



**HAL**  
open science

## Basic biological data and calculations for graduate students

Bernard Lorber

► **To cite this version:**

| Bernard Lorber. Basic biological data and calculations for graduate students. 2025. <hal-04924879>

**HAL Id: hal-04924879**

**<https://hal.science/hal-04924879v1>**

Preprint submitted on 1 Feb 2025

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

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



HAL Authorization

## Basic biological data and calculations for graduate students

**Bernard Lorber\***

*From the research team Traduction mitochondriale et pathologies, Architecture et Réactivité de l'ARN, Université de Strasbourg, CNRS, IBMC, 15 rue René Descartes, F-67084 Strasbourg, France*

The abbreviations used are: cmc, critical micellization concentration; AspRS, Aspartyl-tRNA synthetase; DTT, dithiothreitol; *E. coli*, *Escherichia coli*; DLS, dynamic light scattering; Mr, relative mass; N<sub>A</sub>, Avogadro number; SI, The International System of Units; tRNA, transfer RNA.

### **Keywords**

Laboratory exercise, active learning, student conceptual and reasoning difficulties.

**ABSTRACT**

Many graduate students in biochemistry and molecular biology have no clear representation of the size and mass of biological particles, such as proteins, nucleic acids, nucleoprotein complexes, or viruses. Some do not know how one determines the concentration of these invisible objects, and others have only a vague notion about what is an extinction coefficient or a stoichiometry. This article provides course material and exercises for classroom or individual study to train these students before their first practical session. It deals with the aforementioned essential concepts and recalls some basic calculation and conversion methods. It also uses examples from a research laboratory to illustrate how useful the knowledge of particle properties is for the preparation of the experimental work, and for the analysis and interpretation of the results. This scientific treatment of biological data should enable students to find important information by logical thinking with the help of a few memorized facts and procedures.

*Keywords:* Protein, nucleic acid, virus, size, mass, extinction coefficient, concentration, stoichiometry, molarity, solubility, purity, international system of units.

*Running title:* Biological data and calculations

**The author declares no conflict of interest.**

Scientific research relies on the consistent use and application of a number of well-defined concepts, definitions, methods, and techniques. In the various branches of the mathematics, every word has a precise meaning and every equation obeys strict rules. The same holds for physics, chemistry, and also for biochemistry and molecular biology.

Many graduate students attending a first laboratory practical in the two latter scientific disciplines are not sufficiently familiar with the elementary concepts of particle mass, dimension, concentration, extinction coefficient, or stoichiometry. Some have no idea of the size differences between small molecules, biological macromolecules, multi-macromolecular assemblies, and viruses. Others have difficulties remembering the composition of the biochemical sample they have in their hands, not so much because of the number of ingredients than because of the confusing diversity of concentration units. For a number of students, the conversion of the volume percentage of a chemical to molarity remains an enigma. Finally, a few also ignore how to determine the ratio between the concentrations or the number of molecules of two ingredients, or are unable to estimate how much impurity can be tolerated by a biological macromolecule. Very often it is only a small point that needs to be explained in slightly different words.

This article presents the minimum definitions and procedures any graduate student in biochemistry and molecular biology should know before attending a first practical session or being immersed in a research laboratory. To give a sense of scale, it compares the masses and sizes of proteins, nucleic acids, nucleoprotein assemblies, and viruses, hereafter termed “biological macromolecules”. Examples from a research laboratory demonstrate the usefulness of a few simple equations and elementary calculation methods dealing with the extinction coefficient, concentration, stoichiometry, solubility, and purity. Most of the molecular properties and data discussed here are scattered in laboratory manuals and in study guides accompanying textbooks of biochemistry (e.g. [1-5]), biophysics (e.g. [6]), and physical chemistry (e.g. [7]).

This course material with practical exercises may be used either for the training of students in the classroom or for their individual study. It attempts to make above basic concepts and procedures more intuitive and to give a more realistic view of the structural properties of biological macromolecules at the nanometer scale. It demonstrates that a few facts and procedures are sufficient to find supplementary information by analyzing structural and biochemical data in terms of quantities and ratios.

## The size and mass of proteins, nucleic acids, viruses, and their components

Two inseparable properties, namely size and mass, characterize every molecule or particle handled in biochemistry and molecular biology. The students have learned in chemistry and physics courses that scientists worldwide use a small set of basic units (known as the International System of Units, SI) to make published results more easily understandable and reproducible [8]. In this system, the official units for mass and length are the kilogram (symbol kg) and the meter (symbol m). They have, however, conserved a few of the older units like the micron, the Ångström or the Dalton, for practical reasons. To refresh the students' knowledge, one must add that every unit of the SI may be accompanied by standard prefixes for multiples or fractions, such as dm for decimeter, cm for centimeter, mm for millimeter,  $\mu\text{g}$  for microgram, nm for nanometer, and pm for picometer [9]. **Table 1** lists the fundamental properties of biological macromolecules discussed in this paper (together with their symbols and units) and the information required to apply computational or experimental methods to determine them.

It may be necessary to insist that the mole is the SI unit for the amount of substance and that the rule taught for small inorganic and organic molecules in the chemistry course applies also to biological entities, from macromolecules to multi-molecular assemblies and to larger particles like viruses. Although this may seem paradoxical, it implies that the mass of the biggest virus is based on the mass of one mole of pure carbon isotope with atomic weight 12 (or exactly 12 grams) containing  $N_A$  atoms, the Avogadro number  $N_A$  standing for about  $6.022 \times 10^{23}$  elementary entities of substance (see e.g. [10]). It must be pointed out that the mole is always referred to atoms, molecules, ions, electrons or other particles (see e.g. [11]). The aforementioned Dalton (Da) is equal to 1/12 of the mass of the carbon atom (or about  $1.66053886 \times 10^{-27}$  kg) and is used either alone or combined with prefixes (as kDa for kiloDalton, and MDa for megaDalton). It should be stressed that relative masses have the advantage to be dimensionless. One writes  $M_r = 100\,000$  instead of  $M = 100\,000$  Da or 100 kDa (or about  $1.66 \times 10^{-22}$  kg).

It should be easy for students to remember that the unit of length in the nanoworld is the nanometer ( $1\text{ nm} = 10^{-9}\text{ m}$ ) and that the objects they are studying have dimensions in the 1 to 100 nm range, or between about 1/500 and 1/5 of the wavelength of visible light (e.g.  $\lambda = 550$  nm for the yellow ray of sodium). As said above, the Ångström ( $1\text{ Å} = 10^{-10}\text{ m} = 10^{-1}\text{ nm} = 10^{-2}\text{ pm}$ ) is also not a unit of the SI, but chemists use it traditionally to describe ionic radii (e.g.  $\text{Na}^+$  1.02 Å,  $\text{Cl}^-$  1.81 Å, and 2.3 Å for the radius of the V-shaped  $\text{H}_2\text{O}$  molecule) and crystallographer to express distances between atoms.

**Table 2** lists few popular submicroscopic biological objects. It covers a remarkably wide range of masses and sizes and may be used to test the students' ability to estimate the dimensions of well-known particles. The size classes facilitate comparison. On this scale, the difference between small molecules and macromolecules is evident. The Mr of the small molecules does not exceed a few hundred, and their length is less than 2 nm. The small proteins up to the large multi-macromolecular assemblies (as the iron carrier protein ferritin or the huge ribonucleoprotein complex called ribosome) have relative masses ranging from a few thousand to a few million and dimensions from roughly 3 to 30 nm. It is noteworthy that the Mr of the smallest viruses with diameters of 20 to 30 nm already reaches a few millions. The larger viruses whose diameters range from more than 30 to 200 nm exceeds several millions and is usually not given. The fact that the largest viruses discovered during the last decade have dimensions comparable to those of the smallest bacteria may surprises many students. Many of them may also be astonished to learn that the largest bacteria are visible with the naked eye. These microorganisms can reach up to 750  $\mu\text{m}$  or three quarter of a mm in diameter. They are almost hundred times greater than human red blood cells which measure only 8  $\mu\text{m}$  in diameter and are only 2  $\mu\text{m}$  thick.

**Fig. 1** displays side by side a few images of the smallest objects at the same scale for a better visualization of their architectures. At a glance, the students will notice that the symmetric icosahedral virus and the protein ferritin are almost globular and that the other particles have irregular shapes. Biological macromolecules come in a great diversity of shapes but every shape is unique in its smallest details. The volume of the irregular particles cannot be calculated easily for the same reason as one cannot deduce the volume of a fly, a mouse, or an elephant from the dimensions of the imaginary shoe box in which its fits. One must also stress that the static views in **Fig. 1** do not at all reflect the dynamic nature of the macromolecules. Many biological macromolecules are not just rigid blocks. Some have two or more domains that move one relative to the other to accomplish their function. Many polypeptide chains may possess floppy parts. Some change their shape through subtle deformations when they bind small ligands or larger partners. Every state is called a conformation. Nucleic acids share the same properties.

The students may question about how particle sizes can be determined. **Table 1** lists the most commonly used methods. As an example, **Fig. 2** overlays the particle size distributions stemming from independent dynamic light scattering (DLS) measurements. A DLS analysis derives the particle size from rate at which the particle moves in solution by Brownian motion, large particles being slower. (For details on the principle and experimental aspects, see ref. [12]). It has the advantage to need only small amounts of sample, to cover a broad range of

sizes, and to be simple to use, non-invasive, and faster than size exclusion chromatography or ultracentrifugation. Notice that the x axis is a logarithmic scale and don't forget that some macromolecules are not spherical. In the latter case, the more the shape of the particle departs from that of a sphere, the more the experimental hydrodynamic diameter will be far from reality. Measurements performed in aqueous solvents account the presence of water at the surface of the particles.

All students have heard about the elementary building blocks of proteins and of nucleic acids in the biology course. These molecules are too small to be represented in **Fig. 1**. Knowing the chemical formulas of the 22 natural amino acids and the loss of a water molecule (Mr 18) during the formation of each peptide bond by condensation, one deduces that the average relative mass per residue engaged in peptide bonds is about 118 (min. 75 for Gly, max. 204 for Trp). Biologists indicate the length of a polypeptide chain by the number of its amino acids (from a few tens to over one thousand) and identify every residue by the name of amino acid and its position (e.g. Trp<sub>265</sub> in three letter code or W<sub>265</sub> in a more concise form). The number of building blocks of polypeptides rarely exceeds 1000. The supplementary useful information is that the average distance between the  $\alpha$ -carbon atoms of two consecutive amino acid residues is about 0.35 nm. From this, one finds that the number of amino acid residues in and the length of a polypeptide chain with Mr 10 000 is 85 and about 30 nm.

The students can also compute the average mass of the nucleotides found in ribonucleic and deoxyribonucleic acids (i.e. 320 and 307, respectively with a minimum of 323 for CMP, UMP, and dTMP, and a maximum of 507 for ATP and dGTP) from their chemical formulas. Note that the 23S nucleic acid of the eukaryote ribosomes contains almost 3000 nucleotides. It is convenient to express the length of very long molecules, like DNA or RNA genomes, by their number of bases. The unit is the kilobase (1 kb = 1000 bases or nucleotides). From the mean nucleotide mass, one calculates that one kb corresponds to Mr 320 000 for single-stranded RNA, 307 000 for ssDNA, and twice as much for double-stranded chains. It may be worth remembering that the distance between two consecutive base pairs of DNA or between two consecutive phosphorus atoms is on average 0.35 nm, or about the same as the average distance between two consecutive  $\alpha$ -carbon atoms of amino acid residues. Hence, a nucleotide chain of 1 kb has a linear length of 350 nm, the same as a polypeptide chain encompassing 1000 amino acids (Mr 118 000).

To introduce viruses one may say that they are infectious agents made of proteins and nucleic acids that need living cells to replicate. Viruses with architectures based on that of icosahedra (i.e. solids with 20 equivalent faces) have capsids made by well-defined numbers of copies of one or more proteins which contains one or several single or double stranded nucleic

acid molecules. Some icosahedral viruses have capsids enveloped in a glycoprotein and lipid-rich membrane. Bacteriophages have an icosahedral head and a cylindrical tail. Many viruses contain small molecules (e.g. polyamines) and macromolecules (e.g. enzymes) originating from their host.

### The concentrations of proteins and nucleic acid solutions

The students have learned in their biochemistry course that conjugated double chemical bonds are responsible for the absorbance of cyclic compounds, such as the flavine adenine dinucleotide (FAD) and the nicotinamide adenine dinucleotide (NAD) involved in important reactions of the metabolism. The oxidized form of FAD absorbs light in the visible spectrum and has a yellow color. This absorbance can be measured in a spectrophotometer. The extinction coefficient or absorptivity indicates the amount of light absorbed at a given wavelength. It can be expressed either per mole of substance ( $\epsilon$ , in absorbance units per mole) or by weight (E, in  $\text{mg ml}^{-1}$  for an optical path of 1 cm). The relationship between the two latter is:

$$E = \epsilon / M \quad (1)$$

where M is the molecular mass.

In proteins, the double bonds of the aromatic heterocycle indole of Trp and those of the phenyl ring of Tyr strongly absorb the ultraviolet (UV) light. Their absorbance is maximal at a wavelength close to 280 nm. The absorbance of Cys and Phe is much weaker. The purine and pyrimidine bases of nucleic acids absorb light strongly in the UV around 260 nm.

The conversion from absorbance to concentration is straightforward with the extinction coefficient. According to the Beer-Lambert law of optics, the absorbance is equal to:

$$A = -\log(I_0/I) = \epsilon c l \quad (2)$$

where  $I_0/I$  is the transmittance,  $\epsilon$  the molar extinction coefficient (in  $\text{mol}^{-1} \text{cm}^{-1}$ ),  $c$  the concentration (in  $\text{mol L}^{-1}$ ), and  $l$  the pathway (in cm). In practice, many biologists use the relationship:

$$A = E c l \quad (3)$$

where E is the extinction coefficient by weight (in  $\text{mg ml}^{-1} \text{cm}^{-1}$ ) and  $c$  the concentration in  $\text{mg mL}^{-1}$ . The students should memorize that proteins and nucleic acids have values between a

few thousand and several ten thousand. For proteins an average value of  $E$  is  $1 \text{ mg mL}^{-1} \text{ cm}^{-1}$  at 280 nm, but much lower (e.g. 0.5) or much higher (e.g. 2.6) values are common.

The students may wish to know how  $\epsilon$  or  $E$  can be determined experimentally. This can be done by comparing the absorbance of a known amount of substance to either another property of this substance (e.g. the index of refraction measured with a refractometer) or the weight of a known volume of solution after complete drying under vacuum in a dessicator. The index of refraction ( $n$ ) of a substance is the ratio of the velocity of light in vacuum to its velocity in this substance. It is greater than 1 and concentration dependent. In the case of proteins, it varies with the average amino acid composition but does not change much from one protein to another. An average value of the increment of refractive index with concentration ( $dn/dc$ ) is  $0.19 \text{ mg}^{-1} \text{ mL}^{-1}$  [13]. In both methods, the precision on protein mass may be increased by eliminating all unknown counter-ions of the protein. This can be done by dialyzing the latter against a 10 mM solution of small volatile compound such as ammonium hydroxide. Nowadays,  $E$  is computed automatically from amino acid composition of the polypeptide chain. An empirical relationship for proteins dissolved in water is:

$$E_{280 \text{ nm}} = 5500 X_{\text{Trp}} + 1490 Y_{\text{Tyr}} + 125 Z_{\text{Cys}} \quad (4)$$

where  $X$ ,  $Y$  and  $Z$  are the numbers of Trp, Tyr, and Cys residues and 5500, 1490 and 125 the respective extinction coefficients [14,15].

A simple exercise consists in comparing protein molar absorbances calculated from masses and amino acid compositions. Another one supposes that a protein has an extinction coefficient  $E = 1 \text{ mg mL}^{-1} \text{ cm}^{-1}$  at a wavelength of 280 nm. In other words, a solution of this protein at a concentration of  $1 \text{ mg mL}^{-1}$  has an absorbance of 1 when the optical path of the spectrophotometer cuvette is 1 cm. What is the concentration of this solution if the same absorbance is measured in a cuvette with a path of only 1 mm? Since the path is 10-fold shorter, this means that the concentration of the solution is 10-fold higher (i.e.  $10 \text{ mg mL}^{-1}$ ). One can also predict that if  $E$  is equal to  $0.5 \text{ mg mL}^{-1} \text{ cm}^{-1}$  (i.e. the protein absorbs light only half as much as above), then one absorbance unit corresponds to the double protein concentration (i.e.  $2 \text{ mg/mL}$ ) in a 1 cm-long cuvette. Finally, if  $E$  equals  $2 \text{ mg mL}^{-1} \text{ cm}^{-1}$ , then an absorbance of 1 corresponds to a concentration of only  $0.5 \text{ mg mL}^{-1}$ . Some training is needed to make this procedure intuitive.

As seen above, the purine and pyrimidine bases of the nucleic acids are rich in conjugated double bonds that produce a strong absorption in the UV between 240 and 290 nm with maxima

around 260 nm. For pure nucleic acids, the rule of thumb for single-strand (sd) and double-strand (ss) molecules giving the equivalence between absorbance and concentration is:

$$1 A_{260 \text{ nm}} \sim 40 \mu\text{g ssRNA or } 37 \mu\text{g ssDNA or } 50 \mu\text{g dsDNA per mL.}$$

It must be added that, as for proteins, this is only true in the absence of any other substance that absorbs at the same wavelength. This relationship also means that for single stranded RNA an absorbance of 1 at 260 nm corresponds to a concentration of  $1 / 0.040$  or  $25 \text{ mg mL}^{-1} \text{ cm}^{-1}$ . Hence, the absorbance of nucleic acids is 10 to 50-fold higher at 260 nm than that of proteins at 280 nm [16]. In nucleoprotein complexes and in viruses, the absorbance of the nucleic acids is generally superior to that of the proteins.

#### *Solubility and extreme concentrations*

All students have probably observed once in their life that only a limited amount of substance can be dissolved in a given solvent and at a given temperature, e.g. sugar is more soluble in hot than in cold coffee. A few definitions are required to explain this phenomenon. First, a solution is composed of a solvent (e.g. coffee) and of a solute (e.g. sugar) dissolved in the former. Second, saturation is the highest concentration at which all the sugar is soluble. The coffee is unsaturated at all sugar concentrations below saturation and supersaturated above. The solubility of numerous substances may be varied by changing a physical parameter (e.g. temperature or pressure) or adding other compounds. The crystallization of biological macromolecules exploits the latter property (see e.g. [17,18]). In the previous example, at least part of the insoluble sugar will dissolve when temperature increases. At a higher temperature, however, all sugar molecules may not be soluble and saturation may be reached again at a higher sugar concentration.

The students should not ignore that the solubility of proteins varies with temperature. It is normal if it increases and reverse if it decreases when temperature increases. So far, it is not possible to predict protein solubility from amino acid composition, sequence or three-dimensional structure. There are a few well-documented cases of exceptionally high protein solubilities. For instance, the concentration of hemoglobin can exceed  $300 \text{ mg mL}^{-1}$  (5 mM) inside erythrocytes. The concentration of crystallins can reach  $400 \text{ mg mL}^{-1}$  (4 mM) inside translucent eye lenses (see e.g. [19]), but a single point mutation may be sufficient to lower solubility and lead to cataract. These concentrations are close to that existing inside crystals of biological macromolecules harbored by some living animal or plant cells (see e.g. [20]), or inside crystals grown in vitro for crystallographic analyses (see e.g. [21]). Finally, the word

concentration, used to describe the process used to increase the molarity of a dissolved substance by removing solvent (e.g. using filtration or evaporation), should not misled the students.

### *Molarity as a convenient unit for comparing concentrations*

In the SI, the coherent derived unit for molar concentration is the mole per cubic meter ( $\text{mol m}^{-3}$ ) but the number of moles per liter (or molarity in  $\text{mol L}^{-1}$ ) is more practical for laboratory usage. The official name of the liter that is the cubic decimeter ( $\text{dm}^3$ ) and the word liter (symbol L to avoid confusion with capital i) should be used only with results of high-accuracy measurements.

By definition, a 1 M (one molar) solution contains one mole of substance per liter of solution, i.e. its concentration is  $1 \text{ mol L}^{-1}$ . It should be insisted that the adjective “molar” has a slightly different meaning from that used to describe the energy per mole (molar energy in joule per mole,  $\text{J mol}^{-1}$ ) or the entropy per mole (molar entropy or molar heat capacity in joule per mole Kelvin,  $\text{J mol}^{-1} \text{ K}^{-1}$ ) [6]. The students should also be aware that chemists use another concentration unit called molality, sometimes represented by the symbol *m*. At variance with the one molar solution, one prepares a one molal solution by adding one mole of substance to one kg of solvent so that the final concentration is  $1 \text{ mol kg}^{-1}$ . Everybody agrees that there is less dissolved substance per liter of molal solution than per liter of molar solution. The students can intuitively verify this assertion since the volume of the powder or liquid adds to that of the solvent.

The students must be convinced that a corollary of the use of the mole as a single concentration unit is the ease to compare concentrations. The following exercise illustrates the point well. It proposes to compare how much glycerol is contained in a 10 (v/v) and a 10% (m/v) solution. Knowing that pure glycerol has a density  $d = 1260 \text{ g L}^{-1}$  (the density of water is  $1000 \text{ g L}^{-1}$ ) and  $M_r = 92$ , one finds that the molarity of a 10% (v/v) solution (prepared by adding 100 mL glycerol to 900 mL water) is  $(1260 \text{ g} / 92 \text{ g}) \times (100 / 1000) = 1.36 \text{ M}$ , as compared with the molarity of pure water that is  $1000 / 18 = 55.55 \text{ M}$ . Using the same procedure, the molarity of a 10% (m/v) solution (prepared by adding water to 100 g glycerol to a final volume of 1 liter) is only equal to  $100 \text{ g} / 92 \text{ g} = 1.09 \text{ M}$ . By doing the calculation the opposite way, one finds that one liter of a 10 % (v/v) or 1.36 M solution contains  $92 \times 1.36 = 125 \text{ g}$  glycerol, and that 100 g glycerol occupy a volume of only  $100 \text{ g} / 1.26 \text{ g mL}^{-1}$  or 79.4 mL. At this point it may be interesting to tell that glycerol is 1760-fold more viscous than water at 20°C [22], and that the experimental error on a weight is less than the error on a volume because balances are more

accurate than pipets, beakers, and flasks. So, the error on concentration will be minimal when one weighs the required amount of glycerol and adjusts the volume with water.

Concentrations expressed in percentages represent a source of error for the experimenter when the information is incomplete. Did the person who prepared the solution mean 10% (m/v) or 10% (v/v)? It is not always possible to find out quickly which one is correct when both data make sense. The students must also realize that a w/v percentage should actually be written m/v, because a weight is a force (equal to mass multiplied by acceleration, with SI unit  $\text{m kg s}^{-2}$  or Newton, N). It may be necessary to recall that the reading of a balance is a mass (in kg or a fraction of a kg) and not a weight (in N). Finally, concentrations expressed as percentage of saturation (e.g. 50% sat. ammonium sulfate) without indication of temperature are meaningless because the solubility of a substance may vary with this parameter (as explained above).

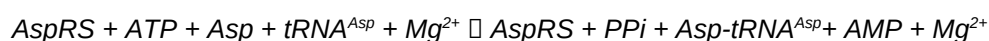
Here are more data to convince students that the use of molar concentrations greatly facilitates the comparison of the concentrations of the constituents of a complex sample. **Table 3** lists the ingredients of a reaction mixture used to assay the activity of an enzyme. The reader can readily appreciate the width of the distribution of the data given in the same unit without doing any calculation. In a similar manner, one could compare the number of molecules that is related to molarity by Avogadro's constant  $N_A$ .

A mnemonic for the conversion from  $\text{mg mL}^{-1}$  to molarity is: a  $100 \text{ mg mL}^{-1}$  solution of a protein with  $M_r$  100 000 has a molarity of 1 mM. To give a feeling of how much is 1 mM, one may say that the concentration of most biological sample solutions used in the laboratory is below 1 mM or very rarely exceeds 10 mM. The concentration of small molecules (e.g. buffer, salt, and most other additives) except ligands, is on average between 10 mM and 1 M. According to the above relationship, a 10 mM solution contains  $100 \text{ mg mL}^{-1}$  of a molecule with  $M_r$  10 000. This is a high concentration for a protein but a very low concentration in the case of a buffering substance (e.g. 50 mM MES buffer in **Table 3**). It may be useful to recall that enzymes are catalysts that can perform the same reaction up to a few million times per second when there is enough substrate under optimal conditions. A concentration in the  $\mu\text{M}$  to nM range may suffice to assay their activity in vitro with a sensitive detection method. In living cells, enzymes may be present in tens or hundreds of copies (e.g. 2000 molecules of glyoxalase [23]), but sometimes a few molecules can fulfill a vital function.

## Stoichiometry

The strange-sounding word stoichiometry scares many young biologists. Chemists and biophysicists use it to speak about a ratio of quantities or of concentrations. A ratio is the relationship between two or more things; it is written x:y or x/y, as 5:1 or 5/1. The stoichiometry is the quotient of this division. In scientific papers, it is written  $n = 5$  (this number has nothing to do with either the index of refraction or the number of samples analyzed in statistics). A stoichiometric mixture of compounds A and B contains as many molecules of A and B. The students must understand why this ratio is important for the reproducible preparation of a functional assembly formed by the association of a well-defined number of protein and nucleic acid molecules. It may also be a crucial variable for the successful preparation of crystals, especially in the case of an association of biological macromolecules. A slight excess of one partner may be necessary to favor the formation of the desired complex but too much of it may be deleterious.

The reaction mixture employed to assay the activity of the enzyme aspartyl-tRNA synthetase (AspRS) in **Table 3** provides other applications of the concept of stoichiometry. The graduate students have certainly heard about the existence of the 20 aminoacyl-tRNA synthetases involved in the first reaction of protein biosynthesis pathway. AspRS first activates aspartic acid in the presence of ATP and magnesium ions and then it loads the activate aspartate moiety onto the cognate transfer RNA. The two steps of the tRNA aminoacylation reaction can be written:



(The footnote in **Table 3** gives more details on the activity assay). It should be highlighted that the enzyme has needs four molecules to perform one cycle: a macromolecule (the tRNA) and three smaller ligands (ATP, aspartic acid, and magnesium ions). The last column in **Table 3** lists the number of molecules of every component per protein molecule, as calculated from respective molarities. The AspRS molecule is composed of two identical subunits. As shown by the result of the titration experiment displayed in **Fig. 3**, the binding of tRNA is stoichiometric, i.e. each of subunits of AspRS binds one tRNA molecule.

The students may wonder about the role of the other molecules listed in **Table 3**. They will probably notice that there is a 5 000-fold excess of amino acid and a 50-fold excess of tRNA per enzyme subunit. The tRNAs may lose their amino acid moiety by chemical hydrolysis (deacylation), and the substantial excess of ATP serves to activate many more amino acid

molecules than those corresponding to one stoichiometry. The  $\text{Mg}^{2+}$  ions are required to perform the catalysis. The buffer molecules stabilize the pH and the salt brings sufficient ionic strength to minimize protein aggregation. Glutathione keeps the thiol groups of the cysteine residues of AspRS in a reduced state so that no undesired disulfide bridges can form between molecules. Albumin prevents that AspRS binds in a non-specific manner to the glass or plastic ware in which the aminoacylation reaction takes place. Finally, who would have guessed that 6 million glycerol molecules stabilize a single protein subunit? The reducing agent (antioxidant) dithiothreitol can replace glutathione in the reaction mixture. The polyamine spermine stabilizes the tRNA by binding in the major groove of the anticodon stem [24]. The non-ionic detergent octyl glucose (Mr 292) binds to hydrophobic regions at the surface of the protein and prevents its aggregation. Above a concentration called the critical micellization concentration ( $\text{cmc} = 20 \text{ mM}$  at  $20^\circ\text{C}$ ), it forms micelles composed of about 75 molecules (total micelle Mr  $\sim 22\,000$ ) that are in equilibrium with soluble detergent monomers (see e.g. [25,26]). Detergent micelles are dynamical objects having the shape of a pumpkin or of a watermelon. They are characterized by a life time and their detergent monomers by a residence time. At a given temperature, the size and shape of micelles vary with the ionic strength and the nature and concentration of the other molecules present in solution.

Another exercise that deals with the stoichiometry consists in determining the number of transfer RNA molecules bound by a dimeric enzyme under real experimental conditions. The tRNAs have an average Mr of 25000. Let us calculate how many molecules of tRNA are present per molecule of enzyme in a typical experiment. If  $5 \mu\text{L}$  tRNA solution at  $10 \text{ mg mL}^{-1}$  is mixed with  $100 \mu\text{L}$  of a  $1 \text{ mg mL}^{-1}$  solution of a dimeric enzyme with Mr 100 000, then the mixture contains  $(5 \mu\text{L} / 1000 \mu\text{L}) \times (10 \text{ mg mL}^{-1} / 25\,000 \text{ mg mL}^{-1}) = 2 \mu\text{mol}$  tRNA and  $(100 \mu\text{L} / 1000 \mu\text{L}) \times (1 \text{ mg mL}^{-1} / 100\,000 \text{ mg mL}^{-1}) = 1 \mu\text{mol}$  enzyme. There are thus 2 tRNA molecules per dimeric enzyme molecule or 1 tRNA per enzyme subunit, as in the case of AspRS in **Fig. 3**. In this equimolecular mixture, all enzyme subunits will bind one tRNA molecule only if their affinity is strong enough. Otherwise, it will be necessary to add an excess of tRNA to saturate all binding sites.

It may be instructive to ask the students to determine the number of impurity molecules added to a pure enzyme with a contaminated chemical. Here is a real case based on the assumption that a non-ionic detergent with Mr 292 contains 1% (m/m) of a contaminant with the same Mr. In other words, there is 1 g contaminant in 100 g detergent. Find out how many molecules of contaminant there are per enzyme molecule (with Mr 60 000) when this detergent is at a final concentration of 1% (m/v) (i.e. 1 g in 100 mL solution or  $10 \text{ g L}^{-1}$ ) in a  $10 \text{ mg mL}^{-1}$  enzyme solution. The question is less difficult than it at first sight appears to be. First, one

calculates the molarity of the detergent  $[(0.01 \times 10 \text{ g L}^{-1}) / 292 \text{ g L}^{-1} = 0.034 \text{ M}$  or 34 mM], then that of the contaminant (1% of 34 mM is 0.34 mM), and finally that of the enzyme ( $10 \text{ mg mL}^{-1} / 60\,000 \text{ g L}^{-1} = 0.00017 \text{ M}$  or 0.17 mM). One immediately sees that there are 0.34 mM / 0.17 mM or 2 molecules of contaminant per enzyme molecule. The activity of the enzyme may be lost instantaneously if the contaminant binds irreversibly to its catalytic site. This example highlights why chemical purity should not be neglected. Rigorous biochemists do this calculation before adding a new chemical to a precious biological macromolecule whose preparation has taken many days or weeks of hard work.

Apparently harmless traces of impurities contained in compounds used at molar concentration may lead to false positive or negative results or to under- or over-estimated data. That is why it is essential to know beforehand the concentration of an impurity an assay can tolerate before it interferes. Chemists express the concentrations of additives and impurities in parts per million (ppm) or in parts per billion (ppb). In the above case, 1% impurity in a compound used a 1% (m/v) corresponds to 10 000 ppm (i.e. 1% of 1 million parts). This is an extremely high level of contamination.

To pique the interest of the students, one can propose a final exercise involving unit conversion. It deals with the production of crystals for the crystallographic analysis of biological macromolecules that remains a challenge because the conditions under which crystals grow cannot be predicted. The initial search for a promising crystallization condition relies on a trial-and-error approach. The experimenter sets up either by hand or with the help of a robot a great number of small droplets containing the target macromolecule and various randomly chosen chemicals (e.g. salts, polymers, alcohols, or mixtures of them) until a beautifully formed crystal or a microcrystal unattractive in appearance is found [27]. To do so, he needs to know how many 200 nL droplets at 5 mg/mL he can prepare with 50 nmol of RNA with Mr 20000. First, one converts the 50 nmol in mg, i.e.  $50 \times 10^{-9} \text{ mol} \times 20000 \text{ g} \times 10^3 = 1 \text{ mg}$ . With this amount one can prepare 0.2 mL RNA solution at 5 mg mL<sup>-1</sup>. This is equal to  $0.2 \times 10^6 = 200\,000 \text{ nL}$  or 1000-fold the volume of a droplet. Hence, the experimenter can theoretically prepare 1000 droplets of 200 nL RNA solution at 5 mg/mL.

## CONCLUSION

A short internship in a research laboratory is a foray into the life of professional scientists for graduate students. To make this first involvement as young scientists most profitable and attractive, the students should possess solid background knowledge. In biochemistry and

molecular biology, they must have a minimum structural knowledge on biological macromolecules and master a few simple procedures.

This article has presented some concepts, definitions, and practical methods which may help convey this knowledge and teach how experimental data can be handled. Such a scientific approach should enable students to develop logical thinking and critical reasoning from a few memorized facts and procedures. The students good at math will enjoy the latter, and those without an inordinate fondness for mathematics will be able to apply them, to extract information on a system beforehand or verify the reliability of experimental results.

To cope with the information overload, the students should learn develop strategies for selecting, sorting, and storing data. They should be trained to use analytical thinking. Common sense and mental techniques (such a mnemonics and calculus) are convenient tools to circumvent tedious and time-consuming operations, check the result of a calculator or decide about the validity of data. The students efficient in estimating, calculating, and checking the reasonableness of answers will have more time for experimenting or analyzing their results. This is particularly true in cross-disciplinary work on biological macromolecules involving biochemical, molecular biological, chemical, and physical studies.

*Acknowledgements* – The author thanks Dr. Richard Giegé, Dr. C. Sauter and Profs. C. Florentz and E. Westhof for their continued interest and support. Special thanks are due to the numerous students with whom the author had the opportunity to work during the last four decades. The author's research was supported by the Centre National de la Recherche Scientifique (CNRS), the Université de Strasbourg, and the French National Program "Investissement d'Avenir, Labex Mitocross".

## REFERENCES

- [1] J.M. Berg, J.L. Tymoczko, L. Stryer (2011) *Biochemistry* 7<sup>th</sup> ed., W.H. Freeman and Company, New York, USA.
- [2] A. Lehninger, D.L. Nelson, M.M. Cox (2008) *Lehninger Principles of Biochemistry*, 5<sup>th</sup> ed., W.H. Freeman and company, New York, USA.
- [3] D. Voet, J. Voet (2011) *Biochemistry*, 4<sup>th</sup> ed. Wiley, Hoboken, NJ, USA.
- [4] N.C. Price, J. Nairn (2009) Oxford Univ. Press, Oxford. *Exploring proteins: A student's guide to experimental skills and methods*.
- [5] A.J. Ninfa, D.P. Ballou, M. Benore (2010) *Fundamental Laboratory Approaches for Biochemistry and Biotechnology*, 2nd Edition, Wiley.
- [6] I.N. Serdyuk, N.R. Zaccai, J. Zaccai (2007) *Methods in Molecular Biophysics: Structure, Dynamics, Function*, Cambridge Univ. Press, Cambridge, UK.
- [7] P. Atkins P., J. de Paula J. (2011) *Physical Chemistry for the Life Sciences*, 2<sup>nd</sup> ed., Oxford Univ. Press, Oxford, UK.
- [8] National Bureau of standards Policy for NBS usage of SI Units (1971) *J. Chem. Educ.* 48, 569-572.
- [9] *The International System of Units* (2006) 8<sup>th</sup> ed., Bureau International des Poids et Mesures, Sèvres, France. Translation "The International System of Units" (2008) NIST special publication 330 Taylor B.N. & Thompson A., Eds. National Institute of standards and technology U.S. Department of Commerce.
- [10] W.B. Jensen (2010) Why has the value of Avogadro's number changed over time?, *J. Chem. Educ.* 87, 1302.
- [11] A. Baranski (2012) The atomic mass, the Avogadro constant, and the mole: A way to understanding, *J. Chem. Educ.* 89, 97-102.

- [12] B. Lorber, F. Fischer, M. Bailly, H. Roy, D. Kern (2012) Protein analysis by dynamic light scattering: Methods and techniques for students, *Biochem. Mol. Biol. Educ.* 40, 372-382.
- [13] H. Zhao, P.H. Brown, P. Schuck (2011) On the distribution of protein refractive index increments. *Biophys. J.* 100, 2309-2317.
- [14] G.R. Grimsley G.R., C.N. Pace (2003) Spectrophotometric determination of protein concentration. *Current Protocols in Protein Science*, 3.1.1-3.1.9.
- [15] Sims P.A. (2012) *J. Chem. Educ.* 89, 738-742. Use of a spreadsheet to help students understand the origin of the empirical equation that allows estimation of the extinction coefficient of proteins.
- [16] H.H. Bull (1964) An introduction to physical chemistry. F.A. Davis, Philadelphia, Pa.
- [17] A. Ducruix, R. Giegé (1999) Crystallization of nucleic acids and proteins: A practical approach, Oxford University Press, UK
- [18] A. McPherson (1999) Crystallization of biological macromolecules, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
- [19] H. Zhao, M. T. Magone, P. Schuck (2011) The role of macromolecular crowding in the evolution of lens crystallins with high molecular refractive index, *Phys. Biol.* 8, 046004 (10pp).
- [20] J.P.K. Doye, W.C.K. Poon (2006) Protein crystallization in vivo, *Current Opinion Colloid and Interface Sci.* 11, 40-46.
- [21] B. Rupp (2009) Biomolecular Crystallography: Principles, practice, and applications to structural biology. Garland Science.
- [22] Handbook of Chemistry and Physics (1974–75) Weast R.C. Ed., 55th ed., CRC Press, Cleveland, OH.
- [23] H.V. Miranda, A.E.N. Ferreira, A. Quintas, C. Cordeiro, A.P. Freire (2008) Measuring intracellular enzyme concentration, *Biochem. Biol. Mol. Educ.* 36, 135-138.
- [24] G.A. Quigley, M.M. Teether, A. Rich (1978) Structural analysis of spermine and magnesium ion binding to yeast phenylalanine transfer RNA, *Proc. Natl. Acad. Sci. U.S.A.* 75, 64-68.
- [25] B. Lorber, J.B. Bishop J.B., L.J. DeLucas (1991) Purification of octyl -D-glucosopyranoside and re-estimation, of its micellar size, *Biochim. Biophys. Acta* 1023, 254-265.
- [26] Marcolongo J.P., Mirenda M. (2011) Thermodynamics of sodium dodecyl sulfate (SDS) micellization: An undergraduate laboratory experiment, *J. Chem. Educ.* 88, 629-633.
- [27] Sauter C., Lorber B., McPherson A., Giegé R. (2012) Crystallization – General methods, International Tables for Macromolecular Crystallography (Rossmann, M, Arnold E., eds), Volume F, 2d Edition, 2012, pp 99-121.

#### References cited in tables and figures

##### Table I

- [28] B.K. Teo, W.-K. Li (2011) The scales of time, length, mass, energy, and other fundamental physical quantities in the atomic world and the use of atomic units in quantum mechanical calculations. *J. Chem. Educ.* 88, 921-928.

##### Table II

- [29] R. Giegé (2008) Toward a more complete view of tRNA biology, *Nat. Struct. & Mol. Biol. Biol.* 15, 1007-1014.
- [30] D.L.D. Caspar, A. Klug (1962). Physical principles in the construction of regular viruses. *Cold Spring Harbor Symp. Quant. Biol.* 27,1-24.
- [31] [http://viralzone.expasy.org/all\\_by\\_protein/260.html](http://viralzone.expasy.org/all_by_protein/260.html)

##### Table III

- [32] B. Lorber, R. Giegé, J.-P. Ebel, C. Berthet, J.-C. Thierry, D. Moras (1983) Crystallization of a tRNA.aminoacyl-tRNA synthetase complex: Characterization and first crystallographic data, *J. Biol. Chem.* 258, 8429-8435.
- [33] R. Giegé, B. Lorber B., J.-P. Ebel, D. Moras, J.-C. Thierry, B. Jacrot, G. Zaccari G. (1982) Formation of a catalytically active complex between tRNA<sup>Asp</sup> and aspartyl-tRNA synthetase from yeast in high concentrations of ammonium sulphate, *Biochimie* 64, 357-362.

##### Figure 1:

- [34] P.W. Rose, C. Bi, W.F. Bluhm, C.H. Christie, D. Dimitripoulos, S. Dutta, R.K., Green, D.S. Goodsell, A. Prlic, M. Quesada, G.B. Quinn, A.G. Ramos, J.D. Westbrook, J. Young, C. Zardecki, H.M. Berman, P.E. Bourne (2013) The RCSB Protein Data Bank: new resources for research and education, *Nucleic Acids Res.* 41 (Database issue) D475-D482.

- [35] M. Carrillo-Tripp, C.M. Shepherd, I.A. Borelli, S. Venkataraman, G. Lander, P. Natarajan, J. E. Johnson, C. L. Brooks III, V. S. Reddy (2009) VIPERdb2: an enhanced and web API enabled relational database for structural virology, *Nuc. Acid. Res.* 37, D436-D442.
- [36] S.I. O'Donoghue, A.-C. Gavin, N. Gehleborg, D.S. Goodsell, J.-K. Hériché, C.B. Nielsen, C. North, A.J. Olson, J.B. Procter, D.W. Shattuck, T. Walter, B. Wong (2010) Visualizing biological data – now and in the future, *Nature Methods* Supplement 7, S2-S4.
- [37] S.I. O'Donoghue, D.S. Goodsell, A.S. Frangakis, F. Jossinet, R.A. Laskowski, M. Nilges, H.R. Saibil, A. Schafferhans, R.C. Wade, E. Westhof, A.J. Olson (2010) Visualization of macromolecular structures, *Nature Methods* Supplement 7, S42-S55.
- [38] D.S. Goodsell (2012) Putting proteins in context, *BioEssay* 343, 718-720.

**Figure 3**

Ref. [33] cited in Table 3.

## LEGENDS OF TABLES AND FIGURES

**Table I. Structural properties of biological molecules.**

**Table II. Mass and diameter of some molecules, particles, and cells.**

**Table III. Ingredients of a protein/nucleic acid reaction mixture.**

**Fig. 1: Comparison of the size and mass of a few biological particles.** The drawings of the smaller particles were generated with PYMOL using atomic coordinates retrieved from Protein Data Bank (PDB). The surface views of the apoferritin shell (PDB ID 2W00) and of the virus capsids were prepared with the VIPERdb Virus Particle Explorer<sup>2</sup>. Such images can be created from atomic coordinates deposited by scientists with the Protein Data Bank [34] and visualized using dedicated software (e.g. [35-37]). See also scientific illustrations that bring together information from various sources and provide an integrative view of the molecular biology of cells (e.g. [38]).

**Fig. 2: The size range of biological macromolecules and particles.** The hydrodynamic size of proteins, tRNA, and viruses was compared to those of the small organic molecule glycerol and of the living enterobacteria *Escherichia coli*. Every sample was analyzed separately by dynamic light scattering and all particle size distributions by intensity were overlaid for comparison. Sample solutions (20  $\mu$ L) were analyzed at 20°C in a quartz cuvette using a Malvern Zetasizer NanoZS instrument. The fluctuations of the intensity of the light scattered by the molecules or particles were recorded and an autocorrelation function was generated from which the diffusion coefficient and the size of the particles were derived. For experimental details, see [12].

**Fig. 3: Titration of yeast cytoplasmic AspRS with tRNA<sup>Asp</sup>.** The variation of the fluorescence of the single Trp residue per enzyme monomer was monitored during the addition of tRNA<sup>Asp</sup>. To perform this experiment three decades ago as much as 1 mL enzyme solution at 0.12 mg/mL (monomer concentration 2  $\mu$ M) was needed. Today, a much smaller volume at a lower concentration would suffice to do the same analysis with a more sensitive spectrofluorimeter. Excitation and emission wavelengths were respectively 295 nm and 350 nm, with corresponding optical paths of 0.4 and 1.0 cm. For details see ref. [33]. Upon tRNA binding a conformational change occurs in the polypeptide chain, and the Trp residue becomes exposed to the solvent. Once all monomers have bound one tRNA molecule the fluorescence does not increase further. The diameter of the symbols indicates the experimental error.

**Table I. Structural properties of biological molecules.**

Property	Symbol	Units	Information required	Experimental method
Molecular mass	M	Da		
Relative molecular mass	Mr	none		
Dry particle			Amino acid and/or nucleotide composition <sup>a</sup>	Computation Mass spectrometry
Hydrated molecule or particle				Size exclusion chromatography Analytical ultracentrifugation Static light scattering
Radius or diameter	r or d	nm		
Dry particle				Transmission electron microscopy X-ray crystallography
Hydrated particle	r <sub>h</sub> or d <sub>h</sub>	nm		Dynamic light scattering, X-ray scattering
Concentration (mass/volume or molarity)	c	mg mL <sup>-1</sup> , mol L <sup>-1</sup>	Absorbance and extinction coefficient	Spectrophotometry
Molar extinction coefficient (at a specified wave length)	ε	mol <sup>-1</sup> cm <sup>-1</sup> <sup>b</sup>	Amino acid and/or nucleotide composition	Spectrophotometry and dry weight Spectrophotometry and refractometry <sup>c</sup>

(a) Amino acid and nucleotide composition either determined experimentally or deduced from that of the corresponding gene.

(b) The extinction coefficient by weight, E, is in mg<sup>-1</sup> mL<sup>-1</sup> cm<sup>-1</sup>. It is sometimes given for a 1% (m/v) or 10 mg/mL solution.

(c) In refractometry, a refractometer is used to measure the index of refraction of a substance. The index of refraction is the quotient of the angle by which light is refracted by the substance on that of air. It is a dimensionless number.

Note: It is interesting to compare these units with those used to describe the atomic world. The mass of the electron at rest is about  $9 \times 10^{-31}$  kg while the atomic mass unit – the Dalton, Da - given by  $1\text{g}/N_A$  is about  $1.66 \times 10^{-27}$  kg; the radius of the  $n = 1$  orbit of hydrogen or Bohr radius is about  $0.5 \text{ \AA}$  or  $1/20 \text{ nm}$  (see ref. [28] for more units).

**Table II. Mass and diameter of some molecules, particles, and cells.**

Particle	Relative mass <sup>a</sup>	Average dimensions (nm) <sup>b</sup>
<b>Small inorganic and organic molecules</b>		
Water molecule (V-shaped)	18	Diameter 0.46
Sodium ion (spherical)	23	Diameter 0.2
Glucose (plane molecule)	180	Diameter 1
<b>Proteins, nucleic acids, and their complexes</b>		
Insulin (egg-shaped)	5800	Diameter 3
Lysozyme (hen egg-white)	14 500	Diameter 3.8
tRNA (boomerang-shaped) <sup>c</sup>	25 000	Diameter 6, diagonal 8, branches 2
Serum albumin (elongated)	66 000	Diameter 7
Yeast cytoplasmic aspartyl-tRNA synthetase (ycDRS, non-globular)	127 000	10 x 7.5 (min 4.5) x 9
IgG (Y-shaped)	160 000	16 x 12 x 4
Ribonucleoprotein complex composed of one ycDRS dimer and two yeast tRNA <sup>Asp</sup>	177 000	12 x 7.5 (min 4.5) x 9
Ferritin, shell either empty or filled with iron (globular) <sup>d</sup>	480 000 / 700 000	Diameter 12 (inner diameter 8)
Ribosomes from procaryotes and eukaryotes (elongated)	a few x 10 <sup>6</sup>	20 nm and 30 nm
<b>Viruses</b>		
Smallest icosahedral viruses <sup>e</sup> (quasi-spherical)	One to a few x 10 <sup>6</sup>	Diameter 20 to 30
Icosahedral head of bacteriophage X-174 Many small icosahedral RNA plant viruses Viruses causing common cold, hepatitis B, and papilloma	A few to several x 10 <sup>6</sup>	Diameter 30 to 40
Simian virus SV40, viruses causing polio and dengue Viruses causing West Nile and sheep bluetongue diseases	n.a.	Diameter 50 Diameter 60 and 80
Icosahedral core of human herpes-1 virus Icosahedral human immunodeficiency virus HIV-1	n.a.	Diameter 100, with envelope 200 Diameter with spikes 140
<b>Largest viruses and bacteria</b>		
Biggest icosahedral mimi- and megaviruses	n.a.	Diameters 400 and 440
Smallest coccoid bacterium <i>Nanoarchaeum equitans</i>	n.a.	Diameter 400
Enterobacterium <i>Escherichia coli</i> (rod-shaped)	n.a.	2000 x 500
Biggest coccoid proteobacterium <i>Thiomargarita namibiensis</i>	n.a.	Diameter up to 750 000

- Relative mass is the dry mass that does not take into account bound water molecules.
- The greater values are rounded to the closest integer.
- Classical tRNAs have the shape of a boomerang. The distance from the tip of the anticodon to that of the CCA end is 7 to 8 nm while the diameter of the helical branches is no more than 2 nm (see e.g. [29]).
- Ferritin is a ubiquitous protein whose function is to regulate the concentration of soluble iron in the body of animals and humans. Its 24 subunits form a spherical shell (called apoferritin) that can contain a ferrite core composed of more than 4000 iron atoms.
- Viruses can be divided two groups according to their architecture: those with cubic symmetry and the others with helical symmetry. The first have the shape of an icosahedron and the second that of a rod. An icosahedron is a highly symmetrical and almost spherical solid composed of 20 facets. Every facet is an equilateral triangle and the assembly has several axes of symmetry. The theoretical basis for the structure of isometric viruses was first described by Caspar and Klug [30]. These authors coined the term "triangulation number" that characterizes the architecture of the icosahedron. See e.g. ref. [31] for drawings of icosahedral capsids with various triangulation numbers.

Abbreviation used: n.a.: not applicable. The relative mass of the large viral particles exceeds several millions. The concept of mass does not apply to the bacteria because their exact chemical composition is unknown.

**Table III. Example of the ingredients of a protein / nucleic acid reaction mixture.**

Compound	Concentration	mM	Mol per mol of DRS monomer
<b>Enzyme and substrates</b>			
Aspartyl-tRNA synthetase (AspRS) from yeast cytoplasm	$2 \times 10^{-8}$ M	$2 \times 10^{-5}$	1
ARNt <sup>Asp</sup> (Mr 25 000)	10 $\mu$ M	$1 \times 10^{-3}$	50
ATP	10 mM	10	500 000
Aspartic acid	0.1 mM	0.1	5 000
MgCl <sub>2</sub>	15 mM	15	750 000
<b>Standard additives</b>			
MES buffer pH 7.5	50 mM	50	2 500 000
KCl	30 mM	30	1 500 000
Glutathione	2.5 mM	2.5	12 500
Bovine serum albumin (M <sub>r</sub> = 66 000)	0.2 mg/mL	$3.0 \times 10^{-3}$	150
<b>Optional additives</b>			
Spermine	1 mM	1	50 000
Glycerol (M <sub>r</sub> = 92, density = 1261 g/L)	10% (v/v)	1370	6 850 000
Dithiothreitol (DTT)	10 mM	10	500 000
Octyl- D-glucopyranoside detergent (M <sub>r</sub> = 292)	0.5% (m/v)	17	850 000

This is a list of the ingredients of the reaction mixture used for the aminoacylation of tRNA<sup>Asp</sup> by yeast cytoplasmic aspartyl-tRNA synthetase (AspRS) in the course of the first crystallization of a tRNA.aminoacyl-tRNA synthetase complex [32,33]. AspRS is a homodimer (M<sub>r</sub> 127 000) composed of two identical subunits having each a catalytic site. The concentration given here is that of the subunit. During a catalytic cycle, one molecule of aspartic acid is activated in the presence ATP and magnesium ions to form a molecule of aspartyl adenylate and inorganic pyrophosphate that is released. The aspartate moiety is then bound covalently onto the 2' OH group of the ribose of the 3' terminal adenine at the CCA-end of tRNA<sup>Asp</sup>. The amount of aspartyl-tRNA formed was determined in following way. The activity assay was performed at 37°C on 200  $\mu$ L reaction mixture containing some radioactive aspartic acid. 40  $\mu$ L aliquots were withdrawn after regular time intervals and deposited on pieces of absorbing paper with a surface area of 2 square cm. Once the solution was absorbed, the paper was immersed in a 5% (v/v) solution of trichloroacetic acid to immobilize the protein and the nucleic acid molecules. The unbound radioactive molecules were washed away and the paper was dried before the bound radioactivity was measured in a scintillation counter.

**Fig. 1: Comparison of the shape, size and mass of some biological particles.**

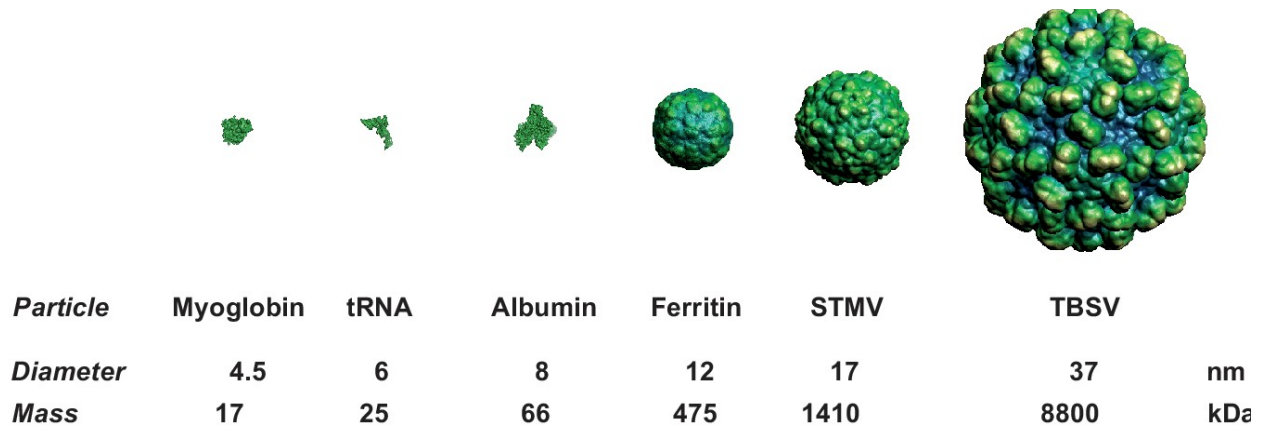


Fig. 2: The size range of some biological macromolecules and particles.

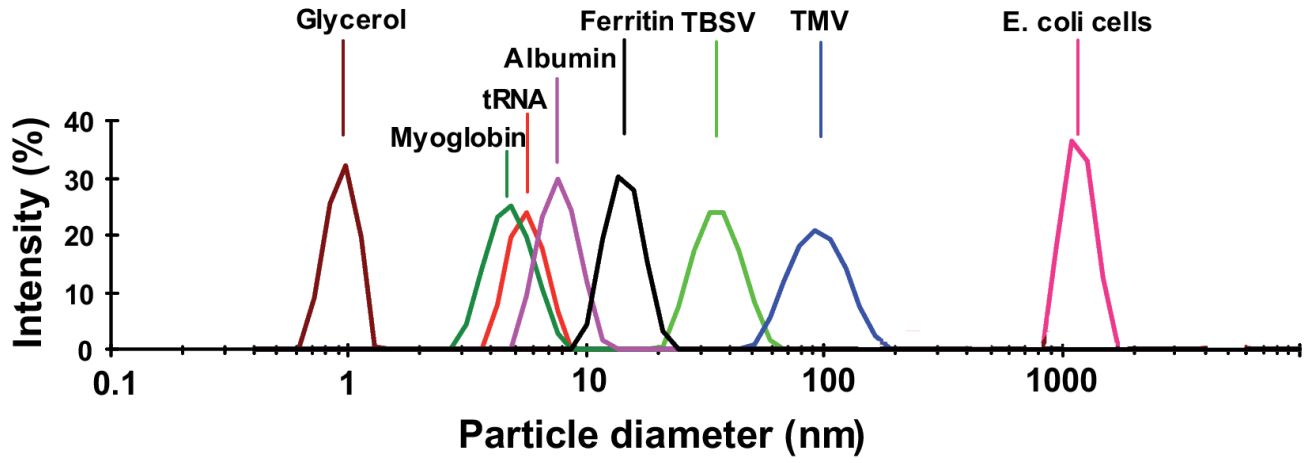


Fig. 3: Titration of yeast cytoplasmic AspRS with tRNA<sup>Asp</sup>.

