**Charles University in Prague**

**First Faculty of Medicine**

Degree program: Human Physiology and Pathophysiology

Field of study: Biomedicine



Formalization of Integrative Physiology

by

**Marek Mateják**

(Dissertation thesis)

Supervisor: Doc. MUDr. Jiří Kofránek, CSc.

Prague, 2015

Content

[1 Introduction 3](#_Toc421377499)

[1.1 Models From Mississippi 3](#_Toc421377500)

[1.2 Formalization of Physiology 5](#_Toc421377501)

[1.3 Goals – integrative model for computer smulation 5](#_Toc421377502)

[2 Methods of base principles formalization 8](#_Toc421377503)

[2.1 Physical principles 8](#_Toc421377504)

[2.2 Modelica Principles 10](#_Toc421377505)

[2.3 Building Modelica Libraries 15](#_Toc421377506)

[3 Results of base principles formalization 20](#_Toc421377507)

[3.1 Chemical domain 20](#_Toc421377508)

[3.2 Thermal domain 26](#_Toc421377509)

[3.3 Hydraulic domain 27](#_Toc421377510)

[3.4 Population domain 29](#_Toc421377511)

[4 Methods of integrative physiology formalization 31](#_Toc421377512)

[6 Result of integrative physiology formalization 36](#_Toc421377513)

[6.1 Cardiovascular system 36](#_Toc421377514)

[6.2 Body Water 41](#_Toc421377515)

[6.3 Hormones 45](#_Toc421377516)

[6.4 Electrolytes and Acid-Base 49](#_Toc421377517)

[6.5 Blood Gases 53](#_Toc421377518)

[6.6 Nutrients and Metabolism 55](#_Toc421377519)

[6.7 Thermoregulation 57](#_Toc421377520)

[6.8 Neural Regulations 58](#_Toc421377521)

[7 Discussion 61](#_Toc421377522)

[7.1 Physiological expandability 61](#_Toc421377523)

[7.2 Expandability in the field of physical chemistry 62](#_Toc421377524)

[7.3 Effort of Integrative Physiology 64](#_Toc421377525)

[8 Conclusion 65](#_Toc421377526)

Motto:

“Science is a method for deciding whether what we choose to believe has a basis in the laws of nature or not.”

Marcia McNutt

# Introduction

Having detailed computer simulation, which can predict the behavior of the system is a dream of many scientist. However, in the world of physics the prediction is already possible. Even in chemistry in year 2013 M. Karplus, M. Levitt and A. Warshell won the Nobel Prize for the development of multiscale models for complex chemical systems, which predicts the chemical reaction properties even for complex molecules. So why this kind of computer prediction should not be possible in complex physiology? What need to be done to reach that goal?

I have a strong belief, that for this purposes the physiological processes must be formalized in the meaning of generalization and integration. Formalization in the level of exact physical description should be the step in the right direction, because the physics generalize the processes in the nature. However, the human physiology is very complex and it must be described with respect to this complexity. So the effective integration of the physiological knowledge could be the clue to the *in silico* prediction.

The formalization of integrative physiology can be understand as finding the base principles from physics and to define the methods how to integrate these base principles together into one complex model of human physiology. This approach already started in Mississippi in sixties 19th century. This work follow up to this development by finding better base principles, finding better way to implement the model and finally by extending the model with better calculation of chemical processes such as presented in hemoglobin allostery calculation.

The problem is, that the model has still a lot of parameters, which need to be formalized and calculated from more detailed physical principles. However, even if we cannot simulate the development of the body the

it draws from more than 400 physiological papers that describe real experiments pertaining to the presented theories.

It describes in detail the integration of physiological knowledge into one complex model in Modelica language.

achieving the readability and error-resistance of complex models is the primary goal of this thesis.

All described physiology is in the above context implemented by the author in [Modelica](https://www.modelica.org/), using best practices and features described in section 2. Section 3 shows that each physical law is represented by the author as an equation in the code hidden behind the library component. Using the [Physiolibrary](http://www.physiolibrary.org/) and [Chemical](https://github.com/MarekMatejak/Chemical/releases) library, the author of this thesis presents an integrated and extended large model, [HumMod](http://hummod.org/) 1.6, the largest integrative physiological model for the year 2012. During implementation, more than 30 corrections were made by the author of this dissertation, as described in the final subsections about each main subsystem of the [Physiomodel](http://www.physiomodel.org/) (sections: 4.1.5, 4.2.6, 4.3.9, 4.4.6, 4.5.5, 4.6.6, 4.7.3 and 4.8.1).

The biggest extension incorporated into the model is the new gases transport and acid-base model, as described in sections 4.4.1, 4.5 and 6.1. Section 4 shows that each part of the model is also a component that can be used and parametrized as many times as necessary. Furthermore, all these components can be created hierarchically using diagrams. Such a graphical representation will show exact mathematical meaning, with the library components from section 3 at the lowest level of the hierarchy. The discussion of physiology in section 4, similar as in physiological texts, provides descriptions of the running mathematical model in the specific hierarchical level. This is a result of the final generation of object-oriented equation-based software tools, which allows for exactly mapping almost all physiological terminologies, for both physics and mathematics. The generalized theoretical approach used in the development of the model is formalized in section 5, where general answers to the primary integration question about validity and usability of the complex models are presented.

In the beginning of this work was reimplementation of physiological models to Matlab/Simulink (Mathworks corp., U.S.). This approach was very time-consuming with only a small additional value, because the models were not simplified and not self-describing. Even more these models was unable to easy extended or modified. That was the reason why we go away from causal simulation tools and we starts to look for new possibilities how to formalize the physiological model for computer simulation (Mateják, et al., 2008). The solution for our criteria of computer language was achieved by Modelica language, where I easily implemented original Guyton model from year 1972 (Mateják, et al., 2009). One year later I implemented to Modelica the model QHP (Mateják and Kofránek, 2010). This model was one of the main result of our very successful national project „e-Golem: mediacal learning simulator of human physiological functions as a backgound of e-learning teaching of critical care medicine“ (2006-2009, MSM/2C, 2C06031). The next model, I implemented in Modelica, was HumMod 1.6 (Mateják and Kofránek, 2011). This implementation has an automatic comparison of output values with original model, so it can describe all experiments for which was designed the original one. Having our implementation it starts to be very easy to extend the model with new acid-base theory, blood gases transport or new cardiovascular details. So, in year 2012 I implemented into the model the new blood oxygen status model of Siggaard-Andersona (Siggaard-Andersen and Siggaard-Andersen, 1990). The Model was able to simulate for example the support of artificial ventilation or even extravascular oxygenation (Mateják, et al., 2012). These and many other inputs such as infusions, dialyses, transfusions or hemorrhage was designed for educational simulations in project „ Virtual patient - Simulator for medical education“ (2011-2014, MPO/FR, FR-TI3/869). In the same manner of educational simulation was in the model tested also different scenarios of acid-base and respiratory disorders, for example the ketoacidosis (Mateják, 2013), where was already implemented the new calculation of acid-base as a result of electroneutrality with calculation of each significant chemical substance. One of the last my significant result of this work was separating of the main principles from the model into the Modelica library called Physiolibrary (Mateják, et al., 2014). This library won the free Modelica library awards at 10th International Modelica Conference in year 2014 in Lund, Sweden. Having these base principles for hydraulic calculation of cardiovascular system, we was implemented and identify the more detailed cardiovascular models {Kulhánek, 2014 #2}. Working on the physical chemistry theory behind physiology I formalized the first principles of electrochemical processes in year 2015. Using these chemical formalizations it is possible to implement almost each equilibrium from physical chemistry. And more, I designed also the general principle of allosteric equilibria, which was used for example to calculate the hemoglobin model with three ligands – oxygen, carbon dioxide and hydrogen ions (Mateják, 2015; Mateják, et al., 2015). As shown in the thesis all these physiological descriptions can be easily integrated into the one model, which I named Physiomodel.

## Models From Mississippi

One of the first integrative mathematical models of human physiology was designed by (Guyton, et al., 1972). This model integrates the most necessary relations in order to describe essential hypertension. Since this pathology includes cardiovascular circulation, renal functions, the renin-angiotensin-aldosterone-system (RAAS), vasopressin and fluid balances, the model was quite complex in terms of describing the state of the patient in periods of minutes, days and months. The model was compared and partially fitted to dog nephrectomy experiments. The model well-describes the regulation of the cardiovascular system as it relates to extended water volume, which is caused by kidney function failure. It is also able to provide answers to matters of increased blood flow during an acute phase, as well as an increase of blood pressure in the chronic phase following hormonal inducted vasoconstriction (Guyton, 1991; Guyton and CE Coleman, 1973).

The power of integration in this model was very strong, leading to it becoming the basis for follow-up ongoing developments. Guyton and coworkers noticed the interactions between regulations at different time scales and elegantly described long-term pathological processes. Guyton’s Textbook of Medical Physiology (Hall, 2010) became a bestseller that was translated into at least 15 languages and is still upgraded today with new editions.

The development of the model continued with more detailed cardio-vascular and body fluid regulations (Guyton, 1981; Guyton and CE Coleman, 1973; Guyton, et al., 1975). In 1983, Thomas G. Coleman, Guyton’s coworker, published the “Human” model (Coleman and Randall, 1983). This model was more interactive and based on standard medical situations and the cures applied by physicians. The user can simulate cardiac failure, renal failure or hemorrhage and can interact during the simulation by applying infusion, transfusion, artificial ventilation, selected drugs and/or dialysis. This model became a virtual simulator designed for teaching medical students. Even its interface was simple; both teacher and student examine the state of the patient (described by physical quantities such as blood pressure, heart rate, ventilation rate, etc.) and apply the necessary response measures. The model was also reimplemented into a [web version](http://placid.skidmore.edu/human/index.php) by Roy D. Meyers, Leo D. Geoffrion and Chris L. Doherty.

This model could be applied for more than simply the description of development of hypertension. Its users were able to examine the microcirculation of tissues, more details regarding the function of kidneys in terms of complex homeostasis, the ventilation system and other areas. The scientific expertise of Mississippi University’s Medical Center also evolved as a result of cooperation with NASA on interesting physiological research, for example, the project Digital Astronaut (Summers and Coleman, 2002). Estimation and validation of experiments in microgravity or in artificial environments remain one of the goals of the model. Though there were many types of simulations available, there was only one model describing one organism in different settings and different conditions.

In 2005, the next version of the model, called “Quantitative Circulatory Physiology” ([QCP](https://www.umc.edu/education/schools/medicine/basic_science/physiology_and_biophysics/core_facilities(physiology)/free_qcp_software_download.aspx)) (Abram, et al., 2007) was released. This model continues in the style of the Human model. Dr. Coleman extracted parameter sets from the model, which allows for creating a patient using separate files with unique parameters. Though there are more than 1000 parameters to the setting, most of them can be copied or scaled.

The next improved version was called “Digital Human” or “Quantitative Human Physiology” ([QHP](https://www.umc.edu/uploadedFiles/UMCedu/Content/Education/Schools/Medicine/Basic_Science/Physiology_and_Biophysics/Facilities/Modeling_Workshop/QHP07PreviewVersion.zip)) (Hester, et al., 2008). This model even separates equations from the compiled executable file.

Dr. Coleman defined a new xml-based language of physiology formalization. Using Coleman’s tags, it is possible to implement a set of differential equations and implicit equations. To read these xml-files, we designed an xsl-template for Mozilla browser called [QHPView](http://patf-biokyb.lf1.cuni.cz/wiki/projekty/qhp2007), which allows for reading the equation in a more readable style. Furthermore, the graphical user interface was implemented using specific xml-language that allows for designing the application simply by editing these files. This mode has more than four thousand variables, defined mostly as real [physical quantities](https://en.wikipedia.org/wiki/Physical_quantity). This model was in 2010 renamed as “[HumMod](http://hummod.org/)” (Hester, et al., 2011; Hester, et al., 2011). With more than 4500 physical quantities definitions, the HumMod is to date the largest model of the complexity of human physiology.

## Formalization of Physiology

The building of mathematical models is strongly connected with the formalization process. The advantage of formalized description of examined reality using mathematical relations is that the deduction of the behavior of examined objects can be done by solving the equations. The equations do not need to be solved manually, because typically, this is done by computers. These solutions to differential equations simulate the behavior of real objects over a period of time. This is called a **computer simulation**.

One of the first formalized mathematical models in physiology was a simplified model of a neuron in 1943 by (McCulloch and Pitts, 1943). Five years later, the compartmental approach for pharmacology was invented by (Sheppard, 1948). Following on, a model of excitable neural membrane was published by (Hodgkin and Huxley, 1952). In 1954, two compartmental models of respiration homeostasis was designed by (Grodins, et al., 1954). In the 1960s, with the development of the first computers, more complex models were developed, e.g., the first complex blood gases transport model (Grodins, et al., 1967) and the regulation of blood circulation with body fluid balance (Coleman, et al., 1967). In the 1970s, complex models such as the aforementioned models by Guyton from Mississippi University (Guyton, et al., 1972) and models of homeostasis and body fluid balance by (Ikeda, et al., 1979) were developed. In the 1980s and 1990s, personal computers rendered mathematical modeling one of the standard methods for use in physiological research.

Nowadays, physiological formalization has a new dimension, thanks to the international project [Physiome](http://www.physiome.org/), the successor of the famous project [Genome](http://web.ornl.gov/sci/techresources/Human_Genome/index.shtml). The goals of Physiome and the European project called Virtual Physiology Human ([VPH](http://physiomeproject.org/about/the-virtual-physiological-human)) is the formalized description of physiological functions. The framework used includes computers models of physiological parts of the human body (Bassingthwaighte, 2000).

Practical use of these formalized models is today concentrated on education, using smart robotic simulators to emulate real situations in critical care medicine; these include mannequins from [Leardal](http://www.laerdal.com/us/) or the [METIMan](http://www.caehealthcare.com/patient-simulators/metiman) from CAE Healthcare. There is also a large set of educational interactive software based on physiological models, starting with [HumMod](http://hummod.org/), [Physiological Atlas](http://www.physiome.cz/atlas/) and Virtual Patient.

New perspectives on the use of complex physiological models include hardware-in-loop devices for the automatic monitoring and curing of patients and improving the product-life-cycle management of pharmacological drugs through, for example, simulated testing prior to conducting clinical trials. Notions pertaining to future use include applying the model for the use of personalized medicine. Simulation of a selected patient is not currently possible, due to a lack of methods for effectively setting all the parameters within the complex model. However, restricting default values or simple scaling to type of pathological problem, height, sex, age, surface of skin or height is in most instances sufficient for observing the primary implemented physiological principles for a particular object.

## Goals – integrative model for computer smulation

This work deals with two hypotheses:

**Hypothesis 1 (formalization):**

***Modelica®, as the most recent generation of object-oriented equation-based computer language designed for the dynamic simulation of large complex systems and machines, is suitable for exact formalization of integrative human physiology.***

**Hypothesis 2 (integrative):**

***Mathematical formalizations of physiological knowledge about one organism can be integrated into one complex physiological model.***

Modelica® is the most recent generation of computer equation-based object-oriented language for physical modeling and is maintained by the Modelica Association (non-profit organization with members from Europe, USA, Canada and Asia). It contains all the necessary support for an exact definition of elementary physical laws, as well as support for the robust integration of complex systems. Many libraries for electrical, mechanical, magnetic and thermal domains already in use in the industry are already included, particularly among automotive companies, such as Audi, BMW, Daimler, Ford, Toyota and Volkswagen. However, usage of Modelica in chemistry or biology is still in its initial phase. Existing Modelica libraries such as Fluid, HemoltzMedia, BioChem, ADGenKinetics, FCSys, FuelCellLib and NeuralNetwork are not yet well-enough designed for all general physiological purposes. As such, the current thesis is a pioneering work regarding the use of Modelica in the field of physiology. Furthermore, because to date no-one knows whether it is really possible, the usability of this language in the field of integrative physiology serves as the primary investigation of this work.

The suitability of Modelica support for the development of the model should be based on giving the user the power to develop complex physical human health simulations simply by dragging, dropping and connecting small amounts of components from prepared libraries into schemes, such as electrical circuits with connectors independent of the causality of calculation. Each scheme should be able to be used many times in many other schemes with different values and parameters for each type of tissue, cell, organelle, receptor, macromolecule or any chemical element. To attain such support, physiology should be considered as a science in which all principles need to be exactly described according to the laws of nature. If mathematical equations are created without a physical background, the results are simply describing data. Without a physical description, the data becomes more valuable than their interpolation, or worse, extrapolation functions. A much better approach is to use physics to describe the experimental setting and results using physical quantities and physical units. When this is the case, well-known and valid relations from physics and physical chemistry can also be used.

Nowadays, physiological research is more focused on unknown elementary interactions, typically concerning one type of gene or molecule. These results should be integrated in order to gain a complex picture of physiological functions. Of course, the goal of the present work is not to integrate all of these interactions; however, it should provide the rules, the basis and the motivation for integrating them within Modelica.

Considering the long history of HumMod development described in section 1.1, integration into one complex model of human physiology appears possible using the approach formulated in the second hypothesis. Therefore, the support designed for physiology in Modelica should be used for re-implementation of the HumMod and even more so in the case of integrating the new acid-base and gases transport into one complex model of human physiology.

The final goal of the thesis is to implement this complex model using hierarchical graphical diagrams in Modelica, which will render it readable at any selected level, even for medical students. Similarly, the graphical diagrams already used only for illustrative purposes in some physiological textbooks, such as in (Kittnar and Mlček, 2009). However, these diagrams in Modelica are the real implementation for attaining a runnable physiological model that can be simulated using selected settings and its dynamical results are the analogies or measurable physical quantities of the body, such as blood pressure, chemical concentrations, temperatures, etc.

# Methods of base principles formalization

## Physical principles

Generalization of physical laws leads to similar principles existing between many physical domains. The motivation is not only to have similar mathematical expressions, but also to apply prepared methodologies to more than one domain. For example, an electrical circuit diagram that can be generalized for chemical, osmotic, hydraulic or other non-electrical systems. To do this, it is necessary to find analogies in physical quantities and physical laws.

Using only two quantities, the state of subsystems at interfaces can be described. One of these variables is flow in term of [Kirchhoff’s law](https://en.wikipedia.org/wiki/Kirchhoff%27s_circuit_laws), i.e., the sum of connected flows is zero at each place in the scheme. The second variable has to be nonflow, i.e., it has the same value in each connected side. The flows are usually changes in some quantity over time such as volumetric flow, molar flow, heat flow, electric current, magnetic flux or mechanical force. The nonflows should be some type of effort such as pressure, electrochemical potential, temperature, electric potential, magnetic potential or space position. Most of the physical laws from the already mentioned physical domains can be represented using equations that mention flow and nonflow physical quantities, for example, hydraulic resistance, diffusion, thermal conduction, Ohm’s law, etc.

### International system of units

As a result of a very long tradition in medicine, values are still represented in “medical” units instead of physical units of international standard ([SI](https://en.wikipedia.org/wiki/International_System_of_Units)). Even recently manufactured medical devices still use mmHg, calories and degrees in Celsius, among others. The problem is that these units primarily indicate their type of their measurement rather than their usability in terms of how they calculate physical laws. However, almost always exist the simple recalculation between “medical” non-SI and physical SI units. In Modelica, there is consensus that the running simulation is always in SI-units and recalculation from/to “medical” units can be done only prior to starting or after finishing the simulation.

Table 1: Selected non-SI units.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Unit conversion table | | | | |
|  | x kcal | = | 4186.8\*x | J |
|  | x kcal/min | = | 69.78\*x | W |
|  | x mmHg | = | 133.322387415\*x | Pa |
|  | x degC | = | 273.15 + x | K |
|  | x meq | = | 96.4853365\*x | C |
|  | x meq/min | = | 1.60808894\*x | A |
|  | x mosm | = | 0.001\*x | mol |
|  | x literSTP | = | 0.044031617\*x | mol |
|  | x literSATP | = | 0.040339548\*x | mol |
|  | x literNIST | = | 0.041571200\*x | mol |
|  | x iu of Erythropoietin | = | ? | mol/m3 |

Energy in medicine and chemistry has been observed for a very long period of time. However, one must not be confused by its different units and definitions. The researcher must be aware of the multiple definitions related to ‘calorie’, such as the international calorie, the 15°C calorie, the thermal calorie or ‘Calorie’ with a capital "C". The origin of this unit lies within the amount of thermal energy needed to heat one gram of water by one degree Celsius. However, because measurement conditions may differ, alternative definitions are necessary. In physiology, it is recommended that only international calorie, as defined in Table 1, be used. The flow of heat/energy is usually calculated in kcal/min; however, in physics, this is called ‘power’ and expressed in the SI unit of watts.

Pressure units in medicine are also primarily based on historical measurements. For many years, blood pressure was measured by the mercury sphygmomanometer, where the pressure is represented by the change in mercury hydrostatic column height. Since the scale of units on the column is in millimeters, the pressure unit is called millimeter of mercury, 'mmHg'. There also exists a very small difference between this unit and torr, which is caused by variance in measurement conditions.

Many physiological processes are based on electrical principles within the human body. The main cause of this is that each cell has a nonconductive membrane with molecular structures called channels, through which the fluxes of electrolytes can be precisely regulated. Even more, the cells use energy created by the metabolism to retain a small amount of electric potential, both internally and externally. This view gave rise to a unit called equivalents or “eq”. A charge of 1eq, for example, has 1mol of sodium cations (Na+). The fluxes of electrically charged ions can be in meq/min; however, in physics, the SI unit ‘ampere’ is more generally used.

Another odd unit describing the amount of substance is the osmol (“osm”), which has the same value as mol, but highlights the property that this substance cannot cross the membrane together with the flux of its solvent.

For gases, it is common to measure the amount as volume, which for specific measurement conditions is equivalent to the number of molecules. The International Union of Pure and Applied Chemistry ([IUPAC](http://www.iupac.org/)) set this standard condition for temperature and pressure ([STP](https://en.wikipedia.org/wiki/Standard_conditions_for_temperature_and_pressure)) precisely at 0°C and 100kPa. However, other standards also exist. For example, SATP is measured at 25°C and 100kPa, or alternatively, at the standard measurement condition of the National Institute of Standards and Technology ([NIST](http://webbook.nist.gov/chemistry/)), which is 20°C and 101.325kPa.

Chemical substances can be quantified in many ways, typically as the amount of substance in moles, which after multiplication by the Avogadro constant (6.02214129(27)×1023 mol−1) yields the number of substance particles. The amount of pure substance can be expressed from its molar mass, as each atom has known and recorded its molar mass in a table of elements, usually in the unit Dalton (gram per one mole). However, each substance has different molar mass and as a result, the conversion from mass to moles is always dependent on type (composition) of substance.

In physiology, units for directly-immeasurable substances are also used. Small concentrations such as 10-12 moles per liter are almost impossible to measure directly and only indirect measurements with immunoreactions or biological effects are known. However, the effect of some substances at these small concentrations can be crucial enough that they nonetheless need to be calculated somehow within physiological models. Most of these substances are called hormones, but some may be also enzymes (renin) or cytokines (erythropoietin). Pharmacological [international units](http://www.who.int/biologicals/expert_committee/en/) of these substances are defined as ratios to some extracted and purified standardized sample, which also has an unknown molar concentration, but known and well-described biological effects. As a result, the pharmacological international unit of substances often has no equivalent in SI units, but nonetheless needs to be used in physiological calculations as is. The danger of using these units is significant, especially in pathologies, because their biological activity is often times species dependent and is usually defined within a “normal” population. For example, in diabetes mellitus type 2 is presented an insulin resistance; therefore, the biological activity of the same amount of insulin for these individuals is entirely different than in a healthy individual.

### Redundant physical quantities

Some standardization should be conducted regarding the definitions of physical quantities. For example, each two variables in reciprocal relation, connected only by the trivial equation a=1/b, where the handling of both does not bring any additional information to the model, because their physical meaning is the same. Even the zero-infinity numerical problem can be easily solved by selecting variables such as the smallest representable floating point number or the highest representable floating point number, which are typically far enough from tolerance limits even for very long simulations.

The above reciprocal quantities are derivable from almost each physiological parameter such as hydraulic conductance, hydraulic resistance, hydraulic compliance, hydraulic elasticity, frequency, time period, solubility, volatility, dissociation coefficient, association coefficient, etc. To simplify this situation, it is better to select only one of each pair and build the physiological and chemical laws above as is usually done in physics, which will assist significantly in the elimination of redundancies inside shared interfaces.

Bad practice also includes using a unitless logarithm or other non-units, as well as non-physical variables in interfaces, even if the user has good documentation of how to convert these values. Values should always have an analogy in physical quantity, as this is more user-friendly and more intuitive for follow-up developments.

### Conservation laws

The next step of physiology formalization is the identification of physiological systems as physical systems. Based on interactions with environment, there exist [closed](https://en.wikipedia.org/wiki/Closed_system) and [open](https://en.wikipedia.org/wiki/Open_system_(systems_theory)) systems. An example of an open system is oxygen transport, where there is non-zero flow of oxygen from the environment to the body. In a closed system, there are not interactions between the body and the environment. For example, the chemical system of elementary particles in all its forms and in all places within the body that partakes in no exchange with the external environment is a closed particle system.

The [laws of conservation](https://en.wikipedia.org/wiki/Conservation_law) applies to closed systems. Neither energy, mass, an amount of substance or electric charge can be created from nothing. In dynamic models, it is very intuitive, because there is a rule stating that input flow to one component is always output flow from another component. However, in a steady state calculation (section 3.3), this system equation must be written explicitly, which is often not as intuitive; this is because in a steady state, flows from/to components are equal to zero.

## Modelica Principles

Modelica is an object-oriented, equation-based computer language that is standardized and maintained by the Modelica Association ([www.modelica.org](http://www.modelica.org)). The non-proprietary standard of this language is supported by many other projects, companies and organizations. As a result, there are a range of different environments available for this language. For example, [Dymola](http://www.3ds.com/products-services/catia/products/dymola/latest-release/), [OpenModelica](https://openmodelica.org/), SimulationX, JModelica, CATIA Systems, CyModelica, MapleSim and Wolfram SystemModeler.

My Modelica extensions called [Chemical](https://github.com/MarekMatejak/Chemical) (section 3.4), [Physiolibrary](https://github.com/MarekMatejak/Physiolibrary) (section 3) and [Physiomodel](https://github.com/physiology/Physiomodel) (section 4) should be able to run in all these environments, which support the [Modelica standard 3.3](https://www.modelica.org/documents/ModelicaSpec33.pdf) and [Modelica Standard Library 3.2.1](https://github.com/modelica/Modelica).

### Floating point numbers

From a mathematical point of view, the domain of real numbers is infinite. How then is it possible that it can be represented by a finite number of small bites, i.e., 32 or 64 and zeros? The answer is that it does so via approximations. There must always be some limits to precisions, some tolerances. [Floating point](https://en.wikipedia.org/wiki/Floating_point) numbers are represented by scientific notation with mantissa (a) and exponent (b) as a\*10b. Both mantissa and exponent are represented by a fixed number of bites. At single-precision floating point format there is one bit for a sign, 8 bits for an exponent and 23 bits for mantissa. This representation provides the smallest number as 10-127, the biggest number as 10127 and eps (the biggest number, e.g., 1.0 + eps = 1.0) as <10-6. This 32-bit precision is sufficient in most common cases; however, for specific calculations, better precision exists. For 64-bit, the double-precision floating-point format has 11 as the exponent (with a theoretical range from 10-1027 to 101027) and 52 bits for mantissa (with eps <10-15).

Though ranges and precisions are limited, floating points calculations can also present the user with other obstacles. The first of these is expressing the equality of real numbers. For example, what does it mean if we say that x is equal to zero, such as the condition x==0? If the number x is set to zero by the user and it does not by calculation change, its value truly remains zero; however, if it is calculated, it is always calculated with some precision. This means that a test of equality has sense only within this tolerance range. If we have set tolerance to 10-3, then we should be satisfied with numbers greater than -0.001 and numbers less than 0.001. Otherwise, the solver may reach the limits of number representations and as a result, will no longer reach equality.

The user tolerance definition for elementary non-iterative mathematical operations is not needed, but it is necessary for iterative numerical methods. The most common of these methods are numerical solving of differential equations (such as the Euler method, DASSL and others) or the numerical solving of non-linear equations by iterative approximations (such as the Newton method). At first glance, it seems that the tolerance for each tested variable is needed in error conditions of the iterative algorithms. However, this can be handled only by one relative tolerance and scaling of the variables. For this scaling, Modelica uses the attribute ‘nominal’, which can be included in all real variables.

### Object-oriented programming

[Object oriented programming](https://en.wikipedia.org/wiki/Object-oriented_programming) has been one of the most significant developments in computer science to date. The programing of large applications and systems becomes more simplified when re-using and extending already defined [objects](https://en.wikipedia.org/wiki/Object_(computer_science)). The notion of an object as a definition is particularly intuitive, because it copies human language and thinking. Each defined term is an object, which can have more occurrences; occurrence of the object is called an [instance](https://en.wikipedia.org/wiki/Instance_(computer_science)).

Each object can have properties. These properties can be a primitive variable such as a number, text, a true/false value or an object. This can create hierarchical decomposition within one system as one object to its subsystems as more and more detailed definitions of the owner parts. Particularly in physiology, these patterns are everywhere. When an object is present for a chemical reaction, a chemical substance, organelle, membrane channel, cell, membrane, tissue or physiological system, it is possible to compose new detailed objects such as large models of physiology, by using already described objects and by choosing the correct parameters of these new instances.

It is not necessary to make decomposition of problem from up to down[[1]](#footnote-2) or vice versa, because object-oriented thinking just support to start everywhere. There is only one condition for effective object-oriented programing: *the minimization of object numbers at the same time as the minimization of instance numbers to describe the same system according to the same rules*. This process is already used in mathematics and physical sciences, where the science can be exactly built from a small number of base rules by a finite minimized number of definitions.

The above idea can also be found in medicine books, where many principles or objects are generalized and finally applied to many parts of the body’s systems. For example, one family of membrane receptors can be used in many pathways and can interact with many different effectors.

Computer language principles are fairly straightforward. As a simple example, we defined two objects: class A and class B. Class B has only one parameter p, which can in each instance of B have a different value. Class A, as an example of class composition, contains two instances of class B, first with the parameter set at 1 and second with the parameter set at a value of 2.

class B "Definition of class B"

  parameter Real p "Real number parameter";

end B;

class A "Definition of class A"

  B b1(p=1) "First instance of class B";

  B b2(p=2) "Second instance of class B";

  B bArray[100](each p=3) "Array of one hundered instances of class B";

end A;

It is good practice to write the names of classes as starting with a capital letter and the names of instances starting in lower case. The object-oriented pattern includes any combination of parameters, variables and instances within the class definition. Other more sophisticated rules of object-oriented programing in Modelica can be described as modifications of this principle. The instances, variables and parameters can be hidden or publish outside the class by using the prefixes ‘private’, ‘protected’ and ‘public’, which provides useful restrictions for other users; this is referred to as [encapsulation](https://en.wikipedia.org/wiki/Encapsulation_(object-oriented_programming)).

Modelica language provides an analogy of these classic textual representations using graphical diagrams. Generally, a definition of each class is accessible as an icon on the left side of the Modelica environment, called the ‘Package Browser’. These classes can be as simple as elementary mathematical operations (see Figure 1) or very complex classes, which may be hierarchically composed from other classes.

Figure 1: Standardized definition of class Gain inside Modelica Standard Library (MSL).

To make an instance from any class in the ‘Package Browser’ it is necessary to have opened your class in diagram mode, then drag the selected class from ‘Package Browser’ and drop it to diagram. In general, it is not possible to modify integrated library classes; therefore, it is first necessary to create a ‘new model’ (using menu command: File > New > Model) with the unique name ‘MyClass’. Any class instance can be added to ‘MyClass’ just by dragging and dropping icons from the ‘Package Browser’ to the new model. Care must be taken, however, because double clicking any class in the ‘Package Browser’ can cause switches in class definitions.

Figure 2: Action sequence of inserting class instance.

The restricted class called ‘model’ without connectors can be flattened, translated and simulated with all its instance trees. This is because they have section ‘equation’, where all the equations and connections between instances are defined that is needed to calculate complete behavior (defined by the same number of equations as the number of unknowns). In the first step of this model, the [Open Modelica Compiler](https://www.ida.liu.se/labs/pelab/modelica/OpenModelica/OpenModelicaDevelopersCourse/resources/slides/4-OMC%20Compiler-overview-slides.pdf) translates structures into a flat model, where the same equation and algorithm are extracted, but without using object-oriented class definitions. This step can be done fully automatic and can generate a large amount of code compared to the original object-oriented representation. Then, the compiler automatically translate this flattened model into lower level computer language such as [C/C++](https://cs.wikipedia.org/wiki/C%2B%2B). This code is then run as a typical computer program; it has inputs such as initial setting and outputs such as the results of the simulation during the simulation time interval.

### Connections

Each library class has some possibilities for connecting their instances together. In the case of restricted classes, referred to as ‘block’ (as ‘Gain’ in Figure 1), they are only causal connectors, which can be ‘input’ or ‘output’ variables. The restricted class, referred to as ‘connector’ is in this case used only as a substitution elementary type for the real number (‘Real’) and with a causality direction prefix. After inserting any block instance to ‘MyClass’, all input and output connectors will be visible. Connections of these types of connectors are intuitive – each output can be connected to many inputs and connected variables will always have the same value.

Figure 3: Action sequence of connecting connectors.

Since the complex parts of a model can have many inputs and outputs, there exists in Modelica a special class called ‘expandable connector’. This connector does not have an explicitly defined list of variables or their causal direction, since this can be automatically generated from connections. For example, if we connect connector ‘c’ to this expandable connector called ‘busConnector’ as variable ‘busConnector.c3’, it automatically creates an implicit definition from the ‘c’ connector. This is designed only for large models and sends values from one branch to another branch of instances. Generally, it does not make sense to use expandable connectors for models where instances at the top level are composed only of elementary classes.

What allows for creating models such as electrical circuits is a connector defined by two variables: nonflow and flow. The flow variable has the prefix ‘flow’. It is possible to connect any number of connector instances of one definition together. These connections generate expected rules of circuits, where connected nonflows are equal and the sum of connected flows is zero.

The best practice in computer science is to use negative flow values for outflowing from the component and positive for inflowing to the component.

### Conditional inputs

The Modelica library for physiology can be designed to have the minimal number of components that are necessary for describing any processes inside the human body. Thanks to the support of steady state interfaces, the same components exist for dynamic and for equilibrium calculations. The conditional Modelica principle is also used for switching between parameters and for input to the block. These inputs are called conditional inputs and they are arranged in the same pattern as some components in MSL, for example, the component “Modelica.Analog.Basic.Resistor”.

## Building Modelica Libraries

The primary result of this work is “[Physiolibrary](http://www.physiolibrary.org/)”, a Modelica library for physiology, as well as a general Modelica library for electrochemical processes called “[Chemical](https://www.modelica.org/libraries)”. The whole of section 3 is dedicated to a description of these libraries, which serve as the basis for the [Physiomodel](http://www.physiomodel.org/) described in section 4.

Table 3: Physical connectors in the Physiolibrary and Chemical libraries, compared to electrical connectors of the Modelica Standard Library.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Connector: | | Nonflow variable |  | Flow variable |  |
| C:\Users\marek\AppData\Local\Microsoft\Windows\INetCache\Content.Word\ChemicalPorts.png | Substance | electrochemical potential | [J.mol-1] | molar flow | [mol.s-1] |
|  | Hydraulic | pressure | [Pa] | volumetric flow | [m3.s-1] |
|  | Thermal | temperature | [K] | heat flow | [W] |
| C:\Users\marek\AppData\Local\Microsoft\Windows\INetCache\Content.Word\PopulationPorts.png | Population | size of population | [1] | change of population | [s-1] |
|  | Electrical | electric potential | [V] | electric current | [A] |

Due to Modelica principles, it is possible to describe with a relatively small amount of physical types the basic rules of selected physical domains. First, we were implemented in Modelica complex models such as Guyton’s ‘Overall Circulation’ (Guyton, et al., 1972; Kofránek, et al., 2009), Ikeda’s ‘Body Fluids’ (Ikeda, et al., 1979; Mateják and Kofránek, 2010), Siggaard’s ‘Oxygen Status Algorithm’ (Mateják, et al., 2012; Siggaard-Andersen and Siggaard-Andersen, 1990), ‘Quantitative Human Physiology’ (Mateják and Kofránek, 2010) and finally Coleman’s ‘[HumMod](http://hummod.org/)’ model (Mateják and Kofránek, 2011). We can argue that reimplementation of models does not bring about new knowledge, but we nonetheless hope that this is not entirely correct and that my methodology will be useful for researchers designing their own theories, as well as for the integration of models.

As proof that new theories can be based on the physical laws already implemented in the Physiolibrary, we presented some of our models in physiological articles. The first concerns the modeling of pulsatile circulations (Kulhánek, et al., 2014a; Kulhánek, et al., 2014b), the second concerns the modeling of oxygen, carbon dioxide and hydrogen ions binding to hemoglobin (Mateják, et al., 2015). The integration of models works well because of object-oriented programing with well-defined interfaces using physical SI units, physical quantities, physical connectors and physical laws. The primary result of the integration of the above-mentioned models is the Physiomodel.

Each connector of the Physio- or Chemical library belongs to one physical domain (see Table 3), where the components can be connected using appropriate connector definitions. As seen in Table 3, most of the components are analogous throughout the domains. For example, the resistor in electrical circuits are analogous to those in the chemical domain (as diffusion), because the molar flow of substance is driven by an electrochemical potential gradient in the same manner as an electric current is driven by voltage gradient. To define these mathematical analogies in Table 3, the symbols ***e*** – as in the effort for connector nonflow variables – and ***f*** – as in the flow for connector flow variables – are selected. If more connectors are present in the component, they are differentiated by index.

Table 4: Analogies of selected Physio- and Chemical library components, based on connectors from Table 3 and with electrical components from the Modelica Standard Library.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Resistance** | **Accumulation** | **Stream** | **Inertia** | **Effort source** |
| *f1=G\*(e1 ‑ e2)*  *f1+f2=0* | *a=C\*e* | *f1+f2=0* | *f1+f2=0* | **e = E** |
| G… conductance | C…capacitance | F…stream flow | L…intertia | E..effort |
| C:\Users\marek\AppData\Local\Microsoft\Windows\INetCache\Content.Word\chemicalDiffusion.png  Chemical diffusion | substance  Chemical substance | Solution stream | Does not exist | Mole Fraction |
| C:\Users\marek\AppData\Local\Microsoft\Windows\INetCache\Content.Word\hydraulicConductor.png  Hydraulic resistance | elasticVessel  Elastic vessel | Does not exist | Inertia | Pressure |
| C:\Users\marek\AppData\Local\Microsoft\Windows\INetCache\Content.Word\thermalConductor.png  Heat convection | heat  Heat accumulation | Heated mass flow | Does not exist | Temperature |
| Does not exist | Population | Growth, Differentiation | Does not exist | Does not exist |
| resistor  Electrical resistor | ec  Electrical capacitor | Does not exist | Inductor | Voltage |

### Types

Most of the variables in mathematical models are real numbers and can be defined only by using an elementary type, i.e., ‘Real’. Why then does the Physiolibrary need so many elementary types for real numbers? Even the ‘Real’ is a simple type, which represents the number as described in the section [Floating point numbers](#_Floating_point_numbers); in Modelica, it can have attributes that differentiate between the meaning of values. This meaning is aimed at being user-friendly through the use of library components. With the help of these attributes, the Modelica environments can:

* find incompatible physical quantities in connections or equations
* recalculate the physical units in dialogs or in outputs
* assert the simulation when the values are not in their domain of definition
* increase the precision of results and speed up calculations

Checking physical quantities is extremely useful, especially for simple input/output connectors, which are in the Physiolibrary specified for each type in the package ‘Types.RealIO’. Using this type of connector instead of simple RealInput/RealOutput ones, a warning can be generated or even an error each time the user tries to connect, for example, an output connector of pressure value with an input connector expecting volume value.

Setting parameters using dialogs during implementation of the model can be simplified by the correct specification of physical units. Some environments can recalculate many non-SI units into expected SI units within models, but they need to know at which SI unit the value is (see the following section: [International system of units](#_International_system_of)). For the dialog setting of only one value within the model, constants are prepared for each type in the package ‘Types.Constants’. However, if the user applies any Physiolibrary types for his parameter or variable, the recalculations of this entire unit should also be automatically recognized.

The min/max assertions are not always set as the default debug feature for environments, but if they are, they can recognize bad results such as negative volumes, negative masses and temperatures less than 0 K. In correct physical models, these values should not be reached; however, the user always has the option to implement whatever equation they wish. Furthermore, because the correctness of each model cannot be decided automatically, any warning or assertion can be very useful.

Due to the compatibility of all Modelica libraries and models, all values should be calculated in SI units. This rule can generate strange dimensions for some values. For example, the SI unit for volume is cubic meter; however, in the body, compartments are measured in volumes of milliliters. Therefore, the numbers used for calculation will be a million times smaller than those generally used by a physiologist. However, this does not matter, because in the Physiolibrary, these types are defined as having ‘nominal’ attributes, which move the tolerance level from SI units to the typical values used in physiology.

### Blocks

The reason why Physiolibrary defines blocks is because they are missing in the Modelica Standard Library 3.2 (MSL). The blocks in the Physiolibrary are graphical implementations of simple mathematical operations such as reciprocity and power, among others, as well as more complex blocks for the interpolation of value by cubic function. This type of interpolation is needed for the implementation of empirical dependences between two variables, for which the physical explanation lies beyond the scope of the model. The interpolation can be implemented in different ways. The linear MSL look-up table approximates the value between known points using linear segments, which generates once derivation discontinuities. To solve this problem, we selected a cubic **spline** interpolation curve, which also has continuous first derivations. The curve is defined by a set of points with the coordinates x, y and slope. Approximated value v (coordinate y) is calculated from u (coordinate x), where point (u,v) lies on the curve. First, a segment of curve is selected, which is defined by the nearest curve points. During initialization, each segment has a prepared coefficient a, b, c and d of cubic equation ax3+bx2+cx+d=y as a means for reaching these definition points at defined coordinates and slopes. However, we must be cautious when using these cubic splines, because in some cases, the segment can have a non-injective function. In other words, inverse calculation of x from y can yield more solutions; which of these solutions is used may depend on a guest/previous value. My recommendation is to always draw this curve before it is used in the model and to modify the slopes in order to minimize the non-injective segments.

Figure 4: Non-invertible segment of cubic interpolation caused by wild slope differences.

#### Factors

Special patterns are used in the Physiolibrary in the form of factors. This idea has been applied in many physiological models in which multiplication effects are relative. In normal conditions, this effect is at value one. When it increases the base value twice its value is two; when it decreases the resulting value to half its value is 0.5. The resulting value can be affected by many factors, because at normal conditions, the product of ones is one. The graphical block for a factor always has one input at the top for an unaffected value and one output on the bottom for a resulting value, which is calculated as the effect multiplied by the unaffected value. The calculation of effects differentiates the factors from one another. The ‘Blocks.Factors’ package contains not only linear or cubic interpolations, there are also factors, which could quickly or slowly adapt the effect in time to the left-located input. This adaptation is called ‘**lag**’ and a simple mathematical filter is defined for it by Equation 1:

|  |  |
| --- | --- |
|  | Equation 1, Lag |

where t is time, x is an input, y is an output as adapted value and k is a parameter. The meaning of parameter k can be solved from a hypothetical situation where x is constant during simulation and y has another initial value than x. The solution to this simplification as a simple differential equation of one unknown function y shows that the halftime of y’s adaptation to x’s value is exactly ln(2)/k, as shown in Figure 5; this is because in the case of constant input x, the result is always y(ln(2)/k) = x + (y(0)-x)/2.

Figure 5: Lag in a specific setting as an exponential adaptation to constant value.

### Steady states

Each integrator is implemented in Physiolibrary 2.3 using a [steady state](https://en.wikipedia.org/wiki/Steady_state) interface. This provides support for changing the convergent system of differential equations to a system without derivations and with direct calculation of the fixed converging state. This feature is not designed for non-convergent systems such as oscillating or divergent systems. Though periodical processes in physiology are common, e.g., heartbeat, breathing, pericardial cycle or menstrual cycle, they can be implemented as a convergent system.

A convergent system does not have typical oscillating behavior; instead, oscillation is generally simplified to mean values and frequencies (frequency equals reciprocal period time). Surprisingly, for most variables, this significant simplification does not change the impact on other processes also calculated in mean values. Another example is if we want to view the specific current points in an oscillating period. This type of calculation delivers significant complexity in the form of additional processes, which can be neglected in a convergent system. For example, if we calculate to include convergent blood circulation, we can successively use mean pressures and mean blood flows with only two types of equation for elastic vessels and hydraulic resistor. However, if we want to calculate the values of pressure and blood flow continuously beat-by-beat, many other physical laws must also be included for precise dynamic calculation such as the opening and closing of valves, the inertia of mass flow, pressure waves with reflections in a 3D net of vessels, a fluid convection model inside the vessels and many others that disappear completely during the time the process takes place. Sometimes, these dynamic effects must also be calculated; however, for the long-term simulation of a typically healthy patient in typical conditions they can be eliminated without loss of generality.

Having a convergent system of differential equations, the point of convergence can be calculated by setting derivations to zero. This static time-independent situation is known as steady state. Typically, it can be used for very quick processes that converge in much shorter times than the time of simulation. Solving these processes dynamically using differential equations leads to stiff-equations, which can cause many problems in numerical solutions. Avoiding these very slow numerical calculations and their uncertain results can be effected by instead calculating the steady state (equilibrium) immediately.

The main problem with defining the steady state is that the swapping of branches in Equation 2 can generate dependent equations, especially when changing from a dynamic state to a steady state. For each dependent equation one additional equation should be added. These additional equations typically describe the state of the system, e.g., conservation laws or the environmental conditions.

|  |  |
| --- | --- |
|  | Equation : Steady state. |

For example, chemical equilibrium is the steady state in the chemical domain. The chemical reaction can be so fast that for long-term simulation, a dissociation constant with sufficient precision is always reached, rendering it unnecessary to calculate the dynamic for reaching a chemical equilibrium within the model. One solution to this issue is to implement the system only as equilibrium; however, the physical reality is the same as it is for models, where this dynamic is necessary. A better approach is therefore to implement the process with the possibility of selecting the option for dynamic or steady state calculation according to parameters prior to simulation. This implementation can be used for both short-term and long-term simulations.

Steady state does not always infer zero flow. For example, the steady state of a cardiovascular system is the state of non-zero, meaning cardiac output typically around five liters per minute. However, the total number of derivations, which increase or decrease the mean volume inside vessels remains zero, as defined by the steady state. Constant mean vessel volumes lead to constant mean pressures, driven only by hydraulic resistances. Furthermore, the systemic or pulmonary circulation at steady state can be calculated as systemic or pulmonary resistance without any dynamic adaptations caused by the spillover of blood volume.

# Results of base principles formalization

## Chemical domain

The chemical connector provides the molar flow “∂nA” of substance A. The [amount of substance](https://en.wikipedia.org/wiki/Amount_of_substance) “nA” can be expressed by the integration of this molar flow, as shown in Equation 3. In equilibrium, changes in substance “∂nA” remain zero and the amount of the substance “nA” remains constant.

|  |  |
| --- | --- |
|  | Equation 3: Amount of the substance. |

From the amount of the substance “nA” its [mole fraction](https://en.wikipedia.org/wiki/Mole_fraction) “xA”, [molar concentration](https://en.wikipedia.org/wiki/Molar_concentration) “cA” or [molality](https://en.wikipedia.org/wiki/Molality) “bA” can be expressed in the solution. If the amount of all particles in the solution is “nT”, the volume of the entire solution is “VT” and the mass of the solvent is “mS”, and the relation between mole fraction, concentration and molarity is Equation 4.

|  |  |
| --- | --- |
|  | Equation 4: Mole fraction of the substance. |

*For example, one liter of typical blood plasma, as presented in (Raftos, et al., 1990), has a total number of particles “nT” of about 51.8 mol and contains water as a solvent in a mass of 0.93 kg. The mole fraction of water is about 0.995 mol/mol and the mole fraction of chloride in the molar concentration 100 mmol/L is 0.00193 mol/mol.*

*Inside the cell, the situation is different. For example, in the red blood cells of (Raftos, et al., 1990), the total number of particles “nT” is roughly 38.7 mole per liter and the mass of water is only 0.69 kg per liter of intracellular fluid. However, these values yield the same mole fraction of water as in plasma (0.995 mol/mol). The mole fraction of chloride, the molar concentration of which is in erythrocyte around 50 mol/L, is 0.0013 mol/mol.*

The current theory of physical chemistry requires a correction coefficient between different ways of measuring a substance within a solution. This correction is called [activity coefficient](https://en.wikipedia.org/wiki/Activity_coefficient) . For the ideal substance, this is 1. However, it can vary for electrolytes, as predicted by Debey and Huckel (Debye and Huckel, 1923), Davies and Robinson-Strokes (Stokes and Robinson, 1948) and others. As a reason of this non-ideal behavior an activity of the substance “” should be inserted to the equation of the chemical potential instead of its mole fraction .

|  |  |
| --- | --- |
|  | Equation 5: Activityof the substance. |

Each chemical process will endeavor to equilibrate the [electrochemical potentials](https://en.wikipedia.org/wiki/Electrochemical_potential) of the substances. Electrochemical potential (Equation 6) describes the free Gibbs energy of one mole of the substance in the solution at defined conditions. This definition is the base equation for the physical chemistry (Mortimer, 2008).

|  |  |
| --- | --- |
|  | Equation 6: Electrochemical potential. |
|  | Equation 7: Chemical potential. |

where T is the temperature of the solution, φ is electric potential of the solution, R is gas constant (8.314), is the number of ion charge (0 is the substance and is not an ion), F is Faraday constant and is the chemical potential of the pure substance.

A typical chemical solution has electric potential equal to zero (φ=0), so that the electrochemical potential is the same as the chemical potential (Mortimer, 2008). Additionally, for all chemical processes in one homogenous chemical solution, the electrical part of electrochemical potential can potentially be neglected, because it is algebraically eliminated (the electric potential is the same in both sides of electrochemical equality for each chemical substance). After removing the electrical part of the electrochemical potential, only the chemical potential remains. However, for processes between different solutions, the electric part must be present. As a result, non-zero electric potentials can be presented between solutions.

### Chemical substance transports

The most intuitive equilibrium of electrochemical potentials is reached by equilibrating diffusion in order to attain one homogenous mixture with the same activities of all substances in all places.

It is somewhat complicated to imagine the equilibrium of uncharged substances through a semipermeable membrane. The electric part of the potential is zero, because the substance has zero charge (z=0). Therefore, only the chemical parts of the potential on both sides of the membrane are equilibrated. As it is the same substance on both sides, equilibrium of both pressure and chemical potential is reached if the activities are the same in both sides of the membrane. This equilibration of permeable particles is usually called [osmosis](https://en.wikipedia.org/wiki/Osmosis) and is the reason why water has the same mole fraction in plasma as in the intracellular fluids of erythrocytes. A second view of the same calculation can sometimes be simplified by impermeable particles, because a higher level of impermeable particles causes lower mole fractions in each permeable substance. In some cases, the mathematical simplification “ln(1-x) ≈ x” can be used, e.g., for such small x as mole fraction of plasmatic proteins on capillary membranes, where ln(xPermeants) = ln(1‑xImpermeants) ≈ xImpermeants. In this case, it appears as if there us equilibration of osmolarities (molar concentration of impermeable substances); however, in reality this is equilibration of electrochemical potentials for each permeable substance as expressed in Equation 8, where must be extended alongside the pressure dependence of its [state of matter](https://en.wikipedia.org/wiki/State_of_matter), as we do with constant molar volume Vm,A in case of incompressible substances.

|  |  |
| --- | --- |
|  | Equation 8: Osmotic pressure gradient across the membrane for uncharged incompressible permeable substance A. |

The other equilibrium is reached for chloride on cellular membrane of erythrocyte. Chloride can freely cross the membrane through a membrane channel, a process called chloride-shift. In contrast with electroneutral water, chloride has a charge number, -1, which takes membrane potential into the equation of the equilibrium. If we reorder this relation, we get the [Nernst membrane potential](https://en.wikipedia.org/wiki/Resting_potential) equation as a relation between the ratio of chloride inside and outside the red blood cells.

|  |  |
| --- | --- |
|  | Equation 9: Algebraic derivation of Nernst equilibrium of passive ion transport across membrane. |

*Donnan’s equilibrium is present on the red blood cell membranes, which generates a Donnan’s ratio of about 0.5 for each permeable anion of charge -1 (Raftos, et al., 1990). This ratio reflects a measurable electric potential of about -12 mV (Gedde and Huestis, 1997). For chloride, it appears that we can assume almost the same activity coefficients for the same substance on both side of the membrane, because the ratio of activities is the same as Donnan’s ratio of concentrations.*

### The chemical substance formations

In all of the above types of substance transport equilibrating, we worked with the same substance in the same phase and in the same solvent. This rendered the equilibrium independent of the base chemical potentials of the pure substance (), because both sides of the equation had the same value. However, if the chemical process created a new substance or changed its phase, was changed to a new substance. We called this process the formation of the substance A; if the reference substances from which the substances formed (e.g. H2(gas), O2(gas), N2(gas), C(graphite), Na(solid), etc.) had been carefully selected, we were able to measure the relative (to these reference substances; marked by the degree symbol “°”) formulate energies from the entire formation process in order to describe for any new formed substance in the chemical system by using Equation 10.

|  |  |
| --- | --- |
|  | Equation 10 : Chemical potential of the pure substance. |

where T is temperature, is [free Gibbs energy of formation](https://en.wikipedia.org/wiki/List_of_standard_Gibbs_free_energies_of_formation) of substance A (relative to selected reference substances), is [free formation enthalpy](https://en.wikipedia.org/wiki/Standard_enthalpy_of_formation) (heat energy consumed by the formation) of substance A (relative to selected reference substances) and is free formation entropy (the function of changed microstates by the formation) of substance A (relative to selected reference substances). The relation between enthalpy H, entropy S and [Gibbs energy](https://en.wikipedia.org/wiki/Gibbs_free_energy) G (G = H - T\*S) is a primary relationship of chemical thermodynamics (Mortimer, 2008). Typical in this situation is the formation of enthalpy , as well as Gibbs energy tabulated value at T0 = 25°C and pressure p0 = 100 kPa.

This description of molar Gibbs energies in substances is a dual way for describing the [equilibrium coefficient](https://en.wikipedia.org/wiki/Equilibrium_constant) of chemical processes. For example, the dissociation constant of chemical reaction A<->B in a solution without electric potential, defined as K=[B]/[A], has a relationship to the Gibbs energy of the reaction. If we set A as the reference substance (), then . This is also the result of the equilibrium of the chemical potentials, which defines the free Gibbs energy of reaction :

|  |  |
| --- | --- |
|  | Equation 11: Algebraic derivation of free Gibbs energy from the reaction of its dissociation coefficient. |

Using free Gibbs energies instead of dissociation constants is much better, because it automatically fulfills the [principle of detailed balance](https://en.wikipedia.org/wiki/Detailed_balance). This fundamental principle can be also translated as follows: the product of dissociation constants must be equal to 1 for each closed chemical cycle (closed system of the reactions, which ends with the same substances, phases, enthalpies and entropies as it starts). This relation of dissociation constants is not as intuitive as thermodynamic meaning, i.e., the free Gibbs energy of each closed chemical cycle is 0, which means that an isolated system does not consume or produce new energy. Mathematically stated, this can be stated as: the logarithm of 1 is 0.

As shown in the example of the simple reaction A<->B, the free Gibbs energy of the reaction () is the difference between the free Gibbs energies of product and substrate (). This rule is called Hess’ law and can be used for any chemical process with any number of substances and with any stoichiometric coefficients “v”. If we use the positive-negative stoichiometric coefficient notations (e.g., vB=2, vA=-3 for reaction 3 A <-> 2 B), we can extend the equilibrium to any number of reactants and products:

|  |  |
| --- | --- |
|  | Equation 12: Hess’s law in the equilibrium of chemical reaction 0<->v1A1 + v2A2 +… |

*For example the equilibrium of the Henderson-Hasselbalch reaction H2O + CO2 <-> HCO3- + H+ in red blood cells at 37°C can be calculated from the* [*tabulated*](http://www.update.uu.se/~jolkkonen/pdf/CRC_TD.pdf) *formation energies (shifted from 25°C to 37°C) of substances, as in ,* ***,*** *, as reaction free Gibbs energy . Thus, the negative decimal logarithm of the dissociation constant is pK=6.3. However, the bicarbonate is an anion and must therefore be corrected by activity coefficient in order* to *reach its physiologically measured intracellular concentration of 11.6 mmol/L (xHCO3=3e-4, aHCO3=1.9e-4) at pH=7.2=-log10(aH+) (aH+=6.31e-8), where the mole fraction of the free dissolved carbon dioxide xCO2=aCO2=3.22e-5 and the mole fraction of water xH2O=aH2O=0.995 as in data presented by Raftos et al. (Raftos, et al., 1990). The same pattern with the same activity coefficient of bicarbonate can be applied to blood plasma in order to reach a typical concentration of bicarbonate 24 mmol/L at pH=7.4 and at the same mole fraction of carbon dioxide xCO2=aCO2=3.22e-5.*

In our example, we calculated the free dissolved gas in liquid. Gases such as carbon dioxide or oxygen are equilibrated in the lungs between their gaseous and aqueous states of matter. The equilibrium of this process is in physical chemistry described as Henry’s law. Thanks to this fixed linear dependence between gaseous and aqueous forms, it is possible to exchange the meaning of the partial pressures of gases with their concentration as a free dissolved form. The relation between mole fraction xA,g and partial pressure pA of the substance A in a gas of total pressure pT is xA,g=pA/pT. Henry’s coefficient can be defined as kH = xA,l / xA,g , where xA,l is a mole fraction of free dissolved substance A in a liquid. Additionally, this coefficient can be rewritten to Gibbs energy in a chemical process:

|  |  |
| --- | --- |
|  | Equation 13: Algebraic derivation of free Gibbs energy of the gas dissolution in liquid using its Henry’s coefficient. |

*The primary problem with these parameters is that they are shifted from mole fraction to molality in water such as in NIST (U.S. Department of Commerce, National Institute of Standard and Technology) tables. Therefore, if we use a value from* [*NIST*](http://webbook.nist.gov/chemistry/)*, e.g., 0.034 mol/kg/bar for CO2, it should be recalculated to mole fraction units using water molar mass 0.018 kg/mol as kH=0.034\*0.018. Furthermore, the calculation of mole fraction of free dissolved CO2 at pCO2=40 mmHg in blood is xCO2=0.034\*0.018\*40/760=3.22e-5. The molar concentration in erythrocytes can be expressed by the multiplication of the total amount of substances 38.7 mol in one liter of intracellular fluid and the resulting value of 1.24 mmol/L is in good agreement with* [*Siggaard’s data*](http://siggaard-andersen.dk/OsaTextbook.htm) *(Siggaard-Andersen and Siggaard-Andersen, 1990).*

### Change of phase of the substance

Change of phase is in physiology primarily connected with evaporation of water in the lungs or as sweat through the skin. In terms of a chemical equation, this also means the equilibration of chemical potentials in the water in the liquid and gaseous phases.

|  |  |
| --- | --- |
|  | Equation 14: Algebraic derivation of equilibrium in water evaporation. |

The free formation enthalpy of liquid water is [-285.8 kJ/mol](http://www.update.uu.se/~jolkkonen/pdf/CRC_TD.pdf) and for gaseous water, it is [-241.8 kJ/mol](http://www.update.uu.se/~jolkkonen/pdf/CRC_TD.pdf). The free formation Gibbs energy of liquid water is [-237.1 kJ/mol](http://www.update.uu.se/~jolkkonen/pdf/CRC_TD.pdf) and for gaseous water [-228.6 kJ/mol](http://www.update.uu.se/~jolkkonen/pdf/CRC_TD.pdf) at 25°C and 100kPa. From these values, free entropy of formation can be expressed as -163.14 J/mol/K for liquid water and -44.27 J/mol/K for gaseous water at 25°C and 100kPa using Equation 15. As a result of almost constant enthalpy and entropy of formation, the free Gibbs energy of vaporization to temperature T at pressure 100kPa can be recalculated.

|  |  |
| --- | --- |
|  | Equation 15: Water evaporation molar energies. |

*The free Gibbs energy of vaporization is -44030+118.867\*T and the vapor pressure of water is xH2O\*exp((-44030/T+118.867)/8.314 bar. Using this relation for pure water at a body temperature of 310 K (37°C) and normal atmospheric pressure of 100kPa we get the vapor pressure 6.22 kPa (47.2mmHg) at equilibrium of the vaporization process. This value is close to the observation of* [*6.28 kPa (47.7mmHg)*](http://www.thunderscientific.com/tech_info/reflibrary/its90formulas.pdf)*.*

### Stream, degradation and clearance

The chemical substance can be transported together with their chemical solution. The component modeling volumetric flow of solution is called the **stream**. Typically, the stream is used alongside the air transportation of oxygen or carbon dioxide during ventilation, as well as for the transportation of substances using blood circulation. The calculated molar flow of an entrained substance labeled as ∂*nA* is in this instance the molar flow of the entire solution ∂*n* multiplied by the mole fraction *xA* in the original stream, as shown in Equation 16.

|  |  |
| --- | --- |
|  | Equation 16: Stream. |
|  | Equation 17: Clearance. |

An analogy of stream calculation is in medicine referred to as **clearance**, which is used for calculating extracting substances from the body such as kidney excretion, liver metabolism, enzymatic processes, etc. For a defined substance, the *Clearance* parameters are measured as an amount of solution flow that has been fully cleared from the substance. Since we use the mole fraction instead of molar concentration, we need to convert mole fraction to molar concentration using the mole density of the solution *ρm* [mol.m-3], which is the total amount of the substances per volume unit. In contrast with stream, there is not loss of solution.

One must be careful, because clearance is not only one possible way of removing substances from the body. In some cases, there is also the passive **degradation** of molecules in the entire solution volume (Equation 18). In contrast with clearance, degradation is dependent on the distribution space of a substance. If there is no other change in the substance and only degradation in the constant amount of solution *n* takes place, then the concentration reverts to half the amount following the amount of time expressed as the parameter *HalfTime*. In the condition of the constant solution amount this can also be rewritten to Clearance = (n/*ρm*)\*ln(2)/HalfTime.

|  |  |
| --- | --- |
|  | Equation 18: Degradation. |

The simplest chemical components for chemical substances can be gained by simply putting a prescribed number as molar flow of substance; this is referred to as **Pump**. This molar flow is usually calculated by user-defined schemes, for example, using normal flow as a parameter affected by factors such as those described in the section [Blocks](#_Blocks).

### Macromolecule equilibria

Macromolecules in physiology are common and include, for example, proteins, DNA and RNA. These molecules are typically polymers composed from small amounts of base elements such as amino acids for proteins or nucleotides for DNA. Through the polymerization of these base elements one strand of macromolecule is created, where not only the order of the base elements are critical for future functions, but where the entire space with all the types of presented bounds determine the function of the macromolecule.

From a chemical point of view, there are many distinguishable forms of macromolecules. For example the side chains of some amino acids can be presented as base or conjugate acids, while some can even be presented in an acid or conjugate base form at the typical cellular or interstitial pH. Fortunately, reactions with ligands can be independent on different sites in macromolecules, which simplifies the calculation of equilibrium. The mole fraction of the specific state “sQ” (defined by selected quaternary conformation “Q” and by the state of each independent site “i” at equilibrium) is calculated by Equation 19. The equation can be read as the probability of the selected form in quaternary conformation (xsQ), which is the probability of quaternary conformation (xQ) multiplied by the probability of each selected site form in the quaternary conformation (xi/xQ).

|  |  |
| --- | --- |
|  | Equation 19: Speciation. |

As a result of these general equations for attaining equilibrium of macromolecules, changes between quaternary conformations can be easily expressed. For xample, the allosteric effect on hemoglobin can be described by tensed and relaxed conformation of hemoglobin tetramer molecules, where the binding of oxygen is in each conformation independent. However, the affinity of oxygen in tensed conformation is much higher than in relaxed conformation. This simplification was first presented in a model by Monod-Wyman-Changeux in 1965 as the basis of all allosteric regulations. However, they are only calculating the very simplified one ligand situation.

Our research pertaining to these allosteric effects is able to extend the model of allostery with many possible quaternary states and with many ligands. The example of usage this extended revolutionary approach is the first paper in appendix {Mateják, 2015 #20} I designed the mathematical description of the hemoglobin model including oxygenation, carboxylation, Bohr’s effects and heat balance. Co-author Stanislav Matoušek completed the review and scientific language support for the paper. Co-author Tomáš Kulhánek identified the parameters of the model in order to fit experimental data. As shown in what follows, the model describes the interconnection of all phenomena that are joined together nonlinearly. Therefore, it is not possible to describe the phenomena separately as independent processes.

## Thermal domain

It should not be surprising that in the thermal domain, heat energy is accumulated as shown in Equation 20. From **accumulated heat,** temperature [K] can be calculated using the properties of materials such as their specific heat [J.kg-1.K-1] and mass [kg] (Equation 21). Since in human physiology temperature is regulated to 37°C (=310.15 K), relative heat is shifted to this value. The negative value of heat means a lack of heat to 37°C while the positive value of relative heat means heat excess and a higher temperature.

|  |  |
| --- | --- |
|  | Equation 20: Heat. |
|  | Equation 21: Temperature. |

The connectors in heat domains use temperature [K] as nonflow and heat flow [J.s-1] in establishing meanings of change in heat energy. The connector was inherited from the package Thermal.HeatTransfer of Modelica Standard Library (MSL 3.2.1), which creates compatibility with all standard thermal components of that package.

**Heat conduction** is driven by a temperature gradient as shown in Equation 22. Heat is transferred from a warmer to colder environment until the temperature is equilibrated. The speed of conduction is determined by the parameter Cond, which can also be expressed as the reciprocal value of heat resistance.

|  |  |
| --- | --- |
|  | Equation 22: Conduction. |

Heat is also transported together with mass. Each loss of mass will decrease absolute heat, but it does not change temperature. This situation is an analogy of substance molar flow when the entire solution is outflowing. Additionally, Equation 16 and Equation 23 are similar, but the meaning of variables is different. The **heat stream** is based on mass flow [kg.s‑1] not molar flow and there is no molar fraction, but a “concentration of heat energy” expressed as the multiplication of temperature [K] with the specific heat of the mass [J.kg-1.K-1].

|  |  |
| --- | --- |
|  | Equation 23: Heat change by water evaporation. |

Typically, the microcirculation is so effective that the outgoing blood from capillary nets has the same temperature as the tissue around the capillaries. The principle of heat transfer from blood to tissue acts like an **ideal radiator**, because the radiator also overflows with heated liquid. The specific heat [J.kg-1.K-1] of this liquid is named *SpecificHeat*. The amount of transferred heat to the environment is proportional to the flow of the liquid inside the radiator, called *massFlow* [kg.s‑1]*.* Maximal heat flow to the environment can be limited by the equilibrium of temperatures of outflowing liquid and the environment around the radiator can be controlled as in Equation 24. Equation 25 states that all heat energy of the inflowing liquid (Ti\*SpecificHeat) is divided only to heat energy transferred to the environment (heatFlowToEnv/massFlow) and the heat energy of the outflowing liquid (To\*SpecificHeat).

|  |  |
| --- | --- |
|  | Equation 24: Ideal radiator. |
|  | Equation 25: Heat flow. |

However, blood can transfer about 5% more heat from working muscles to the lungs than is calculated by Equation 25, due to the endothermic behavior of hemoglobin deoxygenation (Mateják, et al., 2015). This additional heat is not accumulated to mass when temperature changes, but released by chemical reactions when the form of molecules changes as described in the above sections (chemical enthalpy). This type of chemical enthalpy also takes place during sweating, in the water change phase from liquid to gas. This process effectively cools down the skin even when the environment temperature is higher than the temperature of the skin.

## Hydraulic domain

The modeling of the cardiovascular system is based on hydraulic principles where volume [m3] in **elastic vessels** generates pressure [Pa] and the pressure drives blood flow [m3.s‑1] through circulation. The main component in the accumulation of volume is called ElasticVessel and is described with Equation 26 and Equation 27. As a result of the elastic properties of blood vessels, there is an increase in pressure, together with an increase in the volume inside this component. This proportional dependence is set by the parameter Compliance [m3.Pa‑1], which is the property of the wall of the blood vessel. For example, compliance is bigger for systemic veins, where the same additional volume does not increase the pressure as much as in systemic arteries. The walls do not generate positive pressure within when the volume decreases bellow V0 and they lose tension. The result is the same pressure inside as outside the vessel.

|  |  |
| --- | --- |
|  | Equation 26: Volume. |
|  | Equation 27: Elastic Vessel. |

Fortunately, the typical working state of elastic vessels at each place during each phase in the heart period is at the first branch, where volume is >V0. The second additional branch solves critical situations, which may appear, for example, after massive hemorrhage. The external pressure around vessels PExt are typically set to zero with the exception of a local bandage or intrathorax pressure. Negative intrathorax pressure around ‑500 Pa is a result of respiration quotient. Inside the lungs, more oxygen is bound by hemoglobin than carbon dioxide is released from the blood, leading to a lack of molecules within properly working alveoli. This causes a small pressure debt to the intrathorax extravascular pressure during the entire respiration period.

The volumetric flow through segments of vessel is driven by the pressure gradient. This component is called the **conductor** or hydraulic resistor. Flow moves from a higher to a lower pressure; its value is determined with conductance Cond [m3.s‑1.Pa‑1], which can be expressed with a reciprocal value in the form of hydraulic resistance [Pa.s.m‑3].

|  |  |
| --- | --- |
|  | Equation 28: Conductance. |

Conductance is dependent on the current radius of the vessel. Vasoconstriction and vasodilation changes the radius, which in turn changes conductance. Higher conductance means higher flow for the same pressure gradient.

Pressure in liquid is also generated by gravity. Hydrostatic pressure is dependent on depth below the surface, on the density of the liquid and on gravitational acceleration. For example, pressure of one atmosphere is at the bottom of a 0.76 m high column of mercury or at the bottom of a roughly 10 m high column of water. This phenomenon causes additional blood pressure in the lower parts of circulation and lower blood pressure in the upper parts, as expressed by Equation 29. The classic formula (gravity\*density\*height) is here extended with pumping effect (*pumpE*), which significantly helps to break the **hydrostatic column** within the vein’s valves.

|  |  |
| --- | --- |
|  | Equation 29: Hydrostatic pressure gradient. |

Typically, one point is selected for circulation (e.g., heart aortic valve). Height below this point is positive. Height above this point is negative. Change in the orthostatic position of the body during standing or lying is represented by changing the height of computed vessels. Gravitational acceleration (*gravity*) in the earth’s surface is always set to 9.8 [m.s‑1]. The pumping effect changes with movement of the legs, because the segments of leg veins between valves can push the blood up only when the leg’s skeletal muscles periodically contract and relax.

**Ideal valve** is a designed hydraulic component that acts as conductor, but with different resistance for each flow direction. Forward flow has high conductance (low resistance) *Gon* and backward flow has low conductance (high resistance) *Goff*. The second branch of Equation 30 is valid during the opened phase (pressure gradient > 0); otherwise, if the valve is closed, the first branch occurs instead. At the break point defined by pressure gradient 0 both branches with zero are valid.

|  |  |
| --- | --- |
|  | Equation 30: Valve. |

Backward conductance is typically very small; small volumetric flow in the case of a closed valve can be generated. However, this flow can be so small that it can be described by the swelling of the valve membrane without any direct connection between liquids on both sides.

The resistance of mass to any change in motion is called **inertia**. Volumetric flow has the tendency to continue forward and as a result, volumetric flow will continuously react to changes in pressure. The other view to the Equation 31 concerns generating pressure proportionally to the change of flow. The higher parameter Inertance [Pa.m‑3.s2] means the higher pressure gradient answers to the same changes in volumetric flow.

|  |  |
| --- | --- |
|  | Equation 31: Inertia. |

The inertance of fluid in the vessel segment can be expressed as the density\*length/cross-sectional area. Typically, inertia is most important in the aorta, where in each heart cycle blood flow starts and stops from the left ventricle.

## Population domain

Physiology models must also take into account organisms, cells, viruses, bacteria, etc. As in a predator-prey model, there is also an accumulation of the members of the **populations** that can reproduce or die. Even though all the calculations are in real numbers (as in Equation 32), the results can be rounded to integers quite easily. The number of cells is typically very high and this approximation, alongside floating point numbers, is able to count any large amount of members.

|  |  |
| --- | --- |
|  | Equation 32: Population. |

The number of members is called *population(t)*. An increase or decrease in members is called *populationChange(t).* One type of cell is generally selected as representing the population; for example, red cells, which are produced by erythropoiesis in bone marrow. Furthermore, as a population can be implemented, so can one phase of cell maturation, differentiation or reproduction, in which the properties that differentiate these cells from others exist.

Reproduction, mortality and stream are represented by the same equation. The main idea is the proportional dependence of population change on population size, as expressed in Equation 33.

|  |  |
| --- | --- |
|  | Equation 33: Change. |

The parameter *changePerPopulationMember* can be recalculated from lifetime or half-life, where *lifetime = ln(2)\*half-life* and *changePerPopulationMember = 1/lifetime*. Even though these conditions and behavior has been simplified, it can nonetheless illustrate the primary trends of dynamics and can fit the steady states of the system.

# Methods of integrative physiology formalization

The integrative model of human physiology is a representation of physiological theory. As in all natural sciences, physiology must also be based on real experiments and logic. The model of physiological functions must be proven with measurements and the power of deduction regarding proven models can be used to prove new facts or identify questionable situations. All approaches must be deterministic and there cannot be any mathematical controversy.

The following exact logical definitions and theorems claim that it is possible to develop by *integration, reduction* and *extension* from previous models a model that must be *at least as good as* the previous models. Robustness (the ‘at least as good as’ operator) of the model must take into account all *real experiments* behind the models (the ‘be described’ operator) and the dimensions and size of the *image* of the model (number and type of results). This comparison of models can be defined by a comparison of all their real experiments. Integration of new knowledge can be done without losing any potential of the model in order to solve previously described real experimental data. The extension of the image of the model with new physiological definitions of new variables should also be noted as an improvement. On the other hand, an extension with new and unnecessary parameters and with new inputs is unwanted, because it brings correlations and problems needing unique identification to individual objects. The reduction of *domain* (number and type of parameters) will also be exactly defined to gain a model at least as good as the original. As a result of these rules of development, the better model should describe more data with more outputs and less parameters. Having logical proof at each instance that the new version of the model is better than the previous version will indicate that the new version is better than each preceding model in its history.

|  |  |
| --- | --- |
| 1. **Real Experiment** |  |
| ***Real experiment*** *R*in [domain of definition](https://en.wikipedia.org/wiki/Domain_of_a_function) *D* giving data in [image domain](https://en.wikipedia.org/wiki/Image_(mathematics)) *I*, where for each setting ***s*** is the run of the experiment giving measurable data *R(s)*. | |
|  | |

Each experiment must be [reproducible](https://en.wikipedia.org/wiki/Deterministic_system) with the same output at the same settings and without any other assumed conditions. The range and type of these value settings will determine the domain of the definition. For example, if we have only a parameter *s1*, which can reach values from 0 to 1, then the domain of the experiment is one-dimensional interval <0,1>. Outputs must be handled similarly, for example, as values from -10 to 10 and having image as a one-dimensional interval <‑10,10>. Typically, settings remain constant during the experiment and the data can vary during the experiment in the form of measured outputs.

|  |  |
| --- | --- |
| 1. **Default Setting** |  |
| ***Default setting*** are fixed values in domain *D*, which represent the selected state as a set of parameters and inputs for all experiments. | |
|  | |

Default settings must be consistent, which can be achieved by measuring all values for one individual object, at well-defined conditions within the specified fixed time. This snapshot of all values is typically used for parameters, which can be ignored or neglected in experiments to reduce the complexity and dimension of domain *D*. This is very comfortable, because by using default settings experiments can be designed with only a few parameters, which directly determines the outputs. Default settings should be selected to represent the normal state of a normal patient in normal conditions in order to achieve reasonably normal data even for long-term experiment *R(D,I)*. If it is not really necessary, default values should not be changed during development.

|  |  |
| --- | --- |
| 1. **Model** |  |
| ***Theoretical model*** *M* in [domain](https://en.wikipedia.org/wiki/Domain_of_a_function) *D* giving data in [image](https://en.wikipedia.org/wiki/Image_(mathematics)) *I*, where for each setting *s* is the simulation of the model *M* giving the simulation outputs *M(s)*. | |
|  | |

A model is always an approximation of reality. Even if we use fundamental physical relations, we should not be sure that a model will exactly match the reality. There is the same assumption when working with a model as for conducting a real experiment: the unique results for each parameter of the model (reproducibility). In other words, all necessary settings should be known before both real and theoretical experiments. For this reason, parameters and inputs are usually defined as physical quantities with physical units. Additionally, the model setting *s* typically remains constant during the simulation and the outputs can vary over time. This dynamic behavior is caused by a differential equation, which can dynamically react to prescribed changes of the settings. Each model *M*in domain *D* should give the reasonable results for the default values . However, this condition is not part of the model definition, but simply a logical proposition.

|  |  |
| --- | --- |
| 1. Experiment **is described** by model |  |
| Experiment *R* ***is described*** by model *M* in domain with precision *P* using [metric](https://en.wikipedia.org/wiki/Metric_space) *dIR* of image *IR* if and only if the image of the experiment *IR* is a subset of the image of the model *IM* and the distance between data and simulation for each setting is less than *P*. | |
|  | |

Figure 11: Definition domains of model and experiment and the position of partial settings.

The setting *s={sRM, sR}* arises from the design of experiments. The model can reduce this setting only to the primary information necessary for determining the right behavior in each case (as setting *sRM*). Comparison of the measured data with the results of the simulation can be done in different ways. An error calculation called a metric must always be used, which provides the information about how similar the curves are. Any mathematical [norm](https://en.wikipedia.org/wiki/Norm_(mathematics)) can be used. For example, having points in a specific time, the sum of square distances between the simulation and measured points can be used. The value of precision *P* must be selected appropriately in order to select the metric and set the precision of measurements. The ideal value of precision P is 0.5% for the Euclidean norm. However, this precision is too difficult for most physiological measurements; therefore, even a 10% error is currently tolerated by the Physiomodel.

|  |  |
| --- | --- |
| 1. **Model Comparison** |  |
| Model M1 ***is at least as good as*** M2 if and only if it can describe all experiments that can be described by M2. | |
|  | |

Almost all debates about different models that represents the same thing revolves around recognizing which model is better. This definition cannot be used to compare all models, because there may be situations when a second model is better than the first, even if in other situations, the first will work best. This operator, “be at least as good as”, can therefore be applied if all real experiments described by the “worse” model can also be described by the “better” model. From the “is described” operator, this new operator inherits the following useful properties: . Therefore, the better model has a number of output variables at least as big as the worse model. The equality of models occurs when M1(D1,I1) P M2(D2,I2) and M2(D2,I2) P M1(D1,I1). In this situation, the direct result is the equality of images (I1=I2) and the same set of experiments, which are described by both models. The other property of the operator ‘at least as good as’ is transitivity, which means that if and then . This transitive relation is a result of definitions of both “be described” relations as , which is the definition of the “be at least as good as” operator between M1 and M3. Having the transitive "is at least as great as" operator for model comparison is critical for model development, because if the new version is at least as good as the previous version of the model during each phase of development, then the new version must also be at least as good as all versions in the history of the model’s development.

|  |  |
| --- | --- |
| 1. **Incomparable Models** |  |
| Model M1 ***is incomparable*** with M2 if and only if M1 is not at least as good as M2 (precision P) and M2 is not at least as good as M1 (precision P). | |
|  | |

Negation of this relation between two models must render one of the models at least as good as the other model. In other words, if the models are not incomparable then they are comparable. The incomparable operator is commutative, so if then .

|  |  |
| --- | --- |
| 1. **Integration** |  |
| MAB is an ***integration*** of two models MA and MB­ in precision P if and only if MAB describes all data that are described using the model MA or the model MB | |
|  | |

Having models for different types of experiments, there must be an option for how to merge them into one model, which will be able to describe all of these experiments. This integration is the most problematic stage of model development, because for each process, only one relation must be selected, even if it is described in both models differently. The new integrated model must generally be re-implemented with a new theory that describes both types of experiments. If it is not possible to establish a smooth mathematical relation for both groups of behavior, it can be always be implemented as an “if-then-else-” solution. The existence of conditional integration is proof of the integration theorem, which states that there must exist some integration for each of the two models.

Integration theorem:

For example, if we have two models of aortic valve, one for an opened valve and a second for a closed valve, then we can integrate them by using conditions pertaining to the opening and closing of valves. If the integration is based on improved physical theories, then the final model can achieve much higher potential for solving more complicated experiments than the original experiments based on the once separate, but now integrated models. This integration phenomenon was observed when we integrated, for example, the new physical chemistry theory with HumMod 1.6. The original HumMod 1.6 was based on the equilibration of molar concentration and osmolarity. However, using these old relations, it was not possible to describe equilibrium on the erythrocyte membrane. Using a new physical theory, however, we solved not only this problem, but also brought to the model a solution for all electrochemical processes, which is based on new physical chemistry relations, as noted in section 3.4.

|  |  |
| --- | --- |
| 1. **Domain reduction** |  |
| *Mr*is a ***reduction*** of the model *M* in precision *P* to new domain Dr if and only if its domain *Dr* is a subset of the domain *D* and *Mr* describes all experiments described by the original model *M* in the same image. | |
|  | |

The model after domain reduction is at least as good as the original model. The image remains the same. The domain is reduced typically by removing correlating parameters or making parameters more invariant in order to have the same default values during all experiments (parameters become constants). Making values more invariant has long tradition in physiology, typically in terms of scaling values according to weight, skin surface, height, age, sex, etc.

|  |  |
| --- | --- |
| 1. **Image extension** |  |
| *Me*is an ***extension*** of the model *M* if and only if the original image *I* is a subset of the extended image *Ie* and *Me* describes all data described by the original model *M* in the extended image *Ie* with precision *P*. | |
|  | |

The model after extension of the image is at least as good as the original model, as only the image was extended . The domain and the experiments remain the same.

Development using these rules of integration, image extension and domain reduction always consistently provides the next version of the model with all previously integrated knowledge. However, it is necessary to collect the knowledge of integrated real experiments and their measured data in order to know exactly the limits of the current version of the model. These references to scientific research are an inseparable part of any complex model.

# Result of integrative physiology formalization

Diagram 1: Physiomodel subsystems, top-level diagram implementation.

## Cardiovascular system

The cardiovascular system is composed of the heart, pulmonary circulation and systemic circulation, as implemented in Diagram 2. These components are connected using the Physiolibrary’s hydraulic connectors, where pressure and volumetric flow is hidden behind the black line connections. Both pulmonary and systemic circulation has the same behavior during steady state as the simple hydraulic resistor. The heart behaves as a continuous hydraulic pump during steady state. However, during dynamic middle-term simulation, the situation is more complex and blood volumes can dynamically spillover between blood vessels, changing blood pressure and blood flow. The pumping heart is more complexly described in the ‘Heart’ subsection and both dynamic circulations are described in detail in the ‘Circulation’ subsection using tissue arterioles, capillaries and venules from the ‘Microcirculation’ subsection.

As blood volume and hematocrit strongly influences both blood pressure and blood flow in all areas of the cardiovascular system, their implementation is also shown in Diagram 2 as ‘red cells’ and ‘blood properties’. These components calculate the amount of red cells, blood volume, hematocrit and blood viscosity effect on hydraulic conductance, among others and are described in the ‘Blood’ subsection.

Diagram 2: Cardiovascular system; the black line at the top-right represents the pressure and blood flow at the end of pulmonary veins, while the bottom-right black line notes the start of the aorta, the bottom-left line notes the end of systemic veins and the top-left line shows the start of pulmonary arteries.

### Heart

The model for blood pumped by the heart consists of models for heart atriums, ventricles, the sinoatrial node, atrial pressure receptors and atriopeptin. The heart components have four hydraulic connectors that connect the veins and arteries. From systemic veins blood is transferred directly into the right atrium, from which the right ventricle is filled. The right ventricle is emptied into pulmonary arteries using a connector in the left bottom corner of the heart icon (see on Diagram 2). After oxygenation in the lungs, blood moves to the left atrium and left ventricle, from which it is ejected into the aorta (Diagram 2). The pathological state of mixing deoxygenated blood with oxygenated blood, when the foramen ovale is opened, is not yet implemented. Thus, during steady state, the flow in the connector of veins is the same as the flow in the connector of arteries for both halves of the heart.

The sinoatrial node calculates the heart rate and will be described together with low pressure receptors in a section on autonomic neural activity. Atriopeptin, as a hormone produced by the heart in answer to blood pressure inside heart atriums, will be described in a section about hormones.

As long-term heart activity can be modeled using the mean values of pressures and flows, there must not be solved beat by beat. Instead of using dynamic periodical values, it is precisely calculated in values, which make up the arithmetical average of the flow or pressure during each heart period and is referred to as mean variables or mean values. In these conditions, the heart atrium can be implemented using the simple elastic vessels in the Physiolibrary as defined by Equation 26 and Equation 27, and represented by yellow circles in Diagram 3.

Diagram 3: Heart; deoxygenated (oxygenated) blood goes from systemic (pulmonary) veins to pulmonary (systemic) arteries through the right (left) heart atrium and right (left) ventricle.

The heart ventricle, as implemented in Diagram 4, has two hydraulic connectors, which represent the area prior to the input valve and the area following the output valve. Through this area, some blood flow moves and some pressure is always generated, as is typical within the hydraulic connector. Flow going to the arteries is called cardiac output. Cardiac output (CO) as a mean blood flow from the heart ventricle is heart rate (HR) multiplied by stroke volume (SV), where stroke volume is the difference between end diastolic volume (EDV) (Carter, et al., 1998; Gaasch, et al., 1975) and end systolic volume (ESV) (NODA, et al., 1993). The most common descriptions are pressure-volume relations (Sagawa, et al., 1988), as in A-V fistula experiments (Guyton and Sagawa, 1961) or filling pressure experiments (SUGA and SAGAWA, 1974), or less invasive exercise experiments (Little and Cheng, 1993).

This model does not solve the situation where there are very short times for the good filling of the ventricle. However, using Physiolibrary, it is possible to create the beat-by-beat implementations as we have described in articles (Kulhánek, et al., 2014; Kulhánek, et al., 2014). These publications show the opening and closing valves (Equation 30) that simulate the current pressure and flow during the diastolic filling and systolic ejection of ventricles. Moreover, blood flow inertia is also integrated (Equation 31), which plays a significant role in the shape of blood flow and pressure (e.g., generating dicrotic notch) during these short-term events.

Diagram 4: Heart ventricle; the block diastole calculates the end diastolic volume from mean filling pressure, while the block systole calculates the end systolic volume from mean arterial pressure and contractility, which is a function of the beta receptors’ activity.

### Circulation

During pulmonary circulation, blood flows through pulmonary arteries, capillaries and veins. All of these are represented in Diagram 5 by the elastic vessel (Equation 26 and Equation 27) and hydraulic resistor (Equation 28). A special block is used for the calculation of perfusion of ventilated alveoli and is based on total blood flow through pulmonary capillaries, called lungBloodFlow.

Diagram 5: Pulmonary circulation.

The local regulation of vasoconstriction and vasodilation in lungs (Archer and Michelakis, 2002) is not implemented, but can easily be inserted in the follow-up versions.

In systemic circulation, blood flow from systemic arteries (Roach and Burton, 1957) is divided into branches for different tissues. The upper part of Diagram 6 shows coronary (micro)circulation through the heart; next is shown all other circulation for peripheral organs except for splanchnic circulation and the splanchnic circulation, where is the blood from gastrointestinal tract mixed with blood from hepatic arteries. The lower part of Diagram 6 represents the sequestered blood in lower parts of the body caused by the hydrostatic gravitation effect (Equation 29). The characteristics of sequestered blood in leg vessels can be measured using a variety of orthostatic experiments (Bevegärd and Lodin, 1962; Bock, et al., 1930; Henry and Gauer, 1950; Mayerson, et al., 1939; OCHSNER, et al., 1951; Pollack and Wood, 1949; Thompson, et al., 1928). Together with the blood pumping effect (using vein valves during contraction and relaxation of surrounding skeletal muscle) it answers the question of why is so uncomfortable log-term staying at one place without motion in contrast with long-term walking (Armstrong, et al., 1985; LAUGHLIN, 1987; Laughlin and Armstrong, 1983).

After flowing through tissues, blood moves into systemic veins, where zero-pressure-volume is driven by venoconstriction. Venoconstriction is driven by sympathetic neural answer as part of the baroreflex (ECHT, et al., 1974; GAUER, et al., 1956; Shigemi, et al., 1994). The final phenomenon in systemic veins in place of entering intrathorax cavity may be the collapsing of veins. This is caused by small negative intrathorax pressure, which can suck all volume from the veins at the diaphragm and restrict blood flow, which occurs in collapsing vessels if there is not enough blood volume.

Diagram 6: Systemic circulation.

Peripheral circulation is composed of eight types of tissues: bone, neural, adipose, skeletal and respiratory muscle, renal, skin tissues and the rest. These organs are implemented by the same class of microcirculation but with different parametrical settings. The exception to general microcirculation is the renal circulation of kidneys (Diagram 10). This type is very specific, because blood flow after renal arcuate artery and afferent arterioles access the glomerular capillaries net. After the glomeruli and efferent arterioles is blood divided again to the capillary net of the vasa recta or interlobular capillary net. The differences in renal circulation are significant, because renal blood flow is typically around 20% of cardiac output.

Diagram 7: Peripheral circulation.

Splanchnic circulation delivers all blood from the gastrointestinal tract to the liver via the portal vein (Bradley, et al., 1953). In the liver hepatic blood flow is determined by the portal vein and hepatic artery blood flow. Normal hepatic blood flow can vary from 1 to 2.5 l/min (BRADLEY, et al., 1952), depending on gastro-intestinal blood flow. Splanchnic circulation can function as a blood reservoir during hemorrhage or during blood infusion (Greenway and Lister, 1974; Maass-Moreno and Rothe, 1992).

Diagram 8: Splanchnic circulation.

### Microcirculation

Blood flow through blood vessels depends on blood viscosity (Whittaker and Winton, 1933), as shown by the upper factor of Diagram 9. Bellow this factor is shown the vasodilation/vasoconstriction effect of anesthesia, then the effect of angiotensin 2, vasopressin and catecholamines. Catecholamines such as epinephrine and norepinephrine that freely dissolve in extracellular fluids are described in the ‘Hormones’ section and their effect on alpha receptors are calculated as the variable AlphaPool\_Effect. The alpha receptors can also be stimulated by sympathetic neural activity (GangliaGeneral\_NA) or inhibited using alpha blockers (AlphaBlocade\_Effect), as will be described in section addressing neural activity. The next factor applies to skeletal muscles, where a metaboreflex dilates the arterioles to bring more oxygen and nutrients into working tissues. The next factor is an adaptation to a long-term low hypoxic condition by angiogenesis, where new branches between arterioles and venules cause lower resistance for blood flow. The partial pressure of oxygen can also have an acute effect on vasodilation (or local vasoconstriction in lungs). However, the brain must also calculate the effect of carbon dioxide (Kety and Schmidt, 1948), which increases blood conductance in a situation where it needs to be washed out or where it decreases blood conductance, where it must be accumulated to eliminate local rapid pH changes. The local metabolic demand for oxygen is also one of the factors of vascular resistance. The final factor is an embolism, where the perceptual part of tissue circulation can be blocked by an embolus in the form of a blood clot, gas bubble or any solid blockage in the blood stream.

Diagram 9: Microcirculation.

An exception in terms of microcirculation is the renal circulation of kidneys, where only the efferent interlobar part is driven by some of the abovementioned factors. The strictly regulated renal blood flow by both afferent and efferent arterioles (Diagram 10) needs to set an optimal filtration pressure (Manning, 1987; Manning, 1990) and prevent the washout of kidney medulla concentrations. This process is driven by a number of working nephrons, tubule-glomerular-feedback (Ito and Carretero, 1990; Moore and Casellas, 1990), baroreflex-like patterns (Skarlatos, et al., 1993), local mechanoreceptor-myogenic patterns (Aukland, 1989; Drummond, et al., 2008) and by efferent interlobar microcirculation (Heyeraas and Aukland, 1987).

Diagram 10: Renal (micro-) circulation of kidneys.

Hydraulic resistance (reciprocal value of conductance) is regulated by a cross-sectional area of vessels. The higher the cross-area the faster the blood stream will be at the same pressure gradient. The radius of this area is a function of circumference, which is determined by the current length of vascular smooth muscle around. The vascular smooth muscle tone is regulated by many factors, as described previously (Mellander and Bjornberg, 1992; Shigemi, et al., 1994). Vasoconstriction causes increasing resistance and pressure, together with decreasing blood flow. Vasodilation has the opposite effect. These types of vascular regulations are specific for different tissues, where any of the factors can be disabled or enabled, or set to different sensibilities for different tissues.

### Blood properties

Blood volume is calculated as plasma volume plus the volume of red cells. Blood plasma volume is calculated by the water subsystem, but the amount of red cells is integrated inside the cardiovascular subsystem alongside the components shown in Diagram 11. Using population components from the Physiolibrary (Equation 32 and Equation 33), increasing the number of erythrocytes by erythropoiesis or transfusion is increased, while erythrocytes are decreased as a result of their natural mortality or by hemorrhage. The rate of erythropoiesis is determined by the concentration of erythropoietin, which is modeled in section on hormones.

Diagram 11: Red cells.

The final additional component of the cardiovascular system is a block with general blood properties such as total blood volume, hematocrit and viscosity or the viscosity conductance effect. Viscosity of blood is strongly dependent on hematocrit (Begg and Hearns, 1966; Schrier, et al., 1970; Stone, et al., 1968); therefore, the higher the number of red cells, the less ability blood has to move. However, if there are more red cells in hemoglobin, then more oxygen can be connected to hemoglobin. As a result, optimal hematocrit for oxygen transport between these two conditions can experimentally be measured as 40-60% in most tissues (Fan, et al., 1980; Jan and Chien, 1977).

Figure 6: Viscosity conductance effect on hematocrit using the measured data of Fan et al. (Fan, et al., 1980).

### Comparison with HumMod v1.6

The cardiovascular system shows many differences when comparing the Physiomodel and HumMod 1.6, but yields very similar results in terms of simulations.

In our Physiomodel, the accumulation of blood volume in systemic veins is implemented with the same component (Equation 27, ElasticVessel) as in other areas of circulation. The total blood volume is calculated as the sum of these compartments. The original HumMod v1.6 calculates the integration of total blood volume instead of systemic venous blood volume, which is calculated as the rest of the blood volume from all other places in the body. These reformulations lead to the same steady state equations; therefore, in a normal situation when blood volume remains constant, the simulation results are the same. Additionally, after stabilization and equilibration to the new blood volume, the same results must be reached. However, our component-based solution has better properties in a dynamic situation, where rapid blood changes can be applied to the specific parts of circulation with local dynamic responses. Cross-checking in the Physiomodel is done using the conservation law of blood volume alongside known changes of total blood volume. Change in the sum of blood volume of all circulation components must be the same as external changes from/to circulation. These tests can also uncover non-correctly-defined changes in the blood volume of heart ventricles and their connection with end systolic pressures, which are also corrected in the Physiomodel, in contrast with HumMod 1.6.

As each circulation component must be connected in the circuit diagrams, all blood flows are correctly defined by these diagrams in the Physiomodel. The original textual representation of the HumMod has no form of connection checking and as such, it is easy to forget to connect, for example, blood flow from splanchnic circulation to the systemic veins. In Modelica diagrams, it can be observed when some physical connectors are not connected. Thus, the user sees blood outflow from components that are unconnected. Even when noting flows from/to the environment, the user should use the specific component such as flow pumps or fixed pressure source.

The graphical diagrams illustrate the connections. Finding all the equations for blood flows in the original HumMod and applying an understanding of these equations must lead to very similar visual representations. If these representations have mathematical meaning, as in Modelica, understanding and upgrading of the model will be easy. In this diagram, for example, it is immediately evident that the coronary circulation in the HumMod accesses the systemic veins, as in other peripheral blood flows. The Physiomodel changes this to a more anatomically precise idea, i.e., that coronary circulation ends directly into the right atrium. As a result, the resistance parameter of coronary vessels was recalculated to a new pressure gradient (between the aorta and right atrium) in order to reach the same coronary blood flow.

There remains small cardiovascular system disproportions in both models. For example, the changing of pulmonary blood flows through ventilated alveoli is not connected to the circulation circuit. The total pulmonary circulation in these versions is independent of this process, which in reality must be interconnected. Additionally, renal blood flow through the vasa recta is not correctly connected within the model’s cardiovascular system. All of these parts can be upgraded in future versions of the models.

## Body Water

The model for water (Diagram 12), for example, the model for extracellular proteins, is divided into eight main compartments: blood plasma (plasma), red blood cells (RBC) and interstitial (IST)/intracellular (ICF) water of the upper torso (UT), middle torso (MT) and lower torso (LT). These compartments are connected by chemical connectors, which also support the osmotic processes. Selected distribution of body water (41L for 70kg, male) between compartments is shown in Table IV. From these values, the total interstitial, extracellular or intracellular volume used for simplified pharmacokinetic calculations can also be expressed.

Table V: Typical steady state water volume of compartments [L].

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Plasma | RBC | UT\_IST | UT\_ICF | MT\_IST | MT\_ICF | LT\_IST | LT\_ICF |
| 3.0 | 1.6 | 2.3 | 5.0 | 5.7 | 12.5 | 3.4 | 7.5 |

Diagram 12: Water subsystem.

Selected mean water flow between all compartments are listed in Table V, as examined in a range of studies (Eisenhoffer, et al., 1994; Engeset, et al., 1973; Henriksen, 1985; Xie, et al., 1995). The steady state of Table V reverts the sum of each row and each column to zero. Table rows represent flow description while columns refer to different areas. The areas into which water flows are indicated according to whether the value is positive, or whether the places from which the water flows has a negative value. For example, in the first line, the water is absorbed from diet in the gastrointestinal tract; the food comes from the environment (ENV) and goes into the blood plasma (Plasma). In each torso it is metabolically produced and excreted through sweating or by vaporization. Flows such as hemorrhage, transfusion, intravenous drip, flow to the peritoneum, to the lungs and edema are zero in normal conditions. Excretion to urine is modeled by the kidney component.

Table VI: Selected steady state water flows between compartments [ml/min].

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Plasma | UT | MT | LT | ENV |
| From diet | 1.4 |  |  |  | -1.4 |
| Across capillaries | -3.01 | 0.38 | 1.23 | 1.40 |  |
| Lymph | 2.41 | -0.32 | -0.75 | -1.34 |  |
| From metabolism |  | 0.06 | 0.11 | 0.06 | -0.23 |
| Evaporation |  | -0.12 | -0.59 | -0.12 | 0.83 |
| Urination | -0.8 |  |  |  | 0.8 |

### Extracellular proteins

Water distribution between cardiovascular and interstitial spaces occurs concurrently with colloid osmotic pressures, which leads to the calculation of extracellular proteins of the same compartments, as described in the previous section. Usually, proteins are calculated in units of mass; however, our implementation calculates their amount in substance, because mole fractions play a role in osmotic equilibrium (Equation 8). The molar mass of albumin is 66.5 kDa and the mass of albumins is about 60% of the total plasmatic protein mass. The rest of significant colloid proteins are globulins. The typical molar amount of plasmatic proteins is as presented in Table VI. The general approach for recalculating mass-molar units is by joining an osmotic pressure equation as a mass function (Ahlqvist, 2003; Manning, 1987) with Equation 8, where the molar volume of water is around 0.018 L/kg.

Table VII, Typical plasma proteins concentrations

|  |  |  |
| --- | --- | --- |
| Total | Albumin | Globulins |
| 1.44 mmol/L | 0.63 mmol/L | 0.81 mmol/L |
| 28 µmol/mol | 12 µmol/mol | 16 µmol/mol |

As already noted, the model of proteins (Diagram 13) has four main compartments: blood plasma, upper torso interstitial space, middle torso interstitial space and lower torso interstitial space. Normal concentrations in interstitial compartments are listed in Table VII. Normal mean proteins synthesis is the same as it is for protein degradation. Their current values can be changed alongside deviation of their plasmatic concentrations. Movement between compartments is caused by capillary membrane concentration gradient or by lymph flow (Mayerson, et al., 1960) from interstitial spaces to blood, as implemented in the scheme of Diagram 13. Special changes in plasmatic concentrations can be effected by intravenous therapy, hemorrhage or pathological states. Pathological states include proteins entering the peritoneal space or breaking the glomerular membrane as filtration to primary urine.

Table VIII: Typical protein concentrations in interstitium.

|  |  |  |
| --- | --- | --- |
| Upper torso | Middle torso | Lower torso |
| 0.6 mmol/L | 0.48 mmol/L | 0.4 mmol/L |
| 28 µmol/mol | 28 µmol/mol | 28 µmol/mol |

Diagram 13: Subsystem of extracellular proteins.

### Gastrointestinal water absorption

As presented in Table V and Table IV, the mean water in diet should be about 2 l/day, which is the sum of all water in food and drinks. Firstly, water is accumulated in the gastrointestinal lumen (GILumen). From the lumen, water can cross the cellular membrane of gastrointestinal cells (OsmBody\_CellWall) using aquaporins to equilibrate its extracellular and intracellular mole fractions (Diagram 14).

Diagram 14: Water absorption in the gastrointestinal tract.

The absorption of water from the gastrointestinal lumen into the intestinal cells is here driven only by osmotic forces. The typical mean intake of 2 L/day is caused by a mean osmotic pressure gradient of 25 kPa at a temperature of 37°C. From these assumptions, the permeability parameter can easily be expressed as 0.08 L/(kPa.day).

### Upper/middle/lower torso water

Flow between plasma and the interstitium is determined by colloid osmolarity on the capillary walls. Another way is the one directional lymph flow from the interstitium to blood plasma (Eisenhoffer, et al., 1994; Engeset, et al., 1973; Henriksen, 1985), as presented in Table V. These flows can be influenced by the internal pressure in tissues caused by their volume and skin, as examined by Gyuton (Guyton, 1965) and Xie et al. (Xie, et al., 1995). Water crossing the capillary wall is driven by hydrostatic-[oncotic](https://en.wikipedia.org/wiki/Oncotic_pressure) pressure gradients, as expressed by Equation 8.

Table IX: Amount of substances per liter of fluids.

|  |  |  |  |
| --- | --- | --- | --- |
| Substance | UT | MT | LT |
| IST - Total | 21.96 mol | 17.57 mol | 14.64 mol |
| IST - Water | 21.85 mol | 17.48 mol | 14.57 mol |
| IST - Electrolytes | 0.10 mol | 0.08 mol | 0.06 mol |
| IST - Urea | 2.3 mmol | 1.9 mmol | 1.5 mmol |
| IST - Glucose | 2.3 mmol | 1.9 mmol | 1.5 mmol |
| ICF - Total | 39.05 mol | 39.05 mol | 39.05 mol |
| ICF - Water | 38.86 mol | 38.86 mol | 38.86 mol |
| ICF - Electrolytes | 0.15 mol | 0.15 mol | 0.15 mol |
| ICF - Urea | 4.1 mmol | 4.1 mmol | 4.1 mmol |

However, the flow of water between the interstitium and cells is determined by its model’s fraction equilibrium. In cellular membrane, proteins’ osmolarity plays a minor role, because their concentration is only about 1 mol/L. In the extracellular space the total amount of substances is divided into water, electrolytes, urea, glucose and other solutes. In the intracellular space, these substances are water, electrolytes, urea and other solutes. The small total amount of solution in interstitial fluid is caused by large molecules, which take up the highest solution volume (60-74%) with their small amount (less than 10 mmol/L).

Diagram 15: Water exchanges for upper, middle or lower torso.

### Kidney water excretion

In the kidneys, water is delivered by blood to the glomerulus, where blood plasma is filtrated into glomerular filtrate (GFR). Most of this filtrate is reabsorbed in nephrotic parts: the proximal tubule (PT), the loop of Henle (LH), the distal tubule (DT) and the collecting ducts (CD); the rest is accumulated in the bladder as urine.

Table X: Typical average steady state flows through nephrons [ml/min.]

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| GFR | to LH | to DT | to CD | to Bladder |
| 120 ml/min | 57 ml/min | 41 ml/min | 4.6 ml/min | 0.8 ml/min |

Diagram 16: Water excretion by kidney nephrons.

Proximal tubule

Glomerular filtrate in the glomerulus has the same pressure as blood in the glomerulus and this pressure pushes it into nephrons. The reabsorption fraction in the proximal tubule is determined only by sodium reabsorption of the proximal tubule.

Loop of Henle

Only the short coronary nephrons contain the aquaporin channels inside the loop of Henle, which here renders the water reabsorption fraction only 37% of the sodium reabsorption fraction (Gottschalk and Mylle, 1959; Nielsen, et al., 2000).

Distal tubule

Outflow of filtrate to the collecting duct is determined by the outflow of sodium, where it is dependent on ADH nephron concentration, as described in Khokhar et al. (1976) and Atherton et al. (Atherton, et al., 1971; Khokhar, et al., 1976).

Collecting duct

In the collecting duct, the number of active aquaporin channels are driven by ADH; proportionally, this indicates the volumetric flow rate of reabsorbed water by the collecting duct tubules (Jamison and Lacy, 1972; Jamison, et al., 1971). Changing the activity of aquaporin channels is modeled by the integration of inactive channels driven by ADH concentration as simulating the process of their intracellular vesicular storage. There is minimal water outflow to urine, which is determined by sodium outflow to urine and medulla osmolarity.

### Hydrostatic spillover

Orthostatic position also plays a role in water transport. The hydrostatic pressure component can be calculated using Equation 29. Together with hydrodynamic blood pressure and osmotic pressure components, it forms the pressure gradient on the capillary walls of tissues. The values of blood hydrostatic pressure and interstitium hydrostatic pressure are different in a stand-up position. They are caused by the different heights of compact liquid columns, which generate the pressure. Blood vessels are mostly compact and open tubes, where the highest area determines the hydrostatic pressure of the areas below. These hydrostatic pressure components can be calculated only from the height difference, while the interstitial space can be more hydrostatically independent, meaning that the weight of the tissue water can be hanged using a system of cavity membranes, which generates smaller heights of hydrostatic columns. In the lower torso, veins are enabled as a result of motions made by the leg’s skeletal muscle, which has a pumping effect similar to that of the heart. The reason for this is the availability of vein valves, between which accumulates a volume of blood from lower parts during skeletal muscle relaxation, which is ejected to upper parts of systemic veins during skeletal muscle contraction. This pumping effect not only reduces hydrostatic pressure, but also actively increases the blood flow of systemic veins during walking, running or cycling.

### Relational comparison with HumMod 1.6.1

The primary differences between the HumMod and Physiomodel in terms of water calculation is that the latter is more physically-based, but yielding almost the same results as the former.

This means that, for example, osmosis is calculated by the equilibrium of chemical potentials, as indicated in physical chemistry textbooks (Mortimer, 2008). The impermeable proteins of plasma and interstitial fluid are in the Physiomodel recalculated to molar amounts. These molar amounts were selected by mass amounts of the HumMod to reach the same normal values of osmotic pressures on capillary membrane. Oncotic pressure is calculated in both models from these proteins, which cannot freely cross the capillary membrane. Their molar concentrations in Table VI, typical plasma protein concentration in Table VI and in Table VII are consistent with both physical osmotic pressure calculation and with their original mass concentration (Ahlqvist, 2003), where the ratio between mass and molar concentration is the average molar mass of these proteins.

In textual representation, it is easy to misuse one variable in the place of another. As a result, equations can be entirely wrong, which can immediately be observed in simulation results. Worse, if the both variables have similar values, all results will appear in order up to the point where a specific setting is applied. This is the case for regional capillary blood pressure in HumMod 1.6. If the patient is lying down on a bed, all regional pressures will be the same. If the patient stands up, hydrostatic pressures will change according to different heights differences from heart (Equation 29). As a result, there will be higher regional capillary pressure in the lower torso and lower regional capillary pressure in the upper torso. The Physiomodel calculates results by taking these factors into account; however, in the HumMod 1.6, these tissue capillary pressure gradients are always calculated only from the upper torso capillaries’ pressure, or from the middle or lower torso.

The Physiomodel also makes improvements in the relationship regarding losing water by vaporization in respiratory pathways. In the HumMod 1.6, it is assumed that all water in expired air arises from vaporization. However, there can be significant humidity to the inspired air. The Physiomodel inserts this humidity to slow down the expiration of water in hot and humid environments.

The final differences between the Physiomodel and the HumMod 1.6 concerns the behavior of tissue water in an initial state and during steady state. For example, lymph flow in the upper torso was too slow to deliver the same amount of water from the upper torso interstitial space as the amount of water crossing the capillary membranes at the same time and in the same place. Thus, the interstitial water of the upper torso slowly increases from the start of the simulation to some pathophysiological values. To prevent this instability in the model, we recalculated the permeability coefficient of the upper capillaries in order to meet the steady state of these flows. As a result, there are initial steady states of interstitial water in the Physiomodel.

## Hormones

### Anti-Diuretic Hormone (ADH, Vasopressin)

Arginine vasopressin, known as an antidiuretic hormone (ADH), has a molecular weight of 1084 Dalton. ADH as a hypothalamic neurohormone is synthesized in the cell bodies of magnocellular neurons of paraventricular and supraoptic nucleui and is intracellularly transported to the lower side of these neurons in the posterior pituitary.

Diagram 17: Vasopressin.

The model (Diagram 17) accumulates an amount of this hormone in four areas: in the cell bodies of magnocellular neurons (Slow Mass), from where it needs to be transported to the posterior pituitary part of the cells; in the posterior pituitary side of neurons (Fast Mass), where ADH is prepared for secretion into the blood; in the entire body’s extracellular fluid (ECF); in the kidney tissue, where it plays a role in water reabsorption. The normal long-time amounts of ADH in these compartments are listed in Table 10; however, during regulation, their concentration can be increased a hundred- or thousand-fold (Lankford, et al., 1991). The normal long-term mean rate of synthesis, secretion and degradation should be the same at steady state. However, secretion in the form of a short-time process can achieve much larger changes. The effect of various changes and concentrations has been demonstrated in dosage experiments (Atherton, et al., 1971). Internal secretion is determined by osmoreceptors and pituitary activity. Osmoreceptors are the cells in the anterior hypothalamus near the supraoptic nuclei. When the osmolarity increases, the osmoreceptors shrink and send a neural signal to release ADH (Young, et al., 1977). Another possibility for regulating ADH secretion is through cardiovascular centrum reflexes (Erwald and Wiechel, 1978).

Table 11: Selected long-term steady state amounts of vasopressin.

|  |  |  |  |
| --- | --- | --- | --- |
| Slow Mass | Fast Mass | ECF | Kidney Medulla |
| 15.7 nmol | 2.95 nmol | 0.028 nmol | 0.000 057 nmol |

The vasopressin within cells is modeled using instances of the chemical Substance class; the intracellular vesicular mole fraction is 1, because ADH is transported as a pure substance by vesicles down the cell. The degradation is divided into liver, kidney and other tissue blood clearance. To reach the mean constant level of ADH, the sum of all long-term mean losses must be the same as the long-term mean synthesis and secretion. The loss of ADH in these organs as an enzymatic degradation in liver, kidney and other tissues is dependent on blood flow.

A typical concentration in blood plasma and extracellular fluid is in the order of ng/l, pg/ml, pmol/l or mIU/l. An increase in these concentrations causes water reabsorption in the kidneys (Lankford, et al., 1991).

### Atriopeptin

The secretion of atrium natriuretic peptide (ANP) is driven by mean blood pressure in both atria. These pressures are relative to pericardium pressure and noted with the suffix ‘\_TMP’ in Diagram 18. There is an adaptation of secretion to current pressures with half time about fifteen minutes makes from ANP the middle-term regulator of blood pressure and blood volume (Conte, et al., 1992; METZLER, et al., 1986; Mizelle, et al., 1990; Nicholls and Richards, 1987; Renkin and Tucker, 1996; Weidmann, et al., 1986; Yandle, et al., 1986).

Diagram 18: Atrium natriuretic peptide.

### Catecholamines (epinephrine; norepinephrine)

The model of catecholamine accumulation, secretion and clearance is very simple. Driven by sympathetic neural activity, \*\*\*\* is secreted in the adrenal gland. It is then accumulated in extracellular space and continuously degraded by clearance, which at a long-term steady state causes the same mean degradation as the mean secretion for the long-term average of concentration. This model has also been observed by an experiment of 60-minutes’ continuous intravenous epinephrine infusion, where different nominal rates caused different steady state plasma epinephrine concentrations (Clutter, et al., 1980).

Diagram 19: A catecholamines model composed with model of epinephrine and nonepineprhine.

The effect of catecholamine in alpha- or beta-receptors on the effector organs is expressed as a decimal logarithm of the concentration. This effect is combined with sympathetic neural activity on the receptors and can be blocked by alpha- or beta-blockers.

### Erythropoietin (EPO)

Erythropoietin (EPO) secretion is driven by partial oxygen pressure in the kidneys (BAUER, 1993) (Goldberg and Schneider, 1994; Jacobson, et al., 1957; Pagel, et al., 1988; Porter and Goldberg, 1993). In contrast with previous hormones, the distribution space of EPO is not all extracellular fluid, but only about 40% of this fluid (Miller, et al., 1982; Reissmann, et al., 1965). The mean degradation must be the same as the mean secretion during typical mean concentration in a steady state.

The role of erythropoietin is connected with erythropoiesis in the bone marrow (Jacobson, et al., 1957; Roush, 1995; Winearls, et al., 1986).

### Insulin and glucagon

Insulin is one of the most widely studied hormones. His molar mass is 5.808 kDa. The first standard international unit of insulin was noted in 1958 (Standardization and Organization, 1958); the most recently discontinued definition from 1986 has been improved to 38.46 µg/IU (Standardization and Organization, 1987). Using this definition, it is possible to estimate a conversion such as 6.621pmol/IU.

The insulin pharmacokinetic and pharmacodynamics obeys the same principles as the model of glucose-insulin homeostasis put forward by Guyton et al. (Guyton, et al., 1978). Insulin is synthetized and stored in beta-cells and its secretion is driven primarily by glucose and secondarily by keto acids (Imai, et al., 2008; Rutter and Hill, 2006). Portal and peripheral vein insulin has different concentrations (Blackard and Nelson, 1970), because insulin is transported directly after its secretion by portal veins to the liver. Absorbance and clearance have been measured using a range of infusion experiments (Dobson, et al., 1967; DOEDEN and RIZZA, 1987; GINSBERG, et al., 1973).

Problems with insufficient insulin secretion results in type 1 diabetes mellitus and receptor insensibility leads to type 2 diabetes mellitus (George, et al., 2004; Prager, et al., 1987; Summers, et al., 1997), where many differences between normal and obese individuals have been observed (Prager, et al., 1986). Insulin has significant effects on glucose absorption by cells of liver (Iwanishi, et al., 2000; Previs, et al., 2000; Rother, et al., 1998), where glucose is stored and released to/from glycogen (glycogenesis, glycogenolysis), which is created from amino-acids (gluconeogenesis) or transformed to fats (lipogenesis) (Guyton, et al., 1978; Miles, et al., 1995; Prager, et al., 1986). The effect of glucose absorption and storage (similar to that of glycogen) is modeled in skeletal muscle tissue. Insulin also helps in the storage of fatty acids in adipose tissue, as modeled in the lipid submodel of metabolism fraction.

In contrast to the storage effect of insulin, glucagon helps to increase the glucose and fatty acid concentration in the extracellular space. However, the secretion of glucagon is dependent on the insulin concentration (and the glucose concentration), rendering it the secondary regulator of blood glucose concentration.

### Leptin

Leptin is secreted by adipose tissue as a signal from accumulated lipids (JÉQuier, 2002). The notion of curing obesity using leptin is not valid, due to leptin resistance in instances of obesity (Myers Jr, et al., 2010) (Friedman-Einat, et al., 2003). The clearance of leptin is primarily effected by the kidneys (Cumin, et al., 1996). Leptin has multiple effects on higher metabolic centers (Mantzoros, et al., 2011; Wong, et al., 2004), which is modeled primarily according to the influence of diet composition and the amount of food eaten as a result of changes in taste caused by leptin concentration.

### Renin-angiotensin-aldosterone system

The secretion of renin in the kidneys is driven by tubuloglomerular feedback (TGF) (Braam, et al., 1993; Seeliger, et al., 1999), as well as adrenergic receptors (Almgård and Ljungqvist, 1975; WINER, et al., 1969). Clearance is primarily conducted by the liver (Christlieb, et al., 1968). Renin is an enzyme that converts angiotensinogen into angiotensin I. This conversion obeys Michaelis-Menton dynamics, which creates linear dependence between the amount of renin and the rate of conversion (Goldblatt, et al., 1953). The same dynamic is observed in the lungs regarding angiotensin converting enzyme (ACE), where angiotensin I is transformed into angiotensin II. In optimal regulation conditions, it gives linear dependence between renin concentration and angiotensin II concentration (Claassen, et al., 2013).

### Thyroid hormones

The main purpose of thyroid hormones in our model is to maintain the basal metabolism in connection with long-term thermoregulation (Edelman, 1974). The concentrations, secretions and clearance of thyroid hormones are well known, because of the relative easy measurement of iodine radioactive isotopes (HAYS, 1993; Chopra, 1976; Larsen, 1972; Nicoloff, et al., 1972). During cold months, increasing triiodothyronine (T3) (Hesslink, et al., 1992) increases the basal metabolism (Osiba, 1957), which improves heat regulation in cold conditions. The impulse for the production and secretion of T3 and thyroxine (T4) is thyrotropin (TSH) (Jackson, 1982). The secretion of TSH is driven by thermoreceptors and is directly suppressed by T3 (Gross and Pitt-Rivers, 1953; Hesslink, et al., 1992; SURKS and LIFSCHITZ, 1977; SURKS and OPPENHEIMER, 1976). The clearance of TSH happens much quicker than the clearance of T3 or T4 (Ridgway, et al., 1974); as a result, its concentration can be directly estimated from the secretion, which is determined by current thyroxines concentrations and temperature.

### Comparison with HumMod 1.6

There are also small corrections of hormonal equations in the Physiomodel; however, the mean levels of hormones are almost the same as in HumMod 1.6.; this solves, for example, the mish-mash of physical units or the potential confusion of variables.

The primary aim of modeling all organic chemical substances was to calculate the amount of their molecules. This was the impetus for using molar units. The amount of substance in moles multiplied by the Avogadro constant yields the relevant number of particles. However, currently, even the Avogadro constant is approximated and as a decimal number, only the first eight of its 23 digits are exactly determined. This means that the precision of counting the particles of a solution is limited and we do not know exactly how to measure and work with very small molar concentrations (such as piko- or fentomoles per liter). These very small concentrations are biologically significant for some hormones, enzymes or cytokines. Instead of physical units, some pharmacokinetic units must therefore be used (u or iu – international unit defined by [WHO](http://www.who.int/biologicals/reference_preparations/en/)); these are defined by the solution extract as a result of the described purification process. In the Physiomodel, this situation is solved using a redefinition of the Physiolibrary, where all moles are switched to international units of the specific hormone, enzyme or cytokine. These redefined library components must be used for each block calculation with the relevant substance. For example, if we want to create the chemical reaction of this ‘unmeasurable’ hormone with its receptor, the concentration and amount of receptors must also be in the same international units as the hormone. Hormones, enzymes or cytokines with already known conversions should be used in their physical amount of substance rather than in international units.

It is not surprising that physical units can be a source of many mistakes in the models, especially in the case of hormones, where unusual prefixes such as micro-, nano-, piko- and fento are used together with more alternatives for expressing concentrations such as molar concentration, mass concentration, molar fraction, mass fraction and international units. For example, in HumMod 1.6, this combination of units with thyroid hormones, where a concentration is 100 x higher due to switching ‘ug/ml’ for ‘ug/dl’ when compared to data presented by (CHOPRA, et al., 1975). When using the Physiolibrary in some user-friendly Modelica environments, there is automatic support for recalculation of SI units into non-SI units, including those with different prefixes. Using this automated physical unit support for inputs and outputs of simulation the modeling process becomes more error-free in the area of unit conversions.

The other example of changes in variable names in HumMod 1.6 is in the case of intracellular renin fluctuation calculations. Instead of calculating the flux from free renin synthesis to renin granules, the same variable is used for the intake and outtake of the renin granules. Thus, the renin granules have the same concentration every time for each simulation experiment, even if rapid secretion occurs. Using the diagram modeling of the Physiolibrary it becomes very clear *from* which or *to* which compartment the renin flows move. The likelihood of mistakes occurring in this area therefore decreases, because the connections are visually self-descriptive.

Changes in physical units can hide more contextual errors. For example, a change in vasopressin as a result of blood clearance in circulating blood through tissue must be less or equal than the amount of vasopressin inflowing as a result of blood moving to this tissue. Furthermore, because there are multiple coefficients hidden in the flow, in the recalculation from milliliters to liters, values higher than 1 is hidden by dividing of thousand. This confusion is also visible in other instances of tissue clearance of vasopressin in HumMod 1.6; in the Physiomodel, it is resolved by selecting full clearance in this type of tissue, which generates only a small amount of vasopressin degradation.

## Electrolytes and Acid-Base

### Acid-base

The acid-base balance calculation is based on electroneutrality. In the case of plasma, in extracellular and in intracellular fluid, the charges of strong ions that do not significantly change their charge at a pH from 5 to 9 are summarized. This is called strong ion difference (**SID**) (Stewart, 1981). As an analogy of SID, the variable anion gap (**AG**) can be used, which is the same as SID with a charge of bicarbonates (AG = SID – [HCO3-]), where the amounts and properties of other non-bicarbonates acid-base puffers are also not included. The acid-base buffers (HCO3-and other weak ions) are calculated as the **negative** summary charge concentration at normal conditions (prefix N); this is called normal strong ion difference (**NSID**). The **normal conditions** are defined as **plasma pH=7.4, full oxygen saturation, CO2 partial pressure 40mmHg and temperature 37°C**.

NSID describes the potential of acid-base buffers. In normal conditions it has the same value as SID. In a situation where there is a higher value of NSID than SID (for example, there is an excess of strong acids) the arterial pH<7.4 during a normal state of respiration at 37°C. If NSID<SID, then pH>7.4 (for example, an excess of strong bases) during a normal state of respiration at 37°C. Both SID and NSID can be calculated in plasma (suffix P) and inside erythrocytes (suffix E). The titration of one liter of blood in order to reach normal conditions must use the same amount of strong acid as the differences between SID and NSID in plasma and in erythrocytes (Mateják, 2013); this is expressed as **BEox** = Hct\*(SIDE-NSIDE)+(1-Hct)\*(SIDP-NSIDP), where Hct is the hematocrit and BEox is the base excess of oxygenated blood, as defined by (Kofránek, 2009; Kofranek, et al., 2007). This measurable amount of titrant can also be expressed as a negative value, referred to as titratable hydrogen ions of oxygenated blood (**cTHox** = ‑BEox), as applied by Siggaard using the Van-Slyke equation (Siggaard-Andersen, 2005). The BEox and cTHox are independent of blood gases (CO2, HCO3-, O2), which renders them perfect candidates for describing the metabolic parts of acid-base disorders. Respiratory problems or additional regulation of acid-base disorders should be immediately observable from the arterial blood partial pressure of CO2, which should normally be regulated by respiration to 40mmHg.

Diagram 20: Acid-base subsystem.

The acid-base equilibrium is connected to all charged substances by electroneutrality. The charges of substances are calculated in physical units called equivalent (**eq**) or milliequivalent (1 meq = 0.001 eq). A positive value means a positive charge, while a negative value indicates a negative charge. One **equivalent** is the charge of one mol of protons, which is the same as one mol of sodium cations Na+ or the same as half a mole of calcium cations Ca2+. The typical SIDP is composed of Na+, K+, Cl-, SO42-, Lactate‑ and the typical NSIDP is calculated as a negative sum of normal bicarbonate HCO3-, albumin, phosphates and globulins charges at a hypothetical pH=7.4 and a temperature of 37°C. In erythrocytes, the SIDE is typically the sum of the charges K+, Cl-, Na+, Mg2+ and SO42-. The NSIDE is the negative sum of the charges HCO3-, hemoglobin, phosphates such as 2,3-DPG, ATP, ADP and GSH at hypothetical plasma pH=7.4, full oxygen saturation of hemoglobin and a temperature of 37°C. Other electrolytes and buffers are neglected, because of their small concentration and/or small charge. This calculation in the Physiomodel is similar as the calculation of Raftos et al. (Raftos, et al., 1990), Wolf et al. (Wolf, 2013; Wolf and DeLand, 2011).

Diagram 21: Acid buffers (normal strong ion difference).

The calculation of charge of the weak ions (weak acids) is dependent on pH, because they are equilibrated each time such as chemical reactions in Table 11. The first schematic reaction is called the Henderson-Hasselbalch equation and is generally used to calculate the carbonic acid dissociation to bicarbonate, which is often connected with CO2 dissolution in water (Equation 13) and CO2 hydration to H2CO3 and accelerated by carbonic anhydrase inside red cells. The acid-base equilibrium can be calculated as in Equation 11, where the dissociation constant K can be defined using a negative decimal logarithm such as pK = ‑log10(K).

Table 12: Scheme of acid-base reactions.

|  |  |  |
| --- | --- | --- |
| Group of acid | Type of reaction | Example of acids |
| Monoprotic | HA ↔ A- + H+ | HCl, -COOH, some protein side chains |
| Diprotic | H2A ↔ HA- + H+ ↔ A2- + 2H+ | H2SO4, H2CO3 |
| Polyprotic | HnA ↔ Hn-1A- + H+ ↔ … | H3PO4 |
| Brønsted | AH+ ↔ AH + H+ | NH4+, -NH3+, some protein side chains |

Table 13: Dissociation constants (pK) of selected acid-base reactions.

|  |  |  |
| --- | --- | --- |
| Chemical reaction | pK | Temperature of pK |
| CO2(aq) + H2O ↔ H+ + HCO3- | 6.103 | 37°C |
| HCO3- ↔ H+ + CO32- | 10.329 | 25°C |
| AcAc ↔ H+ + AcAc- | 3.6 | 37°C |
| β-Hb ↔ H+ + β-Hb- | 4.7 | 37°C |
| HSO4- ↔ H+ + SO42- | 1.99 | 25°C |
| H3PO4 ↔ H+ + H2PO4- | 1.91 | 37°C |
| H2PO4- ↔ H+ + HPO42- | 6.66 | 37°C |
| HPO42- ↔ H+ + PO43- | 11.78 | 37°C |
| NH­4+ ↔ H+ + NH­3 | 9.25 | 25°C |

### Kidney acid-base regulation

In the kidneys, pH is regulated by the excretion of titratable hydrogen ions H+ and ammonium ions NH4+. In contrast with the H+ of weak acids, the protons connected to NH4+ remains more bounded than separated, as H+ and NH3 pH is lower than 9.2. This is the general situation, because urine pH can vary between 4.6 and 8. To connect the flowing acidity of urine (pHu) with the flow of all charged substances, electroneutrality is used. The total molar flow of each substance is described in the following subsections, but the charge of substances in urine does not always remain the same as in the extracellular fluid. This is caused by different pH levels. For example, in more acidic urine conditions (more H+, lower pH), the H+ joins organic acids and phosphates (H2PO4-), and during more basic urine conditions (less H+, higher pH), the H+ separates from phosphates (HPO42-); some H+ can even be separated from NH4+ and HCO3-. Electrolytes such as HPO42-, PO43-­, CO32- and C2O42- can react with calcium Ca2+ to create solid salts crystals known as kidney stones. The charge of each substance is calculated using its scheme of chemical reaction (Table 11) in equilibrium with Equation 11, using dissociation constants from Table 12.

### Sodium

The sodium (Na+) concentration is modeled in extracellular space to reach a typical value from 140 to 150 mmol/L. Sodium intake results from diet in the gastrointestinal tract, while its outtake to urine is regulated by the kidneys; outtake by sweating is expressed via the sweat glands (Diagram 22). Other mechanisms that can change sodium mass (often together with changes in fluid volumes) are modeled as dialysis, intravenous drip, transfusion or hemorrhage.

Diagram 22: Sodium in extracellular fluid.

Diagram 23: Kidney excretion of sodium.

In the kidneys, sodium cations are filtered by the glomerulus into the primary urine of the nephrons. In each part of the nephrons, the sodium is actively reabsorbed into the kidney medulla (Diagram 23) and together with sodium, is reabsorbed also the water expect the collecting duct and the loop of Henle of juxtamedullary nephrons, where are missing the aquaporins. After glomerular filtration, the sodium is reabsorbed in the proximal tubule, the loop of Henle, the distal tubule and finally in the collecting duct. Reabsorption is driven by aldosterone, atrial natriuretic peptides and angiotensin 2. Reabsorbed sodium is accumulated inside the kidney medulla, where it is the secondary determinant of osmolarity, the first being urea. From the kidney medulla, sodium is washed out by vasa recta blood flow, where the equilibrium between tubular reabsorption and vasa recta outflow sets a high intramedullary sodium concentration.

Diagram 24: Sodium excretion by sweat glands.

The backward reabsorption of sodium from excreted sweat is driven by aldosterone. When the amount of excreted water by sweat glands is high, all sodium from sweat is not reabsorbed and it remains as salt on the surface of the skin.

### Potassium

The highest concentrations of potassium (K+) are stored inside cells; therefore, the potassium model must be composed of at least with two compartments – intracellular and extracellular (Diagram 25). Potassium intake results mainly from the gastrointestinal tract while its main outtake occurs through kidney nephrons to urine. Potassium flow through cellular membrane is regulated by Nernst potential, by aldosterone and by glucose intake to the cells. Additionally, kidney excretion and sweating of levels of potassium is affected by a number of channels, the expression of which is affected by aldosterone.

Diagram 25: Potassium in intracellular and extracellular fluids.

Diagram 26: Cellular membrane potassium transport.

Diagram 27: Kidney potassium excretion.

### Phosphates and Sulfates

Sulfates (SO42-) and phosphates (H­PO42-, H­2PO4-) are accumulated in extracellular fluid. Their intake occurs from diet and unregulated outtake to urine undergoes Donnan’s equilibrium at the glomerular membrane (Equation 9).

### Comparison with HumMod 1.6

Acid-bases in the Physiomodel is entirely different that in HumMod 1.6. The new calculation of the former uses the original electrolyte models of strong ions. However, the pH is calculated from the electroneutrality equation applied to significantly weak ions such as phosphates, proteins and carbonic acid. The charge of these weak ions is dependent on pH, known as titration curve; thus, the idea is to find the value of a pH that generates sufficient charges for the weak ions at electroneutrality with the rest of solution. The state variables are the total amounts of substances, which in the case of weak ions includes all their protonated and deprotonated forms. For example, inorganic extracellular phosphates refer to H3PO4, H2PO4-, HPO42- and PO43- forms. The total amount is independent of pH; however, the total charge of phosphates is pH dependent. The same calculation can also be applied for the total amount of albumin, globulins, hemoglobin or carbon dioxide. For example, if we assume some constant total concentrations of all weak acids in blood plasma, then we can propose a titration curve of the plasma as a function of SID on pH. However, this function is dependent on the type and total concentrations of weak acids; thus, all charged solutes in a solution must be included in the calculation. This renders our acid-base model in the Physiomodel better than the acid-base calculation of HumMod 1.6, because there is no calculation of weak ion charges, except for bicarbonate. Moreover, the inorganic phosphate is calculated as a strong ion, which does not change form alongside a change of pH.

The same style of acid-base model is also applied to urine in the Physiomodel. The acidity of urine was not calculated in HumMod 1.6; however, all the necessary data for acidity of the primary urine was calculated from the electroneutrality of all the outflowing substances. In future, this calculation can be used, for example, for the modeling of kidney stones or for establishing more precise functions of membrane channels in nephrons.

During simulation of metabolic acidosis such as ketoacidosis (Mateják, 2013), we found that the original pH regulation was so strong that even apparently non-existing chloride was excreted from the body. This confusing notion leads to negative chloride concentrations. To prevent this confusion, we added the stopper function to the Physiomodel, which stops chloride excretion if the extracellular concentration falls below 50 mmol/L. We do not have the exact data for this pathological function, but the insufficiency of chloride during acidosis is a known phenomenon (Levitin, et al., 1958). In conscious patients, is this situation can be solved with the salty taste.

## Blood Gases

To support the metabolism of every cell, oxygen (O2) must be delivered to it and carbon dioxide (CO2) must be transported out of the body. Both of these blood gases are critical to life. Blood gas transport begins through lung ventilation in order to achieve optimal alveolar partial pressures of carbon dioxide (pCO2) and oxygen (pO2). These pressures play a role in gases dissolving in the blood. However, the total amount of transported gases is also dependent on blood flow, the binding properties of hemoglobin, temperature and hydrogen ion activity. The blood is delivered so close to cells by tissue microcirculation that no other active delivery is needed and as a result, only diffusion takes place here.

Diagram 28: Gases subsystem.

The submodels of gas transport are: ventilation, where air flow, water vapor dilution, temperatures and pressure effects are calculated; oxygen transport; carbon dioxide transport; and acid-base as hydrogen ion activity calculations.

### Ventilation

Natural ventilation is mainly driven by neural reflexes, the sensors of which are central chemoreceptors, which answer to changes in intracellular pH; peripheral chemoreceptors located in the arterial sinus and aorta, which detects changes in blood gases; receptors of the skeletal muscle metaboreflex. The entire afferent path of respiratory reflexes is summarized in the model as one normalized value called TotalDrive; from within this value, the efferent part calculates the respiratory rate and normalized respiratory center’s motoric nerve activity.

From the lung properties are calculated current tidal volume (for example, 450 ml at body conditions of a temperature of 37°C and 100% humidity) and current dead space volume (for example, 150ml at body conditions). As the temperature and humidity in lungs differ from the air in the surrounding environment, alveolar ventilation is recalculated to the inspired air conditions in a submodel called alveolarVentilation.

Diagram 29: Regulation of ventilation.

### Oxygen

The content of air oxygen in the earth’s atmosphere is typically 21%, with atmospheric pressure at 101325 Pa, which yields a partial pressure in air of around 21 kPa. However, the amount of oxygen molecules is still dependent on temperature driven by an [ideal gas](https://en.wikipedia.org/wiki/Ideal_gas_law) equation. For example, in 0°C (273.15 K), dry air is a molar concentration of oxygen at 9.2 mmol/l, while in 40°C, dry air oxygen molar concentration is only 8.1 mmol/l at the same oxygen partial pressure of 21 kPa.

In respiratory paths, air is heated to body temperature and diluted by water. The volume of inspired air changes, which is reflected in the variable AlveolarVentilation, recalculated to inspired air conditions. Once the air is transported to the alveolus, an exchange takes place. Oxygen dissolves in blood plasma and chemically bounds the hemoglobin molecules inside red cells. The dissolving of oxygen in water is driven by Henry’s law (Equation 13), where body temperature is also regulated.

Diagram 30: Oxygen.

### Hemoglobin

Hemoglobin allosterically binds oxygen, carbon dioxide and hydrogen ions, creating cross‑dependences between concentrations of all three substances in the blood (Mateják, et al., 2015).

The most common hemoglobin in adults is hemoglobin A. As a protein tetramer it is symmetrically composed of four subunits, two alpha and two beta units. In the middle of each subunit is heme with a central ferritin atom (Fe2+), where the oxygen molecule is bounded. The bounding of oxygen (oxygenation) causes small changes in the shape of heme, which increase the probability of relaxed space conformation for the entire tetramer. Otherwise, tensed conformation is more common for fully deoxygenated tetramers. The binding of CO2 into a terminal ‑NH2 group of each subunit is known as carboxylation and competes with H+ binding in the same area to form –NH3+. These reactions also has different dissociation constants in tensed and in relaxed conformations. In beta-cleft the is more than ten other amino acid side chains that bind H+ (Bohr’s protons) with different dissociation constants in relaxed and tensed states. In normal conditions, the release of two oxygen molecules is connected with the binding of one H+ and vice versa.

### Carbon dioxide

Most carbon dioxide is transported by blood from tissues to the lungs as bicarbonate (HCO3-). Even small amounts are bounded to hemoglobin and form a significant part of transported CO2 (about 23%), due to its connection with oxygen binding. As noted in a previous section, changes in hemoglobin also change the binding properties of CO2.

HCO3- is a salt of carbonic acid (H2CO3). It significantly affects the acid-base, as mentioned in section 4.4.1. The hydration of free dissolved CO2 to H2CO3 in blood is enzymatically accelerated by carbonic anhydrases inside the red cells, from which the HCO3- is transported to plasma in exchange for the chloride ion Cl-, using Hamburger shift channels to reach Donnan’s equilibrium (Equation 9).

### Comparison with HumMod 1.6

The primary notion of the blood gas transport subsystem is the same in both models and composed of exchange within the lungs, systemic arteries, exchange in tissues, and through systemic veins back to the lungs. The difference is in the calculation of the total amount of gases in the blood, from their partial pressure and vice versa. The Physiomodel uses more precise calculations, based on Siggaard-Andersen’s OSA (Oxygen Status Algorithm) (Siggaard-Andersen and Siggaard-Andersen, 1990). This calculation of hemoglobin oxygen saturation is dependent on pH, DPG, temperature and carbon dioxide. Additionally, the recalculation between carbon dioxide partial pressure and carbon dioxide content in blood is in the Physiomodel also taken from OSA. This divides the total amount of carbon dioxide in blood into free dissolved carbon dioxide in plasma and in erythrocytes, into bicarbonate in plasma, and into bicarbonate and carboxylated hemoglobin amino-terminals in erythrocytes. These calculations are better than the simplified calculations of HumMod 1.6, where the carbon dioxide is calculated only as one bicarbonate concentration for the blood as a whole, independent of hematocrit or oxygen saturation.

## Nutrients and Metabolism

Almost all mechanical energy in the human body is taken from food, metabolized into small high energy compounds such as ATP, which is used by muscles, membrane channels and vesicular transports. The body can metabolize three groups of organic compounds: saccharides, proteins and lipids. After eating these, they are absorbed in the form of base nutrients such as glucose, lactate, amino acids, fatty acids, triglycerides and keto acids. The regulation of uptake, usage, storage, release and transformation of these nutrients is managed primarily by hormones such as leptin, insulin, glucagon and thyroxine.

Diagram 31: Nutrients and metabolism subsystems.

### Cellular metabolism

Base nutrients can be changed in cellular cytosol by glycolysis or lipolysis into acetyl coenzyme A, which is used directly by the mitochondrial citric acid cycle to produce high energy electrons (bound to NADH or FADH), which helps to move the hydronium ions (H3O+ noted as H+) into the mitochondrial intermembrane space. Finally, to equilibrate electrochemical potentials, the hydronium ions have to go back to the mitochondrial matrix through the ATP synthase. The new synthetized ATP is exchanged for ADP and one phosphate across mitochondrial membranes using the electroneutral symporter mechanism.

The ratio between using the base nutrients can differs according to the type of cell (Randle, 1986). For example, the heart muscle prefers lactate more than other organs, while neural tissue prefers glucose and keto-acids and does not use any fatty acids or triglycerides (Owen, et al., 1967). Amino acids can be metabolized only by the liver or in kidney tubules (Hannaford, et al., 1982), because only there can the toxic ammonia (NH4+) be eliminated.

### Liver metabolism

To support good function of all cells, it is necessary to have balanced extracellular concentrations of the base nutrients, even if food is monotone and does not explicitly contain all types of these nutrients in sufficient ratios. The transforming processes from one base nutrient to another takes place in the liver and is known as gluconeogenesis (Wahren and Ekberg, 2007), ketogenesis (McGarry and Foster, 1976) and lipogenesis (Kotani, et al., 2004). Gluconeogenesis creates new glucose from amino acids, ketogenesis creates keto-acids from lipids and lipogenesis can create triglycerides from glucose or amino acids.

The base nutrients can also be stored as lipids in adipose tissue or as glycogen in the liver or in muscles. Stored lipids are long-time reservoirs of energy in contrast with glucose, which is stored as glycogen and which can be used much faster (Chiasson, et al., 1976). The process of storing glucose in glycogen granules is known as glycogenesis and the reversal process – releasing glucose from glycogen – is known as glycogenolysis (Diagram 32).

Diagram 32: Liver transformations of base nutrients.

### Lipids

Lipids are transported from the gastrointestinal tract by lymph to blood plasma using chylomicrons. Chylomicrons contain primarily triglycerides, which are hydrolyzed into fatty acids. These fatty acids can be stored in lipid deposits or used for the metabolic purposes of cells.

In the capillary walls of adipose tissue or muscles, the enzyme lipoprotein lipase is expressed, which transform the triglycerides of chylomicrons into fatty acids and glycerol. These fatty acids can very easily cross the cellular membrane and be stored in adipose tissue or used as metabolism fuel in the form of energy. Only a small amount of free fatty acids is transported by the cardiovascular system and is typically connected to albumin. However, their turnover is extremely fast: every two-to-three minutes, half of these free fatty acids is used as energy by the metabolism and replaced by new fatty acids from lipid deposits (Frayn, 2002).

Diagram 33: Transformation between triglycerides, free fatty acids and lipid deposits.

### Proteins, amino-acids and urea

Almost all proteins from diet are absorbed as amino-acids. There are only 20 types of amino acids and 10 of these are essential. The essential amino acids cannot be synthetized inside the human body and must therefore be ingested as part of food. The primary role of amino acids is to build new proteins; however, they can also be used as fuel for the metabolism. The degradation of amino acids (deamination) takes place almost entirely in liver hepatocytes, because only there can the toxic byproduct NH4+ be transformed into urea. The other area for deamination is in the kidney nephron tubular cells, from which NH4+ is excreted directly to into urine in the form of an extremely efficient mechanism of H+ excretion; this is done without decreasing urine pH. The deamination of amino acids in the liver will prepare new glucose or triglycerides as a source of energy for other cells. The new synthetized urea diffuses into the blood and takes on a primary role in high kidney medulla osmolarity, which is necessary for water balance (Sands, 1999).

### Keto acids

Keto-acids are not the primary fuel for the body’s metabolism, but in some critical situations they can temporarily act as a substitute for missing primary nutrients, especially for neural tissue. During ketoacidosis, there are elevated levels of acetyl acetate and beta-hydroxybutyrate in the body. Keto acid are synthetized in the liver from acetyl coenzyme A, which is created mainly from free fatty acids, acetic acid or ethanol (McGarry and Foster, 1976). They can be metabolized in various tissues, but the speed of their degradation is limited to the speed of the mitochondrial metabolism. Thus, if their production is higher than these limits, they can cause metabolic acidosis accompanied by their elevated renal excretion (Angielski and Lukowicz, 1978; Mateják, 2013).

Figure 7: Keto-acids.

### Comparison with HumMod 1.6

The nutrient metabolism in the Physiomodel has only small differences to the HumMod 1.6 such as in the case of correcting units, stabilizing the shape of approximated curves and the setting of an optimal diet for long simulations.

For example, the physical unit of glucose consumed during anaerobic metabolism by tissue is in HumMod 1.6 expressed in mg/min and this value is assigned by mistake as a variable given in cal/min. In Modelica, a user is not allowed to connect incompatible physical quantities or physical units without converting them first. Thus, this type of error should automatically be prevented in the Physiomodel, because glucose is in this instance always represented by ‘amount of substance’ and ‘molar flows’ as physical quantities using ‘moles’ and ‘moles per second’ as SI-units. These units are used in each instance of the Physiomodel – in chemical reactions, in glucose transports and in glucose storage.

Some approximations of effects use the cubic interpolation splines to express empirical dependences. Using these splines is easy: the user sets only a small number of points on the x-y graph between the dependent variables. For example, for the definition of the linear line, setting one point is enough; for more nonlinear functions, two or three points are typically used. These points contain the value of the x-axe, the value of the y-axe and the value of the slope. The slope poses the biggest problem, because it determines the shape of the function between two points. If the slope is too sharp, a small fluctuation can occur. This means that the shape of the function can leave the interval defined by the y-coordinates of the points. In this situation, the function loses its invertibility and can be inconsistent for some types of non-linear solvers such as “Newton solver”. Therefore, if it is not necessary, it is better to avoid this fluctuation by correcting the nearest slope to a better comfort value or by adding more definition points. In HumMod 1.6, these fluctuations can be observed, for example, in the mass effect of the glycogen to glycogenesis process in the liver, or in the effect of the glucagon to ketogenesis process in the liver. These situations are solved in the Physiomodel by rotating the slope of the middle point to a slightly more horizontal direction, as shown in Figure 8 and Figure 9.

Figure 8: LM\_Metabolism; Glycogenesis; GlycogenMassEffect.

Figure 9, LM\_Metabolism; KetoAcids; GlucagonEffect

For the Physiomodel, the optimal diet for the default setting (for a 75kg male with a metabolic rate of 2500kcal/day) was prepared. A diet of 2000kcal of carbohydrates, 300kcal of proteins and 200kcal of fats per day was selected. In a long-term simulation including this diet, adipose tissue stabilized at 3.6 kg as 5% of body mass – typical for a healthy state with normal secretions of leptin.

## Thermoregulation

### Heat

The human body works best at a core temperature of around 37°C. All chemical processes are dependent on temperature (Equation 6). If the temperature rises, proteins structures become unstable. Even, actually the gene expression of 394 from 12,600 investigated genes are upregulated or downregulated after 20 minutes’ exposure to 43°C, as examined by Sonna et al. (Sonna, et al., 2002). At higher temperatures, heat-shocked proteins are expressed, while at lower temperatures more cold-shocked proteins are expressed (Katschinski, 2004). Both of these instances can change cellular processes, help with protein refolding and if the situation becomes worse, cellular apoptosis can also occur. Additionally, cellular membrane processes are affected. Therefore, the regulation of body temperature is extremely important. The primary mechanism for regulating body temperature is the regulation of blood flow in the skin (Hardy and Soderstrom, 1938; Hsieh, et al., 1965; Kamon and Belding, 1968) (Equation 24), as well as the amount of clothes worn. There is also long-term (in terms of months) regulation of heat by thyroid hormones (Osiba, 1957), which increases or decreases the speed of the basal metabolism as the source of heat in each cell. Short-term heat production is typically based on the working of muscles (Saltin and Hermansen, 1966) or by shivering (Florez-Duquet and McDonald, 1998). The efficiency of skeletal muscle is about 30%; thus, a significant part of consumed energy is released as heat during motion. It is assumed that for the heat transfer of any microcirculation, the temperature of outgoing blood is the same as the temperature of tissue (Equation 24) and as such, blood flow will directly determine the amount of transferred heat between the body core and body tissue. Typically, heat is conducted from a warmer to a colder area, but heat can be also transferred by chemical processes as evaporation against the temperature gradient.

### Evaporation

Significant loss of heat is connected with the evaporation of water (Equation 23) in upper respiratory pathways during air inspiration (Brebbia, et al., 1957). Cold dry air from the environment is here heated to body temperature and fully saturated with water. Water is also evaporated directly from the surface of the skin. In contrast with water loss by respiration, the function of sweat glands can be regulated (Dodt and Zotterman, 1952; HENSEL, 1953; Piwonka and Robinson, 1967; Sato, 1977; Wyndham, et al., 1966). The regulation of sweating sets the amount of water excreted by the skin. During higher physical activity, this evaporated water bounds heat as enthalpy of vaporization, which is more effective in dry, warm and windy environments. However, it works even if the environmental temperature is higher than the body temperature. Therefore, there must be adequate water intake to prevent dehydration.

### Comparison with HumMod 1.6

Thermoregulation shows only one small difference between the Physiomodel and HumMod 1.6, that is, in the fatigue of sweat glands. The use of fuel in sweat glands is in the Physiomodel dependent on the current rate of sweating water, not on the constant default base value of this rate, as in HumMod 1.6.

## Neural Regulations

The integrative model of human physiology also contains the primary neural regulations, because the autonomic nerves directly drive base processes such as the vasoconstriction of blood vessels, heart rate, heart contractility (SUGA, et al., 1976), kidney functions, secretion of hormones, respiration, sweating, etc. The inputs to the autonomic neural reflexes are from specialized cells, which measure the current state of the system: baroreceptors (carotid sinus, aorta, heart atria), osmoreceptors (hypothalamus), chemoreceptors (carotid sinus, aorta, medulla oblongata) and thermoreceptors (skin). In these cells, the neural impulses start that are used for the calculation of a final answer. There are two autonomic pathways: sympathetic and parasympathetic. When a signal reaches the end of the last neuron in a pathway, noradrenaline is typically released for sympathetic stimulation and acetylcholine is typically released for parasympathetic stimulation. The synapse receptors of the effector cells are typically muscarinic in the parasympathetic pathways and adrenergic in the sympathetic pathways. There are two groups of adrenergic receptors: alpha and beta receptors. Both adrenergic receptor groups react on epinephrine and on norepinephrine. As a result, the model for the alpha/beta receptors may be dependent on sympathetic stimuli, together with the extracellular concentration of these catecholamines and other agonists (e.g., desglymidodrine) or antagonists (alpha/beta-blockers). The model of alpha receptors can be used in many instances such as for the model of microcirculation, which is used for many tissues with different parametrical settings.

Autonomic regulations simply correct the functions of cells, tissues and organs. In many cases, they are not necessary for life. They can even be removed through surgery (vagotomy, endoscopic thoracic sympathectomy). However, quality of life rapidly decreases, because the loss of regulation decreases the limits to where the body can function properly. For example, the loss of regulation of heart rate is critical in instances of increased physical activity, when higher cardiac output is needed to support oxygen transport to muscles. Without external innervation, the heart still beats using autonomic oscillations of sinoatrial node cells (Bootsma, et al., 1994; Warner and Cox, 1962), but without receiving useful information (Xenopoulos and Applegate, 1994) about muscle metabolism, current blood status and about blood pressure (Ferguson, et al., 1985; Takeshita, et al., 1979).

|  |  |  |
| --- | --- | --- |
| Neural pathway | Receptors | Effectors |
| Baroreflex | carotid sinus artery and atrial baroreceptors | heart, vasoconstriction/vasodilation |
| Metaboreflex | skeletal muscle chemoreceptors | heart, vasoconstriction/vasodilation |
| Thermoregulation | skin thermoreceptors, core thermoreceptors | vasoconstriction/vasodilation of skin vessels, sweating |
| Respiration reflexes | central chemoreceptors, peripheral chemoreceptors,  skeletal muscle pH | respiration rate, tidal volume |
| Drivers of kidney function | atrial low pressure receptors | proximal tubule sodium reabsorption |

### Comparison with HumMod 1.6

In the Physiomodel, the reaction of central chemoreceptors is shifted to new and more precise values of their intracellular pH=7.08 (Kintner, et al., 2000) as a result of detailed tissues and respiratory quotients, based on the metabolic consumption of base nutrients. The details of this shift is described in section 6.1.

The next difference is the correction of the shape of the oxygen effect in the same central chemoreceptors, where the changes between the Physiomodel and HumMod curves are plotted in Figure 10.

Figure 10: Correction of oxygen effect on central chemoreceptors.

# Discussion

The integrative description of human physiology using Modelica opens up new possibilities of scientific examination and the interconnection of physiology within many scientific disciplines. One must realize that according to a physiological description there are always some chemical, hydraulic, thermal, osmotic or population-based domains. Furthermore, the laws of these physical domains can be exactly described by mathematical equations. As was shown, the entire human physiology can be described as a large set of these mathematical equations. Without the object-orienting equation-based approach of computer science, it is not possible to easily create, extend, modify or understand such extremely large mathematical models.

This work does not describe how to find, fit or estimate the large number of parameters in the presented model. Some parameters are patient-specific, such as weight or body volume, organs or tissues. However, most parameters are properties of chemical substances, membranes, tissues and organs, which are typically not patient-specific and are invariant in terms of other variables or parameters. On the other hand, there are also interpolation curves, which very simplified empirical observations such as of the hormone effects or neural reflexes. In the following model developments, these interpolations should be more complexly modeled by both the physical and chemical bases of the processes. For example, the shifting of the interpolation curve of hemoglobin saturation with oxygen, as is implemented in HumMod 1.6 does not allow for examining this process together with the binding of Bohr’s protons and with the binding of carbon dioxide, as in our model. It also does not show the transfer of heat using these chemical processes.

## Physiological expandability

After we implemented the HumMod model in Modelica, we realized that improving the model is possible without impacting work conducted previously. Comparing the resulting model with the original one, alongside the correct consultation of measured data yielded better equations and a more stable model in each improving step.

For example, in the original model the production of carbon dioxide is calculated in all tissues using the same global respiration quotient (RQ). The flow of produced carbon dioxide was proportional to the flow of consumed oxygen and was independent according to the types of tissues. The right flow of consumed oxygen with the right type and amount of consumed metabolites had already been calculated. Furthermore, because we knew the respiration quotient for each of these metabolites, we improved the amount of produced carbon dioxide per tissue. As a result, there was higher production of CO2 in neural tissues, because neurons cannot metabolize fats (RQ=0.7) and their primary source of energy is glucose (RQ=1). Comparing the results of the simulation before and after this modification, it was observed that the patient began to hyperventilate. After a short investigation, it was found that this hyperventilation was caused by the neural regulation of respiration. Observing this was quite easy, because the total effect of neural respiratory regulation (TotalDrive) in a normal situation should have the value one such as each other effect. Looking at diagram one, the connected systems from which the TotalDrive is calculated can be seen. After examination of these few systems, it became clear that it goes from central chemoreceptors, which fire the impulses in answer to the intracellular pH of these neural cells in the brain. Logically, there is a shift in the intracellular pH of neurons (pHin), caused by the higher production of CO2. It must be established whether the new value of the calculated pHin reflects the reality of the situation any better. The measured value of pHin in physiological papers were found to be closer to our new value (Kintner, et al., 2000), which was automatically calculated from the model acid-base equations of each intracellular environment. As a result, it was necessary to shift the interpolation curve, which was estimated to lower values of pHin. These effects of central chemoreceptors had to be shifted to have normal value at a normal pHin. After this correction, the model had the same or better neural respiratory regulations and the same or a better normal state of other variables, and with more precise local status of tissues and blood. These results indicated that all changed values had to be consulted and compared with other research, as the pHin had been in our case.

The other example of physiological expandability concerns the addition of the new acid-base module. At first we tried to calculate the blood acidity from the base excess of oxygenated blood (BEox) using the empirical Van Slyke equation, as Siggaard Anderson did with cTH=- BEox. Because the BEox was not calculated in the original model, it had to somehow be defined. The first idea was to create the state variable. Some sources for acids or bases (mainly from the metabolism) and some losses of acids and bases (mainly through the kidneys) were found. The change in bases was the same as the change of BEox; the change of acids inferred the negative change of BEox. However, this implementation failed after a few hours of simulation. After a short examination of the simulation results, the lost electroneutrality of blood plasma was observed. The model had already integrated all main electrolytes and all main acid-base puffers; as such, it was possible each time to calculate strong ion differences (SID) or the charge on acid-base puffers at the calculated pH. As a result of electroneutrality, the sum of all these charges must be zero, which was not the case. A first generator of these problems was assumed to be the changes in BEox; had there been more properly connected flows of electrolytes and flows of all organic acids alongside the changes in BEox, the stability of the model would have increased from hours to days or even months of simulation time. However, it failed because of electroneutrality. This gave rise to a new idea: if the model includes all electrolytes and acid-base buffers as state variables, then the BEox is not a state variable and it can be directly calculated from the equation of electroneutrality, as is described in section 4.4.1. After improving BEox, the model became stable for more than a year of simulation time. As a result of this added equation, it never again lost electroneutrality.

## Expandability in the field of physical chemistry

The main problem with the original model is that calculations are causal and redundant. Almost all physical and chemical laws as equations are repeated as many times as it occurs within the human body. Modelica provides the opportunity to define one law only in one place and to use it simply by referencing this place. It does not matter which variable is used as an output from the equation, because Modelica can automatically effect algebraic manipulation during compilation time.

For example, consider that only one component defines each chemical reaction. Improving this component will improve all chemical reactions in the model. My first implementation of a chemical reaction had two main parameters: dissociation constant and reaction forward rate. At equilibrium, all is calculated only from the dissociation constant, because there is no speed of reaction. The next investigation of chemical bases told us that dissociation constant is not really a constant, but dependent on temperature; thus, the Van’t Hoff’s law was added to this component with a default setting at zero reaction enthalpy. In other words, if reaction enthalpy as a parameter is not set by the user as a specific occurrence, it will have the same behavior as before. However, it also offers the possibility of defining non-zero enthalpy for reactions with equilibrium dependent on temperature. The investigation of the meaning of reaction enthalpy also gave rise to another idea: to calculate the flow of heat energy from/to a reaction and using the conditional thermal heat port, this allowed the chemical reaction to be a multidomain (chemical and thermal) component. Thus, as in the chemical theory, positive reaction enthalpy means the endothermic reaction and the negative value of reaction enthalpy implies the exothermic reaction. Moreover, because the heat port was hidden by default, all instances of chemical reactions in the model remained the same, with the same settings and connections as before these thermal extensions.

The next approach in the field of physical chemistry showed us that we can also calculate the dissociation coefficient at a defined temperature from the thermodynamic properties of substrates and products of the reaction. The idea was to simplify the use of chemical reaction components. The user simply selects the type of substrates and products and the dissociation coefficient is automatically calculated. This approach uses a database of chemical substances alongside their free enthalpies of formation (ΔfH) and free Gibbs energy of formation (ΔfG). The enthalpy of the reaction is the sum of formation enthalpies of products minus the sum of formation enthalpies of substrates ([Hess’s law](https://en.wikipedia.org/wiki/Hess%27s_law)). Having Gibbs energies of all products and all substrates, the Gibbs energy of the reaction (ΔrG) is also the result of Hess’s law (Equation 11), while the dissociation coefficient (K) of the reaction at temperature T is defined from the Gibbs energy of the reaction as in Equation 34, where R is the gas constant.

|  |  |
| --- | --- |
|  | Equation 34: Gibbs free energy of the reaction. |

Chemical processes in the body occur in the water solution of electrolytes. The adaptation to water conditions can be done using activity coefficients. Water surrounds the charged particles and creates solvation shells, which decrease the activity of the substance. This behavior is driven by the Poisson-Boltzmann model, which can be simplified using the Debie-Hückel theory. The creation of metabolic pathways should be simple, i.e., simply by connecting substances with the implementation of chemical reactions. The user simply selects the names of substances instead of odd values of dissociation coefficients for each reaction. If these reactions are all in equilibrium, it is not even necessary to apply the value of kinetic rate coefficients in order to start the simulation.

This type of improvement also guarantees the more sophisticated rule of chemical systems called the “[Principle of Detailed Balance](https://en.wikipedia.org/wiki/Detailed_balance)”. This principle states that in a closed equilibrated chemical system, equilibrium is reached at each chemical reaction. As a result of this law, the product of dissociation constants in chemical circles is equal to 1. Therefore, if the user defines the chemical system as chemical reactions in a circle and wishes to set all dissociation constants as parameters, he must always think about the dependences between them. However, the new proposal based on Gibbs energies does not allow for breaking this fundamental rule of chemical systems. For example, having a closed system with the following chemical reaction: A1<->A2, A2<->A3 and A3<->A1 after an extended time, when the concentrations of A1, A2 and A3 are constant. The K12= A2/A1, K23= A2/A3 and K31= A3/A1 are dissociation constants of the reactions at temperature T. If the user wants to set the dissociation constants, the following requirements must be verified by the user: K123 = K12\* K23\* K31 = 1. However, if the system is calculated from Gibbs energies of formation of each substance ΔfG1, ΔfG2 and ΔfG3, then this relation will be automatically fulfilled as relation 0 = (ΔfG2- ΔfG1) + (ΔfG3- ΔfG2) + (ΔfG1- ΔfG3) = ΔrG12 + ΔrG23 + ΔrG31 = ΔrG123, because ΔrG = 0 if and only if K = 1 as in Equation 34.

## Effort of Integrative Physiology

If we assume that the extension is made using the rule of integration and image extension, then this theoretical view should confirm our hypothesis 1: “**For each physiological model there exists an extension, which is at least as well as the original model”**.

The practical situation is somewhat different, however, because the image extension is often also connected with domain extension, which is not assumed as an improvement of the model. In many instances in the Chemical library and Physiolibrary, this is solved using the “default” setting. This means that the new parameters are designed as unnecessary if the user does not want to use them. In this case, the user does not need to set them, because they have a default value. An example is the kinetics coefficient of chemical processes; this is set to be fast enough to equilibrate the process in a very short time. The user, who is interested only in viewing the chemical equilibrium, will simply ignore setting the kinetics coefficient.

Therefore, if possible, it is better to design parameters that have already been scaled and that are almost independent of other settings. For example, the weight of organs and tissues can be represented by parameters defined as fractions of the entire weight. This automatically scales the size of each organ when we set only one parameter, i.e., the total weight of the individual patient. Furthermore, the default values that solve the switch between preselected genotypes or phenotypes can be considered as part of the model. This can decrease even a complex setting into choosing one type of setting, which is considered a rapid improvement of the model by applying the theory noted above. Modelica has been extremely helpful in solving the initial equations where relations between parameters can be implemented.

The formalization of model properties also partially answers our second hypothesis: “**Mathematical formalization and integration of practical physiological knowledge about one organism can be implemented into one complex physiological model”**.

As an extension, integration also has the same problems in terms of increasing the parameters and inputs of the model. Solving this problem using default values, scaled parameters or predefined datasets is not always a good idea. Mass should be guaranteed, so that during integration, redundant relations do not occur in the model. This state is very dangerous and is not automatically detectable. The redundant equation or redundant variable is an alternative equation of something that is already present in model. If this redundant variable is the state variable, its change as a type of flow is calculated twice; however, there must not be twice the amount of flows in any model. If an entirely new subsystem is inserted to the model that already calculates some parts of that subsystem, then the regulations and answers to the state of the new subsystem will be incorrect. The only clear solution to this problem is to remove all duplicated variables and duplicated relations, even if they calculate the same data but in a different way. In this way, the trusted relations can be selected and relations with lower credibility can be eliminated. An example of this is rendering our extension of the acid-base model as described in sections 6.1, 4.4.1 and 4.4.2.

# Conclusion

The theory of integrative model development has been formalized (Section 5). Using these rules, the [Physiomodel](http://www.physiomodel.org/) (Section 4), a large integrative model of human physiology was created in [Modelica](https://www.modelica.org/), a computer language standardized by the [Modelica Association](https://www.modelica.org/association). The model is an extension and integration of the [HumMod](http://hummod.org/) 1.6 model, created by the Mississippi University of Medical Center. The Physiomodel integrates the original HumMod model with the new acid-base model (Section 4.4.1, Section 4.4.6 and Section 6.1) and new blood gases transport approach (Section 4.5). As part of the acid-base and gases transport model, a new general model of complex chemical equilibria on macromolecules was developed, as demonstrated by the O2-H+-CO2-HemoglobinA system (Section 4.5.3 and Section 3.4.5). The primary results of this work include also the new Modelica libraries called [Physiolibrary](http://www.physiolibrary.org/) and [Chemical library](https://github.com/MarekMatejak/Chemical/releases) (Section 3), which have already been integrated to the open-source Modelica environment called [OpenModelica](https://openmodelica.org/).

The results of the formalized theory of model development (Section 5) are the rules for comparison, reduction, extension and integration of models. The comparison between models was based on real experiments that can be described by the models. Reduction here refers to the elimination of correlated parameters. Extension refers to the addition of new relations with new variables to the model, which does not collide with aspects of the rest of model. Finally, integration refers to the development of a new theory that can describe all real experiments of previously separated models. The rules have been designed to always achieve the best model, better from those models preceding it.

During the compilation of the Physiomodel (Section 4), more than 8000 variables were presented. Most of these variables were generated from graphical connection, so that the number of real physically distinguishable variables was between 4000 and 5000. There were also more than 4000 parameters. However, each of these had a default value and they were designed using the reduction rule to be invariant of other parameters or variables. For example, instead of a parameter containing the absolute weight of the tissue or organ, there is a parameter that contains a fraction of the total body weight for the tissue or organ. Thus, it was easy to simulate the experiments, because they could be described only by a few parameters, while the rest remained set as default values.

During implementation of the HumMod, we corrected more than 30 mistakes in the original implementation, as described in the subsection called “Comparison with HumMod 1.6” when discussing the appropriate position and description of the whole model (4.1.5, 4.2.6, 4.3.9, 4.4.6, 4.5.5, 4.6.6, 4.7.3, 4.8.1). Most of these mistakes were caused by the misuse of non-SI units (e.g., mg/L instead of mg/dl) or by the misuse of physical quantities (e.g., mass concentration instead of molar concentration). Other mistakes included the misuse of variable connections (exchanging variable names) or ignoring known facts (e.g., the charge of inorganic phosphates is dependent on pH). Additionally, some subsystems were designed very simply in order to describe the physiological functions that we wanted to examine (e.g., synergy of blood gas binding to hemoglobin).

As has been shown in the section on physical and Modelica backgrounds (Section 2), the principles can be implemented at a highly error-proof level. The correctly defined basis can generate errors or warnings almost every time when the user wishes to use a variable or class in a questionable way. In Modelica, the user can design predefined submodules with well-defined inputs, outputs, parameters and connectors. For future usage of these components, it is almost impossible to confuse physical units or physical quantities.

The object-oriented equation-based Modelica language allowed us to create extremely robust support of integrative physiology based on physical principles (Section 3). Similar support has already been implemented for technical sciences such as libraries for electrical, magnetic, mechanical and thermal processes. However, in physiology, we need models for more sophisticated domains such as electrochemical (for chemical reactions, membrane transports, water evaporation and diffusion) and hydraulic domains (for cardiovascular systems). By creating an analogy to existing systems we created Physiolibrary 2.3, a library for physiological processes, which was named by the Modelica Association in the Modelica Library Awards as the best free Modelica library in 2014. We have released that the general library for electrochemical processes, called Chemical library, is helpful for more general electrochemical equilibration processes. This library is a by-product of our integration of membrane electrochemical equilibria, based on Donnan’s electrolytes equilibrium and Nernst membrane potential in the field of physical chemistry, as described in section 3.4.

1. Decomposition from up to down means starting with the whole and dividing it to smaller pieces. Decomposition from down to up means starting with many small pieces and joining them together. [↑](#footnote-ref-2)