Drug Metabolites: General Features and Most Applicable Analytical Methods of Studies

Pedro H. J. Batista¹, José C. Quilles Jr²*
¹São Carlos Institute of Chemistry (IQSC), University of São Paulo (USP), Brazil
²Department of Cell and Molecular Biology, Ribeirão Preto Medical School (FMRP), University of São Paulo (USP), Brazil
Email: quilles@usp.br

Abstract: Drug discovery and development is a multi and interdisciplinary process that includes chemistry and biology understanding. Though, drug discovery is an expensive and time-consuming process, promoting the lack of new drugs approval in recent years. In general, most of the drugs are established based on a macromolecular target, which is vital to disease progression. However, in many cases, the mechanism of drug-macromolecular target interaction is complex and not understood. All this missed information is strongly dependent on the physicalal-chemical stability and behavior of the molecule inside the cell, which is correlated to its metabolism resistance and entrance into the cellular membrane, respectively. Thus, bioanalysis often provides enough data about these molecular characteristics and helps to figure out new information about these subjects. In this context, the study of the metabolism of parental compound plays a key role in bioanalysis and, consequently in drug discovery. For a long time, the study of metabolism represented a huge challenge in medicinal chemistry, but with the technological advancements, many powerful techniques were developed and, currently, the metabolomics fields are essential steps in the process of discovering a new drug. Herein, we briefly discuss the biological aspects of the drug metabolism, focusing on the most used analytical tools to better understand the metabolite generation, and consequently, their chemical and biological characteristics.

Keywords: metabolites, drug metabolism, bioanalysis, chromatography

1. Introduction

In the 19 century, the medicinal chemistry emerged with the principle of chemical modifications of small molecules, as well as their validation in biological systems. In addition, researches into biological receptors and enzyme specificity also completed the multidisciplinarity of the medicinal chemistry beginning[1]. As main idea, the medicinal chemistry comprises the creation and modification of chemical molecules which may exhibit biological effects on the treatment, prevention and cure of diseases[2]. Since then, the research and development of new drugs count on the medicinal chemistry contribution, which uses techniques and knowledge of distinct basic areas, such as chemistry and biology[1].

The drug discovery and development begins due to the need of new appropriate bioactive substances to treat or refrain a specific disease, whether new or existing, but still with ineffective treatments[3]. In this way, medicinal chemistry also embodies new technologies for optimizing the physical-chemical properties of a new molecular entity (NME) in order to overcome several issues about drug development. Just as an example, the drug delivery systems are very important in the medicinal chemistry studies, once they may promote a specific drug release, besides decreasing the side effects[4-5]. Furthermore, during the discovery and development of a new drug, the identification and validation processes of a molecular target are essential steps of this course. A “target” is considered as a vital biomolecule for the establishment and development of a disease, which NMEs are designed to interact with such molecule, aiming beneficial effects, with the possibility of becoming a drug[6]. Most of the drugs have as target essential biomolecules or biological pathways, which their interruption may promote the disease regression. NME reaches its objective as a drug when it interacts with the macromolecular target, generally inside the cell, and regulates its function. In many cases, the mechanism of interaction between the ligand and its target is unclear, which can lead to the unspecific interactions, promoting the off-target effect, a mechanism usually known to cause the side effects[7]. One of the most promising strategies to face this problem is to use a known drug for the new treatments, a strategy known as drug repurposing. This process generates an additional charge to a
A drug already approved for some specific disease, that may be also applied to a treatment of another, based on the target similarity\(^9\). Usually, such strategy provides lower costs and time for the drug approval by the regulatory agency, once the drug has its safety in humans already proved\(^9\). Also, advantages over the discovery of new drugs include the possibility of not developing steps, such as synthesis, manufacture, safety assays and pharmacokinetical analysis\(^9\).

In general, drug discovery is an expensive, time-consuming and a high-risk associated process, that involves many steps from the target identification until the drug approval (Fig 1). The costs of drug development associated with the total time spent during this process have decreased the total number of new drugs’ approvals\(^10\). Although potency, efficacy and selectivity of the drugs are significant attributes that make an NME interesting during the drug discovery process, the pharmacokinetics features are indispensable for a complete drug validation\(^11\). A chemical cannot be considered as a commercial drug, even though it presents high efficacy and selectivity, unless it grants considerable absorption with relevant distribution in the organism, besides presenting stable chemical species derived from its metabolism, as well as, no body retention and great elimination\(^10\). Based on that, bioanalysis plays a key role in medicinal chemistry, promoting crucial understanding of the physical-chemical properties and drug metabolism before approving any substance for human tests. Without these essential informations, about 40% of NMEs were canceled due to absorption, distribution, metabolism and excretion (ADME) problems in the past\(^13\). Currently, it is known the importance of considering and analyzing minutely these drug properties\(^14\-15\).

**Pharmacokinetics parameters of NMEs correlate the potency and action duration, which are directly linked to bioavailability and half-life\(^16\).** The ADME analysis aims to understand the NME action, bearing in mind the drug pharmacokinetics and pharmacodynamics parameters, which include: (1-absorption) NME ability to pass through biological barriers, such as intestinal membrane, nasal lining or skin; (2-distribution) how the NME is distributed and accumulated around the organism, identifying its most common accumulation regions; (3-metabolism) the NME breakdown by the body, in which the main issue is the drug-drug interaction and the effects caused by the generated molecules (called as metabolites); and (4-excretion) the process and rate of the NME elimination by the organism, based on its concentration decrease in the body\(^17\-18\) which might also affect drug metabolism and pharmacokinetics. In addition, the elderly population will develop multiple diseases and, consequently, often has to take several drugs. As the hepatic first-pass effect of highly cleared drugs could be reduced (due to decreases in liver mass and perfusion. Besides the drug metabolism, drug-drug interaction (DDI) is another characteristic to be explored in the discovery of NMEs\(^19\). These studies are based on the identification of probable clinical issues that may happen due to safety and altered activity, as a result of the interaction between the drugs\(^20\). Herein, our focus is the metabolite studies, centering on the chemistry features and tools used for the main metabolites studies as show Table 1. Besides discussing the chemical parameters, it is indispensable to have a general idea about how and what happens to a drug when it is administered into an organism. A general overview of the pathways of drug internalization into a cell is briefly covered in the next topic.
Table 1. Applications of analytical techniques in bioanalysis

<table>
<thead>
<tr>
<th>Sample</th>
<th>Application</th>
<th>Sample Preparation</th>
<th>Analytical Technique</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Volatiles compounds in cellular medium</td>
<td>In vitro MDMA metabolism</td>
<td>PPT/HS-SPME</td>
<td>CG-MS</td>
<td>43</td>
</tr>
<tr>
<td>Annona crassiflora extracts</td>
<td>Larvicidal Activity in Natural Products field</td>
<td>LLE</td>
<td>LC-DAD-MS</td>
<td>46</td>
</tr>
<tr>
<td>Microbial colonies</td>
<td>Method development</td>
<td>None</td>
<td>On-line MS (ESI)</td>
<td>47</td>
</tr>
<tr>
<td>Cordyceps militaris metabolites</td>
<td>Investigation of interaction between natural products and treated cancer cells</td>
<td>PPT/LLE</td>
<td>LC-MS/MS</td>
<td>51</td>
</tr>
<tr>
<td>Penicillium nordicum metabolites</td>
<td>Food contamination</td>
<td>LLE</td>
<td>On-line MS (Isotope labeling)</td>
<td>52</td>
</tr>
<tr>
<td>Microbial metabolites</td>
<td>Microbial metabolite exchange factors</td>
<td>None</td>
<td>LC-MS (ESI)</td>
<td>53</td>
</tr>
<tr>
<td>Human serum</td>
<td>Biomarkers in a pancreatic cancer diagnosis</td>
<td>PPT</td>
<td>LC-MS (ESI)</td>
<td>54</td>
</tr>
<tr>
<td>Ayahuasca extracts</td>
<td>Neuroprotective activity from Natural Products (Ayahuasca decoction)</td>
<td>LLE</td>
<td>LC-MS (ESI)</td>
<td>55</td>
</tr>
<tr>
<td>Cycloartanes</td>
<td>Putative annotations in metabolomics studies</td>
<td>LLE</td>
<td>LC-MS (APCI)</td>
<td>56</td>
</tr>
<tr>
<td>Human serum</td>
<td>Biomarkers for diabetes mellitus</td>
<td>PPT</td>
<td>LC-MS (ESI)</td>
<td>37</td>
</tr>
<tr>
<td>Fermented soybean products</td>
<td>Food analysis</td>
<td>None</td>
<td>CG-MS/UHPLC-MS</td>
<td>57</td>
</tr>
<tr>
<td>Animal Fat</td>
<td>Food analysis</td>
<td>Derivatization/Extraction</td>
<td>LC-DAD</td>
<td>58</td>
</tr>
<tr>
<td>Human plasma</td>
<td>Forensic analysis</td>
<td>LLE</td>
<td>UHPLC-MS</td>
<td>68</td>
</tr>
<tr>
<td>Rat plasma</td>
<td>Metabolomics analysis</td>
<td>LLE</td>
<td>UHPLC-MS/MS</td>
<td>69</td>
</tr>
<tr>
<td>Human plasma</td>
<td>Drug metabolism (Pharmacokinetics studies of Odanacatib)</td>
<td>LLE</td>
<td>LC-MS/MS</td>
<td>63</td>
</tr>
<tr>
<td>Bovine serum albumin</td>
<td>Proteome method development</td>
<td>PPT</td>
<td>LC-MS</td>
<td>72</td>
</tr>
<tr>
<td>Human serum</td>
<td>Vitamin A quantification</td>
<td>PPT</td>
<td>HPLC-UV</td>
<td>73</td>
</tr>
<tr>
<td>Cellular medium</td>
<td>Anticancer drugs quantification</td>
<td>PPT</td>
<td>LC-MS/MS</td>
<td>74</td>
</tr>
<tr>
<td>Human plasma</td>
<td>Drug metabolism (Clobazam quantification)</td>
<td>PPT</td>
<td>LC-MS/MS</td>
<td>75</td>
</tr>
<tr>
<td>Cerebrospinal fluids</td>
<td>Analysis of nucleotides</td>
<td>SPE</td>
<td>LC-DAD</td>
<td>76</td>
</tr>
<tr>
<td>Peanuts oil</td>
<td>Method development</td>
<td>SPE</td>
<td>HPLC-UV</td>
<td>81</td>
</tr>
<tr>
<td>Lipids</td>
<td>Lipid biosynthesis</td>
<td>SPE</td>
<td>LC-MS/MS</td>
<td>82</td>
</tr>
<tr>
<td>Marine and freshwater</td>
<td>DOM isolation and characterization</td>
<td>SPE</td>
<td>On-line MS</td>
<td>83</td>
</tr>
</tbody>
</table>

2. General cellular pathways for the molecule absorption

Among the usual pathways used for drug administration, oral delivery is the most common for many types of available drugs. Also, an interesting strategy during drug discovery is the creation of permeable-intestine drugs[21]. All these features are determined by different steps during the drug discovery process, in which the phase II is crucial to the pharmacokinetics parameters. Some of the principles of this phase is to identify the metabolites and promote their characterization based on their biological effects, also determining the bioavailability and accumulation of the drug[22]. However, for studying these parameters, we should keep in mind the types of cellular membrane transport and molecules uptake by the cells. In general, when an NME gets the intestinal epithelium, many are the possible transportation processes through the cell membrane (Fig 2). Passive diffusion is the main and simplest method of molecules uptake, which is promoted by the concentration gradient. The drug transport by passive diffusion occurs as by the cell junctions as by the cytoplasm transport, depending on the chemical characteristic of the compound. While hydrophobic molecules are able to cross the plasm membrane with no difficulty, hydrophilic species are significantly impermeable to the cell membrane, which are mainly transported into the cell interior through the cellular junctions[21].

Besides the easy drug uptake by passive diffusion, some chemical species can get into the cell only by the action of specific receptors, usually present on the cell membrane. This transport mechanism is known as carrier-mediated transport, once the molecule entrance into the cell, as for amino-acids, charged molecules and ions, is completely depending on specific transporters[21]. Calcium, an important ion for a functional cell system which its early accumulation inside the cells promotes the bone mass maintenance, has its permeability made possible only by the action of a specific carrier which
structurally composes the cell membrane\textsuperscript{[23]}. Conversely to the passive transport, the phenomenon by some transporters is a pathway that requires cellular energy, called active transport\textsuperscript{[21]}. On the other hand, some molecules have their transportation induced by their own load, based on the homeostasis (or efflux) from the extra to intracellular environment, and the other way around\textsuperscript{[24]}. Cholesterol, a very known lipid constituent of the mammalian cell membrane, also plays an essential role in the lipids assemble and function, making its concentration control vital to life development\textsuperscript{[25]}. The cholesterol amount inside the cell is controlled by the efflux of its transport between the distinct environments, and its transport has been attracted attention for the cancer research, once its levels are crucial during the cancer development\textsuperscript{[26]} thereby altering the cellular equilibrium. This review addresses the different mechanisms of cholesterol efflux from the cell and highlights their role and regulation in context to tumor development. There are four different routes by which cholesterol can be effluxed from the cell namely, 1. Besides all the cellular pathways for molecules uptake, some physical-chemical and environmental factors are also responsible for modifying their entrance into the cell. Drug transportation through the cell membrane may be affected by its interaction with another drug molecule, or even with some types of food, which may compromise the efficacy and the treatment\textsuperscript{[27]}. However, our focus herein is on the metabolite studies and chemical tools for their identification, and the next sections are dedicated to these chemical parameters.

3. Metabolism and metabolites

Sub products generated into the intracellular environment from a parental drug metabolism play essential roles in the new drug candidates success, once their chemical and biological stability are crucial to determinate the safety and efficacy of them\textsuperscript{[28]}. Due to that, systematic drug metabolism study has become a crucial part of drug development, clinical therapy and mechanism of drug-induced toxicity\textsuperscript{[29]}.  

Drug metabolism, also named biotransformation, is defined as a biochemical process that one substance is converted in other chemical species, usually more hydrophilic, to facilitate its elimination of the organism\textsuperscript{[30]}. Metabolism reactions can be divided into two phases: in phase I, wherein the predominant chemical processes are redox and reduce reactions, hydrolysis among others, a huge range of metabolic enzymes are also involved, which cytochrome P450 is indispensable\textsuperscript{[31]}. In phase II, some chemical conjugations, such as glucoronidation, sulfonation and acetylation are more frequent. Although this terminology may give us the idea that phase II occurs only after phase I, it is not necessarily correct and there is no rule for that\textsuperscript{[32]}. After one or both phases, the metabolized molecule becomes into a more hydrophilic specie or also into two or more chemical species, to facilitate their elimination during the cell metabolism. Mephedrone is a synthetic chemical molecule that is able to inhibit dopamine, norepinephrine and serotonin reuptake by blocking their respective receptors\textsuperscript{[33]}. In Figure 3, some of the main metabolic reactions are presented considering the mephedrone as the parental compound. These species are called metabolites, which are physiologically inactive in many cases. Also, some
papers have already presented many other molecules as possible metabolites, which turns the mephedrone a chemical molecule with more hydrophilic characteristics\[^{34-35}\]. In addition, some metabolites may have beneficial biological properties or even, be toxic to the cells. Other metabolism consequences of parental compounds include rapid drug elimination and some drug-drug interactions\[^{30}\], which both may compromise the drug activity.

Therefore, the metabolic identification of new drug candidates is an essential feature to select the compounds with desirable pharmacokinetic parameters. Also, such information allows the medicinal chemists to promote chemical modification in the original structure and modulate their reactivity and stability\[^{29}\]. Based on the metabolites studies, Obach R.S. has divided them into four categories: 1) metabolites that contribute to the majority activity; 2) metabolites that also contribute to the parental compound; 3) metabolites that have activity but low contribution to in vivo effect and 4) metabolites that have activity on alternative pharmacological targets\[^{36}\]. The comprehension of all these types of metabolites is the main of the metabolomics studies, aiming for a safe and potential drug.

Figure 3. Putative metabolites from the mephedrone metabolism based on the main metabolic pathways, according to the phases I and II. The principal chemical modifications on the mephedrone structure improve its hydrophilicity, as well as, facilitate body elimination. Such metabolites were proposed as the unique products from each pathway, only for representing the main chemical modifications resulting from some of the metabolism reactions.

4. Metabolomics studies

One of the most powerful techniques for screening these metabolites is the metabolomics, which includes studies at target and non-target levels\[^{37}\]. Metabolomics can be defined as a qualitative and/or quantitative analysis of metabolites generated by the chemical entity metabolism in an organism or biological sample. The first relates about a systematic study was published by J.K. Nicholson, in 1999, which describes the concept of metabolomics, the multivariate statistical analysis of biological NMR spectroscopic data and the metabolic response in living systems\[^{38-39}\].

Metabolomics analysis also enables to describe endogenous compounds (natural metabolites, such as steroid hormones and bile acids\[^{40}\]) in a biological matrix, as well as the changes of endogenous compounds before and after stimulations or disorders promoted by any physiological or pathological damages in a sample\[^{38}\]. The metabolomics studies can be applied in toxicity evaluations or even in biomarkers identification. Using NMR-based non-target metabolomics, it was possible to evaluate the effects of fungicide tebuconazole on the earthworm *Eisenia fetida* metabolism, as well as its toxic potential. Based on that, the metabolite disorders were identified from an increment of endogenous metabolites levels, such as adenosine monophosphate, trimethylamine and methylguanine in relation to the decrease of other, as creatine, fumaric acid and urinary acid\[^{41}\]. These results about the tebuconazole metabolic route were crucial to provide several essential informations in terms of mRNA expression, enzymatic activity and toxicity, in addition to the metabolomics
profile. The toxicological level of a Tibetan ethnomedicine called Renqingchangjue in rat’s serum was also determined by metabolomics approaches. Statistical analysis of the $^1$H NMR metabolomics profile revealed 28 potential biomarkers and relevant metabolic pathways induced by the Tibetan prescription, such as disturbances in energy metabolism and amino acids metabolism.\(^{[42]}\) Metabolomics assays were also applied to evaluate the in vitro hepatotoxicity of MDMA (a synthetic drug commonly named as ecstasy) under heat stress conditions. For that, CG-MS platform was used to compare the volatile metabolites present in cell assays to the accessible mass spectral library in the National Institute of Standard Technology (NIST 14). In the end, 125 metabolites were found, 57 unequivocally identified, 31 were identified only by comparison with spectral databases, 10 of them had only the class identified and 27 remain unknown. Biologically, metabolites associated with MDMA are frequently involved in the amino acid metabolism, glutathione metabolism, pyruvate metabolism among further processes, which MDMA triggers significant metabolic hepatic alterations.\(^{[43]}\)

A typical route of metabolomics analysis goes through essential paths of analysis, depending on the purpose (Fig. 4). Initial steps are based on the existence or not of a metabolite as target of interest, which defines the next stages of analysis. Depending on the initial interest, specific or general analytical tools are applied, until to identify the chemical modification in the metabolites structures. Once identified, a metabolic pathway is proposed, aiming to understand the mechanism of metabolite formation.\(^{[38]}\) Most of the papers interested in metabolomics analysis have been focusing their attention on plant, microbial and biological fluids samples.\(^{[45-47]}\) However, the applications for this field of study are numerous, once they encompass topics as biochemical, bioanalytical and chemometric analysis.\(^{[44]}\)

In order to ensure all scientific requirements necessary for this kind of analysis, several platforms of metabolites analysis and identification have been developed, especially those which use hyphenated systems, such as Liquid Chromatography coupled with High-Resolution Mass Spectrometry (LC-HSRMS). Liquid Chromatography coupled with Mass Spectrometry (LC-MS) is not able to provide deep structural information, due to providing the elemental formula but not specific information, as stereoisomer, a crucial property for drug discovery. In these cases, it makes necessary the use of other techniques, such as X-Ray Crystallography and Nuclear Resonance Magnetic (NRM).\(^{[49]}\)

In the last decades, the development of robust and sensitive chromatography techniques compatible with atmospheric ionization sources, especially the ion-spray ionization (ESI) and Atmospheric-Pressure Chemical Ionization (APCI), made possible the direct coupling of LC with MS. Such approaches became this system as the most preeminent analytical tool for detection and/or quantification of metabolites in any step of the drug development.\(^{[30]}\) Furthermore, the use of atmospheric ionization sources has been useful in on-line characterization of known compounds and their metabolites (dereplication).\(^{[49]}\) Dereplication and multivariate analysis of the NMR and mass spectrometry have been widely applied to identify and isolate antimicrobial natural compounds as well as, microbial metabolites.\(^{[51,52]}\)

The first, fast, simple and promising strategy of dereplication was the prediction of possible metabolites. Such method, known as data-dependent acquisition for drug metabolite identification, is based on the knowledge about the biosynthetic route of a defined parental compound, which is used to create a database able to predict a series of chemical analogs generated during its metabolism. Once with the possible metabolites, the information about the fragmentation pathway is used to identify compounds observed in LC-MS experiment, considering the same mass of the constructed database.\(^{[30]}\) On the other hand, the Non-Target analysis is based on simultaneous analysis of analytical techniques and data processing. This strategy identifies all the detectable metabolites in a sample, besides providing the metabolic profile and quantifying these metabolites individually. Thus, it is feasible to identify all compounds at statistically different levels into the sample.

---

*Figure 4. Steps of a typical route of the metabolomics analysis*\(^{[38]}\)

---

---
groups. This technique also provides the fingerprint for the metabolites responsible for the rapid sample classification, highlighting all signals independently of his intensity[49].

Many studies based on the atmospheric ionization mass spectrometry prove its ideal functionality for data-independent acquisition, as well as its excellent discriminatory power of different metabolites types in a biological matrix. A systematic evaluation of microbial metabolic exchange factors was possible by the use of the non-target metabolomics with atmospheric ionization mass spectrometry. As result, seven secreted metabolites were unequivocally identified by the MS precursor, including sugars, alcohols and macrocycles[53]. In addition, 574 metabolites, among them bile acids, amino acids, lysophosphatidylcholines and glycerophosphocholine, were identified in human serum using ESI in a metabolomics study, with 8 potential biomarkers in pancreatic cancer diagnosis. Also, an overexpression of 3 of them was possible to be identified by the metabolomics analysis[54].

These metabolomics approaches present an interesting versatility, being applied to natural products as well. For example, extract from Ayahuasca, a tea produced by the decoction of Amazonian plants, was analyzed by ESI using the non-target metabolomics strategy aiming to explore its chemical composition and neuroprotective activity. In the end, 1447 and 972 peaks were respectively detected in the positive and negative mode of analysis, with the major alkaloids with isolated and quantified and their neuroprotective activity evaluated[55]. Nine cycloartane were also identified by LC-MS-APCI, which proved that the methodology was completely adequate to describe putative annotations in metabolomics[56]. Aiming to identify biomarkers for diabetes mellitus, using a LC-MS-ESI strategy, 74 biomolecules (such as amino acids, organic acids, amines, amides, sugars, carnitines and ceramides) were recognized in human serum of patients, with 21 specific key metabolites. The metabolomics studies presented that the N-N-dimethylglycine and choline downregulation in these patients might be due to the changes in the methylamine metabolism and may possess high clinical interest[57]. In this way, such method provided a better understanding of the metabolomics and effect promoted by the alteration in the biomolecules levels.

Also, other analytical detectors may be used beyond the MS, depending on the kind of sample. If the majority of the compounds are known, the diode array (DAD) can be used on their quantification, in which gas chromatography (CG) can be used if the objective is to analyze volatile compounds[49]. Based on this detection approach, the quality of six traditional fermented soybean products (FSP) using a non-target metabolite profile was identified by multivariate analysis with CG-MS and Ultra High Performance Liquid Chromatography (UHPLC) coupled to mass spectrometry. It was possible to group the six FPS analyzed according to the major metabolites founded, such as amino acids, sugars, isoflavone glycosides, soyasaponins among others. Also, the identification of primary metabolites (amino acids, sugars, fatty acids among others) were significantly affected by the secondary metabolites (isoflavonoids, soyasaponins, capsaicinoids and 13 non-identified metabolites), influenced by the fermentation time[47]. DAD has been also applied as a detector during the study of vegetable oils. The animal fat (lard) adulteration in vegetable oil was analyzed, which methyl myristate, methyl palmitate, methyl oleate and methyl stearate (all fatty acids methyl esters) were recognized as discriminant markers to identify and quantify lard adulteration in 12 commercial olive oils and 3 lard samples[58].

5. Sample preparation

The concern of the metabolomics field is to identify and quantify the metabolites generated by NMEs metabolism in tissues, cells, organs or organisms. The methodology covers an efficient sample preparation, an analytical sensitive technique of analysis and a statistical tool for data processing[59]. However, sample preparation is a crucial step for the success of a reliable and reproducible result. As the first step in bioanalysis assays, the sample preparation procedure is responsible for providing a clean sample with no interferences, which may influence the results obtained and analyte recovery[60]. Besides that, this step is decisive to the analyte transfer from a biological matrix to the chemical environment of analysis, which is usually the solvent used in the method[61]. Furthermore, the analyte integrity must be maintained from the collection instant to the analysis moment[62].

Choosing an analytical technique depends on the analyte and sample characteristics, which is strongly associated with their physical-chemical properties. There are many parameters to be evaluated in order to choose the sample preparation and method for analysis. Among them, the most relevant properties of the analyte are its chemical structure, pKa value, solubility and polarity properties, stability and adsorption[60]. Each biological matrix presents specific and particular challenges that make their sample preparation a case-study, which is justified by the peculiarities present in the most used biological samples. As an example, urine sample shows high levels of salts, plasma contains a high concentration of phospholipids, while the red blood cells are found in the blood, which is commonly lysed and used for many types of biological analysis. Thus, the type, complexity and behaviour of the matrix and analyte are essential to dictate the better
strategy to be used\cite{62}. Some of the most techniques used are briefly discussed below, as well as their main applications.

### 5.1 Liquid-liquid extraction (LLE)

Commonly used for hydrophobic compounds, the LEE is the most traditional sample preparation technique in bioanalysis\cite{63}. The principle of this method is based on the polarity difference of two immiscible solvents, that are mixed and posteriorly partitioned, which supports the separation of polar and non-polar analytes. The solvents are chosen considering their octanol-water coefficient partition, known as log $P$\cite{64}, which describes the compound solubility\cite{65}.

Initially, log P values were first proposed for functional groups, in which about 800 partition coefficients between octanol and water were determined, dividing the compounds into two groups-H donors (acids, phenols, alcohols, nitriles) and H acceptors (aromatic amines, ethers, esters, ketones). The concentration of each analyte was measured in one phase and, by difference, the concentration was obtained in the other phase\cite{66}. In reversed-phase liquid chromatograph, log P values have been used to measure the hydrophobicity of the compounds analyzed\cite{67}. Such methodology provides a relevant analyte recovery and, due to the solvents characteristics, excellent clean samples. However, LLE is a technique with high time-consuming, involving many steps of sample handling, not suitable to be extracted with large difference of polarity form same sample and high environmental costs\cite{68}. Using LLE, 136 compounds such as antidepressants, neuroleptics, benzodiazepines among others were extracted from human plasma and identified by UHPLC-MS for forensic analysis. In this kind of analysis, the monitoring of various drug classes (a multi-analyte procedure) in one single body sample is essential. Recovery values between 9 and 89% were found for these different compounds, proving that LLE is comparable to the other multi-analytes methods, such as SPE\cite{69}. Ginkgo flavonoids, terpene lactones and nimodipine were possible to be quantified using LLE extraction from rat plasm. A Two-step LLE was performed in this study, once some analytes, as nimodipine, are highly sensitive to temperature and acid solutions, while other compounds required only one of these conditions. The final solution was analyzed by UHPLC-MS/MS, in which the method was able to recover up to 95% of analytes in nano concentrations\cite{70}. Odanacatib, a protease inhibitor in development for osteoporosis treatment, was quantified by LC-MS/MS from its extraction from the human plasma by an LLE-based technique using a semi-automatized platform\cite{71}. Osteoporosis is a worldwide common disease and affect about 200 million people\cite{72} and, for the study of the pharmacokinetic parameters of Odanacatib in the human plasma, analytical assays able to quantify the analyte and their metabolites are necessary. By LC-MS/MS, the extraction recovery reached values about 99% for the parental compound and the analytical method was able to quantify the Odanacatib in a very low concentration, within a range of 0.5-500 ng.mL$^{-1}$\cite{73}. Such determination evidences the ability of quantification, as well as the sensibility of the LLE flowed by LC-MS/MS analysis.

### 5.2 Protein precipitation (PPT)

Precipitation is extensively used to recover proteins in biological matrices, such as blood samples, that contain a high amount of analytes. The precipitation phenomenon of proteins consists in separating the proteins from the other analytes based on their isoelectric potential, which these molecules become into an insoluble phase in the sample. In this methodology, the precipitation is induced by the addition of some precipitation inducers, such as miscible organic solvent, salting out, the addition of metallic ions or pH changing. All these inducers are able to create an isoelectric disequilibrium onto the protein potential, promoting its precipitation by the charge annulment\cite{74}. As an advantage, PPT is a fast and simple approach and can be applied to proteins with both hydrophobic and hydrophilic characteristics. However, when the objective is to extract parental compounds and their metabolites, PPT can be followed by LLE or SPE to achieve more efficiency\cite{75}. This rapid and quantitative protein precipitation method was proved for proteome analysis by mass spectrometry, with high protein recovery (=98%) by increments of salt concentration and temperature, incubating the sample in acetone 80% and bovine serum albumin. The PPT process consumed 2 minutes in total\cite{76}, showing its practicality and an effective ability for protein recovery. Also, smaller biomolecules, like vitamins, can be recovered by the PPT and detected using some analytical features. Vitamin A levels in human serum were quantified using PPT and HPLC coupled with Ultraviolet detector (HPLC-UV). According to the World Health Organization (WHO), the analysis of vitamin A in human serum is essential to public health, especially in low and middle-income countries. Based on that, samples of serum were precipitated with acetonitrile for extraction of vitamin A, which demonstrated that the method was robust and sensitive in measuring the analyte amount in low concentrations, nearly 20 µg.dL$^{-1}$\cite{77}. From the point of view of the drug target discovery in proteomic scale, a solvent-induce protein precipitation approach was applied to provide the interaction profile of anticancer drugs with their target proteins, quantified by LC-MS/MS. According to the Western blotting analysis, the target protein DHFR (dihydrofolate reductase) was fully precipitated in 15% of organic solvent and, analyzing the control of drug-target interaction, it was proved its stabilization after binding to the drug\cite{78}. Based on that, as a huge advantage, it was possible to discover drug targets in cell lysates without any drug modification, once the conventional
strategy requires the modification or immobilization of the small drug molecules. This supports the importance of PTT during drug discovery studies, in which the drug-molecules interaction is crucial to its action, distribution and body elimination. Clobazam and its major metabolite (N-desmethylocobazam) present in the human plasm of patients, who were in treatment with the drug, was determined by a simple PPT flowed by LC-MS/MS. The clobazam function is to block the GABA-A receptors in order to promote beneficial effects for epilepsy, and the method used was able to validate its action in a very sensitive range of 2.0-750 ng.mL-1 for clobazam and 0.7-200 ng.mL-1 for the metabolite\[76\]. Besides the rapid and efficient results, the PPT method is essential to bioanalysis, since it does not require high amount of the drug or target, being applicable to the natural treatment conditions in humans.

5.3 Solid phase extraction (SPE)

Currently, SPE is the most advantageous technique in terms of cost-effective, efficiency, reproducibility and easy-automatization. Such methodology is widely used in the metabolomics studies, especially due to provides the clean-up and pre-concentration in the same analyses. Usually, SPE technique is divided into three different mechanisms: reversed phase, normal phase and ion exchange. The reverse phase is commonly applied to drugs in biological fluids and environmental pollutants in water\[60\]. Such technique has been used to analyze nucleotides in human cerebrospinal fluid, followed by LC analysis. In living systems, the purine and pyrimidine nucleotides play important roles in numerous metabolic processes, and it is very important to have a balance of their levels. The method reached efficiency over 90% for the analytes, which led analytes quantification within the range 0-500 pmol\[60\]. During organ transplantation and autoimmune disorders treatment, cyclosporine A is a very common drug used, and its quantification in the human plasm is essential. SPE extraction recovered over 90% of the drug, besides activating with high sensitivity, in a µg range\[77\]. Some of the SPE applications as extraction method also include antibodies\[79\], and simultaneous determination of drugs and their main metabolites, as an example the quantification of Fluoxetine and fluoxetine in human plasm. Fluoxetine is commonly prescribed for depression treatment, but it is also recommended to other disturbs, in which its amount in the bloodstream is very important to determine the therapeutic time. The SPE technique was linear over the concentration range of 0.048-100 ng.mL-1 with a LLOQ of 32 pg.mL-1 for both analytes\[60\]. Such data compare the

As a final comparison, Table 1 summarizes the three most used sample preparations techniques among others as environmental friendly with a potential to be widely used in a near future. This discussion goes beyond this paper, but many studies report and discuss methodologies such as solid-phase micro extraction (SPME), in-tube SPME, stir bar sorptive extraction (SBSE), micro extraction by packed sorbent (MEPS), liquid-phase micro extraction (LPME), single drop micro extraction (SDME) among others as environmental friendly with a potential to be widely used in a near future. As a final comparison, Table 1 summarizes the three most used sample preparations techniques\[60\]. Such data compare the
main parameters between the sample preparation techniques, which proves their better ability for each one, as a resume before starting any analyte assay.

Table 2. Comparison between the most used sample preparation techniques. The symbols resume the high (↑), medium (↔) or low (↓) characteristics for each parameter

<table>
<thead>
<tr>
<th>Parameter</th>
<th>LLE</th>
<th>PPT</th>
<th>SPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample handling</td>
<td>↑</td>
<td>↓</td>
<td>↑</td>
</tr>
<tr>
<td>Selectivity</td>
<td>↔</td>
<td>↓</td>
<td>↑</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>↔</td>
<td>↔</td>
<td>↔</td>
</tr>
<tr>
<td>Ion suppression</td>
<td>↓</td>
<td>↑</td>
<td>↓</td>
</tr>
<tr>
<td>Automation</td>
<td>↓</td>
<td>↓</td>
<td>↑</td>
</tr>
<tr>
<td>Cost</td>
<td>↑</td>
<td>↓</td>
<td>↑</td>
</tr>
<tr>
<td>Analyte available</td>
<td>Lipophilic</td>
<td>Hydrophilic</td>
<td>Both polarity</td>
</tr>
</tbody>
</table>

Adapted from[60]

6. Method validation

Among the indispensable steps for drug development and discovery, another essential part of the process and crucial to the success of bioanalysis is the method validation by the regulatory agency. Worldly, two of the most used guidance for the drug regulation and method validation are the Food and Drug Agency (FDA) and the European Medicines Agency (EMA), with a specific guide for validation of bioanalytical methods proposed by each agency[84,85]. It is important to note that all analytical methods have specifics requirements to be approved by each regulatory agency[85], but all of them aim precisely at the definition of method validation: a sensitive, accurate and robust method that complies with standard specifications. In terms of bioanalysis, this method is usually applied to biological matrices to quantify the parental compounds and other endogenous compounds and/or metabolites[86]. Usually, the figures of merit are common for most guides and each one shows how the figure of merit must be evaluated, as discussed below:

**Limit of Detection (LOD):** LOD is defined as the lowest concentration of analyte which can be detected by the method, but not quantified. Each guide shows the method for evaluating the LOD, in which the most common for the chromatographic methods is the ratio analysis between analyte chromatographic signal for the and line-base of the analysis[86]. Such signal/noise ratio (S/N) is measured by dividing the analytical signal (signal integration) of the analyte by the noise signal produced by that equipment.

**Lowest Limit of Quantification (LOQ):** in contrast to LOD, LOQ is the minimum analyte concentration which can be quantified by the method. As during the LOD parameter, the LOQ is measured by comparing the ratio between the analyte peak and the line base of the analysis. Both LOD and LOQ parameters must be approved by the chosen guide to be properly used in a curve calibration. Usually, the acceptable limits for the LOD and LOQ in the chromatographic methods are 15% and 20%, respectively, considering the error between the replicates[84,86].

**Selectivity:** this criterion is based on the method ability in quantifying precisely the analyte of interest in the presence of interfering species, such as metabolites, products of degradation and endogenous compounds[84-85]. Commonly, selectivity is evaluated by the comparison of the analyte signal-response and a blank of the matrix, which such analyte is studied[86].

**Linearity:** defined as the concentration range for the analyte quantification, the linearity assesses the method ability to obtain results directly proportional to the analyte concentration in a matrix studied[60]. Usually, at least five concentration levels and a minimum of three replicates in each point of the curve are recommended for a statistically reliable analysis (85). The acceptance criteria change according to each guide, but the most acceptable error values are, respectively, 15% and 20% for LOQ and the other points of the curve[84]. Other non-linear models can be also used for the calibration curve construction, as well as the use of another weighting over the x coefficient[60].

**Matrix Effect (ME):** during a bioanalysis, when some endogenous compound present in the matrix studied interferes with the analyte detection, it is commonly known as matrix effect. This effect is most usual in chromatographic methods, even using hyphenated techniques, as LC-MS/MS[84]. The ME is usually determined by the correlation of the signal area in the matrix with the signal area of a standard solution, as presented in Equation 1. Statistically, values below 15% of the coefficient of variation (CV%) are normally accepted for guaranteeing the no matrix effect for the method.

\[
ME(\%) = \left( \frac{\text{Signal area in matrix spiked postextraction}}{\text{Signal area in standard solution}} - 1 \right) \times 100
\]  

(1)
Precision: this parameter describes the proximity of each analyte measurements. The precision, according to the most studied guides, must not exceed 15% of CV except for the LOQ, in which the tolerated CV is up to 20%\(^{[86]}\). Precision usually is analyzed in two different ways: repeatability (intra-assay) and intermediate precision\(^{[85]}\). Intra-assay describes the variations observed in a single run (standard deviation \(\delta_{\text{INTRA}}\)) while the intermediate precision describes the variations observed in different analytical runs (standard deviation \(\delta_{\text{INTER}}\)). The combination of these two standard deviations provides the total method precision (Equation 2), in which \(n\) is the number of replicates obtained for Intra-assay.

\[
\sigma_{\text{Total}} = \sqrt{\frac{\sigma_{\text{intra}}^2}{n} + \sigma_{\text{inter}}^2}^{[85]}
\]

Accuracy: On the other hand, the accuracy evaluates the proximity of the results average acquired by the analytical method to the real analyte concentration. Usually, accuracy is evaluated by the Certified Reference Material (CRM), which analysis of six replicates by the method developed and the real concentration of each one is performed. Then, using the mean, standard error and CV (%) as presented in Equation 3.

\[
\text{Accuracy(\%)} = 100 \left(\frac{\text{average concentration obtained}}{\text{CRM concentration}}\right)^{[85]}
\]

However, for the same cases, CRMs are not available. In these cases, the accuracy can be obtained by the recovery. It must be performed by 18 sample blanks fortified with three different levels of concentration in six replicates (usually 1, 1.5 and 2 times the limit of detection). Last, the mean, standard error and CV (%) are used and applied to Equation 4. The acceptable values for the accuracy usually are 15% of deviation for LOQ and 20% for the other points of the analysis\(^{[86]}\).

\[
\text{Recovery(\%)} = 100 \left(\frac{\text{average concentration obtained}}{\text{fortified concentration}}\right)^{[85]}
\]

Precision and accuracy parameters determine the errors in the analysis, and they are the most important criteria to evaluate the conformity of the experiments\(^{[84]}\). Also, many other figures of merit can be evaluated according to the objective of the study. This review has focused on the major parameters for chromatography methods for bioanalysis. However, several guides concern about the other applications, based on different analytical methods. In general, the type of analyte and the desired result are crucial to determine the analytical strategy for each study. Herein, we summarized a brief review of the major aspects that should be kept in mind when an analytical analysis is aimed. In addition, this review may also guide many researchers, mainly analytical chemists which are starting studies about natural or synthetic NMEs, to the major parameters and concepts before planning their initial trials.

References


