Biochemistry of Carbohydrates

Lloyd Benjamin

BIOCHEMISTRY OF CARBOHYDRATES

BIOCHEMISTRY OF CARBOHYDRATES

Lloyd Benjamin

BIBLIOTEX
Digital Library

Biochemistry of Carbohydrates
by Lloyd Benjamin

www.bibliotex.com

Ebook ISBN: 9781984665836

BIBLIOTEX
Digital Library

Published by:
Bibliotex
Canada
Website: www.bibliotex.com

Contents

Chapter 1

Carbohydrate

A **carbohydrate** is a biomolecule consisting of carbon (C), hydrogen (H) and oxygen (O) atoms, usually with a hydrogen–oxygen atom ratio of 2:1 (as in water) and thus with the empirical formula $C_m(H_2O)_n$ (where *m* may or may not be different from *n*). However, not all carbohydrates conform to this precise stoichiometric definition (e.g., uronic acids, deoxy-sugars such as fucose), nor are all chemicals that do conform to this definition automatically classified as carbohydrates (e.g. formaldehyde and acetic acid).

The term is most common in biochemistry, where it is a synonym of **saccharide**, a group that includes sugars, starch, and cellulose. The saccharides are divided into four chemical groups: monosaccharides, disaccharides, oligosaccharides, and polysaccharides. Monosaccharides and disaccharides, the smallest (lower molecular weight) carbohydrates, are commonly referred to as sugars. The word *saccharide* comes from the Greek word *σάκχαρον* (*sákkharon*), meaning "sugar". While the scientific nomenclature of carbohydrates is complex, the names of the monosaccharides and disaccharides very often end in the suffix -ose, which was originally taken from glucose, from Ancient Greek *γλεῦκος* (*gleûkos*, “wine, must”), and is used for almost all sugars, e.g. fructose (fruit sugar), sucrose (cane or beet sugar), ribose, amylose, lactose (milk sugar), etc.

Carbohydrates perform numerous roles in living organisms. Polysaccharides serve for the storage of energy (e.g. starch and glycogen) and as structural components (e.g. cellulose in

plants and chitin in arthropods). The 5-carbon monosaccharide ribose is an important component of coenzymes (e.g. ATP, FAD and NAD) and the backbone of the genetic molecule known as RNA. The related deoxyribose is a component of DNA. Saccharides and their derivatives include many other important biomolecules that play key roles in the immune system, fertilization, preventing pathogenesis, blood clotting, and development.

Carbohydrates are central to nutrition and are found in a wide variety of natural and processed foods. Starch is a polysaccharide.

It is abundant in cereals (wheat, maize, rice), potatoes, and processed food based on cereal flour, such as bread, pizza or pasta. Sugars appear in human diet mainly as table sugar (sucrose, extracted from sugarcane or sugar beets), lactose (abundant in milk), glucose and fructose, both of which occur naturally in honey, many fruits, and some vegetables. Table sugar, milk, or honey are often added to drinks and many prepared foods such as jam, biscuits and cakes.

Cellulose, a polysaccharide found in the cell walls of all plants, is one of the main components of insoluble dietary fiber. Although it is not digestible, insoluble dietary fiber helps to maintain a healthy digestive system by easing defecation.

Other polysaccharides contained in dietary fiber include resistant starch and inulin, which feed some bacteria in the microbiota of the large intestine, and are metabolized by these bacteria to yield short-chain fatty acids.

Terminology

In scientific literature, the term "carbohydrate" has many synonyms, like "sugar" (in the broad sense), "saccharide", "ose", "glucide", "hydrate of carbon" or "polyhydroxy compounds with aldehyde or ketone". Some of these terms, specially "carbohydrate" and "sugar", are also used with other meanings.

In food science and in many informal contexts, the term "carbohydrate" often means any food that is particularly rich in the complex carbohydrate starch (such as cereals, bread and pasta) or simple carbohydrates, such as sugar (found in candy, jams, and desserts).

Often in lists of nutritional information, such as the USDA National Nutrient Database, the term "carbohydrate" (or "carbohydrate by difference") is used for everything other than water, protein, fat, ash, and ethanol.

This includes chemical compounds such as acetic or lactic acid, which are not normally considered carbohydrates. It also includes dietary fiber which is a carbohydrate but which does not contribute much in the way of food energy (kilocalories), even though it is often included in the calculation of total food energy just as though it were a sugar.

In the strict sense, "sugar" is applied for sweet, soluble carbohydrates, many of which are used in food.

Structure

Formerly the name "carbohydrate" was used in chemistry for any compound with the formula C_m $(H_2O)_n$. Following this definition, some chemists considered formaldehyde (CH_2O) to be the simplest carbohydrate, while others claimed that title for glycolaldehyde. Today, the term is generally understood in the biochemistry sense, which excludes compounds with only one or two carbons and includes many biological carbohydrates which deviate from this formula. For example, while the above representative formulas would seem to capture the commonly known carbohydrates, ubiquitous and abundant carbohydrates often deviate from this. For example, carbohydrates often display chemical groups such as:*N*-acetyl (e.g. chitin), sulphate (e.g. glycosaminoglycans), carboxylic acid (e.g. sialic acid) and deoxy modifications (e.g. fucose and sialic acid).

Natural saccharides are generally built of simple carbohydrates called monosaccharides with general formula $(CH_2O)_n$ where *n* is three or more. A typical monosaccharide has the structure $H–(CHOH)_x(C=O)–(CHOH)_y–H$, that is, an aldehyde or ketone with many hydroxyl groups added, usually one on each carbonatom that is not part of the aldehyde or ketone functional group.

Examples of monosaccharides are glucose, fructose, and glyceraldehydes. However, some biological substances commonly called "monosaccharides" do not conform to this formula (e.g. uronic acids and deoxy-sugars such as fucose) and there are many chemicals that do conform to this formula but are not considered to be monosaccharides (e.g. formaldehyde CH_2O and inositol $(CH_2O)_6$).

The open-chain form of a monosaccharide often coexists with a closed ring form where the aldehyde/ketonecarbonyl group carbon (C=O) and hydroxyl group (–OH) react forming a hemiacetal with a new C–O–C bridge.

Monosaccharides can be linked together into what are called polysaccharides (or oligosaccharides) in a large variety of ways. Many carbohydrates contain one or more modified monosaccharide units that have had one or more groups replaced or removed. For example, deoxyribose, a component of DNA, is a modified version of ribose; chitin is composed of repeating units of N-acetyl glucosamine, a nitrogen-containing form of glucose.

Monosaccharides

Monosaccharides are the simplest carbohydrates in that they cannot be hydrolyzed to smaller carbohydrates. They are aldehydes or ketones with two or more hydroxyl groups. The general chemical formula of an unmodified monosaccharide is $(C•H_2O)_n$, literally a "carbon hydrate". Monosaccharides are important fuel molecules as well as building blocks for nucleic acids. The smallest monosaccharides, for which n=3, are dihydroxyacetone and D- and L-glyceraldehydes.

Classification of monosaccharides

Monosaccharides are classified according to three different characteristics: the placement of its carbonyl group, the number of carbon atoms it contains, and its chiral handedness. If the carbonyl group is an aldehyde, the monosaccharide is an aldose; if the carbonyl group is a ketone,

the monosaccharide is a ketose. Monosaccharides with three carbon atoms are called trioses, those with four are called tetroses, five are called pentoses, six are hexoses, and so on. These two systems of classification are often combined. For example, glucose is an aldohexose (a six-carbon aldehyde), ribose is an aldopentose (a five-carbon aldehyde), and fructose is a ketohexose (a six-carbon ketone).

Each carbon atom bearing a hydroxyl group (-OH), with the exception of the first and last carbons, are asymmetric, making them stereo centers with two possible configurations each (R or S). Because of this asymmetry, a number of isomers may exist for any given monosaccharide formula.

Using Le Bel-van't Hoff rule, the aldohexose D-glucose, for example, has the formula $(C{\cdot}H_2O)_6$, of which four of its six carbons atoms are stereogenic, making D-glucose one of 2=16 possible stereoisomers. In the case of glyceraldehydes, an aldotriose, there is one pair of possible stereoisomers, which are enantiomers and epimers. 1, 3-dihydroxyacetone, the ketose corresponding to the aldose glyceraldehydes, is a symmetric molecule with no stereo centers.

The assignment of D or L is made according to the orientation of the asymmetric carbon furthest from the carbonyl group: in a standard Fischer projection if the hydroxyl group is on the right the molecule is a D sugar, otherwise it is an L sugar. The "D-" and "L-" prefixes should not be confused with "d-" or "l-", which indicate the direction that the sugar rotates plane polarized light. This usage of "d-" and "l-" is no longer followed in carbohydrate chemistry.

Ring-straight chain isomerism

The aldehyde or ketone group of a straight-chain monosaccharide will react reversibly with a hydroxyl group on a different carbon atom to form a hemiacetal or hemiketal, forming a heterocyclic ring with an oxygen bridge between two carbon atoms. Rings with five and six atoms are calledfuranose and pyranose forms, respectively, and exist in equilibrium with the straight-chain form.

During the conversion from straight-chain form to the cyclic form, the carbon atom containing the carbonyl oxygen, called the anomeric carbon, becomes a stereogenic center with two possible configurations: The oxygen atom may take a position either above or below the plane of the ring. The resulting possible pair of stereoisomers is calledanomers. In the *α anomer*, the -OH substituent on the anomeric carbon rests on the opposite side (trans) of the ring from the CH_2OH side branch. The alternative form, in which the CH_2OH substituent and the anomeric hydroxyl are on the same side (cis) of the plane of the ring, is called the *β anomer*.

Use in living organisms

Monosaccharides are the major fuel source for metabolism, being used both as an energy source (glucose being the most important in nature) and in biosynthesis. When monosaccharides are not immediately needed by many cells, they are often converted to more space-efficient forms, often polysaccharides. In many animals, including humans, this storage form is glycogen, especially in liver and muscle cells. In plants, starchis used for the same purpose. The most

abundant carbohydrate, cellulose, is a structural component of the cell wall of plants and many forms of algae. Ribose is a component of RNA. Deoxyribose is a component of DNA. Lyxose is a component of lyxoflavin found in the human heart.Ribulose and xylulose occur in the pentose phosphate pathway. Galactose, a component of milk sugar lactose, is found in galactolipids in plant cell membranes and in glycoproteins in many tissues. Mannose occurs in human metabolism, especially in the glycosylation of certain proteins. Fructose, or fruit sugar, is found in many plants and humans, it is metabolized in the liver, absorbed directly into the intestines during digestion, and found in semen. Trehalose, a major sugar of insects, is rapidly hydrolyzed into two glucose molecules to support continuous flight.

Disaccharides

Two joined monosaccharides are called a disaccharide and these are the simplest polysaccharides. Examples include sucrose and lactose. They are composed of two monosaccharide units bound together by a covalent bond known as a glycosidic linkage formed via a dehydration reaction, resulting in the loss of a hydrogen atom from one monosaccharide and a hydroxyl group from the other. The formula of unmodified disaccharides is $C_{12}H_{22}O_{11}$. Although there are numerous kinds of disaccharides, a handful of disaccharides are particularly notable.

Sucrose, pictured to the right, is the most abundant disaccharide, and the main form in which carbohydrates are transported in plants. It is composed of one D-glucose molecule and one D-fructose molecule. The systematic name

for sucrose, *O*-α-D-glucopyranosyl-(1→2)-D-fructofuranoside, indicates four things:

Its monosaccharides: glucose and fructose
Their ring types: glucose is a pyranose and fructose is a furanose
How they are linked together: the oxygen on carbon number 1 (C1) of α-D-glucose is linked to the C2 of D-fructose.
The *-oside* suffix indicates that the anomeric carbon of both monosaccharides participates in the glycosidic bond.

Lactose, a disaccharide composed of one D-galactose molecule and one D-glucose molecule, occurs naturally in mammalian milk. The systematic name for lactose is *O*-β-D-galactopyranosyl-(1→4)-D-glucopyranose. Other notable disaccharides include maltose (two D-glucoses linked α-1,4) and cellobiose (two D-glucoses linked β-1,4). Disaccharides can be classified into two types: reducing and non-reducing disaccharides. If the functional group is present in bonding with another sugar unit, it is called a reducing disaccharide or biose.

Nutrition

Carbohydrate consumed in food yields 3.87 kilocalories of energy per gram for simple sugars, and 3.57 to 4.12 kilocalories per gram for complex carbohydrate in most other foods. Relatively high levels of carbohydrate are associated with processed foods or refined foods made from plants, including sweets, cookies and candy, table sugar, honey, soft

drinks, breads and crackers, jams and fruit products, pastas and breakfast cereals. Lower amounts of carbohydrate are usually associated with unrefined foods, including beans, tubers, rice, and unrefined fruit. Animal-based foods generally have the lowest carbohydrate levels, although milk does contain a high proportion of lactose.

Organisms typically cannot metabolize all types of carbohydrate to yield energy. Glucose is a nearly universal and accessible source of energy. Many organisms also have the ability to metabolize other monosaccharides and disaccharides but glucose is often metabolized first. In *Escherichia coli*, for example, the lac operon will express enzymes for the digestion of lactose when it is present, but if both lactose and glucose are present the *lac* operon is repressed, resulting in the glucose being used first (see: Diauxie). Polysaccharides are also common sources of energy. Many organisms can easily break down starches into glucose; most organisms, however, cannot metabolize cellulose or other polysaccharides like chitin and arabinoxylans. These carbohydrate types can be metabolized by some bacteria and protists. Ruminants and termites, for example, use microorganisms to process cellulose. Even though these complex carbohydrates are not very digestible, they represent an important dietary element for humans, called dietary fiber. Fiber enhances digestion, among other benefits.

The Institute of Medicine recommends that American and Canadian adults get between 45 and 65% of dietary energy from whole-grain carbohydrates. The Food and Agriculture Organization and World Health Organization jointly recommend that national dietary guidelines set a goal of 55–75% of total

energy from carbohydrates, but only 10% directly from sugars (their term for simple carbohydrates). A 2017 Cochrane Systematic Review concluded that there was insufficient evidence to support the claim that whole grain diets can affect cardiovascular disease.

Classification

Nutritionists often refer to carbohydrates as either simple or complex. However, the exact distinction between these groups can be ambiguous. The term *complex carbohydrate* was first used in the U.S. Senate Select Committee on Nutrition and Human Needs publication *Dietary Goals for the United States* (1977) where it was intended to distinguish sugars from other carbohydrates (which were perceived to be nutritionally superior).However, the report put "fruit, vegetables and whole-grains" in the complex carbohydrate column, despite the fact that these may contain sugars as well as polysaccharides. This confusion persists as today some nutritionists use the term complex carbohydrate to refer to any sort of digestible saccharide present in a whole food, where fiber, vitamins and minerals are also found (as opposed to processed carbohydrates, which provide energy but few other nutrients). The standard usage, however, is to classify carbohydrates chemically: simple if they are sugars (monosaccharides and disaccharides) and complex if they are polysaccharides (or oligosaccharides).

In any case, the simple vs. complex chemical distinction has little value for determining the nutritional quality of carbohydrates. Some simple carbohydrates (e.g. fructose) raise blood glucose rapidly, while some complex carbohydrates

(starches), raise blood sugar slowly. The speed of digestion is determined by a variety of factors including which other nutrients are consumed with the carbohydrate, how the food is prepared, individual differences in metabolism, and the chemistry of the carbohydrate. Carbohydrates are sometimes divided into "available carbohydrates", which are absorbed in the small intestine and "unavailable carbohydrates", which pass to the large intestine, where they are subject to fermentation by the gastrointestinal microbiota.

The USDA's*Dietary Guidelines for Americans 2010* call for moderate- to high-carbohydrate consumption from a balanced diet that includes six one-ounce servings of grain foods each day, at least half from whole grain sources and the rest from enriched.

The glycemic index (GI) and glycemic load concepts have been developed to characterize food behavior during human digestion. They rank carbohydrate-rich foods based on the rapidity and magnitude of their effect on blood glucose levels. Glycemic index is a measure of how quickly food glucose is absorbed, while glycemic load is a measure of the total absorbable glucose in foods. The insulin index is a similar, more recent classification method that ranks foods based on their effects on blood insulin levels, which are caused by glucose (or starch) and some amino acids in food.

Health effects of dietary carbohydrate restriction

Low-carbohydrate diets may miss the health advantages – such as increased intake of dietary fiber – afforded by high-quality carbohydrates found in legumes and pulses, whole grains,

fruits, and vegetables. Disadvantages of the diet might include halitosis, headache and constipation, and in general the potential adverse effects of carbohydrate-restricted diets are under-researched, particularly for possible risks of osteoporosis and cancer incidence.

Carbohydrate-restricted diets can be as effective as low-fat diets in helping achieve weight loss over the short term when overall calorie intake is reduced. An Endocrine Society scientific statement said that "when calorie intake is held constant [...] body-fat accumulation does not appear to be affected by even very pronounced changes in the amount of fat vs carbohydrate in the diet." In the long term, effective weight loss or maintenance depends on calorie restriction, not the ratio of macronutrients in a diet. The reasoning of diet advocates that carbohydrates cause undue fat accumulation by increasing blood insulin levels, and that low-carbohydrate diets have a "metabolic advantage", is not supported by clinical evidence. Further, it is not clear how low-carbohydrate dieting affects cardiovascular health, although two reviews showed that carbohydrate restriction may improve lipid markers of cardiovascular disease risk.

Carbohydrate-restricted diets are no more effective than a conventional healthy diet in preventing the onset of type 2 diabetes, but for people with type 2 diabetes, they are a viable option for losing weight or helping with glycemic control. There is limited evidence to support routine use of low-carbohydrate dieting in managing type 1 diabetes. The American Diabetes Association recommends that people with diabetes should adopt a generally healthy diet, rather than a diet focused on carbohydrate or other macronutrients.

An extreme form of low-carbohydrate diet – the ketogenic diet – is established as a medical diet for treating epilepsy. Through celebrity endorsement during the early 21st century, it became a fad diet as a means of weight loss, but with risks of undesirable side effects, such as low energy levels and increased hunger, insomnia, nausea, and gastrointestinal discomfort. The British Dietetic Association named it one of the "top 5 worst celeb diets to avoid in 2018".

Metabolism

Carbohydrate metabolism is the series of biochemical processes responsible for the formation, breakdown and interconversion of carbohydrates in livingorganisms.

The most important carbohydrate is glucose, a simple sugar (monosaccharide) that is metabolized by nearly all known organisms. Glucose and other carbohydrates are part of a wide variety of metabolic pathways across species: plants synthesize carbohydrates from carbon dioxide and water by photosynthesis storing the absorbed energy internally, often in the form of starch or lipids. Plant components are consumed by animals and fungi, and used as fuel for cellular respiration. Oxidation of one gram of carbohydrate yields approximately 16 kJ (4 kcal) of energy, while the oxidation of one gram of lipids yields about 38 kJ (9 kcal).

The human body stores between 300 and 500 g of carbohydrates depending on body weight, with the skeletal muscle contributing to a large portion of the storage. Energy obtained from metabolism (e.g., oxidation of glucose) is usually stored temporarily within cells in the form of ATP. Organisms

capable of anaerobic and aerobic respiration metabolize glucose and oxygen (aerobic) to release energy, with carbon dioxide and water as byproducts.

Catabolism

Catabolism is the metabolic reaction which cells undergo to break down larger molecules, extracting energy. There are two major metabolic pathways of monosaccharide catabolism: glycolysis and the citric acid cycle.

In glycolysis, oligo- and polysaccharides are cleaved first to smaller monosaccharides by enzymes called glycoside hydrolases. The monosaccharide units can then enter into monosaccharide catabolism.

A 2 ATP investment is required in the early steps of glycolysis to phosphorylate Glucose to Glucose 6-Phosphate (G6P) and Fructose 6-Phosphate (F6P) to Fructose 1,6-biphosphate (FBP), thereby pushing the reaction forward irreversibly. In some cases, as with humans, not all carbohydrate types are usable as the digestive and metabolic enzymes necessary are not present.

Carbohydrate chemistry

Carbohydrate chemistry is a large and economically important branch of organic chemistry. Some of the main organic reactions that involve carbohydrates are:

Carbohydrate acetalisation
Cyanohydrin reaction

Lobry de Bruyn–Van Ekenstein transformation

Amadori rearrangement

Nef reaction

Wohl degradation

Koenigs–Knorr reaction

Carbohydrate digestion

Chapter 2

Carbohydrate Chemistry

Carbohydrate chemistry is a subdiscipline of chemistry primarily concerned with the synthesis, structure, and function of carbohydrates. Due to the general structure of carbohydrates, their synthesis is often preoccupied with the selective formation of glycosidic linkages and the selective reaction of hydroxyl groups; as a result, it relies heavily on the use of protecting groups.

Monosaccharides

Monosaccharides (from Greek*monos*: single, *sacchar*: sugar), also called **simple sugars**, are the simplest form of sugar and the most basic units (monomers) of carbohydrates. The general formula is $C_nH_{2n}O_n$, or [C*n*(H2O)*n*] or { CH2O}*n* albeit not all molecules fitting this formula (e.g. acetic acid) are carbohydrates. They are usually colorless, water-soluble, and crystalline solids. Contrary to their name (sugars), only some monosaccharides have a sweet taste.

Examples of monosaccharides include glucose (dextrose), fructose (levulose), and galactose. Monosaccharides are the building blocks of disaccharides (such as sucrose and lactose) and polysaccharides (such as cellulose and starch). Each carbon atom that supports a hydroxyl group is chiral, except those at the end of the chain. This gives rise to a number of isomeric forms, all with the same chemical formula. For

instance, galactose and glucose are both aldohexoses, but have different physical structures and chemical properties.

The monosaccharide glucose plays a pivotal role in metabolism, where the chemical energy is extracted through glycolysis and the citric acid cycle to provide energy to living organisms. Some other monosaccharides can be converted in the living organism to glucose.

Structure and nomenclature

With few exceptions (e.g., deoxyribose), monosaccharides have this chemical formula: $(CH_2O)_x$, where conventionally $x \geq 3$. Monosaccharides can be classified by the number x of carbon atoms they contain: triose (3), tetrose (4), pentose (5), hexose (6), heptose (7), and so on.

Glucose, used as an energy source and for the synthesis of starch, glycogen and cellulose, is a hexose. Ribose and deoxyribose (in RNA and DNA respectively) are pentose sugars. Examples of heptoses include the ketoses, mannoheptulose and sedoheptulose. Monosaccharides with eight or more carbons are rarely observed as they are quite unstable. In aqueous solutions monosaccharides exist as rings if they have more than four carbons.

Linear-chain monosaccharides

Simple monosaccharides have a linear and unbranched carbon skeleton with one carbonyl (C=O) functional group, and one hydroxyl (OH) group on each of the remaining carbon atoms. Therefore, the molecular structure of a simple monosaccharide

can be written as $H(CHOH)_n(C=O)(CHOH)_mH$, where $n + 1 + m = x$; so that its elemental formula is $C_xH_{2x}O_x$.

By convention, the carbon atoms are numbered from 1 to x along the backbone, starting from the end that is closest to the C=O group. Monosaccharides are the simplest units of carbohydrates and the simplest form of sugar.

If the carbonyl is at position 1 (that is, n or m is zero), the molecule begins with a formyl group H(C=O)– and is technically an aldehyde. In that case, the compound is termed an aldose. Otherwise, the molecule has a keto group, a carbonyl –(C=O)– between two carbons; then it is formally a ketone, and is termed a ketose. Ketoses of biological interest usually have the carbonyl at position 2.

The various classifications above can be combined, resulting in names such as "aldohexose" and "ketotriose".

A more general nomenclature for open-chain monosaccharides combines a Greek prefix to indicate the number of carbons (tri-, tetr-, pent-, hex-, etc.) with the suffixes "-ose" for aldoses and "-ulose" for ketoses. In the latter case, if the carbonyl is not at position 2, its position is then indicated by a numeric infix. So, for example, $H(C=O)(CHOH)_4H$ is pentose, $H(CHOH)(C=O)(CHOH)_3H$ is pentulose, and $H(CHOH)_2(C=O)(CHOH)_2H$ is pent-3-ulose.

Open-chain stereoisomers

Two monosaccharides with equivalent molecular graphs (same chain length and same carbonyl position) may still be distinct stereoisomers, whose molecules differ in spatial orientation.

This happens only if the molecule contains a stereogenic center, specifically a carbon atom that is chiral (connected to four distinct molecular sub-structures). Those four bonds can have any of two configurations in space distinguished by their handedness. In a simple open-chain monosaccharide, every carbon is chiral except the first and the last atoms of the chain, and (in ketoses) the carbon with the keto group.

For example, the triketose H(CHOH)(C=O)(CHOH)H (glycerone, dihydroxyacetone) has no stereogenic center, and therefore exists as a single stereoisomer. The other triose, the aldose $H(C=O)(CHOH)_2H$ (glyceraldehyde), has one chiral carbon — the central one, number 2 — which is bonded to groups –H, –OH, $–C(OH)H_2$, and –(C=O)H. Therefore, it exists as two stereoisomers whose molecules are mirror images of each other (like a left and a right glove). Monosaccharides with four or more carbons may contain multiple chiral carbons, so they typically have more than two stereoisomers. The number of distinct stereoisomers with the same diagram is bounded by 2^c, where *c* is the total number of chiral carbons.

The Fischer projection is a systematic way of drawing the skeletal formula of an acyclic monosaccharide so that the handedness of each chiral carbon is well specified. Each stereoisomer of a simple open-chain monosaccharide can be identified by the positions (right or left) in the Fischer diagram of the chiral hydroxyls (the hydroxyls attached to the chiral carbons).

Most stereoisomers are themselves chiral (distinct from their mirror images). In the Fischer projection, two mirror-image isomers differ by having the positions of all chiral hydroxyls

reversed right-to-left. Mirror-image isomers are chemically identical in non-chiral environments, but usually have very different biochemical properties and occurrences in nature.

While most stereoisomers can be arranged in pairs of mirror-image forms, there are some non-chiral stereoisomers that are identical to their mirror images, in spite of having chiral centers. This happens whenever the molecular graph is symmetrical, as in the 3-ketopentoses $H(CHOH)_2(CO)(CHOH)_2H$, and the two halves are mirror images of each other. In that case, mirroring is equivalent to a half-turn rotation. For this reason, there are only three distinct 3-ketopentose stereoisomers, even though the molecule has two chiral carbons.

Distinct stereoisomers that are not mirror-images of each other usually have different chemical properties, even in non-chiral environments. Therefore, each mirror pair and each non-chiral stereoisomer may be given a specific monosaccharide name. For example, there are 16 distinct aldohexose stereoisomers, but the name "glucose" means a specific pair of mirror-image aldohexoses. In the Fischer projection, one of the two glucose isomers has the hydroxyl at left on C3, and at right on C4 and C5; while the other isomer has the reversed pattern. These specific monosaccharide names have conventional three-letter abbreviations, like "Glu" for glucose and "Thr" for threose.

Generally, a monosaccharide with n asymmetrical carbons has 2 stereoisomers. The number of open chain stereoisomers for an aldose monosaccharide is larger by one than that of a ketose monosaccharide of the same length. Every ketose will have 2 stereoisomers where $n > 2$ is the number of carbons.

Every aldose will have 2 stereoisomers where $n > 2$ is the number of carbons. These are also referred to as epimers which have the different arrangement of –OH and –H groups at the asymmetric or chiral carbon atoms (this does not apply to those carbons having the carbonyl functional group).

Configuration of monosaccharides

Like many chiral molecules, the two stereoisomers of glyceraldehyde will gradually rotate the polarization direction of linearly polarized light as it passes through it, even in solution. The two stereoisomers are identified with the prefixes D- and L-, according to the sense of rotation: D-glyceraldehyde is dextrorotatory (rotates the polarization axis clockwise), while L-glyceraldehyde is levorotatory (rotates it counterclockwise).

The D- and L- prefixes are also used with other monosaccharides, to distinguish two particular stereoisomers that are mirror-images of each other. For this purpose, one considers the chiral carbon that is furthest removed from the C=O group. Its four bonds must connect to –H, –OH, –C(OH)H, and the rest of the molecule. If the molecule can be rotated in space so that the directions of those four groups match those of the analog groups in D-glyceraldehyde's C2, then the isomer receives the D- prefix. Otherwise, it receives the L- prefix.

In the Fischer projection, the D- and L- prefixes specifies the configuration at the carbon atom that is second from bottom: D- if the hydroxyl is on the right side, and L- if it is on the left side.

Note that the D- and L- prefixes do not indicate the direction of rotation of polarized light, which is a combined effect of the arrangement at all chiral centers. However, the two enantiomers will always rotate the light in opposite directions, by the same amount. See also D/L system.

Cyclisation of monosaccharides

A monosaccharide often switches from the acyclic (open-chain) form to a cyclic form, through a nucleophilic addition reaction between the carbonyl group and one of the hydroxyls of the same molecule. The reaction creates a ring of carbon atoms closed by one bridging oxygen atom. The resulting molecule has a hemiacetal or hemiketal group, depending on whether the linear form was an aldose or a ketose. The reaction is easily reversed, yielding the original open-chain form.

In these cyclic forms, the ring usually has five or six atoms. These forms are called furanoses and pyranoses, respectively — by analogy with furan and pyran, the simplest compounds with the same carbon-oxygen ring (although they lack the double bonds of these two molecules).

For example, the aldohexose glucose may form a hemiacetal linkage between the aldehyde group on carbon 1 and the hydroxyl on carbon 4, yielding a molecule with a 5-membered ring, called glucofuranose. The same reaction can take place between carbons 1 and 5 to form a molecule with a 6-membered ring, called glucopyranose. Cyclic forms with a seven-atom ring (the same of oxepane), rarely encountered, are called heptoses.

For many monosaccharides (including glucose), the cyclic forms predominate, in the solid state and in solutions, and therefore the same name commonly is used for the open- and closed-chain isomers. Thus, for example, the term "glucose" may signify glucofuranose, glucopyranose, the open-chain form, or a mixture of the three.

Cyclization creates a new stereogenic center at the carbonyl-bearing carbon. The –OH group that replaces the carbonyl's oxygen may end up in two distinct positions relative to the ring's midplane.

Thus each open-chain monosaccharide yields two cyclic isomers (anomers), denoted by the prefixes α- and β-. The molecule can change between these two forms by a process called mutarotation, that consists in a reversal of the ring-forming reaction followed by another ring formation.

Haworth projection

The stereochemical structure of a cyclic monosaccharide can be represented in a Haworth projection. In this diagram, the α-isomer for the pyranose form of a D-aldohexose has the –OH of the anomeric carbon below the plane of the carbon atoms, while the β-isomer has the –OH of the anomeric carbon above the plane.

Pyranoses typically adopt a chair conformation, similar to that of cyclohexane. In this conformation, the α-isomer has the –OH of the anomeric carbon in an axial position, whereas the β-isomer has the –OH of the anomeric carbon in equatorial position (considering D-aldohexose sugars).

Derivatives

A large number of biologically important modified monosaccharides exist:

Amino sugars such as:
galactosamine
glucosamine
sialic acid
N-acetylglucosamine
Sulfosugars such as:
sulfoquinovose
Others such as:
ascorbic acid
mannitol
glucuronic acid

Carbohydrate synthesis

Carbohydrate synthesis is a sub-field of organic chemistry concerned specifically with the generation of natural and unnatural carbohydrate structures. This can include the synthesis of monosaccharide residues or structures containing more than one monosaccharide, known as oligosaccharides.

Background

Generally speaking, carbohydrates can be classified into two groups, simple sugars, and complex carbohydrates. Simple sugars, also called monosaccharides, are carbohydrates that

can not be converted into smaller sugars by hydrolysis. When two or more monosaccharide units are connected via a glycoside linkage, complex carbohydrates are formed. Complex carbohydrates, according to the different number of monosaccharide units, can be classed into three groups, disaccharides, oligosaccharides, and polysaccharides. A disaccharide is formed from two monosaccharides. Oligosaccharides can be formed by a small number of monosaccharides linked together. Higher oligosaccharides are called polysaccharides. It is now well known that glycoconjugates play an indispensable role in many biological processes. These biological processes in which carbohydrates are involved are typically associated not with monosaccharides, but with oligosaccharides structures of glycoconjugates. Therefore, the oligosaccharide synthesis becomes more and more important in studying biological activities.

Oligosaccharide synthesis

Oligosaccharides have diverse structures. The number of monosaccharides, ring size, the different anomericstereochemistry, and the existence of the branched-chain sugars all contribute to the amazing complexity of the oligosaccharide structures. The essence of the reducing oligosaccharide synthesis is connecting the anomeric hydroxyl of the glycosyl donors to the alcoholic hydroxyl groups of the glycosyl acceptors. Protection of the hydroxyl groups of the acceptor with the target alcoholic hydroxyl group unprotected can assure regiochemical control. Additionally, factors such as the different protecting

groups, the solvent, and the glycosylation methods can influence the anomeric configurations. This concept is illustrated by an oligosaccharide synthesis in Scheme 1. Oligosaccharide synthesis normally consists of four parts: preparation of the glycosyl donors, preparation of the glycosyl acceptors with a single unprotected hydroxyl group, the coupling of them, and the deprotection process.

Building blocks

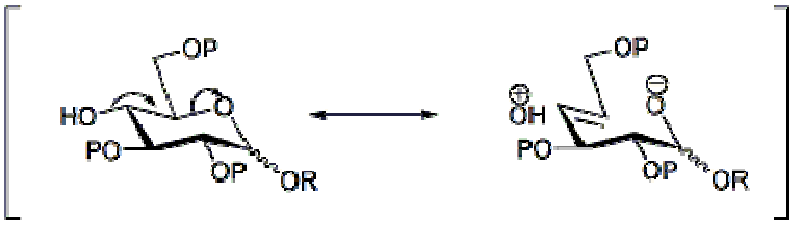
Common donors in oligosaccharide synthesis are glycosyl halides, glycosyl acetates, thioglycosides, trichloroacetimidates, pentenyl glycosides, and glycals. Of all these donors, glycosyl halides are classic donors, which played a historical role in the development of glycosylation reactions. Thioglycoside and trichloroacetimidate donors are used more than others in contemporary glycosylation methods. When it comes to the trichloroacetimidate method, one of the advantages is that there is no need to introduce heavy metal reagents in the activation process. Moreover, using different bases can selectively lead to different anomeric configurations. (Scheme 2) As to the thioglycosides, the greatest strength is that they can offer temporary protection to the anomeric centre because they can survive after most of the activation processes. Additionally, a variety of activation methods can be employed, such as NIS/ AgOTf, NIS/ TfOH, IDCP (Iodine Dicollidine Perchlorate), iodine, and Ph_2SO/ Tf_2O. Furthermore, in the preparation of 1, 2-trans glycosidic linkage,

using thioglycosides and imidates can promote the rearrangement of the orthoester byproducts, since the reaction mixtures are acidic enough.

Stereoselectivity

The structures of acceptors play a critical role in the rate and stereoselectivity of glycosylations. Generally, the unprotected hydroxyl groups are less reactive when they are between bulky protecting groups.

That is the reason why the hydroxyl group at OH-4 in pyranosides is unreactive. Hyperconjugation is involved when OH-4 is anti-periplanar to the ring oxygen, which can also reduce its reactivity. (Scheme 3) Furthermore, acyl protecting groups can reduce the reactivity of the acceptors compared with alkyl protecting groups because of their electron-withdrawing ability. Hydroxyl group at OH-4 of N-acetylglucosamine derivatives is particularly unreactive.



Scheme 3

The glycosidic bondis formed from a glycosyl donor and a glycosyl acceptor. There are four types of glycosidic linkages: 1, 2-trans-α, 1, 2-trans-beta, 1, 2-cis-α, and 1, 2-cis-beta linkages. 1, 2-trans glycosidic linkages can be easily achieved by using 2-O-acylated glycosyl donors (neighboring group

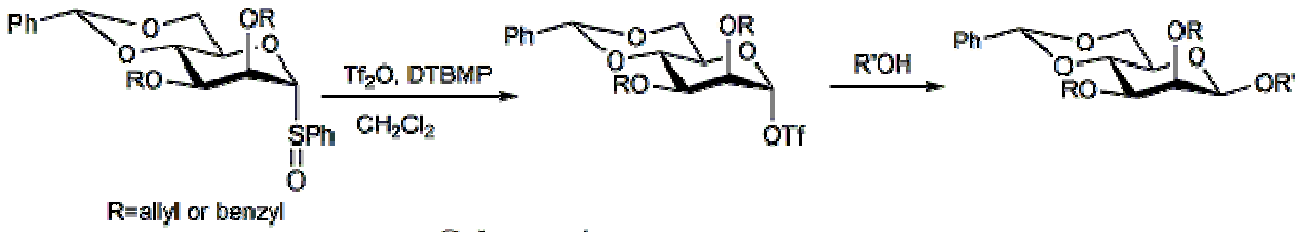
participation). To prevent the accumulation of the orthoester intermediates, the glycosylation condition should be slightly acidic.

Difficult linkages

It is somewhat more difficult to prepare 1, 2-cis-β-glycosidic linkages stereoselectively. Typically, when non-participating groups on O-2 position, 1, 2-cis-β-linkage can be achieved either by using the historically important halide ion methods, or by using 2-O-alkylated glycosyl donors, commonly thioglycosides or trichloroacetimidates, in nonpolar solvents.

In the early 1990s, it was still the case that the beta mannoside linkage was too challenging to be attempted by amateurs.

However, the method introduced by Crich (Scheme 4), with 4,6-benzylidene protection a prerequisite and anomeric alpha triflate a key intermediate leaves this problem essentially solved. The concurrently developed but rather more protracted intramolecular aglycon delivery (IAD) approach is a little-used but nevertheless stereospecific alternative.



Scheme 4

Chapter 3

Glycosidic Bond Formationf

Chemical glycosylation

A **chemical glycosylation reaction** involves the coupling of a glycosyl donor, to a glycosyl acceptor forming a glycoside. If both the donor and acceptor are sugars, then the product is an oligosaccharide. The reaction requires activation with a suitable activating reagent. The reactions often result in a mixture of products due to the creation of a new stereogenic centre at the anomeric position of the glycosyl donor. The formation of a glycosidic linkage allows for the synthesis of complex polysaccharides which may play important roles in biological processes and pathogenesis and therefore having synthetic analogs of these molecules allows for further studies with respect to their biological importance.

Terminology

The glycosylation reaction involves the coupling of a glycosyl donor and a glycosyl acceptor via initiation using an activator under suitable reaction conditions.

A glycosyl donor is a sugar with a suitable leaving group at the anomeric position. This group, under the reaction conditions, is activated and via the formation of an oxocarbenium is eliminated leaving an electrophilic anomeric carbon.

A glycosyl acceptor is a sugar with an unprotected nucleophilichydroxyl group which may attack the carbon of the oxocarbenium ion formed during the reaction and allow for the formation of the glycosidic bond.

An activator is commonly a Lewis acid which enables the leaving group at the anomeric position to leave and results in the formation of the oxocarbenium ion.

Neighbouring group participation

The stereochemical outcome of a glycosylation reaction may in certain cases be affected by the type of protecting group employed at position 2 of the glycosyl donor. A participating group, typically one with a carboxyl group present, will predominantly result in the formation of a β-glycoside. Whereas a non-participating group, a group usually without a carboxyl group, will often result in anα-glycoside.

Below it can be seen that having an acetyl protecting group at position 2 allows for the formation for an acetoxonium ion intermediate that blocks attack to the bottom face of the ring therefore allowing for the formation of the β-glycoside predominantly.

Protecting groups

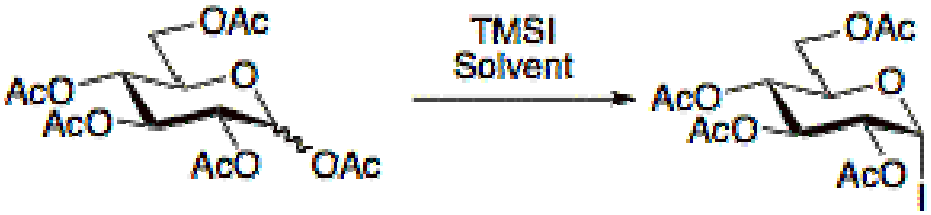
Different protecting groups on either the glycosyl donor or the glycosyl acceptor may affect the reactivity and yield of the glycosylation reaction. Typically,

electron-withdrawing groups such as acetyl or benzoyl groups are found to decrease the reactivity of the donor/acceptor and are therefore termed "disarming" groups. Electron-donating groupssuch as the benzyl group, are found to increase the reactivity of the donor/acceptor and are therefore called "arming" groups.

Current methods in glycoside synthesis

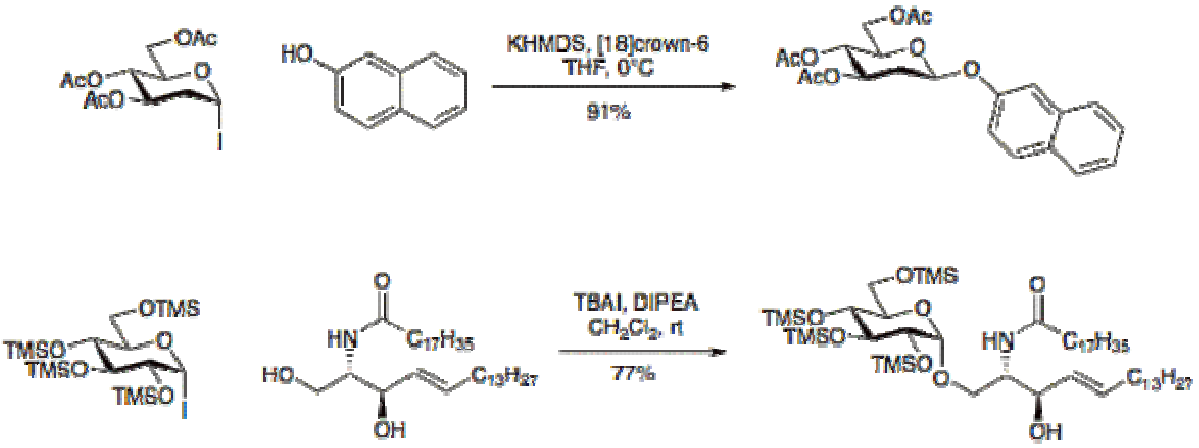
Glycosyl iodides

Glycosyl iodides were first introduced for use in glycosylation reactions in 1901 by Koenigs and Knorr although were often considered too reactive for synthetic use. Recently several research groups have shown these donors to have unique reactive properties and can differ from other glycosyl chlorides or bromides with respect to reaction time, efficiency, and stereochemistry. Glycosyl iodides may be made under a variety of conditions, one method of note is the reaction of a 1-*O*-acetylpyranoside with TMSI.



Iodide donors may typically be activated under basic conditions to give β-glycosides with good selectivity. The use of

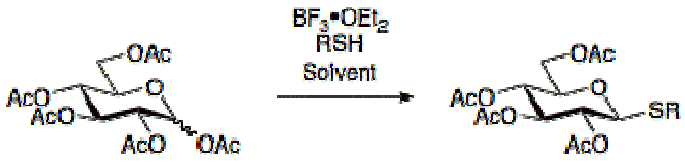
tetraalkylammonium iodide salts such as tetrabutylammonium iodide (TBAI) allows for in-situanomerization of the α-glycosyl halide to the β-glycosyl halide and provides the α-glycoside in good selectivity.



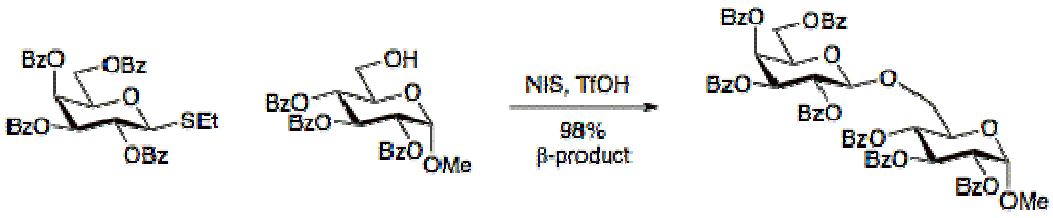
Thioglycosides

Thioglycosides were first reported in 1909 by Fischer and since then have been explored constantly allowing for the development of numerous protocols for their preparation. The advantage of using thioglycosides is their stability under a wide range of reaction conditions allowing for protecting group manipulations.

Additionally thioglycosides act as temporary protecting groups at the anomeric position allowing for thioglycosides to be useful as both glycosyl donors as well as glycosyl acceptors. Thioglycosides are usually prepared by reacting per-acetylated sugars with $BF_3 \bullet OEt_2$ and the appropriate thiol.



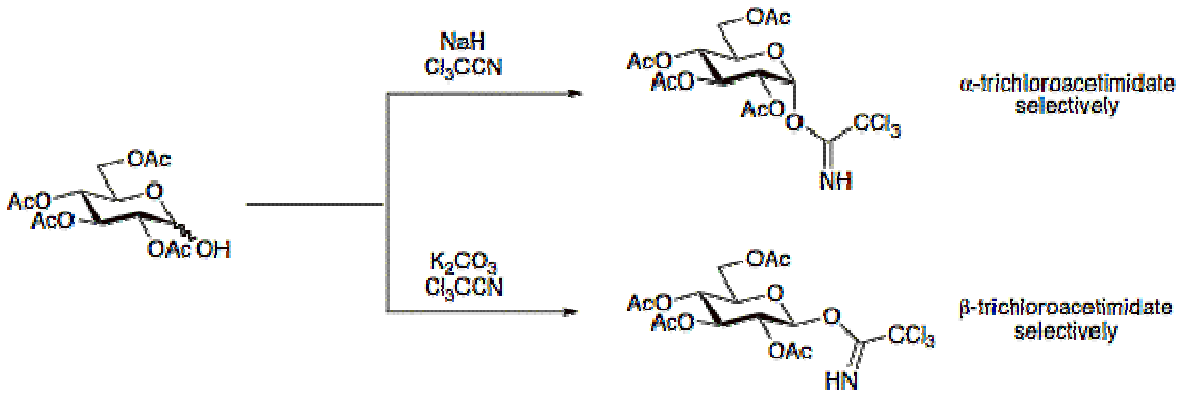
Thioglycosides used in glycosylation reactions as donors can be activated under a wide range of conditions, most notably using NIS/AgOTf.



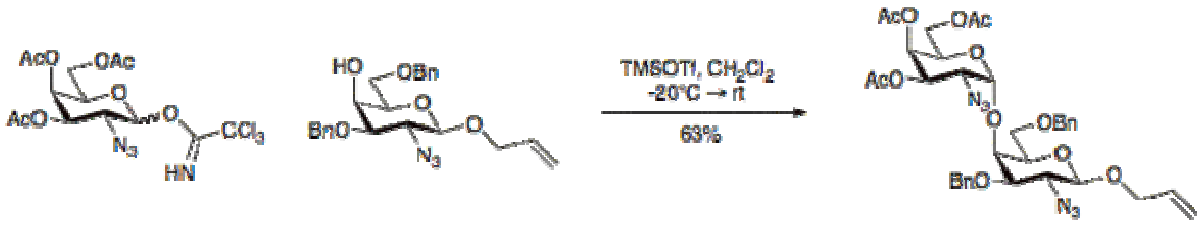
Trichloroacetimidates

Trichloroacetimidates were first introduced and explored by Schmidt in 1980 and since then have become very popular for glycoside synthesis.

The use of trichloroacetimidates provides many advantages including ease of formation, reactivity and stereochemical outcome. *O*-Glycosyl trichloroacetimidates are prepared via the addition of trichloroacetonitrile (Cl_3CCN) under basic conditions to a free anomeric hydroxyl group.



Typical activating groups for glycosylation reactions using trichloroacetimidates are $BF_3 \bullet OEt_2$ or TMSOTf.



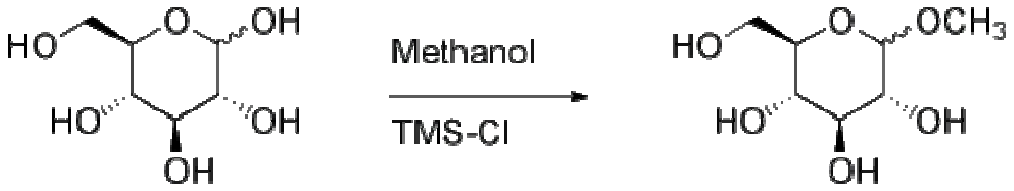
Column chromatographic purification of the reaction mixture can sometimes be challenging due to the trichloroacetamide by-product. This can, however, be overcome by washing the organic layer with 1 M NaOH solution in a separatory funnel prior to chromatography. Acetyl protecting groups were found to be stable during this procedure.

Fischer glycosidation

Fischer glycosidation (or Fischer glycosylation) refers to the formation of a **glycoside** by the reaction of an aldose or ketose with an alcohol in the presence of an acid catalyst. The reaction is named after the German chemist, Emil Fischer, winner of the Nobel Prize in chemistry, 1902, who developed this method between 1893 and 1895.

Commonly, the reaction is performed using a solution or suspension of the carbohydrate in the alcohol as the solvent. The carbohydrate is usually completely unprotected. The Fischer glycosidation reaction is an equilibrium process and can lead to a mixture of ring size isomers, and anomers, plus in some cases, small amounts of acyclic forms. With hexoses, short reactions times usually lead to furanose ring forms, and longer reaction times lead to pyranose forms. With long reaction times the most thermodynamically stable product will

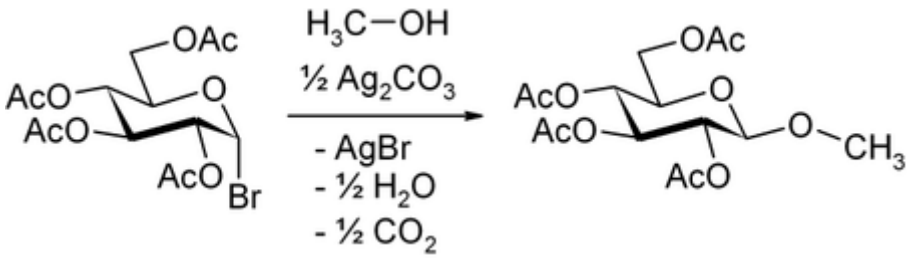
result which, owing to the anomeric effect, is usually the alpha anomer.



Fischer glycosidation of glucose to produce methyl glucoside using trimethylsilylchloride as catalyst.

Koenigs–Knorr reaction

The **Koenigs–Knorr reaction** in organic chemistry is the substitution reaction of a glycosylhalide with an alcohol to give a glycoside. It is one of the oldest glycosylation reactions. It is named after Wilhelm Koenigs (1851–1906), a student of von Bayer and fellow student with Hermann Emil Fischer, and Edward Knorr, a student of Koenigs.



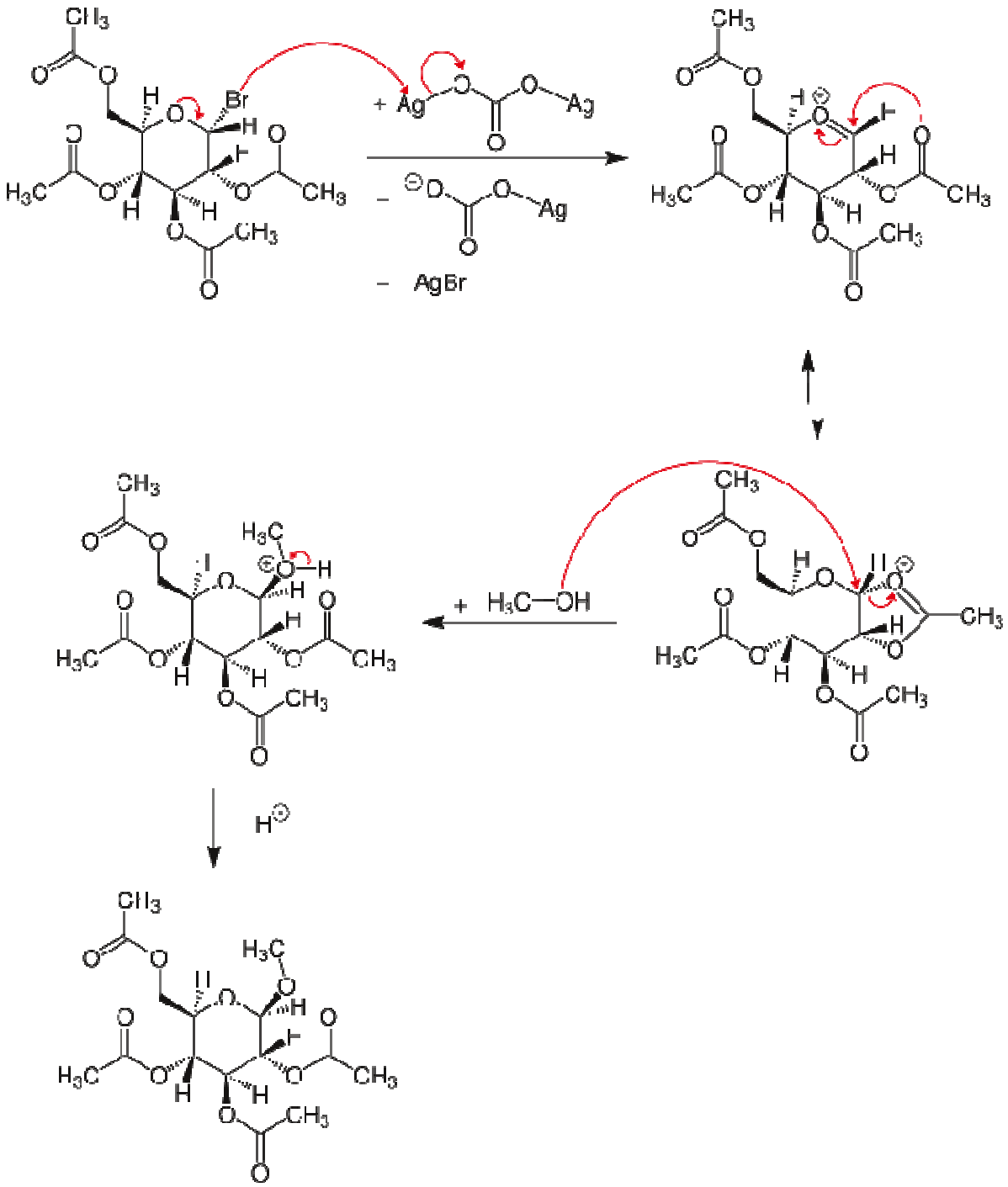
In its original form, Koenigs and Knorr treated *acetobromoglucose* with alcohols in the presence of silver carbonate. Shortly afterwards Fischer and Armstrong reported very similar findings.

In the above example, the stereochemical outcome is determined by the presence of the neighboring group at C2 that lends anchimeric assistance, resulting in the formation of

a 1,2-trans stereochemical arrangement. Esters (e.g. acetyl, benzoyl, pivalyl) generally provide good anchimeric assistance, whereas ethers (e.g. benzyl, methyl etc.) do not, leading to mixtures of stereoisomers.

Mechanism

In the first step of the mechanism, the glycosyl bromide reacts with silver carbonate upon elimination of silver bromide and the silver carbonate anion to the oxocarbenium ion. From this structure a dioxolaniumnium is formed, which is attacked by methanol via an SN_2 mechanism at the carbonyl carbon atom. This attack leads to the inversion. After deprotonation of the intermediate oxonium, the product glycoside is formed.



The reaction can also be applied to carbohydrates with other protecting groups. In the oligosaccharide synthesis in place of the methanol other carbohydrates are used, which have been modified with protective groups in such a way that only one hydroxyl group is accessible.

History

The method was later transferred by Emil Fischer and Burckhardt Helferich to other chloro-substituted purines and produced thus for the first time synthetic nucleosides. It was later improved and modified by numerous chemists.

Alternative reactions

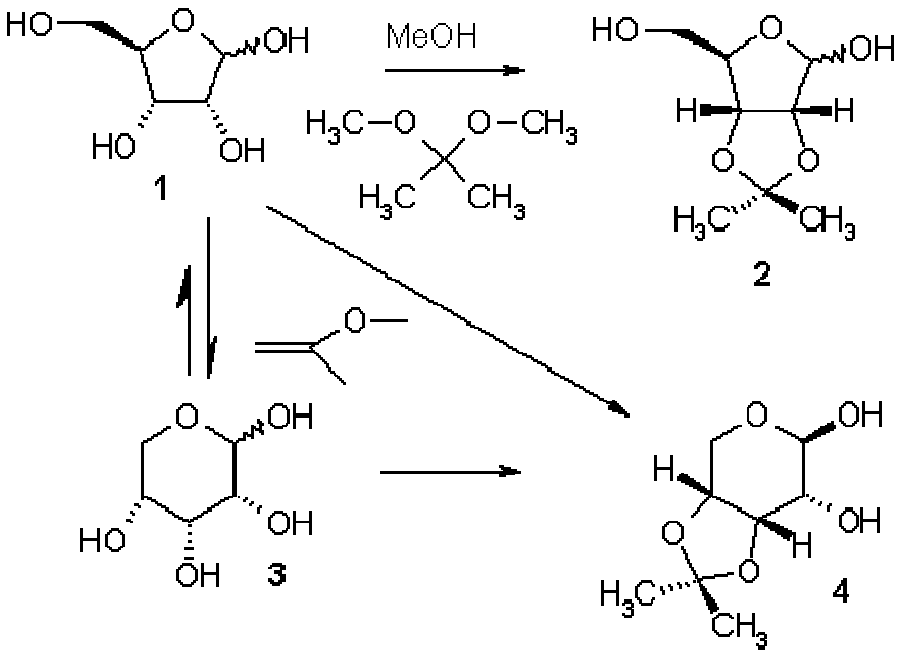
Generally, the Koenigs–Knorr reaction refers to the use of glycosyl chlorides, bromides and more recently iodides as glycosyl donors. The Koenigs–Knorr reaction can be performed with alternative promoters such as various heavy metal salts including mercuric bromide/mercuric oxide, mercuric cyanide and silver triflate. When mercury salts are used, the reaction is normally called the Helferich method. Other glycosidation methods are Fischer glycosidation, use of glycosyl acetates, thioglycosides, glycosyl trichloroacetimidates, glycosyl fluorides or n-pentenyl glycosides as glycosyl donors, or intramolecular aglycon delivery.

Chapter 4

Protecting Groups

Carbohydrate acetalisation

In carbohydrate chemistry**carbohydrate acetalisation** is an organic reaction and a very effective means of providing a protecting group. The example below depicts the acetalisation reaction of D-ribose**1**. With acetone or 2,2-dimethoxypropane as the acetalisation reagent the reaction is under thermodynamic reaction control and results in the pentose**2**. The latter reagent in itself is an acetal and therefore the reaction is actually a **cross-acetalisation**.



Kinetic reaction control results from 2-methoxypropene as the reagent. D-ribose in itself is a hemiacetal and in equilibrium with the pyranose**3**. In aqueous solution ribose is 75% pyranose and 25% furanose and a different acetal **4** is formed.

Selective acetalization of carbohydrate and formation of acetals possessing atypical properties is achieved by using arylsulfonyl acetals. An example of arylsulfonyl acetals as carbohydrate-protective groups are phenylsulfonylethylidene acetals. These acetals are resistant to the acid hydrolysis and can be deprotected easily by classical reductive conditions.

Trimethylsilyl

A **trimethylsilyl group** (abbreviated TMS) is a functional group in organic chemistry. This group consists of three methyl groups bonded to a silicon atom [–$Si(CH_3)_3$], which is in turn bonded to the rest of a molecule. This structural group is characterized by chemical inertness and a large molecular volume, which makes it useful in a number of applications.

A trimethylsilyl group bonded to a methyl group forms tetramethylsilane, which is abbreviated as TMS as well.

Compounds with trimethylsilyl groups are not normally found in nature. Chemists sometimes use a trimethylsilylating reagent to derivatize rather non-volatile compounds such as certain alcohols, phenols, or carboxylic acids by substituting a trimethylsilyl group for a hydrogen in the hydroxyl groups on the compounds. This way trimethylsiloxy groups [–O-$Si(CH_3)_3$] are formed on the molecule. A couple of examples of trimethylsilylating agents include trimethylsilyl chloride and bis(trimethylsilyl)acetamide. Trimethylsilyl groups on a molecule have a tendency to make it more volatile, often making the compounds more amenable to analysis by gas chromatography or mass spectrometry. An example of such trimethylsilylation is mentioned in the Brassicasterol article.

Such derivatizations are often done on a small scale in special vials. When attached to certain functional groups in a reactant molecule, trimethylsilyl groups may also be used as temporary protecting groups during chemical synthesis or some other chemical reactions.

In chromatography, derivitization of accessible silanol groups in a bonded stationary phase with trimethylsilyl groups is referred to as endcapping.

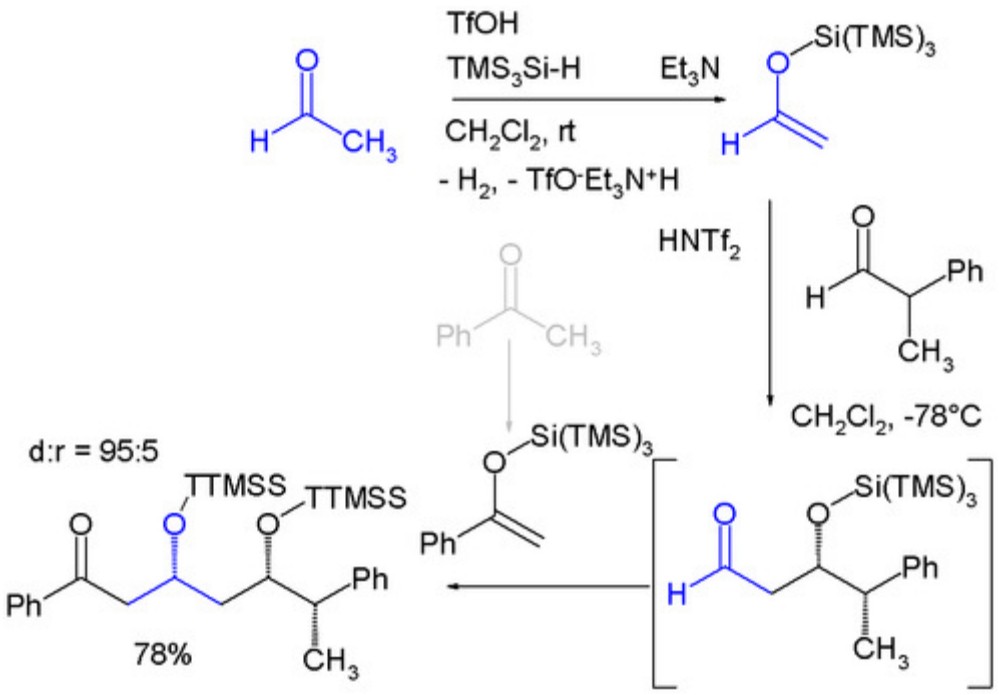
In an NMR spectrum, signals from atoms in trimethylsilyl groups in compounds will commonly have chemical shifts close to the tetramethylsilane reference peak at 0 ppm. Also compounds, such as high temperature silicone "stopcock" grease, which have polysiloxanes (often called silicones) in them will commonly show peaks from their methyl groups (attached to the silicon atoms) having NMR chemical shifts close to the tetramethylsilane standard peak, such as at 0.07 ppm in $CDCl_3$.

Otherwise very reactive molecules can be isolated when enveloped by bulky trimethylsilyl groups. This effect can be observed in tetrahedranes.

Super silyl groups

Related to trimethylsilyl groups are "super" silyl groups of which there exist two varieties: A silicon group connected to three trimethylsilyl groups makes a tri(trimethylsilyl)silyl group (TTMSS or TMS_3Si) and a silicon group connected to three tert-butyl groups. The TTMSS group was proposed in 1993 by Hans Bock. With a van der Waals volume of up to 7

cubic angstrom it surpasses the related TIPS group (around 2) and one potential application is its use as a temporary substituent promoting asymmetric induction for example in this diastereoselectiveone-pot reaction involving two sequential Mukaiyama aldol reactions:

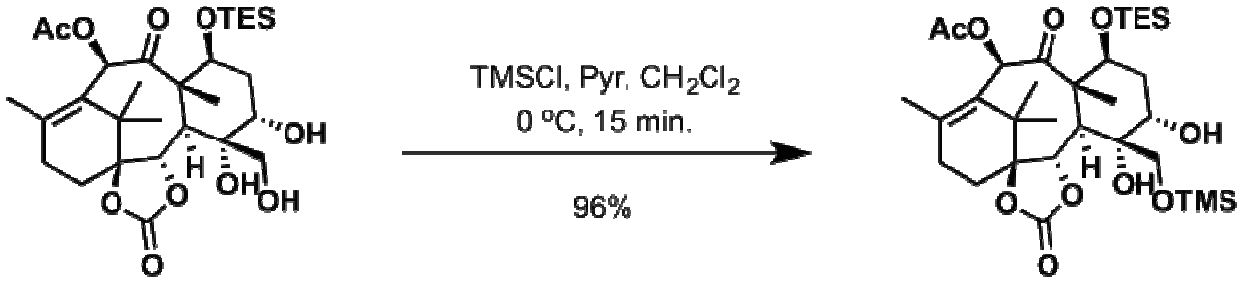


TTMSS can also stand for tris(trimethylsilyl)silane, which is comparable as a chemical reagent to tributyltin hydride without the associated toxicity concern of organotin and tributyltin compounds. The reagent is employed in radical reductions, hydrosilylation and consecutiveradical reactions.

Alcohol protection

In organic synthesis, TMS group is used as a protecting group for alcohols.

Most common protection methods



Trimethylsilyl chloride (TMSCl) or trimethylsilyl trifluoromethanesulfonate (TMSOTf) and base (i.e. pyridine, triethylamine, or 2,6-lutidine) in dichloromethane

TMSCl and lithium sulfide (Li_2S) in acetonitrile

Most common deprotection methods

TMS groups are susceptible to cleavage upon treatment with HF-based reagents

Tetrabutylammonium fluoride (Bu_4NF) in THF

Fluorosilicic acid (H_2SiF_6)

Treatment with HCl in THF/water solution

Benzyl group

In organic chemistry, **benzyl** is the substituent or molecular fragment possessing the structure $C_6H_5CH_2$–. Benzyl features a benzene ring attached to a CH_2 group.

Nomenclature

In IUPAC nomenclature the prefix **benzyl** refers to a $C_6H_5CH_2$ substituent, for example benzyl chloride or benzyl benzoate. Benzyl is not to be confused with phenyl with the formula C_6H_5. The term **benzylic**is used to describe the position of the first carbon bonded to a benzene or other aromatic ring. For example, $(C_6H_5)(CH_3)_2C$ is referred to as a "benzylic" carbocation. The benzyl free radical has the formula $C_6H_5CH_2$. The benzyl cation or phenylcarbenium ion is the carbocation with formula $C_6H_5CH_2$; the benzyl anion or phenylmethanide ion is the carbanion with the formula $C_6H_5CH_2$. None of these species can be formed in significant amounts in the solution phase under normal conditions, but they are useful referents for discussion of reaction mechanisms and may exist as reactive intermediates.

Abbreviations

The abbreviation "Bn" denotes benzyl. For example, benzyl alcoholcan be represented as BnOH. This abbreviation is not to be confused with "Bz", which is the abbreviation for the benzoyl group $C_6H_5C(O)–$, or the phenyl group C_6H_5, abbreviated "Ph". Confusingly, in old literature, "Bz" was also used for benzyl.

Functionalization at the benzylic position

In a few cases, these benzylic transformations occur under conditions suitable for lab synthesis. The Wohl-Ziegler reaction

will brominate a benzylic C–H bond: ($ArCHR_2 \rightarrow ArCBrR_2$). Any non-tertiary benzylic alkyl group will be oxidized to a carboxyl group by aqueous potassium permanganate ($KMnO_4$) or concentrated nitric acid (HNO_3): ($ArCHR_2 \rightarrow ArCOOH$). Finally, the complex of chromium trioxide and 3,5-dimethylpyrazole (CrO_3–dmpyz) will selectively oxidize a benzylic methylene group to a carbonyl: ($ArCH_2R \rightarrow ArC(O)R$).2-iodoxybenzoic acid in DMSO performs similarly.

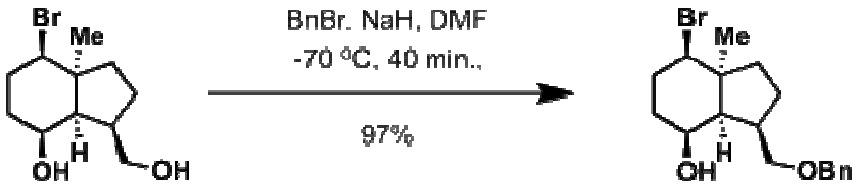
As a protecting group

Benzyl groups are occasionally employed as protecting groups in organic synthesis. Their installation and especially their removal require relatively harsh conditions, so benzyl is not typically preferred for protection.

Alcohol protection

Benzyl is commonly used in organic synthesis as a robust protecting group for alcohols and carboxylic acids.

Treatment of alcohol with a strong base such as powdered potassium hydroxide or sodium hydride and benzyl halide (BnCl or BnBr)



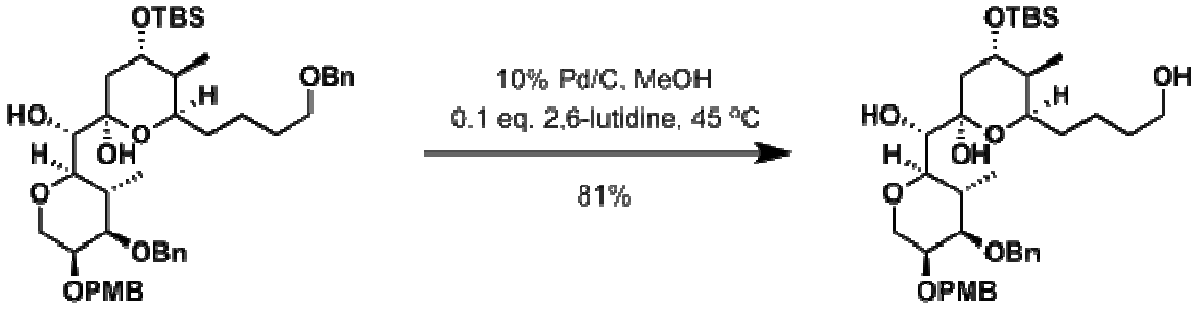
Monobenzylation of diols can be achieved using Ag_2O in dimethylformamide (DMF) at ambient to elevated temperatures

Primary alcohols can be selectively benzylated in presence of phenol functional groups using $Cu(acac)_2$

Deprotection methods

Benzyl ethers can be removed under *reductive conditions*, *oxidative conditions*, and the use of *Lewis Acids*.

Removed using hydrogenolysis



Single electron process with Na/NH_3 or Li/NH_3

Benzyl protecting groups can be removed using a wide range of oxidizing agents including:

CrO_3/acetic acid at ambient temperature

Ozone

N-Bromosuccinimide (NBS)

N-Iodosuccinimide (NIS)

Trimethylsilyl iodide (Me_3SiI) in dichloromethane at ambient temperature (selectivity can be achieved under specific conditions)

The *p*-methoxybenzyl protecting group

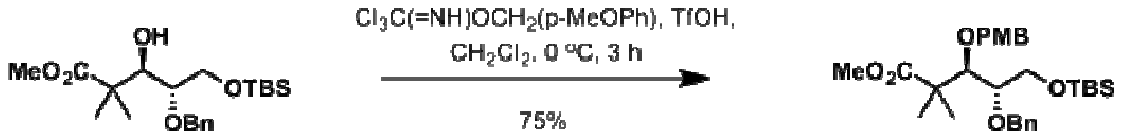
p-Methoxybenzyl (**PMB**) is used as a protecting group for alcohols in organic synthesis (4-Methoxybenzylthiol is used to protect thiols).

Strong base such as powdered potassium hydroxide or sodium hydride and *p*-methoxybenzyl halide (chloride or bromide)

4-methoxybenzyl-2,2,2-trichloroacetimidate can be used to install the PMB group in presence of:

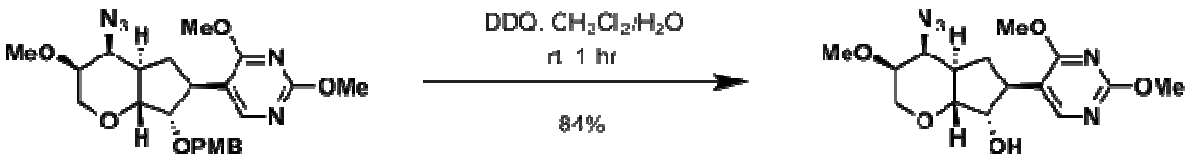
Scandium (III) triflate ($Sc(OTf)_3$) in toluene at 0 °C

Trifluoromethanesulfonic acid (TfOH) in dichloromethane at 0 °C



Deprotection methods

2,3-Dichloro-5,6-dicyano-*p*-benzoquinone (DDQ)

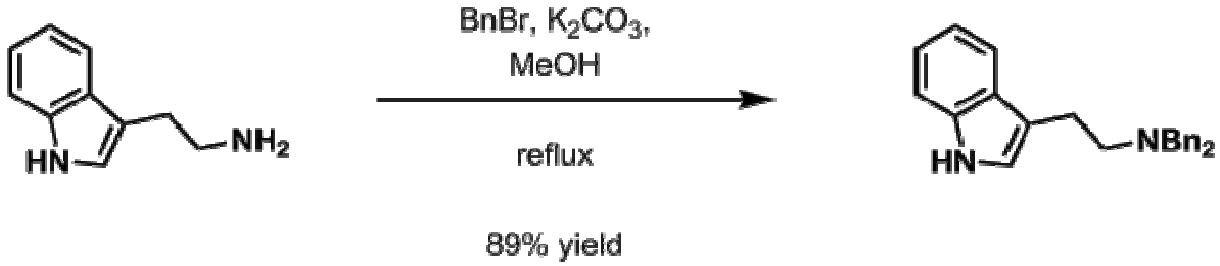


Conditions for deprotection of benzyl group are applicable for cleavage of the PMB protecting group

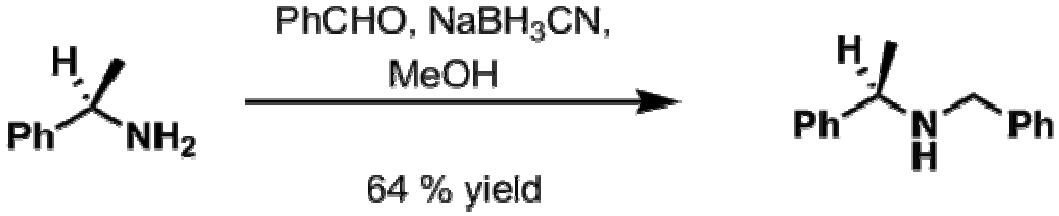
Amine protection

The benzyl group is occasionally used as a protecting group for amines in organic synthesis. Other methods exist.

Aqueous potassium carbonate and benzyl halide (BnCl, BnBr) in methanol



Benzaldehyde, 6 M HCl and $NaBH_3CN$ in methanol



Deprotection methods

Hydrogenation in the presence of the palladium catalyst

Chapter 5

Oligosaccharide

An **oligosaccharide** is a saccharidepolymer containing a small number (typically three to ten) of monosaccharides (simple sugars). Oligosaccharides can have many functions including cell recognition and cell binding. For example, glycolipids have an important role in the immune response.

They are normally present as glycans: oligosaccharide chains are linked to lipids or to compatible amino acid side chains in proteins, by *N*- or *O*-glycosidic bonds. *N*-Linked oligosaccharides are always pentasaccharides attached to asparagine via a beta linkage to the amine nitrogen of the side chain.

Alternately, *O*-linked oligosaccharides are generally attached to threonine or serine on the alcohol group of the side chain. Not all natural oligosaccharides occur as components of glycoproteins or glycolipids.

Some, such as the raffinose series, occur as storage or transport carbohydrates in plants. Others, such as maltodextrins or cellodextrins, result from the microbial breakdown of larger polysaccharides such as starch or cellulose.

Glycosylation

In biology, glycosylation is the process by which a carbohydrate is covalently attached to an organic molecule, creating structures such as glycoproteins and glycolipids.

N-Linked oligosaccharides

N-Linked glycosylation involves oligosaccharide attachment to asparagine via a beta linkage to the amine nitrogen of the side chain. The process of *N*-linked glycosylation occurs cotranslationally, or concurrently while the proteins is being translated. Since it is added cotranslationally, it is believed that *N*-linked glycosylation helps determine the folding of polypeptides due to the hydrophilic nature of sugars. All *N*-linked oligosaccharides are pentasaccharides: five monosaccharides long.

In *N*-glycosylation for eukaryotes, the oligosaccharide substrate is assembled right at the membrane of the endoplasmatic reticulum. For prokaryotes, this process occurs at the plasma membrane. In both cases, the acceptor substrate is an asparagine residue. The asparagine residue linked to an *N*-linked oligosaccharide usually occurs in the sequence Asn-X-Ser/Thr, where X can be any amino acid except for proline, although it is rare to see Asp, Glu, Leu, or Trp in this position.

O-Linked oligosaccharides

Oligosaccharides that participate in *O*-linked glycosylation are attached to threonine or serine on the hydroxyl group of the side chain.*O*-linked glycosylation occurs in the Golgi apparatus, where monosaccharide units are added to a complete polypeptide chain. Cell surface proteins and extracellular proteins are *O*-glycosylated. Glycosylation sites in *O*-linked oligosaccharides are determined by the secondary and tertiary structures of the polypeptide, which dictate where glycosyltransferases will add sugars.

Glycosylated biomolecules

Glycoproteins and glycolipids are by definition covalently bonded to carbohydrates. They are very abundant on the surface of the cell, and their interactions contribute to the overall stability of the cell.

Glycoproteins

Glycoproteins have distinct Oligosaccharide structures which have significant effects on many of their properties, affecting critical functions such as antigenicity, solubility, and resistance to proteases. Glycoproteins are relevant as cell-surface receptors, cell-adhesion molecules, immunoglobulins, and tumor antigens.

Glycolipids

Glycolipids are important for cell recognition, and are important for modulating the function of membrane proteins that act as receptors. Glycolipids are lipid molecules bound to oligosaccharides, generally present in the lipid bilayer. Additionally, they can serve as receptors for cellular recognition and cell signaling. The head of the oligosaccharide serves as a binding partner in receptor activity. The binding mechanisms of receptors to the oligosaccharides depends on the composition of the oligosaccharides that are exposed or presented above the surface of the membrane. There is great diversity in the binding mechanisms of glycolipids, which is what makes them such an important target for pathogens as a site for interaction and entrance. For example, the chaperone

activity of glycolipids has been studied for its relevance to HIV infection.

Functions

Cell recognition

All cells are coated in either glycoproteins or glycolipids, both of which help determine cell types.Lectins, or proteins that bind carbohydrates, can recognize specific oligosaccharides and provide useful information for cell recognition based on oligosaccharide binding.

An important example of oligosaccharide cell recognition is the role of glycolipids in determining blood types. The various blood types are distinguished by the glycan modification present on the surface of blood cells. These can be visualized using mass spectrometry. The oligosaccharides found on the A, B, and H antigen occur on the non-reducing ends of the oligosaccharide. The H antigen (which indicates an O blood type) serves as a precursor for the A and B antigen. Therefore, a person with A blood type will have the A antigen and H antigen present on the glycolipids of the red blood cell plasma membrane. A person with B blood type will have the B and H antigen present. A person with AB blood type will have A, B, and H antigens present. And finally, a person with O blood type will only have the H antigen present. This means all blood types have the H antigen, which explains why the O blood type is known as the "universal donor".

How do transport vesicles know the final destination of the protein that they are transporting?

Vesicles are directed by many ways, but the two main ways are:

The sorting signals encoded in the amino acid sequence of the proteins.
The Oligosaccharide attached to the protein.

The sorting signals are recognised by specific receptors that reside in the membranes or surface coats of budding vesicles, ensuring that the protein is transported to the appropriate destination.

Cell adhesion

Many cells produce specific carbohydrate-binding proteins known as lectins, which mediate cell adhesion with oligosaccharides.Selectins, a family of lectins, mediate certain cell–cell adhesion processes, including those of leukocytes to endothelial cells. In an immune response, endothelial cells can express certain selectins transiently in response to damage or injury to the cells. In response, a reciprocal selectin–oligosaccharide interaction will occur between the two molecules which allows the white blood cell to help eliminate the infection or damage. Protein-Carbohydrate bonding is often mediated by hydrogen bonding and van der Waals forces.

Dietary oligosaccharides

Fructo-oligosaccharides (FOS), which are found in many vegetables, are short chains of fructose molecules. They differ from fructans such as inulin, which as polysaccharides have a much higher degree of polymerization than FOS and other

Oligiosaccharides, but like inulin and other fructans, they are considered soluble dietary fibre. Galactooligosaccharides (GOS), which also occur naturally, consist of short chains of galactose molecules. Human milk is an example of this and contains oligosaccharides, known as human milk oligosaccharides (HMOs), which are derived from lactose. These oligosaccharides have biological function in the development of the gut flora of infants. Examples include lacto-N-tetraose, lacto-N-neotetraose, and lacto-N-fucopentaose. These compounds cannot be digested in the human small intestine, and instead pass through to the large intestine, where they promote the growth of *Bifidobacteria*, which are beneficial to gut health.

Mannan oligosaccharides (MOS) are widely used in animal feed to improve gastrointestinal health. They are normally obtained from the yeast cell walls of *Saccharomyces cerevisiae*. Mannan oligosaccharides differ from other oligosaccharides in that they are not fermentable and their primary mode of actions include agglutination of type-1 fimbria pathogens and immunomodulation

Sources

Oligosaccharides are a component of fibre from plant tissue. FOS and inulin are present in Jerusalem artichoke, burdock, chicory, leeks, onions, and asparagus. Inulin is a significant part of the daily diet of most of the world's population. FOS can also be synthesized by enzymes of the fungus *Aspergillus niger* acting on sucrose. GOS is naturally found in soybeans and can be synthesized from lactose. FOS, GOS, and inulin are also sold as nutritional supplements.

Fructooligosaccharide

Fructooligosaccharides (**FOS**) also sometimes called **oligofructose** or **oligofructan**, are oligosaccharidefructans, used as an alternative sweetener. FOS exhibits sweetness levels between 30 and 50 percent of sugar in commercially prepared syrups. It occurs naturally, and its commercial use emerged in the 1980s in response to consumer demand for healthier and calorie-reduced foods.

Chemistry

Two different classes of fructooligosaccharide (FOS) mixtures are produced commercially, based on inulin degradation or transfructosylation processes.

FOS can be produced by degradation of inulin, or polyfructose, a polymer of D-fructose residues linked by β(2→1) bonds with a terminal α(1→2) linked D-glucose. The degree of polymerization of inulin ranges from 10 to 60. Inulin can be degraded enzymatically or chemically to a mixture of oligosaccharides with the general structure Glu–Fru_n (abbrev. GF_n) and Fru_m (F_m), with n and m ranging from 1 to 7. This process also occurs to some extent in nature, and these oligosaccharides can be found in a large number of plants, especially in Jerusalem artichoke, chicory and the blue agave plant. The main components of commercial products are kestose (GF_2), nystose (GF_3), fructosylnystose (GF_4), bifurcose (GF_3), inulobiose (F_2), inulotriose (F_3), and inulotetraose (F_4).

The second class of FOS is prepared by the transfructosylation action of a β-fructosidase of *Aspergillus niger* or *Aspergillus* on sucrose. The resulting mixture has the general formula of GF_n, with n ranging from 1 to 5. Contrary to the inulin-derived FOS, not only is there β(1→2) binding but other linkages do occur, however, in limited numbers.

Because of the configuration of their glycosidic bonds, fructooligosaccharides resist hydrolysis by salivary and intestinal digestive enzymes. In the colon they are fermented by anaerobic bacteria.

In other words, they have a lower caloric value, while contributing to the dietary fiber fraction of the diet. Fructooligosaccharides are more soluble than inulins and are, therefore, sometimes used as an additive to yogurt and other (dairy) products. Fructooligosaccharides are used specially in combination with high-intensity artificial sweeteners, whose sweetness profile and aftertaste it improves.

Food sources

FOS is extracted from the blue agave plant as well as fruits and vegetables such as bananas, onions, chicory root, garlic, asparagus, jícama, and leeks.

Some grains and cereals, such as wheat and barley, also contain FOS. The Jerusalem artichoke and its relative yacón together with the blue agave plant have been found to have the highest concentrations of FOS of cultured plants.

Health benefits

FOS has been a popular sweetener in Japan and Korea for many years, even before 1990, when the Japanese government installed a "Functionalized Food Study Committee" of 22 experts to start to regulate "special nutrition foods or functional foods" that contain the categories of fortified foods (e.g., vitamin-fortified wheat flour), and is now becoming increasingly popular in Western cultures for its prebiotic effects. FOS serves as a substrate for microflora in the large intestine, increasing the overall gastrointestinal tract health. It has also been proposed as a supplement for treating yeast infections.

Several studies have found that FOS and inulin promote calcium absorption in both the animal and the human gut. The intestinal microflora in the lower gut can ferment FOS, which results in a reduced pH. Calcium is more soluble in acid, and, therefore, more of it comes out of food and is available to move from the gut into the bloodstream.

FOS can be considered a small dietary fibre with (like all types of fibre) low caloric value. The fermentation of FOS results in the production of gases and acids. The latter provide some energy to the body.

Side-effects

All inulin-type prebiotics, including FOS, are generally thought to stimulate the growth of *Bifidobacteria* species. Bifidobacteria are considered beneficial bacteria.

This effect has not been uniformly found in all studies, either for bifidobacteria or for other gut organisms. FOS are also fermented by numerous bacterial species in the intestine, including *Klebsiella*, *E. coli* and many *Clostridium* species, which can be pathogenic in the gut. These species are responsible mainly for the gas formation (hydrogen and carbon dioxide), which results after ingestion of FOS. Studies have shown that up to 20 grams/day is well tolerated.

Regulation

US FDA regulation

FOS is classified as generally recognized as safe (GRAS).

NZ FSANZ regulation

The Food Safety Authority warned parents of babies that a major European baby-formula brand made in New Zealand does not comply with local regulations (because it contains fructo-oligosaccharides (FOS)), and urged them to stop using it.

EU regulation

FOS use has been approved in the European Union; allowing addition of FOS in **restricted** amounts to baby formula (for babies up to 6 months) and follow-on formula (for babies between 6 and 12 months). Infant and follow-on formula products containing FOS have been sold in the EU since 1999.

Canadian regulations

FOS is currently not approved for use in baby formula.

2'-Fucosyllactose

2′-Fucosyllactose (2′-FL) is an oligosaccharide, more precisely, fucosylated, neutral trisaccharide composed of L-fucose, D-galactose, and D-glucose units. It is the most prevalent human milk oligosaccharide (HMO) naturally present in human breast milk, making up about 30% of all of HMOs. It was first discovered in the 1950s in human milk. The oligosaccharide's primary isolation technique has been in use since 1972.

Uses

As with other oligosaccharides, a widely regarded characteristic of 2'-fucosyllactose is its ability to protect against infectious diseases namely in preventing epithelial level adhesions of toxins and pathogens. The 2FL stimulates the growth of certain bifidobacteria and receptor analogons which lends to toxic and pathogenic protection, all this being most prevalent in infants. Among the pathogens that 2FL is known to protect against are *Campylobacter jejuni*, *Salmonella enterica* serotype Typhimurium, *Helicobacter pylori*, among others.

Galactooligosaccharide

Galacto-oligosaccharides (**GOS**), also known as **oligogalactosyllactose**, **oligogalactose**, **oligolactose** or **transgalactooligosaccharides** (**TOS**), belong to the group of prebiotics. Prebiotics are defined as non-digestible food ingredients that beneficially affect the host by stimulating the growth and/or activity of beneficial bacteria in the colon. GOS occurs in commercially available products such as food for both infants and adults.

Chemistry

The composition of the galacto-oligosaccharide fraction varies in chain length and type of linkage between the monomer units. Galacto-oligosaccharides are produced through the enzymatic conversion of lactose, a component of bovine milk.

A range of factors come into play when determining the yield, style, and type of GOS produced. These factors include:

enzyme source
enzyme dosage
feeding stock (lactose) concentration
origins of the lactose
process involved (e.g. free or immobilized enzyme)
reaction conditions impacting the processing situation
medium composition

GOS generally comprise a chain of galactose units that arise through consecutive transgalactosylation reactions, with a

terminal glucose unit. However, where a terminal galactose unit is indicated, hydrolysis of GOS formed at an earlier stage in the process has occurred. The degree of polymerization of GOS can vary quite markedly, ranging from 2 to 8 monomeric units, depending mainly on the type of the enzyme used and the conversion degree of lactose.

Health effects

Because of the configuration of their glycosidic bonds, galacto-oligosaccharides (GOS) largely resist hydrolysis by salivary and intestinal digestive enzymes. Galacto-oligosaccharides are classified as prebiotics, defined as non-digestible food ingredients that beneficially affect the host by stimulating the growth and/or activity of beneficial bacteria in the colon. The increased activity of these health-promoting bacteria results in a number of effects, both directly by the bacteria themselves or indirectly by the organic acids they produce via fermentation. Examples of effects are stimulation of immune functions, absorption of essential nutrients, production of the powerful anti-oxidant H2 gas and syntheses of certain vitamins.

Stimulating bacteria

Galacto-oligosaccharides are a substrate for bacteria, such as *Bifidobacteria* and lactobacilli. Studies with infants and adults have shown that foods or drinks enriched with galacto-oligosaccharides result in a significant increase in *Bifidobacteria*. These sugars can be found naturally in human milk, known as human milk oligosaccharides. Examples

include lacto-N-tetraose, lacto-N-neotetraose, and lacto-N-fucopentaose.

Immune response

Human gut microbiota play a key role in the intestinal immune system. Galacto-oligosaccharides support natural defenses of the human body via the gut microflora, indirectly by increasing the number of bacteria in the gut and inhibiting the binding or survival of *Escherichia coli*, *Salmonellatyphimurium* and *Clostridia*. GOS can positively influence the immune system indirectly through the production of antimicrobial substances, reducing the proliferation of pathogenic bacteria.

Constipation

Constipation is a potential problem, particularly among infants, elderly and pregnant women. In infants, formula feeding may be associated with constipation and hard stools. Galacto-oligosaccharides may improve stool frequency and relieve symptoms related to constipation.

Glycan

The terms **glycan** and polysaccharideare defined by IUPAC as synonyms meaning "compounds consisting of a large number of monosaccharides linked glycosidically". However, in practice the term glycan may also be used to refer to the carbohydrate portion of a glycoconjugate, such as a glycoprotein, glycolipid, or a proteoglycan, even if the carbohydrate is only an oligosaccharide. Glycans usually consist solely of O-glycosidic linkages of monosaccharides. For example, cellulose is a glycan

(or, to be more specific, a glucan) composed of β-1,4-linked D-glucose, and chitin is a glycan composed of β-1,4-linked *N*-acetyl-D-glucosamine. Glycans can be homo- or heteropolymers of monosaccharide residues, and can be linear or branched.

Glycans and proteins

Glycans can be found attached to proteins as in glycoproteins and proteoglycans. In general, they are found on the exterior surface of cells. O- and N-linked glycans are very common in eukaryotes but may also be found, although less commonly, in prokaryotes.

N-Linked glycans

Introduction

N-Linked glycans are attached in the endoplasmic reticulum to the nitrogen (N) in the side chain of asparagine (Asn) in the sequon. The sequon is an Asn-X-Ser or Asn-X-Thr sequence, where X is any amino acid except proline and the glycan may be composed of *N*-acetylgalactosamine, galactose, neuraminic acid, *N*-acetylglucosamine, fucose, mannose, and other monosaccharides.

Processing, modification, and diversity

Once transferred to the nascent peptide chain, N-linked glycans, in general, undergo extensive processing reactions, whereby the three glucose residues are removed, as well as several mannose residues, depending on the N-linked glycan in question. The removal of the glucose residues is dependent on

proper protein folding. These processing reactions occur in the Golgi apparatus. Modification reactions may involve the addition of a phosphate or acetyl group onto the sugars, or the addition of new sugars, such as neuraminic acid. Processing and modification of N-linked glycans within the Golgi does not follow a linear pathway. As a result, many different variations of N-linked glycan structure are possible, depending on enzyme activity in the Golgi.

Functions and importance

N-linked glycans are extremely important in proper protein folding in eukaryotic cells. Chaperone proteins in the endoplasmic reticulum, such as calnexin and calreticulin, bind to the three glucose residues present on the core N-linked glycan. These chaperone proteins then serve to aid in the folding of the protein that the glycan is attached to. Following proper folding, the three glucose residues are removed, and the glycan moves on to further processing reactions. If the protein fails to fold properly, the three glucose residues are reattached, allowing the protein to re-associate with the chaperones. This cycle may repeat several times until a protein reaches its proper conformation. If a protein repeatedly fails to properly fold, it is excreted from the endoplasmic reticulum and degraded by cytoplasmic proteases.

N-linked glycans also contribute to protein folding by steric effects. For example, cysteine residues in the peptide may be temporarily blocked from forming disulfide bonds with other cysteine residues, due to the size of a nearby glycan. Therefore, the presence of a N-linked glycan allows the cell to control which cysteine residues will form disulfide bonds.

N-linked glycans also play an important role in cell-cell interactions. For example, tumour cells make N-linked glycans that are abnormal. These are recognized by the CD337 receptor on Natural Killer cells as a sign that the cell in question is cancerous.

Within the immune system the N-linked glycans on an immune cell's surface will help dictate that migration pattern of the cell, e.g. immune cells that migrate to the skin have specific glycosylations that favor homing to that site. The glycosylation patterns on the various immunoglobulins including IgE, IgM, IgD, IgE, IgA, and IgG bestow them with unique effector functions by altering their affinities for Fc and other immune receptors. Glycans may also be involved in "self" and "non self" discrimination, which may be relevant to the pathophysiology of various autoimmune diseases; including rheumatoid arthritis and type 1 diabetes.

The targeting of degradative lysosomal enzymes is also accomplished by N-linked glycans. The modification of an N-linked glycan with a mannose-6-phosphate residue serves as a signal that the protein to which this glycan is attached should be moved to the lysosome. This recognition and trafficking of lysosomal enzymes by the presence of mannose-6-phosphate is accomplished by two proteins: CI-MPR (cation-independent mannose-6-phosphate receptor) and CD-MPR (cation-dependent mannose-6-phosphate receptor).

O-Linked glycans

Introduction

In eukaryotes, *O*-linked glycans are assembled one sugar at a time on a serine or threonine residue of a peptide chain in the Golgi apparatus. Unlike *N*-linked glycans, there is no known consensus sequence yet. However, the placement of a proline residue at either -1 or +3 relative to the serine or threonine is favourable for O-linked glycosylation.

Functions and importance

Sialyl lewis x is important in ABO blood antigen determination.

SLex is also important to proper immune response. P-selectin release from Weibel-Palade bodies, on blood vessel endothelial cells, can be induced by a number of factors. One such factor is the response of the endothelial cell to certain bacterial molecules, such as peptidoglycan. P-selectin binds to the SLex structure that is present on neutrophils in the bloodstream and helps to mediate the extravasation of these cells into the surrounding tissue during infection. *O*-linked glycans, in particular mucin, have been found to be important in developing normal intestinal microflora. Certain strains of intestinal bacteria bind specifically to mucin, allowing them to colonize the intestine.

Examples of *O*-linked glycoproteins are:

Glycophorin, a protein in erythrocytecell membranes
Mucin, a protein in saliva involved in formation of dental plaque
Notch, a transmembrane receptor involved in development and cell fate decisions
Thrombospondin
Factor VII

Factor IX
Urinary type plasminogen activator

Glycosaminoglycans

Another type of cellular glycan is the glycosaminoglycans (GAGs). These comprise 2-aminosugars linked in an alternating fashion with uronic acids, and include polymers such as heparin, heparan sulfate, chondroitin, keratan and dermatan. Some glycosaminoglycans, such as heparan sulfate, are found attached to the cell surface, where they are linked through a tetrasacharide linker via a xylosyl residue to a protein (forming a glycoprotein or proteoglycan).

Glycoscience

A 2012 report from the U.S. National Research Council calls for a new focus on glycoscience, a field that explores the structures and functions of glycans and promises great advances in areas as diverse as medicine, energy generation, and materials science. Until now, glycans have received little attention from the research community due to a lack of tools to probe their often complex structures and properties. The report presents a roadmap for transforming glycoscience from a field dominated by specialists to a widely studied and integrated discipline.

Glycan-protein interactions

Glycan-Protein interactions represent a class of biomolecular interactions that occur between free or protein-bound glycans

and their cognate binding partners. Intramolecular glycan-protein (protein-glycan) interactions occur between glycans and proteins that they are covalently attached to. Together with protein-protein interactions, they form a mechanistic basis for many essential cell processes, especially for cell-cell interactions and host-cell interactions. For instance, SARS-CoV-2, the causative agent of COVID-19, employs its extensively glycosylated spike (S) protein to bind to the ACE2 receptor, allowing it to enter host cells. The spike protein is a trimeric structure, with each subunit containing 22 N-glycosylation sites, making it an attractive target for vaccine search.

Glycans, a generic name for monosaccharides and oligosaccharides, represent one of the major post-translational modification of proteins contributing to the enormous biological complexity of life. Indeed, three different hexoses could theoretically produce from 1056 to 27,648 unique trisaccharides in contrast to only 6 peptides or oligonucleotides formed from 3 amino acids or 3 nucleotides respectively. In contrast to template-drivenprotein biosynthesis, the "language" of glycosylation is still unknown, making glycobiology a hot topic of current research given their prevalence in living organisms.

The study of glycan-protein interactions provides insight into the mechanisms of cell-signaling and allows to create better-diagnosing tools for many diseases, including cancer. Indeed, there are no known types of cancer that do not involve erratic patterns of protein glycosylation.

Carbohydrate-binding partners

There are many proteins capable of binding to glycans, including lectins, antibodies, microbial adhesins, viral agglutinins, etc.

Lectins

Lectins is a generic name for proteins with carbohydrate-recognizing domains (CRD). Although it became almost synonymous with glycan-binding proteins, it does not include antibodies which also belong to the class.

Lectins found in plants and fungi cells have been extensively used in research as a tool to detect, purify, and analyze glycans. However, useful lectins usually have sub-optimal specificities. For instance, *Ulex europaeus* agglutinin-1 (UEA-1), a plant-extracted lectin capable of binding to human blood type Oantigen, can also bind to unrelated glycans such as 2'-fucosyllactose, GalNAcα1-4(Fucα1-2)Galβ1-4GlcNAc, and Lewis-Y antigen.

Antibodies

Although antibodies exhibit nanomolar affinities toward protein antigens, the specificity against glycans is very limited. In fact, available antibodies may bind only <4% of the 7000 mammalian glycan antigens; moreover, most of those antibodies have low affinity and exhibit cross-reactivity.

Carbohydrate-binding partners

There are many proteins capable of binding to glycans, including lectins, antibodies, microbial adhesins, viral agglutinins, etc.

Lectins

Lectins is a generic name for proteins with carbohydrate-recognizing domains (CRD). Although it became almost synonymous with glycan-binding proteins, it does not include antibodies which also belong to the class.

Lectins found in plants and fungi cells have been extensively used in research as a tool to detect, purify, and analyze glycans. However, useful lectins usually have sub-optimal specificities. For instance, *Ulex europaeus* agglutinin-1 (UEA-1), a plant-extracted lectin capable of binding to human blood type Oantigen, can also bind to unrelated glycans such as 2'-fucosyllactose, GalNAcα1-4(Fucα1-2)Galβ1-4GlcNAc, and Lewis-Y antigen.

Antibodies

Although antibodies exhibit nanomolar affinities toward protein antigens, the specificity against glycans is very limited. In fact, available antibodies may bind only <4% of the 7000 mammalian glycan antigens; moreover, most of those antibodies have low affinity and exhibit cross-reactivity.

A close look at the crystal structure of *VLRB.aGPA.23* reveals a tryptophan residue at position 187 right over the carbohydrate binding pocket.

Multivalency in structure

Many glycan binding proteins (GBPs) are oligomeric and typically contain multiple sites for glycan binding (also called carbohydrate-recognition domains).

The ability to form multivalent protein-ligand interactions significantly enhances the strength of binding: while values for individual CRD-glycan interactions may be in the mM range, the overall affinity of GBP towards glycans may reach nanomolar or even picomolar ranges. The overall strength of interactions is described as *avidity* (in contrast with an *affinity* which describes single equilibrium). Sometimes the *avidity*is also called an *apparent* to emphasize the non-equilibrium nature of the interaction.

Common oligomerization structures of lectinsare shown below. For example, galectinsare usually observed as dimers, while intelectins form trimers and pentraxins assemble into pentamers.

Larger structures, like hexameric Reg proteins, may assemble into membrane penetrating pores. Collectins may form even more bizarre complexes: bouquets of trimers or even cruciform-like structures (e.g. in SP-D).

Current Research

Given the importance of glycan-protein interactions, there is an ongoing research dedicated to the a) creation of new tools to detect glycan-protein interactions and b) using those tools to decipher the so-called sugar code.

Glycan Arrays

One of the most widely used tools for probing glycan-protein interactions is glycan arrays. A glycan array usually is an NHS- or epoxy-activated glass slides on which various glycanswere printed using robotic printing. These commercially available arrays may contain up to 600 different glycans, specificity of which has been extensively studied.

Glycan-protein interactions may be detected by testing proteins of interest (or libraries of those) that bear fluorescent tags. The structure of the glycan-binding protein may be deciphered by several analytical methods based on mass-spectrometry, including MALDI-MS, LC-MS, tandem MS-MS, and/or 2D NMR.

Bioinformatics driven research

Computational methods have been applied to search for parameters (e.g. residue propensity, hydrophobicity, planarity) that could distinguish glycan-binding proteins from other surface patches.

For example, a model trained on 19 non-homologous carbohydrate binding structures was able to predict

carbohydrate-binding domains (CRDs) with an accuracy of 65% for non-enzymatic structures and 87% for enzymatic ones. Further studies have employed calculations of Van der Waals energies of protein-probe interactions and amino acid propensities to identify CRDs with 98% specificity at 73% sensitivity.More recent methods can predict CRDs even from protein sequences, by comparing the sequence with those for which structures are already known.

Sugar code

In contrast with protein studies, where a primary protein structure is unambiguously defined by the sequence of nucleotides (the genetic code), the glycobiology still cannot explain how a certain "message" is encoded using carbohydrates or how it is "read" and "translated" by other biological entities.

An interdisciplinary effort, combining chemistry, biology, and biochemistry, studies glycan-protein interactions to see how different sequences of carbohydrates initiate different cellular responses.

Glycosyl

A **glycosyl group** is a univalentfree radical or substituent structure obtained by removing the hemiacetalhydroxyl group from the cyclic form of a monosaccharide and, by extension, of a lower oligosaccharide. Glycosyl also reacts with inorganic acids, such as phosphoric acid, forming an ester such as glucose 1-phosphate.

Examples

In cellulose, glycosyl groups link together 1,4-beta-D-glucosyl units to form chains of $(1,4\text{-beta-D-glucosyl})_n$. Other examples include ribityl in 6,7-Dimethyl-8-ribityllumazine, and glycosylamines.

Alternative substituent groups

Instead of the hemiacetal hydroxyl group, a *hydrogen* atom can be removed to form a substituent, for example the hydrogen from the C3 hydroxyl of a glucose molecule. Then the substituent is called D-glucopyranos-3-*O*-yl as it appears in the name of the drug Mifamurtide.

Recent detection of the Au in living organism was possible through the use of C-glycosyl pyrene, where it's permeability through cell membrane and fluorescence properties were used to detect Au.

Human milk oligosaccharide

Human milk oligosaccharides, also known as **human milk glycans**, are short polymers of simple sugars that can be found in high concentrations exclusively in human breast milk.

Occurrence

Human milk oligosaccharides (HMOs) form the third most abundant solid component (dissolved or emulsified or suspended in water) of human milk, after lactose and fat. HMOs are present in a concentration of 0.35–0.88 ounces (9.9–24.9 g)/L. Approximately 200 structurally different human milk oligosaccharides are known. The composition of human milk oligosaccharides in breast milk is individual to each mother and varies over the period of lactation. The dominant oligosaccharide in 80% of all women is 2′-fucosyllactose, which is present in human breast milk at a concentration of approximately 2.5 g/L. Other oligosacchardies include lacto-*N*-tetraose, lacto-*N*-neotetraose, and lacto-*N*-fucopentaose.

Characteristics

In contrast to the other components of breast milk that are absorbed by the infant through breastfeeding, HMOs are indigestible for the nursing child. However, they have a prebiotic effect and serve as food for intestinal bacteria, especially bifidobacteria. The dominance of these intestinal bacteria in the gut reduces the colonization with pathogenicbacteria (probiosis) and thereby promotes a healthy intestinal microbiota and reduces the risk of dangerous intestinal infections. Recent studies suggest that HMOs significantly lower the risk of viral and bacterial infections and thus diminish the chance of diarrhoea and respiratory diseases.

This protective function of the HMOs is activated when in contact with specific pathogens, such as certain bacteria or viruses. These have the ability to bind themselves to the glycan receptors (receptors for long chains of connected sugar molecules on the surface of human cells) located on the surface of the intestinal cells and can thereby infect the cells of the intestinal mucosa. Researchers have discovered that HMOs mimic these glycan receptors so the pathogens bind themselves to the HMOs rather than the intestinal cells. This reduces the risk of an infection with a pathogen. In addition to this, HMOs seem to influence the reaction of specific cells of the immune system in a way that reduces inflammatory responses. It is also suspected that HMOs reduce the risk of premature infants becoming infected with the potentially life-threatening disease necrotizing enterocolitis (NEC).

Some of the metabolites directly affect the nervous system or the brain and can sometimes influence the development and behavior of children in the long term. There are studies that indicate certain HMOs supply the child with sialic acid residues. Sialic acid is an essential nutrient for the development of the child's brain and mental abilities.

HMOs are used as supplements in baby food to provide babies that are not being breastfed with this important component of human milk.

Evolution

In experiments designed to test the suitability of HMOs as a prebiotic source of carbon for intestinal bacteria it was discovered that they are highly selective for a commensal

bacteria known as *Bifidobacteria longum* biovar *infantis*. The presence of genes unique to *B. infantis*, including co-regulated glycosidases, and its efficiency at using HMOs as a carbon source may imply a co-evolution of HMOs and the genetic capability of select bacteria to utilize them.

Enzymatic synthesis and large-scale production

Enzymatic synthesis of HMOs through transgalactosylation is an efficient way for the large-scale production. Various donors, including *p*-nitrophenyl-β-galactopyranoside, uridine diphosphate galactose and lactose, can be used in transgalactosylation. In particular, lactose may act as either a donor or an acceptor in a variety of enzymatic reactions and is available in large quantities from the whey produced as a co-processing product from cheese production. There is a lack of published data, however, describing the large-scale production of such galactooligosaccharides.

Idraparinux

Idraparinux sodium is an anticoagulant medication in development by Sanofi-Aventis.

It has a similar chemical structure and the same method of action as fondaparinux, but with an elimination half-life about five to six times longer (an increase from fondaparinux's 17 hours to approximately 80 hours), which means that the drug should only need to be injected once a week.Supriya Dey et al

recently reported shortest chemical synthesis of Idraparinux for the large scale production[8].

Sanofi discontinued the development of idraparinux sodium.

Negative clinical trial

A phase III trial of idraparinux sodium for stroke prevention in patients with AF (AMADEUS) was halted prematurely due to excessive clinically relevant and intracranial bleeding. Bleedings were particularly increased in elderly patients and those with renal impairment. Sanofi discontinued the development of idraparinux sodium in favour of a biotinylated formulation of the drug called idrabiotaparinux sodium.

Method of action

Idraparinux selectively blocks coagulation factor Xa.

See Heparin: Mechanism of anticoagulant action for a comparison of the mechanism of heparin, low-molecular-weight heparins, fondaparinux and idraparinux.

Idrabiotaparinux

Idrabiotaparinux sodium is also administered once-weekly. It has the same pentasaccharidic structure as idraparinux sodium, but with biotin attached, which allows its neutralisation with avidin, an egg-derived protein with low antigenicity. Sanofi conducted three phase III trials of idrabiotaparinux sodium between 2006 and 2008 in

approximately 13,550 patients. In one phase III trial (the BOREALIS-AF study), idrabiotaparinux sodium was non-inferior to warfarin in preventing the recurrent venous thromboembolism at three months in patients with pulmonary embolism and the incidence of clinically relevant bleeding was lower in the idrabiotaparinux sodium arm. The drug was expected to be filed for stroke prevention in patients with atrial fibrillation in 2011. The study was prematurely terminated, and the reason remains unclear. However, Sanofi has since discontinued its development. The company announced in May 2011 that the drug is available for licensing. A systematic review found that until now there is not sufficient evidence to clarify whether idraparinux or idrabiotaparinux are as effective and safe as the standard warfarin treatment for venous thromboembolism prevention. Idraparinux or idrabiotaparinux decreased major bleeding rate significantly but had a trend to increase the all-cause mortality compared with warfarin.

Isomaltooligosaccharide

Isomaltooligosaccharide (**IMO**), more commonly known as **isomalto-oligosaccharide**, is a mixture of short-chain carbohydrates which may have a digestion-resistant property. IMO is found naturally in some foods, as well as being manufactured commercially. Although isomaltose is found in some foods, such as honey, as a disaccharide, it behaves like all other disaccharides and is easily digested. Because of a confusion over nomenclature, such disaccharides have frequently been described as IMO, however to truly be called an "oligosaccharide" the molecules must have a degree of polymerization (DP) of three or more. Historically, the best

documented source of IMO was found in sourdough breads wherein proper oligosaccharides are produced. IMO is currently produced by two distinctly different methods. One is based on a conversion of starch using enzymes. The raw material used for manufacturing IMO is starch, which is enzymatically converted into a mixture of isomaltooligosaccharides. However, IMOs produced via this method result in a very high proportion of isomaltose disaccharides (approximately 50%) and a majority of panose (IMO DP3) as the end product. These IMO preparations have been shownto be highly digestible and generally do not exhibit much if any, digestion resistance. The second method uses bacterial fermentation and bio-conversion of sugar and starch to create an IMO variant maltosyl-isomaltooligosaccharide (MIMO). Typically the end product ranges in molecular complexity from DP3 (<10%) to DP9 with an average molecular weight near DP5. MIMO preparations exhibit very good digestion resistance and also intestinal microbiota selectivity.

Chemistry

The term "oligosaccharide" encompasses carbohydrates that are larger than simple di- or tri-saccharides, but smaller than polysaccharides (greater than 10 units). Isomalto-oligosaccharides (IMO) are glucose oligomers with α-D-(1,6)-linkages, including isomaltose, panose, isomaltotriose, isomaltotetraose, isomaltopentaose, nigerose, kojibiose, and higher branched oligosaccharides. Depending on production method, the structure of the IMO molecules can vary significantly. While human intestinal enzymes readily digest α(1,4)-glycosidic bonds, longer change IMO (e.g. >= DP4) with

α(1,6)-linkages are not easily hydrolyzed and exhibit a digestion-resistant property. Therefore, some IMO preparations are only partially digested in the upper gastrointestinal tract.

Isomalto-oligosaccharides are a normal part of the human diet and occur naturally in fermented foods, such as fermented sourdough breads and kimchi. The disaccharide isomaltose is also present in rice miso, soy sauce, and sake. Isomaltose, one of the α(1,6)-linked disaccharide components of IMO, has been identified as a natural constituent of honey and although chemically related, it is not and IMO . IMO is a sweet-tasting, high-density syrup which could be spray-dried into powder form.

Manufacturing

For manufacturing IMO on a commercial scale, food industries use starch processed from cereal crops like wheat, barley, pulses (peas, beans, lentils), oats, tapioca, rice, potato and others. This variety in sources could benefit consumers who have allergies or hypersensitivity to certain cereal crops. The manufacturing process controls the degree of polymerization (dp) and the α(1,6)-linkages to ensure a consistent quality of IMO from different starch sources. The starch is first converted, by means of simple enzymatic hydrolysis, into high maltose syrup with di-, tri and oligosaccharides (2, 3 or more glucose units) having α(1,4)-glycosidic linkages which are readily digestible in the human intestine. These α(1,4)-glycosidic linkages are further converted into digestion-resistant α(1,6)-glycosidic linkages, creating "iso" linkages between glucose moieties and forming Isomalto-oligosaccharide

(IMO). The majority of oligosaccharides found in IMO consist of three to six monosaccharide (glucose) units linked together. However, disaccharides, as well as longer polysaccharides (up to nine glucose units), are also present. The disaccharide fraction of IMO consists mainly of α(1,6)-linked isomaltose, while maltotriose, panose, and isomaltotriose make up the trisaccharide fraction. A mixture of isomaltotetraose, isomaltopentaose, maltohexaose, maltoheptaose, and small amounts of oligomers with 8 or more degrees of polymerization, comprise the remaining oligomers in IMO. Longer oligomers do not have 100% α(1,6)-linkages; the ratio of α(1,4)- to α(1,6)-linkages is variable.

Health claims for oligosaccharides

Health claims for the various classes of oligosaccharides have been investigated by the European Food Safety Authority (EFSA) and found to be insufficiently substantiated. Therefore, health claims for oligosaccharides and prebiotics are prohibited in the European Union.

Health benefits

IMO is a multifunctional molecule which exerts positive effects on human digestive health; it acts as a prebiotic, decreases flatulence [this is true for the MIMO variant, however, the literature only indicated that the enzymatically produced IMO "induce the least amount of gas compared to other prebiotics, which should not be confused with reducing flatulence], has a low glycemic index [this is not accurate for the enzymatically produced IMO. In fact, it has been shown that these IMO

preparations behave similarly to glucose syrup with respect to blood glucose absorption], and prevents dental caries in animals.

Prebiotics are defined as "non-digestible food ingredients that may beneficially affect the host by selectively stimulating the growth and/or activity of a limited number of bacteria in the colon". Oligosaccharides that are not digested and absorbed in the small intestine, pass through to the colon where they are fermented by Bifidobacteria, thus enhancing the proliferation of the bacteria. In this respect, fermentable oligosaccharides may be consideredprebiotics. The oligosaccharides in IMO mixtures are, at least partially, fermented by bacteria in the colon and may, therefore, stimulate the growth of bacterial subpopulations.

Short chain oligosaccharides which confer prebiotic properties also produce short-chain fatty acids (like acetate, propionate and butyrate) as end-products of fermentation. These molecules decrease the intra-luminal pH, directly inhibiting the growth and activity of harmful micro-organisms (enteropathogens). This stimulates the growth of Bifidobacteria, which compete with the enteropathogens for nutrients and epithelial adhesion sites. The beneficial effects of IMO have been found in infants, children, and the elderly.

Dental cariesis caused by the formation of insoluble glucan (plaque) on the surface of teeth, and the production of acids by bacteria in the plaque. These acids attack the hard tissues of the teeth. Studies with animal models showed that IMO, in place of sucrose, reduces the amount of plaque formed and

also reduces the amount of enamel-attacking acids formed. Therefore, IMO acts as an anti-caries agent.

The reported Glycemic Index (GI) for IMO is 34.66±7.65 (on a scale of 1–100) which represents a low GI. Consumption of IMO effectively improved bowel movements, stool output and microbial fermentation in the colon without any adverse effects in elderly people.

The American Association of Cereal Chemists (AACC) defines soluble fiber as "the edible parts of plants or similar carbohydrates resistant to digestion and absorption in the human small intestine with complete or partial fermentation in the large intestine". Dietary fiber consists of many plant components including oligosaccharides. For a dietary substrate to be classified as a fiber, it must be resistant to digestion and absorption in upper GI tract, and cause a bulking effect in defecation. IMO is considered a dietary fiber for the following reasons: it consists of glucose units linked together (mostly) by digestion-resistant linkages; it has a prebiotic effect; it retains moisture, producing a bulking effect and helping to move the stool forward.

Usage

IMO is finding global acceptance by food manufacturers for use in a wide range of food products, especially beverages and snack/nutrition bars. In the United States, IMO is used mostly as a source of dietary fiber. However, IMO is also used as a low calorie sweetener in a variety of foods like bakery and cereal products. Since IMO is about 50% as sweet as sucrose (sugar), it cannot replace sugar in a one-to-one ratio. However, IMO

has few side effects compared to other oligosaccharides of the same class.Therefore this carbohydrate molecule is receiving growing attention by food manufacturers across North America, as well as in Europe.

Side-effects

Generally, all digestion-resistant oligosaccharides, including IMO, have adverse side effects when consumed in amounts greater than permissible levels. The maximum permissible dose of IMO is 1.5 g/kg body weight, which is higher than for any other sugar substitute. However, the U.S. Food and Drug Administration (FDA) has recommended a maximum consumption of 30 g/day for IMO. Higher dosages (greater than 40 g/day), can cause gastrointestinal symptoms like flatulence, bloating, soft stool or diarrhea.

Regulatory information

IMO and other oligosaccharides have long been approved in China and Japan. In Japan, IMO is on the list of Foods for Specified Health Use (FOSHU) for more than 10 years. In 2002, over 50% of the FOSHU foods in Japan incorporated oligosaccharides as the functional component. The list includes many types of foods: soft drinks and other beverages, frozen yogurt, confectionery products, sweeteners, cookies, coffee drink mixes, bread, tofu, chocolate, and soup mixes. IMO has been imported into the United States for the last few years but has never been manufactured there or formally approved by the FDA. In 2009, a Canadian-based company, BioNeutra, received FDA-GRAS and Health Canada approval for IMO. The

European Food Safety Agency (EFSA) recently authorized xylo-oligosaccharides (XOS) as a novel food (NF) pursuant to Regulation (EU) 2015/2283.

Commercial availability

IMO is commercially manufactured mostly in China and Japan. However, most of this product is consumed locally or exported to neighboring Asian countries. In Japan, Meiji Dairies (Meiji Food Company) is one of the biggest IMO producers. IMO is marketed under several trade names like IMO-900 and IMO-800. Being a novel food ingredient, there wasn't a producer of IMO in North America and Europe until BioNeutra North America, Inc. began to manufacture this product with the VitaFiber IMO trademark, which was approved for use in Canada by Health Canada in 2012. US-based companies have been in producing other kinds of oligosaccharides, like GOS, FOS, and XOS.

Maltodextrin

Maltodextrin is a polysaccharide that is used as a food ingredient. It is produced from vegetable starch by partial hydrolysis and is usually found as a white hygroscopicspray-dried powder. Maltodextrin is easily digestible, being absorbed as rapidly as glucose and may be either moderately sweet or almost flavorless (depending on the degree of polymerisation). It can be found as an ingredient in a variety of processed foods.

Structure

Maltodextrin consists of D-glucose units connected in chains of variable length. The glucose units are primarily linked with α(1→4) glycosidic bonds, like that seen in the linear derivative of glycogen (after the removal of α1,6- branching). Maltodextrin is typically composed of a mixture of chains that vary from three to 17 glucose units long.

Maltodextrins are classified by DE (dextrose equivalent) and have a DE between 3 and 20. The higher the DE value, the shorter the glucose chains, the higher the sweetness, the higher the solubility, and the lower heat resistance. Above DE 20, the European Union's CN code calls it glucose syrup; at DE 10 or lower the customs CN code nomenclature classifies maltodextrins as dextrins.

Production

Maltodextrin can be enzymatically derived from any starch. In the US, this starch is usually corn; in Europe, it is common to use wheat. In the European Union, wheat-derived maltodextrin is exempt from wheat allergen labeling, as set out in Annex II of EC Regulation No 1169/2011. In the United States, however, it is not exempt from allergen declaration per the Food Allergen Labeling and Consumer Protection Act, and its effect on a voluntary gluten-free claim must be evaluated on a case-by-case basis per the applicable Food and Drug Administration policy.

Food uses

Maltodextrin is used to improve the mouthfeel of food and beverage products. It is also used in some snacks such as potato chips and jerky. It is used in "light" peanut butter to reduce the fat content but maintain the texture. Maltodextrin is also sometimes taken as a dietary supplement by athletes, in powder form, gel packets, or energy drinks.

Maltodextrin is used as an inexpensive additive to thicken food products such as infant formula. It is also used as a filler in sugar substitutes and other products. Maltodextrin has a glycemic index ranging from 85 to 105.

In animal studies, there is evidence to suggest that maltodextrin may exacerbate intestinal inflammation.

Other uses

Maltodextrin is used as a horticultural insecticide both in the field and in greenhouses. It has no biochemical action. Its efficacy is based upon spraying a dilute solution upon the pest insects, whereupon the solution dries, blocks the insects' spiracles and causes death by asphyxiation.

Nod factor

Nod factors (**nodulation factors** or **NF**), are signaling molecules produced by soil bacteria known as rhizobia in response to flavonoid exudation from plants under nitrogen

limited conditions. Nod factors initiate the establishment of a symbiotic relationship between legumes and rhizobia by inducing nodulation. Nod factors produce the differentiation of plant tissue in root hairs into nodules where the bacteria reside and are able to fix nitrogen from the atmosphere for the plant in exchange for photosynthates and the appropriate environment for nitrogen fixation. One of the most important features provided by the plant in this symbiosis is the production of leghemoglobin, which maintains the oxygen concentration low and prevents the inhibition of nitrogenase activity.

Chemical Structure

Nod factors structurally are lipochitooligosaccharides (LCOs) that consist of an *N*-acetyl-D-glucosamine chain linked through β-1,4 linkage with a fatty acid of variable identity attached to a non reducing nitrogen in the backbone with various functional group substitutions at the terminal or non-terminal residues.

Nod factors are produced in complex mixtures differing in the following characteristics:

- Length of the chain can vary from three to six units of *N*-acetyl-D-glucosamine with the exception of *M. loti* which can produce Nod factors with two unit only.
- Presence or absence of strain-specific substitutions along the chain
- Identity of the fatty acid component
- Presence or absence of unsaturated fatty acids

Nod gene expression is induced by the presence of certain flavonoids in the soil, which are secreted by the plant and act ad an attractant to bacteria and induce Nod factor production. Flavonoids activate NodD, a LysR family transcription factor, which binds to the *nod* box and initiates the transcription of the nod genes which encode the proteins necessary for the production of a wide range of LCOs.

Function

Nod factors are potentially recognized by plant receptors made of two histidine kinases with extracellular LysM domain, which have been identified in *L. japonicus*,soybean, and *M. truncatula.* Binding of Nod factors to these receptors depolarizes the plasma membrane of root hairs via an influx of Ca which induce the expression of early nodulin (ENOD) genes and swelling of the root hairs. In M. truncatula, the signal transduction initiates by the activation of *dmi1*, *dmi2*, and *dmi3* which lead to the deformation of root hairs, early nodulin expression, cortical cell division and bacterial infection. Additionally, *nsp* and *hcl* genes are recruited later and aid in the process of early nodulation expression, cortical cell division, and infection.Genes*dmi1, dmi2, and dmi3* have also been found to aid in the establishment of interactions between *M. truncatula* and arbuscular mycorrhiza, indicating that the two very different symbioses may share some common mechanisms. The end result is the nodule, the structure in which nitrogen is fixed. Nod factors act by inducing changes in gene expression in the legume, most notable the nodulin genes, which are needed for nodule organogenesis.

Nodulation

Rhizobia bind to host specific lectins present in root hairs which together with Nod factors lead to the formation of nodulation. Nod factors are recognized by a specific class of receptorkinases that have LysM domains in their extracellular domains. The two LysM (lysin motif) receptor kinases (NFR1 and NFR5) that appear to make up the Nod factor receptor were first isolated in the model legume *Lotus japonicus* in 2003. They now have been isolated also from soybean and the model legume *Medicago truncatula*. NFR5 lacks the classical activation loop in the kinase domain. The *NFR5* gene lacks introns. First the cell membrane is depolarized and the root hairs start to swell and cell division stops. Nod factor cause the fragmentation and rearrangement of actin network, which coupled with the reinstitution of cell growth lead to the curling of the root hair around the bacteria. This is followed by the localized breakdown of the cell wall and the invagination of the plant cell membrane, allowing the bacterium to form an infection thread. As the infection thread grows the rhizobia travel down its length towards the site of the nodule. During this process the pericycle cells in plants become activated and cells in the inner cortex start growing and become the nodule primordium where the rhizobia infect and differentiate into bacteroids and fix nitrogen. Activation of adjacent middle cortex cells leads to the formation of nodule meristem.

Oligosaccharide nomenclature

Oligosaccharides and **polysaccharides** are an important class of polymeric carbohydrates found in virtually all living entities.

Their structural features make their nomenclature challenging and their roles in living systems make their nomenclature important.

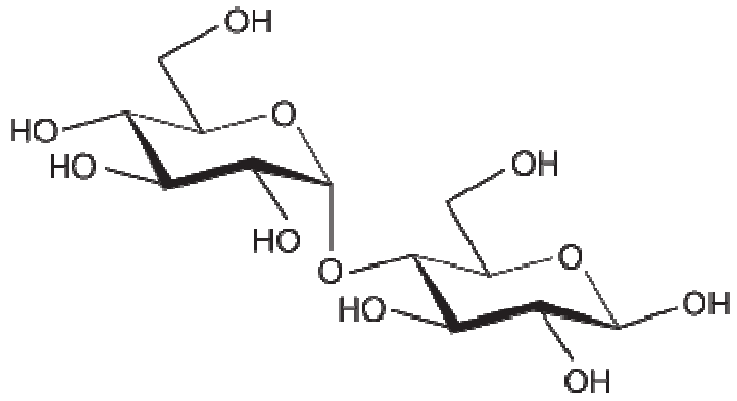
Oligosaccharides

Oligosaccharides are carbohydrates that are composed of several monosaccharide residues joined through glycosidic linkage, which can be hydrolyzed by enzymes or acid to give the constituent monosaccharide units. While a strict definition of an oligosaccharide is not established, it is generally agreed that a carbohydrate consisting of two to ten monosaccharide residues with a defined structure is an oligosaccharide.

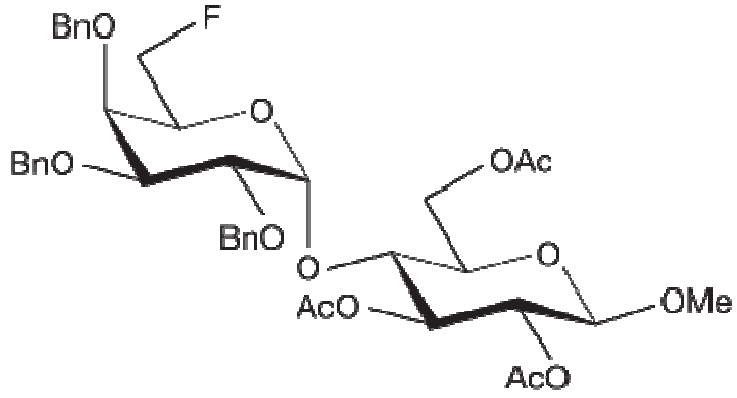
An oligosaccharide has both a reducing and a non-reducing end. The reducing end of an oligosaccharide is the monosaccharide residue with hemiacetal functionality, thereby capable of reducing the Tollens' reagent, while the non-reducing end is the monosaccharide residue in acetal form, thus incapable of reducing the Tollens' reagent. The reducing and non-reducing ends of an oligosaccharide are conventionally drawn with the reducing-end monosaccharide residue furthest to the right and the non-reducing (or terminal) end furthest to the left.

Naming of oligosaccharides proceeds from left to right (from the non-reducing end to the reducing end) as glycosyl [glycosyl]$_n$ glycoses or glycosyl [glycosyl]$_n$ glycosides, depending on whether or not the reducing end is a free hemiacetal group. In parentheses, between the names of the monosaccharide residues, the number of the anomeric carbon atom, an arrow symbol, and the number of the carbon atom bearing the

connecting oxygen of the next monosaccharide unit are listed. Appropriate symbols are used to indicate the stereochemistry of the glycosidic bonds (α orβ), the configuration of the monosaccharide residue (D orL), and the substitutions at oxygen atoms (*O*).Maltose and a derivative of sucrose illustrate these concepts:

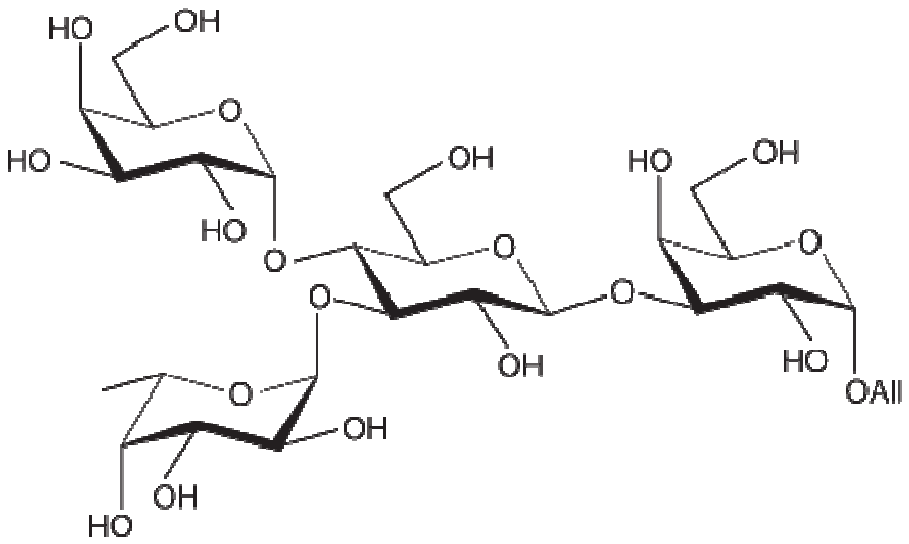


Maltose: α-D-Glucopyranosyl-(1→4)-β-D-glucopyranose



Methyl 2,3,4-tri-*O*-benzyl-6-deoxy-6-fluoro-α-D-galactopyranosyl-(1→4)-2,3,6-tri-*O*-acetyl-β-D-glucopyranoside

In the case of branched oligosaccharides, meaning that the structure contains at least one monosaccharide residue linked to more than two other monosaccharide residues, terms designating the branches should be listed in square brackets, with the longest linear chain (the parent chain) written without square brackets. The following example will help illustrate this concept:



Allyl α-L-fucopyranosyl-(1→3)-[α-D-galactopyranosyl-(1→4)]-α-D-glucopyranosyl-(1→3)-α-D-galactopyranoside

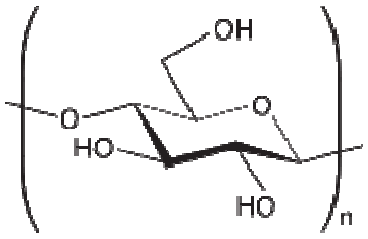
These systematic names are quite useful in that they provide information about the structure of the oligosaccharide. They do require a lot of space, however, so abbreviated forms are used when possible. In these abbreviated forms, the names of the monosaccharide units are shortened to their corresponding three-letter abbreviations, followed by *p* for pyranose or *f* for furanose ring structures, with the abbreviated aglyconic alcohol placed at the end of the name. Using this system, the previous example would have the abbreviated name α-L-Fuc*p*-

(1→3)-[α-D-Gal*p*-(1→4)]-α-D-Glc*p*-(1→3)-α-D-Gal*p*OAll.General Formula_Cn+1(H2o)n.. Structure Formula..C12'H22'O11.

Polysaccharides

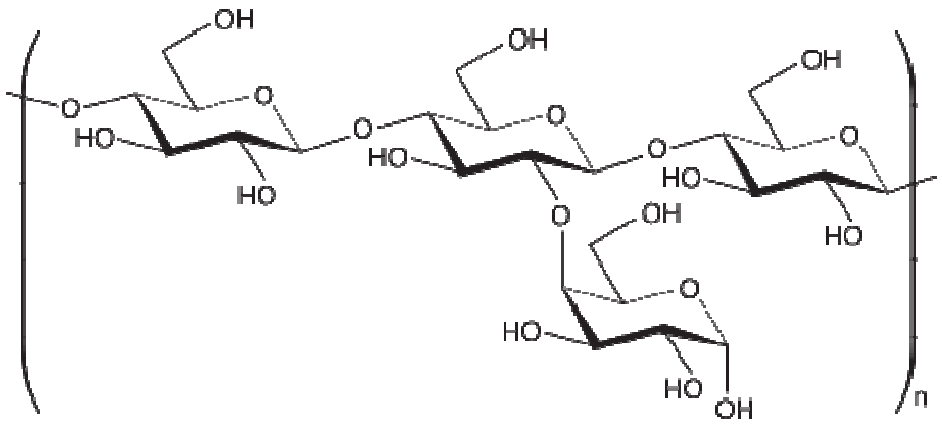
Polysaccharides are considered to be polymers of monosaccharides containing ten or more monosaccharide residues. Polysaccharides have been given trivial names that reflect their origin. Two common examples are cellulose, a main component of the cell wall in plants, and starch, a name derived from the Anglo-Saxon stercan, meaning to stiffen.

To name a polysaccharide composed of a single type of monosaccharide, that is a homopolysaccharide, the ending "-ose" of the monosaccharide is replaced with "-an". For example, a glucose polymer is named glucan, a mannose polymer is named mannan, and a galactose polymer is named galactan. When the glycosidic linkages and configurations of the monosaccharides are known, they may be included as a prefix to the name, with the notation for glycosidic linkages preceding the symbols designating the configuration. The following example will help illustrate this concept:



(1→4)-β-D-Glucan

A heteropolysaccharide is a polymer containing more than one kind of monosaccharide residue. The parent chain contains only one type of monosaccharide and should be listed last with the ending "-an", and the other types of monosaccharides listed in alphabetical order as "glyco-" prefixes. When there is no parent chain, all different monosaccharide residues are to be listed alphabetically as "glyco-" prefixes and the name should end with "-glycan". The following example will help illustrate this concept:



((1→2)-α-D-galacto)-(1→4)-β-D-Glucan

Validamycin

Validamycin is an antibiotic and fungicide produced by *Streptomyces hygroscopicus*. It is used as an inhibitor of trehalase. It is used for the control of sheath blight of rice and damping-off of cucumbers.

Xylooligosaccharide

Xylooligosaccharides (**XOS**) are polymers of the sugarxylose. They are produced from the xylan fraction in plant fiber. Their C5 (where C is a quantity of carbon atoms in each monomer) structure is fundamentally different from other prebiotics, which are based upon C6 sugars. Xylooligosaccharides have been commercially available since the 1980s, originally produced by Suntory in Japan. They have more recently become more widely available commercially, as technologies have advanced and production costs have fallen. Some enzymes from yeast can exclusively convert xylan into only xylooligosaccharides-DP-3 to 7.

Xylooligosaccharides act as a prebiotic, selectively feeding beneficial bacteria such as bifidobacteria and lactobacilli within the digestive tract. A large number of clinical trials have been conducted with XOS, demonstrating a variety of health benefits, including improvements in blood sugars and lipids, digestive health benefits, laxation, and beneficial changes to immune markers. These health benefits have typically been observed at 1–4 g/d, a lower dose than required for prebiotics such as fructooligosaccharides and inulin.

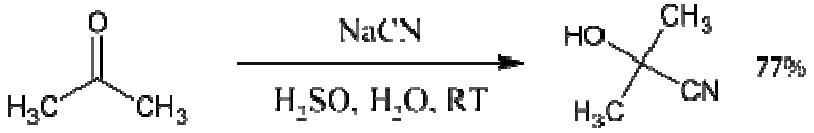
Chapter 6

Reactions of Carbohydrates

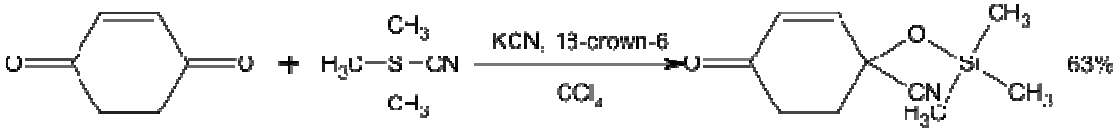
Cyanohydrin reaction

A **cyanohydrin reaction** is an organic chemical reaction by an aldehyde or ketone with a cyanideanion or a nitrile to form a cyanohydrin. This nucleophilic addition is a reversible reaction but with aliphaticcarbonyl compounds equilibrium is in favor of the reaction products. The cyanide source can be potassium cyanide, sodium cyanide or trimethylsilyl cyanide. With aromatic aldehydes such as benzaldehyde, the benzoin condensation is a competing reaction. The reaction is used in carbohydrate chemistry as a chain extension method for example that of D-xylose.

Examples

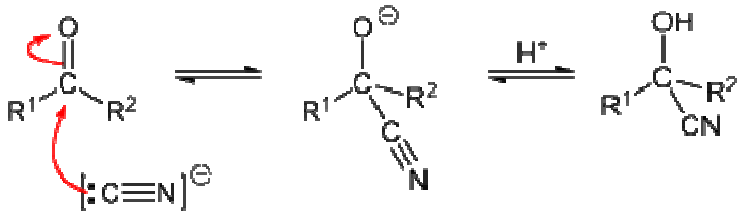


Reaction of acetone with sodium cyanide to hydroxyacetonitrile



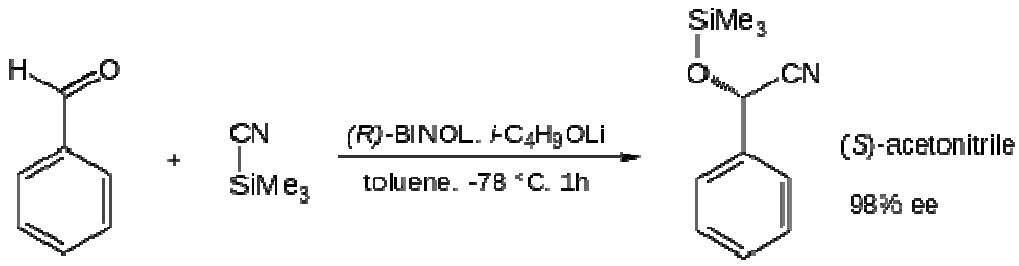
Reaction of benzoquinone with trimethylsilylcyanide, catalyst KCN is introduced as a 1:1 complex with the Crown ether 18-crown-6

Reaction mechanism



Asymmetric synthesis

The asymmetric cyanohydrin reaction of benzaldehyde with trimethylsilylcyanide is made possible by employment of (R)-Binol at 1–10% catalyst loading. This ligand firsts reacts with a lithium alkoxy compound to form a lithium binaphtholate Complex.



Asymmetric reaction of benzaldehyde with (R)–Binol–lithium(i-propyloxy) gives (S)-acetonitrile with 98% ee

The chemist Urech in 1872 was the first to synthesize cyanohydrins from ketones with alkali cyanides and acetic acid and therefore this reaction also goes by the name of **Urech cyanohydrin method**. With HCN in acidic conditions – i.e. the cyanohydrin is the functional group CN–C–OH.

Lobry de Bruyn–Van Ekenstein transformation

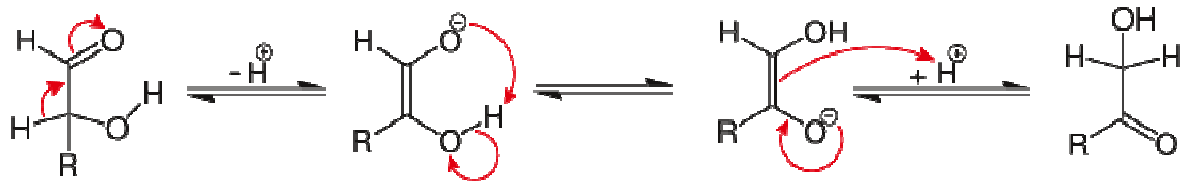
In carbohydrate chemistry, the **Lobry de Bruyn–Van Ekenstein transformation** also known as the **Lobry de Bruyn–Alberda van Ekenstein transformation** is the base or acid catalyzed transformation of an aldose into the ketose isomer or vice versa, with a tautomericenediol as reaction intermediate. Ketoses may be transformed into 3-ketoses, etcetera. The enediol is also an intermediate for the epimerization of an aldose or ketose.

The reactions are usually base catalyzed, but can also take place under acid or neutral conditions. A typical rearrangement reaction is that between the aldose glyceraldehyde and the ketose dihydroxyacetone in a chemical equilibrium.

The Lobry de Bruyn–Van Ekenstein transformation is relevant for the industrial production of certain ketoses and was discovered in 1885 by Cornelis Adriaan Lobry van Troostenburg de Bruyn and Willem Alberda van Ekenstein.

Aldose-ketose transformation

The following scheme describes the interconversion between an aldose and a ketose, where R is any organic residue.

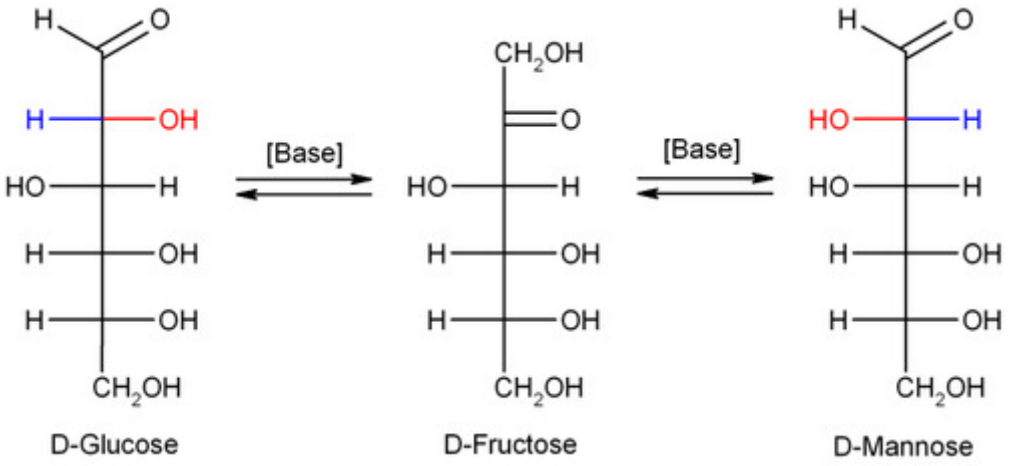


The equilibrium or the *reactant to product ratio* depends on concentration, solvent, pH and temperature. At equilibrium the aldose and ketose form a mixture which in the case of the glyceraldehyde and dihydroxyacetone is also called **glycerose**.

A related reaction is the alpha-ketol rearrangement.

Epimerization

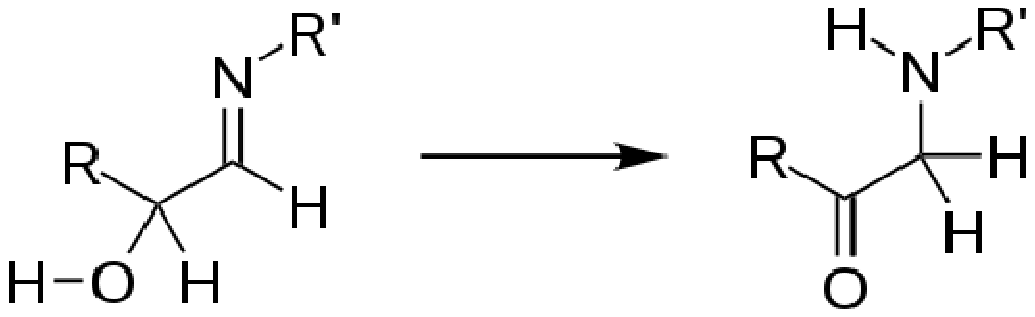
The carbon atom at which the initial deprotonation takes place is a stereocenter. If, for example, D-glucose (an Aldose) rearranges to D-fructose, the ketose, the stereochemical configuration is lost in the enol form. In the chemical reaction the enol can be protonated from two faces, resulting in the backformation of glucose or the formation of the epimerD-mannose. The final product is a mix of D-glucose, D-fructose and D-mannose.



Amadori rearrangement

The **Amadori rearrangement** is an organic reaction describing the acid or base catalyzedisomerization or rearrangement reaction of the *N*-glycoside of an aldose or the glycosylamine to the corresponding 1-amino-1-deoxy-ketose. The reaction is important in carbohydrate chemistry, specifically the glycation of hemoglobin (as measured by the HbA1c test).

The rearrangement is usually preceded by formation of a α-hydroxyimine by condensation of an amine with an aldose sugar. The rearrangement itself entails intramolecular redox reaction, converting this α-hydroxyimine to anα-ketoamine:



The formation of imines is generally reversible, but subsequent to conversion to the keto-amine, the attached amine is fixed

irreversibly. This Amadori product is an intermediate in the production of advanced glycation end-products (AGE)s. The formation of an advanced glycation end-product involves the oxidation of the Amadori product.

Food chemistry

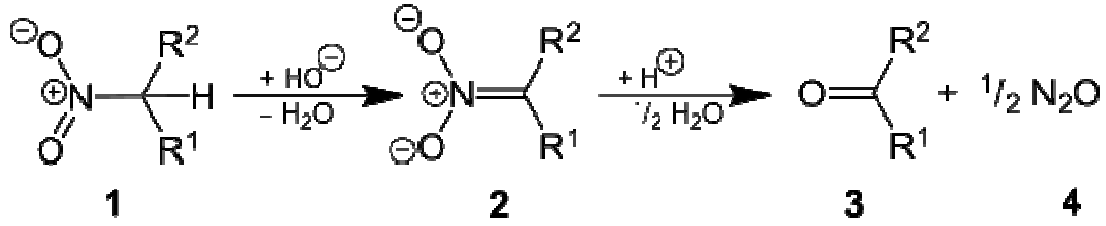
The reaction is associated with the Maillard reaction in which the reagents are naturally occurring sugars and amino acids. One study demonstrated the possibility of Amadori rearrangement during interaction between oxidized dextran and gelatine.

History

The Amadori rearrangement was discovered by the organic chemist Mario Amadori (1886–1941), who in 1925 reported this reaction while studying the Maillard reaction.

Nef reaction

The **Nef reaction** is an organic reaction describing the acidhydrolysis of a salt of a primary or secondary nitroalkane (**1**) to an aldehyde or a ketone (**3**) and nitrous oxide (**4**). The reaction has been the subject of several literature reviews.

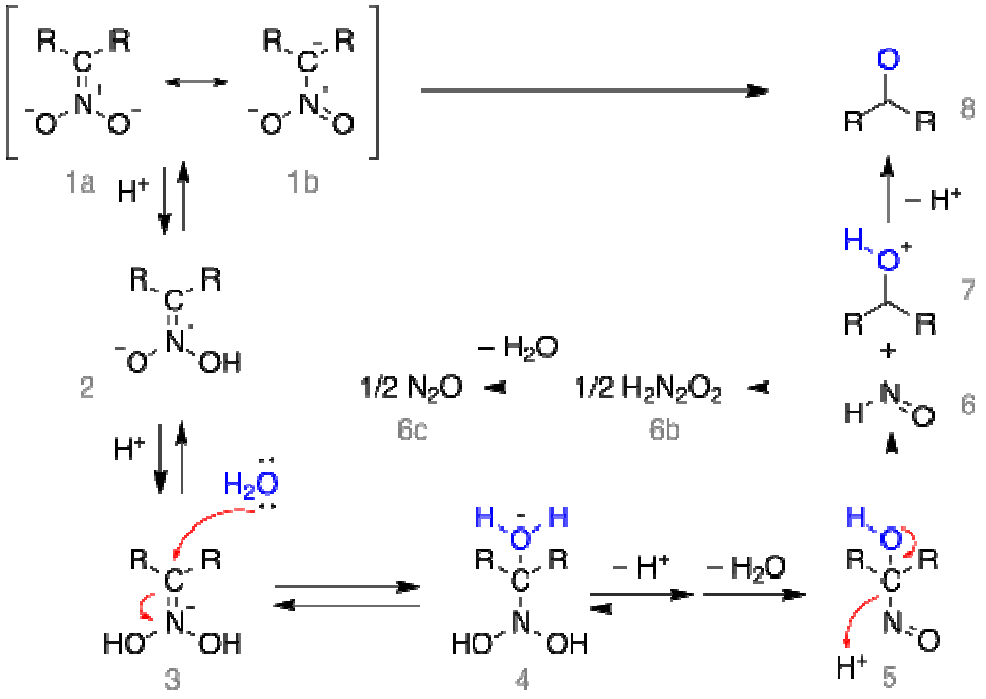


The reaction was reported in 1894 by the chemist John Ulric Nef, who treated the sodium salt of nitroethane with sulfuric acid resulting in an 85–89% yield of nitrous oxide and at least 70% yield of acetaldehyde. However, the reaction was pioneered a year earlier in 1893 by Konovalov, who converted the potassium salt of 1-phenylnitroethane with sulfuric acid to acetophenone.

The *Nef reaction*should not be confused with the Nef synthesis.

Reaction mechanism

The reaction mechanism starting from the nitronate salt as the resonance structures**1a** and **1b**is depicted below:



The salt is protonated forming the nitronic acid**2** (in some cases these nitronates have been isolated) and once more to the iminium ion **3**. This intermediate is attacked by water in a

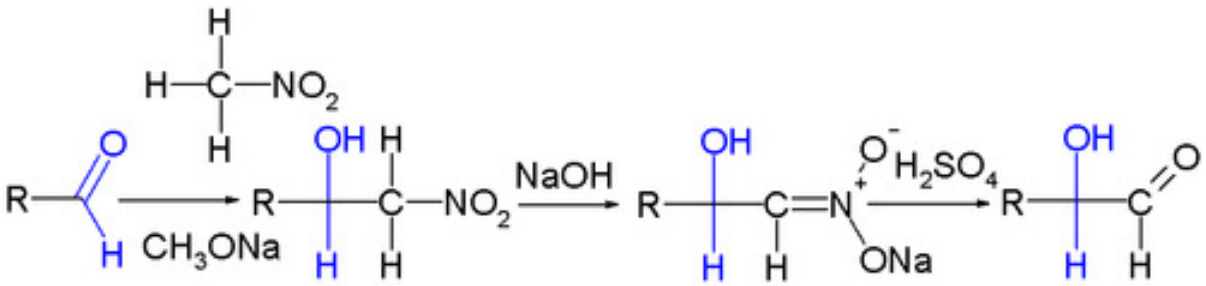
nucleophilic addition forming **4** which loses a proton and then water to the 1-nitroso-alkanol **5** which is believed to be responsible for the deep-blue color of the reaction mixture in many Nef reactions.

This intermediate rearranges to hyponitrous acid**6** (forming nitrous oxide**6c** through **6b**) and the oxonium ion**7** which loses a proton to form the carbonyl compound.

Note that formation of the nitronate salt from the nitro compound requires an alpha hydrogen atom and therefore the reaction fails with tertiary nitro compounds.

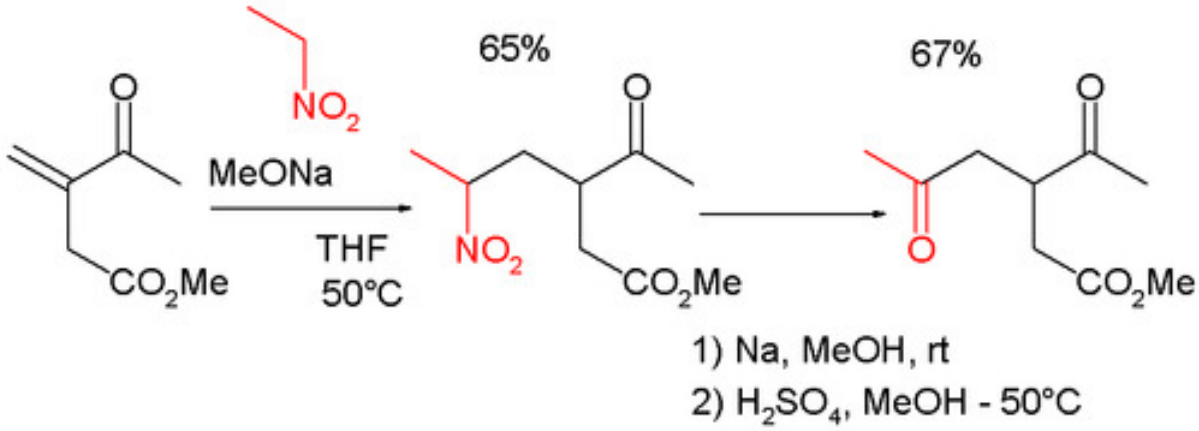
Scope

The Nef reaction is frequently encountered in organic synthesis. It has been applied in carbohydrate chemistry as a chain-extension method for aldoses for example in the isotope labeling of C-D-mannose and C-D-glucose from D-arabinose and C14-nitromethane (the first step here is a Henry reaction):



The opposite reaction is the Wohl degradation.

The reaction is also used in combination with the Michael reaction in the synthesis of γ-keto-carbonyls such as:



or*2,5-heptanedione*

Hydrolysis of nitro compounds with strong acid without the intermediate salt stage results in the formation of carboxylic acids and hydroxylamine salts.

The hydrolysis step of the Nef reaction can also be performed with Lewis acids such as tin(IV) chloride and iron(III) chloride or oxidizing agents, such as oxone.

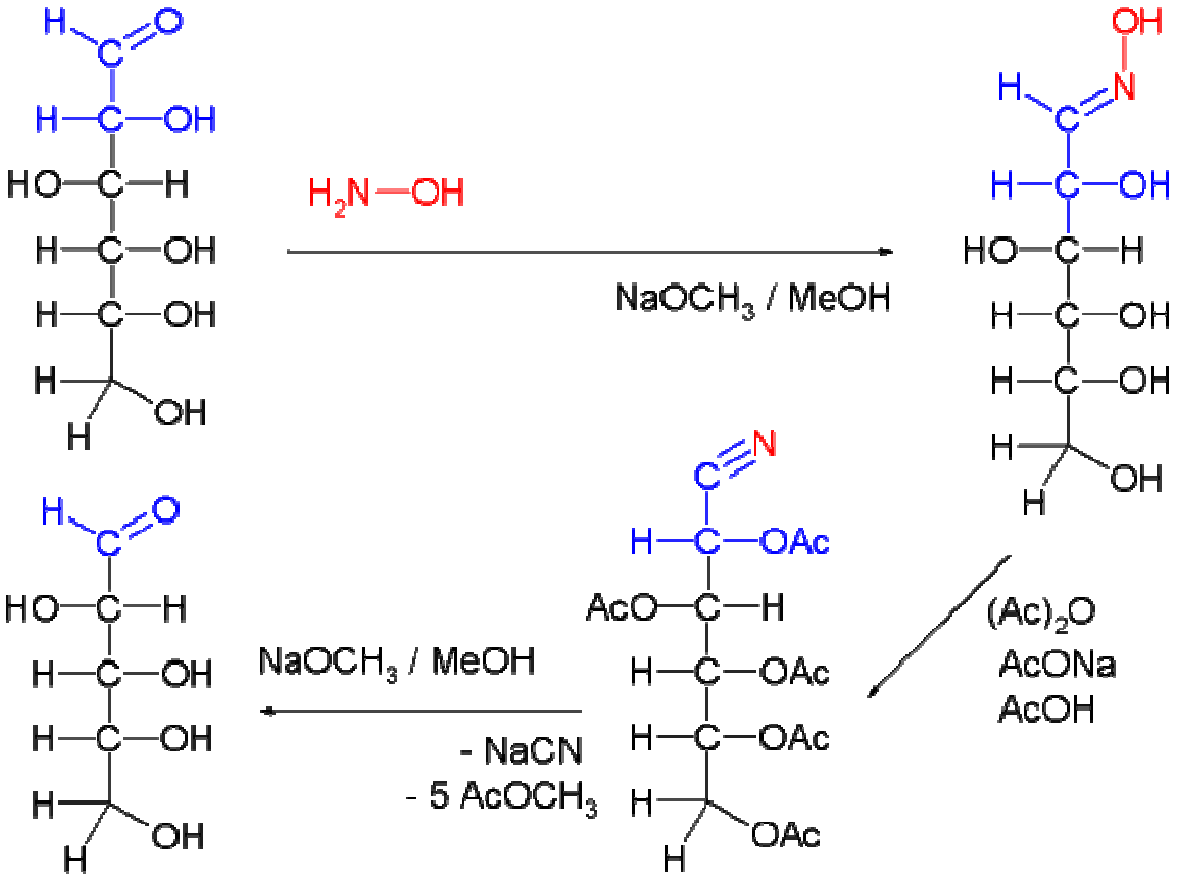
Wohl degradation

The **Wohl degradation** in carbohydrate chemistry is a chain contraction method for aldoses.

The classic example is the conversion of glucose to arabinose as shown below. The reaction is named after the German chemist Alfred Wohl (1863–1939).

In one modification, d-glucoseis converted to the glucose oxime by reaction with hydroxylamine and sodium methoxide. In the second step the pentaacetyl glycononitrile is formed by reaction with acetic anhydride in acetic acid with sodium

acetate. In this reaction step the oxime is converted into the nitrile with simultaneous conversion of all the alcohol groups to acetate groups.



In the final stepsodium methoxide in methanol is added, leading to removal of all the acetate groups and ejection of the nitrile group and collapse of the second carbon from a tetrahedral structure to an aldehyde.

Ruff–Fenton degradation

In a variation, the **Ruff–Fenton degradation** (Otto Ruff 1898, H.J.H. Fenton 1893) converts the aldose first to the alpha-hydroxy-carboxylic acid with bromine and calcium hydroxide

and then to the shortened aldose by reaction with Iron(III) sulfate and hydrogen peroxide.

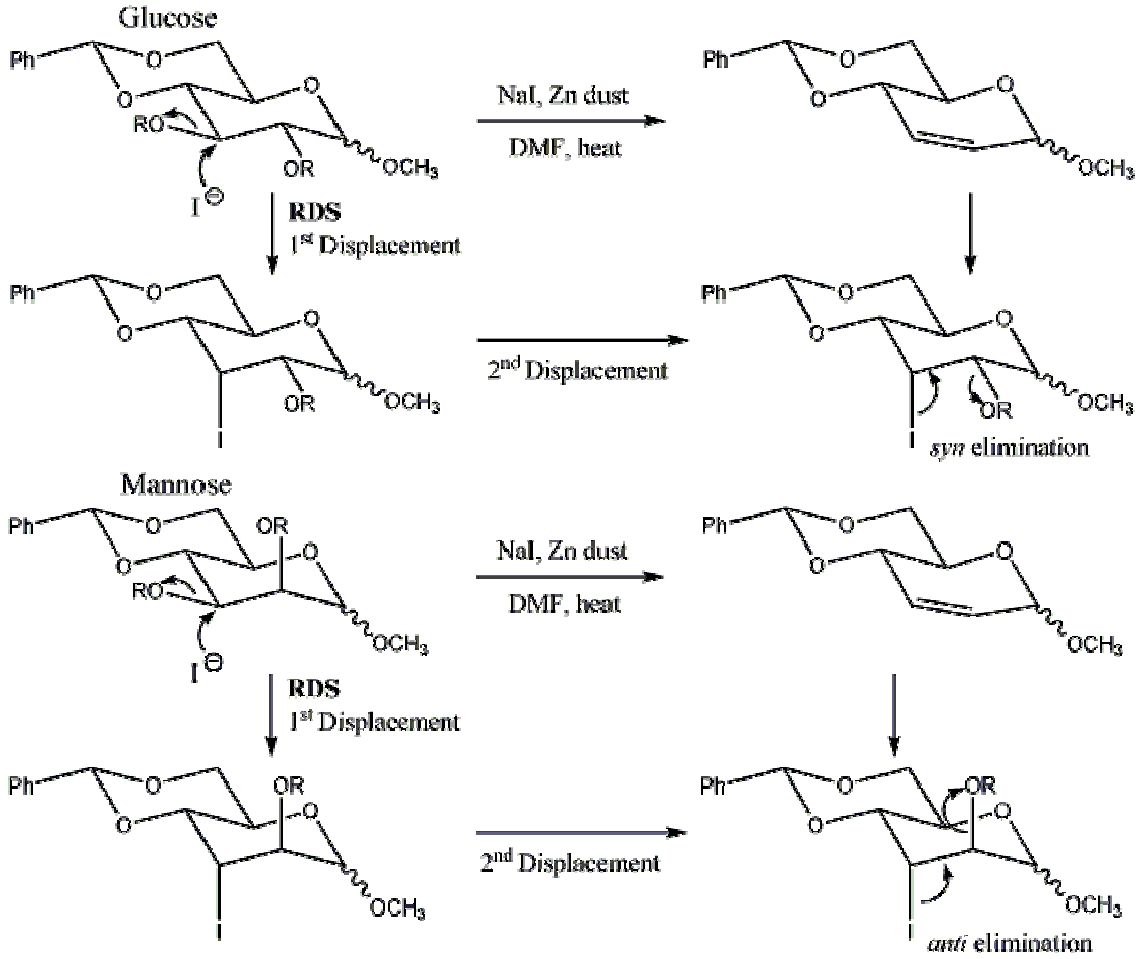
Tipson–Cohen reaction

The **Tipson–Cohen reaction** is a name reaction first discovered by Stuart Tipson and Alex Cohen at the National Bureau of Standards in Washington D.C. The Tipson–Cohen reaction occurs when two neighboring secondary sulfonyloxy groups in a sugar molecule are treated with zinc dust (Zn) and sodium iodide (NaI) in a refluxing solvent such as *N*,*N*-dimethylformamide (DMF) to give an unsaturated carbohydrate.

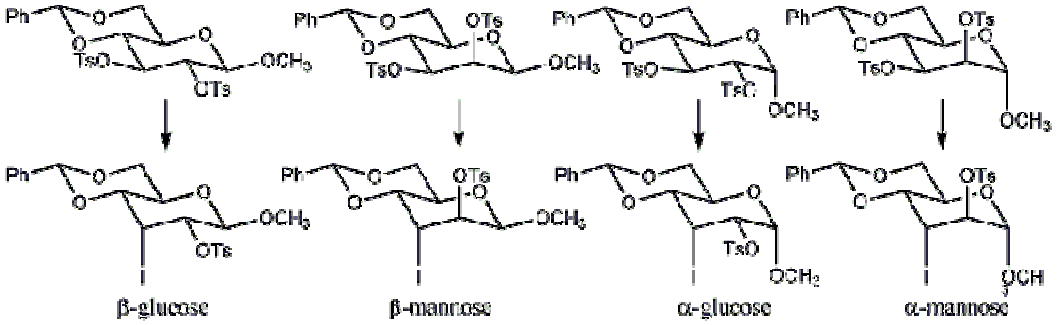
Background

Unsaturated carbohydrates are desired as they are versatile building blocks that can be used in a variety of reactions. For example, they can be used as intermediates in the synthesis of natural products, or as dienophiles in the Diels-Alder reaction, or as precursors in the synthesis of oligosaccharides. The Tipson–Cohen reaction goes through a *syn* or *anti* elimination mechanism to produce an alkene in high to moderate yields. The reaction depends on the neighboring substituents. A mechanism for glucopyranosides and mannooyranosides is shown below.

Scheme 1:*Syn* elimination occurs with the glucopyranosides. Galactopyranosides follows a similar syn mechanism. Whereas, *anti* elimination occurs with mannopyranosides. Note that R could be a methanesulfonyl CH_2O_2S (Ms), or a toluenesulfonyl $CH_3C_6H_4O_2S$ (Ts).



Reaction mechanism



Scheme 3: The scheme illustrates the first displacement, the rate determining step and slowest step, where the starting material is converted to the iodo-intermediate. The intermediate is not detectable as it is rapidly converted to the unsaturated sugar. Experiments with azide instead of the

iodide confirmed attack occurs at the C-3 as nitrogen-intermediates were isolated. The order of reactivity from most reactive to least reactive is:β-glucopyranosides >β-mannopyranosides >α-glucopyranosides>α-mannopyranosides.

The reaction of β–mannopyranosides gives low yields and required longer reaction times than with β-glucopyranosides due to the presence of a neighboring axial substituent (sulfonyloxy) relative to C-3 sulfonyloxy group in the starting material. The axial substituent increases the steric interactions in the transition state, causing unfavorable eclipsing of the two sulfonyloxy groups. α-Glucopyranosides possess a β-*trans*-axial substituent relative to C-3 sulfonyloxy (anomeric OCH_3 group) in the starting material. The β-*trans*-axial substituent influences the transition state by also causing an unfavorable steric interaction between the two groups. In the case of α-mannopyranosides, both a neighboring axial substituent (2-sulfonyloxy group) and a β-*trans*-axial substituent (anomeric OCH_3 group) are present, therefore significantly increasing the reaction time and decreasing the yield.

Reaction scope

The reaction has been attempted in the microwave, improving yields with the α-glucopyranoside to 88% and reducing the reaction time significantly to 14 minutes.

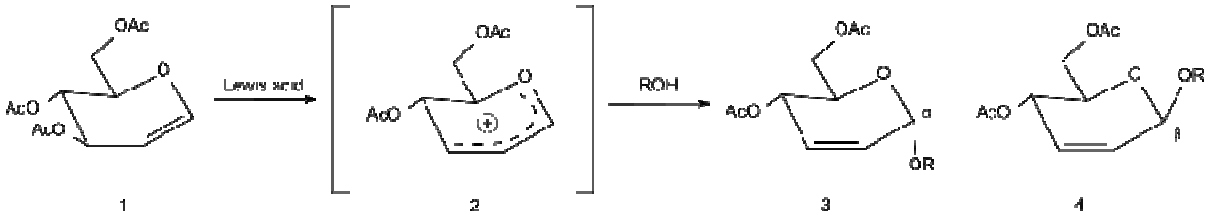
The original paper by Tipson and Cohen also used acyclic sugars to illustrate the utility of the reaction.

Thus the reaction is not limited to cyclic carbohydrate derivatives.

Sulphonoxy groups such as methanesulfonyl and toluenesulfonyl were both used, however it was found that substrates with toluenesulfonyl groups gave higher yields and lower reaction times.

Ferrier rearrangement

The **Ferrier rearrangement** is an organic reaction that involves a nucleophilic substitution reaction combined with an allylic shift in a glycal (a 2,3-unsaturatedglycoside). It was discovered by the carbohydrate chemist Robert J. Ferrier.



Mechanism

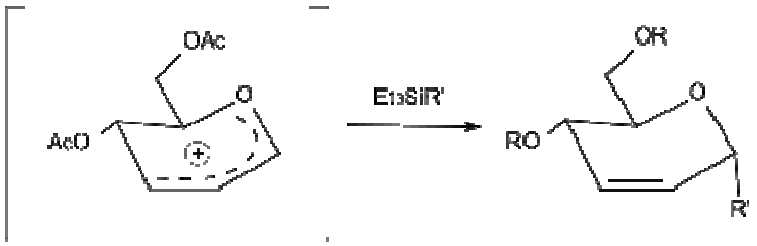
In the first step, a delocalized allyloxocarbenium ion (**2**) is formed, typically with the aid of a Lewis acid like indium(III) chloride or boron trifluoride.

This ion reacts in situ with an alcohol, yielding a mixture of the α (**3**) and β (**4**) anomers of the 2-glycoside, with the double bond shifted to position 3,4.

Modifications

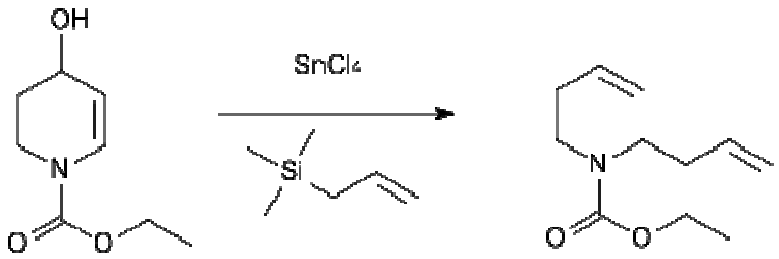
Forming of C-glycosides

By replacing the alcohol with a silane, C-glycosides can be formed. With triethylsilane (R'=H), the reaction yields a 2,3-unsaturated deoxy sugar.



Nitrogen analogue

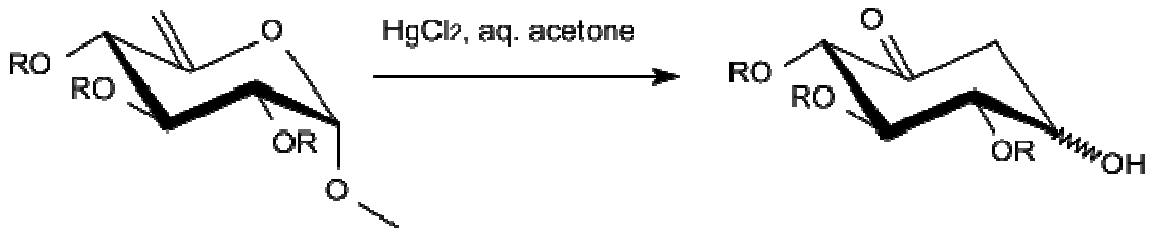
An analogous reaction with nitrogen as the heteroatom was described in 1984 for the synthesis of the antibiotic substance streptazolin.



Ferrier carbocyclization

The **Ferrier carbocyclization** (or **Ferrier II reaction**) is an organic reaction that was first reported by the carbohydrate

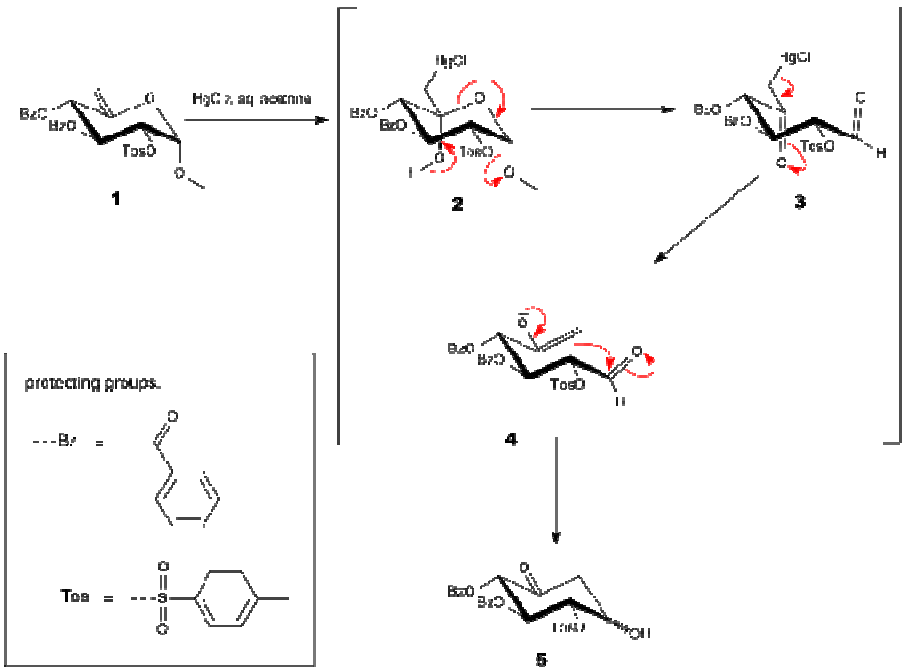
chemist Robert J. Ferrier in 1979. It is a metal-mediated rearrangement of enol etherpyrans to cyclohexanones. Typically, this reaction is catalyzed by mercury salts, specifically mercury(II) chloride.



Several reviews have been published.

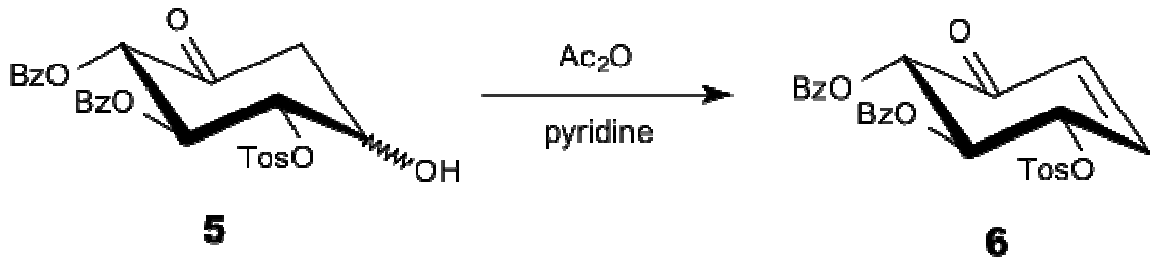
Reaction mechanism

Ferrier proposed the following reaction mechanism:



In this mechanism, the terminal olefin undergoes hydroxymercuration to produce the first intermediate, compound **2**, a hemiacetal. Next, methanol is lost and the dicarbonyl compound cyclizes through an attack on the electrophilic aldehyde to form the carbocycle as the product. A downside to this reaction is that the loss of CH_3OH at the anomeric position (carbon-1) results in a mixture of α- and β-anomers. The reaction also works for substituted alkenes (e. g. having an -OAc group on the terminal alkene).

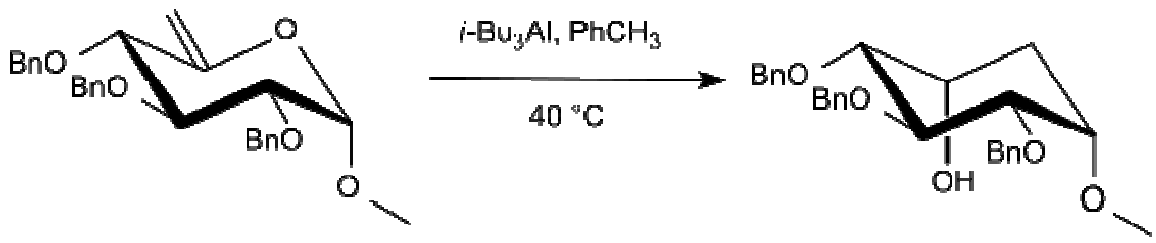
Ferrier also reported that the final product, compound **5**, could be converted into a conjugated ketone (compound **6**) by reaction with acetic anhydride (Ac_2O) and pyridine, as shown below.



5 **6**

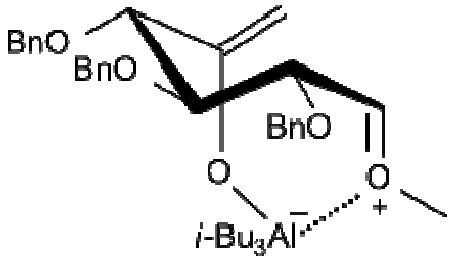
Modifications

In 1997, Sinaÿ and co-workers reported an alternative route to the synthesis (shown below) that did not involve cleavage of the bond at the anomeric position (the glycosidic bond). In this case, the major product formed had maintained its original configuration at the anomeric position.



(Bn = benzyl, *i*-Bu = isobutyl)

Sinaÿ proposed this reaction went through the following transition state:



Sinaÿ also discovered that titanium (IV) derivatives such as [$TiCl_3$(O*i*Pr)] worked in the same reaction as a milder version of the Lewis acid, *i*-Bu_3Al,which goes through a similar transition state involving the retention of configuration at the anomeric center.

In 1988, Adam reported a modification of the reaction that used catalytic amounts of palladium (II) salts, which brought about the same conversion of enol ethers into carbosugars in a more environmentally friendly manner.

Applications

The development of the Ferrier carbocyclization has been useful for the synthesis of numerous natural products that contain the carbocycle group. In 1991, Bender and co-workers reported a synthetic route to pure enantiomers of *myo*-inositol derivatives using this reaction. It has also been applied to the synthesis of aminocyclitols in work done by Barton and co-workers. Finally, Amano *et al.* used the Ferrier conditions to synthesise complex conjugated cyclohexanones in 1998.

Chapter 7

Functions of Carbohydrates

Glucose

Glucose is a simple sugar with the molecular formula$C_6H_{12}O_6$. Glucose is the most abundant monosaccharide, a subcategory of carbohydrates. Glucose is mainly made by plants and most algae during photosynthesis from water and carbon dioxide, using energy from sunlight, where it is used to make cellulose in cell walls, the most abundant carbohydrate in the world.

In energy metabolism, glucose is the most important source of energy in all organisms. Glucose for metabolism is stored as a polymer, in plants mainly as starch and amylopectin, and in animals as glycogen. Glucose circulates in the blood of animals as blood sugar. The naturally occurring form of glucose is d-glucose, while l-glucoseis produced synthetically in comparatively small amounts and is of lesser importance. Glucose is a monosaccharide containing six carbon atoms and an aldehyde group, and is therefore an aldohexose. The glucose molecule can exist in an open-chain (acyclic) as well as ring (cyclic) form. Glucose is naturally occurring and is found in fruits and other parts of plants in its free state. In animals, glucose is released from the breakdown of glycogen in a process known as glycogenolysis.

Glucose, as intravenous sugar solution, is on the World Health Organization's List of Essential Medicines, the safest and most

effective medicines needed in a health system. It is also on the list in combination with sodium chloride.

The name glucose is derived from Ancient Greekγλε□κος (gleûkos, “wine, must”), from γλυκύς (glykýs, “sweet”). The suffix "-ose" is a chemical classifier, denoting a sugar.

History

Glucose was first isolated from raisins in 1747 by the German chemist Andreas Marggraf. Glucose was discovered in grapes by Johann Tobias Lowitz in 1792, and distinguished as being different from cane sugar (sucrose). Glucose is the term coined by Jean Baptiste Dumas in 1838, which has prevailed in the chemical literature. Friedrich August Kekulé proposed the term dextrose (from Latin dexter = right), because in aqueous solution of glucose, the plane of linearly polarized light is turned to the right. In contrast, d-fructose (a ketohexose) and l-glucose turn linearly polarized light to the left. The earlier notation according to the rotation of the plane of linearly polarized light (*d* and *l*-nomenclature) was later abandoned in favor of the d- and l-notation, which refers to the absolute configuration of the asymmetric center farthest from the carbonyl group, and in concordance with the configuration of d- or l-glyceraldehyde.

Since glucose is a basic necessity of many organisms, a correct understanding of its chemical makeup and structure contributed greatly to a general advancement in organic chemistry. This understanding occurred largely as a result of the investigations of Emil Fischer, a German chemist who received the 1902 Nobel Prize in Chemistry for his findings.

The synthesis of glucose established the structure of organic material and consequently formed the first definitive validation of Jacobus Henricus van 't Hoff's theories of chemical kinetics and the arrangements of chemical bonds in carbon-bearing molecules. Between 1891 and 1894, Fischer established the stereochemical configuration of all the known sugars and correctly predicted the possible isomers, applying Van 't Hoff's theory of asymmetrical carbon atoms. The names initially referred to the natural substances. Their enantiomers were given the same name with the introduction of systematic nomenclatures, taking into account absolute stereochemistry (e.g. Fischer nomenclature, d/l nomenclature).

For the discovery of the metabolism of glucoseOtto Meyerhof received the Nobel Prize in Physiology or Medicine in 1922.Hans von Euler-Chelpinwas awarded the Nobel Prize in Chemistry along with Arthur Harden in 1929 for their "research on the fermentation of sugar and their share of enzymes in this process". In 1947, Bernardo Houssay (for his discovery of the role of the pituitary gland in the metabolism of glucose and the derived carbohydrates) as well as Carl and Gerty Cori (for their discovery of the conversion of glycogen from glucose) received the Nobel Prize in Physiology or Medicine. In 1970, Luis Leloirwas awarded the Nobel Prize in Chemistry for the discovery of glucose-derived sugar nucleotides in the biosynthesis of carbohydrates.

Chemical properties

Glucose forms white or colorless solids that are highly soluble in water and acetic acid but poorly soluble in methanol and ethanol. They melt at 146 °C (295 °F) (α) and 150 °C (302 °F)

(*β*), and decompose starting at 188 °C (370 °F) with release of various volatile products, ultimately leaving a residue of carbon.

With six carbon atoms, it is classed as a hexose, a subcategory of the monosaccharides. d-Glucose is one of the sixteen aldohexosestereoisomers. The d-isomer, d-glucose, also known as *dextrose*, occurs widely in nature, but the l-isomer, l-glucose, does not. Glucose can be obtained by hydrolysis of carbohydrates such as milk sugar (lactose), cane sugar (sucrose), maltose, cellulose, glycogen, etc. Dextrose is commonly commercially manufactured from cornstarch in the US and Japan, from potato and wheat starch in Europe, and from tapioca starch in tropical areas. The manufacturing process uses hydrolysis via pressurized steaming at controlled pH in a jet followed by further enzymatic depolymerization. Unbonded glucose is one of the main ingredients of honey. All forms of glucose are colorless and easily soluble in water, acetic acid, and several other solvents. They are only sparingly soluble in methanol and ethanol.

Structure and nomenclature

Glucose is usually present in solid form as a monohydrate with a closed pyran ring (dextrose hydrate). In aqueous solution, on the other hand, it is an open-chain to a small extent and is present predominantly as α- or β-pyranose, which interconvert (see mutarotation). From aqueous solutions, the three known forms can be crystallized: α-glucopyranose, β-glucopyranose and β-glucopyranose hydrate. Glucose is a building block

of the disaccharides lactose and sucrose (cane or beet sugar), of oligosaccharides such as raffinose and of polysaccharides such as starch and amylopectin, glycogen or cellulose. The glass transition temperature of glucose is 31 °C and the Gordon–Taylor constant (an experimentally determined constant for the prediction of the glass transition temperature for different mass fractions of a mixture of two substances) is 4.5.

Open-chain form

The open-chain form of glucose makes up less than 0.02% of the glucose molecules in an aqueous solution. The rest is one of two cyclic hemiacetal forms. In its open-chain form, the glucose molecule has an open (as opposed to cyclic) unbranched backbone of six carbon atoms, where C-1 is part of an aldehyde group H(C=O)–. Therefore, glucose is also classified as an aldose, or an aldohexose. The aldehyde group makes glucose a reducing sugar giving a positive reaction with the Fehling test.

Cyclic forms

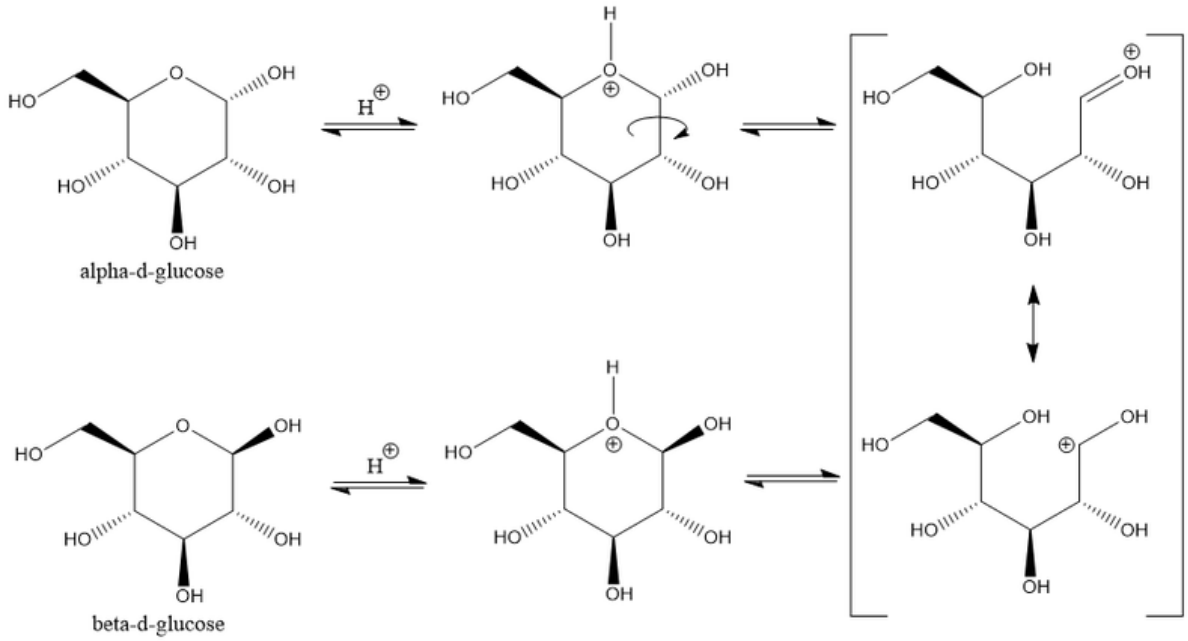
In solutions, the open-chain form of glucose (either "D-" or "L-") exists in equilibrium with several cyclic isomers, each containing a ring of carbons closed by one oxygen atom. In aqueous solution, however, more than 99% of glucose molecules exist as pyranose forms. The open-chain form is limited to about 0.25%, and furanose forms exist in negligible amounts. The terms "glucose" and "D-glucose" are

generally used for these cyclic forms as well. The ring arises from the open-chain form by an intramolecular nucleophilic addition reaction between the aldehyde group (at C-1) and either the C-4 or C-5 hydroxyl group, forming a hemiacetal linkage, –C(OH)H–O–.

The reaction between C-1 and C-5 yields a six-membered heterocyclic system called a pyranose, which is a monosaccharide sugar (hence "-ose") containing a derivatised pyran skeleton. The (much rarer) reaction between C-1 and C-4 yields a five-membered furanose ring, named after the cyclic ether furan. In either case, each carbon in the ring has one hydrogen and one hydroxyl attached, except for the last carbon (C-4 or C-5) where the hydroxyl is replaced by the remainder of the open molecule (which is $-(C(CH_2OH)HOH)-H$ or $-(CHOH)-H$ respectively).

The ring-closing reaction can give two products, denoted "α-" and "β-" When a glucopyranose molecule is drawn in the Haworth projection, the designation "α-" means that the hydroxyl group attached to C-1 and the $-CH_2OH$ group at C-5 lies on opposite sides of the ring's plane (a *trans* arrangement), while "β-" means that they are on the same side of the plane (a *cis* arrangement).

Therefore, the open-chain isomer D-glucose gives rise to four distinct cyclic isomers: α-D-glucopyranose, β-D-glucopyranose, α-D-glucofuranose, and β-D-glucofuranose. These five structures exist in equilibrium and interconvert, and the interconversion is much more rapid with acid catalysis.



The other open-chain isomer L-glucose similarly gives rise to four distinct cyclic forms of L-glucose, each the mirror image of the corresponding D-glucose.

The glucopyranose ring (α orβ) can assume several non-planar shapes, analogous to the "chair" and "boat" conformations of cyclohexane. Similarly, the glucofuranose ring may assume several shapes, analogous to the "envelope" conformations of cyclopentane.

In the solid state, only the glucopyranose forms are observed.

Some derivatives of glucofuranose, such as 1,2-*O*-isopropylidene-D-glucofuranose are stable and can be obtained pure as crystalline solids. For example, reaction of α-D-glucose with *para*-tolylboronic acid$H_3C–(C_6H_4)–B(OH)_2$ reforms the normal pyranose ring to yield the 4-fold ester α-D-glucofuranose-1,2:3,5-bis(*p*-tolylboronate).

Mutarotation

Mutarotation consists of a temporary reversal of the ring-forming reaction, resulting in the open-chain form, followed by a reforming of the ring. The ring closure step may use a different –OH group than the one recreated by the opening step (thus switching between pyranose and furanose forms), or the new hemiacetal group created on C-1 may have the same or opposite handedness as the original one (thus switching between the α and β forms). Thus, though the open-chain form is barely detectable in solution, it is an essential component of the equilibrium.

The open-chain form is thermodynamically unstable, and it spontaneously isomerizes to the cyclic forms. (Although the ring closure reaction could in theory create four- or three-atom rings, these would be highly strained, and are not observed in practice.) In solutions at room temperature, the four cyclic isomers interconvert over a time scale of hours, in a process called mutarotation. Starting from any proportions, the mixture converges to a stable ratio of α:β 36:64. The ratio would be α:β 11:89 if it were not for the influence of the anomeric effect. Mutarotation is considerably slower at temperatures close to 0 °C (32 °F).

Optical activity

Whether in water or the solid form, d-(+)-glucose is dextrorotatory, meaning it will rotate the direction of polarized light clockwise as seen looking toward the light source. The effect is due to the chirality of the molecules, and indeed the mirror-image isomer, l-(–)-glucose, is levorotatory (rotates

polarized light counterclockwise) by the same amount. The strength of the effect is different for each of the five tautomers.

Note that the d- prefix does not refer directly to the optical properties of the compound. It indicates that the C-5 chiral centre has the same handedness as that of d-glyceraldehyde (which was so labelled because it is dextrorotatory). The fact that d-glucose is dextrorotatory is a combined effect of its four chiral centres, not just of C-5; and indeed some of the other d-aldohexoses are levorotatory.

The conversion between the two anomers can be observed in a polarimeter since pure α-dglucose has a specific rotation angle of +112.2°·ml/(dm·g), pure β- D- glucose of +17.5°·ml/(dm·g). When equilibrium has been reached after a certain time due to mutarotation, the angle of rotation is +52.7°·ml/(dm·g). By adding acid or base, this transformation is much accelerated. The equilibration takes place via the open-chain aldehyde form.

Isomerisation

In dilute sodium hydroxide or other dilute bases, the monosaccharides mannose, glucose and fructose interconvert (via a Lobry de Bruyn–Alberda–Van Ekenstein transformation), so that a balance between these isomers is formed. This reaction proceeds via an enediol:

Biochemical properties

Glucose is the most abundant monosaccharide. Glucose is also the most widely used aldohexose in most living organisms. One

possible explanation for this is that glucose has a lower tendency than other aldohexoses to react nonspecifically with the amine groups of proteins. This reaction—glycation—impairs or destroys the function of many proteins, e.g. in glycated hemoglobin. Glucose's low rate of glycation can be attributed to its having a more stable cyclic form compared to other aldohexoses, which means it spends less time than they do in its reactive open-chain form. The reason for glucose having the most stable cyclic form of all the aldohexoses is that its hydroxy groups (with the exception of the hydroxy group on the anomeric carbon of d-glucose) are in the equatorial position. Presumably, glucose is the most abundant natural monosaccharide because it is less glycated with proteins than other monosaccharides. Another hypothesis is that glucose, being the only D-aldohexose that has all five hydroxy substituents in the equatorial position in the form of β-D-glucose, is more readily accessible to chemical reactions, for example, for esterification or acetal formation. For this reason, D-glucose is also a highly preferred building block in natural polysaccharides (glycans). Polysaccharides that are composed solely of glucose are termed glucans.

Glucose is produced by plants through the photosynthesis using sunlight, water and carbon dioxide and can be used by all living organisms as an energy and carbon source. However, most glucose does not occur in its free form, but in the form of its polymers, i.e. lactose, sucrose, starch and others which are energy reserve substances, and cellulose and chitin, which are components of the cell wall in plants or fungi and arthropods, respectively. These polymers are degraded to glucose during food intake by animals, fungi and bacteria using enzymes. All animals are also able to produce glucose themselves from

certain precursors as the need arises. Neurons, cells of the renal medulla and erythrocytes depend on glucose for their energy production. In adult humans, there are about 18 g of glucose, of which about 4 g are present in the blood. Approximately 180 to 220 g of glucose are produced in the liver of an adult in 24 hours.

Many of the long-term complications of diabetes (e.g., blindness, kidney failure, and peripheral neuropathy) are probably due to the glycation of proteins or lipids. In contrast, enzyme-regulated addition of sugars to protein is calledglycosylation and is essential for the function of many proteins.

Uptake

Ingested glucose initially binds to the receptor for sweet taste on the tongue in humans. This complex of the proteins T1R2 and T1R3 makes it possible to identify glucose-containing food sources. Glucose mainly comes from food - about 300 g per day are produced by conversion of food, but it is also synthesized from other metabolites in the body's cells. In humans, the breakdown of glucose-containing polysaccharides happens in part already during chewing by means of amylase, which is contained in saliva, as well as by maltase, lactase, and sucrase on the brush border of the small intestine. Glucose is a building block of many carbohydrates and can be split off from them using certain enzymes. Glucosidases, a subgroup of the glycosidases, first catalyze the hydrolysis of long-chain glucose-containing polysaccharides, removing terminal glucose. In turn, disaccharides are mostly degraded by specific glycosidases to glucose. The names of the degrading enzymes

are often derived from the particular poly- and disaccharide; inter alia, for the degradation of polysaccharide chains there are amylases (named after amylose, a component of starch), cellulases (named after cellulose), chitinases (named after chitin), and more. Furthermore, for the cleavage of disaccharides, there are maltase, lactase, sucrase, trehalase, and others. In humans, about 70 genes are known that code for glycosidases. They have functions in the digestion and degradation of glycogen, sphingolipids, mucopolysaccharides, and poly(ADP-ribose). Humans do not produce cellulases, chitinases, or trehalases, but the bacteria in the gut flora do.

In order to get into or out of cell membranes of cells and membranes of cell compartments, glucose requires special transport proteins from the major facilitator superfamily. In the small intestine (more precisely, in the jejunum), glucose is taken up into the intestinal epithelium with the help of glucose transporters via a secondary active transport mechanism called sodium ion-glucose symport via sodium/glucose cotransporter 1 (SGLT1). Further transfer occurs on the basolateral side of the intestinal epithelial cells via the glucose transporter GLUT2, as well uptake into liver cells, kidney cells, cells of the islets of Langerhans, neurons, astrocytes, and tanycytes. Glucose enters the liver via the portal vein and is stored there as a cellular glycogen. In the liver cell, it is phosphorylated by glucokinase at position 6 to form glucose 6-phosphate, which cannot leave the cell. Glucose 6-phosphatase can convert glucose 6-phosphate back into glucose exclusively in the liver, so the body can maintain a sufficient blood glucose concentration. In other cells, uptake happens by passive transport through one of the 14 GLUT proteins. In the other

cell types, phosphorylation occurs through a hexokinase, whereupon glucose can no longer diffuse out of the cell.

The glucose transporter GLUT1is produced by most cell types and is of particular importance for nerve cells and pancreatic β-cells.GLUT3 is highly expressed in nerve cells. Glucose from the bloodstream is taken up by GLUT4 from muscle cells (of the skeletal muscle and heart muscle) and fat cells.GLUT14is expressed exclusively in testicles. Excess glucose is broken down and converted into fatty acids, which are stored as triglycerides. In the kidneys, glucose in the urine is absorbed via SGLT1 and SGLT2 in the apical cell membranes and transmitted via GLUT2 in the basolateral cell membranes. About 90% of kidney glucose reabsorption is via SGLT2 and about 3% via SGLT1.

Biosynthesis

In plants and some prokaryotes, glucose is a product of photosynthesis. Glucose is also formed by the breakdown of polymeric forms of glucose like glycogen (in animals and mushrooms) or starch (in plants). The cleavage of glycogen is termed glycogenolysis, the cleavage of starch is called starch degradation.

The metabolic pathway that begins with molecules containing two to four carbon atoms (C) and ends in the glucose molecule containing six carbon atoms is called gluconeogenesis and occurs in all living organisms. The smaller starting materials are the result of other metabolic pathways. Ultimately almost all biomolecules come from the assimilation of carbon dioxide in plants during photosynthesis. The free energy of formation

of α-d-glucose is 917.2 kilojoules per mole. In humans, gluconeogenesis occurs in the liver and kidney, but also in other cell types. In the liver about 150 g of glycogen are stored, in skeletal muscle about 250 g. However, the glucose released in muscle cells upon cleavage of the glycogen can not be delivered to the circulation because glucose is phosphorylated by the hexokinase, and a glucose-6-phosphatase is not expressed to remove the phosphate group. Unlike for glucose, there is no transport protein for glucose-6-phosphate. Gluconeogenesis allows the organism to build up glucose from other metabolites, including lactate or certain amino acids, while consuming energy. The renal tubular cells can also produce glucose.

Glucose degradation

In humans, glucose is metabolised by glycolysis and the pentose phosphate pathway.Glycolysis is used by all living organisms, with small variations, and all organisms generate energy from the breakdown of monosaccharides. In the further course of the metabolism, it can be completely degraded via oxidative decarboxylation, the citric acid cycle (synonym *Krebs cycle*) and the respiratory chain to water and carbon dioxide. If there is not enough oxygen available for this, the glucose degradation in animals occurs anaerobic to lactate via lactic acid fermentation and releases less energy. Muscular lactate enters the liver through the bloodstream in mammals, where gluconeogenesis occurs (Cori cycle). With a high supply of glucose, the metabolite acetyl-CoA from the Krebs cycle can also be used for fatty acid synthesis. Glucose is also used to replenish the body's glycogen stores, which are mainly found in liver and skeletal muscle. These processes are hormonally

regulated. In other living organisms, other forms of fermentation can occur. The bacterium *Escherichia coli* can grow on nutrient media containing glucose as the sole carbon source. In some bacteria and, in modified form, also in archaea, glucose is degraded via the Entner-Doudoroff pathway.

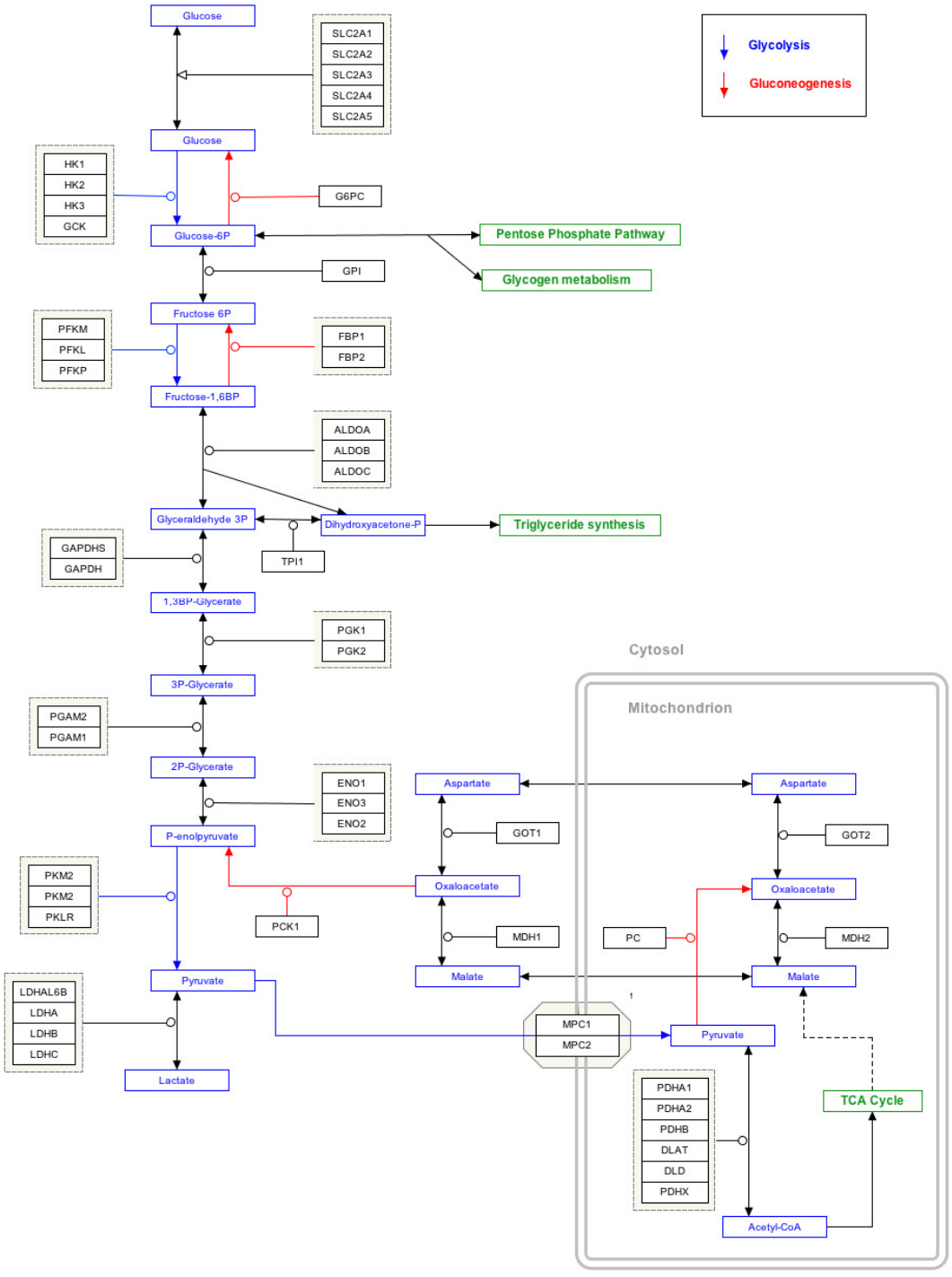
Use of glucose as an energy source in cells is by either aerobic respiration, anaerobic respiration, or fermentation. The first step of glycolysis is the phosphorylation of glucose by a hexokinase to form glucose 6-phosphate. The main reason for the immediate phosphorylation of glucose is to prevent its diffusion out of the cell as the charged phosphate group prevents glucose 6-phosphate from easily crossing the cell membrane. Furthermore, addition of the high-energy phosphate group activates glucose for subsequent breakdown in later steps of glycolysis. At physiological conditions, this initial reaction is irreversible.

In anaerobic respiration, one glucose molecule produces a net gain of two ATP molecules (four ATP molecules are produced during glycolysis through substrate-level phosphorylation, but two are required by enzymes used during the process). In aerobic respiration, a molecule of glucose is much more profitable in that a maximum net production of 30 or 32 ATP molecules (depending on the organism) through oxidative phosphorylation is generated.

Click on genes, proteins and metabolites below to link to respective articles.

[[File:

Title: Glycolysis and Gluconeogenesis
Availability: CC BY 2.0
Last modified: 2/21/2013
Organism: Homo sapiens



[[

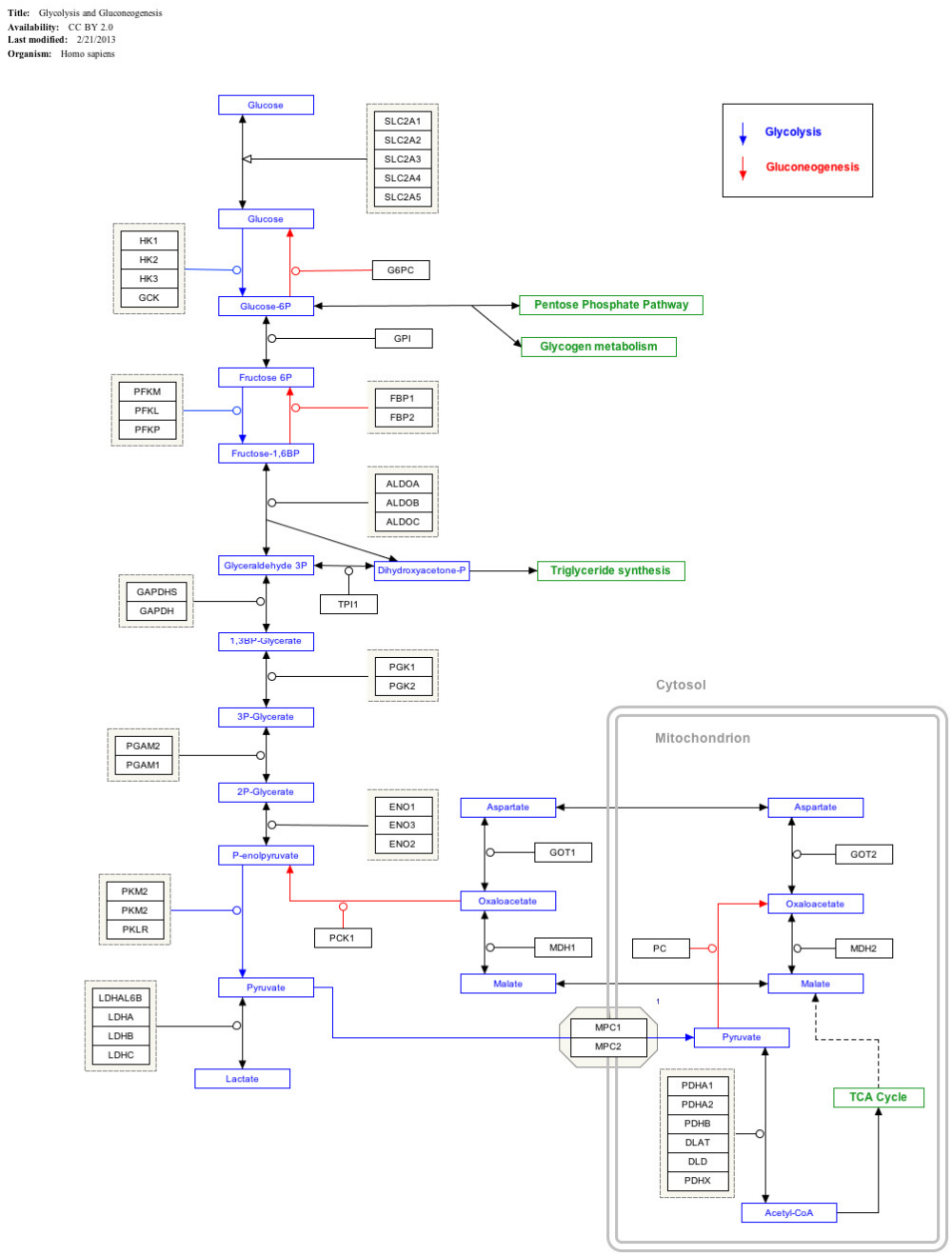
]]

[[

]]

[[
]]
[[
]]
[[
]]
[[
]]
[[
]]
[[
]]
[[
]]
[[
]]
[[
]]
[[
]]
[[
]]
[[
]]
[[
]]
[[
]]
[[
]]
[[
]]
[[
]]
[[
]]
[[
]]
[[
]]
[[
]]
[[
]]

[[
]]
[[
]]
[[
]]
[[
]]
[[
]]
[[
]]
[[
]]
[[
]]
[[
]]
[[
]]
[[
]]
[[
]]
[[
]]
[[
]]
[[
]]
[[
]]
[[
]]
[[
]]
[[
]]
[[
]]
[[
]]
[[
]]



|alt=Glycolysis and Gluconeogenesis edit]]

Glycolysis and Gluconeogenesis edit

The interactive pathway map can be edited at WikiPathways: *"GlycolysisGluconeogenesis_WP534"*.

Tumor cells often grow comparatively quickly and consume an above-average amount of glucose by glycolysis, which leads to the formation of lactate, the end product of fermentation in

mammals, even in the presence of oxygen. This effect is called the Warburg effect. For the increased uptake of glucose in tumors various SGLT and GLUT are overly produced.

In yeast, ethanol is fermented at high glucose concentrations, even in the presence of oxygen (which normally leads to respiration but not to fermentation). This effect is called the Crabtree effect.

Energy source

Glucose is a ubiquitous fuel in biology. It is used as an energy source in organisms, from bacteria to humans, through either aerobic respiration, anaerobic respiration (in bacteria), or fermentation. Glucose is the human body's key source of energy, through aerobic respiration, providing about 3.75 kilocalories (16 kilojoules) of food energy per gram. Breakdown of carbohydrates (e.g., starch) yields mono- and disaccharides, most of which is glucose. Through glycolysis and later in the reactions of the citric acid cycle and oxidative phosphorylation, glucose is oxidized to eventually form carbon dioxide and water, yielding energy mostly in the form of ATP. The insulin reaction, and other mechanisms, regulate the concentration of glucose in the blood. The physiological caloric value of glucose, depending on the source, is 16.2 kilojoules per gram and 15.7 kJ/g (3.74 kcal/g), respectively. The high availability of carbohydrates from plant biomass has led to a variety of methods during evolution, especially in microorganisms, to utilize the energy and carbon storage glucose. Differences exist in which end product can no longer be used for energy production. The presence of individual genes, and their gene products, the enzymes, determine which

reactions are possible. The metabolic pathway of glycolysis is used by almost all living beings. An essential difference in the use of glycolysis is the recovery of NADPH as a reductant for anabolism that would otherwise have to be generated indirectly.

Glucose and oxygen supply almost all the energy for the brain, so its availability influences psychological processes. When glucose is low, psychological processes requiring mental effort (e.g., self-control, effortful decision-making) are impaired. In the brain, which is dependent on glucose and oxygen as the major source of energy, the glucose concentration is usually 4 to 6 mM (5 mM equals 90 mg/dL), but decreases to 2 to 3 mM when fasting.Confusion occurs below 1 mM and coma at lower levels.

The glucose in the blood is calledblood sugar. Blood sugar levels are regulated by glucose-binding nerve cells in the hypothalamus. In addition, glucose in the brain binds to glucose receptors of the reward system in the nucleus accumbens. The binding of glucose to the sweet receptor on the tongue induces a release of various hormones of energy metabolism, either through glucose or through other sugars, leading to an increased cellular uptake and lower blood sugar levels.Artificial sweeteners do not lower blood sugar levels.

The blood sugar content of a healthy person in the short-time fasting state, e.g. after overnight fasting, is about 70 to 100 mg/dL of blood (4 to 5.5 mM). In blood plasma, the measured values are about 10–15% higher. In addition, the values in the arterial blood are higher than the concentrations in the venous blood since glucose is absorbed into the tissue

during the passage of the capillary bed. Also in the capillary blood, which is often used for blood sugar determination, the values are sometimes higher than in the venous blood. The glucose content of the blood is regulated by the hormones insulin, incretin and glucagon. Insulin lowers the glucose level, glucagon increases it. Furthermore, the hormones adrenaline, thyroxine, glucocorticoids, somatotropin and adrenocorticotropin lead to an increase in the glucose level. There is also a hormone-independent regulation, which is referred to as glucose autoregulation. After food intake the blood sugar concentration increases. Values over 180 mg/dL in venous whole blood are pathological and are termed hyperglycemia, values below 40 mg/dL are termed hypoglycaemia. When needed, glucose is released into the bloodstream by glucose-6-phosphatase from glucose-6-phosphate originating from liver and kidney glycogen, thereby regulating the homeostasis of blood glucose concentration. In ruminants, the blood glucose concentration is lower (60 mg/dL in cattle and 40 mg/dL in sheep), because the carbohydrates are converted more by their gut flora into short-chain fatty acids.

Some glucose is converted to lactic acid by astrocytes, which is then utilized as an energy source by brain cells; some glucose is used by intestinal cells and red blood cells, while the rest reaches the liver, adipose tissue and muscle cells, where it is absorbed and stored as glycogen (under the influence of insulin). Liver cell glycogen can be converted to glucose and returned to the blood when insulin is low or absent; muscle cell glycogen is not returned to the blood because of a lack of enzymes. In fat cells, glucose is used to power reactions that synthesize some fat types and have other purposes. Glycogen is

the body's "glucose energy storage" mechanism, because it is much more "space efficient" and less reactive than glucose itself.

As a result of its importance in human health, glucose is an analyte in glucose tests that are common medical blood tests. Eating or fasting prior to taking a blood sample has an effect on analyses for glucose in the blood; a high fasting glucose blood sugar level may be a sign of prediabetes or diabetes mellitus.

The glycemic index is an indicator of the speed of resorption and conversion to blood glucose levels from ingested carbohydrates, measured as the area under the curve of blood glucose levels after consumption in comparison to glucose (glucose is defined as 100). The clinical importance of the glycemic index is controversial, as foods with high fat contents slow the resorption of carbohydrates and lower the glycemic index, e.g. ice cream. An alternative indicator is the insulin index, measured as the impact of carbohydrate consumption on the blood insulin levels. The glycemic load is an indicator for the amount of glucose added to blood glucose levels after consumption, based on the glycemic index and the amount of consumed food.

Precursor

Organisms use glucose as a precursor for the synthesis of several important substances. Starch, cellulose, and glycogen ("animal starch") are common glucose polymers (polysaccharides). Some of these polymers (starch or glycogen) serve as energy stores, while others (cellulose and chitin,

which is made from a derivative of glucose) have structural roles. Oligosaccharides of glucose combined with other sugars serve as important energy stores. These include lactose, the predominant sugar in milk, which is a glucose-galactose disaccharide, and sucrose, another disaccharide which is composed of glucose and fructose. Glucose is also added onto certain proteins and lipids in a process called glycosylation. This is often critical for their functioning. The enzymes that join glucose to other molecules usually use phosphorylated glucose to power the formation of the new bond by coupling it with the breaking of the glucose-phosphate bond.

Other than its direct use as a monomer, glucose can be broken down to synthesize a wide variety of other biomolecules. This is important, as glucose serves both as a primary store of energy and as a source of organic carbon. Glucose can be broken down and converted into lipids. It is also a precursor for the synthesis of other important molecules such as vitamin C (ascorbic acid). In living organisms, glucose is converted to several other chemical compounds that are the starting material for various metabolic pathways. Among them, all other monosaccharides such as fructose (via the polyol pathway), mannose (the epimer of glucose at position 2), galactose (the epimer at position 4), fucose, various uronic acids and the amino sugars are produced from glucose. In addition to the phosphorylation to glucose-6-phosphate, which is part of the glycolysis, glucose can be oxidized during its degradation to glucono-1,5-lactone. Glucose is used in some bacteria as a building block in the trehalose or the dextran biosynthesis and in animals as a building block of glycogen. Glucose can also be converted from bacterial xylose isomerase to fructose. In addition, glucose metabolites produce all

nonessential amino acids, sugar alcohols such as mannitol and sorbitol, fatty acids, cholesterol and nucleic acids. Finally, glucose is used as a building block in the glycosylation of proteins to glycoproteins, glycolipids, peptidoglycans, glycosides and other substances (catalyzed by glycosyltransferases) and can be cleaved from them by glycosidases.

Pathology

Diabetes

Diabetes is a metabolic disorder where the body is unable to regulate levels of glucose in the blood either because of a lack of insulin in the body or the failure, by cells in the body, to respond properly to insulin. Each of these situations can be caused by persistently high elevations of blood glucose levels, through pancreatic burnout and insulin resistance. The pancreas is the organ responsible for the secretion of the hormones insulin and glucagon. Insulin is a hormone that regulates glucose levels, allowing the body's cells to absorb and use glucose. Without it, glucose cannot enter the cell and therefore cannot be used as fuel for the body's functions. If the pancreas is exposed to persistently high elevations of blood glucose levels, the insulin-producing cells in the pancreas could be damaged, causing a lack of insulin in the body. Insulin resistance occurs when the pancreas tries to produce more and more insulin in response to persistently elevated blood glucose levels. Eventually, the rest of the body becomes resistant to the insulin that the pancreas is producing, thereby requiring more insulin to achieve the same blood glucose-

lowering effect, and forcing the pancreas to produce even more insulin to compete with the resistance. This negative spiral contributes to pancreatic burnout, and the disease progression of diabetes.

To monitor the body's response to blood glucose-lowering therapy, glucose levels can be measured. Blood glucose monitoring can be performed by multiple methods, such as the fasting glucose test which measures the level of glucose in the blood after 8 hours of fasting. Another test is the 2-hour glucose tolerance test (GTT) – for this test, the person has a fasting glucose test done, then drinks a 75-gram glucose drink and is retested. This test measures the ability of the person's body to process glucose. Over time the blood glucose levels should decrease as insulin allows it to be taken up by cells and exit the blood stream.

Hypoglycemia management

Individuals with diabetes or other conditions that result in low blood sugar often carry small amounts of sugar in various forms. One sugar commonly used is glucose, often in the form of glucose tablets (glucose pressed into a tablet shape sometimes with one or more other ingredients as a binder), hard candy, or sugar packet.

Commercial production

Glucose is produced industrially from starch by enzymatichydrolysis using glucose amylase or by the use of acids. The enzymatic hydrolysis has largely displaced the acid-catalyzed hydrolysis. The result is glucose syrup (enzymatically

with more than 90% glucose in the dry matter) with an annual worldwide production volume of 20 million tonnes (as of 2011). This is the reason for the former common name "starch sugar". The amylases most often come from *Bacillus licheniformis* or *Bacillus subtilis* (strain MN-385), which are more thermostable than the originally used enzymes. Starting in 1982, pullulanases from *Aspergillus niger* were used in the production of glucose syrup to convert amylopectin to starch (amylose), thereby increasing the yield of glucose. The reaction is carried out at a pH = 4.6–5.2 and a temperature of 55–60 °C.Corn syrup has between 20% and 95% glucose in the dry matter. The Japanese form of the glucose syrup, Mizuame, is made from sweet potato or rice starch.Maltodextrin contains about 20% glucose.

Many crops can be used as the source of starch. Maize, rice,wheat,cassava,potato,barley, sweet potato,corn husk and sago are all used in various parts of the world. In the United States, corn starch (from maize) is used almost exclusively. Some commercial glucose occurs as a component of invert sugar, a roughly 1:1 mixture of glucose and fructose that is produced from sucrose. In principle, cellulose could be hydrolysed to glucose, but this process is not yet commercially practical.

Conversion to fructose

In the USA almost exclusively corn (more precisely: corn syrup) is used as glucose source for the production of isoglucose, which is a mixture of glucose and fructose, since fructose has a higher sweetening power — with same physiological calorific value of 374 kilocalories per 100 g. The annual world

production of isoglucose is 8 million tonnes (as of 2011). When made from corn syrup, the final product is high fructose corn syrup (HFCS).

Commercial usage

Glucose is mainly used for the production of fructose and in the production of glucose-containing foods. In foods, it is used as a sweetener, humectant, to increase the volume and to create a softer mouthfeel. Various sources of glucose, such as grape juice (for wine) or malt (for beer), are used for fermentation to ethanol during the production of alcoholic beverages. Most soft drinks in the US use HFCS-55 (with a fructose content of 55% in the dry mass), while most other HFCS-sweetened foods in the US use HFCS-42 (with a fructose content of 42% in the dry mass). In the neighboring country Mexico, on the other hand, cane sugar is used in the soft drink as a sweetener, which has a higher sweetening power. In addition, glucose syrup is used, inter alia, in the production of confectionery such as candies, toffee and fondant. Typical chemical reactions of glucose when heated under water-free conditions are the caramelization and, in presence of amino acids, the maillard reaction.

In addition, various organic acids can be biotechnologically produced from glucose, for example by fermentation with *Clostridium thermoaceticum* to produce acetic acid, with *Penicilium notatum* for the production of araboascorbic acid, with *Rhizopus delemar* for the production of fumaric acid, with *Aspergillus niger* for the production of gluconic acid, with *Candida brumptii* to produce isocitric acid, with *Aspergillus terreus* for the production of itaconic acid, with *Pseudomonas*

fluorescens for the production of 2-ketogluconic acid, with *Gluconobacter suboxydans* for the production of 5-ketogluconic acid, with *Aspergillus oryzae* for the production of kojic acid, with *Lactobacillus delbrueckii* for the production of lactic acid, with *Lactobacillus brevis* for the production of malic acid, with *Propionibacter shermanii* for the production of propionic acid, with *Pseudomonas aeruginosa* for the production of pyruvic acid and with *Gluconobacter suboxydans* for the production of tartaric acid. Potent, bioactive natural products like triptolide that inhibit mammalian transcription via inhibition of the XPB subunit of the general transcription factor TFIIH has been recently reported as a glucose conjugate for targeting hypoxic cancer cells with increased glucose transporter expression.

Analysis

Specifically, when a glucose molecule is to be detected at a certain position in a larger molecule, nuclear magnetic resonance spectroscopy, X-ray crystallography analysis or lectinimmunostaining is performed with concanavalin A reporter enzyme conjugate (that binds only glucose or mannose).

Classical qualitative detection reactions

These reactions have only historical significance:

Fehling test

The Fehling test is a classic method for the detection of aldoses. Due to mutarotation, glucose is always present to a small extent as an open-chain aldehyde. By adding the Fehling

reagents (Fehling (I) solution and Fehling (II) solution), the aldehyde group is oxidized to a carboxylic acid, while the Cu tartrate complex is reduced to Cu and forms a brick red precipitate (Cu_2O).

Tollens test

In the Tollens test, after addition of ammoniacal $AgNO_3$ to the sample solution, Ag is reduced by glucose to elemental silver.

Barfoed test

In Barfoed's test, a solution of dissolved copper acetate, sodium acetate and acetic acid is added to the solution of the sugar to be tested and subsequently heated in a water bath for a few minutes. Glucose and other monosaccharides rapidly produce a reddish color and reddish brown copper(I) oxide (Cu_2O).

Nylander's test

As a reducing sugar, glucose reacts in the Nylander's test.

Other tests

Upon heating a dilute potassium hydroxide solution with glucose to 100 °C, a strong reddish browning and a caramel-like odor develops. Concentrated sulfuric acid dissolves dry glucose without blackening at room temperature forming sugar sulfuric acid. In a yeast solution, alcoholic fermentation produces carbon dioxide in the ratio of 2.0454 molecules of glucose to one molecule of CO_2. Glucose forms a black mass

with stannous chloride. In an ammoniacal silver solution, glucose (as well as lactose and dextrin) leads to the deposition of silver. In an ammoniacal lead acetate solution, white lead glycosideis formed in the presence of glucose, which becomes less soluble on cooking and turns brown. In an ammoniacal copper solution, yellow copper oxide hydrate is formed with glucose at room temperature, while red copper oxide is formed during boiling (same with dextrin, except for with an ammoniacal copper acetate solution). With Hager's reagent, glucose forms mercury oxide during boiling. An alkaline bismuth solution is used to precipitate elemental, black-brown bismuth with glucose. Glucose boiled in an ammonium molybdate solution turns the solution blue. A solution with indigo carmine and sodium carbonate destains when boiled with glucose.

Instrumental quantification

Refractometry and polarimetry

In concentrated solutions of glucose with a low proportion of other carbohydrates, its concentration can be determined with a polarimeter.

For sugar mixtures, the concentration can be determined with a refractometer, for example in the Oechsle determination in the course of the production of wine.

Photometric enzymatic methods in solution

The enzyme glucose oxidase (GOx) converts glucose into gluconic acid and hydrogen peroxide while consuming oxygen.

Another enzyme, peroxidase, catalyzes a chromogenic reaction (Trinder reaction) of phenol with 4-aminoantipyrine to a purple dye.

Photometric test-strip method

The test-strip method employs the above-mentioned enzymatic conversion of glucose to gluconic acid to form hydrogen peroxide. The reagents are immobilised on a polymer matrix, the so-called test strip, which assumes a more or less intense color. This can be measured reflectometrically at 510 nm with the aid of an LED-based handheld photometer. This allows routine blood sugar determination by laymen. In addition to the reaction of phenol with 4-aminoantipyrine, new chromogenic reactions have been developed that allow photometry at higher wavelengths (550 nm, 750 nm).

Amperometric glucose sensor

The electroanalysis of glucose is also based on the enzymatic reaction mentioned above. The produced hydrogen peroxide can be amperometrically quantified by anodic oxidation at a potential of 600 mV. The GOx is immobilised on the electrode surface or in a membrane placed close to the electrode. Precious metals such as platinum or gold are used in electrodes, as well as carbon nanotube electrodes, which e.g. are doped with boron. Cu–CuO nanowires are also used as enzyme-free amperometric electrodes. This way a detection limit of 50 μmol/L has been achieved. A particularly promising method is the so-called "enzyme wiring". In this case, the electron flowing during the oxidation is transferred directly from the enzyme via a molecular wire to the electrode.

Other sensory methods

There are a variety of other chemical sensors for measuring glucose. Given the importance of glucose analysis in the life sciences, numerous optical probes have also been developed for saccharides based on the use of boronic acids, which are particularly useful for intracellular sensory applications where other (optical) methods are not or only conditionally usable. In addition to the organic boronic acid derivatives, which often bind highly specifically to the 1,2-diol groups of sugars, there are also other probe concepts classified by functional mechanisms which use selective glucose-binding proteins (e.g. concanavalin A) as a receptor. Furthermore, methods were developed which indirectly detect the glucose concentration via the concentration of metabolised products, e.g. by the consumption of oxygen using fluorescence-optical sensors. Finally, there are enzyme-based concepts that use the intrinsic absorbance or fluorescence of (fluorescence-labeled) enzymes as reporters.

Copper iodometry

Glucose can be quantified by copper iodometry.

Chromatographic methods

In particular, for the analysis of complex mixtures containing glucose, e.g. in honey, chromatographic methods such as high performance liquid chromatography and gas chromatography are often used in combination with mass spectrometry. Taking into account the isotope ratios,it is also possible to reliably detect honey adulteration by added sugars with these methods.

Derivatisation using silylation reagents is commonly used.Also, the proportions of di- and trisaccharides can be quantified.

In vivo analysis

Glucose uptake in cells of organisms is measured with 2-deoxy-D-glucose or fluorodeoxyglucose. (F)fluorodeoxyglucose is used as a tracer in positron emission tomography in oncology and neurology, where it is by far the most commonly used diagnostic agent.

Ketosis

Ketosis is a metabolic state characterized by elevated levels of ketone bodies in the blood or urine. Physiologic ketosis is a normal response to low glucose availability, such as low-carbohydrate diets or fasting, that provides an additional energy source for the brain in the form of ketones. In physiologic ketosis, ketones in the blood are elevated above baseline levels, but the body's acid–base homeostasisis maintained. This contrasts with ketoacidosis, an uncontrolled production of ketones that occurs in pathologic states and causes a metabolic acidosis, which is a medical emergency. Ketoacidosis is most commonly the result of complete insulin deficiency in type 1 diabetes or late-stage type 2 diabetes. Ketone levels can be measured in blood, urine or breath and are generally between 0.5 and 3.0 millimolar (mM) in physiologic ketosis, while ketoacidosis may cause blood concentrations greater than 10 mM.

Trace levels of ketones are always present in the blood and increase when blood glucose reserves are low and the liver

shifts from primarily metabolizing carbohydrates to metabolizing fatty acids. This occurs during states of increased fatty acid oxidation such as fasting, starvation, carbohydrate restriction, or prolonged exercise. When the liver rapidly metabolizes fatty acids into acetyl-CoA, some acetyl-CoA molecules can then be converted into ketone bodies: acetoacetate, beta-hydroxybutyrate, and acetone. These ketone bodies can function as an energy source as well as signalling molecules. The liver itself cannot utilize these molecules for energy, so the ketone bodies are released into the blood for use by peripheral tissues including the brain.

When ketosis is induced by carbohydrate restriction, it is sometimes referred to as nutritional ketosis. A low-carbohydrate, moderate protein diet that can lead to ketosis is called a ketogenic diet. Ketosis is well-established as a treatment for epilepsy and is also effective in treating type 2 diabetes. The possible effect on a range of neurological diseases, metabolic syndrome, cancer, and other conditions is currently under investigation.

Definition

Physiologic ketosis

Physiologic ketosis is a physiologic state characterized by elevated serum ketones and normal blood glucose and blood pH. Increasing production of ketone bodies is a response to low glucose availability that creates an alternate energy source for the brain. Physiologic ketosis can result from any state that increases fatty acid oxidation including fasting, prolonged

exercise, or very low-carbohydrate diets such as the ketogenic diet. When physiologic ketosis is induced by carbohydrate restriction, it is sometimes referred to as nutritional ketosis. Ketone levels generally remain below 3 mM.

Ketoacidosis

Ketoacidosis is a pathological state of uncontrolled production of ketones that results in a metabolic acidosis. Ketoacidosis is most commonly caused by a deficiency of insulin in type 1 diabetes or late stage type 2 diabetes but can also be the result of chronic heavy alcohol use, salicylate poisoning, or isopropyl alcohol ingestion. Ketoacidosis causes significant metabolic derangements and is a life-threatening medical emergency. Ketoacidosis is distinct from physiologic ketosis as it requires failure of the normal regulation of ketone body production.

Causes

Elevated blood ketone levels are most often caused by accelerated ketone production but may also be caused by consumption of exogenous ketones or precursors.

When glycogen and blood glucose reserves are low, a metabolic shift occurs in order to save glucose for the brain which is unable to use fatty acids for energy. This shift involves increasing fatty acid oxidation and production of ketones in the liver as an alternate energy source for the brain as well as the skeletal muscles, heart, and kidney. Low levels of ketones are always present in the blood and increase under circumstances of low glucose availability. For example, after an

overnight fast, 2-6% of energy comes from ketones and this increases to 30-40% after a 3-day fast.

The amount of carbohydrate restriction required to induce a state of ketosis is variable and depends on activity level, insulin sensitivity, genetics, age and other factors, but ketosis will usually occur when consuming less than 50 grams of carbohydrates per day for at least three days.

Neonates, pregnant women and lactating women are populations that develop physiologic ketosis especially rapidly in response to energetic challenges such as fasting or illness. This can progress to ketoacidosis in the setting of illness, although it occurs rarely. Propensity for ketone production in neonates is caused by their high-fat breast milk diet, disproportionately large central nervous system and limited liver glycogen.

Biochemistry

The precursors of ketone bodies include fatty acids from adipose tissue or the diet and ketogenic amino acids. The formation of ketone bodies occurs via ketogenesis in the mitochondrial matrix of liver cells.

Fatty acids can be released from adipose tissue by adipokine signaling of high glucagon and epinephrine levels and low insulin levels. High glucagon and low insulin correspond to times of low glucose availability such as fasting. Fatty acids bound to coenzyme A allow penetration into mitochondria. Once inside the mitochondrion, the bound fatty acids are used as fuel in cells predominantly through beta oxidation, which

cleaves two carbons from the acyl-CoA molecule in every cycle to form acetyl-CoA. Acetyl-CoA enters the citric acid cycle, where it undergoes an aldol condensation with oxaloacetate to form citric acid; citric acid then enters the tricarboxylic acid cycle (TCA), which harvests a very high energy yield per carbon in the original fatty acid.

Acetyl-CoA can be metabolized through the TCA cycle in any cell, but it can also undergo ketogenesis in the mitochondria of liver cells. When glucose availability is low, oxaloacetate is diverted away from the TCA cycle and is instead used to produce glucose via gluconeogenesis. This utilization of oxaloacetate in gluconeogenesis can make it unavailable to condense with acetyl-CoA, preventing entrance into the TCA cycle. In this scenario, energy can be harvested from acetyl-CoA through ketone production.

In ketogenesis, two acetyl-CoA molecules condense to form acetoacetyl-CoA via thiolase. Acetoacetyl-CoA briefly combines with another acetyl-CoA via HMG-CoA synthase to form hydroxy-β-methylglutaryl-CoA. Hydroxy-β-methylglutaryl-CoA form the ketone body acetoacetate via HMG-CoA lyase. Acetoacetate can then reversibly convert to another ketone body—D-β-hydroxybutyrate—via D-β-hydroxybutyrate dehydrogenase. Alternatively, acetoacetate can spontaneously degrade to a third ketone body (acetone) and carbon dioxide, which generates much greater concentrations of acetoacetate and D-β-hydroxybutyrate. The resulting ketone bodies cannot be used for energy by the liver so are exported from the liver to supply energy to the brain and peripheral tissues.

In addition to fatty acids, deaminated ketogenic amino acidscan also be converted into intermediates in the citric acid cycle and produce ketone bodies.

Measurement

Ketone levels can be measured by testing urine, blood or breath. There are limitations in directly comparing these methods as they measure different ketone bodies.

Urine testing

Urine testing is the most common method of testing for ketones. Urine test strips utilize a nitroprusside reaction with acetoacetate to give a semi-quantitative measure based on color change of the strip. Although beta-hydroxybutyrate is the predominant circulating ketone, urine test strips only measure acetoacetate.

Urinary ketones often correlate poorly with serum levels because of variability in excretion of ketones by the kidney, influence of hydration status, and renal function.

Serum testing

Finger-stick ketone meters allow instant testing of beta-hydroxybutyrate levels in the blood, similar to glucometers. Beta-hydroxybutrate levels in blood can also be measured in a laboratory.

Medical uses

Epilepsy

Ketosis induced by a ketogenic diet is a long-accepted treatment for refractory epilepsy. It was first used in the 1920s and is now widely implemented for pediatric and adult patients.

Type 2 diabetes

Ketosis induced by a very low-carbohydrate diet is supported as an effective treatment for type 2 diabetes and can improve blood glucose metrics even with reduction or discontinuation of antidiabetic medications.

Ketosis can be an effective treatment for this population by reducing dietary glucose load, increasing insulin sensitivity as well as reducing hepatic glucose output. This results in reduction in fasting glucose and insulin and hemoglobin A1c as well as reduced need for exogenous insulin.

Obesity and metabolic syndrome

Ketosis can improve markers of metabolic syndrome through reduction in serum triglycerides, elevation in high-density lipoprotein (HDL) as well as increased size and volume of low-density lipoprotein (LDL) particles. These changes are consistent with an improved lipid profile despite potential increases in total cholesterol level.

Research

Ketosis is being investigated for a growing number of conditions, however clinical recommendations for these conditions cannot yet be made based on the current level of evidence.

Neurological disorders: In addition to its use for epilepsy, ketosis is being investigated in other neurological disorders because of its proposed neuroprotective effects including Alzheimer's disease, amyotrophic lateral sclerosis (ALS), autism, migraine headache, neurotrauma, pain, Parkinson's disease, and sleep disorders.

Cancer: Preclinical studies have indicated ketosis may have anti-tumor effects, although clinical trials have been limited by small sample sizes and have not shown conclusive benefit.

Glycogenosis Ketosis has been reported to alleviate symptoms related to glycogenosis as in some glycogen storage diseases. GSDs which are characterised by the inability of tissues to utilise glycogen stores, such as McArdle disease, may be managed with a ketogenic diet by ensuring that tissues are using ketone bodies for energy and not the impaired pathways of glycogen utilisation for glycolysis.

Other conditions: There is emerging evidence for the use of ketosis in other conditions such as type 1 diabetes, non-alcoholic fatty liver disease, acne, polycystic kidney disease and polycystic ovary syndrome, although evidence quality is limited by small sample sizes.

Safety

The safety of ketosis from low-carbohydrate diets is often called into question by clinicians, researchers and the media. A common safety concern stems from the misunderstanding of the difference between physiologic ketosis and pathologic ketoacidosis. There is also continued debate whether chronic ketosis is a healthy state or a stressor to be avoided. Some argue that humans evolved to avoid ketosis and should not be in ketosis long-term. The counter-argument is that there is no physiologic requirement for dietary carbohydrate as adequate energy can be made via gluconeogenesis and ketogenesis indefinitely.

Alternatively, the switching between a ketotic and fed state has been proposed to have beneficial effects on metabolic and neurologic health. The effects of sustaining ketosis for up to two years are known from studies of people following a strict ketogenic diet for epilepsy or type 2 diabetes; these include short-term adverse effects leading to potential long-term ones. However, literature on longer term effects of intermittent ketosis is lacking.

Medication considerations

Some medications require attention when in a state of ketosis, especially several classes of diabetes medication. SGLT2 inhibitor medications have been associated with cases of euglycemic ketoacidosis- a rare state of high ketones causing a metabolic acidosis with normal blood glucose levels. This usually occurs with missed insulin doses, illness, dehydration

or adherence to a low-carbohydrate diet while taking the medication.Additionally, medications used to directly lower blood glucose including insulin and sulfonylureas may cause hypoglycemia if they are not titrated prior to starting a diet that results in ketosis.

Adverse effects

The most common side effects of ketosis include headache, fatigue, dizziness, insomnia, difficulty in exercise tolerance, constipation, and nausea, especially in the first days and weeks after starting a ketogenic diet. Breath may develop a sweet, fruity flavor via production of acetone that is exhaled because of its high volatility.

Most adverse effects of long-term ketosis reported are in children because of its longstanding acceptance as a treatment for pediatric epilepsy. These include compromised bone health, stunted growth, hyperlipidemia, and kidney stones.

Contraindications

Ketosis induced by a ketogenic diet should not be pursued by people with pancreatitis because of the high dietary fat content. Ketosis is also contraindicated in pyruvate carboxylase deficiency, porphyria, and other rare genetic disorders of fat metabolism.

Veterinary medicine

In dairy cattle, ketosis commonly occurs during the first weeks after giving birth to a calf and is sometimes referred to as

acetonemia. This is the result of an energy deficit when intake is inadequate to compensate for the increased metabolic demand of lactating. The elevated β-hydroxybutyrate concentrations can depress gluconeogenesis, feed intake and the immune system, as well as have an impact on milk composition. Point of care diagnostic tests can be useful to screen for ketosis in cattle.

In sheep, ketosis, evidenced by hyperketonemia with beta-hydroxybutyrate in blood over 0.7 mmol/L, is referred to as *pregnancy toxemia*. This may develop in late pregnancy in ewes bearing multiple fetuses and is associated with the considerable metabolic demands of the pregnancy. In ruminants, because most glucose in the digestive tract is metabolized by rumen organisms, glucose must be supplied by gluconeogenesis. Pregnancy toxemia is most likely to occur in late pregnancy due to metabolic demand from rapid fetal growth and may be triggered by insufficient feed energy intake due to weather conditions, stress or other causes. Prompt recovery may occur with natural parturition, Caesarean section or induced abortion. Prevention through appropriate feeding and other management is more effective than treatment of advanced stages of pregnancy toxemia.

Major histocompatibility complex

The **major histocompatibility complex** (**MHC**) is a large locus on vertebrate DNA containing a set of closely linked polymorphic genes that code for cell surface proteins essential for the adaptive immune system. These cell surface proteins are called**MHC molecules**.

This locus got its name because it was discovered via the study of transplanted tissue compatibility. Later studies revealed that tissue rejection due to incompatibility is an experimental artifact masking the real function of MHC molecules: binding an antigen derived from self-proteins, or from pathogens, and bringing the antigen presentation to the cell surface for recognition by the appropriate T-cells. MHC molecules mediate the interactions of leukocytes, also called white blood cells (WBCs), with other leukocytes or with body cells. The MHC determines donor compatibility for organ transplant, as well as one's susceptibility to autoimmune diseases via cross-reacting immunization.

In a cell, protein molecules of the host's own phenotype or of other biologic entities are continually synthesized and degraded. Each MHC molecule on the cell surface displays a small peptide (a molecular fraction of a protein) called an epitope. The presented self-antigens prevent an organism's immune system from targeting its own cells. The presentation of pathogen-derived proteins results in the elimination of the infected cell by the immune system.

Diversity of an individual's self-antigen presentation, mediated by MHC self-antigens, is attained in at least three ways: (1) an organism's MHC repertoire is polygenic (via multiple, interacting genes); (2) MHC expression is codominant (from both sets of inherited alleles); (3) MHC gene variants are highly polymorphic (diversely varying from organism to organism within a species).Sexual selectionhas been observed in male mice making mate choices of females with different MHCs and thus demonstrating sexual selection. Also, at least for MHC I presentation, there has been evidence of antigenic peptide

splicing, which can combine peptides from different proteins, vastly increasing antigen diversity.

Discovery

The first descriptions of the MHC were made by British immunologistPeter Gorer in 1936. MHC genes were first identified in inbred mice strains. Clarence Little transplanted tumors across differing strains and found rejection of transplanted tumors according to strains of host versus donor.

George Snell selectively bred two mouse strains, attained a new strain nearly identical to one of the progenitor strains, but differing crucially in histocompatibility—that is, tissue compatibility upon transplantation—and thereupon identified an MHC locus. Later Jean Dausset demonstrated the existence of MHC genes in humans and described the first human leucocyte antigen, the protein which we call now HLA-A2.

Some years later Baruj Benacerraf showed that polymorphic MHC genes not only determine an individual's unique constitution of antigens but also regulate the interaction among the various cells of the immunological system. These three scientists have been awarded the 1980 Nobel Prize in Physiology or Medicine for their discoveries concerning "genetically determined structures on the cell surface that regulate immunological reactions".

The first fully sequenced and annotated MHC was published for humans in 1999 by a consortium of sequencing centers from the UK, USA and Japan in *Nature*. It was a "virtual MHC" since it was a mosaic from different individuals. A much shorter

MHC locus from chickens was published in the same issue of *Nature*. Many other species have been sequenced and the evolution of the MHC was studied, e.g. in the gray short-tailed opossum (*Monodelphis domestica*), a marsupial, MHC spans 3.95 Mb, yielding 114 genes, 87 shared with humans. Marsupial MHC genotypic variation lies between eutherian mammals and birds, taken as the minimal MHC encoding, but is closer in organization to that of nonmammals. The IPD-MHC Databasewas created which provides a centralised repository for sequences of the Major Histocompatibility Complex (MHC) from a number of different species. The database contains 77 species for the release from 2019-12-19.

Genes

The MHC locus is present in all jawed vertebrates, it is assumed to have arisen about 450 million years ago. Despite the difference in the number of genes included in the MHC of different species, the overall organization of the locus is rather similar. Usual MHC contains about a hundred genes and pseudogenes, not all of them are involved in immunity. In humans, the MHC region occurs on chromosome 6, between the flanking genetic markers*MOG* and *COL11A2* (from 6p22.1 to 6p21.3 about 29Mb to 33Mb on the hg38 assembly), and contains 224 genes spanning 3.6 megabase pairs (3 600 000 bases). About half have known immune functions. The human MHC is also called the HLA (human leukocyte antigen) complex (often just the HLA). Similarly, there is SLA (Swine leukocyte antigens), BoLA (Bovine leukocyte antigens), DLA for dogs, etc. However, historically, the MHC in miceis called the Histocompatibility system 2 or just the H-2, in rats - RT1, and

in chicken - B-locus. The MHC gene family is divided into three subgroups: MHC class I, MHC class II, and MHC class III. Among all those genes present in MHC, there are two types of genes coding for the proteins MHC class I molecules and MHC class II molecules that directly involved in the antigen presentation. These genes are highly polymorphic, 19031 alleles of class I HLA, and 7183 of class II HLA are deposited for human in the IMGT database.

Proteins

MHC class I

MHC class I molecules are expressed in all nucleated cells and also in platelets—in essence all cells but red blood cells. It presents epitopes to killer T cells, also called cytotoxic T lymphocytes (CTLs). A CTL expresses CD8 receptors, in addition to T-cell receptors (TCR)s. When a CTL's CD8 receptor docks to a MHC class I molecule, if the CTL's TCR fits the epitope within the MHC class I molecule, the CTL triggers the cell to undergo programmed cell death by apoptosis. Thus, MHC class I helps mediate cellular immunity, a primary means to address intracellular pathogens, such as viruses and some bacteria, including bacterial L forms, bacterial genus*Mycoplasma*, and bacterial genus *Rickettsia*. In humans, MHC class I comprises HLA-A, HLA-B, and HLA-C molecules.

The first crystal structure of Class I MHC molecule, human HLA-A2, was published in 1989. The structure revealed that MHC-I molecules are heterodimers, they have polymorphic heavy α-subunit whose gene occurs inside the MHC locus and

small invariant β_2 microglobulin subunit whose gene is located usually outside of it. Polymorphic heavy chain of MHC-I molecule contains N-terminal extra-cellular region composed by three domains, α1, α2, and α3, transmembrane helix to hold MHC-I molecule on the cell surface and short cytoplasmic tail. Two domains, α1 and α2 form deep peptide-binding groove between two long α-helices and the floor of the groove formed by eight β-strands. Immunoglobulin-like domain α3 involved in the interaction with CD8 co-receptor. β_2 microglobulin provides stability of the complex and participates in the recognition of peptide-MHC class I complex by CD8 co-receptor. The peptide is non-covalently bound to MHC-I, it is held by the several pockets on the floor of the peptide-binding groove. Amino acid side-chains that are most polymorphic in human alleles fill up the central and widest portion of the binding groove, while conserved side-chains are clustered at the narrower ends of the groove.

Classical MHC molecules present epitopes to the TCRs of CD8+ T lymphocytes. **Nonclassical molecules** (MHC class IB) exhibit limited polymorphism, expression patterns, and presented antigens; this group is subdivided into a group encoded within MHC loci (e.g., HLA-E, -F, -G), as well as those not (e.g., stress ligands such as ULBPs, Rae1, and H60); the antigen/ligand for many of these molecules remain unknown, but they can interact with each of CD8+ T cells, NKT cells, and NK cells. The evolutionary oldest nonclassical MHC class I lineage in human was deduced to be the lineage that includes the CD1 and PROCR (alias EPCR) molecules, and this lineage may have been established before the origin of tetrapod species.However, the only nonclassical MHC class I lineage for which evidence exists that it was established before the

evolutionary separation of Actinopterygii (ray-finned fish) and Sarcopterygii (lobe-finned fish plus tetrapods) is lineage Z of which members are found, together in each species with classical MHC class I, in lungfish and throughout ray-finned fishes; why the Z lineage was well conserved in ray-finned fish but lost in tetrapods is not understood.

MHC Class II

MHC class II can be conditionally expressed by all cell types, but normally occurs only on "professional" antigen-presenting cells (APCs): macrophages, B cells, and especially dendritic cells (DCs). An APC takes up an antigenic protein, performs antigen processing, and returns a molecular fraction of it—a fraction termed the epitope—and displays it on the APC's surface coupled within an MHC class II molecule (antigen presentation). On the cell's surface, the epitope can be recognized by immunologic structures like T-cell receptors (TCRs). The molecular region which binds to the epitope is the paratope.

On surfaces of helper T cells are CD4 receptors, as well as TCRs. When a naive helper T cell's CD4 molecule docks to an APC's MHC class II molecule, its TCR can meet and bind the epitope coupled within the MHC class II. This event primes the naive T cell. According to the local milieu, that is, the balance of cytokines secreted by APCs in the microenvironment, the naive helper T cell (Th_0) polarizes into either a memory Th cell or an effector Th cell of phenotype either type 1 (Th_1), type 2 (Th_2), type 17 (Th_{17}), or regulatory/suppressor (T_{reg}), as so far identified, the Th cell's terminal differentiation.

MHC class II thus mediates immunization to—or, if APCs polarize Th_0 cells principally to T_{reg} cells, immune tolerance of—an antigen. The polarization during primary exposure to an antigen is key in determining a number of chronic diseases, such as inflammatory bowel diseases and asthma, by skewing the immune response that memory Th cells coordinate when their memory recall is triggered upon secondary exposure to similar antigens. B cells express MHC class II to present antigens to Th_0, but when their B cell receptors bind matching epitopes, interactions which are not mediated by MHC, these activated B cells secrete soluble immunoglobulins: antibody molecules mediating humoral immunity.

Class II MHC molecules are also heterodimers, genes for both α and β subunits are polymorphic and located within MHC class II subregion. Peptide-binding groove of MHC-II molecules is forms by N-terminal domains of both subunits of the heterodimer, α1 and β1, unlike MHC-I molecules, where two domains of the same chain are involved. In addition, both subunits of MHC-II contain transmembrane helix and immunoglobulin domains α2 or β2 that can be recognized by CD4 co-receptors. In this way MHC molecules chaperone which type of lymphocytes may bind to the given antigen with high affinity, since different lymphocytes express different T-Cell Receptor (TCR) co-receptors.

MHC class II molecules in humans have five to six isotypes. **Classical molecules** present peptides to CD4+ lymphocytes. **Nonclassical molecules**, accessories, with intracellular functions, are not exposed on cell membranes, but in internal membranes, assisting with the loading of antigenic peptides onto classic MHC class II molecules. The important

nonclassical MHC class II molecule DM is only found from the evolutionary level of lungfish, although also in more primitive fishes both classical and nonclassical MHC class II are found.

Class III

Class III molecules have physiologic roles unlike classes I and II, but are encoded between them in the short arm of human chromosome 6. Class III molecules include several secreted proteins with immune functions: components of the complement system (such as C2, C4, and B factor), cytokines (such as TNF-α, LTA, and LTB), and heat shock proteins.

Function

MHC is the tissue-antigen that allows the immune system (more specifically T cells) to bind to, recognize, and tolerate itself (autorecognition). MHC is also the chaperone for intracellular peptides that are complexed with MHCs and presented to T cell receptors (TCRs) as potential foreign antigens. MHC interacts with TCR and its co-receptors to optimize binding conditions for the TCR-antigen interaction, in terms of antigen binding affinity and specificity, and signal transduction effectiveness.

Essentially, the MHC-peptide complex is a complex of auto-antigen/allo-antigen. Upon binding, T cells should in principle tolerate the auto-antigen, but activate when exposed to the allo-antigen. Disease states occur when this principle is disrupted.

Antigen presentation: MHC molecules bind to both T cell receptor and CD4/CD8 co-receptors on T lymphocytes, and the antigen epitope held in the peptide-binding groove of the MHC molecule interacts with the variable Ig-Like domain of the TCR to trigger T-cell activation

Autoimmune reaction: Having some MHC molecules increases the risk of autoimmune diseases more than having others. HLA-B27 is an example. It is unclear how exactly having the HLA-B27 tissue type increases the risk of ankylosing spondylitis and other associated inflammatory diseases, but mechanisms involving aberrant antigen presentation or T cell activation have been hypothesized.

Tissue allorecognition: MHC molecules in complex with peptide epitopes are essentially ligands for TCRs. T cells become activated by binding to the peptide-binding grooves of any MHC molecule that they were not trained to recognize during positive selection in the thymus.

In sexual mate selection

MHC molecules enable immune system surveillance of the population of protein molecules in a host cell, and greater MHC diversity permits greater diversity of antigen presentation. In 1976,

Yamazaki *et al* demonstrated a sexual selectionmate choice by male mice for females of a different MHC. Similar results have been obtained with fish. Some data find lower rates of early pregnancy loss in human couples of dissimilar MHC genes.

MHC may be related to mate choice in some human populations, a theory that found support by studies by Ober and colleagues in 1997, as well as by Chaix and colleagues in 2008. However, the latter findings have been controversial. If it exists, the phenomenon might be mediated by olfaction, as MHC phenotype appears strongly involved in the strength and pleasantness of perceived odour of compounds from sweat. Fatty acid esters—such as methyl undecanoate, methyl decanoate, methyl nonanoate, methyl octanoate, and methyl hexanoate—show strong connection to MHC.

In 1995, Claus Wedekind found that in a group of female college students who smelled T-shirts worn by male students for two nights (without deodorant, cologne, or scented soaps), by far most women chose shirts worn by men of dissimilar MHCs, a preference reversed if the women were on oral contraceptives. Results of a 2002 experiment likewise suggest HLA-associated odors influence odor preference and may mediate social cues. In 2005 in a group of 58 subjects, women were more indecisive when presented with MHCs like their own, although with oral contraceptives, the women showed no particular preference. No studies show the extent to which odor preference determines mate selection (or vice versa).

Evolutionary diversity

Most mammals have MHC variants similar to those of humans, who bear great allelic diversity, especially among the nine classical genes—seemingly due largely to gene duplication—though human MHC regions have many pseudogenes. The most diverse loci, namely HLA-A, HLA-B, and HLA-C, have roughly 6000, 7200, and 5800 known alleles, respectively. Many HLA

alleles are ancient, sometimes of closer homologyto a chimpanzee MHC alleles than to some other human alleles of the same gene.

MHC allelic diversity has challenged evolutionary biologists for explanation. Most posit balancing selection (see polymorphism (biology)), which is any natural selection process whereby no single allele is absolutely most fit, such as frequency-dependent selection and heterozygote advantage. Pathogenic coevolution, as a type of balancing selection, posits that common alleles are under greatest pathogenic pressure, driving positive selection of uncommon alleles—moving targets, so to say, for pathogens. As pathogenic pressure on the previously common alleles decreases, their frequency in the population stabilizes, and remain circulating in a large population.Genetic drift is also a major driving force in some species. It is possible that the combined effects of some or all of these factors cause the genetic diversity.

MHC diversity has also been suggested as a possible indicator for conservation, because large, stable populations tend to display greater MHC diversity, than smaller, isolated populations. Small, fragmented populations that have experienced a population bottleneck typically have lower MHC diversity. For example, relatively low MHC diversity has been observed in the cheetah (*Acinonyx jubatus*),Eurasian beaver (*Castor fiber*), and giant panda (*Ailuropoda melanoleuca*). In 2007 low MHC diversity was attributed a role in disease susceptibility in the Tasmanian devil (*Sarcophilus harrisii*), native to the isolated island of Tasmania, such that an antigen of a transmissible tumor, involved in devil facial tumour disease, appears to be recognized as a *self antigen*. To offset

inbreeding, efforts to sustain genetic diversity in populations of endangered species and of captive animals have been suggested.

In ray-finned fish like rainbow trout, allelic polymorphism in MHC class II is reminiscent of that in mammals and predominantly maps to the peptide binding groove. However, in MHC class I of many teleost fishes, the allelic polymorphism is much more extreme than in mammals in the sense that the sequence identity levels between alleles can be very low and the variation extends far beyond the peptide binding groove. It has been speculated that this type of MHC class I allelic variation contributes to allograft rejection, which may be especially important in fish to avoid grafting of cancer cells through their mucosal skin.

The MHC locus (6p21.3) has 3 other paralogous loci in the human genome, namely 19pl3.1, 9q33-q34, and 1q21-q25. It is believed that the loci arouse from the two-round duplications in vertebrates of a single ProtoMHC locus, and the new domain organizations of the MHC genes were a result of later cis-duplication and exon shuffling in a process termed "the MHC Big Bang." Genes in this locus are apparently linked to intracellular intrinsic immunity in the basal Metazoan*Trichoplax adhaerens*.

In transplant rejection

In a transplant procedure, as of an organ or stem cells, MHC molecules themselves act as antigens and can provoke immune response in the recipient, thus causing transplant rejection. MHC molecules were identified and named after their role in

transplant rejection between mice of different strains, though it took over 20 years to clarify MHC's role in presenting peptide antigens to cytotoxic T lymphocytes (CTLs).

Each human cell expresses six MHC class I alleles (one HLA-A, -B, and -C allele from each parent) and six to eight MHC class II alleles (one HLA-DP and -DQ, and one or two HLA-DR from each parent, and combinations of these). The MHC variation in the human population is high, at least 350 alleles for HLA-A genes, 620 alleles for HLA-B, 400 alleles for DR, and 90 alleles for DQ. Any two individuals who are not identical twins will express differing MHC molecules. All MHC molecules can mediate transplant rejection, but HLA-C and HLA-DP, showing low polymorphism, seem least important.

When maturing in the thymus, T lymphocytes are selected for their TCR incapacity to recognize self antigens, yet T lymphocytes can react against the donor MHC's peptide-binding groove, the variable region of MHC holding the presented antigen's epitope for recognition by TCR, the matching paratope. T lymphocytes of the recipient take the incompatible peptide-binding groove as nonself antigen.

Transplant rejection has various types known to be mediated by MHC (HLA):

Hyperacute rejection occurs when, before the transplantation, the recipient has preformed anti-HLA antibodies, perhaps by previous blood transfusions (donor tissue that includes lymphocytes expressing HLA molecules), by anti-HLA generated during pregnancy (directed at the father's HLA

displayed by the fetus), or by previous transplantation;

Acute cellular rejection occurs when the recipient's T lymphocytes are activated by the donor tissue, causing damage via mechanisms such as direct cytotoxicity from CD8 cells.

Acute humoral rejection and chronic disfunction occurs when the recipient's anti-HLA antibodies form directed at HLA molecules present on endothelial cells of the transplanted tissue.

In all of the above situations, immunity is directed at the transplanted organ, sustaining lesions. A cross-reaction test between potential donor cells and recipient serum seeks to detect presence of preformed anti-HLA antibodies in the potential recipient that recognize donor HLA molecules, so as to prevent hyperacute rejection. In normal circumstances, compatibility between HLA-A, -B, and -DR molecules is assessed. The higher the number of incompatibilities, the lower the five-year survival rate. Global databases of donor information enhance the search for compatible donors.

The involvement in allogeneic transplant rejection appears to be an ancient feature of MHC molecules, because also in fish associations between transplant rejections and (mis-)matching of MHC class I and MHC class IIwere observed.

HLA biology

Human MHC class I and II are also called human leukocyte antigen (HLA). To clarify the usage, some of the biomedical literature uses HLA to refer specifically to the HLA protein

molecules and reserves MHC for the region of the genome that encodes for this molecule, but this is not a consistent convention.

The most studied HLA genes are the nine classical MHC genes: *HLA-A, HLA-B, HLA-C, HLA-DPA1, HLA-DPB1, HLA-DQA1, HLA-DQB1, HLA-DRA*, and *HLA-DRB1*. In humans, the MHC gene cluster is divided into three regions: classes I, II, and III. The A, B and C genes belong to MHC class I, whereas the six D genes belong to class II.

MHC alleles are expressed in codominant fashion. This means the alleles (variants) inherited from both parents are expressed equally:

- Each person carries 2 alleles of each of the 3 class-I genes, (*HLA-A*, *HLA-B* and *HLA-C*), and so can express six different types of MHC-I (see figure).
- In the class-II locus, each person inherits a pair of HLA-DP genes (DPA1 and DPB1, which encode α and β chains), a couple of genes *HLA-DQ* (*DQA1* and *DQB1*, for α andβ chains), one gene *HLA-DRα* (*DRA1*), and one or more genes *HLA-DRβ* (*DRB1* and *DRB3, -4* or *-5*). That means that one heterozygous individual can inherit six or eight functioning class-II alleles, three or more from each parent. The role of *DQA2* or *DQB2*is not verified. The *DRB2, DRB6, DRB7, DRB8* and *DRB9* are pseudogenes.

The set of alleles that is present in each chromosome is called the MHC haplotype. In humans, each HLA allele is named with a number. For instance, for a given individual, his haplotype might be HLA-A2, HLA-B5, HLA-DR3, etc... Each heterozygous

individual will have two MHC haplotypes, one each from the paternal and maternal chromosomes.

The MHC genes are highly polymorphic; many different alleles exist in the different individuals inside a population. The polymorphism is so high, in a mixed population (nonendogamic), no two individuals have exactly the same set of MHC molecules, with the exception of identical twins.

The polymorphic regions in each allele are located in the region for peptide contact. Of all the peptides that could be displayed by MHC, only a subset will bind strongly enough to any given HLA allele, so by carrying two alleles for each gene, each encoding specificity for unique antigens, a much larger set of peptides can be presented.

On the other hand, inside a population, the presence of many different alleles ensures there will always be an individual with a specific MHC molecule able to load the correct peptide to recognize a specific microbe. The evolution of the MHC polymorphism ensures that a population will not succumb to a new pathogen or a mutated one, because at least some individuals will be able to develop an adequate immune response to win over the pathogen. The variations in the MHC molecules (responsible for the polymorphism) are the result of the inheritance of different MHC molecules, and they are not induced by recombination, as it is the case for the antigen receptors.

Because of the high levels of allelic diversity found within its genes, MHC has also attracted the attention of many evolutionary biologists.