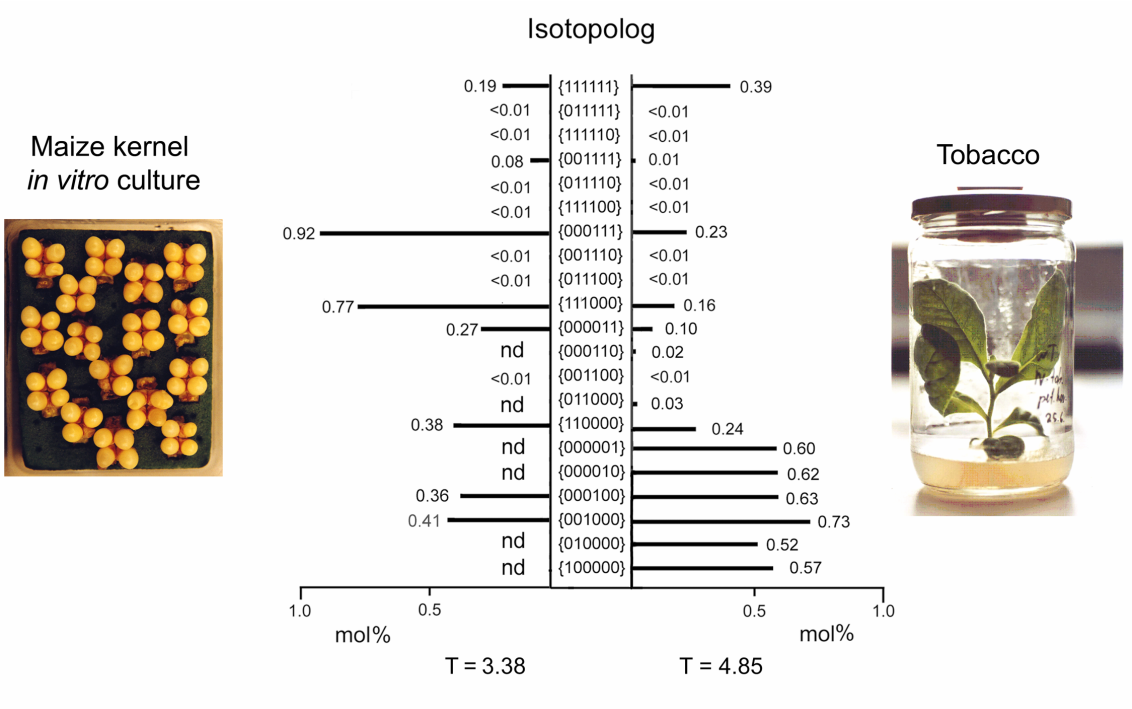

Figure 3.11: ^{13}C excess in experiments with kernels of *Zea mays* (Glawischnig et al., 2002) and growing plants of *Nicotiana tabacum* using [U-$^{13}C_6$]glucose as perturbant. *T* denotes total deviation relative to the natural ^{13}C abundance in glucose.

The maize kernel experiment was dominated by {111000} and {000111}, which result from glycolysis or the transaldolase reaction in the pentose phosphate cycle (cf. chapter 3.4), whereas in the tobacco experiment the high ^{13}C abundance of single-labeled glucose isotopologs already hints at extensive cycling through the pentose phosphate pathway or the Calvin cycle in the leaf tissue.

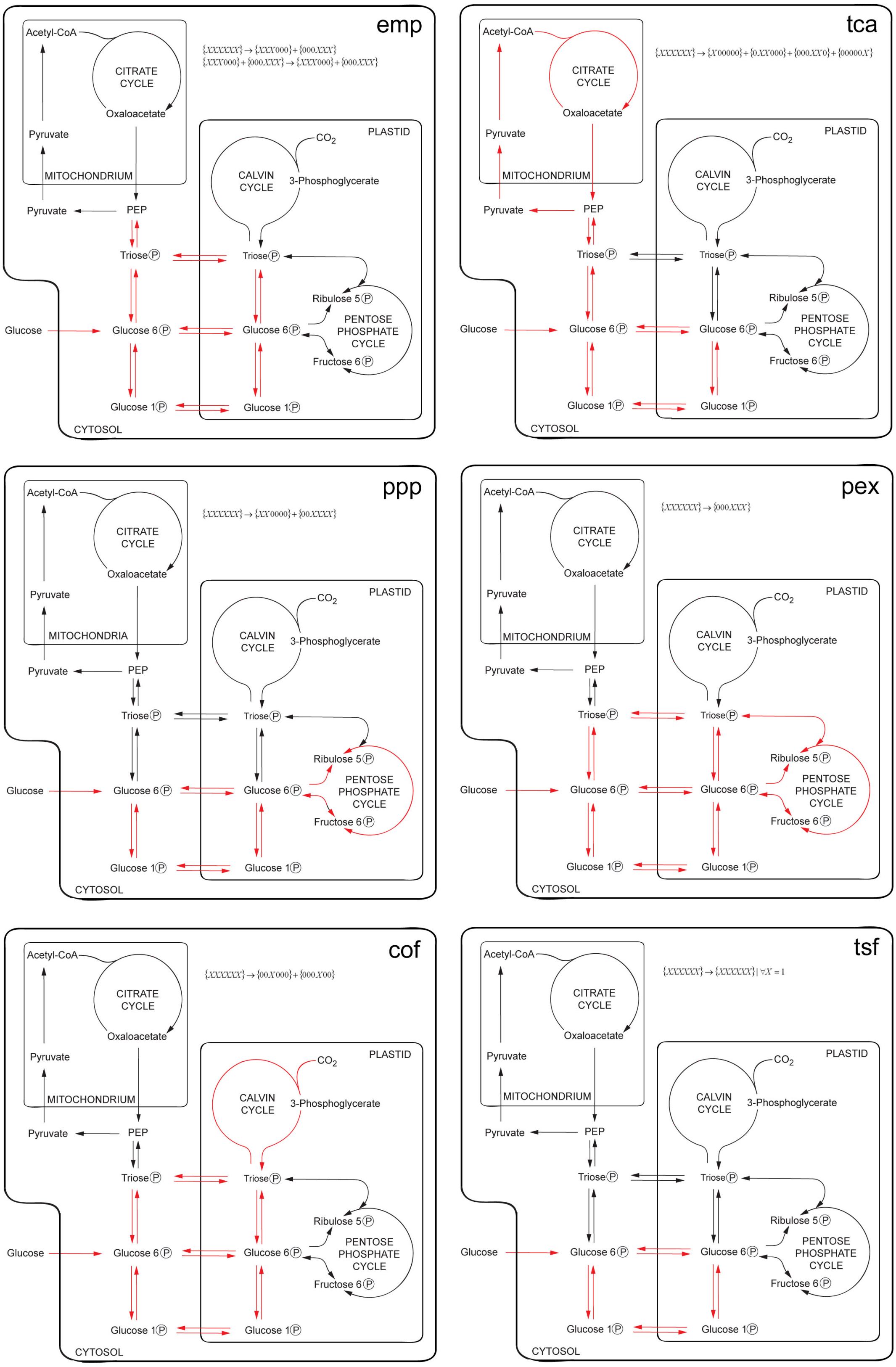

Figure 3.12: Metabolic reactions involved in the formation of single and multiple ^{13}C-labeled isotopologs of glucose isolated from *Nicotiana tabacum*. Pathways shown in red correspond to the 4F transfer rules summarized in chapter 6.4.2.

In silico simulation of the ^{13}C excess of the glucose isotopologs generated by solely catabolic pathways is only able to produce 82 % similarity compared to the experimentally obtained isotopolog patterns. The deviation between the simulated values and experimental values is maximal for {001000} and {000100}. These isotopologs can be generated by refixation of $^{13}CO_2$ produced by cell respiration of the plant tissue. After adding the $^{13}CO_2$ refixation process (cof, Figure 3.12) and the Calvin cycle to the 4F simulation, 97 % similarity between simulated and experimentally determined glucose isotopologs was obtained. It can be concluded that the six metabolic processes implemented in the 4F software (i. e. glycolysis/glucogenesis, transketolase and transaldolase reaction in the pentose phosphate cycle, citric acid cycle, Calvin cycle, direct transfer of exogenous glucose to leaf glucose) are sufficient to describe the key reactions involved in glucose metabolism of tobacco. Cycling of glucose molecules *via* the pentose phosphate cycle and the citric acid cycle was identified as dominant process in reshaping the isotopolog composition of the glucose molecules (28.6 % and 28.3 % of all metabolic transformations, Figure 3.13). These processes are responsible for the particularly high ^{13}C excess of the corresponding single-labeled glucose isotopologs. The refixation of $^{13}CO_2$ in the Calvin cycle takes place in significant amounts, affording 23.1 % of all metabolic transformations on glucose. Fragmentation of glucose through glycolysis/glucogenesis occurred at lower frequency (10.3 %, Figure 3.13).

While it is obvious that the refixation process only occurs in the leaf tissue, the source of the remaining metabolic transformations cannot be localized. Photosynthesis was artificially down-regulated, due to reduced illumination and catabolite repression enforcing heterotrophic growth of the plant. The exemplified technique appears suited to study metabolism on photosynthetically active plants, provided the plants can be grown on a defined medium (Walkey and Woolfitt, 1968). Experimental studies on the regulatory carbon flux modulations, sink metabolites and environmental factors seem possible.

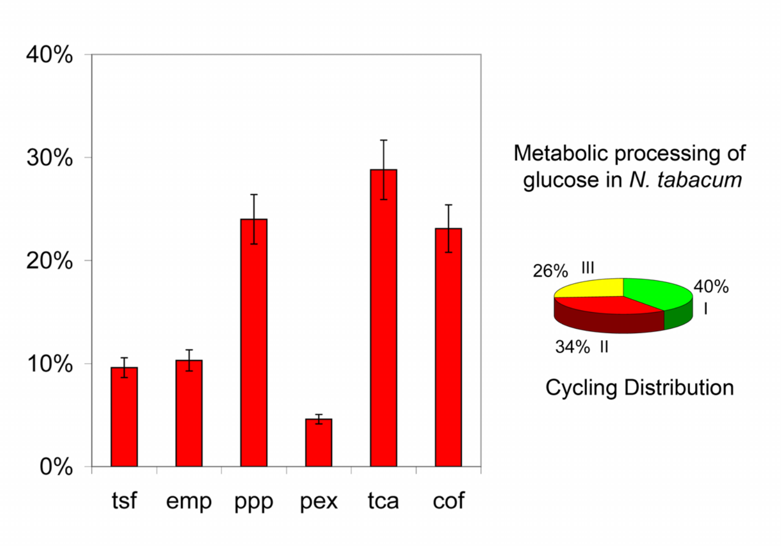

Figure 3.13: Computer simulation of the metabolic network in tobacco plants using the 4F algorithm. Pathways include glycolysis/glucogenesis (emp), the transketolase reaction in the pentose phosphate pathway (ppp), the transaldolase reaction in the pentose phosphate pathway (pex), the citric acid cycle (tca), the $^{13}CO_2$ refixation process *via* the Calvin cycle (cof) and the direct incorporation of [U-$^{13}C_6$]glucose in the plant tissue (tsf).

3.7.3 Temporal metabolic changes in developing maize kernels

Zea mays is one of the oldest agricultural crop plants, being cultured for over 7000 years. It belongs to the sweet grass family (*Poaceae*) and is derived by extensive plant breeding of teosinte (*Zea diploperennis*; Goff and Salmeron, 2004). The maize kernel harvest affords 638 x 10^6 t per year (Goff and Salmeron, 2004), more than rice (589 x 10^6 t/y) or wheat (585 x 10^6 t/y). Classical plant breeding and biotechnology tries to improve yield, resistance, growing time, and temperature tolerance (for review, see Paterson et al., 2003). Understanding the metabolic network in maize may help to design transgenic plants with increased agricultural value rationally.

The kernels of maize are the most valuable product of the plant. Kernels contain approximately 10 % protein, 4 % fat and 70 % starch (for additional information, see Johnson et al., 1999; Moose et al., 2004). Starch is synthesized and stored in a specialized cell compartment, the amyloplast. The precursor of starch is ADP-activated glucose, which can be derived from the hexose pools in the cytosol and the amyloplast of the cell (Figure 3.14). Hexose phosphate units or triose phosphate precursors are shuffled by specific transporters between cytosol and plastid. The coupling of ADP to glucose is the rate-limiting step for the incorporation of these glucose units into starch.

^{13}C perturbation experiments in this study focus on changes of glucose metabolism in developing maize kernels. It will be shown that maize kernels show a high degree of metabolic plasticity during maturation.

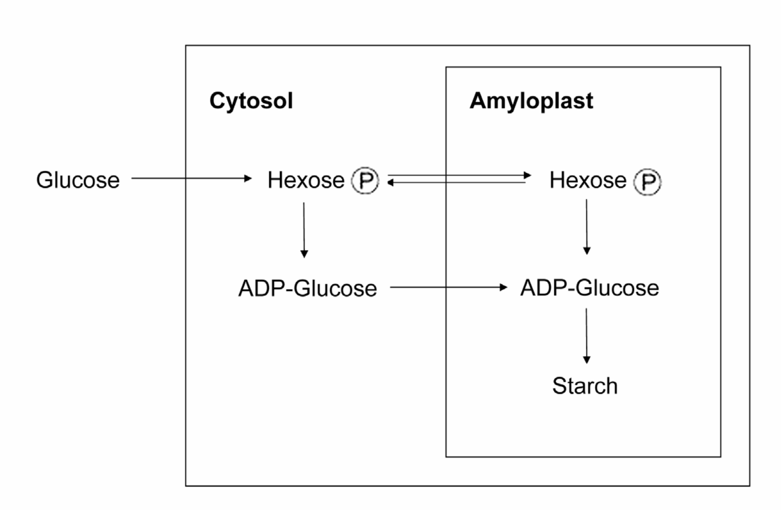

Figure 3.14: Biosynthesis of starch. The hexose phosphate pool of the cell is remodeled by intermediary processes. Hexose phosphates can be shuffled between the cytosol and amyloplast. The conversion of glucose phosphate to ADP-glucose is the committing in starch biosynthesis.

The experiment was conducted using maize cob tissue with attached kernels. Eight days after manual pollination, the tissue was transferred on agar-based medium containing unlabeled glucose (80 g/L). At 11, 18 and 28 days after pollination, cob tissue and maize kernels were transferred to medium supplemented with [U-$^{13}C_6$]glucose (2.7 g/L) and glucose with natural ^{13}C abundance (77.3 g/L). After 7 days on the ^{13}C-labeled medium, the kernels were harvested (Figure 3.15 A). Starch was enzymatically hydrolyzed and glucose was isolated and purified as described in chapter 5.3.1. Samples of glucose were prepared by Gertrud Spielbauer, Chair of Genetics, Technical University of Munich, Weihenstephan. The project was carried out in a cooperation between the Chair of Organic Chemistry and Biochemistry and the Chair of Genetics, Technical University of Munich.

Total ^{13}C abundance in glucose was obtained with high accuracy using IRMS (chapter 5.6). Glucose samples from time intervals 11-18 DPP, 18-25 DPP and 25-32 DPP showed a ^{13}C abundance of 2.53 mol %, 2.08 mol %, and 1.87 mol %. The successive decrease in the total ^{13}C abundance results from mixture of ^{13}C-labeled glucose precursor molecules with a growing number of unlabeled glucose molecules already incorporated in starch. The starch accumulation in the experimental period of time was shown to be approximately linear in the time interval 9 DPP to 39 DPP (cf. Figure 3.15 B).

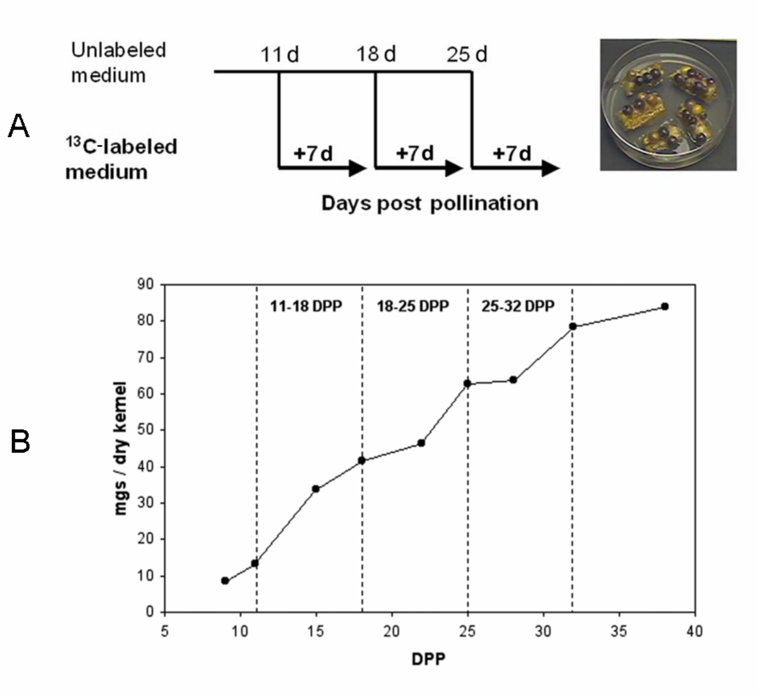

Figure 3.15: A. Experimental schedule for maize kernel transfers and growth curve of cultured kernels. At 11, 18 or 25 days after pollination maize kernels were transferred onto culture medium containing [U-$^{13}C_6$]glucose for 7 days. B. Increase of kernel dry weight in sterile culture from 9 to 38 DPP. The three intervals chosen for ^{13}C-labeling are indicated.

Glucose was analyzed by high-resolution ^{13}C NMR spectroscopy as described (chapter 1.1.3 and chapter 3.1). 28 to 32 X groups were unequivocally extracted from the ^{13}C NMR spectra. The spectra show an increasing ratio between the integral of the central peak and the total integral of a single carbon atom in the experimental time intervals (for examples, see signals C-1β, C-4α and C-4β in Figure 3.16). This information indicates a high degree of metabolic plasticity in developing maize kernels. Deconvolution of the X groups of glucose to their corresponding isotopologs was performed with an error of 0.03 to 0.05 mol % per X group and a standard deviation of 0.06 to 0.08. Figure 3.17 illustrates the impact of X group quality (depending on the signal-to-noise ratio in the ^{13}C NMR experiment) on the deconvolution procedure. X groups of the glucose sample 25-32 DPP showed the best convergences, with only minor deviations in the optimization courses. The deconvolution procedure for X groups of glucose sample 25-32 DPP showed in two out of five iterations exactly the same course of optimization. Four pathways to an optimal solution of the glucose isotopolog space were found in total. The optimum seems to be surrounded with other local minima, since one of the optimization processes afforded a similar solution with nearly identical glucose isotopolog abundances. X groups of the glucose sample 11-18 DPP, showed a similar course of convergence in four out of five repetitions with varying starting conditions (Figure 3.17). The remaining optimization process chose another path through the isotopolog search space, but resulted in the same optimum (Figure 3.17).

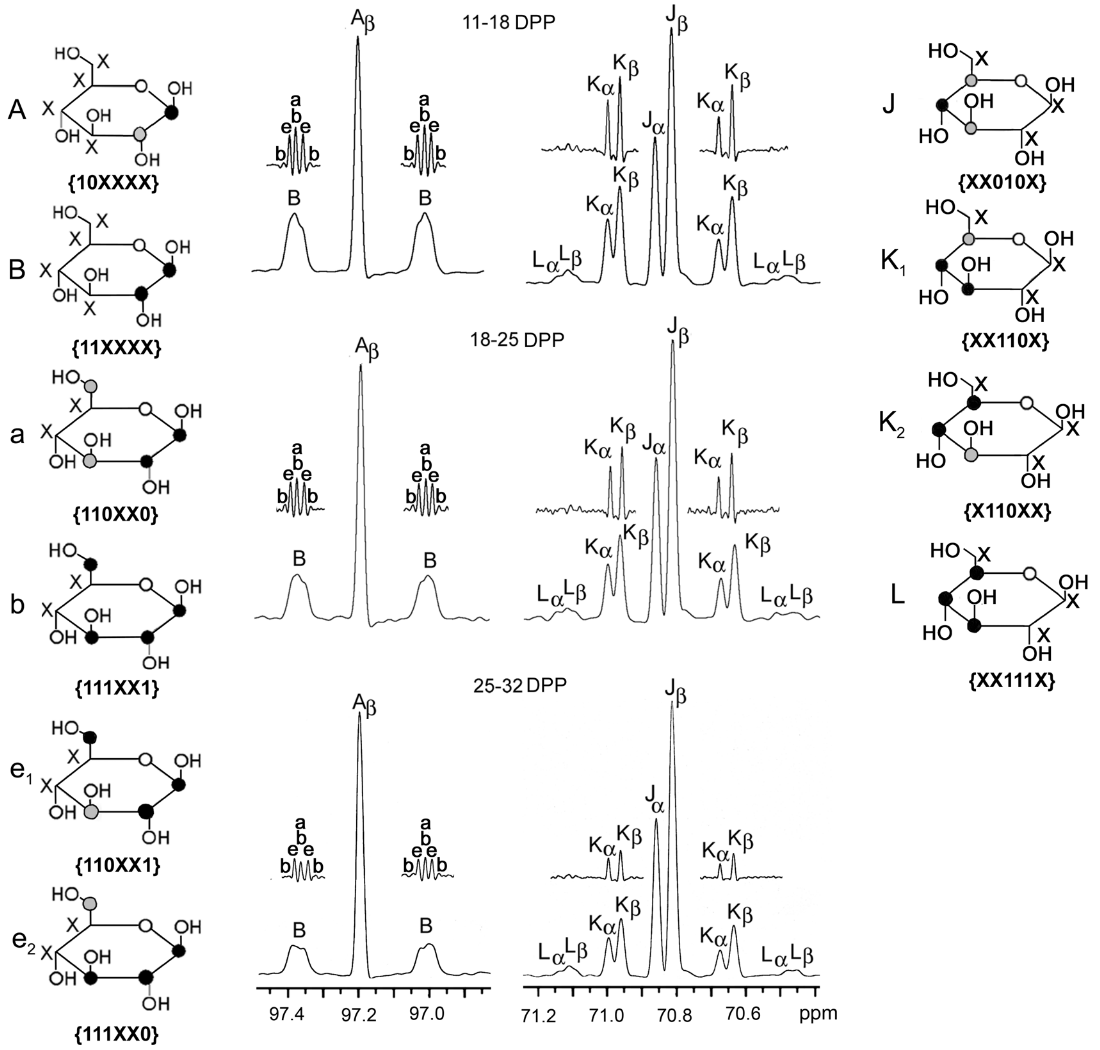

Figure 3.16: ^{13}C NMR spectra of glucose isolated from starch hydrolysate of kernels 11-18 and 18-25 days after pollination show the high metabolic plasticity of the maize kernel tissue. Signals of C-1β (97.20 ppm), C-4α (70.87 ppm), and C-4β (70.83 ppm) are shown with their corresponding X groups.

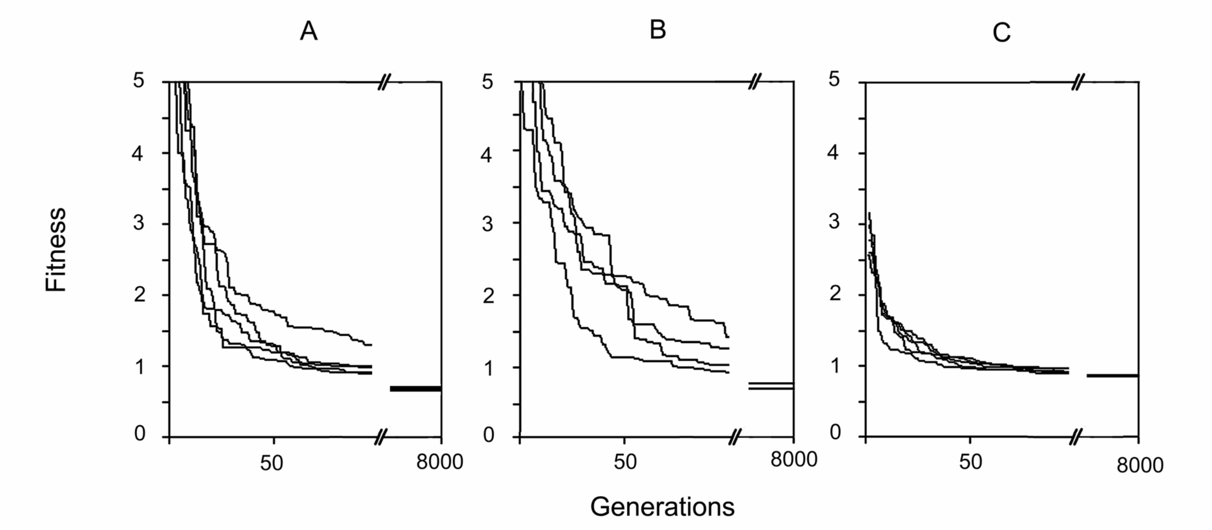

Figure 3.17: The evolution processes for all deconvolutions show a continuous approximation to the target function. A low initial value in iteration one identifies a tightly defined system, while a high initial value hints for a more complex search space. High complex searches can be optimized by additional mutation pulses (> 0.02; cf. chapter 6.3.4) to the search vector. All processes converge in 6,000 to 8,000 generations. A. Deconvolution of glucose from 11-18 DPP. B. Deconvolution of glucose from 18-25 DPP. C. Deconvolution of glucose from 25-32 DPP.

The ^{13}C abundance of all single-labeled glucose isotopologs was found to be between 1.18 mol % and 1.58 mol % (Figure 3.18). In all experiments, all multiple-labeled glucose isotopologs generated by catabolic pathways (cf. chapter 3.1) showed increased ^{13}C abundance (0.02 mol % to 0.77 mol %; Figure 3.18). The relative abundance of all single-labeled glucose isotopologs decreased significantly between day 11 and day 25 after pollination, while increasing between day 25 and day 32 (Figure 3.18). Notably, related pairs of single-labeled glucose isotopologs show a related course of their ^{13}C abundance in the experimental time period (cf. {100000} and {000001}, {010000} and {000010}, and {001000} and {000100} in Figure 3.18). Furthermore, single-labeled glucose isotopologs generated in one cycle (for examples, see {100000} or {001000}) increase less than single-labeled glucose isotopologs generated by multiple cycling (for example, {010000} and {000010}) in the late phase of starch accumulation. Correspondingly, multiple-labeled isotopologs are decreasing continuously in the three samples, with the exception of {011000}, {000011} and {111111}. {111000} and {000111} show the steepest descent in the time period of the experiment. Notably, {111111} shows its maximal ^{13}C abundance at 18-25 DPP. From the function courses of single-labeled glucose isotopologs and {111111}, a minimum of metabolic activity can be determined at 18-25 DPP.

Computer simulations confirm the initial conclusions. Simulating the metabolic excess with the 4F algorithm, the isotopologs can be reconstructed with 91 % to 96 % similarity (Figure 3.19). Glycolysis/glucogenesis was found to be the dominating process, affording a change in the isotopolog signature between 33 % and 48 % of all glucose molecules. This corresponds well with the evaluation of multiple-labeled glucose isotopologs, which contain {111000} and {000111} as most abundant glucose isotopolog species (Figure 3.18). Influence of the transketolase reaction in the non-oxidative branch of the pentose phosphate pathway was similar in 11-18, 18-25 and 25-32 DPP. The fraction of the transaldolase reaction on the total processes reforming the glucose molecules was found minimal (7 %) at 18-25 DPP. The influence of the transaldolase reaction can also be observed in the ratios of {111000} and {000111} which correspond to 1.26, 1.06 and 1.30. Cycling of glucose *via* the citric acid cycle, which could be best observed in {100000}, {010000}, {000010} and {000001}, is minimal in the middle phase of development (11 % *versus* 15 % at 11-18 DPP and 17 % at 25-32 DPP; Figure 3.19). As observed at the {111111} isotopolog, the direct incorporation of glucose molecules into starch is maximal at 11-18 DPP.

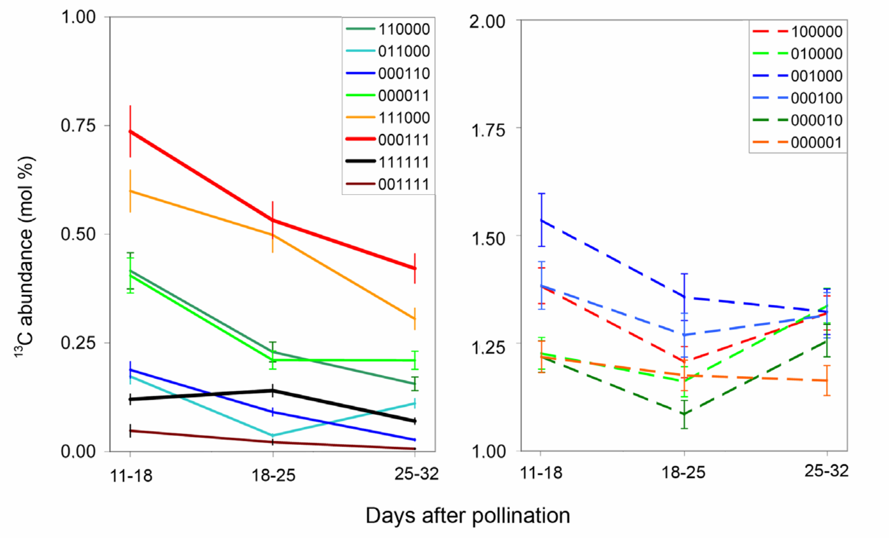

Figure 3.18: Abundance of multiple-labeled (left) and single-labeled (right) glucose isotopolog species in developing maize kernels. The ^{13}C quasi-equilibrium was perturbed by growth on B5 medium supplemented with [U-$^{13}C_6$]glucose for 7 days, after 11, 18 or 25 DPP.

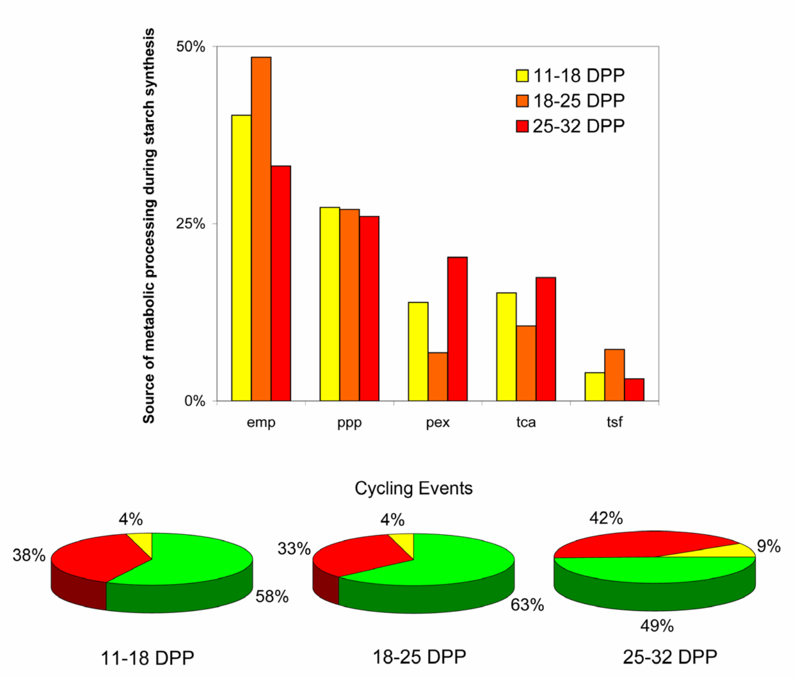

Figure 3.19: Simulation of intermediary metabolism in developing maize kernels. The topology of the simulated pathways is shown in detail in Figure 3.12. In the early and late phase of starch accumulation in maize kernels, metabolic activity is maximal as illustrated by the lower fraction of single cycling events (green). In the middle phase of starch accumulation, multiple cycling events (red, double cycling; yellow, triple cycling) are minimal indicating a decline in metabolic activity.

Simulating the direct incorporation as distinct process, this affords 7 % of all metabolic events at the sample from 11-18 DPP (Figure 3.19). Evaluation of the distribution of single cycling of glucose shows a maximum of metabolic activity in the time interval 18-25 days after pollination (58 % at 11-18 DPP, 63 % at 18-25 DPP, and 49 % at 25-32 DPP). In contrast, triple cycling of glucose molecules remained similar in the early and middle phase of starch accumulation and increased in the late phase of starch accumulation. Using the simulation data for the direct incorporation of glucose molecules, it can be deduced that starch accumulation was maximal in the middle time period (in agreement with literature, cf. Singletary et al., 1997). A larger fraction of glucose is activated with ATP and incorporated into starch without prior remodeling by other primary metabolic pathways. This is in agreement with minimal asymmetry between the triose phosphate moieties, which indicates a reduced carbon flux *via* the pentose phosphate cycle. Furthermore, the interpretation of the cycling distributions directly shows the plasticity of metabolism in maize kernels. Multiple cycling events in the early and late phase of starch accumulation reflect a metabolically highly active state of the maize kernel tissue in these time periods. A possible, although entirely speculative explaination, may be based on the different accumulation rates of storage compounds in maize kernels. While starch is accumulated in a linear fashion, synthesis of lipids and proteins is not constant. Protein accumulation is maximal in the early phase of maize kernel development while biosynthesis of lipids is maximal in the middle phase of maize kernel development (Ingle et al., 1965). The different biosynthetic sink metabolites withdraw biosynthetic precursors from the intermediate carbohydrate network in the maize kernel tissue. Therefore, the observed metabolic plasticity of developing maize kernels may originate from fluctuations in the demand for precursor molecules for different storage compounds.

3.7.4 Biosynthesis of phytoglycogen

The accumulation of phytoglycogen at the expense of starch has been observed in mutants of maize (*su1*, Boyer and Hannah, 2001), rice (Nakamura et al., 1997) and *Arabidopsis thaliana* (Zeeman et al., 1998). The *su1* mutant of maize shows a non-functional isoamylase-type debranching enzyme, which seems to be involved in the biosynthesis of starch (James et al., 1995). However, the exact function of the enzyme is not known and the mechanism leading to the accumulation of phytoglycogen is still under discussion (Ball and Morell, 2003). Currently two models are discussed, the glucan trimming model (Myers et al., 2000; Ball et al., 1996) and the WSP clearing model (Zeeman et al., 1998).

The glucan trimming model states that isoamylase activity removes α-1,6-branches from amylopectin which do not fit into the semicrystalline structure of amylopectin polymers. The model assumes that amylopectin is biosynthesized in the stroma of the amyloplast, trimmed through isoamylase activity, and subsequently incorporated into the starch granula. Absence of glucan trimming prevents amylopectin from crystallization and effects the accumulation of phytoglycogen (Myers et al., 2000).

The WSP clearing model states isoamylase does not directly affect the biosynthesis of starch. Phytoglycogen is thought to be a normal byproduct during starch biosynthesis. Isoamylase activity helps to degrade the accumulated phytoglycogen (Zeeman et al., 1998).

Maize mutant *su1* is derived from the inbreed line Wisconsin 64A (w64a). Maize kernels from *su1* and w64a were transferred to media containing [U-$^{13}C_6$]glucose or [U-$^{13}C_{12}$]sucrose 15 days after pollination, therefore being in the linear phase of starch accumulation (Figure 3.15 B). They remained on their respective medium for 7 days and were then harvested. Glucose was isolated and purified from starch and phytoglycogen as described in chapter 5.3.1 by Gertrud Spielbauer, Chair of Genetics, Technical University of Munich, Weihenstephan. The project was carried out in a cooperation between the Chair of Organic Chemistry and Biochemistry and the Chair of Genetics, Technical University of Munich.

High-resolution ^{13}C NMR spectra afforded 29 to 32 X groups, which were computationally deconvoluted to 14 isotopologs (cf. chapter 3.3). The deconvolution procedure fitted isotopologs to X groups with an average error of 0.02 mol % to 0.03 mol % per X group and a standard deviation of 0.02 to 0.04. Glucose isotopologs from *su1* phytoglycogen *versus su1* starch showed significant higher levels in single ^{13}C-labeled isotopologs and correspondingly lower levels of multiple ^{13}C-labeled isotopologs (Figure 3.20). Significant differences in single-labeled glucose isotopologs were found in {100000} (0.18 mol %), {000100} (0.23 mol %) and {000001} (0.22 mol %). In the set of multiple-labeled glucose isotopologs, {111000} and {000111} were decreased by 0.07 mol % in glucose derived from *su1* phytoglycogen compared to glucose derived from w64a starch. {110000}, {000011} and {111111} showed only smaller differences in the range of 0.15 mol % to 0.03 mol %. In contrast, glucose isotopologs obtained from starch of *su1* and w64a were found to be nearly similar: the RMSD between this sample pair is a three-fold smaller then the RMSD between glucose isotopologs from starch of *su1* or w64a and phytoglycogen of *su1* (for RMSD, see Table 3.2).

The significance of the deviation between glucose isotopologs derived from starch or phytoglycogen was positively confirmed using a χ^2 test (see chapter 6.2.3). The information entropy of the isotopolog space as defined in chapter 3.5 was calculated to 0.81 for glucose derived from starch and 0.92 for glucose derived from phytoglycogen.

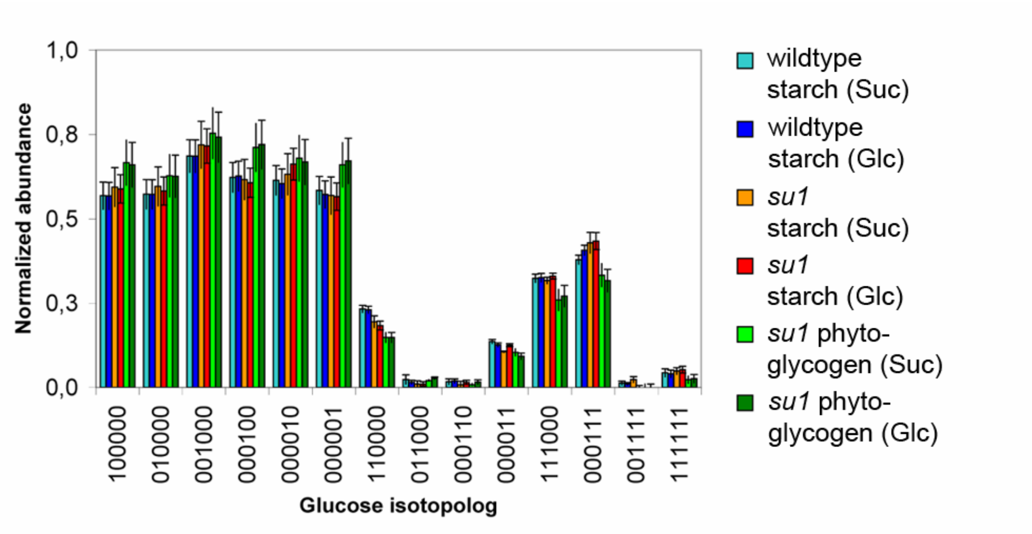

Figure 3.20: Isotopolog space of glucose from starch isolated from inbreed maize line w64a and *su1* mutant. Glucose derived from phytoglycogen was obtained from the *su1* mutant. The ^{13}C abundances have been normalized in respect to their total ^{13}C abundance determined by IRMS.

RMSD (10^{-2})	w64a-starch/ *su1*-starch	w64a-starch/ *su1*-phytoglycogen	w64a-starch/ *su1*-phytoglycogen
[U-$^{13}C_6$]glucose	2.45	6.37	6.12
[U-$^{13}C_{12}$]sucrose	2.42	5.98	5.54

Table 3.2: RMSD between glucose isotopologs of starch derived from maize w64a wild type and starch or phytoglycogen derived from maize *su1* mutant.

Based on the isotopologs, the intermediary metabolic network within the maize kernel tissue was simulated using the 4F algorithm. The simulated glucose isotopologs show 94 % to 96 % similarity compared to the experimentally obtained glucose isotopolog abundances. The simulation indicates that the direct incorporation of glucose to starch is similar in wild type and *su1* kernels (3 % to 4 %), but lower as compared to the direct incorporation in phytoglycogen (2 %). Glycolysis afforded 49 % of all metabolic transformations during starch synthesis in w64a wild type, and is significantly lowered to 43 % in the starch formation of the *su1* mutant. In the biosynthesis of phytoglycogen, glycolysis seems to affect only 33 % of all cycled glucose molecules (Figure 3.21). The contribution of the transketolase reaction in the pentose phosphate pathway is determined in a range of 30 % to 33 % (Figure 3.21). Contribution of the transaldolase reaction in the non-oxidative branch of the pentose phosphate pathway to the generation of the observed isotopolog space affords 9 % in glucose isotopologs isolated from starch of w64a wild type. In glucose isolated from the *su1* mutant, both glucose from starch and phytoglycogen show an increased level of metabolic transformations generated by transaldolase (13 % to 16 %). Contribution of the citric acid cycle is increased by comparison with glucose from starch to glucose from phytoglycogen. In wild type and *su1* mutant, 8 % to 10 % of the glucose was cycled *via* the citric acid cycle. In phytoglycogen from *su1* mutant, this fraction affords 19 %. Furthermore, analyzing the distribution of multiple-cycling events it can be determined that starch *versus* phytoglycogen shows 7 % fewer ternary metabolic cycling events.

The metabolic simulation (Figure 3.21) shows significant differences in the metabolic history of starch and phytoglycogen. Glucose molecules from *su1* phytoglycogen show a higher degree of fragmentation than from starch of both, wild type and *su1* mutant. The similarity of the isotopolog space of glucose molecules derived from starch of w64a and *su1* indicates no significant difference in starch biosynthesis between wild type and mutant. The experiment shows also no significant difference between the labeling patterns obtained from maize kernels grown in medium supplemented with [U-$^{13}C_6$]glucose or [U-$^{13}C_{12}$]sucrose, respectively.

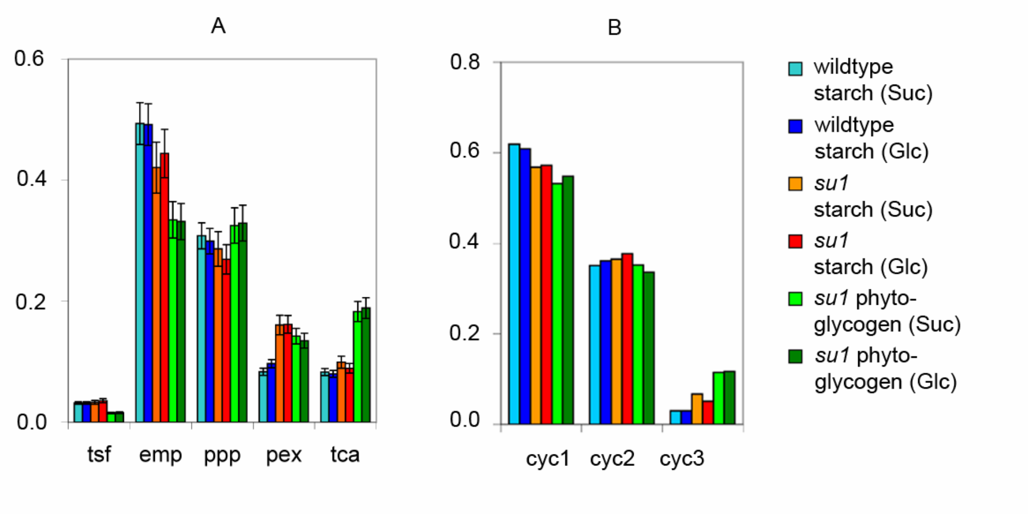

Figure 3.21: Computer simulation of the intermediary network in w64a and *su1* maize occurring in the biosynthesis of starch and phytoglycogen. Contributions of metabolic transfer rules as defined in chapter 6.4.2 and analysis of multiple cycling events show clear differences in the metabolic history of starch and phytoglycogen.

The observed differences between glucose derived from starch *versus* glucose derived from phytoglycogen exclude the possibility, that the two polyglucans were biosynthesized from the same precursor in the same compartment. Different precursors for starch and phytoglycogen seem theoretically possible, but available evidence suggests ADP-glucose as the common precursor: since mutations in genes associated with starch biosynthesis afford differences in phytoglycogen accumulation too, it can be assumed that the biosynthesis of both polyglucans uses a joint set of enzyme (Boyer and Hannah, 2001). Examples are mutations in starch synthase (*waxy, dull1*), branching enzyme (*amylose extender*) or ADP-glucose pyrophosphorylase (*shrunken2*). A model assuming different cell compartments with different metabolic turnover, would therefore provide a plausible explanation. Indeed, development of abnormal plastids containing only phytoglycogen was observed in light microscopy by Boyer et al., 1977.

Based on these facts, a spatial-separated metabolic model can be concluded: type A plastid will refer to a plastid accumulating starch, type B plastid will refer to a plastid accumulating phytoglycogen. Hexose phosphate units and ADP-glucose are available in the cytosol. Assuming a hexose phosphate translocater protein in type A and B plastids, but only an ADP-glucose translocater in type A plastid we can provide a spatial model to explain the generation of glucose isotopologs with different isotopolog entropies (Figure 3.22). Inactivation of the plastid-based *su1* isoamylase would prevent WSPs from being degraded and action of starch synthase and branching enzyme leads to the formation of phytoglycogen in type B plastids (Figure 3.22). Assuming two types of plastids for the biosynthesis of starch and phytoglycogen is not consistent with the glucan trimming model, but with the WSP clearing model. If the main function of the isoamylase would be the trimming of preamylopectin, no differences in the isotopolog patterns of glucose derived from starch or phytoglycogen would be found in the experiment. Reduced starch synthesis in *su1* leads to a higher abundance of polyglucan precursors, which subsequently affords an increased phytoglycogen content in the maize kernel tissue. This is not consistent with finding of Boyer and Hannah, 2001, showing reduced starch synthesis in many mutants without the accumulation of phytoglycogen. On the other hand, assuming WSPs as main substrate for the isoamylase enzyme explains the accumulation of phytoglycogen in different types of compartments independent of the capacity for starch synthesis.

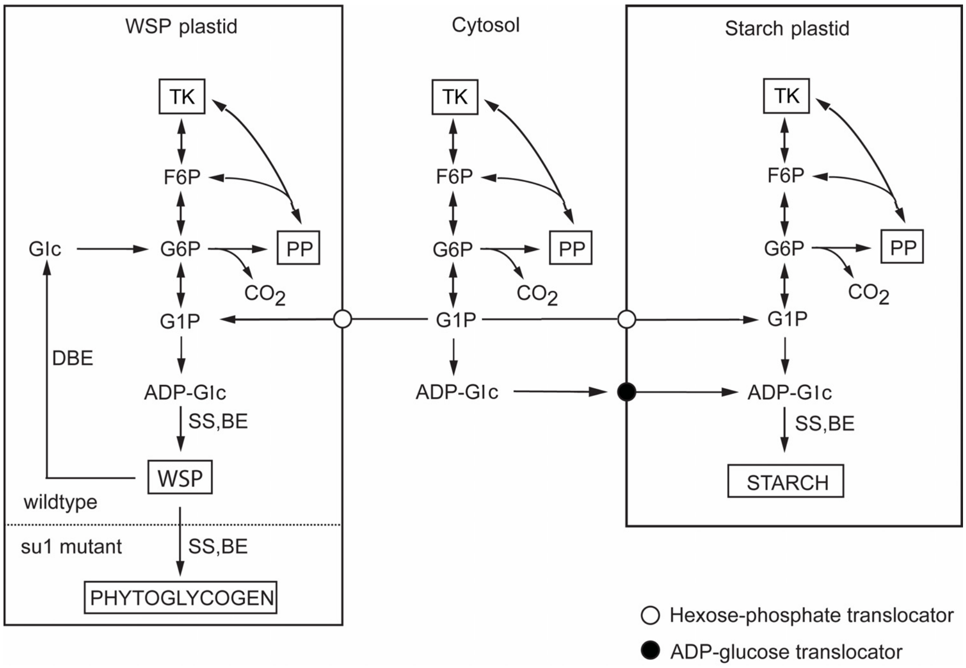

Figure 3.22: Model of starch and phytoglycogen synthesis in the cereal endosperm of maize. In the absence of isoamylase-type debranching enzyme, WSP plastids synthesize phytoglycogen from plastidal ADP-glucose while starch plastids synthesize starch from ADP-glucose imported from the cytosol. Different steady-state isotopolog patterns in the cytosolic and plastidal pools of glucose 1-phosphate then translate into different ^{13}C-labelling patterns of starch or phytoglycogen. Abbreviations: BE, branching enzyme; DBE, debranching enzyme; SS, starch synthase; PP, Pentose phosphate pathway; TK, Transketolase reaction.

3.7.5 Hybrid vigor in maize kernels

Heterosis in crop sciences describes the peculiarity of an agricultural valuable property in a hybrid plant. Hybrid vigor can manifast in increased crop yields, resistance, or other functional gains. The effect has already been discussed for two centuries in agricultural and husbandry sciences (Darwin, 1859; Shull, 1908; Jones, 1917; Hansson and Westerberg, 2002), but the mechanism is still the subject of ongoing discussions (for review, see Birchler et al., 2003). Formerly, single molecular events have been though to afford hybrid vigor. However, current research focuses on a multigenetic origin for the heterosis effect (Birchler et al., 2003). The two dominant theories consider the heterosis effect as either result of up-regulation of different alleles of functional metabolic genes (Duvick, 1999), or as allele-specific interaction of regulatory genes (Bingham et al., 1994).

Two inbred maize lines, b73 and mo17, represent a well established model for heterosis research (Lee et al., 2002; Hazen et al., 2003). Using the illustrated computer-based evaluation of ^{13}C NMR data of glucose, the heterosis effect can be studied on the level of intermediary metabolism in maize kernels.

For the metabolic study of heterosis, kernels from b73 and mo17 (generation F_0), the cross-pollinated kernels (mo17 x b73 or b73 x mo17, generation F_1), and the offspring plant of the crossing from line mo17 and b73 (mo17/b73, generation F_2) were transfered to media containing [U-$^{13}C_6$]glucose or [U-$^{13}C_{12}$]sucrose, respectively, 15 days after pollination. They remained in labeled medium for 7 days and were harvested. Glucose was isolated from starch hydrolysate as described in chapter 5.3.1. Samples of glucose were prepared by Gertrud Spielbauer, Chair of Genetics, Technical University of Munich, Weihenstephan. The project was carried out in a cooperation between the Chair of Organic Chemistry and Biochemistry and the Chair of Genetics, Technical University of Munich.

High-resolution ^{13}C NMR spectra of glucose afforded 30 to 32 X groups, which were computationally deconvoluted to 14 isotopologs. The deconvolution procedure fitted isotopologs to X groups with an average error of 0.02 to 0.05 mol % per X group and a standard deviation of 0.02 to 0.06. The isotopolog space contained 13 of 22 glucose isotopologs with a ^{13}C abundance > 0.02 mol % (Figure 3.23).

The inbreed lines b73 and mo17 show a highly similar glucose isotopolog space without significant differences. Crossing maize lines, b73 x mo17 and mo17 x b73, affords a slightly increased abundance of single ^{13}C-labeled glucose species in the F_1 generation. {001000} and {000001} were increased by 0.10 mol % and 0.15 mol %. The remaining single-labeled glucose isotopologs showed increased ^{13}C abundances between 0.05 mol % and 0.10 mol %. Correspondingly the $^{13}C_3$-labeled isotopologs {111000} and {000111} were found with decreased ^{13}C abundance (0.05 mol %) in the F_1 generation (b73 x mo17 and mo17 x b73). The direct incorporation of glucose as represented by {111111} was found slightly decreased by 0.04 mol %. The F_2 generation, b73/mo17 showed glucose isotopolog patterns similar to the F_0 generation of inbreed maize lines b73 and mo17. Maize kernels grown on medium supplemented with [U-$^{13}C_6$]glucose or [U-$^{13}C_{12}$]sucrose show similar glucose isotopolog patterns in the experiment (b73, mo17, and b73/mo17; Figure 3.23).

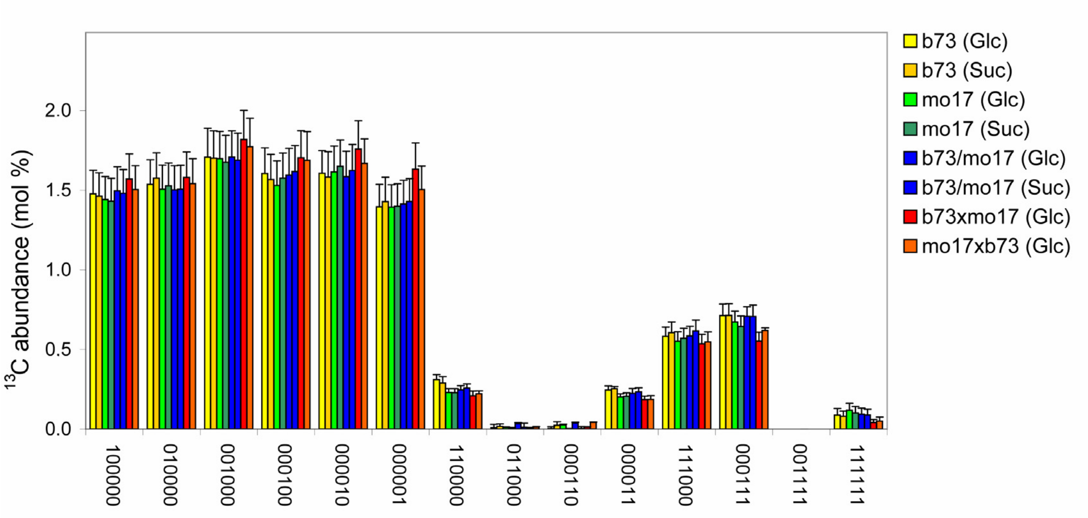

Figure 3.23: Isotopolog space of glucose for catabolic processes (cf. chapter 3.1) obtained from the F_0, F_1 and F_2 generation of the maize lines b73 and mo17. 13 glucose isotopologs show a ^{13}C abundance > 0.02 mol %.

Based on the glucose isotopologs, the intermediary metabolic network within the maize kernel tissue was simulated using the 4F algorithm with 95 % to 98 % similarity between experimental and simulated values. The simulation mapped the differences found in the isotopolog ^{13}C abundances to simulated metabolic processes. Corresponding to the isotopolog-based interpretation of the experiment, no significant differences were found in the F_0 generation between b73 and mo17, or in generation F_0 compared to F_2 (Figure 3.24 B). The metabolic cycling processes showed small differences in the F_1 generation (Figure 3.24 B). The direct incorporation of glucose molecules was found to be decreased by 1.2 % to 2.1 %. Correspondingy, the cycling of glucose molecules *via* glycolysis/ glucogenesis was decreased by 2.7 % to 6.4 %. The transketolase reaction in the non-oxidative branch of the pentose phosphate pathway was decreased by 3 to 4.4 %, while cycling *via* the transaldolase remained similar in the F_0, F_1 and F_2 generation. Glucogenesis of glucose *via* oxaloacetate after passage through the citric acid cycle was increased by 1.6 % to 2.5 %. In accordance with these results, the distribution analysis of multiple metabolic glucose cycling showed a shift of 2.5 % from single cycling events to triple cycling events (Figure 3.24 A).

Glucose isotopolog patterns and computer simulation of glucose cycling showed an increased metabolic plasticity in the F_1 generation (b73 x mo17, mo17 x b73). Notably, the effect vanished in the F_2 generation. The current theories of heterosis state the hybrid vigor results from the up-regulation of functional metabolic genes (Duick, 1999) or allele-specific interaction of regulatory genes (Bingham et al., 1994). While the first theory is in agreement with the increased metabolic plasticity in the F_1 generation, it fails to explain the similarity between generation F_0 and F_2. A genetic determination of hybrid vigor would be carried forward in subsequent generations. Epigenetical effects, i. a. genetic imprinting, may be an attractive explanation for the results (for review, see Scott and Spielman, 2004; Vinkenoog et al., 2003). The genetic imprint is mediated by methylation of DNA, which is biologically reset after fertilization of the ovule. This theory would explain the differences between the metabolic patterns in maize kernels of the F_0 and F_1 generation, as well as the similarity between the F_0 and F_2 generation.

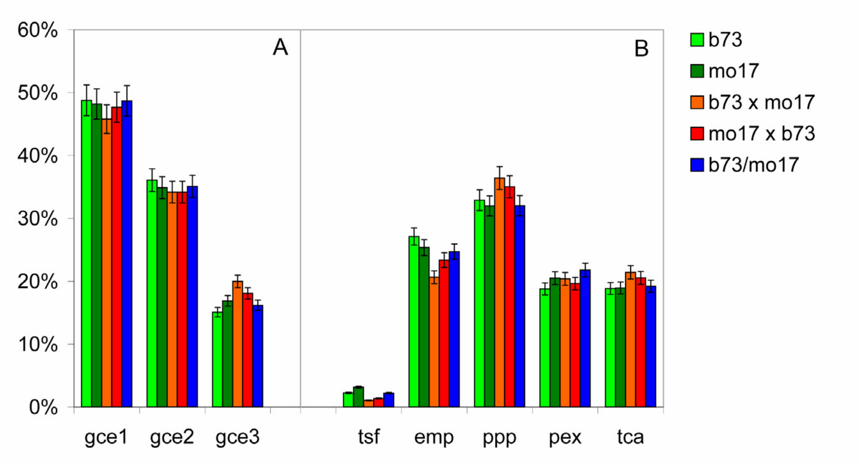

Figure 3.24: Computer simulation of the intermediary metabolism of the F_0, F_1 and F_2 generation of maize lines b73 and mo17. The F_1 generation (b73 x mo17 and mo17 x b73) shows minor modulations in the processes emp and ppp, which vanish in the F_2 generation (b73/mo17). F_1 also shows a minor increase in ternary glucose cycling events at expanse of single cycling events. The topology of the simulated metabolic network is illustrated in Figure 3.12. For the computer simulation, redundant experimental glucose isotopolog abundances generated by using [U-$^{13}C_6$]glucose and [U-$^{13}C_{12}$]sucrose as tracer have been averaged.

3.7.6 Starch synthesis in maize mutants

Isogenic maize lines with defined mutations in functional or regulatory enzymes related to the biosynthesis of starch are a valuable tool for rational, biotechnological optimization of crop yields (Raizada, 2003). Evaluation of the glucose carbon fluxes in mutants associated with starch formation may lead to new approaches for plant breeding to optimize starch properties of industrial importance, e. g. turgidity and gelation (Rindlav-Westling et al., 2002; Biliaderis, 1991).

For the experiments, the mutant maize lines *waxy* (*wx*), *amylose extender* (*ae*), *shrunken-1* (*sh1*), *brittle-1* (*bt1*), and *brittle-2* (*bt2*) on a isogenetic Wisconsin 64A (w64a) background were chosen. Maize kernels from mutants *wx* and *ae* show a different starch structure than starch from the wild type w64a. The *waxy* mutation leads to starch which is composed solely of amylopectin (Eisses et al., 1997), while the *amylose extender* mutation leads to an accumulation of amylose in the starch granula (Nishi et al., 2001). The *wx* mutation affords the inactivation of a granula bound starch synthase (Okagaki et al., 1991) and the *ae* mutation alters the activity of a gene encoding a branching enzyme (Nishi et al., 2001). Mutation *sh1* inactivates a sucrose synthase (Koch et al., 1992). Sucrose is the dominant transport form of hexoses in plants. Mutation *bt1* inactivates an amyloplast bound ADP-glucose translocater (Sullivan and Kaneko, 1995). The *bt2* mutation affects the small subunit of the ADP-glucose pyrophosphorylase, which is the rate limiting enzyme in starch synthesis (Greene and Hannah, 1998). ADP-glucose pyrophosphorylase has an $\alpha_2\beta_2$ quarternary structure. The phenotype analysis of the *bt2* mutant showed shrunken, brittle and collapsed maize kernels with significantly decreased starch content (Giroux and Hannah, 1994; Greene and Hannah, 1998).

Mutant lines of maize were grown in the summer of 2003 on a field in Gainesville, Florida. Kernels from mutant lines were grown for 7 days on media containing [U-$^{13}C_6$]glucose or [U-$^{13}C_{12}$]sucrose, respectively, 15 days after pollination. Glucose was isolated from starch hydrolysate as described (chapter 5.3.1). Samples of glucose were prepared by Gertrud Spielbauer, Chair of Genetics, Technical University of Munich, Weihenstephan. The project was carried out in a cooperation between the Chair of Organic Chemistry and Biochemistry and the Chair of Genetics, Technical University of Munich.

The ^{13}C abundance of glucose was determined by heteronuclear $^{1}H^{13}C$ coupling from the H-1β signal in the ^{1}H spectrum (2.5 mol %). Evaluation of ^{13}C NMR spectra afforded 28 to 32 X groups, which were computationally deconvoluted in 21 dimensions (cf. chapter 3.3). The deconvolution procedure fitted glucose isotopologs to X groups with an average error of 0.03 mol % to 0.06 mol % per X group and a standard deviation of 0.03 to 0.06. The isotopolog space contained 14 glucose isotopologs with a ^{13}C abundance > 0.01 mol % (Figure 3.25). Glucose isotopolog patterns of w64a, *ae*, *wx*, *sh1*, and *bt1* showed similar ^{13}C abundances. Differences were found when comparing w64a with the *bt2* mutant. Single-labeled glucose isotopolog species, {100000}, {010000}, {001000}, {000100}, {000010} and {000001} were found to be increased by 0.21 mol % to 0.38 mol % in glucose isolated from kernels of the *bt2* mutant. {111000} and {000111} were found to be decreased by 0.20 mol % and 0.27 mol % in glucose from *bt2* compared to w64a. {110000}, {000011} and {111111} were decreased by 0.19 mol %, 0.12 mol %, and 0.04 mol % in the *bt2* sample. The remaining glucose isotopolog patterns were not changed significantly.

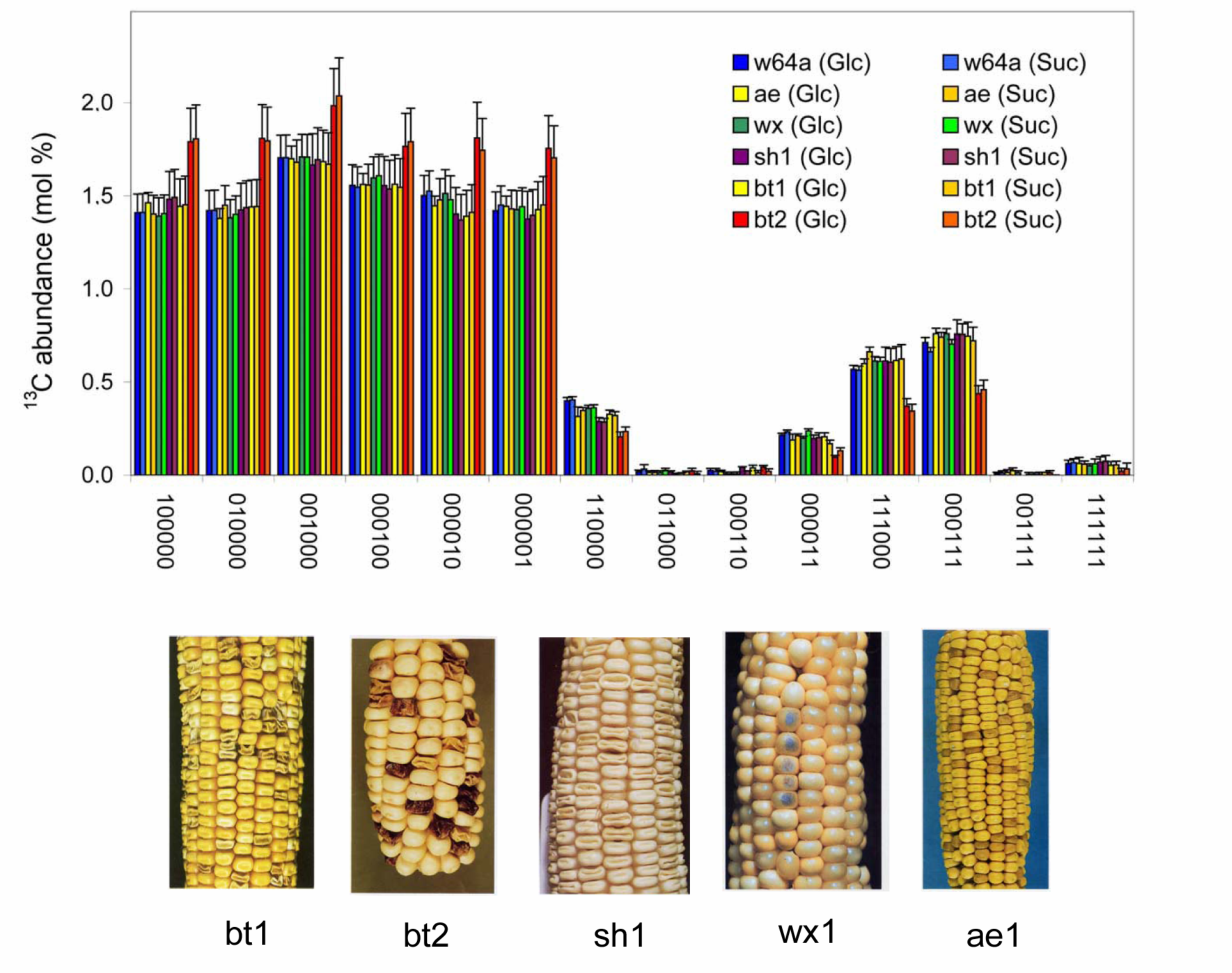

Figure 3.25: Glucose isotopolog space resulting from the mutant maize lines *ae*, *wx*, *sh1*, *bt1* and *bt2* derived from an isogenetic w64a wild type inbreed maize line. Based on the increased level of single-labeled glucose isotopologs, it can be deduced that *bt2* shows significant increased cycling of glucose in its intermediary metabolic network. The deviation between mutant phenotype and wild type of maize kernels (pictures were obtained from Neuffer et al., 1997) are not observed on the metabolic level. The phenotype of maize mutant *bt2* is characterized by brittled and shrunken kernels with decreased starch content.

Computer simulations of w64a and *bt2* were able to reconstruct the observed isotopolog space with 92 % and 95 % similarity, respectively (Figure 3.26). *bt2* showed a decreased glucose cycling *via* glycolysis/glucogenesis by 18 %. Correspondingly, glucose cycling in the *bt2* mutant *via* the transketolase reaction and transaldolase reaction in the pentose phosphate pathway and *via* oxaloacetate after passage through the citric acid cycle was increased by 3 %, 7 % and 9 %. The direct incorporation of glucose to starch was decreased by only 1 % in the *bt2* mutant compared to the w64a wild type.

The computer simulation allows a metabolic interpretation for the *bt2* maize kernel phenotype. The decreased starch content of the kernel is well in line with the lowered direct incorporation rate of glucose to starch and the lowered flux *via* glycolysis/glucogenesis. The enzyme ADP-glucose pyrophosphorylase catalyzes the rate-limiting step in starch biosynthesis. It can be deduced that the decreased activity of the small subunit of the enzyme removes less glucose units from the cellular metabolite equilibrium, thus leading to longer cycling of glucose molecules in other intermediary pathways.

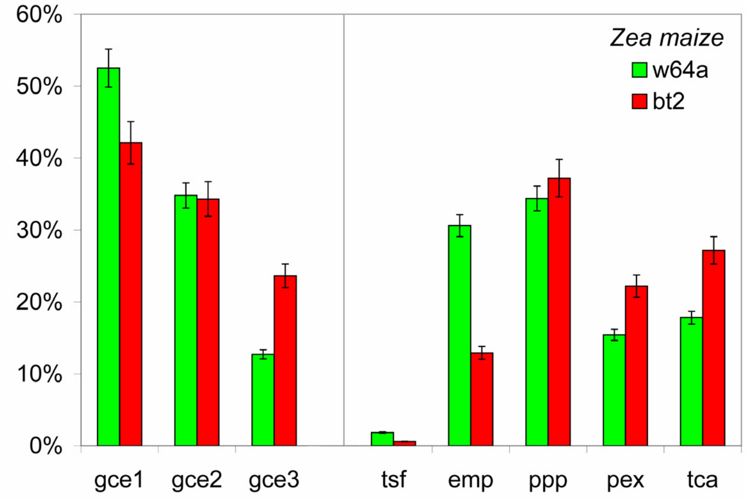

Figure 3.26: Simulation of the dominant metabolic processes generating the glucose isotopolog signatures illustrated in Figure 3.12 for maize mutant line *bt2* (reduced ADP-glucose pyrophosphorylase activity) compared to its *w64a* background. Definitions for the single rules of the *in silico* inference process are listed in chapter 6.4.2.

3.7.7 Metabolic influence of endosymbionts in *Drosophila melanogaster*

Cooperation and parasitism between species are ubiquitous in nature (Herre et al., 1993). Symbiosis affording synergetic effects between species can be observed in the utilization of nutrients. Currently, the best studied example of cooperation between host and endosymbionts in the animal kingdom is the utilization of cellulose in the rumen and cecum of cattle by *Fibrobacter succinogenes* and *Ruminococcus albus* (Xu et al., 2004; Michalet-Doreau et al., 2002).

In genetic and biotechnological model organisms, guest species and metabolic guest/host interactions are poorly characterized. *Drosophila melanogaster* offers a suitable system for studying the intermediary metabolism (chapter 3.7.1). The insect can be used to analyze the influence of endosymbionts in the gut of the fly on its dominant metabolic processes by using [U-$^{13}C_6$]glucose as tracer and glucose isolated from the hemolymph as analyte. As shown by Harcombe and Hoffman, 2004 and Ebbert et al., 2003 the dominant endosymbiotic species of *Drosophila melanogaster* are of *Trypanosomatidae*, *Wolbachia* spp. and the fungus *Coccidiascus legeri*. It is not known, if these endosymbionts are in a synergistic or parasitic relationship to the fly.

The flies were grown in a controlled environment (in contrast to the initial *Drosophila melanogaster* experiment described in chapter 3.7.1) at 27° C, 70 % air humidity and a 12 h day/night rhythm. The medium contained 5.6 g/L [U-$^{13}C_6$]glucose. The normal gut flora of the fly was eliminated by adding clotrimazole, rifampicine and erythromycine to the medium (for details, see chapter 5.2.1). Glucose was isolated from flies grown on either of the media as described (chapter A.3.1). The ^{13}C abundance of glucose was determined at 2.25 mol % by heteronuclear $^1H^{13}C$ coupling in signal H-1β.

^{13}C NMR spectra of these glucoses afforded 30 X groups. Computational deconvolution of the glucose X groups resulted in isotopolog spaces with 14 isotopologs > 0.03 mol %. The averaged error is 0.02 mol % per X group with a standard deviation of 0.05 (Figure 3.27). Cycling of glucose molecules was found higher in flies which were treated with antimycotic and antibacterial drugs. Each single ^{13}C-labeled glucose isotopolog is increased between 0.15 mol % and 0.76 mol %.

Multiple ^{13}C-labeled glucose isotopologs show very specific changes: *Drosophila melanogaster* with normal gut flora showed a similar glucose isotopolog pattern compared to *Drosophila melanogaster* treated with antibiotics, albeit with the exception of the {110000} and {000111} isotopolog (Figure 3.27) which were increased by 0.20 mol % and 0.17 mol %, respectively.

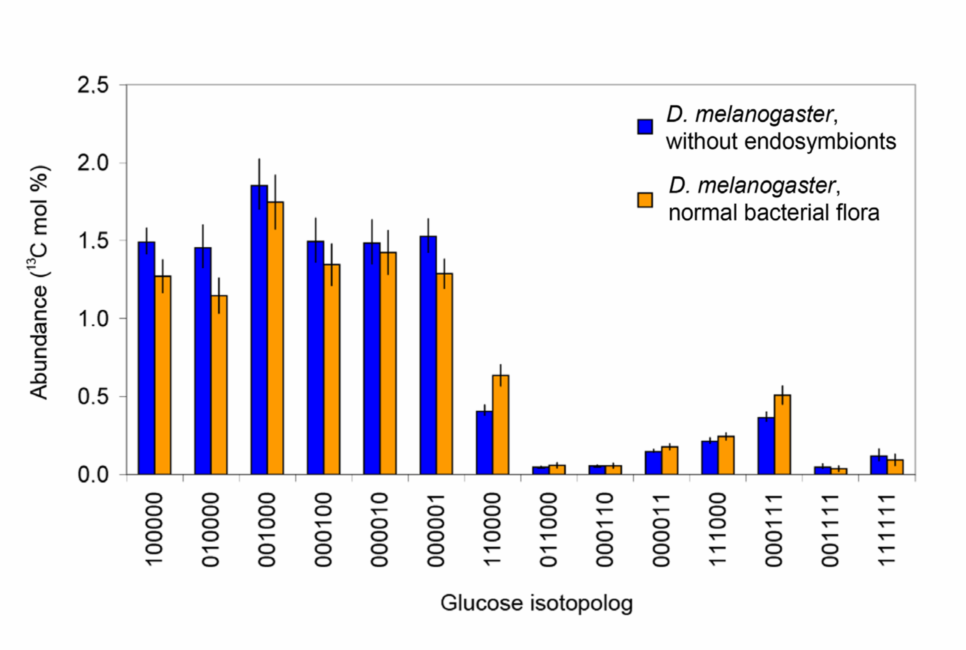

Figure 3.27: Result of the computational deconvolution of glucose X groups obtained from ^{13}C NMR spectroscopy of glucose isolated from *Drosophila melanogaster*. The flies were grown on medium containing [U-$^{13}C_6$]glucose. The insects were raised with normal gut flora and without endosymbionts in a semi-sterile, controlled environment.

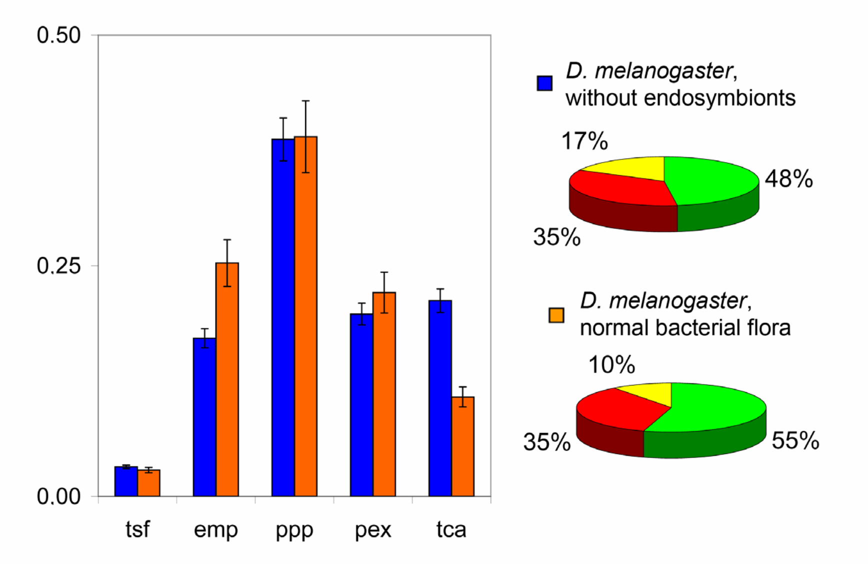

Figure 3.28: Computer simulation using the 4F algorithm of metabolic glucose cycling in *Drosophila melanogaster* grown on medium supplied with [U-$^{13}C_6$]glucose with its normal gut flora and without endosymbionts.

Computer simulations using the 4F algorithm afforded 96 % similarity between the experimentally observed and simulated glucose isotopolog patterns (Figure 3.28). The intermediary metabolic fluxes of the flies treated with anti-fungal and anti-bacterial ingredients in the medium showed an 8 % lower glucose cycling *via* glycolysis/glucogenesis and a 10 % higher glucose cycling *via* the citric acid cycle, or the Cori cycle, or the proline cycle (Figure 3.28).

Correspondingly, the analysis of the distribution of the total glucose cycling events showed an increase of ternary glucose cycling by 7 %, while single glucose cycling decreased by 7 %. Therefore, the metabolic shift is by far the most selective modulation of the intermediary glucose metabolism observed in the conducted experiments (chapters 3.7.1 to 3.7.6). The increased rate of glucose regeneration *via* oxaloacetate may result from stress which is caused by disturbance of the normal bacterial and mycotic flora of the fly or drug treatment but also other effects seem possible. The experiment shows, that glucose as tracer and analyte combined with the illustrated computer-assisted evaluation approaches can be used to analyze guest/host-interactions in animal species.

4. Anabolic metabolism and CO_2 fixation

The analysis of catabolic intermediary networks using glucose as universal analyte can be extended further by including anabolic carbon fluxes. Experimental systems not satisfying the dissociative conjecture (defined in chapter 3.1) require the deconvolution of the complete 64-dimensional isotopolog space of glucose. In the experimental setup described in this chapter, $^{13}CO_2$ replaces [U-$^{13}C_6$]glucose as tracer molecule. The glucose isotopologs are generated by anabolic and catabolic carbon fluxes within the metabolic network of photosynthetically active plants grown in an environment with controlled atmosphere. The deconvolution of the complete glucose isotopolog space requires additional constraints in the form of X groups to obtain a sufficiently determined equation system. Two additional experimental techniques will be introduced to achieve the generation novel X groups: the enzymatic conversion of glucose to 6,7-dimethyl-8-ribityllumazine and the analysis of glucose isotopomers by GC/MS. X groups obtained by high-resolution ^{13}C NMR experiments with glucose and 6,7-dimethyl-8-ribityllumazine as well as the isotopomer sets obtained by GC/MS can afford a sufficiently determined system for the computational deconvolution of the 64-dimensional glucose isotopolog space.

4.1 Extension of the glucose isotopolog space using enzymatic transformation of glucose to 6,7-dimethyl-8-ribityllumazine

Glucose can be converted to 6,7-dimethyl-8-ribityllumazine in a one-pot synthesis procedure using enzymes from the riboflavin biosynthesis pathway (Figure 4.1; Illarionov et al., 2004; Römisch et al., 2002). Enzymatic conversion of glucose to 6,7-dimethyl-8-ribityllumazine was performed by Lilla Margl, Chair of Organic Chemistry and Biochemistry, Technical University of Munich. 6,7-dimethyl-8-ribityllumazine contains C-2, C-3, C-4 and C-6 of glucose at corresponding positions 6α, 6, 7 and 7α (Figure 4.1). C-1 and C-5 of glucose are lost during the *in vitro* biosynthetic procedure. The spin system of 6,7-dimethyl-8-ribityllumazine affords a set of X groups not obtainable by ^{13}C NMR spectroscopy of glucose.

Enzymatic conversion of glucose to glucose 6-phosphate is performed by hexokinase using ATP as phosphoryl donor. The resulting ADP is regenerated by the enzyme pyruvate kinase converting phosphoenolpyruvate to pyruvate. Glucose 6-phosphate is converted to phosphogluconate by the enzyme glucose 6-phosphate dehydrogenase. The reaction requires $NADP^+$. The resulting NADPH is regenerated by the enzyme glutamate dehydrogenase, which converts 2-ketoglutarate to glutamate. 6-phosphogluconate is converted to ribulose 5-phosphate by oxidative decarboxylation through catalytic action of the enzyme 6-phosphogluconate dehydrogenase using $NADP^+$. 3,4-dihydroxy-2-butanone 4-phosphate synthase cleaves formic acid from ribulose 5-phosphate converting it in a thermodynamically irreversible reaction to 3,4-dihydroxy-2-butanone 4-phosphate (Volk and Bacher, 1990). 6,7-dimethyl-8-ribityllumazine synthase condenses 3,4-dihydroxy-2-butanone 4-phosphate with 5-amino-6-ribitylamino-2,4(1H,3H)-pyrimidinedione (obtained by chemical synthesis according to Plaut and Harvey, 1971) to 6,7-dimethyl-8-ribityllumazine in an also thermodynamically irreversible reaction (Kis et al., 1995).

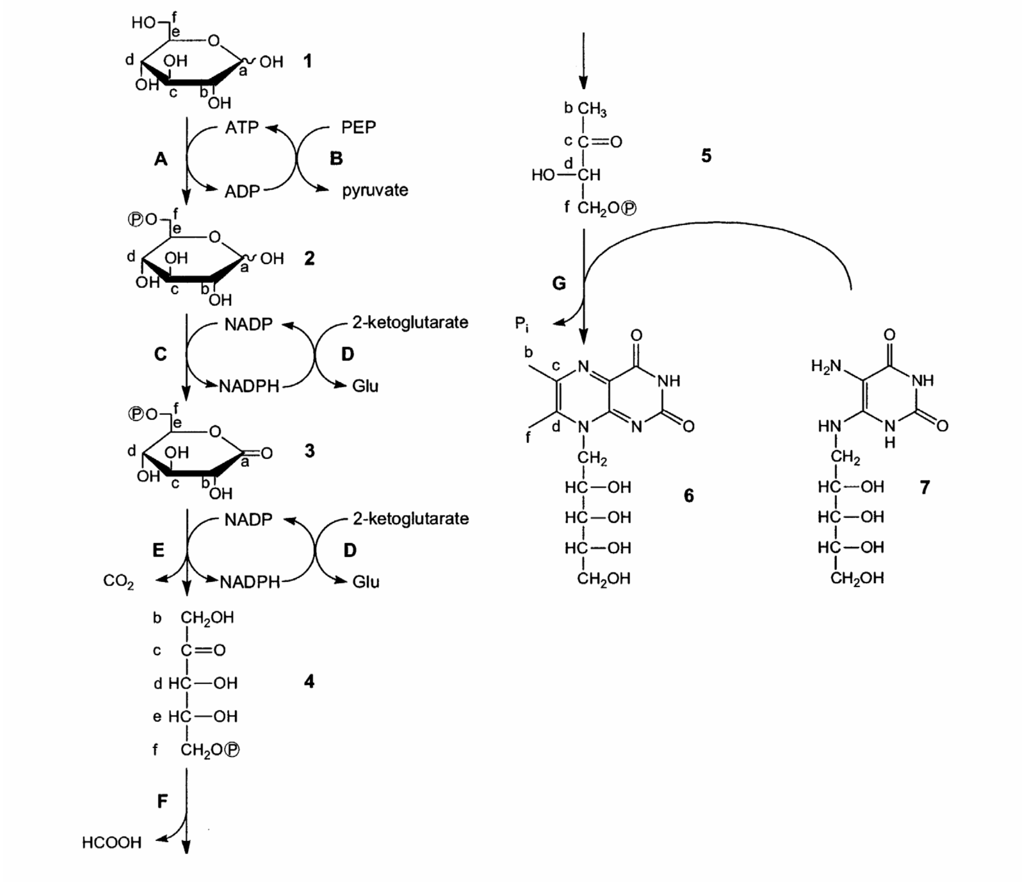

Figure 4.1: Enzymatic synthesis of 6,7-dimethyl-8-ribityllumazine (modified from Römisch et al., 2002). A, hexokinase; B, pyruvate kinase; C, glucose 6-phosphate dehydrogenase; D, glutamate dehydrogenase; E, 6-phosphogluconate dehydrogenase; F, 3,4-dihydroxy-2-butanone 4-phosphate synthase; G, 6,7-dimethyl-8-ribityllumazine synthase. 1, glucose; 2, glucose 6-phosphate; 3, 6-phosphoglucono-δ-lactone; 4, ribulose 5-phosphate; 5, 3,4-dihydroxy-2-butanone 4-phosphate; 6, 6,7-dimethyl-8-ribityllumazine; 7, 5-amino-6-ribitylamino-2,4(1H,3H)-pyrimidinedione.

6,7-dimethyl-8-ribityllumazine affords a maximum of 4^3 eq. 81 theoretically possible X groups. 29 of these X groups are detectable experimentally under optimal conditions (Appendix A.4). These X groups can be translated to glucose X groups by adding the wildcard symbol for positions C-1 and C-5. Therefore, each X group obtained from 6,7-dimethyl-8-ribityllumazine adds a distinct constraint for the computational deconvolution of the complete isotopolog space of glucose.

Reliability and consistency of the transformation of glucose to 6,7-dimethyl-8-ribityllumazine were tested using artificial glucose mixtures of commercially available ^{13}C-labeled glucoses. Glucose mixture model-1 contained 95 % natural abundance glucose and 5 % [U-$^{13}C_6$]glucose. Glucose mixture model-2 contained 95 % natural abundance glucose, 1 % [1,2-$^{13}C_2$]glucose, 1 % [2,3-$^{13}C_2$]glucose, 1 % [3,4-$^{13}C_2$]glucose, 1 % [5,6-$^{13}C_2$]glucose, and 1 % [U-$^{13}C_6$]glucose. Both mixtures were reconstructed using the X groups obtained from ^{13}C NMR spectroscopy of glucose or 6,7-dimethyl-8-ribityllumazine. The similarity between the weight values and the ^{13}C abundances obtained by computational deconvolution of X groups was > 94 % in all test sets. Each set of X groups, derived from glucose or 6,7-dimethyl-8-ribityllumazine was sufficient to reconstruct a glucose isotopolog space with 22 dimensions. Combination of the X group sets of glucose and 6,7-dimethyl-8-ribityllumazine for the deconvolution of model-2 afforded the same isotopolog space as the use the single X group sets caused by the high-abundance of unoccupied isotopolog species.

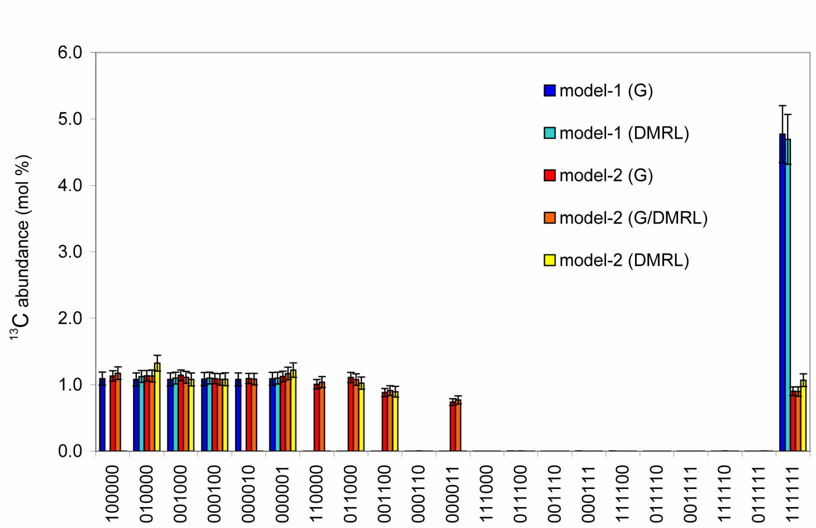

Figure 4.2: Consistency test of X groups derived from glucose (G) and 6,7-dimethyl-8-ribityllumazine (DMRL) by deconvolution in a glucose isotopolog space with 22 dimensions. Mixtures for model-1 and model-2 were based on commercially available ^{13}C-labeled glucoses. 6,7-dimethyl-8-ribityllumazine cannot yield information on C-1 and C-5 of glucose.

4.2 Extension of the glucose isotopolog space using mass spectroscopy

GC/MS analysis of glucose requires derivatization of glucose molecules with ethyl acetate. The mass spectrum was evaluated using two-dimensional cross-calibration and linear regression. Commercially available glucose isotopologs with known ^{13}C abundances were used for evaluation of the GC/MS spectrum (Figure 4.3; for details, see chapter 5.7). GC/MS affords maximal 7 isotopomer sets ([U-^{12}C$_6$]glucose, [^{13}C$_1$]glucose, [^{13}C$_2$]glucose, [^{13}C$_3$]glucose, [^{13}C$_4$]glucose, [^{13}C$_5$]glucose, and [U-^{13}C$_6$]glucose). The evaluation was able to afford all isotopomer sets, but [U-^{13}C$_4$]glucose and [U-^{13}C$_5$]glucose. For these glucose isotopomers no calibration values were available. Therefore, four additional X groups can be determined with high accuracy by GC/MS.

Fragmentation would increase the number of available X groups further, but would also require more complex derivatization procedures and evaluation techniques. With the X group set obtained by ^{13}C NMR spectroscopy of glucose and 6,7-dimethyl-8-ribityllumazine, an approach using mass spectrometry as first line analytical technique will be set aside for consecutively considerations.

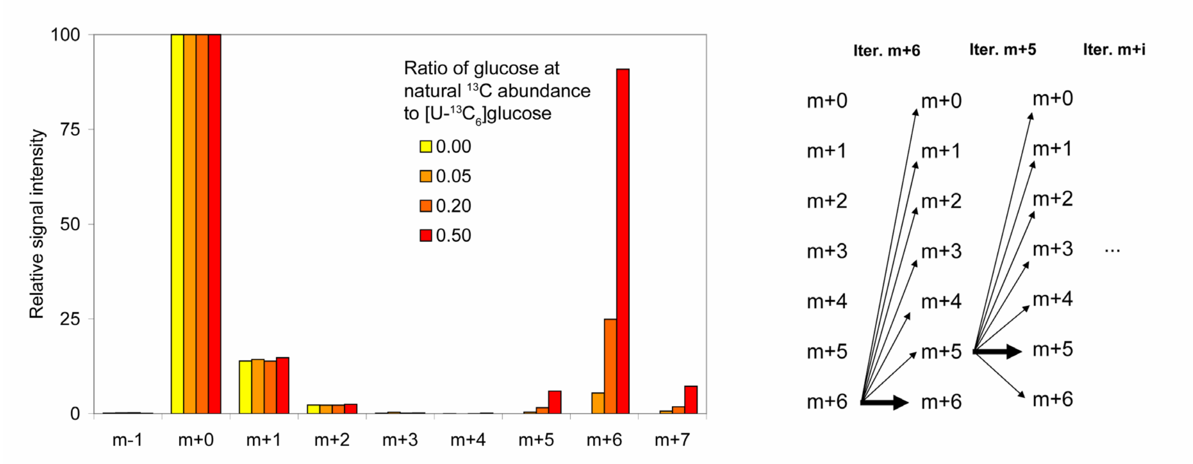

Figure 4.3: GC/MS spectrum of glucose with natural ^{13}C abundance and glucose with natural ^{13}C abundance containing 5 %, 20 % or 50 % [U-^{13}C$_6$]glucose. The determination of glucose isotopomer m_i requires the correction of peak m_i relative to the intensities of its adjacent peaks. Using defined mixtures of glucose isotopomers, the spectrum can be corrected and evaluated by cross-calibration (chapter 5.7). The iterative correction procedure of glucose can be initiated using peak intensities *m+0* or *m+6*.

4.3 Deconvolution of the 64-dimensional glucose isotopolog space

For the deconvolution of the 64-dimensional glucose isotopolog space, a similar approach as for the 22-dimensional glucose isotopolog space was chosen. Inconsistencies and experimental noise require a genetic algorithm to resolve local X group constraints obtained by NMR and GC/MS to the complete glucose isotopolog space. The matrix environment of the genetic algorithm described in chapter 3.3 was adapted to the increased number glucose isotopologs and X groups. Isotopolog definition vector *S* and isotopolog state vector *T* were expanded from 21 dimensions to 63 dimensions (cf. Figure 3.3). Integration of X groups from 6,7-dimethyl-8-ribityllumazine ^{13}C NMR spectroscopy and GC/MS led to a massive expansion of binary occupation matrix *D* and state matrix *H* (cf. Figure 3.3). The runtime parameters of the fitting process had to be modified in order to suppress oscillation and identification of local minima. Briefly, insertion of stochastic elements in the search process was reduced by 0.2 % and the resolution of the isotopolog state vector was increased to 16 bit or 32 bit, respectively (for details, see chapter 6.3.4).

For a definite, unequivocal deconvolution of the X groups to 63 glucose isotopologs, X groups from the glucose ^{13}C NMR experiment, the 6,7-dimethyl-8-ribityllumazine ^{13}C NMR experiment, and the glucose GC/MS experiment must be available. The ^{13}C NMR calibration value has to be the same for all experimental approaches, and is best determined by GC/MS or an additional IRMS experiment. Data consistency must be tested by evaluation of the course of the fitting process. Consistent and equally calibrated data has to be deconvoluted with different initial values for *T* to identify the stability of the minimum and isotopolog error bounds. If all deconvolution results for different initial values of *T* afford the same glucose isotopolog space, the solutions seems valid within its error bounds.

4.4 Metabolic interpretation of the complete glucose isotopolog space

The metabolic interpretation of a glucose isotopolog space with 64 dimensions will be focused on the metabolic network occurring in photosynthetic active plants and has to include all dominant anabolic and catabolic processes. The transformation of ^{13}C-labeling patterns of glucose by intermediary pathways has been extensively described in chapter 3.4 (formally encoded by transfer rules summarized in chapter 6.4.2). In the interpretation the possibility of generation of glucose from ^{13}C-labeled precursor molecules must be considered for each metabolic process (glycolysis/glucogenesis, pentose phosphate pathway, citric acid cycle). More specifically, ^{13}C isotopolog signatures generated by the Calvin cycle occur as non-homogenous population of glucose tracers must be included in the interpretation.

In the Calvin cycle, the enzyme ribulose 1,5-bisphosphate carboxylase converts $^{13}CO_2$ and ribulose 1,5-bisphosphate to two units 3-phosphoglycerate, one ^{13}C-labelled at C-1. The labeled precursors are converted to 1,3-bisphosphoglycerate by catalytic activity of phosphoglycerate kinase and to glyceraldehyde 3-phosphate by the enzyme glyceraldehyde 3-phosphate dehydrogenase. The triose phosphate moiety can be shuffled in the Calvin cycle (in reactions analog to the pentose phosphate cycle), or exported from the plastid to other metabolic processes (e. g. glucogenesis). The Calvin cycle can also generates $^{13}C_2$- and $^{13}C_3$-labeled fragments. The enzyme transaldolase can generate C-3 and C-4 labeled fructose 1,6-bisphosphate from glyceraldehyde 3-phosphate and dihydroxy-acetonephosphate each labeled at C-1. Within the Calvin cycle, one [U-$^{13}C_3$]labeled triose moiety will be generated after generation of two ^{13}C-3-labeled ribulose 1,5-bisphosphate by the transketolase reaction. $^{13}C_3$-labeled fragments can also generate $^{13}C_2$ fragments by photorespiration, which takes place in approximately a quarter of the reactions catalyzed by ribulose 1,5-bisphosphate carboxylase. In this reaction, ribulose 1,5-bisphosphate affords 3-phosphoglycerate and 2-phosphoglycolate by addition of O_2 instead of CO_2. This by-product regenerated in the cytosol and mitochondria and will only conserve labels at C-2 and C-3 of 3-phosphoglycerate.

4.5 Application example – metabolic fluxes in *Ruta graveolens*

Ruta graveolens is a traditional herb grown in mediterranean climate zones. It belongs to the *Rutaceae* family and its ingredients are used in food and drug industries (e. g., Pathak et al., 2003; Oliva et al., 2003). The major components are the bitter agent rutine (7 to 8 % w/w) and ethereal oils consisting of 2-hendecanone and 2-nonanone (1 to 2 % w/w).

The plant grows fast and shows a high resistance to environmental stress. It will be shown that *Ruta graveolens* is a suitable plant model for studying anabolic and catabolic metabolic systems using glucose as analyte.

Ruta graveolens was aerated for 5 hours with normal atmosphere containing 50 mol % $^{13}CO_2$, instead of the natural abundance of 1.1 % mol %. The CO_2 fraction in the atmosphere was determined at 500 ppb to 2000 ppb. After the ^{13}C perturbation period, the plants were frozen and ground. Glucose was isolated by water/methanol extraction from the leaf and the phylum of the plant. Purification of glucose was performed as described (chapter 5.3.1).

High-resolution ^{13}C NMR spectroscopy of glucose and 6,7-dimethyl-8-ribityllumazine afforded 56 X groups and 28 X groups, respectively (Figure 4.4 and 4.5). The ^{13}C NMR spectroscopy experiment of glucose was optimized by variation of the measurement temperature, compared to the experiments used for the evaluation of catabolic intermediary metabolic networks described in chapter 3. Signals 1α, 1β, 2β, 3α, 3β, 5β and 6β are best resolved at 300 K temperature. Signals 2α, 4α, 4β, 5α and 6α are best resolved at 350 K temperature. The 56 glucose X groups contain information from ^{13}C NMR experiments at both temperatures. ^{13}C perturbation of the system with $^{13}CO_2$ afforded 20 glucose X groups, which were not observed in the experiments using [U-$^{13}C_6$]glucose as tracer. Adjacent to the central peaks of 1α (93.28 ppm), 1β (97.20 ppm), 2β (75.41 ppm), 3α (74.05 ppm) and 6α/β (61.85 ppm, 61.97 ppm) a number of new X groups was identified, among the most informative (in term of non-redundancy to known X groups) and best-resolved are {10XX00}, {10XX11}, {101XX1}, {1100XX} and {0X1X11, 1X0X11} (Figure 4.4). The glucose X groups derived from 6,7-dimethyl-8-ribityllumazine contained only 6 redundant X groups already obtained from the direct NMR spectroscopy of glucose, including the 4 supersets for C-2 (21.04 ppm), C-3 (144.95 ppm), C-4 (152.30 ppm) and C-6 (18.09 ppm), {X010XX}, and {X111XX}. The remaining 22 X groups derived from the ^{13}C NMR spectra of 6,7-dimethyl-8-ribityllumazine are not available from the ^{13}C NMR experiment with glucose (Figure 4.5). GC/MS afforded 4 additional X groups for isotopomer sets containing one ^{13}C atom, two ^{13}C atoms, three ^{13}C atoms or six ^{13}C atoms. All X groups obtained from the ^{13}C NMR experiments were calibrated using the total ^{13}C abundance obtained in the mass spectroscopy experiment (3.87 mol %).

In total, 88 X groups were available for the computational deconvolution process. 64 dimensions constraint by 88 X groups results in an isotopolog/X group occupation matrix with 5544 elements. The X groups were resolved in 909 positive and 4635 negative constraints within the occupation matrix. The positive constraints can be divided in 544 constraints derived from the ^{13}C NMR spectrum of glucose, 324 constraints from the ^{13}C NMR spectrum of 6,7-dimethyl-8-ribityllumazine and 41 constraints derived from the GC/MS experiment with glucose. The deconvolution using a genetic algorithms afforded an average error of 0.07 mol % per X group. Tests including X groups exclusively derived from glucose or glucose and 6,7-dimethyl-8-ribityllumazine afforded only regional stabilities in the 63-dimensional glucose isotopolog space, even if 88 X groups (76 non-redundant X groups) should be sufficient to for the deconvolution process. This contradiction can be explained by the number of X groups consumed for the correction of the background noise level of the experimental system.

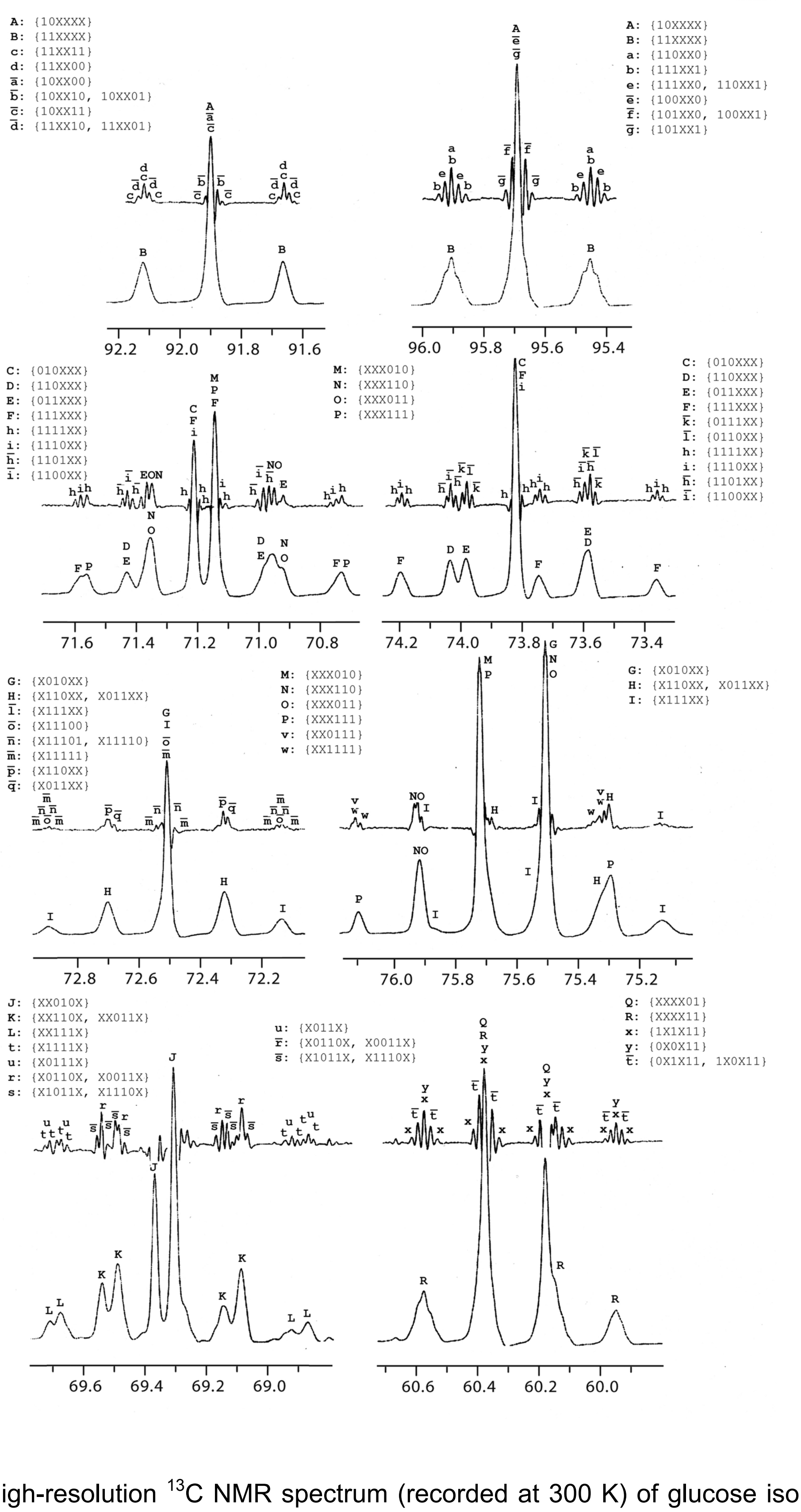

Figure 4.4: High-resolution ^{13}C NMR spectrum (recorded at 300 K) of glucose isolated from the leaf and phylum of *Ruta graveolens*. The ^{13}C quasi-equilibrium of the plant was perturbed using $^{13}CO_2$. Overlined letters indicate X groups, which could not be identified in experimental setups using [U-$^{13}C_6$]glucose as perturbant.

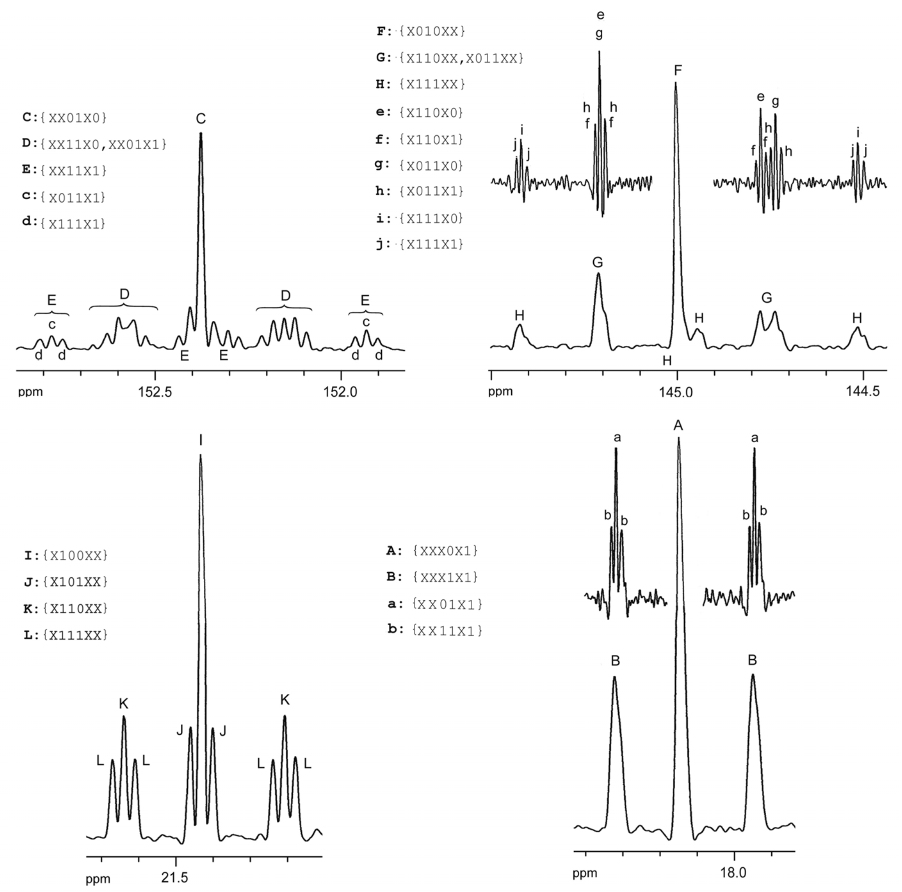

Figure 4.5: High-resolution ^{13}C NMR spectrum of 6,7-dimethyl-8-ribityllumazine, which resulted from enzymatic conversion of glucose isolated from the leaf and phylum of *Ruta graveolens*. The glucose X groups indicate the translation of the information available from 6,7-dimethyl-8-ribityllumazine by inserting the wild card symbol at the carbon positions C-1 and C-5, which are lost during the enzymatic conversions.

The share of these X groups can be estimated to $\sqrt{n}$ with *n* being the number of available X groups in total. Therefore, only 67 non-redundant X groups would be available for a 63-dimensional deconvolution problem leading to only regionally stable solutions, which in each case depend on the degree of consistency between the remaining X groups.

The deconvolution process of all 88 X groups afforded 63 occupied ^{13}C glucose isotopologs (Figure 4.6). The glucose isotopolog space includes 6 isotopomers carrying one ^{13}C atom (6.96 mol %), 15 isotopomers carrying two ^{13}C atoms (1.64 mol %), 20 isotopomers carrying three ^{13}C atoms (1.50 mol %), 15 isotopomers carrying four ^{13}C atoms (1.46 mol %), 6 isotopomers carrying five ^{13}C atoms (0.16 mol %) and one isotopomer carrying six ^{13}C atoms (0.31 mol %). The measured ^{13}C glucose equilibrium shaped by the biochemical network in the plant is far away from the natural abundance equilibrium. The impact of the perturbation can be estimated to a *Z* value of 0.056 (cf. chapter 3.5). The distribution of the ^{13}C excess is differing between the glucose isotopologs. For example, double-labeled glucose isotopologs are dominated by {110000} (0.18 mol %) and {000011} (0.40 mol %), triple-labeled glucose isotopologs are dominated by {111000} (0.20 mol %), {000111} (0.21 mol %), and {001110} (0.20 mol %), fourfold-labeled glucose isotopologs are dominated by {100111} (0.31 mol %) and {111001} (0.15 mol %). The ^{13}C excess over natural abundance of the two glucose isotopologs dominating the isotopolog space, {000011} and {011000}, was identified at 0.40 mol % and 0.26 mol % respectively. The isotopologs {110000}, {101000}, {000111}, {111000}, {001011}, {100011}, {110001}, {001101}, {111100}, {011110}, {100111}, {111001}, and {111111} were found in a ^{13}C concentration range between 0.13 and 0.25 mol %. The remaining glucose isotopologs were found > 0.13 mol %. These set also includes the isotopologs {000100}, {010000}, {000110}, {001011}, {101010}, {100101}, {101001}, {010110}, {011010}, {010111}, {101011}, {110011}, {011111}, {101111}, and {111011} which were found with ^{13}C abundance only slightly above the noise level of the system. Within the error bound, it is possible that these glucose isotopologs might not show a significant ^{13}C excess over natural abundance. Error bounds depend on the number of constraints available for a single isotopolog: {111111} was determined very exactly at 0.31 ± 0.01 mol % having 30 constraints available, while glucose isotopologs {000011} or {011000} were determined by 8 and 10 constraints and therefore showing higher error bounds, in detail 0.40 ± 0.07 mol % and 0.25 ± 0.05 mol %. Notably, many single-labeled and multiple-labeled glucose isotopologs show significant deviations in symmetry between C-1 to C-3 and C-4 to C-6. For example, {000100} *versus* {001000} differs by 0.10 mol % and {110000} *versus* {000011} differs by 0.22 mol %.

The metabolic interpretation of the detected glucose isotopologs is extensively discussed in chapters 3.4 and 4.4. The experimental approach generates significant amounts of [U-$^{13}C_6$]glucose and other multiple-labeled glucose isotopologs. The complete glucose isotopolog space can be deconvoluted by ^{13}C NMR spectroscopy of glucose, glucose enzymatically converted to 6,7-dimethyl-8-ribityllumazine, and mass spectroscopy. The glucose isotopologs are dominantly generated by reactions in the Calvin cycle. The results indicate a high retention time of Calvin cycle intermediates in the plastid during the biosynthesis of trioses and hexoses. The oxygenase activity of the enzyme ribulose 1,5-bisphosphate carboxylase can be estimated to approximately 5 % by the ratio of the glucose isotopologs {001000} and {000100}. Short retention times in the photosynthetic active compartment would result in high levels of single-labeled glucose isotopologs, since they will become highly diluted with the unlabeled metabolites in the plant.

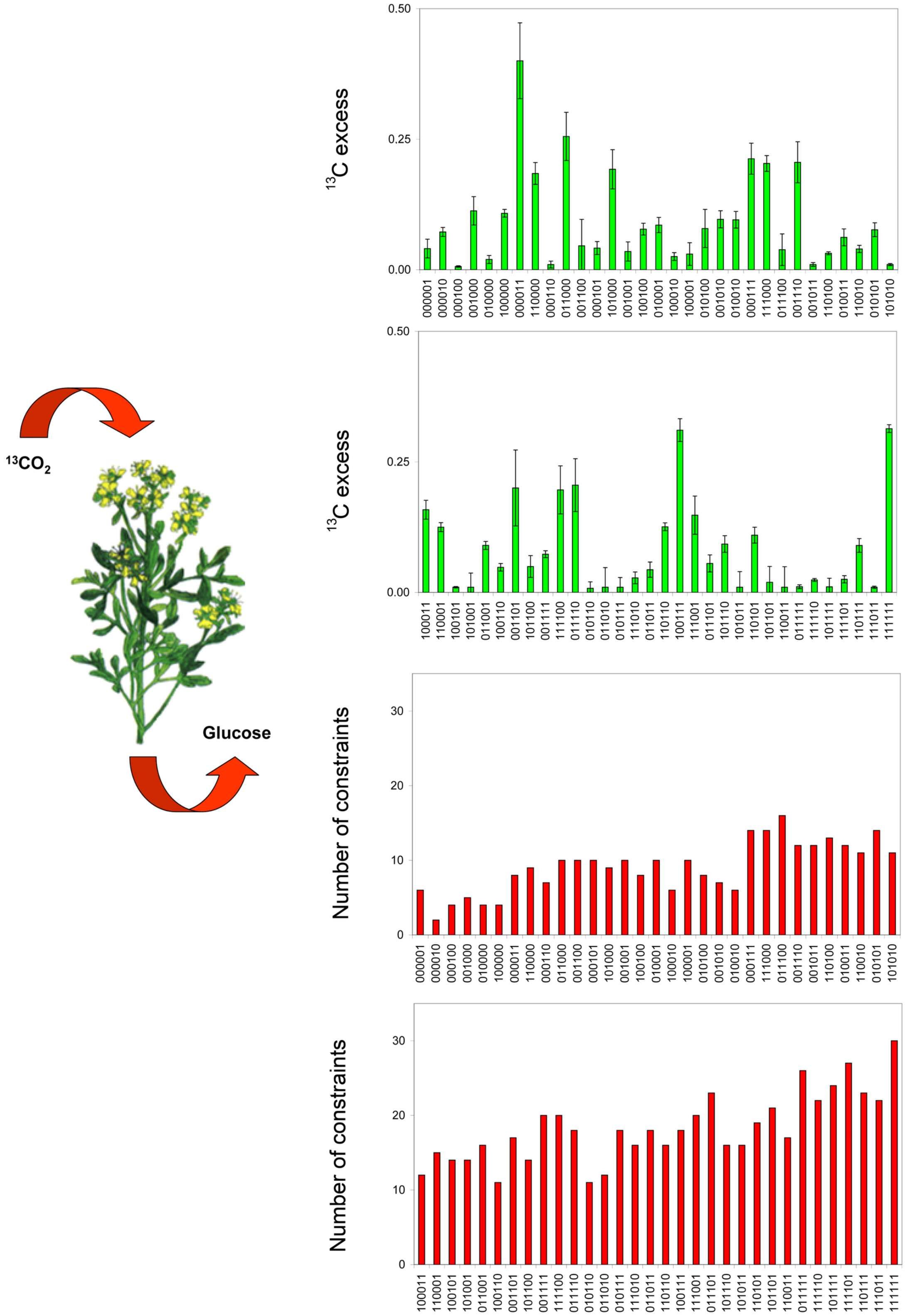

Figure 4.6: Complete glucose isotopolog space for *Ruta graveolens* (shown left). The metabolic network of the plant was perturbed by aeration with $^{13}CO_2$ for 5 h. The number of constraints per isotopolog indicates the degree of their determination by the available X groups obtained from glucose and 6,7-dimethyl-8-ribityllumazine .

The glucose isotopolog space is extensively shaped by the biochemical processes of the intermediary metabolism and creates a new ^{13}C equilibrium in the plant system.

Ruta graveolens is has been established as well-suited system for studying plant metabolism by perturbation of the ^{13}C quasi-equilibrium. The ^{13}C pulse-labeling techniques illustrated are not restricted to glucose, but may be a valuable tool for identifying biosynthetic routes and quantitative carbon fluxes in other plant metabolites. It can also be used to study the impact of environmental factors (CO_2 concentration in the atmosphere, heat, aridness, soil salinity) on an arbitrary photosynthetic active plant system.

5. Biochemical materials and methods

5.1 Materials and organisms

Chemicals were purchased from: Aldrich Chemie, Steinheim, Germany; Sigma-Chemie, Deisenhofen, Germany; Merck, Darmstadt, Germany; Serva Feinbiochemika, Heidelberg, Germany.

Enzymes were purchased from Aldrich Chemie, Steinheim, Germany and New England Biolab, Frankfurt/Main, Germany.

Isotope-enriched glucoses were purchased from Omicron, South Bend, Indiana. Isotope-enriched metabolites were purchased from Promochem, Wesel, Germany.

Standard laboratory equipment, NMR spectrometer and GC-MS were provided by the Chair of Organic Chemistry and Biochemistry, Technical University of Munich. IRMS was provided by Isolab GmbH, Schweitenkirchen, Germany.

Drosophila melanogaster Oregon-R* strain was obtained from the Max-Planck-Institut for Developmental Biology, Department of Genetics, Tübingen, Germany.

Nicotiana tabacum was obtained from the Department of Botany, Faculty of Biology, Ludwigs-Maximilians University, Munich.

Ruta graveolens was obtained from the Botanical Garden, Augsburg, Germany.

Tagetes patula and isolated benzofuran compounds were obtained from Lilla Margl, Chair of Organic Chemistry and Biochemistry, Technical University of Munich, Germany.

Taxus chinensis cell culture and isolated taxol were obtained from Joo Sun Son, Samyang Genex Biotech Research Institut, Daejeon, Korea, and the Department of Pharmacy, Martin-Luther-University, Halle, Germany.

Zea mays was obtained from the Chair of Genetics, Technical University of Munich, Weihenstephan, and the Horticultural Sciences Department, University of Florida.

5.2 Experimental cultures

5.2.1 *Drosophila melanogaster*

The standard culture medium (modified from Roberts, 1998) for preservation culture at ambient laboratory conditions contained, per liter: 80 g of corn flour, 10 g of soy flour, 8 g of dry yeast powder, 4 g of agar, 11 g of beet syrup, 40 g of malt extract, 5.6 g of [U-$^{13}C_6$]glucose, 6.3 ml of propionic acid, and 750 mg of methyl parabene. The culture medium for an incubator is modified and contains 10 g/L agar and 40 mg/L clotrimazole. Medium for *Drosophila melanogaster* with endosymbionts removed contained, per liter: 100 mg clotrimazole instead of 40 mg in standard medium, 100 mg rifampicine, 80 mg erythromycine. The medium was autoclaved. Glucose, propionic acid and methyl parabene and, where required, clotrimazole, rifampicine and erythromycine were added as filter-sterilized solutions.

In the Innova 4230 incubator (New Brunswick), cultures were kept at 27° C and 70 % air humidity in a 12 hour day/night rhythm. Flies were harvested one to two days after hatching and were stored at -20 °C.

5.2.2 *Nicotiana tabacum*

Tobacco plants (*Nicotiana tabacum* L, *cv. Petite Havana*) were grown under aseptic conditions from sterilized seeds or shoot cutting explanted on B5 medium (Manandhar and Gresshoff, 1980) containing per liter: 2 % (w/v) sucrose, 0.7 % (w/v) bacteriological agar, 4.0 g [U-$^{13}C_6$]glucose, 19.8 g of unlabelled sucrose.

The plants were grown in a new PsycroTherm incubator (New Brunswick) at 25 °C with 8/16 h dark/light illumination cycles at a light intensity of 0.5 - 1 W/m^2 (Osram L85W/25 Universal White fluorescent lamps). Tobacco leaves were harvested after a growth period of 20 days.

5.2.3 *Zea mays*

Maize kernels of inbred line w64a, b17 and mo17, mutants of w64a, and hybrids of b17 and mo17 were grown on a field in Gainesville, Florida, in summer 2003 and in a green house in Weihenstephan, Germany, respectively (see Tables 5.1 and 5.2).

Ears were harvested 8 to 10 days after pollination and processed as described (Glawischnig et al., 2002) with the following modifications. Protruding silks were cut off and 2 layers of outer husks were removed. The ear was sprayed with ethanol and flamed. After removing the inner husks and silks, the ear was dissected into two-row pieces with a total of four to ten attached kernels. The kernel blocks were placed in 25 x 100 mm Petri dishes containing 50 ml culture medium.

Abbreviation	**Genetic background**	**Name**	**Lesion**	**Growing environment**
w64a	w64a	Wisconsin 64A	Wild type	Field
ae1	w64a	*amylose extender*	Branching enzyme	Field
wx	w64a	*waxy*	Granule-bound starch synthase	Field
su1	w64a	*sugary1*	Debranching enzyme (isoamylase)	Field
sh1	w64a	*shrunken1*	Sucrose synthase	Field
bt1	w64a	*brittle1*	ADP-Glucose translocator	Field
bt2	w64a	*brittle2*	ADP-Glucose pyrophosphorylase (small subunit)	Field

Table 5.1: Maize wild type Wisconsin 64A (w64a) and its mutants were used in experiments for elucidating metabolic changes in kernel development and starch synthesis. Plants were grown in Gainesville, Florida, in summer 2002. Each type of kernels was grown on two types of media, supplemented with [U-$^{13}C_6$]glucose or [U-$^{13}C_{12}$]sucrose, respectively.

Abbreviation	**Genetic background**	**Name**	**Generation**	**Growing environment**
h1	b73	Iowa Stiff Stalk Synthetik 73 (BSSS)	P	Sterile culture
h2	b73 x mo17	Hybrid	F_1	Sterile culture
h3	mo17	Missouri 17	P	Sterile culture
h4	mo17 x b73	Hybrid	F_1	Sterile culture
h5	b73/mo17	Hybrid	F_2	Sterile culture
h6	b73	Iowa Stiff Stalk Synthetik 73 (BSSS)	P	Sterile culture
h7	mo17	Missouri 17	P	Sterile culture
h8	b73/mo17	Hybrid	F_2	Sterile culture

Table 5.2: Maize types used in experiments to determine metabolic changes occurring by the heterosis effect. P, F_1, F_2 indicate, whether the sample belongs to the parental generation or the first and second filial generation. Kernels from the samples h1 to h5 were grown on medium supplemented with [U-$^{13}C_6$]glucose. Kernels from the samples h6 to h8 were grown on medium containing [U-$^{13}C_{12}$]sucrose.

The culture medium contained per liter: 4.33 g Murashige-Skoog salts, 1.0 g casein hydrolysate, 20 mg gentamycine, 3.5 g agar, 1.0 mg 2,4-dichlorophenoxyacetic acid and 80 g glucose. The pH was adjusted to 5.8 prior to autoclaving.

Kernel blocks were initially cultured on unlabeled medium and incubated at 28°C in the dark. For studying metabolic changes in kernel development, at 11, 18 and 25 days after pollination, individual kernel blocks were transferred to fresh Petri dishes containing per liter of medium: 2.7 g [U-$^{13}C_6$]glucose and 77.3 g unlabeled glucose. They were harvested after incubation for 7 days. For studying starch and phytoglycogen biosynthesis, at 15 DPP kernel blocks were transferred onto fresh culture media containing per liter medium: 77.4 g/l glucose and 2.6 g/l [U-$^{13}C_6$]glucose or 77.4 g/l sucrose and 2.6 g/l [U-$^{13}C_{12}$]sucrose, respectively. Kernel blocks were harvested after incubation for 7 days. Tissue samples were frozen in liquid nitrogen and stored at –70°C.

5.3 Isolation and purification of metabolites

5.3.1 Glucose

Glucose was obtained either from plant or animal cell mass, starch, or phytoglycogen. All require the production of a crude extract and subsequent purification by HPLC.

Glucose was isolated from plant or animal cells by grounding the cell mass with liquid N_2. The cell mass was allowed to warm up to room temperature and suspended in 300 ml of water. The suspension was stirred at 37 °C for 1h and centrifuged at 6,000 rpm (GS3 rotor) for 20 minutes. The pellet was re-suspended in 300 ml water and stirred at 37 °C for 1h and again centrifuged. The extracts were combined, filtrated, and concentrated under reduced pressure.

Glucose was isolated from starch storage tissues and phytoglycogen by drying the kernels in a lyophilizer for 5 days. The pericarp and embryo were removed by dissection. The endosperm tissue was ground in liquid N_2. For the isolation of phytoglycogen (Inouchi et al., 1983), the endosperm was extracted with hexane to remove soluble sugars. The residue was extracted with water. Phytoglycogen was precipitated from the watery extract by the addition of three volumes of methanol. The mixture was centrifuged and the pellet was washed with methanol. Phytoglycogen was diluted in 0.1 M citrate buffer (pH 4.6) and then degraded to glucose by the addition of amyloglucosidase (60 U) for 5 h at 55°C. For isolating glucose from the remaining starch, the remaining endosperm residue was boiled with 80% (v/v) ethanol for 15 min to remove soluble sugars. The mixture was centrifuged. The procedure was repeated twice. The residue was re-suspended in 2 ml of 0.5 M NaOH for 1h at 65°C to gelatinize starch. The mixture was dissolved in water and adjusted to pH 4.5 with 1 M acetic acid. Starch was degraded to glucose by the addition of 60 U amyloglucosidase for 5 h at 55°C. Subsequently glucose was purified by affinity chromatography as described (Glawischnig et al. 2002).

All extracts as obtained by one of the three above steps, were further purified by HPLC using a Luna 5u NH2 column (250 x 21.2 mm, 40 ml/min, 80°C, Phenomenex) and differential refractrometric monitoring (Gamma Analysen Technik GmbH, Bremerhaven, Germany).

Glucose eluted at 8.2 min in 80 % (v/v) acetonitrile. Pure glucose was concentrated under reduced pressure. The final concentration of glucose was determined using an AccuChek blood glucose monitor device (Roche Diagnostics). The solution was evaporated to dryness under reduced pressure and glucose was stored at -80° C.

5.3.2 Nucleosides

Purification protocols for nucleosides were obtained from Eisenreich, 1991 and modified. Grinded tissues were extracted with acetone and methanol (1:1 v/v). The remaining cell mass was suspended in 1 M sodium hydroxide solution and stirred for 24 h at room temperature. Cellular proteins were precipitated by adding a mixture (1:1 v/v) of 10 M hydrochloric acid and 25 % trichloride acetic acid with a ratio of 1:10 to the suspension volume. Precipitated proteins were separated by centrifugation (SS 34 rotor, 16,000 rpm, 4° C, 15 minutes).

The supernatant contains ribonucleotides. The purification protocols for nucleosids are modified from Eisenreich, 1991. The pH had to be adjusted to 8.0 by adding 25 % ammonia. The solution became diluted 1:8 with ethanol (p. a.) and was cooled to 4° C. By adding 1:16 (v/w) barium acetate and letting the solution rest for 12 h, nucleotides were precipitated as barium salts. The salts became separated by centrifugation (SS 34 rotor, 16,000 rpm, 4° C, 15 minutes). The pellet was resuspended in 50 ml 50 mM hydrochloric acid. 2 ml of a solution saturated with sodium sulfate was added. Precipitated barium sulfate was separated by centrifugation (SS 34 rotor, 16,000 rpm, 4° C, 15 minutes). The supernatant was purified by ion-exchange chromatography using a column with Dowex 50 WX8 (H^+ form, 200 – 400 mesh, 2 x 34 cm). The column was developed with 160 ml 50 mM hydrochloric acid and further eluated with water (flow rate 100 ml/h). Fractions of 20 ml were collected and analyzed photometrically (254 nm) and by thin film chromatography (cf. Table X). For the given column the retention volumes were: UMP, 100 ml; GMP, 300 ml; CMP and AMP, 1000 ml. CMP and AMP were separated by reverse phase HPLC using a Lichrosorb RP_{18} (16 x 250 mm) column. The column was developed with 100 mM ammonia formiate/100 mM formic acid. Nucleotides were detected photometrically. CMP was eluated at 100 ml, AMP at 260 ml. Solvent was removed under reduced pressure.

The nucleotides were suspended in 5 ml 0.5 mM magnesium chloride in 0.1 M ammonia hydrogen carbonate (pH 10.4). The solution was incubated for 7 hours at 37° C with basic phosphatase (CMP, UMP, GMP with 30 U basic phosphatase II and AMP with 100 U basic phosphatase). Nucleosides were purified by reverse phase HPLC with a Lichrosorb RP_{18} (16 x 250 mm) column. For purification of uridine, cytidine and guanosine the column was eluated with water. For purification of adenosine the column was eluated with 5 % methanol (p. a.). The retention volumes were: uridine, 135 ml; cytidine, 145 ml; guanosine, 420 ml; adenosine, 530 ml. The nuclesides were concentrated under reduced pressure.

R_f values of nucleotides, nucleosides and amino acids			
Lysine	0.36	Tyrosine	0.62
Glycine	0.41	Valine	0.65
Aspartic acid	0.42	2'-AMP	0.65
Serine	0.43	3'-GMP	0.65
Histidine	0.43	Methionine	0.69
Arginine	0.44	Phenylalanine	0.71
Adenosine	0.45	2'-GMP	0.73
Glutamic acid	0.48	Cytidine	0.77
Threonine	0.49	Uridine	0.79
Guanosine	0.51	Isoleucine	0.86
Alanine	0.53	2’-/3’-UMP	0.92
3'-AMP	0.56	2’-/3’-CMP	0.95

Table 5.3: Detection of nucleotides, nucleosides and amino acids by thin layer chromatography (modified from Eisenreich, 1991). For nucleotides and nucleosides, cellulose F_{254} layers with 3 % ammonia chloride as solvent were used. Detection was performed optically by 254 nm. For amino acids, cellulose F_{254} layers with n-butanol, glacial acetic acid and water (50/20/30 v/v/v) as solvent were used. Detection was performed using a solution of 5 % ninhydrine.

5.3.3 Amino acids

Proteins from grinded tissues were precipitated as described (cf. chapter 5.3.2). The purification protocols for amino acids are obtained in modified form from Eisenreich, 1991. Tyrosine, phenylalanine, lysine, histidine, arginine, glutamic acid and aspartic acid, methionine, isoleucine, serine, threonine, glycine, alanine, valine and proline can be obtained by acidic hydrolysis of the cellular proteins. Tryptophane can be obtained by alkaline hydrolysis of the cellular proteins.

For acidic hydrolysis, the pellet was lyophilized and resuspended in 200 ml 6 M hydrochloric acid and 8 ml thioglycolic acid. The mixture was stirred for 24 hours at 110° C in N_2 atmosphere. The resulting solution was filtered and the fraction of hydrochloric acid and thioglycolic acid was minimized under reduced pressure resulting from a water-jet pump.

For alkaline hydrolysis, the lyophilized pellet was suspended in 580 ml 2 M barium hydroxide. The mixture was stirred for 24 hours at 120° C. After cooling down, the mixture was diluted with 3.5 liter water. The mixture was stirred and neutralized by aerating with CO_2. Precipitated barium carbonate was separated by centrifugation (SS 34 rotor, 6,000 rpm, 4° C, 10 minutes). The supernatant was concentrated under reduced pressure resulting from a water-jet pump.

Further purification by reverse phase and ion-exchange chromatography yields the single amino acids and is described in the following section. Note that the retention volumes can vary in a broad range, while the order of the elution of the amino acids remains the same.

5.3.3.1 Purification of tyrosine, phenylalanine, lysine, histidine and arginine

The mixture of amino acids resulting from acidic hydrolysis became separated by cation-exchange chromatography. The mixture was suspended in 50 ml water and piled on a column with Dowex 50 WX8 (ammonia form, 200 – 400 mesh, 3.2 x 28 cm). The column was developed with 3.5 liter 0.2 M ammonia formiate (pH 4.5), with 1.5 liter 0.2 M ammonia formiate (pH 6.5) and with 1 liter 0.5 M ammonia formiate (pH 6.5). 20 ml fractions were collected. Amino acids were detected by thin layer chromatography (Table X) and the chemical reaction with ninhydrine. Amino acids were eluated in the following order: neutral and acidic amino acids, tyrosine and phenylalanine, lysine, histidine, arginine. Related fractions were merged and lyophilized.

Tyrosine and phenylalanine were separated by reverse phase HPLC using a Lichrosorb RP_{18} column (16 x 250 mm). Tyrosine and phenylalanine were eluated with water and detected photometrically at 280 nm. The retention volume was: tyrosine, 85 – 120 ml; phenylalanine, 140 – 180 ml. Tyrosine and phenylalanine were concentrated under reduced pressure.

5.3.3.2 Purification of glutamic and aspartic acid

Glutamic and aspartic acid were separated from neutral amino acids by anion-exchange chromatography. The amino acid mixture was piled on a column with Dowex WX8 (formiate form, 200 – 400 mesh, 15 x 340 mm). The column was developed with 500 ml water and 500 ml 100 mM formic acid. Neutral amino acids were eluated with water, glutamic and aspartic acid were eluated with 100 mM formic acid. The charged amino acids were detected by photometry at 206 nm and refractrometry. The retention volume was: glutamic acid, 80 – 120 ml; aspartic acid, 100 – 220 ml. Glutamic and aspartic acid were concentrated under reduced pressure.

5.3.3.3 Purification of methionine and isoleucine

Neutral amino acids were purified by cation-exchange chromatography. The amino acid mixture was piled on a column with Dowex 50 WX8 (ammonia form, 200-400 mesh, 31 x 280 mm). The column was developed with 0.2 M ammonia formiate (pH 3.0). 20 ml fractions were collected. At a retention volume of 1500 – 1700 ml methionine and isoleucine were eluated. Methionine and isoleucine were separated by reverse phase HPLC using a column with Lichrosorb RP_{18} (16 x 250 mm). The column was developed with water. Detection occurred with photometry at 206 nm and refractrometry. The retention volume was: methionine, 80 – 120 ml; isoleucine, 100 – 130 ml. Methionine and isoleucine were concentrated under reduced pressure.

5.3.3.4 Purification of serine, threonine, glycine and alanine

Serine, threonine, glycine and alanine were purified by cation-exchange chromatography. The amino acid mixture was piled on a column with Dowex 50 WX8 (H^+ form, 200-400 mesh, 23 x 400 mm). The column was developed with 200 ml water, and subsequently with a linear gradient of hydrochloric acid (0 – 3 M, 2 liter total volume). The retention volumes were: serine and threonine, 720 – 800 ml. glycine, 790 - 890 ml; alanine, 860 – 900 ml. Related fractions were merged and concentrated under reduced pressure. Serine and threonine were further purified by anion-exchange chromatography using a column with Nucleosil SB10 (formiate form, 16 x 250 mm). The column was developed with water. The amino acids were detected photometrically at 206 nm and by refractrometry. The retention volume for serine and threonine was 30 – 60 ml. The fractions were concentrated under reduced pressure.

5.3.3.5 Purification of tryptophane

The mixture resulting from alkaline hydrolysis of cellular protein was purified by reverse phase HPLC. It was piled on a column with Lichrosorb RP_{18} (16 x 250 mm). The column was developed with 100 mM ammonia formiate. Tryptophane was detected photometrically at 280 nm. The retention volume was 415 ml. The solution was concentrated under reduced pressure. For further purification the reverse phase HPLC was repeated using 5 % methanol as eluant. The retention volume was 240 ml. The solution was concentrated under reduced pressure.

5.4 Enzymatical conversion of glucose to 6,7-dimethyl-8-ribityllumazine

A solution containing 100 mM tris hydrochloride (pH 7.8), 10 mM $MgCl_2$, 10 mM DTT, 0.02 % sodium azide, 3 mM ^{13}C-labeled glucose, 0.05 mM ATP, 3 mM phosphoenolpyruvate, 0.05 mM $NADP^+$, 6 mM α-ketoglutarate, 6 mM NH_4Cl, 1.5 mM 5-amino-6-ribitylamino-2,4(1*H*,3*H*)-pyrimidinedione (freshly prepared by hydrogenation of 5-nitroso-6-ribitylamino-2,4(1*H*,3*H*)-pyrimidinedione according to Plaut and Harvey, 1971), 100 U of hexokinase, 100 U of pyruvate kinase, 50 U of glucose 6-phosphate dehydrogenase, 50 U of 6-phosphogluconate dehydrogenase, 100 U of glutamate dehydrogenase, 10 U of phosphoriboisomerase, 3 U 3,4-dihydroxy-2-butanone 4-phosphate synthase from *Escherichia coli*, and 3 U 6,7-dimethyl-8-ribityllumazine synthase from *Bacillus subtilis* in a total volume of 200 ml, was incubated in the dark for 72 hours under an atmosphere of argon at room temperature. During the first hours of incubation, the pH was repeatedly adjusted to 7.8 by the addition of 3 M NaOH.

The solution was applied to a Florisil (100 x 28 mm) column which was washed with 400 ml of water and developed with 200 ml of acetone/water (1:4 v/v) followed by 600 ml of acetone/water (1:1, v/v). The solution was evaporated to dryness under reduced pressure. The residue was dissolved water and centrifuged. The supernatant was applied to an AG 50 WX8 (H^+ form, 100 x 28 mm) column which was developed with bidistilled water. Fractions containing 6,7-dimethyl-8-ribityllumazine were combined and evaporated to dryness.

5.5 NMR spectroscopy

All spectra were recorded using a four-channel Avance Bruker DRX500 spectrometer (Bruker Instruments, Karlsruhe, Germany) equipped with an inverse triple resonance probe head. Data was initially stored and processed with Bruker XWinNMR 3.0. Detailed evaluation of the 1D spectrum was performed with Bruker WinNMR 2.6. Detailed evaluation of the 2D spectra was performed with Bruker XWinNMR 3.0 and Aurelia 2.8.11.

Samples were dissolved in 0.55 ml D_2O. ^1H and ^{13}C NMR spectra were recorded at 500.13 MHz and 125.76 MHz, respectively, using the Bruker DRX500 spectrometer at 27 °C. Water suppression was achieved by presaturation of the residual water signal at the lowest possible power level. For ^1H-decoupling, a composite pulse sequence (WALTZ) was used in the ^{13}C NMR experiments. In processing the spectra, prior to Fourier transformation the FID was zero-filled to 256,000 and multiplied with a Gaussian function (GB, -2; LB, 0.4). Prior to integration, the baseline of the spectra was corrected and the chemical shift was calibrated. The evaluation of NMR spectra for determination of single isotopologs and groups of isotopologs was described in detail in chapter 1.1.4 and 3.7.1.

5.6 IRMS

Total ^{13}C-enrichments were determined by ^{13}C-isotope ratio mass spectrometry by Isolab GmbH, Germany. The ^{13}C abundance A in a compound is calculated by referencing the $^{13}C/^{12}C$ ratio of the sample R_S to the Pee Dee Belemnite (PDB) standard R_R (for details confine Schoell et al., 1983), thus $A = \frac{R_S - R_R}{R_R}$.

5.7 GC mass spectroscopy

Glucose must be derivatized before GC/MS measurements. Glucose with a concentration of 5 mg/ml were mixed 1:10 (v/v) with 0.38 M sulphuric acid. The mixture was dissolved in acetone (1:40 v/v) and incubated for 1h at room temperature. Subsequent it is mixed 1:1 (v/v) with saturated NaCl solution and 1:2 (v/v) with sodium carbonate. Ethyl acetate is added 1:1 (v/v) to the mixture. After shaking, the glucose derivate is solubilized in the ethyl acetate phase. The solvent is evaporated under N_2. The residue is resuspended in 200 µl ethyl acetate/acetic acid anhydride (1:1, v/v). The mixture is incubated by 60° C for 12 h before it can be used for the measurement.

The total ^{13}C abundance for a carbon atom of an arbitrary compound is calculated by weighting the each mass m_i by its peak intensity p_i and correcting the obtained values by the ^{13}C natural abundance peak intensity of the sample. By normalizing the value to the *n* carbon atoms of the compound and adding the natural abundance, the total ^{13}C abundance for the compound is obtained.

$$A = \frac{1}{n} \cdot \left(\frac{\sum_{i=0}^{n} m_i \cdot p_{E,i}}{\sum_{i=0}^{n} p_{E,i}} - \frac{\sum_{i=0}^{n} m_i \cdot p_{NA,i}}{\sum_{i=0}^{n} p_{NA,i}} \right) + A_{NA}$$

Abundances for single isotopomer groups were calculated by two-dimensional cross-calibration with known isotopolog glucose compostions from commercially available glucoses. Calibration function were calculated by linear regression of GC/MS signal intensities derived from measurements of natural abundance glucose and 5 %, 20 % or 50 % of the commercially available glucose isotopologs ([$^{13}C_1$]glucose, [$^{13}C_2$]glucose mixtures, [$^{13}C_3$]glucose mixtures and [U-$^{13}C_6$]glucose).

6. Computational methods

6.1 Software environment

Numerical and computational algorithms were implemented in C++ and C# using the Microsoft Development Environment 7.1 and .NET Framework 1.1 on the Microsoft Windows platform.

The rule-based algorithms were tested in SWI-Prolog 5.2.13, and ported to C# for performance and integrity.

For genetic algorithms and neural networks, the software packages GeneHunter, NeuroShell predictor and NeuroShell classifier from Ward System, Frederick, USA, were utilized.

Complex calculations of standard algorithms were performed with MatLab 6.5 R12 including the statistics toolbox 5.0.1 and signal processing toolbox 6.2.1.

High-performance computing was carried out on a Hitachi SR8000-F1 in the Leibniz supercomputing center (project t363102/1). The following development tools were used on this platform: GNU C/C++ compiler gcc/g++ 2.95.2 and TotalView 5.0. The MPI 1.1 library was used for process distribution and coordination.

6.2 Statistical methods

This chapter defines the statistical methods applied within the scope of this thesis. All statistical calculations were carried out using the software packages MatLab 6.5 R12, Microsoft Excel 10 and SPSS 13.0.

6.2.1 Basic definitions and error propagation

The root mean square deviation (RMSD) σ is a unit for measuring the similarity of two sets $X = \{x_1,...,x_i,...x_n\}$ and $Y = \{y_1,...,y_i,...y_n\}$ of the same magnitude *n* and the same codomain range, $\delta = \sqrt{\frac{1}{n} \cdot \sum (x_i - y_i)^2}$. Since the sets *X* and *Y* can be arbitrary normalized, σ can only be used for relative conclusions.

The variance $\delta(X)^2$ of a set *X* is defined by $\delta(X)^2 = \sum_{i=1}^{n} (x_i - \hat{x})^2 \cdot p_{x_i}$, if each element *x*_{*i*} can be attributed with a significance value p_{x_i} in the range [0;1]. The standard deviation δ is calculated by $\delta = \sqrt{\delta(X)^2}$. The covariance ϕ of two sets *X* and *Y* indicates the degree of dissimilarity between the sets and is defined as $\phi(X,Y) = \sum_{i=1}^{n} (x_i - \hat{x})^2 \cdot p_{x_i} \cdot (y_i - \hat{y})^2 \cdot p_{y_i}$. Alike sets are characterized by $\phi(x) = 0$.

Errors are propagated if value *y* is calculated from a parameter set *X*. The propagated error Δ*y* generated by *n* elements x_i each affected with an error Δx_i is calculated according $\Delta y = \sqrt{\sum_{i=1}^{n} \left(\frac{\partial y}{\partial x_i} \right)^2 \cdot \Delta x_i}$ (Gaussian error propagation rule).

.

6.2.2 Bravais-Pearson correlation

The Bravais-Pearson correlation coefficient τ shows the statistical significance of the assumption of a linear correlation between two sets *X* and *Y*. The sets must both normally distributed and the size of the sets must be in the same order of magnitude. For the calculation of τ, the covariance $\phi(X,Y)$ is normalized by the variance of the two sets, $\tau = \frac{\phi(X,Y)}{\sqrt{\delta(X)^2 \cdot \delta(Y)^2}}$. The degree of correlation is defined by three intervals: $|\tau| < 0.5$ shows a weak correlation, $0.5 \le |\tau| < 0.8$ shows a medial correlation, and $0.8 \le |\tau|$ shows a strong correlation between the two sets.

6.2.3 χ^2 test

The χ^2 test shows the correlation of two classified distributions *X* and *Y*. It is used to test if the empirical distribution *Y* matches theoretical distribution *X* or other another arbitrary empirical distributions. Both sets must have the same magnitude *n*. The test can be calculated by the formula $\chi^2 = \sum_{i=0}^{n} \frac{(x_i - y_i)^2}{x_i}$. The statistical significance of χ^2 can be verified by comparing with the critical value for χ^2 with *n-1* degree of freedom (for tables, see Boertz et al, 2000).

6.2.4 t-test

The t-test is possible for non-connected sets and interconnected sets *X* and *Y* of the magnitudes n_x and n_y. Precondition is the Gaussian distribution of both sets. For non-connected, classified sets the *t* value is calculated by:

$$t_{nc} = \frac{|\hat{x} - \hat{y}|}{\sqrt{\frac{n_x + n_y}{n_x \cdot n_y} \cdot \frac{(n_x - 1)\cdot \delta(X)^2 + (n_y - 1)\cdot \delta(Y)^2}{n_x + n_y - 2}}}$$

For interconnected sets of the same magnitude *n*, the *t* value of the test must be calculated according to the formula:

$$t_{nc} = \frac{|\hat{x} - \hat{y}|}{\sqrt{\frac{\sum_{i=1}^{n}(x_i - y_i)^2 - \frac{1}{n}\cdot\left(\sum_{i=1}^{n}(x_i - y_i)\right)^2}{n\cdot(n-1)}}}$$

The obtained value *t* must be compared with the critical value with *n-1* degree of freedom to test for significance of the result (for tables, see Boertz et al, 2000).

6.3 Function optimization techniques

Minimization and maximization of functions are equivalent problems and are interconvertable by exchanging the sign of an arbitrary function. To simplify the following descriptions, only the identification of minima will be illustrated. For nonlinear functions with one minimum the steepest descent, line search and conjugated gradient algorithms have been implemented (for additional information and source code examples, see Press et al, 1992). For optimization of non-linear functions with multiple minima, genetic and evolutional optimization techniques have been used to identify minima (see chapter 3.3).

6.3.1 Steepest descent algorithm

The steepest descent algorithm is the most suitable optimization algorithm for an arbitrary convex function *f*. A function is convex, if for two arbitrary positions x_1 and x_2 in an *n*-dimensional space, the condition $f(\lambda \cdot x_1 + (1-\lambda) \cdot x_2) \leq \lambda f(x_1) + (1-\lambda) \cdot f(x_2)$ holds true within a range λ, $0 \leq \lambda \leq 1$, of the minima. The steepest descent algorithm follows the gradient from a random starting position x_0 to the minimum x_n by iteratively shifting x_i by Δx_i, $\Delta x_n = -\alpha \cdot \nabla f(x_n)$. The parameter α indicates the stepwidth of the descent. α has to be adapted to the steepness of the descent and the size of broadness of the minima to avoid loss of performance and oscillation.

6.3.2 Line search algorithm

The line search algorithm can be used for convex functions and functions with one non-pathological minimum. It descents along the gradient of a target function similar to the steepest descent algorithm. After setting an initial value for the iteration stepwidth, the parameter α is automatically adapted to the course of the target function applying $\alpha_{n+1} = \min(-\nabla f(x_n + \alpha_n))$ for the calculation of α in the next iteration.

6.3.3 Conjugated gradient algorithm

The conjugated gradient algorithm descents the target function to a minimum by following the path of minimal changes in the gradient of the target function. Starting from an arbitrary position x_0 and with an initial stepwidth $\alpha_0 = -\nabla f(x_0)$, target function *f* is minimized by $\Delta x = -\alpha_0 \cdot \nabla f(x_0)$. In the post-initial steps parameter β is calculated by $\beta_n = \frac{\nabla f(x_{n+1})^T \cdot (\nabla f(x_{n+1}) - \nabla f(x_n))}{\nabla f(x_n)^T \cdot \nabla f(x_n)}$. The target function is then minimized in the direction of the conjugated gradient $\Delta x_{n+1} = -\nabla f(x_{n+1}) + \beta_n \nabla f(x_{n+1})$. The algorithm terminates if $f(x_n)$ does not improve further.

6.3.4 Genetic techniques

Optimization procedures can be formulated as search problem in genetic and evolutional algorithms. Optimization through genetic techniques is performed by using the GeneHunter library (WardSystems Inc., Frederick, Maryland). Deconvolution of the search space is modeled by creating a binary occupation matrix *D* which assigns the search patterns to resolved signatures with unknown abundances encoded in Vector *T*. A state evaluation matrix of the search space, can thus defined by $H_{ij} = D_{ij} \cdot T_i$. By calculating the similarity *F* between the provided constraints *X* and the state evaluaiton matrix *H*, $F = \sum_j \left| X_j - \sum_i H_{ij} \right|$ the current iteration state can be evaluated. The target function of the algorithm is expressed by $\lim F \to 0$. The chosen resolution μ of *T,* the magnitude of the domain of the search space in each dimension E_i, and the magnitude of dimensions define the size of the complete search space $S = \prod_{i=0}^{\dim(S)} (\mu \cdot E_d)$. The search space can be contracted in the course of iteration by modifying the size of the binary representation of T_i. For this purpose, the algorithm has to start with a low-resolution search and increase the resolution of T_i in course of the optimization (Table 6.1).

The iteration is started by creating random sets of *T*. All occupations of *T* are evaluated by *F*. Based on the best occupations of *T*, a new *T* set is created (elite strategy) by generating new vectors of *T* by random exchange of elements T_i (crossover rate, Table 6.1). Furthermore, some of the best elements of the preceding *T* set are chosen randomly and inserted in the educed *T* set (generation gap, Table 6.1) to contain solutions with high degrees of similarities between constraints and solution for subsequent iterations.

The magnitude of similarity *F* of each *T* in the solution domain will thus increase by a growing number of iterations. Exchanges of single elements T_i by random values (mutation rate, Table 6.1) enhances solution space and avoids that the algorithm will remain in local minima. Each optimization must therefore continue for at least 2,000 iterations after a stable optimum has been identified. Mutations in *T* occurring in this time period can be used to test if the minimum is a local or global minimum of the search space.

Parameter	21 glucose isotopologs, raw search process	21 glucose isotopologs, refinement process	63 glucose isotopologs, raw search process	63 glucose isotopologs, refinement process
$\|T\|$	50	50	75	75
Crossover rate	90 %	90 %	85 %	85 %
Mutation rate	1.0 %	1.0 %	0.8 %	0.8 %
Generation gap	2.0 %	2.0 %	2.0 %	2.0 %
Resolution of T_i	2^8	2^{16}	2^8	$2^{16}/2^{32}$

Table 6.1: Parameters for a genetic algorithm (software package GeneHunter, Wardsystems, Frederick, USA) to deconvolute glucose isotopologs containing one or more ^{13}C atoms resulting from catabolic (21 isotopologs) or catabolic and anabolic processes (63 isotopologs).

6.4 Metabolic simulation techniques

Metabolic simulation was performed using the software package 4F developed within the scope of this thesis (cf. chapter 3.6). 4F supports pattern propagation and rule-based deconvolution of metabolic processes. Standard simulation software were evaluated (cf. chapter 6.4.1) for qualitative and quantitative metabolic simulations and compared to 4F.

6.4.1 Standard simulation software

Evaluation of software for metabolic simulations by functional approaches was performed with ^{13}C Flux Explorer on Aurora Linux 8.1 (Petersen et al, 2000), OPPP Explorer on Excel 10/Microsoft Windows 2000 (Schwender et al, 2003) and the web-based software JWS Online (Brett et al, 2004). The software was tested using records from recycling processes of nutritional glucose from *Drosophila melanogaster* standard culture in controlled environment (cf. chapter 3.7.7). Standard simulation software was found suitable for the simulation of directed flows from a metabolic source compound to a target compound. The software was unable to simulate inherent circular flows without additional metabolic drain nodes.

6.4.2 Simulation of glucose metabolism

The 4F algorithm was described in detail in chapter 3.6. Table 6.2 summarizes the formal rules infered to reconstruct metabolic processes. The software package was implemented in C# (Microsoft Development Environment 7.1) and is available as .NET software service at http://xsystem.org.chemie.tu-muenchen.de. Experimental data can be submitted as XML document containing the single glucose isotopologs, the ^{13}C excess, the calibration value (obtained by IRMS or NMR), and control tags. The software consists of 4 major building blocks for (I). user management and XML submission, (II). parsing of the submitted XML document containing the experimental results, (III). the initialization of the metabolic network for the simulation, (IV). the inference process using the defined rules, and (V). the output of the simulation results. The installation of the software implementation requires Microsoft Windows XP, the .NET environment, and an Apache 2.0 web server.

6.4.3 Simulation of amino acid metabolism

The rule set for simulating amino acid metabolism can be derived from the rule set defined in Table 6.2 for glucose metabolism assuming glucose consists of two equivalent triose moieties and the same experimental setup using as tracer [U-$^{13}C_6$]glucose highly diluted with ^{13}C natural abundance glucose. Therefore, biosynthetic C_3 precursors in amino acid anabolism carry the full information for a simulation of the intermediary circular fluxes in the metabolic network of the experimental system. In detail, rules emp1 and emp2 (Table 6.2) can be summarized to a single rule generating a labeled $^{13}C_3$ unit from a proffered [U-$^{13}C_6$]glucose precursor, emp: $\{XXX\} \rightarrow \{XXX\}$. Asymmetries between C_3 units used in glucose to assess the influence of the transaldolase reaction in the pentose phosphate cycle cannot be exploited in available amino acid information. Thus, the pentose phosphate cycle must be characterized by the influence of the transketolase reaction by rule ppp: $\{XXX\} \rightarrow \{XX0\} + \{00X\}$. The rule tca for the citric acid cycle remains similar to rule tca defined for glucose: $\{XXX\} \rightarrow \{X00\} + \{0XX\}$.

Rule	Operation
emp	Cycling *via* glycolysis/glucogenesis from glucose 6-phosphate to triose phosphate and reverse. Triose phosphate units can also be generated by the transaldolase reaction. The first part of the rule 'emp' formalizes the cleavage and regeneration of glucose. The second part of the rule 'emp' formalizes the possible interchange of the triose moieties. emp1: $\{XXXXXX\} \rightarrow \{XXX000\} + \{000XXX\}$ emp2: $\{XXX000\} + \{000XXX\} \rightarrow \{XXX000\} + \{000XXX\}$
ppp	Cycling *via* the transketolase reaction of the pentose phosphate pathway. ppp1: $\{XXXXXX\} \rightarrow \{XX0000\} + \{00XXXX\}$
pex	Glyceraldehyde 3-phosphate and sedoheptulose 7-phosphate yield an asymmetric isotopolog distribution in catabolic processes. The resulting fructose 6-phosphate connects glycolysis and transaldolase reaction in the non-oxidative branch of the pentose phosphate pathway. pex1: $\{XXXXXX\} \rightarrow \{000XXX\}$
tsf	Transfer of exogenous glucose into leaf glucose without further metabolic cycling events. This rule is only valid if all X contain the value one. It implements an identity transformation. tsf1: $\{XXXXXX\} \rightarrow \{XXXXXX\}$
tca	Glucogenesis *via* oxaloacetate in the citric acid cycle. tca1: $\{XXXXXX\} \rightarrow \{X00000\} + \{0XX000\} + \{000XX0\} + \{00000X\}$
cof	$^{13}CO_2$ refixation yields labeled 3-phosphoglycerate through the Calvin cycle. The $^{13}CO_2$ can be generated from any position of the proffered ^{13}C-labelled glucose. ppp1: $\{XXXXXX\} \rightarrow \{00X000\} + \{000X00\}$

Table 6.2: Metabolic transfer rules used to simulate isotopolog compositions in experiments using [U-$^{13}C_6$]glucose as tracer analyte and glucose isotopologs as tracee. For a detailed description of the X group nomenclature, see chapter 3. All rules except cof are mandatory.

Rule tsf can be removed from the rule set, since amino acid isotopolog signatures carry no information about the direct incorporation of [U-$^{13}C_6$]glucose into the experimental system. The inference of the rules for simulating amino acid metabolism is performed by the 4F algorithm according to the described procedures in chapter 3.6.

6.5 Similarity in *n*-dimensional sets

Similarity analysis in *n*-dimensional sets, i. a. isotopolog spaces, requires two separate procedures: an algorithm to quantify the similarity between a single pair of sets, and an algorithm to cluster the resulting data. Within the scope of this thesis, RMSD (chapter 6.2.1), next neighborhood analysis (chapter 6.5.1) and Voronoi polygons (chapter 6.5.2), are regarded as suitable algorithms to create a similarity matrix between *n*-dimensional sets. UPGMA (rooted tree), NJ (unrooted tree) and neural networks can be applied as clustering algorithms using the obtained similarity coefficient.

6.5.1 Next neighborhood

The similarity *D* of *n*-dimensional sets $A = \{a_1,...,a_i,...a_n\}$ and $B = \{b_1,...,b_i,...b_n\}$ can be quantified by their RMSD (Appendix 6.2.1), if the magnitude of the sets is the same. For sets with a different number of elements, the correlation can be evaluated by the sum of deviations between the next neighbors $a_i \in A$ and $b_i \in B$ using the Euclidean distance function $d_E = \sum_{i=1}^{n} |a_i - b_i|$ or the Manhatten distance function $d_M = \sum_{i=1}^{n} |a_i - b_i|$.

6.5.2 Voronoi polygons

A Voronoi polygon for a point set *Q* in space *S*, partitions *S* in regions which enclose all points which are nearer to $q_i \in Q$ then to any other point $q_j \in Q$. It also represents the dual space of the Delauney triangulation: $V(a_i) = \{x : |q_i - x| \le |q_j - x| \forall j \ne i\}$. The Voronoi polygon *V* is calculated using Fortunes algorithm (Fortune, 1992), which is based on the construction of hypercones on each point in the hyperplane containing all points of *Q*. The root of the cones is placed in the points $q_j \in Q$. The hypercone extend with an angle of π/2 relative to the hyperplane. Proceeding to +∞ the cones will merge, where each pair of cones defines two merging points r_1, r_2. By defining a hyperplane orthogonal to the hyperplane containing *Q* and r_1 and r_2, an intersection line of the two orthogonal hyperplanes is defined. Viewing the intersection lines set from –∞, the Voronoi space of *Q* can be constructed (Guibas and Stolfi, 1985; Guibas et al., 1992). Given two sets *A* and *B*, the similarity can be defined as perimeter difference of the edge set of the Voronoi polygon. Edges with infinite length can be removed from the set.

6.5.3 UPGMA algorithm

The UPGMA cluster algorithm clusters sets by creating a rooted tree *T*. The algorithm requires a similarity matrix *D* which can be created by the algorithms described in 6.5.1 and 6.5.2. UPGMA consists of 4 iteration steps: (I). Calculate $D_{ij} = \min(D)$. (II) Create node $n = \{i, j\}$ and include *n* in *T*. (III) Matrix *D* is transformed according to $D \leftarrow (D - i \cap D - j) \cup \{n\}$. (IV). If $\dim(D) = 2$ terminate the algorithm, else recalculate *D* using an arbitrary distance function *d* according to $D_{nk} = \frac{d(i,n) + d(j,k)}{2}$ and continue with step (I).

6.5.4 Neighbor joining algorithm

The NJ algorithm clusters an *n*-dimensional sets by creating an unrooted tree *T* (Saitou and Nei, 1987). Like UPGMA, it requires a similarity matrix *D* and the algorithm consists of 4 iteration steps: (I). Calculate the divergence r_i of leaf *i* to all other leaf in the *n*-dimensional cluster set, $r_i = \sum_{k=1}^{n} D_{ik}$. (II). Find the pair *i* and *j* minimizing distance *A*, $A_{ij} = D_{ij} - \frac{r_i - r_j}{n-2}$. (III). Create new node *k*, which connects *i* and *j* in *T* with edge lengths s_{iu} and s_{ju} defined by $s_{iu} = \frac{D_{ij}}{2} + \frac{r_i - r_j}{2(n-2)}$ and $s_{ju} = D_{ij} - s_{iu}$. (IV). Recalculate D, $D \leftarrow (D - i \cap D - j) \cup \{n\}$. If $\dim(D) \leq 2$ terminate the algorithm, else continue with step (I).

6.5.5 Cluster correlation with neural networks

Cluster correlation with neural networks was performed using the NeuroShell Classifier 2.6 software package (WardSystems Inc., Frederick, Maryland), which implements a neural network on the basis of the QuickProp algorithm. QuickProp is a backpropagation-derived algorithm, optimizing the modification of the dendral weight vector *w* in the neural network using an arbitrary error function *E*, $\Delta_t w_{ij} = \left(\frac{\partial E(t)}{\partial w_{ij}} \right) \cdot \left(\frac{\partial E(t-1)}{\partial w_{ij}} - \frac{\partial E(t)}{\partial w_{ij}} \right)^{-1} \cdot \Delta_{t-1} w_{ij}$ by approximation of the partial differentials through a hyperparable (Specht, 1991).

6.6 Simulation of NMR spectra

NMR spectra were simulated using the software NMRSim 2.4.2 (Bruker Instruments, Karlsruhe, Germany). For the simulation the pulse-sequence and spin-system is provided. The software calculates a simulated FID, which must be processed by WinNMR 2.6.

7. References

1. Adams MD et al., *The genome sequence of Drosophila melanogaster*, Science, 2000, **287**, 2185-2195

2. Ainscough R et al., *Genome sequence of the nematode C. elegans: a platform for investigating biology*, Science, 1998, **282**, 2012-2018

3. Almaas E, Kovacs B, Vicsek T, Oltvai ZN, Barabasi AL, *Global organization of the metabolic fluxes in the bacterium Escherichia coli*, Nature, 2004, **427**, 839-843

4. Amos LA, Löwe J, *How Taxol stabilises microtubule structure*, Chem. Biol., 1999, **6**, 65-69

5. Arabidopsis Genome Initiative, *Analysis of the genome sequence of the flowering plant Arabidopsis thaliana*, Nature, 2000, **408**, 796-815

6. Ball S, Guan HP, James M, Myers A, Keeling P, Mouille G, Buleon A, Colonna P, Preiss J, *From glycogen to amylopectin: a model for the biogenesis of the plant starch granule*, Cell, 1996, **86**, 349-352

7. Ball S, Morell MK, *From bacterial glycogen to starch: understanding the biogenesis of the plant starch granule*, Annu. Rev. Plant Biol., 2003, **54**, 207-233

8. Barber J, Andersson B, *Revealing the blueprint of photosynthesis*, Nature, 1994, **370**, 31-34

9. Barnes SJ, Weitzman PD, *Organization of the citric acid cycle enzymes into a multienzyme cluster*, FEBS Letters, 1986, **201**, 267-270

10. Bender DA, *Amino acid metabolism*, 1985, Wiley, New York

11. Besumbes O, Sauret-Gueto S, Phillips MA, Imperial S, Rodriguez-Concepcion M, Boronat A, *Metabolic engineering of isoprenoid biosynthesis in Arabidopsis for the*

production of taxadiene, the first committed precursor of Taxol, Biotechnol. Bioeng., 2004, **88**, 168-172

12. Biemann K, *The mass spectra of isotopically labeled molecules*, 1962, in Mass Spectrometry: Organic Chemical Applications, McGraw-Hill, NY, 223-227

13. Bigi F, Casiraghi G, Casnati G, Sartori G, *Modifikation of the Nickl reaction. A general synthetic approach to 2-vinyl-2,3-dihydrobenzofurans*, Tetrahedron, 1983, **39**, 169-174

14. Biliaderis CG, *The structure and interactions of starch with food constituents*, Can. J. Physiol. Pharmacol., 1991, **69**, 60-78

15. Bingham ET, Grosse RW, Woodfield DR, Kidwell KK, *Complementary gene interactions in alfalfa are greater in tetraploids than diploids*, Crop Sci., 1994, **34**, 823-829

16. Birchler JA, Auger DL, Riddle NC, *In search of the molecular basis of heterosis*, Plant Cell, 2003, **15**, 2236-2239

17. Bogani P, Lio P, Intrieri MC, Buiatti M., *A physiological and molecular analysis of the genus Nicotiana*, Mol. Phylogenet. Evol., 1997, **7**, 62-70

18. Bohlmann F, Grenz M, *Neue Isopentenyl-Acetophenon-Derivate aus Helianthella uniflora*, Chem. Ber., 1970, **103**, 90–96

19. Bohlmann F, Zdero C, 1979, *Über dimere Terpenketone aus Tagetes gracilis*, Phytochem., 1979, **I8**, 341-343

20. Bortz J, Lienert, GA, Boehnke, K, Verteilungsfreie Methoden in der Biostatistik, 1990, Springer, Heidelberg

21. Bovey FA, *NMR data tables of organic compounds*, Wiley-Interscience, 1967

22. Boyer CD, Daniels RR, Shannon JC, *Starch granule (amyloplast) development in endosperm of several Zea mays L. genotypes affecting kernel polysaccharides*, Amer. J. Bot., 1977, **64**, 50-56

23. Boyer CD, Hannah LC, *Kernel mutants of corn*, in: Specialty Corns, 2001, CRC Press, Boca Raton, Florida

24. Bray D, *Molecular networks: the top-down view*, Science, 2003, **301**, 1864-1866

25. Brett G, Snoep O, Snoep L, *Web-based kinetic modelling using JWS Online*, Bioinformatics, 2004, **20**, 2143–2144

26. Brevard C, Granger P, *Handbook of high resolution mulitnuclear NMR*, John Wiley, 1981

27. Christiensen WI, *Milk sickness: a review of the literature*, Econ. Bot., 1965, **19**, 293-300

28. Clifton SW, Roe BA, *The beautifully simple but intriguingly complex world of small genomes*, Genome Res., 1998, **8**, 331-333

29. Crawford IP, *Evolution of a biosynthetic pathway: the tryptophan paradigma*, Ann. Rev. Microbiol., 1989, 43, 567-600

30. Dandekar T, Schuster S, Snel B, Huynen M, Bork P, *Pathway alignment: Application to the comparative analysis of glycolytic enzymes*, Biochem. J., 1999, **343**, 115-124

31. Darwin C, The origin of species, John Murray Press, 1859

32. DeBoer D, Gainza-Bernal ME, von Ooyen RD, Maes RAA, *The analysis of trenbolone and the human urinary metabolites of trenbolone acetate by gas chromatography/mass spectrometry and gas chromatography/tandem mass spectrometry*, Biological Mass Spectrometry, 1991, **20**, 459-466

33. Depre C, Rider MH, Hue L, *Mechanism of control of heart muscle glycolysis*, Eur. J. Biochm., 1998, **258**, 277-290

34. Doblin MS, Kurek I, Jacob-Wilk D, Delmer DP, *Cellulose biosynthesis in plants: from genes to rosettes*, Plant Cell Physiol., 2002, **43**, 1407-1420

35. Domingo GJ, Chauhan HJ, Lessard IA, Fuller C, Perham, RN, *Self-assembly and catalytic activity of the pyruvate dehydrogenase multienzyme complex from Bacilus stearothermophilus*, Eur. J. Biochem., **266**, 1136-1146

36. Drysch A, El Massaoudi M, Mack C, Takors R, de Graaf AA, Sahm H, *Production process monitoring by serial mapping of microbial carbon flux distributions using a novel Sensor Reactor approach: II--(13)C-labeling-based metabolic flux analysis and L-lysine production*, Metab Eng., 2003, **5**, 96-107

37. Duvick DN, *Heterosis: feeding people and protecting natural resources*, in: *Genetics and exploitation of heterosis in crops* (Ed.: Madison WI), 1999, American Society of Agronomy, 19-29

38. Eastmond PJ, Graham IA, *Re-examining the role of the glyoxylate cycle in oilseeds*, Trends Plant Sci., 2001, **6**, 72-8

39. Ebbert MA, Marlowe JL, Burkholder JJ, *Protozoan and intracellular fungal gut endosymbionts in Drosophila: prevalence and fitness effects of single and dual infections*, J. Invertebr. Pathol., 2003, **83**, 37-45

40. Eisenreich W, *NMR-Untersuchungen zur CO_2-Fixierung und zu Coenzymbiosynthesen in anaeroben Bakterien*, 1991, Thesis, Technical University of Munich

41. Eisenreich W, Bacher A, *Elucidation of biosynthetic pathways by retrodictive/predictive comparison of isotopomer patterns determined by NMR spectroscopy*, Genetic Engineering, 2000, **22**, 121-153

42. Eisenreich W, Bacher A, Arigoni D, Rohdich F, *Biosynthesis of isoprenoids via the non-mevalonate pathway*, Cell. Mol. Life Sci., 2004, **61**, 1401-1426

43. Eisenreich W, Menhard B, Hylands PJ, Zenk MH, Bacher A, *Studies on the biosynthesis of taxol: the taxane carbon skeleton is not of mevalonoid origin*, Proc. Natl. Acad. Sci. U S A, 1996, **93**, 6431-6436

44. Eisses JF, Lafoe D, Scott LA, Weil CF, *Novel, developmentally specific control of Ds transposition in maize*, Mol. Gen. Genet., 1997, **256**, 158-68

45. Ettenhuber C, *Longevity of Drosophila melanogaster via caloric restriction*, MSc Thesis in Molecular Biotechnology, Technical University of Munich, 2004

46. Elf J, Doncic A, Ehrenberg M, *Mesoscopic reaction-diffusion in intracellular signaling*, Proceedings of SPIE 5110, 2003, 114-124

47. Fellermeier M, Eisenreich W, Bacher A, Zenk MH, *Biosynthesis of cannabinoids. Incorporation experiments with (13)C-labeled glucoses*, Eur. J. Biochem., 2001, **268**, 1596-604

48. Fiaux J, Andersson CJY, Holmberg N, Bülow L, Kallio PT, Szyperski T, Bailey JE, Wüthrich K, *^{13}C NMR flux ratio analysis of Escherichia coli central carbon metabolism in microaerobic bioprocesses*, J. Am. Chem. Soc., 1999, **121**, 1407-1408

49. Flachmann R, Sauer M, Schopfer CR, Klebsattel M, Pfeiffer AM, Luck T, *Use of astaxanthin-containing plants or parts of plants of the genus Tagetes as feed additives for improvement of pigmentation of animals, especially aquatic animals*, 2004, PCT Int. Appl. CODEN: PIXXD2 WO 2004017749 A2 20040304 CAN 140

50. Fleming PE, Knaggs AR, Xian-Guo H, Mocek U, Floss HG, *Biosynthesis of Taxoids. Mode of Attachment of the Taxol Side Chain*, J. Am. Chem. Soc., 1994, **116**, 4137-4138

51. Fleming PE, Mocek U, Floss HG, *Biosynthesis of Taxoids. Mode of Formation of the Taxol Side Chain*, J. Am. Chem. Soc., 1993, **115**, 805-807

52. Fraser CM et al., *The minimal gene complement of Mycoplasma genitalium*, Science, 1995, **270**, 397-403

53. Friedmann DI, Court DL, *Transcription antitermination, the λ paradigm updated*, Mol. Microbiol., 1995, **18**, 191-200

54. Fujiyama A et al., *Construction and analysis of a human-chimpanzee comparative clone map,* Science, 2002, **295**, 131-134

55. Follstadt BD, Stephanopoulos G, *Effect of reversible reaction on isotope label redistribution. Analysis of the pentose phosphate pathway*, Eur. J. Biochem., 1998, **202**, 360-371

56. Fortune S, *Voronoi Diagrams and Delauney Triangulations*, Euclidean Geometry and Computers, 1992, 193-233

57. Gabb HA, Jackson RM, Sternberg MJE, *Modelling protein docking using shape complementarity, electrostatics, and biochemical information*, J. Mol. Biol., 1997, **272**, 106-120

58. Garg SN, Verma SK, Gupta MM, *An improved process for the extraction of xanthophyll from marigold flowers (Tagetes erecta) for use as a coloring agent for the food and drug industries*, 2000, CODEN: INXXAP IN 184978 A 20001014 CAN 141

59. Gasteiger E, Jung E, Bairoch A, *Swiss-Prot: connecting biological knowledge via a protein database*, Curr. Issues Mol. Biol., 2001, **3**, 47-55

60. Gibbs RA et al., *Genome sequence of the Brown Norway rat yields insights into mammalian evolution*, Nature, 2004, **428**, 493-521

61. Gibson, MA, Bruck J, *Efficient exact stochastic simulation of chemical systems with many species and many channels*, J. Phys. Chem., 2000, **104**, 1876-1889

62. Gilbert EN, *Coding with digits of unequal costs*, Journal of the Association for Computing Machinery, IEEE Trans. Inform. Theory, 1995, **41**, 596-600

63. Gillespie DT, *Exact stochastic simulation of coupled chemical reactions*, J. Phys. Chem., 1977, **81**, 2340-2361

64. Gillespie DT, *Approximate accelerated stochastic simulation of chemically reacting systems*, J. Chem. Phys., 2001, **115**, 1716-1733

65. Gillespie DT, Petzold LR, *Improved leap-size selection for accelerated stochastic simulation*, J. Chem. Phys., 2004, **119**, 8229-8234

66. Giroux MJ, Hannah LC, *ADP-glucose pyrophosphorylase in shrunken-2 and brittle-2 mutants of maize*, Mol. Gen. Genet., 1994, **243**, 400-408

67. Glawischnig E, Gierl A, Tomas A, Bacher A, Eisenreich W, *Retrobiosynthetic nuclear magnetic resonance analysis of amino acid biosynthesis and intermediary metabolism. Metabolic flux in developing maize kernels*, Plant Physiol., 2001, **125** 1178-86

68. Goff SA et al., *A draft sequence of the rice genome (Oryza sativa L. ssp. japonica),* Science, 2002, **296**, 92-100

69. Goff SA, Salmeron JM, *Back to the future of cereals. Genomic studies of the world's major grain crops, together with a technology called marker-assisted breeding, could yield a new green revolution*, Sci. Am., 2004, **291**, 42-49

70. Granner D, Pilkis S, *The genes of hepatic glucose metabolism*, J. Biol. Chem., 1990, **256**, 10173-10176

71. Greenblatt J, Nodwell JR, Mason SW, *Transcriptional antitermination*, Nature, 1993, **364**, 401-406

72. Greene TW, Hannah LC., *Enhanced stability of maize endosperm ADP-glucose pyrophosphorylase is gained through mutants that alter subunit interactions*, Proc. Natl. Acad. Sci. USA, 1998, **95**, 13342-13347

73. Grivet JP, Delort AM, Portais JC, *NMR and microbiology: from physiology to metabolomics*, Biochimie, 2003, **85**, 823-840

74. Guibas LJ, Knuth DE, Sharir M, *Randomized Incremental Construction of Delaunay and Voronoi diagrams*, Algorithmica, 1992, **7**, 381-413

75. Guibas LJ, Stolfi J, *Ruler, compass and computer: the design and analysis of geometric algorithms*, Theoretical Foundations of Computer Graphics an CAD, 1985, **40**, 111-165

76. Gupta SK, *Skin firming anti-aging cosmetic compositions*, 2004, U.S. Pat. Appl. Publ. CODEN: USXXCO US 20040161435 A1 20040819

77. Hansson B, Westerberg L, *On the correlation between heterozygosity and fitness in natural populations*, Mol. Ecol., 2002, **11**, 2467-2474

78. Harcombe W, Hoffmann AA, *Wolbachia effects in Drosophila melanogaster: in search of fitness benefits*, J. Invertebr. Pathol., 2004, **87**, 45-50

79. Hazen SP, Hawley RM, Davis GL, Henrissat B, Walton JD, *Quantitative trait loci and comparative genomics of cereal cell wall composition*, Plant Physiol., 2003, **132**, 263-271

80. Hemmerlin A, Hoeffler JF, Meyer O, Tritsch D, Kagan IA, Grosdemange-Billiard C, Rohmer M, Bach TJ, *Cross-talk between the cytosolic mevalonate and the plastidial methylerythritol phosphate pathways in tobacco bright yellow-2 cells*, J. Biol. Chem., 2003, **278**, 26666-26676

81. Herre EA, *Population structure and the evolution of virulence in nematode parasites of fig wasps*, Science, 1993, **259**, 1442-1445

82. Hills MJ, *Control of storage-product synthesis in seeds*, Curr. Opin. Plant Biol., 2004, **7**, 302-308

83. Hillier LW et al., *Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution*, Nature, 2004, **432**, 695-716

84. Holland JH, *Adaptation in natural and artificial systems*, 1975, University of Michigan Press, Ann. Arbor.

85. Huffman DA, *A method for the construction of minimum redundancy codes*, Proc. IRE 40, 1952, **10**, 1098-1101

86. Illarionov B, Fischer M, Lee CY, Bacher A, Eisenreich W, *Rapid preparation of isotopolog libraries by in vivo transformation of 13C-glucose. Studies on 6,7-dimethyl-8-ribityllumazine, a biosynthetic precursor of vitamin B2*, J. Org. Chem., 2004, **69**, 5588-5594

87. Ingle JD, Hageman RH, *Changes in composition during development and maturation of maize seeds*, Plant Physiol., 1965, **40**, 832-835

88. International Human Genome Sequencing Consortium, *Finishing the euchromatic sequence of the human genome*, Nature, 2004, **431**, 931-945

89. IUPAC, *Compendium of Chemical Terminology*, 1997, 2nd Edition, http://www. iupac.org

90. Izard T, Evarsson A, Allan MD, Westphal AH, Perham RN, DeKok A, Hol WG, *Principles of quasi-equivalence and Euclidean geometry govern the assembly of cubic and dodecahedral cores of pyruvate dehydrogenase complex*, Proc. Natl. Acad. Sci. USA, 1999, **96**, 1240-1245

91. James MG, Denyer K, Myers AM, *Starch synthesis in the cereal endosperm*, Curr. Opin. Plant Biol., 2003, **6**, 215-222

92. James MG, Robertson DS, Myers AM, *Characterization of the maize gene sugary1, a determinant of starch composition in kernels*, *Plant Cell*, 1995, **4**, 417-429

93. Jitrapakdee S, Wallace JC, *Structure, function and regulation of the pyruvate carboxylase*, Biochem. J., 1999, **340**, 1-16

94. Johnson L, Harrison JH, Hunt C, Shinners K, Doggett CG, Sapienza D, *Nutritive value of corn silage as affected by maturity and mechanical processing: a contemporary review*, J. Dairy Sci., 1999, **82**, 2813-2825

95. Jones DF, *Dominance of linked factors as means of accounting for heterosis*, Genetics, 1917, **2**, 466-479

96. Kastner J, Solomon J, Fraser S, *Modeling a hox gene network in silico using a stochastic simulation algorithm*, Developmental Biology, 2002, **246**, 122-131

97. Katoh A, Hashimoto T, *Molecular biology of pyridine nucleotide and nicotine biosyn thesis*, Front. Biosci., 2004, **9**, 1577-1586

98. Katz J, Tayek JA, *Recycling of glucose and determination of the Cori Cycle and gluconeogenesis*, Am. J. Physiol., 1999, **277**, 401-407

99. Kelleher JK, *Estimating gluconeogenesis with [U-13C]glucose: molecular condensation requires a molecular approach*, Am. J. Physiol., 1999, **277**, 395-400

100. Kikuchi J, Shinozaki K, Hirayama T, *Stable isotope labeling of Arabidopsis thaliana for an NMR-based metabolomics approach*, Plant Cell Physiol., 2004, **45**, 1099-1104

101. Kis K, Volk R, Bacher A, *Biosynthesis of riboflavin. Studies on the reaction mechanism of 6,7-dimethyl-8-ribityllumazine synthase*, Biochemistry, 1995, **34**, 2883-2892

102. Koch KE, Nolte KD, Duke ER, McCarty DR, Avigne WT, *Sugar Levels Modulate Differential Expression of Maize Sucrose Synthase Genes*, Plant Cell., 1992, **4**, 59-69

103. Lander ES et al., *Initial sequencing and analysis of the human genome,* Nature, 2001, **409**, 860-921

104. Lee M, Sharopova N, Beavis WD, Grant D, Katt M, Blair D, Hallauer A, *Expanding the genetic map of maize with the intermated B73 x Mo17 (IBM) population*, Plant Mol. Biol., 2002, **48**, 453-461

105. Lee JM, Sonnhammer EL, *Genomic gene clustering analysis of pathways in eukaryotes*, Genome Res., 2003, **13**, 875-882

106. Lewis EB, *A gene complex controlling segmentation in Drosophila*, Nature, 1978, **276**, 565-570

107. Laule O, Furholz A, Chang HS, Zhu T, Wang X, Heifetz PB, Gruissem W, Lange M., *Crosstalk between cytosolic and plastidial pathways of isoprenoid biosynthesis in Arabidopsis thaliana*, Proc Natl Acad Sci U S A, 2003, **100**, 6866-71

108. Lin TJ, Heinstein P, *In vitro biosynthesis of isopentenylacetophenones in Eupatorium rugosum*, Phytochem., 1974, **13**, 1817–1823

109. Lin X, Hezari M, Koepp AE, Floss HG, Croteau R, *Mechanism of the Taxadiene Synthase, a Diterpene Cyclase that Catalyzes the First Step of the Taxol Biosynthesis in Pacific Yew*, Biochemistry, 1996, **35**, 2968-2977

110. Lin TJ, Ramstad E, Heinstein P, *In vivo biosynthesis of isopentenylacetophenones in Eupatorium rugosum*, Phytochem., 1974, **13**, 1809–1815

111. Lopaschuk GD, *Advantages and limitations of experimental techniques used to measure cardiac energy metabolism*, J. Nucl. Cardiol., 1997, **4**, 316-328

112. Mahmud T, Bode HB, Silakowski B, Kroppenstedt RM, Xu M, Nordhoff S, Hofle G, Muller R, *A novel biosynthetic pathway providing precursors for fatty acid biosynthesis and secondary metabolite formation in myxobacteria*, J Biol Chem., 2002, **277**, 32768-32774

113. Manandhar A, Gresshoff PM, *Blue spruce (Picea pungens) tissue and cell culture*, Cytobios., 1980, **29**, 175-182

114. Martin W, Schnarrenberger C, *The evolution of the Calvin cycle from prokaryotic to eucaryotic chromosomes: a case study of functional redundancy in ancient pathways through endosymbiosis*, Curr. Genet., 1997, **32**, 1-18

115. Melendez-Hevia E, Wadell TG, Cascante M, *The puzzle of the Krebs citric acid cycle: assembling the pieces of chemicall feasible reactions, and opportunism in the design of metabolic pathways in evolution*, J. Mol. Evol., 1996, **43**, 293-303

116. Menelaou MA, Fronzek FR, Hjortso MA, Morrison AF, Foroozesh M, Thibodeaux TM, Flores HE, Fischer N, *H-NMR spectral data of benzofurans and bitiophenes from hairy root cultures of Tagetes patula and the molecular structure of isoeuparin*, Spectroscopy Lett., 1993, **24**, 1405–1413

117. Meng TC, Sandeep S, Pawan D, *Modeling and simulation of biological systems with stochasticity*, In Silico Biol., 2004, **4**, 24-42

118. Merzendorfer H, Zimoch L, *Chitin metabolism in insects: structure, function and regulation of chitin synthases and chitinases*, J. Exp. Biol., 2003, **206**, 4393-412

119. Michalet-Doreau B, Fernandez I, Fonty G, *A comparison of enzymatic and molecular approaches to characterize the cellulolytic microbial ecosystems of the rumen and the cecum*, J. Anim. Sci., 2002, **80**, 790-796

120. Miller SL, *Which organic compounds could have occurred on prebiotic earth?*, Cold Spring Harbor Sym. Quant. Bio., 1987, **52**, 17-27

121. Monir H, Proksch P, *Yeast induced accumulation of benzofurans in cell suspension cultures of Ageratina adenophora*, Phytochem., 1989, **28**, 2999–3002

122. Morton-Firth CJ, *Stochastic Simulation of Cell signaling Pathways*, PhD thesis, 1998, University of Cambridge, Cambridge, UK

123. Morton-Firth CJ, Bray D, *Predicting temporal fluctuations in an intracellular signaling pathway*, J. Theor. Biol., 1998, **192**, 117-128

124. Moose SP, Dudley JW, Rocheford TR, *Maize selection passes the century mark: a unique resource for 21st century genomics*, Trends Plant Sci., 2004, **9**, 358-364

125. Myers AM, Morell MK, James MG, Ball SG, *Recent progress toward understanding biosynthesis of the amylopectin crystal*, Plant Physiol., 2000, **122**, 989-9973

126. Nakamura Y, Kubo A, Shimamune T, Matsuda T, Harada K, Satoh H, *Correlation between activities of starch debranching enzyme and polyglucan structure in endosperms of sugary-1 mutants of rice*, Plant J., 1997, **12**, 143-153

127. Neuffer G, Coe EH, Wessler R, *Mutants of maize*, Cold Spring Harbor Laboratory Press, 1997

128. Nier, AO, *A mass spectrometer for isotope and gas analyses.* Rev. Sci. Instr., 1947, **18**, 398

129. Nishi A, Nakamura Y, Tanaka N, Satoh H, *Biochemical and genetic analysis of the effects of amylose-extender mutation in rice endosperm*, Plant Physiol., 2001, **127**, 459-472

130. Notaro R, Afolyan A, Luzzatto L, *Human mutations in glucose-6-phosphate dehydrogenase reflect evolutionary history*, FASEB J., 2000, **14**, 485-494

131. Nüsslein-Volhard C, Wieschaus E, *Mutations Affecting Segment Number and Polarity in Drosophila*, Nature, 1980*,* **287**, 795-801

132. Okagaki RJ, Neuffer MG, Wessler SR, *A deletion common to two independently derived waxy mutations of maize*, Genetics, 1991, **128**, 425-431

133. Paul EA, Frey B, *Carbon isotope techniques*, Elsevier, 1991

134. Parodi FJ, Fischer NH, Flores HF, *Benzofuran and bitiophenes from root culture of Tagetes patula*, J. Nat. Prod., 1988, **51**, 594–595

135. Paterson AH, Bowers JE, Peterson DG, Estill JC, Chapman BA, *Structure and evolution of cereal genomes*, Curr. Opin. Genet. Dev., 2003, **13**, 644-50

136. Park SM, Klapa MI, Sinskey AJ, Stephanopoulos G, *Metabolite and isotopomer balancing in the analysis of metabolic cycles: II. Applications*, Biotechnol. Bioeng., 1999, **62**, 392-401

137. Peters J, Einführung in die allgemeine Informationstheorie, Springer, Heidelberg, 1967

138. Petersen S, de Graaf AA, Eggeling L, Mollney M, Wiechert W, Sahm H, *In vivo quantification of parallel and bidirectional fluxes in the anaplerosis of Corynebacterium glutamicum*, J Biol Chem., 2000, **275**, 35932-35941

139. Pilkis SJ, Granner DK, *Molecular physiology of the regulation of hepatic gluconeogenesis and glycolysis*, Ann. Rev. Physiol., 1992, **54**, 885-909

140. Plaut GWE, Harvey RA, *The enzymatic synthesis of riboflavin*, Methods Enzymol., 1971, **18**, 515-539

141. Powel-Douglass B, *Real Time UML*, 2004, 3th Edition, Addison-Wesley

142. Press WH, Teukolsky SA, Vetterlin WT, Flannery BP, *Numerical Recipes in C*, 1992, Cambridge University Press

143. Previs SF, Brunengraber H, *Methods for measuring gluconeogenesis in vivo*, Curr. Opin. Clin. Nutr. Metab. Care, 1998, **1**, 461-465

144. Raizada MN, *RescueMu protocols for maize functional genomics*, Methods Mol. Biol., 2003, **236**, 37-58

145. Rechenberg I, *Optimierung technischer Systeme nach Prinzipien der biologischen Evolution*, 1973, frommann-holzboog, Stuttgart

146. Rees WA, Weitzel SE, Yager TD, Das A, von Hippel PH, *Bacteriophage λ N protein alone can induce transcription antitermination in vitro*, Proc. Natl. .Acad. Sci. USA, 1993, **93**, 342-346

147. Rindlav-Westling A, Stading M, Gatenholm P, *Crystallinity and morphology in films of starch, amylose and amylopectin blends*, Biomacromolecules, 2002, **3**, 84-91

148. Römisch W, Eisenreich W, Richter G, Bacher A, *Rapid one-pot synthesis of riboflavin isotopomers*, J. Org. Chem., 2002, **67**, 8890-8894

149. Roessner-Tunali U, Liu J, Leisse A, Balbo I, Perez-Melis A, Willmitzer L, Fernie AR, *Kinetics of labelling of organic and amino acids in potato tubers by gas chromatography-mass spectrometry following incubation in (13)C labelled isotopes*, Plant J., 2004, **39**, 668-679

150. Rodriguez-Concepcion M, Fores O, Martinez-Garcia JF, Gonzalez V, Phillips MA, Ferrer A, Boronat A, *Distinct light-mediated pathways regulate the biosynthesis and exchange of isoprenoid precursors during Arabidopsis seedling development*, Plant Cell., 2004, **16**, 144-156

151. Roscher A, Kruger NJ, Ratcliffe RG, *Strategies for metabolic flux analysis in plants using isotope labelling*, J Biotechnol., 2000, **77**, 81-102

152. Rossmann A, Butzenlechner M, Schmidt HL, *Evidence for a nonstatistical carbon isotope distribution in natural glucose*, Plant Physiol., 1991, **96**, 609-614

153. Ryang BS, Kobori T, Matsumoto T, Kosaka Y, Ohki ST, *Cucumber mosaic virus 2b protein compensates for restricted systemic spread of Potato virus Y in doubly infected tobacco*, J. Gen. Virol., 2004, **85**, 3405-3414

154. Saito T, Hyamizu K, Yanagisawa M, Yamamoto O, *SDBS – Spectral database for organic compounds*, www.aist.go.jp/RIODB/SDBS, 2004

155. Saitou N, Nei M, *The neighbor-joining method : a new method for reconstructing phylogenetic trees*, Molecular Biology and Evolution, 1987, **4**, 406-425

156. Sasaki T et al., *The genome sequence and structure of rice chromosome 1*, Nature, 2002, **420**, 312-316

157. Scaraffia PY, Wells MA, *Proline can be utilized as an energy substrate during the flight of Aedes aegypti females, Journal of Insect Physiology*, 2003, **49**, 591-601

158. Sharma OP, Dawra RJ, Kurade NP, Sharma PD, *A rewiev of the toxicosis and biological properties of the genus Eupatorium*, Nat. Toxins, 1998, **6**, 1-14

159. Shi QW, Sauriol F, Mamer O, Zamir LO, *New minor taxanes analogues from the needles of Taxus canadenis*, Bioorg. Med. Chem., 2003, **11**, 293-303

160. Shull GH, *The composition of a field of maize*, Am. Breeders Assoc. Rep., 1908, **4**, 296-301

161. Sieberzt R, Proksch P, Wray V, Witte L, *Accumulation and biosynthesis of benzofurans in root cultures of Eupatorium cannabinum*, Phytochem., 1989, **28**, 789–793

162. Silverman M, S*tructure and function of hexose transporters*, Ann. Rev. Biochem., 1991, **60**, 757-794

163. Schauder P, Wahren J, Paoletti R, Bernardi R, Rinetti M, *Branched-chain amino acids: biochemistry, physiopathology, and clinical sciences*, 1992, Raven Press, New York

164. Schmidt HL, *Fundamentals and systematics of the non-statistical distributions of isotopes in natural compounds*, Naturwissenschaften, 2003, **90**, 537-552

165. Schmidt K, Carlsen M, Nielsen J, Villadsen J, *Modeling isotopomer distributions in biochemical networks using isotopomer mapping matrices*, Biotechnol. Bioeng., 1977, **55**, 831-840

166. Schmidt K, Marx A, deGraaf AA, Wiechert W, Sahm H, Nielsen J, Villadsen J, *^{13}C Tracer experiments and metabolite balancing for metabolic flux analysis: comparing two approaches*, Biotechnol. Bioeng., 1998, **58**, 254-257

167. Schoell M, Faber E, Coleman ML, *Carbon and hydrogen isotope composition of the NBS 22 and NBS 21 stable isotope reference materials: An inter-laboratory comparison,* Org. Geochem., 1983, **5**, 3-6

168. Schwehm M, *Parallel stochastic simulation of whole-cell models*., Proc. 2nd Int. Conf Systems Biology, 1996, Omnipress, 333-341

169. Schwender J, Seemann M, Lichtenthaler HK, Rohmer M, *Biosynthesis of isoprenoids (carotenoids, sterols, prenyl side-chains of chlorophylls and plastoquinone) via a novel pyruvate/glyceraldehyde 3-phosphate non-mevalonate pathway in the green alga Scenedesmus obliquus*, Biochem J., 1996, **316**, 73-80

170. Schwender J, Ohlrogge JB, Shachar-Hill Y, *A Flux Model of Glycolysis and the Oxidative Pentosephosphate Pathway in Developing Brassica napus Embryos*, J. Biol. Chem., 2003, **278**, 29442-29453

171. Scott AI, *Applications of 13C NMR in the study of biosynthetic mechanism*, J. Nat. Prod., 1985, **48**, 689-707

172. Scott RJ, Spielman M, *Epigenetics: imprinting in plants and mammals - the same but different?*, Curr. Biol., 2004, **14**, 201-203

173. Shannon CE, *A Mathematical Theory of Communication*, The Bell System Technical Journal, **27**, 1948, 379-423

174. Shimizu, TS, *The spatial organisation of cell signaling pathways - a computer based study*, PhD thesis, 2002, University of Cambridge, Cambridge, UK.

175. Singeltary GW, Banisadr R, Keeling PL, Influence of gene dosage on carbohydrate synthesis and enzymatic activities in endosperm of starch-deficient mutants of maize, Plant Physiol., 2004, **113**, 293-304

176. Specht DF, *A general regression neural network*, IEEE Transactions on Neural Networks, 1991, **2**, 568-576

177. Sriram G, Fulton DB, Iyer VV, Peterson JM, Zhou R, Westgate ME, Spalding MH, Shanks JV, *Quantification of Compartmented Metabolic Fluxes in Developing Soybean Embryos by Employing Biosynthetically Directed Fractional 13C Labeling, Two-Dimensional [13C, 1H] Nuclear Magnetic Resonance, and Comprehensive Isotopomer Balancing*, Plant Physiol., 2004, electronic publication ahead of print

178. Stundzia AB, Lumsden CJ, *Stochastic simulation of coupled reaction-diffusion processes*, J. Comput. Phys., 1996, **127**, 196-207

179. Sütfeld R, Balza F, Towers GHN, *A benzofuran from Tagetes patula seedlings*, Phytochem., 1985, **24**, 876–877

180. Sullivan TD, Kaneko Y, *The maize brittle 1 gene encodes amyloplast membrane polypeptides*, Planta, 1995, **196**, 477-484

181. Tang CS, Wat CK, Towers GHN, *Thiophenes and benzofurans in the undisturbed rhisosphere of Tagetes patula L*, Plant and Soil, 1987, **98**, 93–98

182. Tekoah Y, Ko K, Koprowski H, Harvey DJ, Wormald MR, Dwek RA, Rudd PM, *Controlled glycosylation of therapeutic antibodies in plants*, Arch. Biochem. Biophys., 2004, **426**, 266-278

183. Thorell S, Gergely P, Banki K, Perl A, Schneider G, *The three-dimensional structure of human transaldolase*, FEBS Letters, 2000, **475**, 205-208

184. Urey HC, Epstein S, McKinney C, McCrea J, M*ethod for measurement of paleotemperatures*, Bulletin of the Geological Society of America, 1948, 59, 1359-1360.

185. van Winden W, Heijnen JJ, Verheijen PJT, *Cumulative bondomers: a new concept in flux analysis from 2D ^{13}C, ^{1}H COSY NMR data*, Biotechn. Bioeng., 2002, **80**, 731-745

186. Velot C, Mixon MB, Teige M, Srere PA, *Model of a quniary structure between Krebs TCA cycle enzymes: a model for the metabolon*, Biochemistry, 1997, **36**, 14271-14276

187. Venter JC et al., *The sequence of the human genome*, Science, 2001, **291**, 1304-1351

188. Vinkenoog R, Bushell C, Spielman M, Adams S, Dickinson HG, Scott RJ, *Genomic imprinting and endosperm development in flowering plants*, Mol. Biotechnol., 2003, **25**, 149-184

189. Volk R, Bacher A, *Studies on the 4-carbon precursor in the biosynthesis of riboflavin. Purification and properties of L-3,4-dihydroxy-2-butanone-4-phosphate synthase*, J. Biol. Chem., 1990, **265**, 19479-19485

190. Walkey DG, Woolfitt JM, *Clonal multiplication of Nicotiana rustica L. from shoot meristems in culture*, Nature, 1968, **220**, 1346-1347

191. Waterston RH, *Initial sequencing and comparative analysis of the mouse genome*, Nature, 2002, **420**, 520-562

192. Wiechert W, de Graaf AA, *Bidirectional reaction steps in metabolic networks: I. Modeling and simulation of carbon isotope labeling experiments*, Biotechnol. Bioeng., 1997, **55**, 101-117

193. Wiechert W, Möllney M, Isermann N, Murzel M, de Graaf AA, *Bidirectional reaction steps in metabolic networks: III. Explicit solution and analysis of isotopomer labeling systems.* Biotechnol. Bioeng., 1999, **66**, 69-85

194. Wiechert W, Murzel M, *Metabolic isotopomer labeling systems. Part I: global dynamic behaviour*, Math. Biosci., 2001, **169**, 173-205

195. Wiechert W, Siefke C, de Graaf AA, Marx A, *Bidirectional reaction steps in metabolic networks: II. Flux estimation and statistical analysis*, Biotechnol. Bioeng., 1997, **55**, 118-135

196. Wiener N, *Cybernetics*, 1948, MIT Press, Cambridge

197. Xu Q, Morrison M, Nelson KE, Bayer EA, Atamna N, Lamed R, *A novel family of carbohydrate-binding modules identified with Ruminococcus albus proteins*, FEBS Lett., 2004, **566**, 11-16

198. Yamada K et al., *Empirical analysis of transcriptional activity in the Arabidopsis genome*, Science, 2003, **302**, 842-846

199. Yourdon E, *Modern Structured Systems Analysis*, 1989, Englewood Cliffs, NJ, Prentice Hall

200. Zeeman SC, Smith SM, Smith AM, The priming of amylose synthesis in Arabidopsis leaves, Plant Physiol., 2002, **128**, 1069-1076

201. Zeeman SC, Tiessen A, Pilling E, Kato KL, Donald AM, Smith AM, *Starch synthesis in Arabidopsis. Granule synthesis, composition, and structure*, Plant Physiol., 2002, **129**, 516-529

202. Zeeman SC, Umemoto T, Lue WL, Au-Yeung P, Martin C, Smith AM, Chen J, *A mutant of Arabidopsis lacking a chloroplastic isoamylase accumulates both starch and phytoglycogen*, Plant Cell, 1998, **10**, 1699-1712

Appendix A. NMR assignments

A.1 Isoeuparin

Pos.	Chemical shift (ppm)		Coupling Constants (Hz)			Coupling Patterns			
	δ^{13}C	δ^{1}H	J_{CC}	J_{HH}	J_{CH}	INADE	COSY	NOESY	HMBC
10	204.08		42.9 (11)						6, 11
8	159.51		69.0 (7), 58.6 (9), 5.5 (4), 2.0 (6)						7, 3
4	158.53		70.6 (9), 63.2 (5), 5.5 (8), 2.6 (7)						6, 7, 3
2	156.75		74.3 (3), 6.4 (14)						3, 13, 13’, 14
12	132.38		71.9 (13), 43.6 (14)			13			14, 13’, 3
6	127.37	7.62	60.8 (5,7) 2.0 (8)	8.8 (7)	159.2 (6), 12.3 (8)	7, 5	7	7, 11	11
13	113.46	5.74	71.9 (12)		159.4 (13)	12	13, 14	13	14, 3
7	103.40	6.97	69.2 (8), 61.5 (6), 6.2 (9) 1.8 (4)	9.0 (6)	167.7 (7)	6	6	6	
3	100.86	6.82	74.1 (2)		178.8 (3), 9.5 (2)			14	13’
11	26.87	2.63	43.1 (10)		128.0 (11), 5.8 (10)			6	
14	19.24	2.10	43.6 (12), 6.2 (2), 1.1 (3)		127.6 (14)			3, 13	13, 13’
9	114.36		65.2 (4), 60.6 (8)						
5	119.01		60.6 (4,6)						

A.2 4-hydroxytremetone

Pos.	**Chemical shift (ppm)**		**Coupling Constants (Hz)**			**Coupling Patterns**			
	δ^{13}C	**δ^{1}H**	**J_{CC}**	**J_{HH}**	**J_{CH}**	**INADE**	**COSY**	**NOESY**	**HMBC**
10	202.65		43.1 (11)						6
8	166.85		68.3 (7), 60.8 (9), 4.0 (4), 3..9 (6)						3, 3’, 6, 7
4	160.32		69.6 (9), 61.0 (5), 5.1 (8), 4.8 (7)						3, 3’, 6
2	89.97	5.29	33.8 (3), 6.1 (14)	8.0 (3), 9.6 (3’)	151.6 (2)	3	3, 3’	13’	14, 3, 3’, 13, 13’
12	143.27		73.0 (13), 43.1 (14)			13			14, 3, 3’, 2,
6	133.25	7.58	61.1 (5), 61.1 (7), 1.8 (8), 3.1 (9),	8.5 (7)	158.1 (6)	7, 5	7	11, 7	11
13		5.06			159.4 (13’)		13, 14	2, 13	
7	101.81	6.38	69.2 (8), 61.5 (6), 6.2 (9), 3.4 (4)	8.5 (6)	166.6 (7)	6	6	6	3, 3’
3	30.95	2.98	33.6 (2)	15.6 (3’), 9.9 (2)	134.3 (3)	2	2, 3’	14	2
11		3.32		15.6 (3), 7.7 (2)	136.0 (3’)		2, 3		
14	26.30	2.53	42.9 (10)		127.8 (11), 5.8 (10)			6	
9	17.00	1.74	43.1 (12), 6.2 (2), 1.1 (3)		127.0 (14)		13, 13’	13, 3	2, 13, 13’
5	113.46		61.0 (4), 60.6 (8)						3, 3’, 7

A.3 Glucose

A.3.1 General assignment

Position	Chemical shift (ppm)	^{13}C^{13}C coupling (Hz)	^{13}C isotope shift (ppb)
1α	93.28	46.3 (2α), 3.0 (6α), 2.0 (5α)	8 (2α)
2α	72.79	45.6 (1α), 38.2 (3α), 2.7 (4α)	11 (1α), 29 (3α), 40 (1α, 3α)
3α	74.05	38.2 (2α), 38.7 (4α), 3.5 (6α), 1.4 (5α)	-7 (2α), -3 (2α, 4α)
4α	70.90	40.9 (5α), 39.2 (3α), 3.0 (2α)	22 (5α), 42 (3α, 5α)
5α	72.55	43.4 (6α), 40.3 (4α), 2.0 (1α), 1.7 (3α)	6 (6α)
6α	61.85	43.4 (5α), 3.5 (1α), 3.6 (3α)	11 (6α)
1β	97.20	46.1 (2β), 4.2 (3β), 4.1 (6β)	8 (2β)
2β	75.41	45.8 (1β), 38.9 (3β), 2.3 (4β)	12 (1β), 22 (3β), 34 (1β, 3β)
3β	77.10	39.4 (4β), 38.9 (2β), 2.3 (5β), 4.0 (6β), 4.2 (1β)	-8 (2β)
4β	70.83	40.5 (5β), 38.7 (3β), 3.5 (2β)	13 (5β), 30 (3β, 5β)
5β	77.19	43.2 (6β), 40.8 (4β), 2.3 (3β)	6 (6β), 5 (4β)
6β	61.97	43.2 (5β), 4.1 (1β), 4.2 (3β)	12 (5β)

A.3.2 X group translation

Signal	α-glucose		β-glucose	
	Chemical shift (ppm)	X group	Chemical shift (ppm)	X group
1	93.28	10XXXX	97.20	10XXXX
	93.458, 93.094	11XXXX	97.376, 97.012	11XXXX
	93.473, 93.458, 93.447, 93.429, 93.109, 93.092, 93.067	11XX11	97.376, 97.012	110XX0
	93.458, 93.092	11XX00	97.409, 97.376, 97.337, 97.045, 97.010, 96.975	111XX1
			97.395, 97.356, 97.031, 96.992	111XX0, 110XX1
2	72.79	010XXX	75.410	010XXX
	72.966, 72.598	110XXX	75.582, 75.217	110XXX
	72.914, 72.610	011XXX		1101XX
	73.087, 72.790, 72.725, 72.417	111XXX	75.582, 75.217	1100XX
		1110XX	75.540, 75.234	011XXX
	73.097, 73.068, 72.791, 72.766, 72.734, 72.700, 72.431, 72.401	1111XX	75.540, 75.234	0110XX
			75.712, 75.410, 75.348, 75.041	111XXX
			75.723, 75.698, 75.410, 75.359, 75.336, 75.051, 75.028	1111XX
			75.712, 75.41, 75.348, 75.041	1110XX
3	74.05	X010XX	77.100	X010XX
	74.210, 73.908	X110XX	77.250, 76.950	X110XX
	74.209, 73.906	X011XX	77.261, 76.950	0110XX
	74.359, 74.050, 73.755	X111XX	77.279, 77.243, 76.970, 76.933	1110XX

	74.382, 74.363, 74.353, 74.336, 74.079, 74.064, 74.050, 74.033, 73.779, 73.762, 73.749, 73.733	X11111	77.250, 76.950	X011XX
	74.359, 74.05, 73.755	X11100	77.411, 77.100, 76.790	X111XX
			77.455, 77.441, 77.420, 77.401, 77.383, 77.369 77.147, 77.100, 77.077, 77.056, 76.836, 76.819,	111111
			77.420, 77.399, 77.10, 76.803, 76.782	011110
			77.411, 77.10, 76.790	011100
4	70.870	XX010X	70.830	XX010X
	71.037, 70.717	XX110X	70.978, 70.654	XX110X
	71.037, 70.717	XX011X	70.978, 70.654	XX011X
	71.165, 70.852, 70.548	XX111X	71.117, 70.788, 70.480	XX111X
	71.183, 71.153, 70.852, 70.561, 70.532	X1111X	71.131, 71.104, 70.788, 70.496, 70.469	X1111X
	71.165, 70.852, 70.548	X0111X	71.118, 70.788, 70.484	X0111X
5	72.55	XXX010	77.190	XXX010
	72.719, 72.393	XXX110	77.350, 77.023	XXX110
	72.719, 72.372	XXX011	77.352, 77.012	XXX011
	72.877, 72.55, 72.218	XXX111	77.517, 77.190, 76.847	XXX111
			77.519, 77.501, 77.19, 76.847, 76.835	XX1111
			77.515, 77.19, 76.845	XX0111
6	61.850	XXXX01	61.970	XXXX01
	62.008, 61.667	XXXX11	62.132, 61.787	XXXX11
	62.036, 62.008, 61.982, 61.695, 61.667, 61.637	1X1X11	62.165, 62.132, 62.098, 61.815, 61.789, 61.756	1X1X11
	62.008, 61.667	0X0X11	62.132, 61.787	0X0X11
			62.146, 62.112, 61.803, 61.769	1X0X11
				0X1X11

A.4 6.7-dimethyl-8-ribityllumazine

Chemical shift (ppm)	**X group (6.7-dimethyl-8-ribityl-lumazine)**	**X group (glucose)**
18.09	XX01	XXX0X1
17.89, 18.27	XX11	XXX1X1
18.09	X001	XX00X1
18.08, 18.10	X101	XX10X1
17.89, 18.27	X011	XX01X1
17.88, 17.91, 18.25, 18.28	X111	XX11X1
152.30	X010	XX01X0
152.08, 152.50	X110	XX11X0 XX01X1
151.86, 152.70	X111	XX11X1
151.86, 152.70	0111	X011X1
151.83, 151.89, 152.67, 152.73	1111	X111X1
144.95	010X	X010XX
144.71, 145.16	110X	X110XX X011XX
144.47, 145.15, 145.17	111X	X111XX
144.95	0100	X010X0
144.73, 145.16	1100	X110X0
144.71, 144.74, 145.15, 145.17	1101	X110X1
144.69, 145.16	0110	X011X0
144.67, 144.70, 145.15, 145.17	0111	X011X1
144.47, 145.37	1110	X111X0
144.45, 144.48, 145.36, 145.38	1111	X111X1
21.40	100X	X100XX
21.37, 21.43	101X	X101XX
21.18, 21.61	110X	X110XX
21.15, 21.21, 21.58, 21.64	111X	X111XX

Appendix B. X group notation represents an optimal code set

The X group notation contains the elements 0, 1 or X to describe arbitrary isotopologs of carbon compounds. Symbol 0 designates ^{12}C, symbol 1 designates ^{13}C, and X designates ^{12}C or ^{13}C. An open curly brace designates the begin of an isotopolog signature, and the corresponding closed curly brace designates the end of an isotopolog signature. The symbols 0, 1 or X represent the carbon atoms of a compound in descending order relative to the magnitude of the number position (for examples of X groups, see chapters 3 and 4).

If an arbitrary code is assumed, which encodes *b* possible states for a single carbon atom position in a compound with *c* carbon atoms in total, the complexity *W* of the encoding scheme is proportional to *b* and *c* and can be formulated by introducing a constant factor α, $W = \alpha \cdot b \cdot c$. The dimensionality *D* of the state space of a single compound is $D = b^c$. Because of $c = \frac{W}{\alpha} \cdot b$, with $\frac{W}{\alpha}$ being constant, an optimal code set requires a basis *k* which maximizes $k^{\frac{W}{\alpha} \cdot \frac{1}{b}}$ to represent an isotopolog space a minimum of complexity. As shown by Wiener, 1948, the maximal magnitude of presentable states against the basis of an arbitrary state space is the Euler constant *e*. Therefore *k* is maximal at value *e*. Since isotopolog spaces cannot be fractal, *k* must be an integer. It can be concluded that the alphabet of the X group notation ($e \approx 3$) is an optimal code set to represent isotopologs an arbitrary carbon compound.