Amino Acids (2005) 29: 341–353
DOI 10.1007/s00726-005-0215-8

Amino Acids
Printed in Austria

Principles of quantitative positron emission tomography

Review Article

J. van den Hoff

PET Center, Institute of Bioinorganic and Radiopharmaceutical Chemistry, Research Center Rossendorf, Dresden, Germany

Received March 2, 2005
Accepted May 13, 2005
Published online July 12, 2005;

Summary. The central distinguishing feature of positron emission tomography (PET) is its ability to investigate quantitatively regional cellular and molecular transport processes *in vivo* with good spatial resolution. This review wants to provide a concise overview of the established principles underlying quantitative data evaluations of the acquired PET images. Especially, the compartment modelling framework is discussed on which virtually all quantification methods utilized in PET are based. The aim of the review is twofold: first, to provide the reader with an idea of the theoretical framework and mathematical tools and second, to enable an intuitive grasp of the possibilities and limitations of a quantitative approach to PET data evaluation. This should facilitate an understanding of how PET measurements translate into quantities such as regional blood flow, volume of distribution, and metabolic rates of specific substrates.

Keywords: Positron emission tomography – PET – Tracer kinetics – Compartment modelling

1 Introduction

Positron emission tomography (PET) aims at non-invasive investigations of cellular and molecular processes *in vivo*. To this end the PET technique assesses the three-dimensional spatial distribution of special tracers incorporating positron emitting radioactive isotopes.

These positron emitting isotopes – the ''label'' – can substitute the corresponding stable isotopes in relevant biomolecules (e.g., by replacing the stable ^{12}C by the positron emitter ^{11}C) leading to a wide class of PET tracers whose *in vivo* behaviour is unaltered in comparison to their naturally occurring counterparts.

The label can also replace other chemical elements in the molecular structure – for example, the widely used ^{18}F can replace hydrogen in many organic compounds. This approach leads to PET tracers whose chemical properties are deviating from those of systemic substances. In this way it is possible to create tracers which desirable special properties such as enhanced trapping in the target regions. A notable example of this approach is [^{18}F]-2-fluoro-2-deoxy-D-glucose (FDG), which is currently the most widely used PET tracer for clinically oriented oncological investigations.

Over the last 20–25 years the PET technique has steadily evolved from a promising new experimental approach to a well established tool for basic research as well as for clinical applications in humans. A recent overview of the historical development is given in (Jones, 2003).

The ever increasing interest in PET can be traced back to a fundamental difference between this method and the other major tomographic techniques, X-ray computed tomography (CT) and magnet resonance imaging (MRI). The latter methods utilize physical properties of the organism under investigation such as tissue dependent absorption of X-rays (CT) and proton densities or relaxation times of an externally induced nuclear magnetization (MRI). As such, the images provided by these methods only allow more or less indirect conclusions concerning biological processes. PET, on the other hand, detects the distribution – and dynamic redistribution – of specific tracers, whose properties are tailored to be sensitive to special biochemical and physiological transport steps. Moreover, the radioactive tracer (the ''imaging agent'') is applied in true tracer amounts, usually at the pico- or nanomolar level, which guarantees that the system under investigation is not perturbed by the applied tracers. CT and MRI contrast agents, on the other hand, have to

be applied in macroscopic quantities. Overall, CT and MRI can be seen to provide complementary information, notably a much better definition of anatomic structures than is achievable with PET. Thus a combination of the different tomographic techniques is, if feasible, the best strategy.

Consequently, the primary signal detected by PET – the time dependent regional concentration of the utilized tracer – always yields truly functional information from the living organism. Phenomena becoming accessible in this way include fundamental physiologic information such as, for instance, regional blood flow. Qualifying this information as ''fundamental'' is not to say that this information can equally well be provided by the other tomographic imaging techniques. Notwithstanding the unquestionable progress, especially in functional MRI, PET remains to date the only modality capable of providing this information in a truly quantitative and unbiased way for volumetric data.

Other phenomena which are accessible by PET measurements encompass, to name the most important, transport processes across diffusion barriers (with respect to clinical applications the most relevant example being glucose transport into the cell), substrate metabolism, protein synthesis, enzyme activity and receptor density/affinity. In this way, it is possible to obtain very specific information concerning a variety of selected molecular processes with the PET imaging technique (''molecular imaging'').

This review wants to provide a concise overview of the established principles of quantitative data evaluation of PET investigations. The aim is twofold: first, to provide the reader with an idea of the theoretical framework and second, to enable an intuitive grasp of the possibilities and limitations of a quantitative approach to PET data evaluation.

Readers interested in more extensive or complementary presentations are referred to the literature. Much of the utilized methodology is derived from, or closely related to classical tracer kinetic techniques which are discussed thoroughly in (Lassen and Perl, 1979). Discussions of the special aspects of PET tracer kinetics together with collections of useful equations frequently encountered in standard applications of the formalism can be found, for instance, in (Carson et al., 2003; Carson, 2003; Willemsen and van den Hoff, 2002).

2 Mode of operation

In PET investigations, the radioactive tracer typically is injected intravenously. Other application schemes (intra-arterial, inhalation, or even orally) are possible – and might be necessary – depending on the properties of the tracer and/or the question addressed by the investigation.

The measurement of the regional tracer distribution with the PET method rests on the detection of the decay of individual atomic nuclei of the radioactive label as follows:

Upon the decay of the radioactive label a positron is emitted from the atomic nucleus that is rapidly stopped in the surrounding tissue (mean range in water: 1.1 mm, 1.5 mm, 0.6 mm for ^{11}C, ^{13}N, ^{18}F, respectively). The positron then annihilates with an electron accompanied by collinear emission of 2 photons, i.e. electromagnetic radiation, in opposing directions.

These high energy (511 keV) photons are detected with scintillation detectors arranged in a ring coaxial to the patient's body axis. The coincident detection of two photons within a time window of a few nanoseconds defines for each detected pair of photons a so-called Line of Response along which the primary decay of the radioactive label has occurred. Accepting only these coincident events yields the necessary directional information that is required for subsequent tomographic image reconstruction of the tracer distribution.

It should be noted, that PET always provides three-dimensional tomographic data sets, i.e. stacks of tomographic images covering a substantial volume of the investigated organism. Modern PET tomographs achieve nearly constant spatial resolution within the imaged volume.

To derive quantitative images, corrections for effects such as scattering and attenuation of the annihilation radiation in the traversed tissue are included in the data processing. Inclusion of these corrections ensures that the resulting images directly yield the regional concentration of the applied tracer substance in absolute terms (usually in units of [Bq/ml], i.e. amount of radioactivity per unit volume).

Data acquisition can be performed either as a so-called static scan which provides a single snapshot of the tracer distribution or as a so-called dynamic scan which represents a contiguous series of successive scans over the whole interesting time range. The latter type of data acquisition is generally a prerequisite of quantitative evaluation of the tracer kinetics.

3 Data evaluation methods

As has been explained in the last section, a PET investigation essentially provides quantitatively the three-dimensional spatial distribution of a radiopharmaceutical

as a function of time. Depending on the question to be answered evaluation of the acquired data can be performed at various levels of sophistication.

The most elementary approach to evaluation of the tracer distribution is visual inspection of the reconstructed tomographic images. In many cases, deviations of the regional distribution of the radio-pharmaceutical from the distribution observed in a reference group occur. Detection of such deviations might suffice for diagnosis, therapy evaluation, or for the assessment of drug effectiveness.

Of course, this approach uses only a small fraction of the information, namely the regional contrast within the given tomographic data set averaged over a certain time interval. Moreover, it does so in a qualitative way only. This is a serious shortcoming of the method. An increase of the image contrast of a target region relative to some ''distant'' reference region within the image volume – the increase being evaluated by comparison to some standard, for instance a baseline investigation prior to pharmacological intervention – may be caused by an increased tracer uptake in the target tissue or a reduction of uptake in the reference region or by a combination of both effects. Moreover, global alterations of the uptake are not detectable at all by this approach. In order to improve the situation, visual analysis is often combined with simple quantitation procedures, especially regional tracer concentration calculations and relative uptake determination. These methods are sufficient for many investigations, in particular for clinical diagnosis.

However, the above-mentioned methods are not always adequate or are even inapplicable because the tissue concentration of a radiopharmaceutical is varying with time. This variation is dependent on the way the plasma concentration varies with time as well as on the kinetics of the tracer in tissue. Consequently, the relation between the measured distribution at a fixed time point and the parameters used to describe the tracer kinetics may be strongly non-linear. Sometimes a one-to-one relation between the single parameters and the distribution may not even exist: different kinetic behaviour might lead to identical tracer concentration levels at a certain time point: the time activity curves can intersect. In these cases it is essential to employ more advanced data analysis methods, which utilize the full dynamic information and are based on an adequate modelling of the underlying transport processes. The overall procedure is depicted in Fig. 1.

First of all, the PET data of the tracer accumulation in tissue have to be combined with a separate measurement of the blood activity of the tracer. It is further necessary to have sufficient a priori knowledge of the behaviour of the tracer to set up a realistic and complete model of transport and metabolism of the tracer.

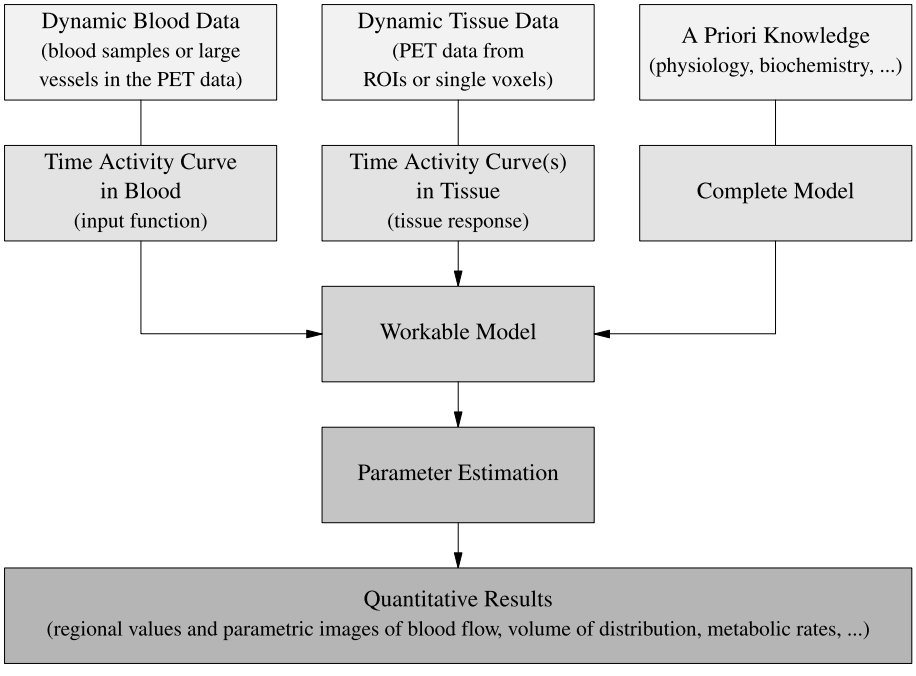

Fig. 1. Steps towards a quantitative description of a PET investigation

The next important step is model reduction, the process of simplifying the complete – and therefore usually complex – model in such a way that data evaluation becomes feasible. This step is necessary, because the PET data itself usually do provide only restricted information concerning the microscopic transport and metabolism of the tracer. We will come back to this point later.

After having set up a ''workable'' model, the next step is the estimation of the model parameters by standard numerical techniques such as least squares fitting.

On ends up with quantitative regional values of the model parameters which correspond to phenomena such as blood flow, metabolic rates, and so forth. This information is provided at the level of spatial resolution of the tissue time activity curves used in the data evaluation. If data evaluation is performed individually for each voxel in the tomographic data set, so-called parametric images of the parameters are obtained.

Today, essentially all approaches to quantitative evaluation of PET investigations are based directly or implicitly on the concepts of compartment modelling. The central postulate of compartment modelling is the assumption that the process/system under investigation can be decomposed into a usually small number of functional units, called compartments, which can be treated as structureless pools containing the tracer in distinct states. Thus, by definition no concentration gradients can be present within a compartment. It is important to realize that a compartment does not need to have an anatomical correlate. Rather, it might describe the tracer in a certain identifiable chemical modification. Due to the tracer principle one usually ends up with pure first order kinetics, i.e. with a mathematical description by first order differential equations. In the majority of cases, this description is adequate. However, in the special case of carrier-added experiments it may no longer hold.

The processes described by the model can be either the actual migration from one physical space to another, e.g., from plasma to intracellular fluid, or a biochemical transformation within a single physical space, e.g., the phosphorylation of FDG to FDG-6-phosphate in the cell. Although each of these steps may be composed of many intermediate processes, it is possible to combine the intermediate processes into one compartment if they proceed much faster than the ''input'' and ''output'' steps. For example, a tracer may be transported first from plasma to extracellular fluid, then into the intracellular fluid and finally into the mitochondria. Nevertheless this can be modelled as a single step process provided that these subsystems approach steady state sufficiently rapid.

4 Theoretical framework

4.1 Conservation of mass

Before concentrating on details of the compartment modelling approach, some terminology and the basic assumptions common to quantitative evaluation of most tracer experiments are summarized.

The principle of every tracer kinetic investigation of a single organ or the living organism as a whole – the ''system'' under consideration – is the determination of the relation between time dependent tracer delivery (the ''input'') to the system, in general via the arterial blood, and the time dependent reaction of the system (the ''response''). Figure 2 visualizes this situation. The tracer concentration in the arterial blood is given by $c_a(t)$, the venous outflow by $c_v(t)$ and the concentration in the system itself by $c_s(t)$. It is a characteristic feature of PET measurements that, quite generally, all measured quantities are normalized to a unit volume of the system: PET measures not tracer amounts, but rather tracer concentrations. These can of course be translated into amounts/masses, if the volume of interest is specified. For the same reason, when setting up the mass balance equations describing the behaviour of the tracer, these equations usually take on the form of apparent ''concentration balance'' equations, although there is no law of ''conservation of concentration''.

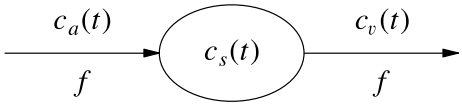

Fig. 2. Principle of a tracer kinetic experiment. The tracer is delivered with the time dependent concentration $c_a(t)$ to the system indicated by the ellipse. Inside the system the tracer concentration is given by $c_s(t)$ and the tracer leaves the system with a time dependent concentration $c_v(t)$. f is the local perfusion, i.e. the blood flow F divided by the system volume V

Because appreciating this fact is of principle importance to a correct understanding of a PET measurement, let's look at the situation depicted in Fig. 2: the mass balance is given by

$$m_s(t) = \int_0^t F c_a(s) ds - \int_0^t F c_v(s) ds,$$

where $m_s(t)$ is the amount of tracer present in the system at time t, and F is the blood flow (usually in units of [ml/min]). This equation expresses the simple fact that the instantaneous $m_s(t)$ is given by the difference of the cumulative input and the cumulative outflow up to time t.

Dividing this equation on both sides by the system volume V leads to

$$c_s(t) = \int_0^t f c_a(s)ds - \int_0^t f c_v(s)ds \quad \text{with } f = \frac{F}{V}. \tag{1}$$

One ends up with an equation containing only "specific" quantities, i.e. quantities referring to a unit of the system volume: the tracer amount m_s present in the system is replaced by the tracer concentration c_s and blood flow F is replaced by perfusion f, i.e. the blood flow per unit volume of perfused tissue (usual units: ml[blood]/min/ml[tissue]). In passing, it should be noted that Eq. (1) is the basis of the classical Kety-Schmidt method for determination of the global cerebral perfusion (Kety and Schmidt, 1948).

4.2 Linear tracer kinetics: impulse response and convolution

In the next step towards a quantitative description use is made of the fact that in tracer experiments we are dealing with a so-called linear system: due to the very high detection sensitivity of the PET method, PET-tracers are usually applied in the picomolar concentration range (we will in this context ignore the possibility of sizable amounts of "cold", non-radioactive compound which might be present – intentionally or inadvertently – in certain circumstances). Therefore, the tracer principle is usually strictly valid: there is no perturbation of the investigated processes by the tracer. This implies that the kinetics of the tracer is always linear with respect to alterations of tracer concentration even if the traced process obeys nonlinear kinetics, for instance of the Michaelis-Menten type.

More formally, this means, that the system under investigation is completely described by its so-called "unit impulse response" $R(t)$ which is the reaction of the system to an idealized, infinitely short input of unit dose at time zero, $c_a(t) = \delta(t)$, called the delta function.

If the impulse response $R(t)$ is known, the system's response to an arbitrary input $c_a(t)$ is given by a mathematical operation called the convolution of $c_a(t)$ and $R(t)$ (denoted by the symbol $\otimes$) which is defined as:

$$c_s(t) = (c_a \otimes R)(t) = \int_0^t c_a(s)R(t-s)ds. \tag{2}$$

This equation mathematically expresses the fact, that the system's behaviour is linear with respect to $c_a(t)$: the system response $c_s(t)$ is constructed as an infinite sum – the integral – of the contributions from separate responses to a succession of fractional delta inputs with doses $c_a(s)ds$ at differing time points s between time zero and the present observation time t.

Equation (2) implies

$$c_s^{(a+b)}(t) = c_s^{(a)}(t) + c_s^{(b)}(t)$$

and, especially,

$$c_s^{(N\cdot a)}(t) = N \cdot c_s^{(a)}(t),$$

i.e., the response to a combination of two inputs $a(t)$ and $b(t)$ equals the sum of the separate responses to both inputs, and the response to N times input $a(t)$ equals N

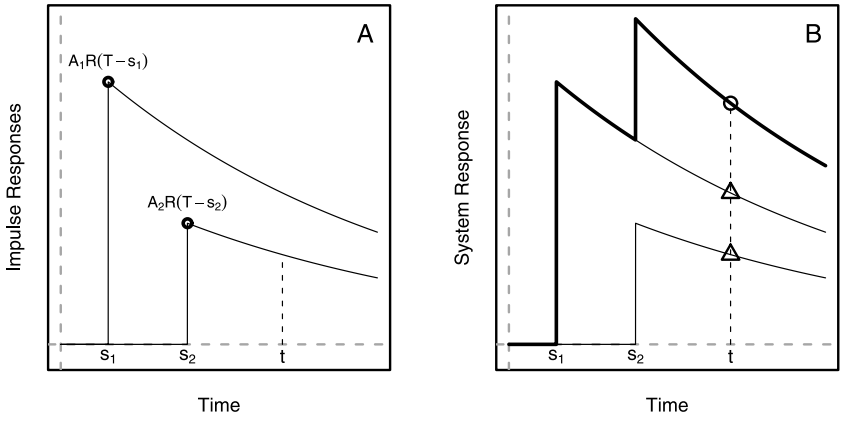

Fig. 3. Principle of convolution. Two impulse inputs of dose A_1 and A_2, respectively, are administered at time points $T = s_1$ and $T = s_2$ (i.e. $c_a(T) = A_1\delta(T - s_1) + A_2\delta(T - s_2)$). **A** The responses to the separate impulse inputs. These are given by the unit impulse response $R(T)$ after a time shift by s_1 and s_2, respectively, and multiplication with the corresponding amplitudes (doses) of the impulse inputs. **B** The system response to the combined inputs. This is given by addition of the separate responses shown in **A**. Taking into account that $R(T - s < 0) = 0$, the system response $c_s(t)$ at an arbitrary time point $T = t$ is given by

$$c_s(t) = A_1 \cdot R(t - s_1) + A_2 \cdot R(t - s_2).$$

Cf. Fig. 4 for the generalisation to multiple inputs. In the limit of a continuous input, one is let to the general convolution equation Eq. (2)

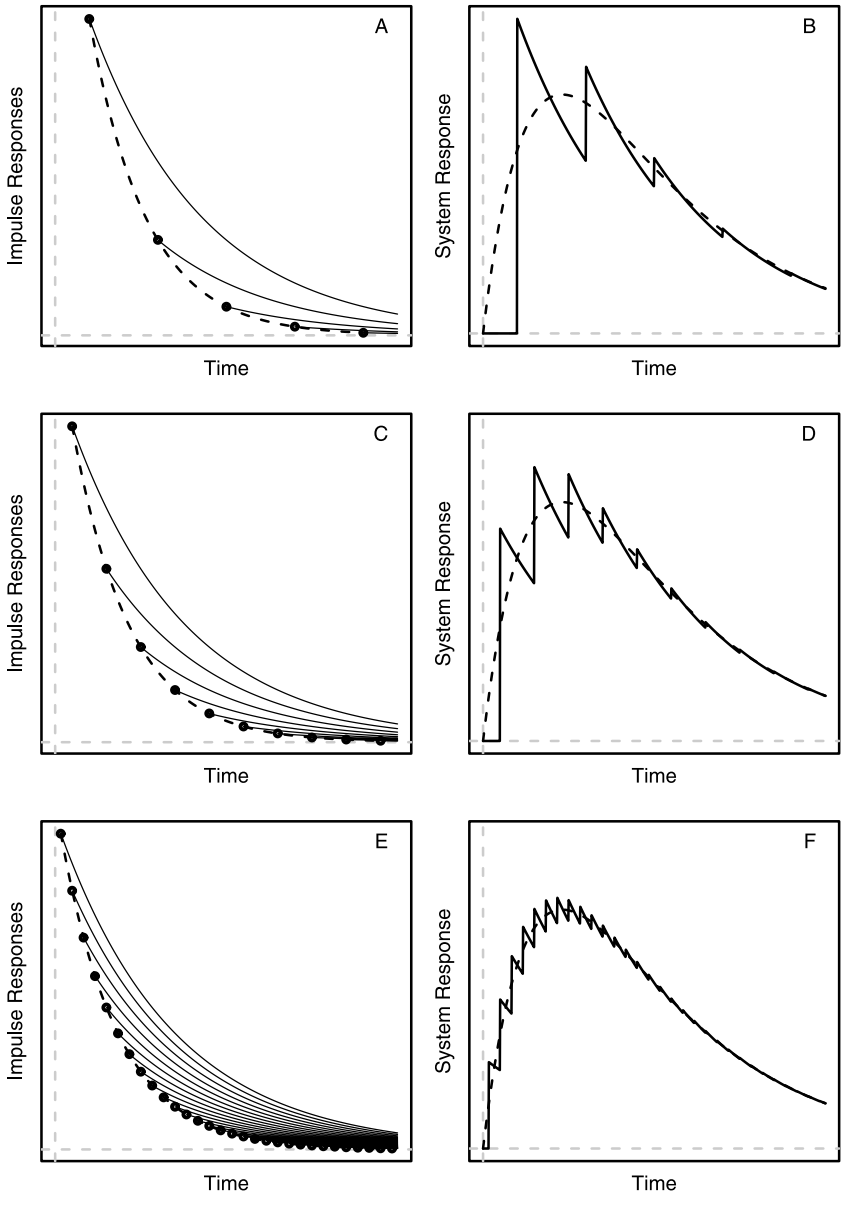

Fig. 4. Generalisation of Fig. 3 to multiple impulse inputs. **A**–**B** 5 inputs, **C**–**D** 10 inputs, **E**–**F** 30 inputs. *Left column*: Location of the impulse inputs (indicated by circles) approximating an exponentially decreasing continuous input function (indicated by the dashed line) and the corresponding impulse responses (solid lines). The doses (amplitudes) of the impulse inputs are equal to the product of instantaneous value of the continuous input and the time step width between the inputs, which ensures the correct total dose (area under the input curve) in the continuous limit. Thus, the doses of the impulse inputs are proportional to the time step width and, accordingly, the scales in **A**, **C**, **E** are different. *Right column*: Resulting discrete system response (solid line). For comparison, the system response to the continuous input is shown, too (dashed line). With increasing number of impulse inputs, the system response converges to the continuous limit given by Eq. (2)

times the response to input $a(t)$. This is, what is meant by the term ''linear'' tracer kinetics. Figures 3 and 4 visualize the situation described by Eq. (2) for a simple case.

It should be stressed again that the assumption of linear tracer kinetics is valid even if the investigated process is non-linear (obeying, for instance, saturation kinetics). A good example of this fact is the investigation of glucose metabolism using tracers such as ^{18}F-fluorodeoxyglucose (FDG). Figuratively speaking, the tracer probes a single point on the Michaelis-Menten curve, the point being defined by the level of the systemic glucose concentration: the total amount of glucose (systemic plus tracer) available for metabolization remains essentially unaltered by the very small amount of injected tracer.

4.3 Some useful observations

Apart from the assumption of linearity and mass conservation no further assumptions have been made up to now. Nevertheless some interesting conclusions can be drawn.

If the tracer is not irreversibly trapped in the system and the input (more precisely, the area under the input curve) is finite, c_s will tend asymptotically to zero at sufficiently late times. In this case, a direct consequence of Eq. (1) is

$$\int_0^\infty c_a(t)dt = \int_0^\infty c_v(t)dt \tag{3}$$

(equality of the areas under input and outflow curve). Moreover, one gets the following relation which follows from the assumption of linearity:

$$\int_0^\infty c_s(t)dt = p\int_0^\infty c_a(t)dt, \tag{4}$$

i.e. proportionality between the areas under the response and the input curve. The proportionality factor p – which can either be taken as being a dimensionless number or having units of ml[tissue]/ml[blood] – is known as the volume of distribution (or partition coefficient, depending on the context). It can be shown to be identical to the area under the impulse response:

$$p = \int_0^\infty R(t)dt. \tag{5}$$

Under equilibrium (''constant infusion'') conditions, the volume of distribution is simply given by the asymptotic ratio of tissue to blood concentrations:

$$p = \frac{c_s(\infty)}{c_a(\infty)} = \frac{c_s(\infty)}{c_v(\infty)}.$$

A further useful parameter is the so-called mean transit time, defined as difference of the ''centre of masses'' of input and outflow:

$$\bar{t} = \bar{t}_v - \bar{t}_a = \frac{\int_0^\infty c_v(t)t\ dt}{\int_0^\infty c_v(t)dt} - \frac{\int_0^\infty c_a(t)t\ dt}{\int_0^\infty c_a(t)dt}. \tag{6}$$

For an impulse input $c_a(t) = \delta(t)$ this reduces to $\bar{t} \equiv \bar{t}_v$. The mean transit time can be viewed as the average residence time of a single tracer molecule in the system and is related to the volume of distribution p and perfusion f by:

$$\bar{t} = \frac{p}{f}. \tag{7}$$

This becomes obvious when considering the fact, that $1/f$ can be interpreted, loosely speaking, as the time necessary for a ''displacement'' of the complete system volume. p can be visualized as the (artificial) fraction of the system volume accessible to the tracer at the same concentration as it is found in the blood space. Therefore, multiplying p by $1/f$ yields the mean transit time. (If the tissue equilibrium concentration is higher than the blood concentration, this fraction is greater than One, a situation which is more adequately described using the concept of partition coefficient for p.)

Classical tracer kinetic experiments generally deal with a situation where the system response $c_s(t)$ is not directly accessible. Instead, information is derived by measuring input c_a as well as venous outflow c_v. This, of course, becomes increasingly difficult or even impossible if one aims at assessing local behaviour *in vivo*. The problem does not lie in determination of the arterial input $c_a(t)$ – which usually can be reliably measured at some artery remote from the target region-, but in the potential inaccessibility of the local outflow $c_v(t)$.

With PET, on the other hand, the situation is somewhat reversed (and improved): PET provides no information on the outflow c_v but allows direct non-invasive assessment of the local system response $c_s(t)$. The combination of this information with a separate measurement of the arterial input c_a (which might be derivable from the PET data, too, if a large vessel or the heart is in the field of view) is the basis of the quantitative description of the system.

The presentation up to now was focused on a general outline of some basic concepts used in tracer kinetic experiments. Useful information can be derived using only these general principle. But to assess more specifically physiological and biochemical transport processes it is necessary to specify an explicit model of the tracer transport within the system. There are several possible approaches ranging from purely phenomenological or statistical approaches – principal component and factor analysis, for instance – to increasingly sophisticated descriptions of the microscopic transport steps. Especially the attempts to include the effects of finite diffusion velocities into the models have a large literature of their own, see, for instance (Bassingthwaighte et al., 1992, 1997; Bassingthwaighte and Goresky, 1984; Bassingthwaighte and Holloway, 1976; Bassingthwaighte, 1997). Quite generally, it can be said, that these complex models become important, when the time scale of the observations is very short, as can be the case in invasive experiments or in certain MRI applications. For PET, it has never been conclusively shown that there is a benefit of using these models, either in terms of the amount of information extracted from the data or in terms of the achieved systematic or statistical accuracy. This is related to the fact that the effective time resolution of tracer kinetic data measured with PET is on the order of a few seconds, at best, and typically bolus durations are of the order of 10–30 seconds. On this time scale, it has turned out that quantitative data evaluation can reliably be based on compartment models which will be discussed in some detail in the following.

4.4 Compartment models

The basic concept underlying the compartment modelling approach has already been mentioned: decomposition of

the investigated system into a (small) number of functional units (compartments), which can be assumed to be internally ''well stirred'' (no concentration differences within the compartment) and description of the transport between compartments by first order (linear) tracer kinetics. Compartments in PET investigations are usually not spatially distinct but, more often, represent the tracer in different chemical modifications. Even in situations, where there is a microscopic anatomical correlate for a compartment, for instance, the extracellular space as opposed to the intracellular space or the different intracellular functional units (e.g., mitochondria), a PET measurement is not able to differentiate these compartments spatially.

An important consequence of this situation is, that the signal measured by PET has to be interpreted as a sum over all the contributing compartments. A measurement of the tracer concentrations in the individual compartments is generally not possible. The normalisation of all tracer amounts to a common volume implies, that the total tracer concentration is simply the sum of the concentrations in all compartments:

$$c_t(t) = \sum_{n=1}^{N} c_n(t).$$

This relation would of course not be true, if the concentrations where specified with respect to separate compartmental spaces with different volumes.

Compartment models are represented mathematically by coupled inhomogeneous linear first order differential equations with constant coefficients. The coefficients are directly related to the parameters describing tracer exchange between the compartments. The validity of the assumption of constant coefficients, and, hence, transport parameters, can usually be guaranteed – or at least be justified – for the duration of a PET investigation. It should be noted, however, that this assumption is by no means self-evident in all circumstances.

In PET, simple model configurations – especially one- and two-compartment systems – are usually adequate for actual data evaluation. Investigation of the properties of more realistic (and generally more complex configurations) is nevertheless an important step in assessing the effects of model simplifications. In the following we will concentrate on demonstrating the general properties of compartment models, using the configurations which are most frequently encountered as illustrative examples.

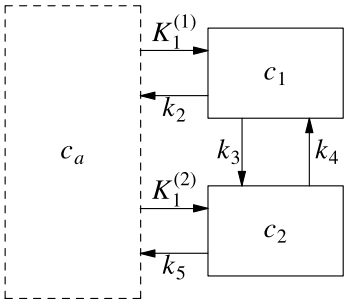

Fig. 5. The most general two-compartment model with direct tracer exchange of both compartments with the blood pool

A graphical representation of the most general two-compartment model is given in Fig. 5, which corresponds to the model equation:

$$\begin{aligned} \dot{c}_1 &= -(k_2 + k_3)c_1 + k_4 c_2 + K_1^{(1)} c_a \\ \dot{c}_2 &= k_3 c_1 - (k_4 + k_5)c_2 + K_1^{(2)} c_a, \end{aligned} \tag{8}$$

where $\dot{c}_m$ denotes the time derivative of c_m.

In this general form the model is of little practical relevance due to the direct input of the tracer from the blood space into two distinct pools: usually in PET only one compartment receives input from the blood space and further compartments describe metabolization/binding of the tracer. (On the contrary, the situation, where metabolites accumulate in the blood space and do enter the target tissue is quite common and needs separate treatment by setting up an independent model for the metabolite kinetics.)

Special cases of a given model configuration are obtained by setting certain parameters to zero. For instance, by setting $K_1^{(2)} = k_5 = 0$ one obtains the model depicted in Fig. 6 where the boxes representing the compartments are intentionally arranged differently to stress the fact that this arrangement is arbitrary and does, of course, not alter the structure of the model.

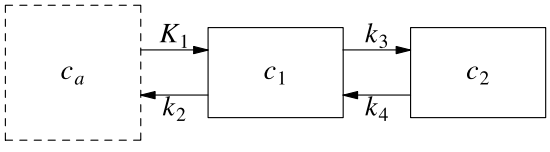

Fig. 6. The reversible two-compartment model with input into the first compartment and no direct clearance from the second compartment to the blood pool. This model is a special case of Fig. 5 obtained by setting $K_1 \equiv K_1^{(1)}$ and $K_1^{(2)} = k_5 = 0$. This special configuration occurs frequently in PET investigations, usually for describing transport of unmetabolized tracer from blood into the tissue compartment c_1 and a reversible binding or metabolization of the tracer in a separate pool c_2

In a rather general form, the compartment equations of the N-compartment model can be written as:

$$\dot{c}_m = \sum_{n=1}^{N} k_{mn} c_n + K_1^{(m)} c_a, \quad m = 1 \cdots N. \tag{9}$$

A word on notation and units: The coefficients k_{mn} in Eq. (9) form a $N \times N$ array which will be called the ''rate matrix'' $\hat{k}$. For example, in the case of Eq. (8) it is given by

$$\hat{k} = \begin{pmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{pmatrix} = \begin{pmatrix} -(k_2 + k_3) & k_4 \\ k_3 & -(k_4 + k_5) \end{pmatrix}.$$

In PET, where the number of model parameters is modest, it is more customary to use the successively enumerated rate constants instead of the k_{mn}. Moreover, for the usual situation of input only into the first compartment, the uptake rate is simply denoted by K_1. This is the notation used in Fig. 6. The rationale for this notation is, that K_1 plays a different role than the other parameters: K_1 is a ''scale'' parameter determining the overall scaling of the system response without influencing its shape. The units are [ml(blood)/min/ml(tissue)]. Its value can be understood as representing the volume of blood space whose tracer content is *completely* transferred to one millilitre of tissue per minute. The other model parameters are ''real'' rate constants which determine the shape of the tissue response curve. They all have units of inverse time [1/min]. Their values can be understood as the fraction of tracer removed per minute from the source compartment through this ''channel''.

The following list summarizes some important facts following from the well developed theory:

1. The input function (the blood space) is not one of the compartments, since it can be treated as an ''externally'' defined quantity which does not alter the impulse response of the system. This means, that the ''feedback'' from the tissue, which obviously alters the time course of the tracer concentration in blood, is not part of the model. The situation in PET, thus, is different from classical pharmacokinetics, where the blood space usually has to be included in the compartments to be modelled. The difference stems from the fact, that compartment models in PET are ''local'' in the sense that they do not describe the behaviour of the tracer in the organism as a whole, but rather only locally in the target region. Because of this fact, it is more precise to speak, for instance, of a ''two-tissue-compartment model'' when referring to the model shown in Fig. 6.
2. The impulse response $r_m(t)$ of compartment number m in a given model is a sum of exponentials with differing amplitudes a_{mn} and decay constants λ_n:
$$r_m(t) = \sum_{n=1}^{N} a_{mn} e^{-\lambda_n t}.$$
3. The number of exponentials N is identical to the number of tissue compartments in the model.
4. The different exponentials in the impulse responses $r_m(t)$ of the single compartments usually have different amplitudes, some of which might be zero or negative. In contrast, the decay constants are identical across all compartments. Therefore, the impulse response $R(t)$ of the complete model – which is the sum over the $r_m(t)$ – is also given by a sum of N exponentials with the same decay constants whereas the amplitudes are obtained from the sum over the partial amplitudes in the different compartments:
$$\begin{aligned} R(t) &= \sum_{m=1}^{N} r_m(t) = \sum_{m=1}^{N} \sum_{n=1}^{N} a_{mn} e^{-\lambda_n t} \\ &= \sum_{n=1}^{N} \left[\sum_{m=1}^{N} a_{mn} \right] e^{-\lambda_n t} = \sum_{n=1}^{N} a_n e^{-\lambda_n t}. \end{aligned}$$
5. It is a rather common error – which still occurs by and now in the literature – to identify the separate exponentials (the a_n and λ_n) with individual compartments. According to what has been said above, this is usually wrong: in general, *all* exponentials contribute with different amplitudes to each compartment's impulse response.
6. The individual compartments are not accessible to the PET investigation. Accessible is only the impulse response $R(t)$ of the complete system which determines the behaviour of the observed tracer concentration. Therefore, the PET measurement does allow to identify the number of kinetic compartments (which equals the number of exponentials present in the impulse response), but not the detailed structure of the model (which would require determination of all the $r_m(t)$ separately). Consequently, additional information is necessary to fix the structure of the model a priori in order to allow derivation of the interesting parameters. This fact is one aspect of the problem of model identifiability.
7. The amplitudes a_{mn} and the decay constants λ_n are well defined functions of the coefficients k_{mn} and $K_1^{(n)}$ occurring in Eq. (9). More specifically, the decay

constants are identical to the eigenvalues of the rate matrix $\hat{k}$, whereas the amplitudes are directly related to the eigenvectors of $\hat{k}$ and the uptake parameters $K_1^{(n)}$. Both sets of $(N^2 + N)$ parameters – either the a_{mn}, λ_n or the k_{mn}, $K_1^{(n)}$ – yield a complete description of the general N-compartment model.

8. The maximum number of parameters ''visible'' to the PET measurement is only $2 \cdot N$ (the number of amplitudes a_n and decay constants λ_n). As long as the total number of non-zero parameters in the compartment model does not exceed this number, all model parameters can be derived – at least in principle – from the PET measurement. For most models of practical importance, e.g., the reversible 2-tissue compartment model in Fig. 6, this assumption is fulfilled. But in the general case, the number of transport parameters of the model exceeds this number. For instance, in Fig. 5 there are six transport parameters but only two exponentials – with four parameters – in the impulse response. Derivation of all the model parameters from the measurement is not be possible in this case. This situation is an example of complete ''parameter unidentifiability'': several sets of model parameters result in exactly the same tissue response.

Practically, parameter identifiability more often is a problem for purely numerical reasons, when completely different sets of model parameters result in only minor changes of the tissue response function and the statistical accuracy of the measured data is not sufficient to discriminate the alternatives. Inclusion of some kind of statistical error analysis into the parameter estimation procedure should, therefore, be considered mandatory.

Determination of the impulse response of a compartment model, i.e. determination of the functional dependence of the a_{mn} and λ_n on the transport constants of the model, is equivalent to determination of the eigenvalues and eigenvectors of the rate matrix $\hat{k}$, both of which can be derived – sometimes analytically, otherwise numerically – by standard mathematical techniques. For the two-compartment models shown in Figs. 5 and 6, examples of typical response curves are given in Fig. 7.

4.4.1 Example: the reversible one-tissue-compartment model

To give some more insight into the mathematics involved one of the simplest compartment models used in PET, the reversible one-tissue-compartment model, is discussed in a bit more detail.

The structure of the reversible one-tissue-compartment model is given in Fig. 8.

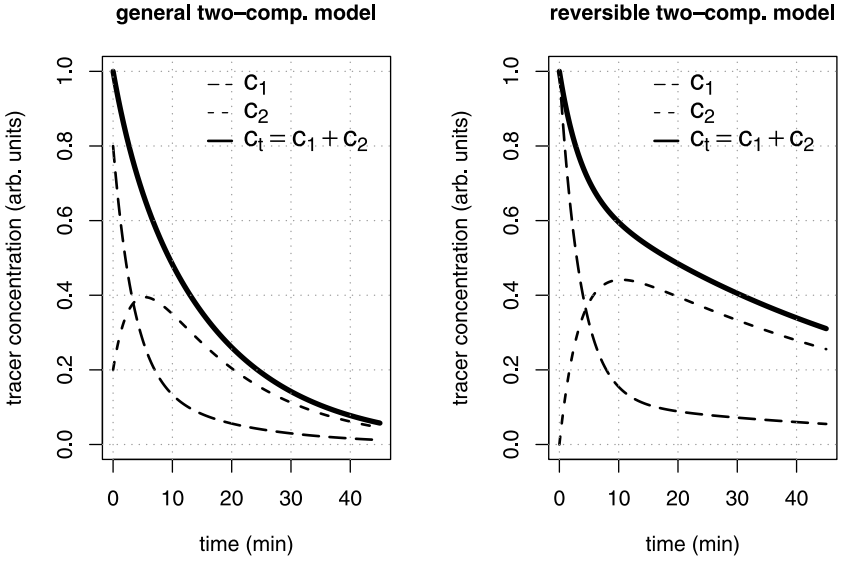

Fig. 7. Example of unit impulse response curves for the models shown in Figs. 5 and 6. On the left hand side, the following values for the parameters were used (units: K_1 [ml/min/ml], otherwise [1/min]): $K_1^{(1)} = 0.8$, $k_2 = 0.1$, $k_3 = 0.15$, $k_4 = 0.05$, $K_1^{(2)} = 0.2$, $k_5 = 0.05$. On the right hand side, $K_1^{(2)}$ and k_5 were set to zero and $K_1 = 1$ was used (the sum of the $K_1^{(1)}$ and $K_1^{(2)}$ values used on the left hand side). At $t = 0$, the response curves start with values numerically equal to the uptake constants ($K_1^{(1)}$ and $K_1^{(2)}$, respectively). Note the similar but not identical shape of the response curve (long dashed) in compartment c_1 for both cases (which is influenced by $K_1^{(2)}$ and k_5 only via a modified backflow from c_2). Note also the different behaviour of c_2 (short dashed), reflecting the different kind of input into the system in both cases

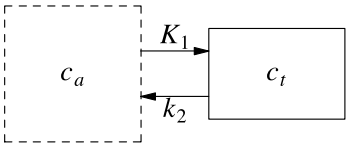

Fig. 8. The reversible one-tissue-compartment model. c_a: arterial input, c_t: tissue concentration

The corresponding model equation

$$\dot{c}_t = K_1 c_a - k_2 c_t \tag{10}$$

has the solution

$$c_t(t) = c_a \otimes K_1 e^{-k_2 t}, \tag{11}$$

i.e. the unit impulse response is

$$R(t) = K_1 e^{-k_2 t}.$$

As can be seen by comparison with the general formulas presented earlier, the one-compartment model trivially allows direct interpretation of amplitude and decay constant of the impulse response: they are identical with the model parameters.

This monoexponential impulse response has already been used in Fig. 4 which might be consulted again to appreciate how the tissue response of this model typically looks like for some kind of finite bolus injection: the washout at late times – or generally: the approach to equilibrium – is controlled by the ''shape parameter'' k_2, whereas the overall scaling is controlled by K_1 (as is obvious from Eq. (11)).

Within the framework of compartmental analysis a definition of extraction is used which essentially coincides with the definition of unidirectional extraction used in the context of compartment-free approaches. The unidirectional extraction is given by

$$E_u = \frac{K_1}{f}, \tag{12}$$

where f designates the regional perfusion. In comparison to the determination of the extraction fraction within the context of compartment-free analysis, Eq. (12) has the advantage that the determination does not necessitate extremely short bolus durations, because the tracer backflow from tissue is accounted for by the corresponding transport constants of the model. As can be seen, PET is readily able to yield information concerning the extraction fraction – across the blood brain barrier, for instance – if the tissue perfusion is already known, possibly from a separate PET investigation with a freely diffusible tracer.

The reversible one-compartment model includes the special case of freely diffusible tracers by identifying K_1 with tissue perfusion, $K_1 = f$, so that the unidirectional extraction equals 100%. Realising that at equilibrium the tissue concentration no longer varies, one concludes from Eq. (10),

$$\dot{c}_t \equiv 0 = K_1 c_a - k_2 c_t \Rightarrow \frac{K_1}{k_2} = \frac{c_t}{c_a} = p,$$

that is, the ratio of K_1 to k_2 is identical to the volume of distribution p. Note, that the assumption of free diffusion implies that p is a constant, hence in this case k_2 varies in proportion to K_1. Thus, for freely diffusible tracers, the distinction between ''shape'' and ''scaling'' parameters does not make sense.

Equation (11) allows derivation of the model parameters from the experimental data – the measurement of both, $c_a(t)$ and $c_s(t)$ – by using standard nonlinear least squares techniques, cf., for instance, Chap. 15 in (Press et al., 1992). Since $c_a(t)$, as a measured quantity, is not known analytically, the convolution in Eq. (11) has generally to be computed numerically. (A suitable parametrisation of c_a might allow to perform the convolution analytically, but this does not present any significant advantage.) Since k_2 enters the equation in the exponent of the exponential function, the fit is nonlinear with respect to this parameter and iterative algorithms are needed. For a small number of data sets the resulting computational overhead is irrelevant and, given an adequate level of statistical accuracy, convergence problems of the iterative search are usually not an issue.

If one aims at utilizing the full spatial resolution of the PET data, the parameter estimation can be performed separately for each voxel of the volumetric data set. In this case, one can generate so-called ''parametric images'' of the model parameters (here: K_1 and k_2) by mapping the computed parameters to the correct position within the tomographic data set. This allows both, an easy visual evaluation of the regional variation of the parameters as well as direct access to the quantitative values in selected Regions of Interest (ROIs). Because 3D data sets in PET typically contain about 10^6 voxels, direct utilisation of Eq. (11) is not feasible. Luckily, there are ways to circumvent the problem and to use linear least squares minimization techniques instead, which can be applied to several model configurations important in PET (Blomqvist, 1984; Logan et al., 1990; Patlak et al., 1983; van den Hoff et al., 1993). These techniques either use the asymptotic behaviour at later times, when a steady state is reached (Logan et al., 1990; Patlak et al., 1983) or derive exact integral equations from the model equations which avoid the necessity of using the analytic solutions of the equations (Blomqvist, 1984; van den Hoff et al., 1993).

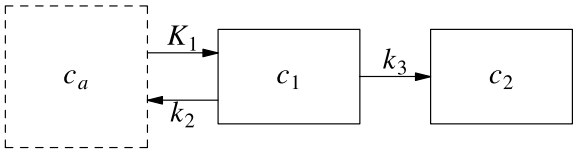

Fig. 9. The irreversible two-compartment model. This model is a special case of Fig. 6, obtained by setting $k_4 = 0$. This configuration is often adequate to describe transport of free tracer from blood to tissue and its subsequent irreversible metabolization/trapping in the tissue. Notably, this description is valid for the tracer [^{18}F]-2-fluoro-2-deoxy-D-glucose (FDG). In this case, K_1, k_2 describe bidirectional transport of the tracer across the cell membrane and k_3 corresponds to the hexokinase catalysed phosphorylation to FDG-6-phosphate which – for this deoxyglucose – is irreversible in most tissues (contrary to the behaviour of glucose)

4.4.2 Example: the irreversible two-tissue-compartment model

As a second example, the important case of a two-compartment model with irreversible binding is presented in Fig. 9.

The system response of this model is

$$c_t(t) = c_a \otimes \frac{K_1}{k_2 + k_3}(k_3 + k_2 e^{-(k_2+k_3)t}). \tag{13}$$

As one can see, one of the two exponentials of the impulse response is degenerated to a constant ($\lambda_1 = 0$). This is a consequence of the fact, that part of the tracer (the metabolized fraction) is irreversibly trapped in the system. The trapping rate is given by

$$K_m = \frac{K_1 k_3}{k_2 + k_3},$$

which can easily be interpreted as the product of initial uptake rate, K_1, and the branching ratio $k_3/(k_2 + k_3)$ which is the fraction of the free tracer in tissue actually undergoing metabolization. Determination of this trapping rate is possible without knowledge of the individual rate constants (Patlak et al., 1983). This quantity is of special importance because, in the case of FDG, it allows to determine the regional metabolic rate of glucose by multiplying K_m with the serum glucose concentration. (Corrections have to be made to account for differences in the detailed behaviour of FDG and glucose.)

4.5 Influence of blood space

The discussion up to now did not account for the fact that during a PET scan there will always be a certain fraction of the blood space – especially the capillary blood space – included which will, thus, contribute to the detected radioactive decays. This leads to two consequences.

For one, the tracer concentrations measured in PET always refer to the total volume (measurement volume), a certain fraction of which is occupied by the capillary space. Therefore, even after complete elimination of the tracer from the blood the measured tracer concentration cannot simply be identified with the "true" tracer concentration (the amount of tracer per unit volume of tissue).

Furthermore, in the early phase of an investigation, the tracer is dominantly located in the capillary space and the concentration measured by PET is in this phase essentially controlled by the intravasal contribution.

In order to account for both effects in the compartment model, the measured tracer concentration in the system, c_s, has to be related to the true tissue concentration c_t. This is achieved by noting that, due to the partial inclusion of the blood space, c_s (as detected by PET) is given by:

$$c_s(t) = (1 - fbv)c_t(t) + fbv\, c_a(t). \tag{14}$$

fbv specifies the relative fraction of the blood space within a unit of the measurement volume (*Fractional Blood Volume*). Eq. (14) is simply the correctly volume weighted average of the arterial input function and the tissue response function: this weighting reflects the fact that the concentrations themselves are of course not additive. (Here, they really refer to different spaces, contrary to the situation we encountered when summing over the different compartments to obtain the total tissue concentration.)

For bolus injections the shape of $c_a(t)$ and $c_t(t)$ usually is sufficiently different, and the additional parameter *fbv* can be derived together with the transport constants K_1 and k_2 from the measured data. Neglecting the *fbv* influence can lead to massively biased parameter estimates, especially concerning the uptake parameter K_1, therefore inclusion of this parameter in the operational equation usually is mandatory. Moreover, knowledge of the regional *fbv* can be a valuable information in itself (e.g., the frequently elevated *fbv* values in solid tumours).

Nevertheless, the *fbv* parameter can usually not be interpreted too literally: it comprises not only true blood space, but also that fraction of tissue space, which behaves kinetically indistinguishable from the blood space (at least at the level of temporal resolution realized in the measurement). This might be true, for instance, for the extracellular tissue space. *fbv* should therefore be viewed as yielding that fraction of the volume which exchanges very rapidly with the blood (the volume fraction which lies "outside" of the first significant diffusion barrier).

4.6 Influence of metabolites

In the above equations, $c_a(t)$ denoted the true time dependent arterial tracer concentration. If the injected labelled compound is a substrate of systemic metabolism, the measured arterial concentration – which only yields the concentration

of the radioactive label itself – has to be corrected for the accumulation of labelled metabolites in the blood, and the compartment modelling has to be modified accordingly.

The time dependency of the metabolite accumulation has therefore to be determined, too. Based on such a measurement, the true arterial concentration $c_a(t)$ is derived from the measured whole blood concentration $c_b(t)$ by multiplication with the instantaneous *Parent Fraction* (i.e. the time dependent unmetabolized fraction of the substance) $P(t)$:

$$c_a(t) = P(t) \cdot c_b(t).$$

The metabolite corrected curve is than used as the input function in the compartment model and Eq. (14) is replaced by

$$c_s(t) = (1 - fbv)c_t(t) + fbv\, c_b(t). \quad (15)$$

Other corrections of the model equations which might become necessary include the effects of shape modifications and timing errors (dispersion and delay) introduced by measuring the input function at a peripheral site remote from the target organ.

By including all the necessary corrections in the model equations, the number of parameters which have to be determined from the data is increased. This can deteriorate the statistical accuracy but avoids bias in the estimate of the derived parameters.

5 Conclusion

By its very nature, PET is a quantitative technique. This fact can be exploited in various ways. In research applications every effort should be undertaken "to make the most of the data". Compartment modelling of dynamically acquired PET data generally is the method of choice to achieve this goal. The mathematical techniques involved are fairly standard but practical problems might still be considerable. Possible pitfalls are numerous, examples of which can still be found in the journals. These include attempts to use overdetermined or simply inadequate models, misinterpretation of the derived parameters and missing sensitivity or statistical error analysis resulting in over-interpretation of the derived results. These pitfalls can be avoided by a careful analysis of the precise conditions under which the experiment is performed. Special care has to be taken to choose an adequate workable model and to include all the necessary corrections required for obtaining unbiased parameter estimates.

If these caveats are observed, valuable quantitative information concerning a wide variety of physiological and biochemical transport phenomena is obtainable by application of the discussed methods to PET data evaluation. This assessment has already been proven to be correct in clinical and pre-clinical research and, more recently, in drug development and evaluation. The future might well see a similar development in other fields – for instance in the food sciences.

References

Bassingthwaighte JB (1977) Physiology and theory of tracer washout technique for the estimation of myocardial blood flow estimation from tracer washout. Prog Cardiovasc Dis 20: 165–189

Bassingthwaighte JB, Goresky CA (1984) Modeling in the analysis of solute and water exchange in the microvasculature. In: Handbook of physiology, sect. 2. The cardiovascular system, vol. IV. The microcirculation, chap. 13. American Physiological Society, Bethesda, pp 549–626

Bassingthwaighte JB, Holloway GA Jr (1976) Estimation of blood flow with radioactive tracers. Semin Nucl Med 6: 141–161

Bassingthwaighte JB, Chan IS, Wang CY (1992) Computationally efficient algorithms for convection-permeation-diffusion models for blood-tissue exchange. Ann Biomed Eng 20: 687–725

Bassingthwaighte JB, Winkler B, King RB (1997) Potassium and thallium uptake in dog myocardium. J Nucl Med 38: 264–274

Blomqvist G (1984) On the construction of functional maps in positron emission tomography. J Cereb Blood Flow Metab 4: 629–632

Carson RE (2003) Tracer kinetic modeling in PET. In: Valk PE, Bailey DL, Townsend DW, Maisey MN (eds) Positron emission tomography. Springer, London, pp 147–179

Carson RE, Cunningham V, Gunn RN, van den Hoff J, Knudsen GM, Lammertsma AA, Leenders KL, Maguire RP, Müller-Schauenburg W (2003) PET pharmacokinetic course. In: Maguire RP, Leenders KL (eds) PET Pharmacokinetic course manual, University of Groningen, Groningen, The Netherlands and McGill University, Canada Montreal

Jones T (2003) Historical development of functional *in vivo* studies using positron-emitting tracers. In: Valk PE, Bailey DL, Townsend DW, Maisey MN (eds) Positron emission tomography. Springer, London, pp 3–40

Kety SS, Schmidt CF (1948) The nitrous oxide method for the quantitative determination of cerebral blood flow in man: theory, procedure and normal values. J Clin Invest 27: 476–483

Lassen N, Perl W (1979) Tracer kinetic methods in medical physiology. Raven Press, New York

Logan J, Fowler JS, Volkow ND, Wolf AP, Dewey SL, Schlyer DJ, MacGregor RR, Hitzemann R, Bendriem B, Gatley SJ, et al (1990) Graphical analysis of reversible radioligand binding from time-activity measurements applied to [N-11C-methyl]-(−)-cocaine PET studies in human subjects. J Cereb Blood Flow Metab 10: 740–747

Patlak CS, Blasberg RG, Fenstermacher JD (1983) Graphical evaluation of blood-to-brain transfer constants from multiple-time uptake data. J Cereb Blood Flow Metab 3: 1–7

Press WH, Teukolsky SA, Vetterling WT, Flannery BP (1992) Numerical recipes in C. Cambridge University Press, Cambridge

van den Hoff J, Burchert W, Müller-Schauenburg W, Meyer GJ, Hundeshagen H (1993) Accurate local blood flow measurements with dynamic PET: fast determination of input function delay and dispersion by multilinear minimization. J Nucl Med 34: 1770–1777

Willemsen ATM, van den Hoff J (2002) Fundamentals of quantitative PET data analysis. Curr Pharm Des 8: 1513–1526

Author's address: Joerg van den Hoff, PhD, PET Center, Institute of Bioinorganic and Radiopharmaceutical Chemistry, Research Center Rossendorf, P.O. Box 510119, 01314 Dresden, Germany,
E-mail: j.van_den_hoff@fz-rossendorf.de