Top Metabolic Pathways in the Beyond Bacteria Cohort

The whole-genome shotgun sequencing protocol used for the Beyond Bacteria samples can be used to provide insight into potential molecular pathways, or the functional potential, of a sample. The basic idea is to compare the sequencing fragments produced to existing characterized reference databases in order to identify what genes are likely present. Sometimes just knowing specific genes is informative. However, much of the biochemical processes contained in your gut (and in your own body) are the result of complex networks of gene interactions known as pathways. Once we know what genes are present in a sample, we can begin to ask whether there are any combinations of genes (from within the same organism) that can be combined together into existing known molecular pathways. It is these pathways that provide for the production of vitamins, essential proteins, and many other processes critical not just for the life of the individual organism, but also for the health and well-being of the human in which the organism lives.

You can find the specific top pathways per Beyond Bacteria sample here, and heatmap comparing each Beyond Bacteria sample by pathways here.

One of the common interactions between hosts and their microbial symbionts is the exchange of chemicals. Hosts can contribute molecules to this exchange through their diets. Microbes are also able to modify host molecules. In return, the microbiome produces molecules that the host would otherwise not be able to use. For example, the human genome encodes less than twenty enzymes capable of breaking down complex sugars. In contrast, an assembled metagenomic library of 104 organisms encoded more than 15,000 enzymes for digesting these sugars (ref). Other metabolites produced by the microbiome that influence host health include the following: serotonin, with 95% of the body's supply made by gut bacteria; short-chain fatty acids, which regulate metabolism and brain function; and trimethylamine N-oxide (TMAO), which plays a role in the development of cardiovascular disease (ref).

Although microbes contribute many functions, human hosts and their microbes share common molecular mechanisms. Nucleic acids (DNA and RNA) store and communicate the function of the cell by determining what proteins can be made. To reproduce, cells need a way to make more DNA. They also need ways to repair DNA if it gets damaged. Proteins are required to carry out many cellular functions. They serve primarily as enzymes, which make the reactions necessary for life fast enough to make them practical. However, proteins are also involved in the structure of the cell, movement, and communication both within and outside the cell. Lipids provide a structure and may also send signals inside the cell. Sugars are useful for storing energy and may be modified to help intracellular communication. Small molecules and cofactors are used for communication. All of these necessary macromolecules are built from smaller units. DNA and RNA chains are

built from single nucleotide units. Proteins are made up of amino acids. Membrane lipids are typically triglycerides, made of carboxylic acids connected by a bridge molecule. And, long chain polysaccharides are made of single sugars.

Building and breakdown the molecular components of the cell requires energy. Energy is typically released when bonds are broken, like during carbohydrate metabolism, and consumed when bonds are made, such as during nucleotide synthesis. Two of the primary ways cells transfer energy are through oxidation/reduction (redox) intermediates and through phosphate bonds. Nicotinamide adenine dinucleotide (NAD or NADP) is a common is a common redox electron carrier. In its oxidized form, it can accept two electrons and a proton (H+). This can then be carried to a reaction where the electrons are needed, and be used to reduce another molecule. High-energy phosphate bonds are also used to store energy, like a battery. It takes energy to combine phosphate groups together, since they contain negatively charged oxygens that like to repel each other. When the bond between two phosphate groups is broken, the energy of keeping the oxygens together can be donated to another reaction. The most common way to find these phosphate bonds is in the form of adenosine-triphosphate (ATP).

The metagenomic pathways found within the human gut are a predictor of the metabolic intermediates which will be found there (ref). Some pathways provide unique functions to the host, while others are required for the maintenance of cellular life. We've described the most abundant pathways found in the Beyond Bacteria cohort, as well as a few pathways of interest which provide molecules you could not otherwise synthesize.

Carbon Fixation

Carbon is the molecular backbone of life, and all of the carbon in organisms was at one point carbon dioxide. That carbon dioxide becomes organic carbon (sugars, lipids, proteins, and other molecules) via a process called carbon fixation. Most of the organic carbon that we eat comes from carbon fixation done by plants; either we eat those plants directly, or we eat animal products that got their energy-rich carbon from plants. By extension, most of the microbes in our guts get their organic carbon from what we eat; however, some of them can instead synthesize organic carbon via their own carbon fixation pathways. In total, six kinds of carbon fixation have been identified in nature. The pathway that fixes carbon in green plants (the Calvin cycle), and therefore most of our food, is not found in our gut bacteria, but two other pathways are.

The reductive citrate cycle (also called the Arnon-Buchanan cycle) is a low-energy alternative to the Calvin cycle found in some Proteobacteria and a few other lineages. Its enzymes are sensitive to oxygen, and therefore it is found only in the anaerobic or microaerophilic regions of the gut. The dicarboxylate-hydroxybutyrate cycle is found in the crenarchaeota Thermoproteales and Desulfurococcales (orders of archaea, not bacteria). These archaea are also anaerobic. Thus, these two major pathways of carbon fixation require low-oxygen conditions, which describes the majority of the human gut.

Carbon Metabolism

As most of the organic carbon required by gut bacteria comes from the food we eat, pathways for utilizing high-energy carbon compounds are highly represented in the gut metagenomic data. Core functions for breaking down sugar for energy are represented in the carbon metabolism genes in the Beyond Bacteria metagenomes.

Glycolysis (also called the Embden-Meyerhof pathway), including the part involving three-carbon compounds, is the central pathway in heterotrophic metabolism (breakdown of organic carbon for energy). It is found in diverse organisms, from E. coli to yeast to humans. Glycolysis (literally "breakdown of glucose") is the conversion of one molecule of glucose into two molecules for pyruvate (a three-carbon compound), with two molecules each of NADH and ATP generated as by-products. Gluconeogenesis is essentially the reverse of glycolysis, converting two three-carbon compounds like pyruvate back into glucose. Why would a cell want to do both? This gives the cell flexibility to break down carbon stores to make ATP (a mode of metabolism called catabolism) or make biomass to grow using available energy (called anabolism). Using many of the same reactions, a cell can shift its metabolism in either direction depending on its needs.

The citrate cycle (also called the tricarboxylic acid (TCA) cycle or Krebs cycle) is the next step of energy production after glycolysis. First, pyruvate is converted to acetyl-CoA, which then feeds into the citrate cycle, where it is converted into three molecules of NADH, one molecule of GTP (similar to ATP), and two molecules of carbon dioxide. The NADH produced here (and in glycolysis) is really important, because it can be used to make much more ATP. Using a large membrane-bound complex called the electron transport chain, the electrons stored in NADH are passed from donor molecules to acceptor molecules, generating a gradient of protons (H+), with the electrons eventually accepted by oxygen to form water. The proton gradient is used to power the synthesis of ATP. Because oxygen is such a good electron acceptor, a lot of ATP can be made. Of course, since much of the gut is anaerobic, other less-efficient terminal electron acceptors have to be used. Compounds like fumarate, nitrate, and nitrite can all serve as final electron acceptors if oxygen isn't available. If all of this sounds complicated, that's because it is; scientists still don't understand how parts of the process work!

The serine pathway for formaldehyde assimilation is found in methanotrophic ("methane eating") bacteria. It is one of two ways for these bacteria to assimilate formaldehyde, produced from the oxidation of methane and methanol, into biosynthetic pathways. In other words, methanotrophs consume one-carbon methane or methanol to make energy, but rather than throw away the product (formaldehyde), they incorporate it into their cellular building blocks.

Nucleotide Synthesis

Nucleotides are the building blocks of DNA and RNA. DNA, of course, is the code or blueprint for making an organism, providing the instructions for making all the proteins in a cell. RNA is the molecule that takes those instructions and makes the proteins, fulfilling a variety of roles in the process. The structures of DNA and RNA nucleotides are very similar. The main difference is that deoxyribonucleotides (as in deoxyribonucleic acid, DNA) have just a proton (H+) on a particular part of the sugar, whereas ribonucleotides (as in ribonucleic acid,

RNA) have an -OH group there. The other difference is that the two molecules use slightly different nitrogenous bases to encode their messages. DNA uses the famous ACGT bases: adenine, cytosine, guanine, and thymine. RNA uses, instead of thymine, uracil (U). The bases are grouped into two categories based on their structures: purines (adenine, guanine) and pyrimidines (cytosine, thymine, uracil). The base-pairing in DNA and RNA generally follows these rules: A pairs with T (or U), and C pairs with G.

The pathways pyrimidine deoxyribonuleotide biosynthesis and pyrimidine ribonucleotide biosynthesis are, as their names suggest, involved in the synthesis of both DNA and RNA versions of cytosine, thymine, and uracil. The names change slightly when they are attached to a sugar (deoxyribose or ribose): cytidine, thymidine, and uridine. Hence, uridine monophosphate biosynthesis is the synthesis of a nucleotide with a uracil base, ribose sugar, and single phosphate molecule. Nucleotides can be either monophosphate, diphosphate, or triphosphate (think ATP), each one progressively more energy-rich. Based on our discussion here, Guanine ribonucleotide biosynthesis should be pretty self-explanatory. Inosine monophosphate biosynthesis produces a nucleotide found in certain types of RNA called transfer RNA. Inosine is interesting in that it can form so-called non-Watson-Crick base pairs; inosine can pair with cytosine, uracil, and adenine. Inosine can also be converted to guanine and adenosine.

Finally, it's not surprising that pathways for nucleotide synthesis are highly represented in human gut metagenomes. There generally isn't enough DNA or RNA in the food we eat to supply the nucleotide requirements for bacteria, and transporting them across membranes is difficult. Therefore, they need to synthesize most of their nucleotides de novo (from scratch).

Amino Acid Metabolism

Amino acids are the building blocks of proteins, just like nucleotides are the building blocks of DNA and RNA. Proteins are built from a combination of 20 different amino acids, each with different chemical properties that give proteins an amazing diversity of three-dimensional shapes and functions. Not all organisms can make all 20 amino acids. For example, humans can make only 11 of the 20; the other 9 are called "essential" amino acids because they must be consumed as part of our food. Bacteria generally are able to make more amino acids than humans.

For gut bacteria, amino acids in our diet are an important source of food. Gut bacteria can ferment (degrade) amino acids, obtaining energy and carbon. This fermentation of dietary amino acids has significant effects on host gut health and beyond. The products of amino acid breakdown affect host immunity and cell function, and also feed back on microbial composition and metabolism (ref).

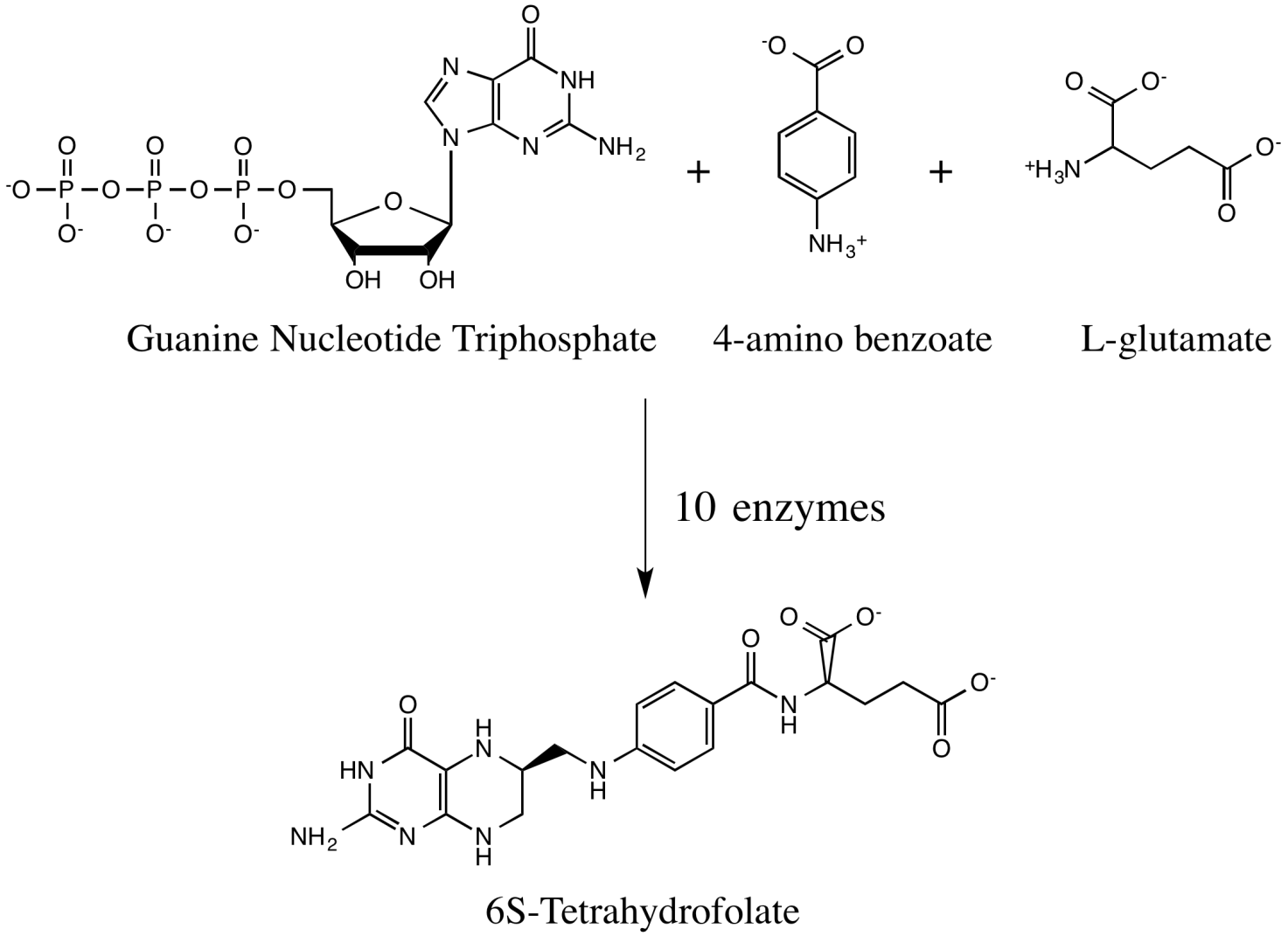
The amino acid pathways most abundant in the Beyond Bacteria cohort were cysteine biosynthesis, methionine degradation, and methionine salvage. Methionine is an essential amino acid, meaning it's required from our diet. These pathways indicate methionine in our gut is consumed (salvaged) by bacteria and also degraded, likely to make energy, other amino acids, or to harvest nitrogen or sulfur. Cysteine is not an

essential amino acid, as humans can synthesize it. However, the sulfur in our cysteine comes from methionine. Therefore, it's possible that methionine degradation by the microbiome is contributing to cysteine biosynthesis, either used by the bacteria or harvested by the host (you!). What's clear is that both bacteria and humans need a full suite of 20 amino acids to make functional proteins, which do almost all the major functions in cells, and also that bacteria use dietary amino acids for food. There is an intricate interplay between microbiome, host, and diet in the synthesis and breakdown of amino acids.

Vitamin Synthesis

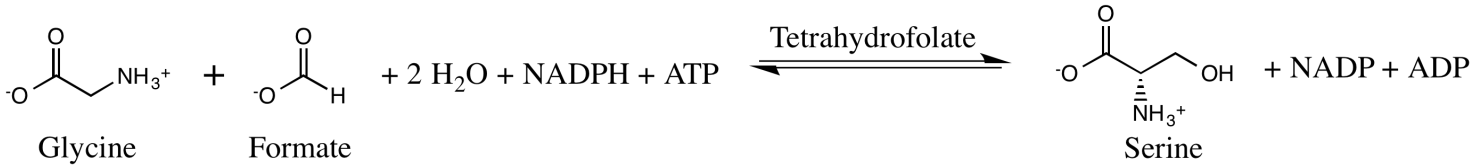
The most common vitamin production pathway in the human gut metagenome is folate synthesis. Tetrahydrofolate (the basic form of tetrahydrofolic acid), is also known as B9. Tetrahydrofolate is necessary for DNA repair and synthesis. Low tetrahydrofolate levels can lead to anemia, and there is some evidence that folic acid may be involved in depression (ref). Two pathways involving tetrahydrofolate were identified in the Beyond Bacteria samples: the synthesis of tetrahydrofolate from GTP and C1-unit interconversion by prokaryotes.

The first pathway is less abundant, and converts a guanine-nucleotide triphosphate, L-guanine, and a molecule of aminobenzoate to a molecule of tetrahydrofolate (Schema 1). Tetrahydrofolate is biologically available.



Schema 1, Molecular components of tetrahydrofolate biosynthesis.

The C1-interconversion pathway utilizes tetrahydrofolate to catalyze the conversion of glycine to serine (Schema 2). This pathway uses energy in the form of ATP to add a formate group first to the folate body and then transfer it to the glycine. The transfer is able to regenerate tetrahydrofolate.



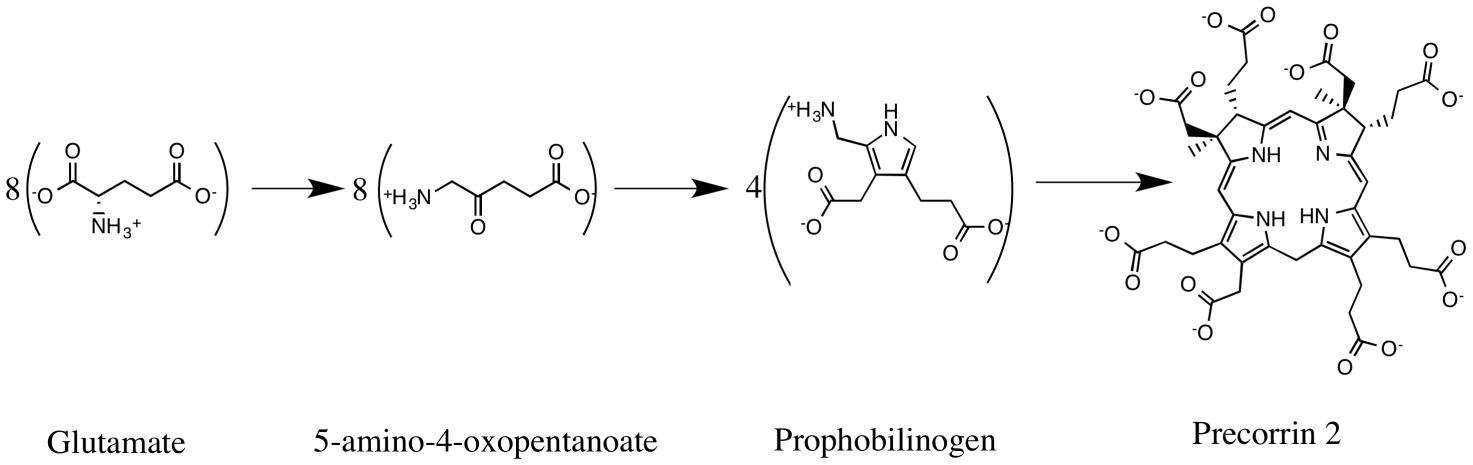
Schema 2, C1 conversion reaction. NADPH donates electrons to the reaction, and ATP provides energy.

Low-Abundance Pathways of Interest

Other Vitamin Synthesis

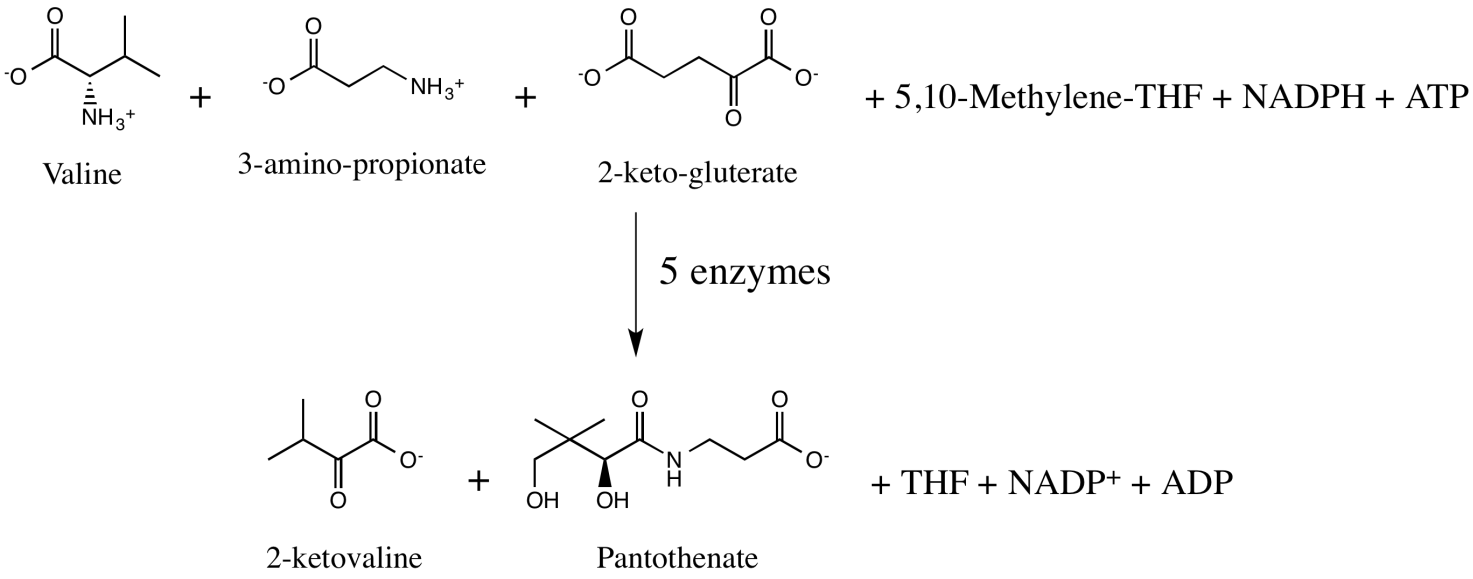
Two other B-vitamins can be synthesized by the microbiome: Vitamin B12, called Cobalamin, and Vitamin B5, or Pantothenate. Cobalamin is one of the most complex "small molecules" utilized by the body. It is involved in DNA synthesis, fatty acid metabolism, and amino acid metabolism. It is critical to neurological function and red blood cell formation. The structure of Cobalamin is unique among vitamins. It consists of a ring made out of interconnected planner carbons and nitrogens. The creates a binding pocket in the center of the molecule which is able to capture a Cobalt (I) ion. The captured cobalt ion means that cobalamin can bind negatively charged molecules above and below the plane of the cobalamin ring, which creates an important chemical interface.

Cobalamin is synthesized in four steps (ref). First, eight glutamates are combined to create the aromatic ring structure called Precorrin-2 (Schema 3). This is a general pathway used in Heme Biosynthesis (M00121). To make a form of heme, precorrin-2 would gain an Iron (II) ion at its center. For vitamin B12 Precorrin-2 is converted to cobamaide. Cobamaide, among other modifications, gains a Cobalt ion at its center. Cobamide is converted to Cob(I)yrinate diamide, which enters the cobalamine pathway found in the American Gut (M00121). This pathway converts Cob(I)yrinate diamide to cobalamin using dimethylbenzimadizole and ATP-energy. The cobalamin can then be utilized within the bacteria, or shared with their host.



Schema 3, insert figure Schema 3. Precorrin-2 biosynthesis from Glutamate.

Pantothenate is required for the synthesis of the co-enzyme, Acetyl CoA. Acetyl CoA is critical for carbon metabolism and fixation, as well as aiding fatty acid synthesis. Pantothenate is a long-chain of carbons which is synthesized from valine, a form of the amino acid alanine where the molecules have been rearranged called 3-amino-propionate, and 2-ketogluterate, which is converted to 2-ketovaline and Pantothenate. The five step processes uses a form of tetrahydrofolate, 5,10-methylenetetrahydrofolate which transfers one carbon into the creation, an electron donor, and ATP for energy (Schema 4).



Schema 4, insert figure Schema 4: Pantothenate Synthesis. THF is an abbreviation for Tetrahydrofolate. NADP+/NADPH are an oxidation/reduction pair which donate electrons to the reaction. ATP stores energy.

Credits

- Embriette Hyde
- Luke Thompson
- Justine Debelius
- Daniel McDonald