BIOCHEMISTRY

Thomas Traut
Department of Biochemistry & Biophysics, University of North Carolina at Chapel Hill, USA

Keywords: Amino acid, amyloid, cancer, carbohydrate, catalyst, cell, cofactor, DNA, DNA fingerprinting, drug, enzyme, epigenetic, exon, fat, gene, gene expression, intron, isozyme, lipid, metabolism, microarray, mutation, natural clock, northern blot. nucleotide, origin of life, protein, protein structure, regulation, ribosome, RNA, RNA world, Southern blot, sugar, transcription, translation, western blot.

Contents

Summary

All living organisms, whether single-celled bacteria or complex mammals, must operate within the limits of their environment, which defines temperature, pressure, acidity, and the availability of nutrients. Any structure or process that is not prohibited by thermodynamic principles could have occurred, and natural selection has favored those that are most efficient and enable the organism using them to be adequately adapted to its environment. The central metabolism of cells was optimized very early, since the same metabolic networks are found in almost all life forms, with some limited variations for special situations. The interconversion of sugars and amino acids form the core for central metabolism, with nucleotides and lipids closely related to this core. Nucleic acids store information in the form of DNA for most organisms, and RNA for a limited subset of retroviruses. RNA, as the transcribed copy of a gene, generally serves as a transient template for protein synthesis on ribosomes. Many small RNAs have a regulatory function to control gene expression. Proteins are the key functioning agents of a cell. They form the transporters on the cell membrane that regulate metabolite transport into or out of the cell, and also the receptors that enable communication between cells by hormones. Proteins also include the many enzymes that catalyze thousands of specific chemical reactions that define the metabolic activity of a cell. The origin of life has not been clearly defined. The ability of RNAs to store information, and also show some very limited catalytic activity with phospho-ester bonds, made an 'RNA world' a popular model for the earliest self-propagating proto life form. A more probable origin involves all components that could have emerged under abiotic conditions.

1. Introduction

The twentieth century produced an immense amount of biochemical information, for all aspects of cellular function, and for many species on this planet. Almost every metabolic pathway has been defined. More than 30,000 proteins have had their structure determined. The entire genome for humans as well as for over 200 other species has been sequenced. Several hundred medical drugs have been developed, and tested in animals and humans. This has been an amazingly successful period for acquiring facts about all aspects of life.

Remarkably, this represents only the prelude in the scientific development of the understanding of life. Until now many individual pieces in the great jigsaw puzzle of life have been discovered and carefully studied. But one remains uncertain as to how many pieces compose the whole, and as yet there is only a limited understanding of how some subsets of these pieces fit together. This is an exciting and promising time for biochemistry.

The field of biochemistry expanded in the 19th century as early studies demonstrated the digestion of foods and the consumption of oxygen to oxidize carbohydrates (sugars) and hydrocarbons (fats) to water and carbon dioxide. The energy released was converted to cellular energy in the form of adenosine triphosphate (ATP), which is the immediate co-substrate for thousands of phospho-transfer reactions, by which metabolic intermediates are activated. Phosphate makes a good leaving group and is readily displaced in reactions where the phosphorylated metabolite can be aminated, acetylated, methylated etc. Enzymes that use ATP to phosphorylate some metabolite are normally called kinases, and more than 2000 kinases have been identified in the human genome. Ten per cent of all human enzymes are kinases.

It has consistently been demonstrated that most features of cellular biochemistry are universal (the genetic code), or nearly universal (central metabolism). This permitted the use of chosen organisms for appropriately specific studies, with a confidence that the results had a wider application. Economy requires that such ideal organisms be small, easily maintained, and readily manipulated. Fruit flies have short life spans and many generations are easily followed in a few months, so they are ideal for studying genetic variations. *Escherichia coli*, the common intestinal bacteria, are readily maintained in a laboratory culture and are easily transfected with plasmids carrying novel, experimental genes, and they have been widely used for studying introduced proteins. *Caenorhabditis elegans*, a small worm, has about a thousand somatic cells and about one hundred different nerve cells, and became the model for studying development. And mice have served for studies on the effects of medicinal drugs on mammalian metabolism.

1.1 Basic Concepts and Definitions

A few initial points must be emphasized before considering cellular biochemistry in a broader context.

a) The chemistry of life is dominated by thermodynamic constraints.
   - The energy involved for any biochemical reaction limits the possibilities for how it is catalyzed. Cellular reactions occur at steady state, and are naturally sensitive to mass action. Available energy is often used to favor reactions.
   - Entropy is important.
   - If some mechanism or structure is not thermodynamically prohibited, it will probably occur, although it may be uncommon.

b) Occam's Razor applies in most cases. The simpler solution for any process is normally observed.

- Among the most common requirements are that a process be reliable and reproducible, while not being energetically too costly. Apparent exceptions may frequently be explained by novel circumstances for that organism, such as bacterial extremophiles which may live in niche environments at 100 °C, or at a pH of 1.
- Complexity is favored only when the increased benefit outweighs the cost. Proteins are synthesized from 20 different amino acids, instead of from only 10, because the ensemble of possible protein structures with 20 amino acids makes every needed protein structure possible.

c) Physiological conditions for mammals/humans will be emphasized.
- The pH is assumed to be 7.4, if no other value is stated. At this pH carboxyl and phosphate groups will be ionized, with a negative charge. Amino groups will be ionized, with a positive charge.

d) Brief definitions for biochemical focus areas. To give emphasis to the total number of constituent members of any field, descriptive terms have been coined. These routinely use the suffix *ome* (from the Greek *soma* = body) plus a prefix to designate the topic area.
- Allosterome: all the enzymes that are subject to allosteric regulation.
- Genome: all the genes of a single species. Also, all the genes from all species.
- Interactome: all the genes that may be influenced by a single transcription factor.
- Metabolome: all the enzymes and metabolites in one organism.
- Proteome: all the proteins in one organism. Also, all the proteins from all species. For mammals this number is expected to be much larger than the genome, because mammals often process gene transcripts to produce a diversity of resulting proteins.

It is also important to have a sense of normal sizes for proteins and nucleic acids, and of the rates at which chemical reactions occur. These values are summarized in Tables 1-3. Protein domains are stable folded units of protein structure, and always contain at least one ligand-binding site. Frequently this is a catalytic site, but it may also be a regulatory site for the binding of an activator or an inhibitor. Domains are seldom much larger than 30 kDa. Protein subunits may be much larger than a simple domain, because many proteins contain two or more domains, and some have twenty or more domains.

Macromolecules	
Protein domain	10 - 30 kDa = 90 – 270 amino acids
Protein subunit	10 - 1500 kDa = 90 – 14,000 amino acids
Genomes	Total bases# genes
viruses	649 - 407,339≥ 4 genes
smallest bacterium	160,000 362 genes
E. coli	4.6 million ~4,500 genes
S. cerevisiae	12 million ~6,000 genes
human	3.2 billion ~20,000 genes
Amoeba dubia	670 billion?

Table 1. Normal Sizes for Macromolecules

Genome sizes also vary (Table 1). Minimal sizes simplify the process of duplicating the genomic DNA for cell division and permit the organism to reproduce rapidly under favorable conditions. This benefits viruses and bacteria. Large genomes normally define a greater number of genes and endow the organism with a more sophisticated repertoire of proteins. This is standard for multicellular organisms. There is a poor correspondence between the total DNA content of a genome and the actual number of genes. A simple amoeba, *A. dubia* has two hundred times more DNA than humans, but is expected to have only about 6,000 genes.

The rates for biochemical reactions show an enormous range of values. The uncatalyzed rates (k_{uncat}) for a few reactions are shown (Table 2). The observed time for one half of the molecules to become chemically altered ($t_{1/2}$) can be converted to the number of such events per second (k_{uncat}). The catalytic power of the enzyme is then expressed by the extent to which the enzymatic rate (k_{cat}) amplifies the uncatalyzed rate (k_{cat}/k_{uncat}). While enzymes are amazing for the extent of this rate enhancement, there is a correspondence between the true chemical rate of a reaction and the enzymatic rate. The more difficult the chemical reaction, the slower is the enzyme-catalyzed reaction.

Uncatalyzed Reactions	**$t_{1/2}$**	**k_{uncat} (s^{-1})**	**Enzymatic Reactions**	**k_{cat} (s^{-1})**	**k_{cat}/k_{uncat}**
hydration of CO_2	5 sec	0.2	carbonic anhydrase	~1,000,000	5×10^6
peptide *cis-trans* isomerization	23 sec	0.043	cyclophilin	13,000	3×10^5
triose isomerization	2 days	5.8×10^{-6}	triose-P isomerase	4300	7.4×10^8
peptide hydrolysis	450 years	7.0×10^{-11}	carboxypeptidase	580	8.2×10^{12}
phosphodiester hydrolysis	13 million years	2.4×10^{-15}	Staphylococcal nuclease	95	3.9×10^{16}

Table 2. Normal Rates for Biochemical Reactions*

Also, it is evident in the current databases that normal metabolic enzymes have catalytic rates between 1 and 1,000,000 s^{-1} (Table 3). Enzymes that are slower than 1 s^{-1} are those enzymes with a need to be very specific, such as DNA modifying enzymes. Also, there are enzymes that act as regulatory switches with a built in delay for a conformational change, and the delay is produced by the remarkably slow catalytic rate at which they hydrolyze a bound regulatory effector. Enzymes known as "G proteins" are in this category. Binding of GTP (guanosine triphosphate) causes a conformational change that permits this enzyme to activate a target enzyme. This active conformation may continue for many minutes or even hours, as the site, at which GTP is bound, slowly hydrolyzes it to GDP. For the enzyme p21*ras*, k_{cat} is about 1.2×10^{-4} s^{-1}.

Enzyme Reactions	**k_{cat} (s^{-1})**
metabolic enzymes	1×10^0 - 1×10^6
DNA modifying enzymes	1×10^{-3} - 1

regulatory switches	$1 \times 10^{-5} - 1 \times 10^{-2}$

Table 3. Normal Rates for Enzyme Reactions

2. Central Metabolism

2.1. Sugars and Carbohydrates

There currently is a good understanding of the metabolism (from the Greek *metabola* = change) of the four major groups of molecules (Figure 1). Carbohydrate metabolism centers on glucose (from the Greek *glykys* = sweet). This simple sugar is the center for carbohydrate metabolism, and it must be maintained at concentrations higher than most metabolites, due to its extensive usage. Other important hexoses (Figure 2) are normally converted to glucose, and therefore have many of the same metabolic results.

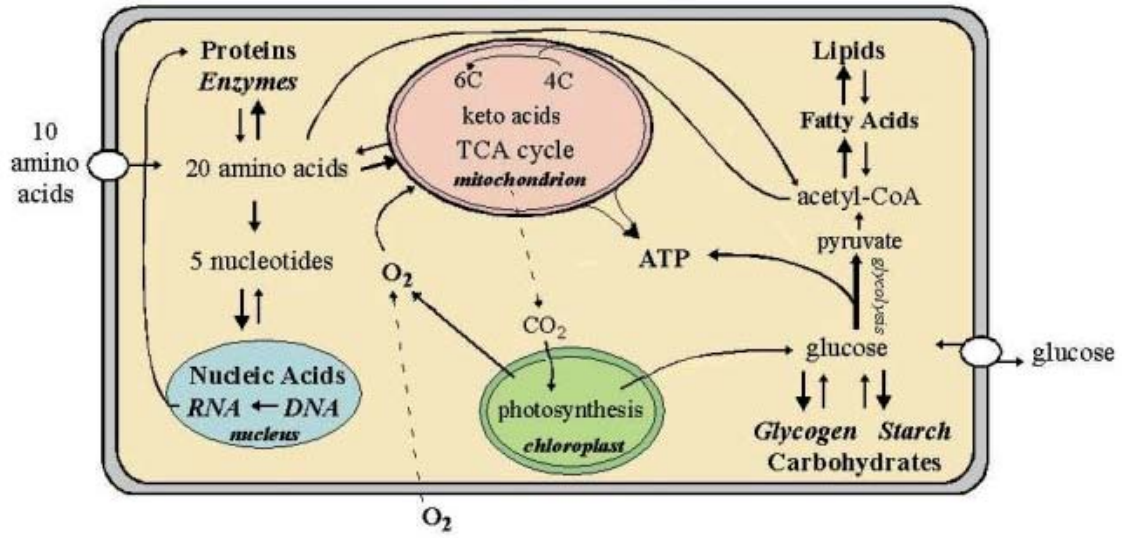

Figure 1. Major areas of central metabolism. The diagram depicts aspects for both animal and plant cells. Bold arrows define synthetic pathways (anabolism) and smaller arrows define degradative pathways (catabolism) or interconversions. Each arrow represents numerous discrete enzyme catalyzed reactions. While gases may diffuse through a membrane, most important dietary metabolites are imported by special membrane transporters, of which only two are depicted. Although all amino acids from the diet are imported, ten are emphasized because these cannot be synthesized by humans, and are essential in our diet.

After glucose molecules enter a cell they can be linked into a large polysaccharide to provide a stored energy source: glycogen in animals, starch in plants. These are polymers of α-glucose. By comparison, cellulose, the principal structural component of plant cell walls, is a polymer of β-glucose. Two disaccharides are important because of their abundance in the human diet. Sucrose, the principal dietary sugar in fruits and plants, is formed from glucose plus fructose. Lactose, the major disaccharide in milk, is formed from galactose plus glucose.

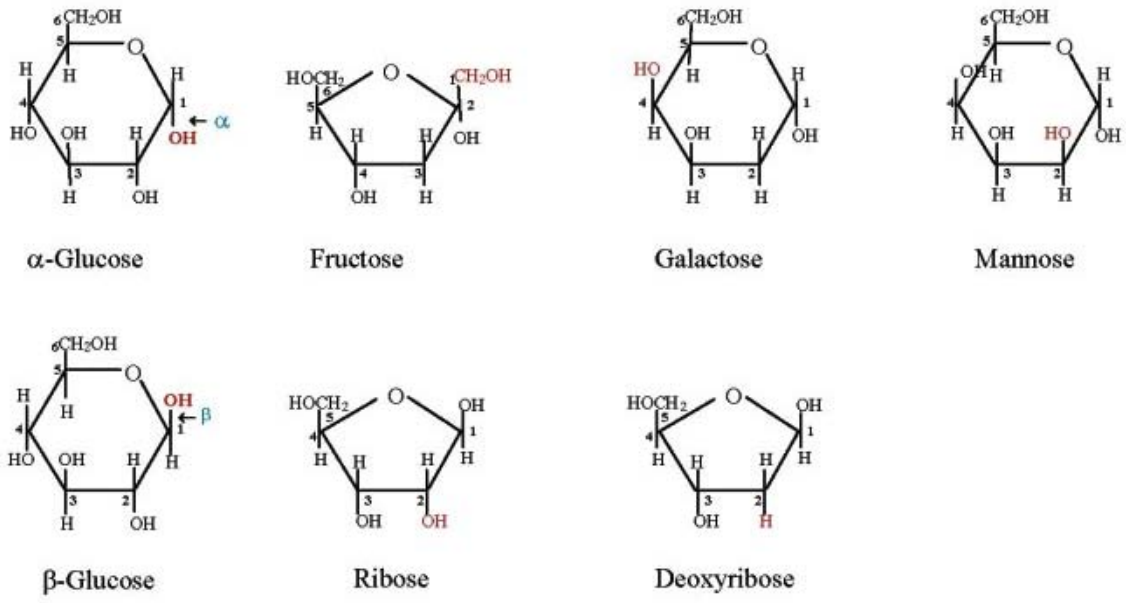

Figure 2. Common sugars. Hexoses (top) are important in polymers (starch, glycogen, cellulose) or in dietary disaccharides (sucrose, and lactose). Ribose and deoxyribose are an important component of nucleotides and nucleic acids.

2.1.1. Problems of Hyperglycemia

Cell membranes normally contain various glycoconjugated proteins. These are formed by enzymes in specific reactions to produce a standard set of glycoproteins. Due to its central role glucose is maintained at a high concentration, and this makes possible the formation of unwanted glycoconjugates. These result when glucose and other sugars become covalently joined, by an uncatalyzed and slow chemical reaction, to various membrane and cellular proteins. This often leads to loss of function for the modified protein. Human blood glucose concentrations are normally maintained at about 4-7 mM. Less than 3 mM is *hypoglycemia*, and greater than 10 mM is *hyperglycemia*. Diabetics, who frequently have glucose concentrations ten times higher than normal (≥ 50 mM), may experience damage due to a greater extent of glyco-conjugates that cause cataracts in the lens of the eye, and occlusion of small capillaries in the extremities. The disaccharide sucrose is cleaved into fructose plus glucose by the enzyme *invertase*. Lactose is cleaved into the two hexoses by the enzyme *lactase*. The inability of many adults to continue their production of the enzyme lactase leads to an inability to metabolize the disaccharide. Accumulation of lactose in the gut supports strains of microbes different from the normal intestinal flora. Side effects produced by these microbes cause lactose intolerance.

The 6-carbon glucose can also be converted to two 3-carbon pyruvates. Glycolysis (from *glykis* + the Greek *lysis* = separate) is the earliest pathway for the utilization of a hexose to provide energy, and is found in all living organisms. It functions in the absence of oxygen, and is presumed to have existed when the earliest life began in a world devoid of oxygen. However, animals are all dependent on oxygen. This metabolic change evolved more than 2 billion years ago when simple cyanobacteria developed

photosynthesis, a process that utilizes single carbon compounds in the form of carbon dioxide, and with ultraviolet energy joins these carbons to make the 6-carbon glucose. Oxygen is a 'waste' product of photosynthesis.

Because oxygen is so easily converted to superoxide, O_2^-, and the very reactive oxygen radical, $O_2^{\bullet -}$, the initial appearance of oxygen was a dangerous side effect of the ability to utilize free solar energy. Those earliest photosynthetic bacteria, or their immediate descendants, quickly evolved novel proteins whose major function was the ability to sequester the newly formed oxygen, by binding it very tightly. These proteins included the first hemoglobin ancestors, and their initial benefit was the removal of oxygen, by binding it to the heme cofactor that all globins and cytochromes contain.

The ability to control oxygen then permitted these cells to evolve by producing additional proteins and enzymes leading to the use of oxygen as the ultimate electron acceptor, by which a much larger amount of potential energy is trapped in the form of ATP as carbon compounds are oxidized. The anaerobic conversion of one glucose to two pyruvate molecules produces 2 net ATP (glycolysis; see Figure 1). The additional combustion of the two pyruvates yields up to 34 more ATP. It is this remarkable increase in the derived energy, from the same starting food source, that made possible the rise of multicellular organisms. When compared to anaerobic bacteria, oxygen-requiring animals have more than 1000 extra enzymes that either function in the control of superoxide and oxygen radical production, or that have added newer oxidative reactions to our metabolism.

-
-
-

Bibliography

Mattick, J.S. and Makunin, I.V. (2006) Non-coding RNA. *Hum. Mol. Gen.* 15, R17-R29. [A review of the many regulatory RNAs.]

Spector, D.L. (2003) The dynamics of chromosome organization and gene regulation. *Annu. Rev. Biochem.* **72**, 573-608. [A review of gene regulation.]

Traut, T. (2008) *Allosteric Regulatory Enzymes*, 250 pp. Springer, NY. [A review of enzyme kinetics and enzyme regulation.]

Voet, G., Voet, J.G., Pratt, C.W. (1999) *Fundamentals of Biochemistry*, John Wiley, NY. [This book includes a clear discussion of central metabolism.]