IMI – oral biopharmaceutics tools project – evaluation of bottom-up PBPK prediction success part 1: Characterisation of the OrBiTo database of compounds

DOI:
10.1016/j.ejps.2016.09.027

Document Version
Accepted author manuscript

Published in:
European Journal of Pharmaceutical Sciences

IMI – Oral biopharmaceutics tools project – Evaluation of bottom-up PBPK prediction success part 1: Characterisation of the OrBiTo database of compounds

Alison Margolskee, Adam S. Darwich, Xavier Pepin, Shriram M. Pathak, Michael B. Bolger, Leon Aarons, Amin Rostami-Hodjegan, Jonas Angstenberger, Franziska Graf, Loic Laplanche, Thomas Müller, Sara Carlert, Pankaj Daga, Dónal Murphy, Christer Tannergren, Mohammed Yasin, Susanne Greschat-Schade, Wolfgang Mück, Uwe Muenster, Dorina van der Mey, Kerstin Julia Frank, Richard Lloyd, Lieve Adriaenssen, Jan Bevernage, Loeckie De Zwart, Dominique Swerts, Christophe Tistaert, An Van Den Bergh, Achiel Van Peer, Stefania Beato, Anh-Thu Nguyen-Trung, Joanne Bennett, Mark McAllister, Mei Wong, Patricia Zane, Céline Ollier, Pascale Vicat, Markus Kolhmann, Alexander Marker, Priscilla Brun, Florent Mazuir, Stéphane Beilles, Marta Venczel, Xavier Boulenc, Petra Loos, Hans Lennernäs, Bertil Abrahamsson

PII: S0928-0987(16)30392-X
DOI: doi: 10.1016/j.ejps.2016.09.027
Reference: PHASCI 3733

To appear in:

Received date: 10 May 2016
Revised date: 12 August 2016
Accepted date: 17 September 2016

IMI – Oral Biopharmaceutics Tools project – Evaluation of Bottom-up PBPK Prediction Success Part 1: Characterisation of the OrBiTo Database of Compounds

Alison Margolskee[1], Adam S. Darwich[1], Xavier Pepin[2,3], Shriram M. Pathak[4], Michael B. Bolger[5], Leon Aarons[1], Amin Rostami-Hodjegan[1,4], Jonas Angstenberger[6], Franziska Graf[6], Loic Laplanche[6], Thomas Müller[6], Sara Carlert[7], Pankaj Daga[8], Dónal Murphy[2], Christer Tannergren[7], Mohammed Yasin[2], Susanne Greschat-Schade[9], Wolfgang Mück[9], Uwe Muenster[9], Dorina van der Mey[9], Kerstin Julia Frank[10], Richard Lloyd[11], Lieve Adriaenssen[12], Jan Bevernage[12], Loeckie De Zwart[12], Dominique Swerts[12], Christophe Tistaert[12], An Van Den Bergh[12], Achiel Van Peer[12], Stefania Beato[13], Anh-Thu Nguyen-Trung[13], Joanne Bennett[14], Mark McAllister[14], Mei Wong[14], Patricia Zane[15], Céline Ollier[3], Pascale Vicat[3], Markus Kolhmann[16], Alexander Marker[16], Priscilla Brun[3], Florent Mazuir[3], Stéphane Beilles[3], Marta Venczel[16], Xavier Boulenc[3], Petra Loos[16], Hans Lennernäs[17], Bertil Abrahamsson[7],

[1]University of Manchester, United Kingdom

[2]AstraZeneca, United Kingdom

[3]Sanofi, France

[4]Simcyp Ltd, United Kingdom

[5]Simulations Plus, Inc., United States

[6]AbbVie, Germany

[7]AstraZeneca, Sweden

[8]AstraZeneca, United States

[2]AstraZeneca, United Kingdom

[9]Bayer Pharma AG, Germany

[10]Boehringer Ingelhheim Pharma GmbH & Co. KG, Germany

[11]GlaxoSmithKline, United Kingdom

[12]Janssen, Belgium

[13]Novartis, Switzerland

[14]Pfizer, United Kingdom

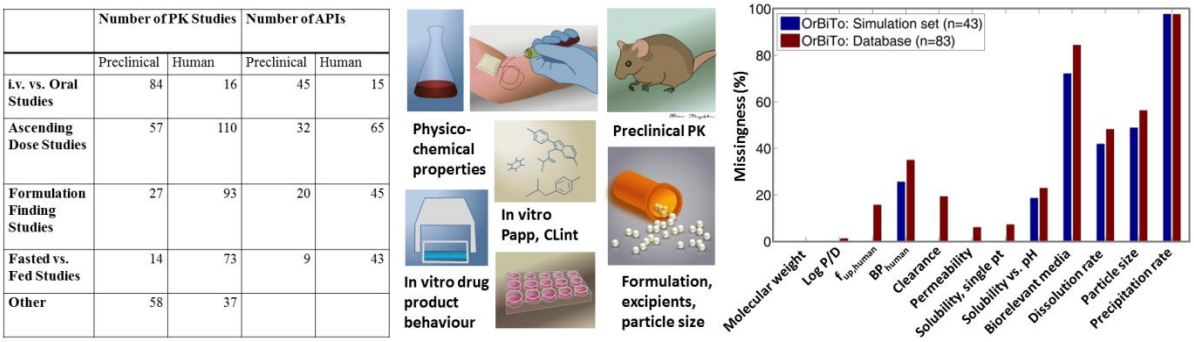
[15]Sanofi, United States

[3]Sanofi, France

[16]Sanofi, Germany

[17]Uppsala University, Sweden

Graphical Abstract



	Number of PK Studies		Number of APIs	
	Preclinical	Human	Preclinical	Human
i.v. vs. Oral Studies	84	16	45	15
Ascending Dose Studies	57	110	32	65
Formulation Finding Studies	27	93	20	45
Fasted vs. Fed Studies	14	73	9	43
Other	58	37		

Abstract

Predicting oral bioavailability (F_{oral}) is of importance for estimating systemic exposure of orally administered drugs. Physiologically-based pharmacokinetic (PBPK) modelling and simulation have been applied extensively in biopharmaceutics recently. The Oral Biopharmaceutical Tools (OrBiTo) project (Innovative Medicines Initiative) aims to develop and improve upon biopharmaceutical tools, including PBPK absorption models. A large-scale evaluation of PBPK models may be considered the first step. Here we characterise the OrBiTo active pharmaceutical ingredient (API) database for use in a large-scale simulation study.

The OrBiTo database comprised 83 APIs and 1,475 study arms. The database displayed a median logP of 3.60 (2.40-4.58), human blood-to-plasma ratio of 0.62 (0.57-0.71), and fraction unbound in plasma of 0.05 (0.01-0.17). The database mainly consisted of basic compounds (48.19%) and Biopharmaceutics Classification System class II compounds (55.81%). Median human intravenous clearance was 16.9 L/h (interquartile range: 11.6 – 43.6 L/h; n=23), volume of distribution was 80.8 L (54.5 – 239 L; n=23). The majority of oral formulations were immediate release (IR: 87.6%). Human F_{oral} displayed a median of 0.415 (0.203 – 0.724; n=22) for IR formulations.

The OrBiTo database was found to be largely representative of previously published datasets. 43 of the APIs were found to satisfy the minimum inclusion criteria for the simulation exercise, and many of these have significant gaps of other key parameters, which could potentially impact the interpretability of the simulation outcome. However, the OrBiTo simulation exercise represents a unique opportunity to perform a large-scale evaluation of the PBPK approach to predicting oral biopharmaceutics.

Keywords:

Physiologically-based pharmacokinetics (PBPK); modelling and simulation (M&S); absorption; oral bioavailability (F_{oral}); biopharmaceutics; drug database

Abbreviations:

API = Active pharmaceutical ingredient,

AUC = Area under the curve,

BCS = Biopharmaceutics classification system,

BP = Blood-to-plasma ratio,

C_{max} = Maximum concentration,

CL = Clearance,

D_o = Dose number according to BCS,

DDI = Drug-drug interaction,

EFPIA = European Federation of Pharmaceutical Industries and Associations,

F_{oral}= Absolute oral bioavailability,

F_{rel}= Relative bioavailability,

fu_p= Fraction unbound in plasma,

IMI = Innovative Medicines Initiative,

LogP = Logarithm of the octanol/water partition coefficient,

$LogD_{pH}$ = Logarithm of the octanol/water partition coefficient at a given pH,

M&S = modelling and simulation,

MW = Molecular weight,

OrBiTo = Oral Biopharmaceutical Tools,

P_{eff} = Effective permeability,

PBPK = Physiologically-based pharmacokinetic,

PK = Pharmacokinetics

t_{max} = Time at maximum concentration,

V_d = Volume of distribution,

1. Introduction

The oral route remains the preferred route of administration due to its ease of use and minimal invasiveness. However, an oral formulation will undergo many processes prior to entering systemic circulation, from release from formulation to dissolution of solid particles, potential precipitation, permeation, and first-pass metabolism. The prediction of absolute and relative oral bioavailability (F_{oral} and F_{rel}, respectively) is of great importance for anticipating the systemic exposure of orally administered formulations. For example, the ability to assess *a priori* when an altered state, *i.e.* a novel oral formulation or a change in prandial state, is likely to cause a significant change in oral drug exposure is increasingly considered of importance to successful biopharmaceutical development. F_{oral} is governed to a large extent by the dissolution in the gastrointestinal tract, absorption, and first pass metabolism of the active pharmaceutical ingredient (API), thus schemes such as the Biopharmaceutics Classification System (BCS), which classifies a drug based on its permeability and solubility characteristics, and the Biopharmaceutics Drug Disposition Classification System (BDDCS), which classifies drugs based on its solubility and extent of metabolism (Benet et al., 2011), have proven valuable in streamlining the experimental and/or clinical study design and ultimately providing the basis for biowaivers for freely soluble, highly permeable drugs (FDA, 2000).

However, for more complex drugs, formulations, or drug disposition characteristics, *e.g.* the involvement of intestinal metabolism, active transport, or unusual dissolution behaviour,

schemes such as the BCS might be considered an oversimplification, further not providing a basis for quantitative predictions. *In silico* physiologically-based pharmacokinetic (PBPK) modelling and simulation (M&S) have been employed for more complex biopharmaceutical problems with promising results which can be found in abundance in the literature (Sjogren et al., 2013; Wu et al., 2013; Patel et al., 2014). The application of PBPK M&S has been particularly successfully in recent years for the prediction of metabolic drug-drug interactions (DDI) and has provided the basis for waiving clinical DDI studies (Huang et al., 2013; Thondre et al., 2013), as well as aiding in dose selection by providing drug exposure predictions for special populations such as paediatrics and organ-impaired subjects (Futagami et al., 2013). There is therefore a great incentive for an increased use of model-based methods throughout drug development.

The OrBiTo (Oral biopharmaceutics Tools) project, funded by IMI JU (Innovative Medicines Initiative Joint Undertaking), aims to develop and improve on existing tools in biopharmaceutical development. The OrBiTo project is a five year IMI consortium, bringing together nine European universities, one regulatory agency, one non-profit research organization, four small-medium enterprises and thirteen pharmaceutical companies with the common goal of proposing innovative tools for oral biopharmaceutics. The consortium is organised in four work packages (WP): physico-chemical tools (WP-1), *in vitro* tools (WP-2), *in vivo* tools (WP-3), and *in silico* models (WP-4). Central to the consortium is the creation of a novel database organising physicochemical, physiological, and pharmacokinetic data for evaluation of current gaps in biopharmaceutical tools, and the improvement of these and new tools throughout the project. (Lennernas et al., 2013)

A significant part of the OrBiTo initiative is the improvement of current *in silico* tools (Figure 1). A detailed review of the current state of PBPK absorption modelling can be found elsewhere (Kostewicz et al., 2013). As a part of OrBiTo, the systematic large-scale evaluation

of current PBPK absorption models is considered an essential first step in identifying where models perform, or underperform, to guide further model development. Previous large-scale efforts have been carried out to assess the predictability of PBPK models (Chang and Leblond, 1971; Parrott and Lave, 2002; Jones et al., 2006; De Buck et al., 2007; Jones et al., 2011; Poulin et al., 2011). The OrBiTo simulation exercise intended to extend the evaluation of PBPK model performance to different software packages and absorption models, focusing on aspects relevant to oral biopharmaceutical drug development, such as a range of oral formulations, bioavailability and food effects studies, and attempting to reveal areas where *in silico* models and the model building process could be improved.

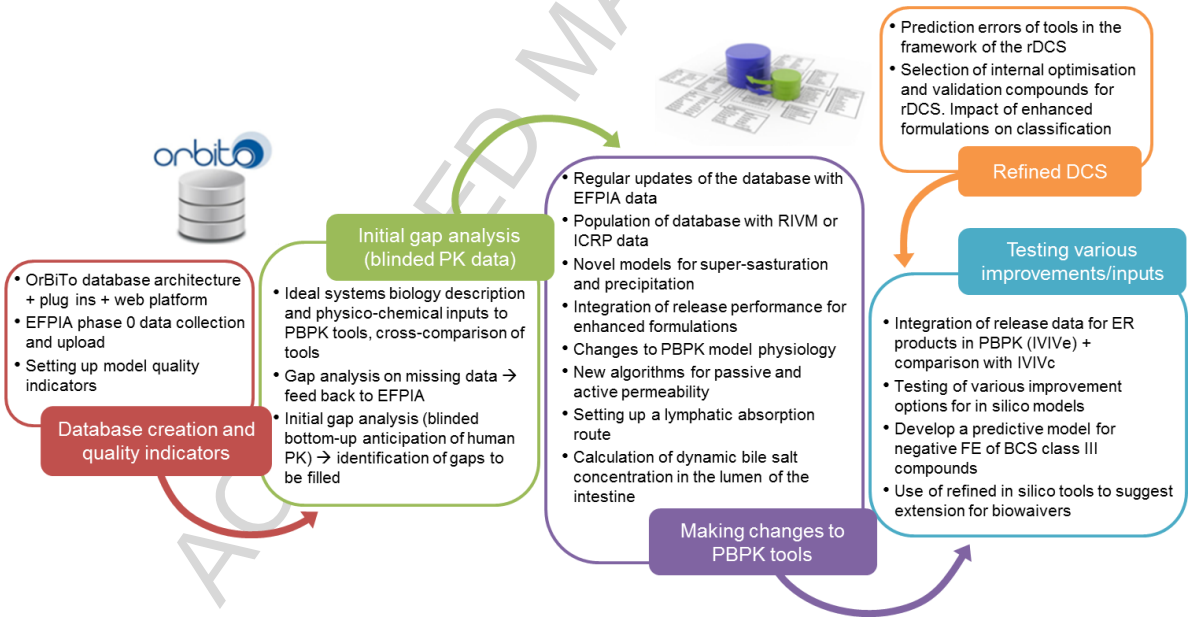


Figure 1: Overall aims of the *in silico* modelling activities in OrBiTo

The OrBiTo database of APIs was intended to combine the knowledge and resources from across different EFPIA members for the purposes of enhancing drug and formulation development, specific to oral drug products and anticipation of bioavailability. It was set up in a manner where selective information could be maintained as anonymous, while providing other members of OrBiTo with useful information to further their research goals. The API

datasets comprised information on the drug substances, drug moiety, formulations, and preclinical and clinical pharmacokinetic studies. Typical blinded information included the international nonproprietary and brand names of the APIs, their chemical structures and other selected compound properties, as well as the identity of the companies uploading the information. This structure allowed for companies to share information that would otherwise have been kept confidential, while still maintaining a level of anonymity. An anonymous messaging feature built into the database interface allowed for a communication link between modellers and compound owners throughout the simulation exercise for clarification of input data sources and values (Lacy-Jones et al. – Submitted). This structure also allowed for the blinding of clinical pharmacokinetic data for the duration of the OrBiTo simulation exercise, ensuring that pure bottom-up predictions could be produced in order to analyse the predictive performance of existing *in silico* methods.

This manuscript summarises the current state of the OrBiTo database, and criteria for compounds to be included in the OrBiTo simulation exercise, a large scale evaluation of selected current PBPK absorption model platforms with focus on the prediction of absolute and relative oral bioavailability and oral pharmacokinetics. Details of the setup of the simulation exercise and analysis of the prediction performance of the simulations can be found in the companion manuscripts (Margolskee et al. – Part 2 – Submitted; Darwich et al. – Part 3 – Submitted).

2. Methods

2.1. Gap Analysis & API Selection Criteria

The gap analysis was intended to reveal any gaps in information that may impact the results of the simulation exercise. For example, any lack of particular administered formulation

types, or imbalance of relevant BCS classes. The APIs were analysed for their physicochemical properties (molecular weight, acid-base nature, logP, logD, BCS class), blood and plasma binding properties (fu_p, BP), formulation characteristics, and availability and type of *in vitro* and preclinical clearance information. The distributions of these properties were compared with similar databases from the literature, and a subset of APIs was selected for inclusion in the OrBiTo simulation exercise, based on availability of parameter information and clinical data. Analysis in accordance to BCS classification was carried out using BCS class as reported in the OrBiTo database, the criteria for which depended upon the reporting company. If BCS class was not available in the database, then an estimated BCS classification was assigned based on fraction absorbed (f_a) estimated from scaled effective permeability (P_{eff}) and dose number (D_o) (see Appendix for further details).

Comparison of the database to other sources in the literature was intended to establish the balance or reveal any imbalance of the drugs in the OrBiTo database, and to allow for prioritisation of further supplementation of the database throughout the project. Datasets used for comparison to the OrBiTo database included: The PhRMA dataset published by Poulin and co-workers (2011) (n=108 compounds), the BDDCS database as published by Benet and co-workers (n=927), Hosea *et al*. (2009) dataset for validation of allometric scaling of clearance (n=50), the database of the WHO list of essential medicines as compiled by Kasim *et al*. (2004) (n=123), the Pfizer compound set for testing of PBPK modelling as published by Jones *et al*. (2011) (n=21), the Johnson and Johnson dataset for testing PBPK modelling as published by de Buck and co-workers (2007) (n=26), Obach *et al*. (2008) database of intravenous pharmacokinetic parameters (n=670) and Bu (2006) dataset over cytochrome P450 3A4 substrates (n=113) (Kasim et al., 2004; Bu, 2006; De Buck et al., 2007; Obach et al., 2008; Hosea et al., 2009; Benet et al., 2011; Jones et al., 2011; Poulin et al., 2011).

Potential compound overlaps between the published datasets were not considered during comparative analysis.

The minimum inclusion criteria for the simulation exercise was selected to be the availability of the following information: Molecular weight, at least one logP or logD value (measured or calculated), at least one solubility point estimate or dissolution profile, measured *in vitro* permeability along with reference compounds necessary for scaling to P_{eff} where applicable, any form of human *in vitro* clearance (e.g. in human liver microsomes, human hepatocytes, or recombinant CYP) or preclinical i.v. (allowing for allometric scaling of clearance), and human fu_p.

3. Results

The OrBiTo database consisted of a total of 83 APIs, as submitted by EFPIA members, as of March 2014. Of these, 43 were found to satisfy the selection criteria. The 43 APIs chosen represent over 165 human studies, and over 600 human study arms. A summary of the OrBiTo database compounds and some of their physico-chemical, blood and plasma binding properties can be found in Table 1, with check marks indicating inclusion in the simulation set. Discussion of the degree of missingness of key parameters is included in the next section. Comparison of the properties of the simulation set with the overall database, as well as with other databases from the literature, can be found in following sections. In the remaining sections, we discuss the PK study designs described in the database (species, administration routes, formulations) as well as observed *in vivo* characteristics (F_{oral}, clearance (CL), and volume of distribution (V_d)).

3.1. Missingness and Compound Selection

The degree of missingness of certain compound properties and *in vitro* measurements reported for the APIs in the OrBiTo database was repeatedly analysed throughout the selection process of APIs to be utilised in the simulation exercise. The main criteria restricting the inclusion of APIs in the simulation set was the availability of scalable clearance (19.3% missingness), followed by fu_p measured in humans (15.7%), any solubility measurement (7.23%), and permeability with reference compounds (6.02%). A large degree of missingness was also observed relating to formulation, dissolution and solubility properties, with 22.9% of the database missing solubility vs. pH profiles, 84.3% biorelevant solubility information, 97.6% degradation rate and 56.3% particle size of solid formulations.

As the simulation inclusion criteria allowed for estimation of BP_{human}, 25.6% of the simulation set was missing this parameter. Further, certain solubility and formulation related parameters were also allowed to be missing in the simulation set, including: solubility measurements in biorelevant media (72.1% missing), particle size information (48.8%) and degradation rate (97.7%; Figure 2).

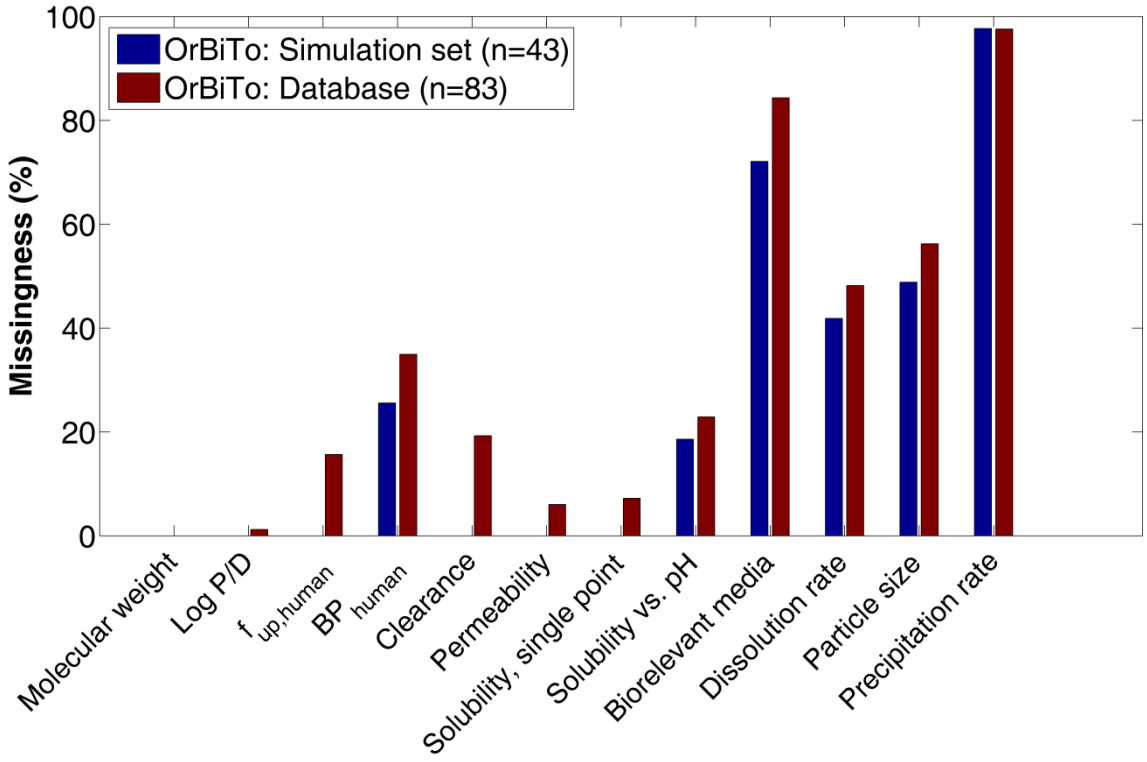


Figure 2: Degree of missingness for different parameters of interest. fu_p = fraction unbound in plasma; BP = blood to plasma ratio

The OrBiTo database contained a diverse set of clearance sources, where 28.9% of APIs were provided with clearance information from human liver microsomes, 19.3% from human hepatocytes, and 2.41% contained recombinant CYP enzyme information. A total of 60.2% of APIs came with pre-clinical data allowing for allometric scaling of clearance, of which 76.1% (48.5% of the database) included information facilitating allometric scaling using multiple pre-clinical species (Figure 3).

As simulation exercise inclusion criteria stipulated that any included API should include a human *in vitro* clearance source (such as human liver microsomes or hepatocytes) or preclinical i.v. study arms available for allometric scaling, there was a higher frequency of

clearance information in the simulation set as compared to the database as a whole. In the simulation set, 39.5% of APIs had human liver microsome information, 30.2% had human hepatocyte information, and 4.65% contained recombinant CYP enzyme information (Figure 3). These frequencies are not mutually exclusive as 16.3% of APIs had clearance information from multiple *in vitro* systems.

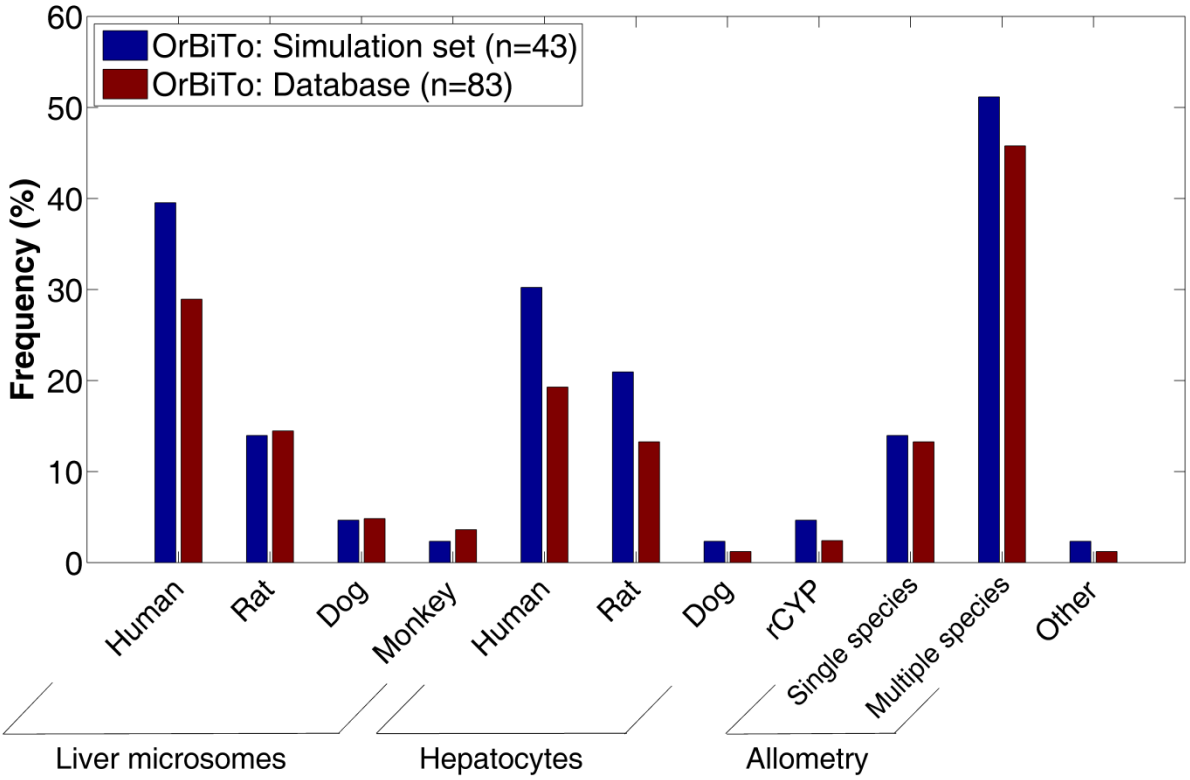


Figure 3. Frequency of APIs in the OrBiTo database providing different *in vitro* sources of clearance (microsomes, hepatocytes and recombinant CYP systems) and allometric scaling of clearance through single or multiple preclinical species.

3.2. Characterisation of the OrBiTo Database Compound Specific Properties

Table 1: Summary of OrBiTo database APIs

	APIcode	MW	BCS Class	Acid/Base Nature	Highest Basic pKa	Lowest Acidic pKa	LogP	LogD pH 7.4	fu_p	BP
	A7014	140	1	Zwitterion	8.6	2	-5.1			

✓	**A1897**	150	1	Strong Base	10.74		1.995	-0.8		
	A7566	160	1	Strong Base	8.8			-0.7	0.554	
	A5766	190	3	Strong Base	9.42			-1.03	0.487	
	A8075	200	1	Strong Base	9.06			3.19 (6.8***)	0.302*	0.598*
	A9208	240	2	Strong Base	11		2.3	2 (6.8***)	0.24	
✓	**A8791**	250	2	Acid		7.7		0.8 (6.7***)	0.25	0.66
✓	**A2352**	290	1	Acid	7.6		5.2	5.8	0.01	
✓	**A0048**	310	2	Acid		4.5	2.14	-0.3	0.0206*	0.538*
✓	**A0608**	320	1	Acid		7.27	4.4		0.0408*	
✓	**A2437**	320	1	Strong Base	9.31			1.1	0.362	1.9
✓	**A6099**	330	1	Zwitterion	7.57	3.69	3.23	2.97 (6.8***)	0.074	
✓	**A7415**	330	2	Strong Base	8.6		-0.72		0.74	1.81*
✓	**A2733**	350	1	Strong Base	10.2		2.79	1.5	0.0563*	3.3
	A8653	350	2	Acid		0.8		4.99	0.00283*	
	A9590	350	2	Neutral			3.12		0.05	
✓	**A5262**	360	2	Neutral			2.329	3.42	0.031	0.6
✓	**A6135**	360	2	Acid		3.85	2.11	-1.45	0.00746*	
	A4684	370	1	Zwitterion	9.1	1		1.92*	0.308*	1.39*
✓	**A9606**	370	2	Weak Base	6.1			4	0.007	0.625
	A3491	370	4**	Ampholyte	2.6	4.2	1.36	-7	0.55	0.64****
	A4010	370		Strong Base	7.6		5.6	5.27 (8***)	0.0767	0.647
✓	**A7513**	390	2	Weak Base	1.81		2.645	2.5 (6.5***)	0.017*	0.703****
✓	**A6555**	390	4	Acid		2.64	4.58	-0.22	0.029	0.56
✓	**A4460**	390	2**	Strong Base	8.8		4.7	2.7	0.05	2
✓	**A0855**	400	2	Neutral			3.23		0.059	0.9
✓	**A2853**	400	2	Weak Base	3.1		2.99		0.0185	0.68
✓	**A8942**	400	2**	Weak Base	6.1		2.55		0.0638*	
	A6257	410	1	Acid		4.5	3.9	1.6 (6.8***)	0.02	
✓	**A4356**	410	2	Ampholyte	2.8	7.5		2.6 (6.8***)	0.0013	0.6
	A2050	410		Acid		8.18	2.72		0.034	0.776*
✓	**A7597**	420	2	Weak Base	3.5		3.09	3.09 (7***)	0.167	0.539*
	A2284	420	4	Ampholyte	1.46	5.66	0.88	-0.62 (7***)	0.311*	0.58
	A2720	430	2	Ampholyte	5.26	12.38	3.626	5.4	0.1	0.6
✓	**A2101**	430	2**	Strong Base	7.8		4.01	3.74 (8***)	0.08	0.712
	A3622	430		Zwitterion	8.2	3.9		-38 (6***)	0.546	0.677*
✓	**A3837**	440	1	Acid		7.9	-0.08		0.52	1.044
	A1150	440	2	Strong Base	13.67		3.65	2.51	0.2	0.5
	A1876	440	2	Weak Base	3.3		6.87		0.0027	0.56
	A6197	440	2	Ampholyte	3.5	4	3.6	3.6	0.17	0.5
	A6646	440	2	Neutral			1.35		0.0752*	0.689*

✓	**A2771**	440	4	Neutral			3.96		0.0044	0.7
	A2092	440		Weak Base	5.8		2.4	2.4		
	A5744	450	1	Strong Base	9.06			3.19 (6.8***)	0.349*	0.598*
	A0772	450	2	Neutral			3.5	3.5	0.994	0.6
✓	**A6939**	450	2	Acid		5.4	5.51		0.0002	0.55
✓	**A3078**	450	4	Neutral			2.54		0.06	
	A8379	460	1	Zwitterion	7.64	5.94	4.2	3.3	0.004	0.49
	A5616	460	2	Zwitterion	14	-1		1.48*	0.129*	0.71
✓	**A9995**	460	2	Acid		11.1		2.7	0.0788*	
	A3877	460		Zwitterion	9.2	4.5	1.94	2.8	0.09	1
✓	**A0619**	470	2	Weak Base	3.8		4.15		0.026	0.625
	A8734	470	2	Strong Base	8.4		6.79	5.14		
✓	**A0765**	480	2	Weak Base	1.5			2.38	0.056	0.58
✓	**A1260**	480	4	Acid		5	7.75		0.00138*	0.517
✓	**A3609**	480	2**	Strong Base	7.9		4.22	4 (8***)	0.025	0.64
✓	**A3336**	480	4**	Strong Base	7.71		5.78	5.74 (6.97***)	0.0082	0.602
	A1476	480		Zwitterion	12	1.8		-13	0.75	0.5
	A6215	480		Strong Base	7.27		3.55		0.011	0.886*
✓	**A9081**	490	2	Weak Base	4			2.2	0.05	0.61
	A4494	490	4	Neutral			2.24		0.0226	
✓	**A6882**	520	4	Neutral			6.88		0.0006	0.625
✓	**A6598**	530	2	Weak Base	5.96		4.14		0.0203*	
✓	**A0799**	530	2**	Strong Base	8.6		3.9		0.015	0.627****
	A4955	540	4	Strong Base	7.98			2.9	0.2	0.822*
	A1149	540		Ampholyte	5.5	10.2	8.77	7.4 (3.5***)	0.007	0.76
	A0714	550	2	Ampholyte	0.37	7.03	3.9	3.73	0.01	0.58
✓	**A4492**	550	4	Weak Base	2.34		1.984	1.5		
	A6229	560		Zwitterion	10.2	6.6		3.6		
✓	**A7294**	570	4	Strong Base	8.9		5.85		0.00392	0.6
	A0851	570		Weak Base	6.2			4.25 (4***)	0.034	0.715
	A8541	580	2	Weak Base	6.78		4.765	4.7	0.07	
	A2450	600	2	Weak Base	3.2		6.1	6.1	0.0018	0.63
✓	**A3427**	600	2	Weak Base	3.3		5.6	5.6	0.0055	0.67
	A9530	600	2	Zwitterion	7.4	3.5		3.4 (6.8***)	0.023	0.57
✓	**A7651**	600	3	Strong Base	9.98			2.6	0.104*	
	A1020	620		Weak Base	6.3		4.1	3		
✓	**A3028**	630	1**	Ampholyte	6.5	10.4	2.6	2.5	0.1	0.73****
	A2452	630		Neutral					0.99	0.587
✓	**A0633**	640	3	Strong Base	8.39		2.24	1.355	0.24	
	A0815	650	3	Weak Base	5.4		4.9	4.87 (6.8***)	0.0529*	

✓	**A2276**	720	2	Strong Base	8		3.3	2.7	0.1	0.708****
✓	**A5637**	870	4**	Zwitterion	10.7	6.3	4.13	0.35*	0.3	
	✓ Meets minimum criteria for inclusion in the simulation exercise * geomean (for fup or BP) or mean (for logP or logD) of multiple values ** Calculated based on fa scaled up from caco-2 and Do *** pH associated with listed LogD value (if different from 7.4) **** converted from Kp erythrocytes (Kpe) using formula Kpe*Ht + (1 – Ht), assuming haematocrit (Ht) of 0.45									

3.2.1. Physicochemical properties

The OrBiTo database displayed a median molecular weight (MW) of 440 g/mol (interquartile range [IQR]: 370-527.5; n=83 APIs). The selected simulation set displayed similar molecular weights (Median: 440, IQR: 370-525; n=43). The MW properties of the OrBiTo database were very similar to that of the PhRMA database, which displayed a median MW of 444 g/mol (range: 171-725; mean: 442.9 SD: 102, n=108). Compared to Benet and WHO datasets both the OrBiTo and PhRMA databases displayed a higher MW (Figure 4) (Kasim et al., 2004; Bu, 2006; Hosea et al., 2009; Benet et al., 2011; Poulin et al., 2011).

In terms of logP, the OrBiTo database showed similar tendencies to the PhRMA database and Bu (2006) dataset, demonstrating a median logP of 3.60 (IQR: 2.40-4.58; mean: 3.57 SD: 2.13; Figure **5**). However, the logP of the OrBiTo database was higher than the overall average logP of 2.43 from all the datasets, suggesting the OrBiTo database consisted of slightly more lipophilic compounds compared to other databases in the literature (Kasim et al., 2004; Bu, 2006; Hosea et al., 2009; Benet et al., 2011; Poulin et al., 2011).

Dividing the OrBiTo database compounds based on acid-base nature revealed basic compounds to be the most represented group constituting 48.2% of the total database, while 16.9% of the database were acidic compounds, 22.9% ampholytes and 12.1% neutral compounds. The selected simulation set displayed a majority of basic compounds (55.8%) and a reduction in the frequency of ampholytic compounds (9.3%) compared to the overall

database. The frequency distribution of acid-base nature in the database was consistent with remaining comparative datasets, where all comparators contained a majority of basic compounds ranging between 46.3 and 92.3% of the total datasets (Figure 6).

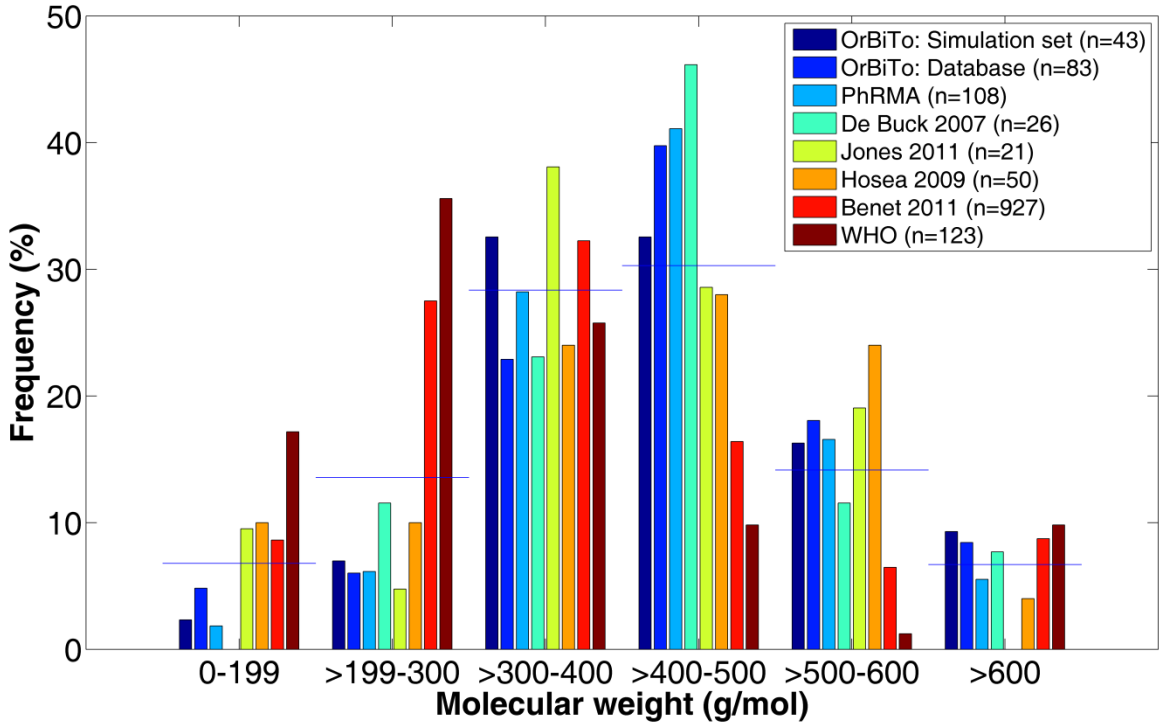


Figure 4. Frequencies of molecular weights of OrBiTo APIs as compared to the WHO essential drugs list, PhRMA Benet and Hosea databases (Kasim et al., 2004; Hosea et al., 2009; Benet et al., 2011; Poulin et al., 2011).

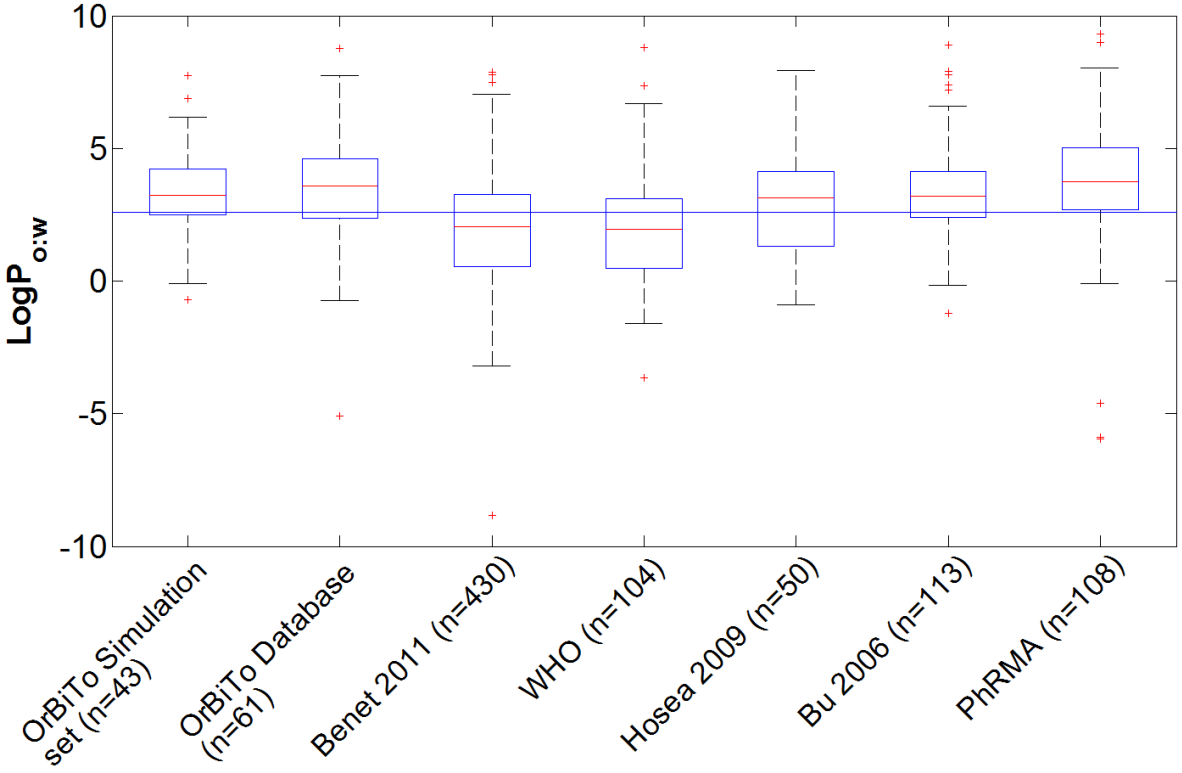


Figure 5. LogP of the OrBiTo APIs as compared to the WHO essential drugs list, PhRMA, Benet, Hosea and Bu databases (Kasim et al., 2004; Bu, 2006; Hosea et al., 2009; Benet et al., 2011; Poulin et al., 2011) .

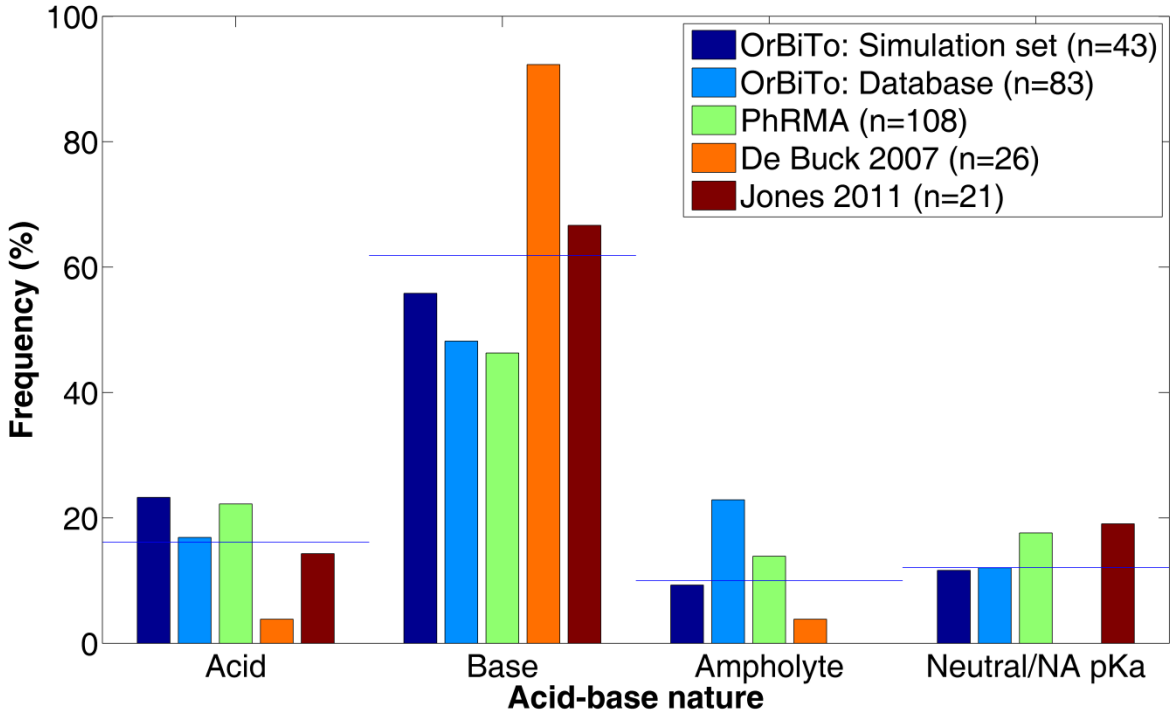


Figure 6. Frequencies of acid-base nature of OrBiTo APIs as compared to the PhRMA initiative dataset (De Buck et al., 2007; Jones et al., 2011; Poulin et al., 2011).

3.2.2. Permeability, solubility and BCS classification

The OrBiTo database mostly consisted of BCS class I and II compounds, representing 18.6% and 55.8% respectively of active substances in the drug library. Although the BCS class composition of the OrBiTo database was very similar to the PhRMA dataset there was a great underrepresentation of BCS class III compounds as compared to the WHO essential drugs list (OrBiTo: 6.78% vs. WHO: 38.5%). The final selection of APIs for the simulation exercise displayed a slightly higher representation of BCS IV compounds (simulation set: 20.9% vs. database: 17.0%), whereas BCS class I and II displayed slightly lower representation in the simulation set (Figure **7**) (Kasim et al., 2004; Poulin et al., 2011).

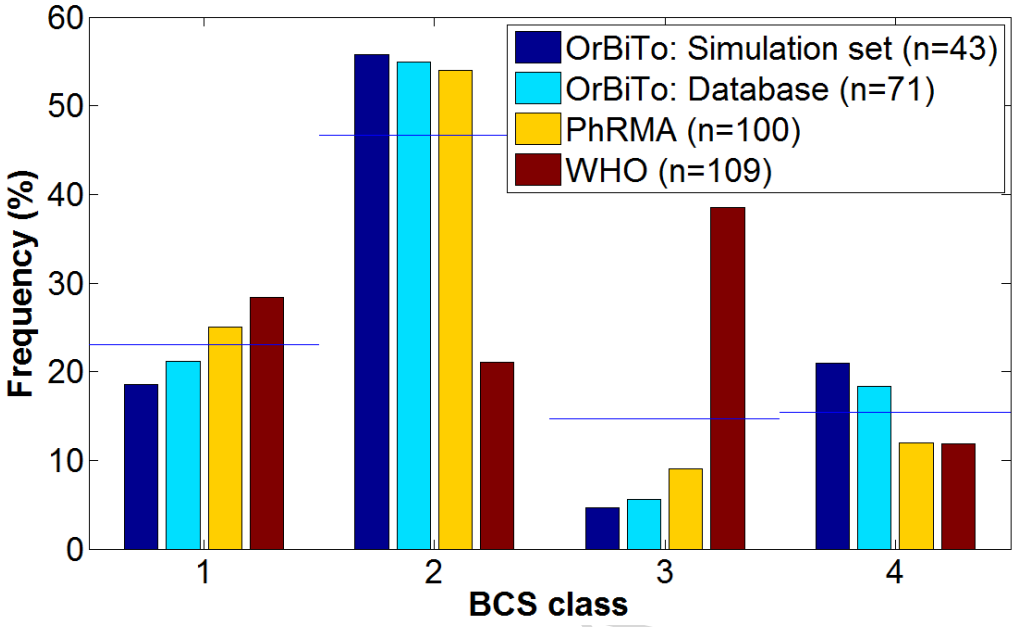


Figure 7**.** BCS (Biopharmaceutics Classification System) class of OrBiTo APIs as compared to the WHO essential drugs list and the PhRMA initiative dataset (Kasim et al., 2004; Poulin et al., 2011).

Preliminary estimations of D_o based on the highest oral dose in the clinical data, showed a majority of OrBiTo drug substances displaying a $D_o \geq 1$, namely 72.9%, suggesting most compounds in the database and simulation set were solubility limited. A majority of the compounds (74.4%) displayed what could be classified as high permeability with an estimated fa $\geq$ 90% based on IVIVE of Caco -2 permeability information provided in the database. This is consistent with the analysis of the BCS classification, with most compounds displaying high permeability-low solubility properties (Figure 8).

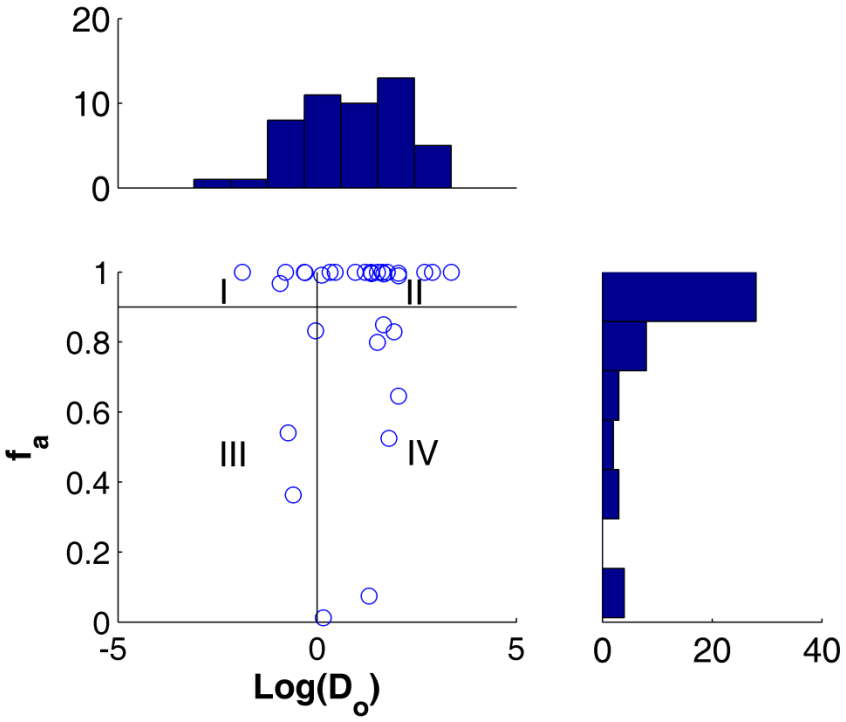


Figure 8: Distributions of fraction absorbed (fa) and dose number (Do) for the different APIs of the OrBiTo database.

3.2.3. Blood and plasma binding properties

The OrBiTo database contained a total of 108 reported blood-to-plasma ratios (BP) from multiple species, of which 38.0% came from human *in vitro* experiments. The most commonly presented pre-clinical species sources of BP were rat and dog representing 25.0% and 20.4% respectively of total reported BP values. The OrBiTo database displayed a median human BP of 0.62 (IQR: 0.57-0.71), with a mean of 0.73 (SD: 0.48). On average, human BP for the OrBiTo database APIs fell below other databases from the literature, with the overall average BP of 0.86 (Figure 9).

The OrBiTo database contained 193 reported values of fraction unbound in plasma (fu_p) based on human and pre-clinical *in vitro* experiments. Human fu_p was the most prevalent reported value representing 31.1% of all reported fu_p's, followed by rat and dog representing 24.4% and 19.2% respectively. Human fu_p values displayed a high variability with a median

fu_p of 0.05 and IQR of (0.01 , 0.17). A similarly wide range was seen in comparative datasets. The average fu_p of the OrBiTo database was 0.16, which was lower as compared to comparative data sets, displaying an overall average fu_p of 0.30 for all datasets (Figure 10) (De Buck et al., 2007; Hosea et al., 2009; Jones et al., 2011).

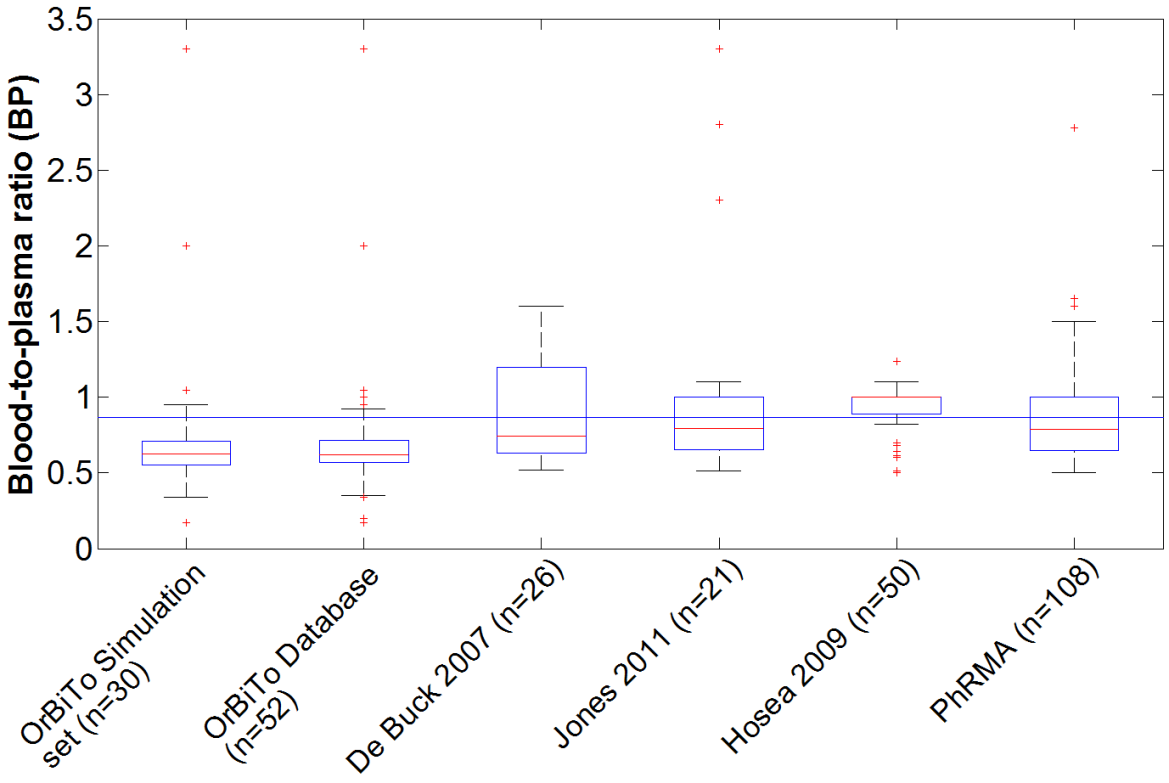


Figure 9**.** Blood-to-plasma ratios (BPs) of OrBiTo APIs as compared to De Buck, Jones and Hosea databases (De Buck et al., 2007; Hosea et al., 2009; Jones et al., 2011).

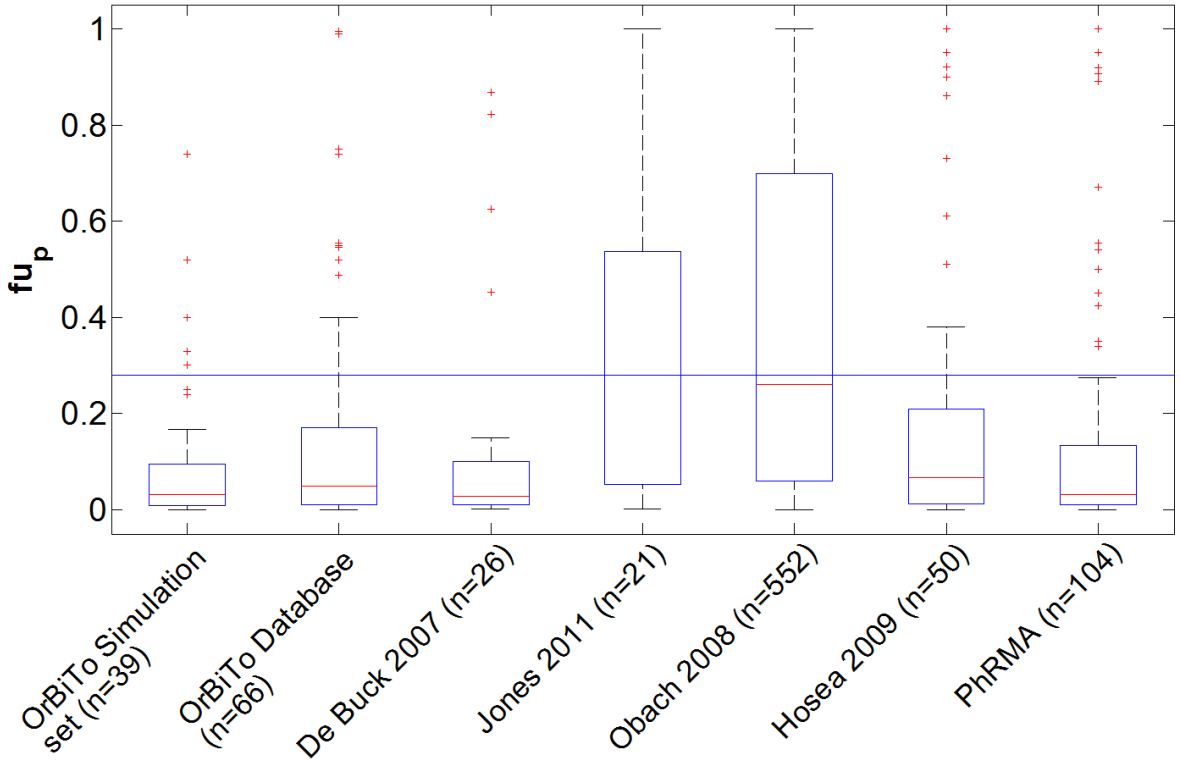


Figure 10. Fraction unbound in plasma (fu_p) of OrBiTo APIs as compared to De Buck, Jones, Obach, Hosea PhRMA databases (De Buck et al., 2007; Obach et al., 2008; Hosea et al., 2009; Jones et al., 2011; Poulin et al., 2011).

3.3. Characterisation of the OrBiTo Database Studies

3.3.1. Studies designs

At the time of the simulation exercise, the database contained 455 clinical and preclinical PK studies, consisting of a total of 1,476 study arms. Of these study arms, 964 were human and represented a total of 26,469 individuals studied. The human study arms comprised close to two thirds of the total number of PK study arms, with rats and dogs representing the next most frequent species studied at 16% and 12% respectively (Table 2).

The number of PK study arms with orally administered drug clearly outweighed those administering intravenous (i.v.) formulations, for both human and preclinical species. While humans represented the majority of the study arms in the database (65%), only 25% of the i.v. study arms were in humans. The most frequent species in i.v. study arms was the rat at 35%, while dogs represented 23%, monkeys 10% and mice 5% (Table 2).

Table 2: Number of PK study arms by species

Species	Number of PK Study Arms	Number of i.v. Study Arms	Number of p.o. Study Arms
human	964	50	892
rat	241	69	161
dog	183	45	135
monkey	43	20	23
mice	21	10	10
minipig	15	4	11
rabbit	8	0	8
mormoset	1	1	0
PK = pharmacokinetic; i.v. = intravenous; p.o. = per oral			

Each API file in the database contained a varied number of PK study arms, ranging from 0 to 57 study arms (0 to 46 human and 0 to 35 preclinical study arms) per API (Figure **11**). The median number of PK study arms per API was 16 (10 human, 6 preclinical), and the mean was 17.8 (11.6 human, 6.17 preclinical). 4 APIs contained no human study arms, and 26 APIs contained no preclinical study arms (3 APIs contained neither human nor preclinical study arms). Of the APIs that contained at least one preclinical PK study, the median number of preclinical study arms per API was 8 and the mean was 8.98. Figure **11** displays the distributions of preclinical and human PK study arm counts per API.

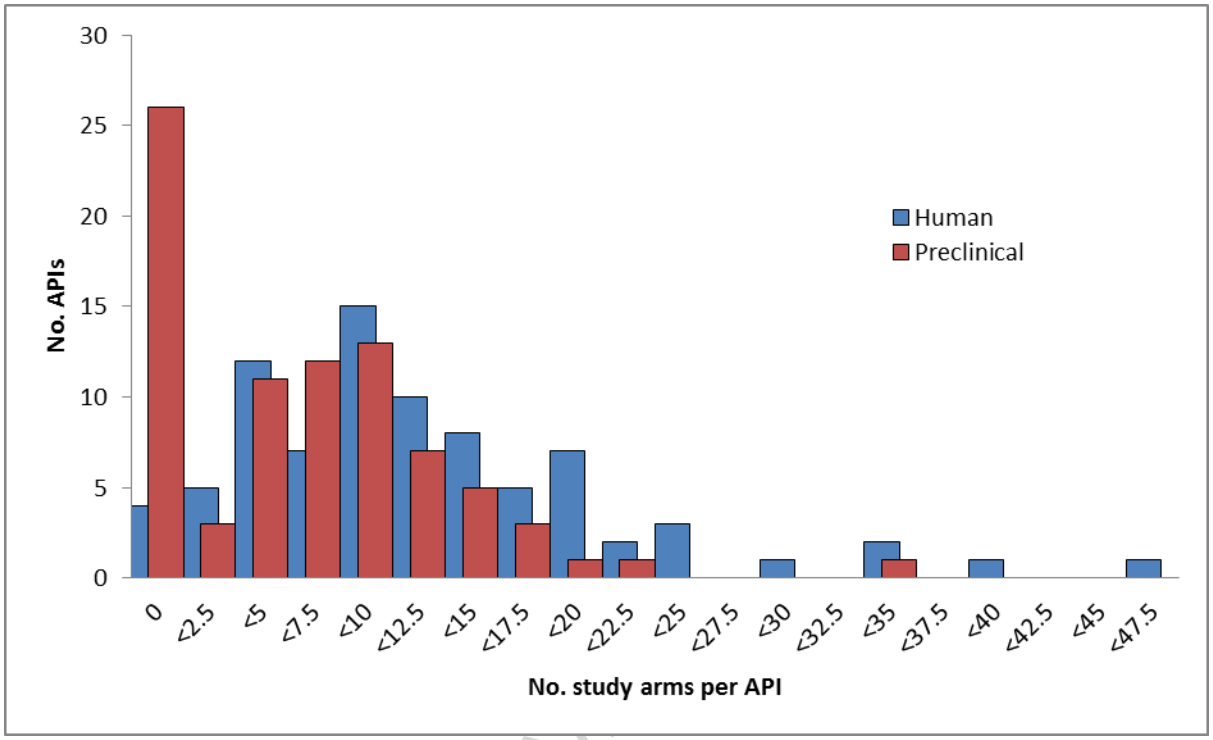


Figure 11. Distribution of preclinical and human PK study arm counts per API. The horizontal axis represents the count of PK study arms per API and the vertical axis represents the number of APIs in each range.

Four different PK study types were identified in the database: i.v. vs. oral studies, ascending dose studies, formulation finding studies, and fasted vs. fed studies. i.v. vs. oral studies were defined as studies containing at least one intravenous and one oral study arm. Ascending dose studies were defined as studies containing two or more administrations of the same formulation in the same prandial state at different doses. Formulation finding studies were defined as studies containing two or more administrations of different formulations in the same prandial state. Fasted vs. fed studies were defined as studies containing at least one fasted and one fed study arm. 360 of the 455 studies fell into one or more of these categories (Table 3). Most of the remaining studies contained single arms with no comparison arm (77 out of the 95 unclassified study arms, or 81%). The number of APIs containing each type of study was also determined, and the results are displayed in Table 3 and Figure **11**.

Table 3. Number of PK studies of different types in the OrBiTo database and the number of APIs containing each study type.

	Number of PK Studies		Number of APIs	
	Preclinical	Human	Preclinical	Human
i.v. vs. Oral Studies	84	16	45	15
Ascending Dose Studies	57	110	32	65
Formulation Finding Studies	27	93	20	45
Fasted vs. Fed Studies	14	73	9	43
Not classified	58	37		
Total*	188	267	57	79

*Note that some studies fell into more than one category, and most APIs contained more than one study type, thus the sum of each column is greater than the total number of studies or APIs.i.v. = intravenous; PK = pharmacokinetic; API = active pharmaceutical ingredient

The most frequent preclinical PK study type was i.v. vs. oral (45%) followed by ascending dose (30%), formulation finding (14%) and fasted vs. fed (7.4%). In contrast, the most frequent human PK study type was ascending dose (63%) followed by formulation finding (35%) and fasted vs. fed (27%), with i.v. vs. oral studies comprising only 6%. 45 of the 83 (54%) APIs in the database contained i.v. vs. oral studies in preclinical species, whereas only 15 (18%) contained human i.v. vs. oral studies (5 additional APIs contained preclinical and 12 contained human i.v. study arms without oral study arms in the same study, thus a total of 50 APIs (60%) had preclinical and 27 APIs (32%) had human i.v. studies). The frequency of APIs containing ascending dose studies was 38.6% and 78.3% for preclinical and clinical studies, respectively. For formulation finding studies, 24.1% and 54.2%, and fasted vs. fed studies 10.8% and 51.8%.

3.3.2. Formulations properties and observed bioavailability

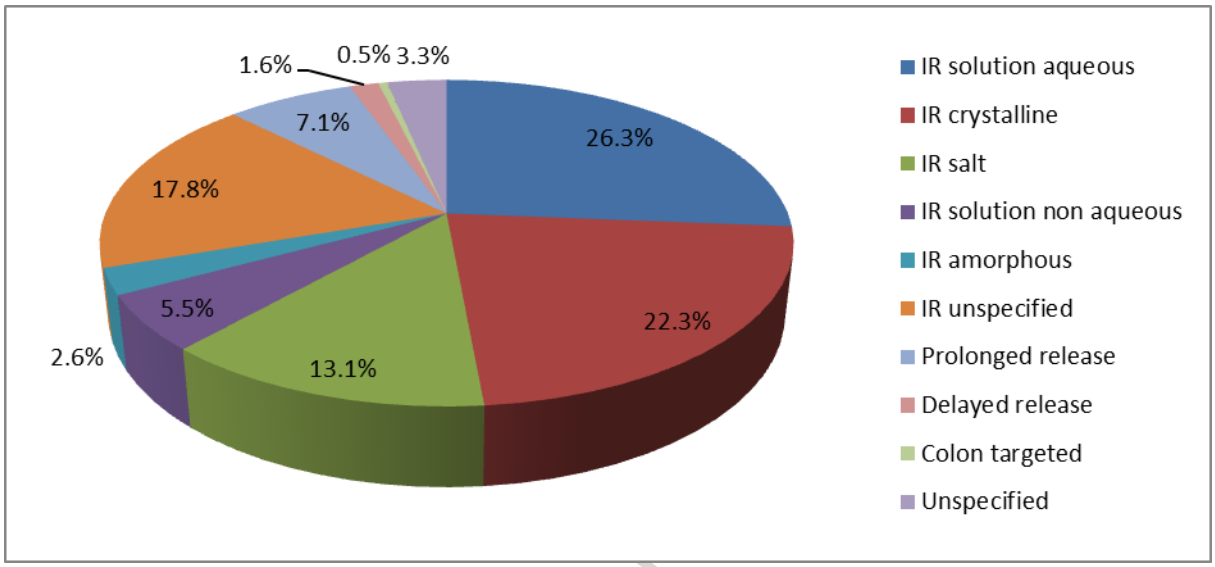


Figure 12: Physical states of orally administered formulations utilised in study arms of the database as of March 2014. IR = immediate release

The database contained a variety of oral formulations administered in the different studies. Immediate release made up the majority, 87.6%, of these formulations with only 7.1% prolonged release, 1.6% delayed release, and 0.5% colon targeted formulations (Figure 12). Of the immediate release formulations, aqueous solutions made up 30%, followed by 25% crystalline and 15% salt. The median bioavailability of immediate release formulations and immediate release solutions for APIs in the OrBiTo database is displayed in Table 4. The bioavailability for immediate release formulations was similar in rat, dog and human (0.446, 0.456, and 0.416 respectively) compared with the PhRMA initiative (0.385, 0.497, and 0.48). However, median bioavailability in monkeys was somewhat higher in the OrBiTo database (0.629) compared with PhRMA (0.356) (Poulin et al. (2011)).

Table 4. Fasted state bioavailability for immediate release formulations of APIs in the database, organised by species.

	Bioavailability (F)					
	IR Formulations			Solutions		
	n	Median	(25^{th}, 75^{th})	n	Median	(25^{th}, 75^{th})
Database						
Rat	34	0.446	(0.281,0.704)	27	0.554	(0.286,0.738)
Dog	32	0.456	(0.332,0.79)	24	0.634	(0.383,0.854)
Monkey	6	0.629	(0.31,0.757)	6	0.533	(0.3,0.719)
Human	22	0.415	(0.203,0.724)	15	0.399	(0.155,0.735)
Simulation Set						
Rat	19	0.581	(0.303,0.778)	16	0.588	(0.27,0.796)
Dog	17	0.598	(0.365,0.78)	14	0.673	(0.391,0.878)
Monkey	5	0.704	(0.554,0.774)	5	0.554	(0.513,0.774)
Human	19	0.45	(0.245, 0.793)	12	0.51	(0.175,0.911)
IR = immediate release; n = number of APIs; (25^{th}, 75^{th}) = (25^{th}, 75^{th}) percentiles						

3.3.3. *In Vivo* Clearance and Volume of Distribution

Clearance and volumes of distribution were calculated for all i.v. study arms, in both preclinical and human subjects, following the prospective PBPK exercise and unblinding of plasma profiles. For APIs which had more than one i.v. study arm for a particular species, the geometric means of the parameters were taken. The median, 25^{th} and 75^{th} percentiles of the parameter values for the APIs in the database and the simulation set are displayed in Table 5.

Table 5. Geometric mean of clearance and volumes of distributions for preclinical i.v. studies in the database compared with the simulation set, organised by species. Values expressed as median and (25^{th}, 75^{th}) percentiles and n represents number of APIs.

		CL_{iv} (L/h/kg or L/h)*		**$V_{d,ss}$ (L/kg or L)***	
	n	Median	(25^{th}, 75^{th})	Median	(25^{th}, 75^{th})
Database					
Rat	45	1.224	(0.46,2.369)	2.728	(1.171,5.258)
Dog	40	0.533	(0.239,1.726)	3.092	(1.304,7.799)
Monkey	9	0.699	(0.293,0.721)	6.672	(0.9,8.32)
Human**	23	16.9	(11.6,43.6)	80.8	(54.5,239)
Simulation Set					
Rat	27	1.42	(0.642,2.956)	3.832	(1.691,5.443)
Dog	21	0.677	(0.311,1.89)	3.333	(1.657,8.225)
Monkey	7	0.684	(0.219,0.72)	6.672	(1.09,8.439)
Human**	19	17.9	(11.9,45.1)	88.2	(54.5,277)

***All preclinical clearances are expressed as L/h/kg and volumes as L/kg. All human clearances are expressed as L/h and volumes as L.**
****Note that human clearance values and volumes of distribution were not available during the simulation exercise, as they were calculated based on plasma profiles revealed after the unblinding of the database.**

Human CL values for the entire database (median 16.9, 25^{th} and 75^{th} percentiles 11.6 and 43.6 L/h, respectively) and the simulation set (median 17.9, 25^{th} and 75^{th} percentiles 11.9 and 45.1 L/h) were comparable to the database of 670 compounds reported in Obach et al. (2008), for which the median, 25^{th} and 75^{th} percentiles were 18, 7.65 and 45 L/h, (converted from mL/min/kg to L/h assuming 75kg body weight). CL values were also comparable to those in Poulin et al. (2011), which reported median i.v. CL in humans of 21.1 L/h (converted from % of liver blood flow), and De Buck et al. (2007) which reported median i.v. CL of 25.4 L/h and 25^{th}, 75^{th} percentiles of 16.5 and 42.1. Preclinical CL values in the OrBiTo database (median 1.224, 0.533, 0.699 L/h/kg for rat, dog and monkey) were also comparable to those in Poulin et al. (2011), which reported median i.v. CL of 1.24, 0.745, 0.756 L/h/kg in rat, dog,

and monkey (converted from % of liver blood flow using liver blood flows of 1.014, 18.54, and 13.08 L/h, and assuming body weights of 0.25, 10, and 4.5 kg for rat, dog and monkey).

Human V_d for the entire database (median 80.8, 25th and 75th percentiles 54.5 and 239 L, respectively) and the simulation set (median 88.2, 25th and 75th percentiles 54.5 and 277 L) were slightly elevated compared to those reported in Obach et al. (2008), for which the median, 25th and 75th percentiles were 72, 22.5 and 195 L (converted from L/kg to L assuming 75kg body weight). However, the median V_d in the OrBiTo database was less than that reported for humans in Poulin et al. (2011) (127.5 L, converted from L/kg assuming 75kg body weight).

4. Discussion

The setup of OrBiTo database represents a major effort by the involved institutions to create a database of pharmaceutical compounds along with drug- and formulation-specific parameter information, clinical and preclinical data with a particular focus on oral biopharmaceutics formulations and studies. The database had several design features that allowed for recording a multitude of relevant information related to drug substances, formulations, and preclinical and clinical studies. One such example was the recording of different drug substances and formulations of the same compound, such as the possibility of different solubilities for different salt forms or different excipients used in different formulations. Experimental solubility measurements were supplemented with information on the composition of solubility media such as the specifications used for simulated fasted or fed state intestinal fluid. While experimental caco-2 measurements were accompanied by values for references compounds, users of the database should be wary that other *in vitro* assays (e.g. clearance, solubility, dissolution) did not have such information to account for inter-lab differences or allow for the establishment of *in vitro in vivo* correlations. The database also

allowed for recording of co-administered medication and altered disease state in clinical studies, however, most of the studies recorded at the time of this gap analysis were in healthy volunteers.

The criteria for inclusion of APIs for simulation in the prospective PBPK simulation exercise were selected partly based on the minimum required parameters to run the simulation software programs, and do not represent in the least the exhaustive list of input parameters that may be necessary for accurate PBPK predictions. Considerable missingness was still apparent in the simulation set. Out of a total of 83 APIs only three initially fulfilled the minimum inclusion criteria set up independently by modellers at the University of Manchester, Sanofi and SimCYP Ltd. Relaxation of the inclusion criteria expanded the simulation set to 43 APIs by allowing estimation of BP, allowing for preclinical i.v. informed CL estimation, and allowing APIs without i.v. PK studies. Apart from these relaxed inclusion criteria, many APIs also had significant missingness in the area of biopharmaceutics related drug and formulation-specific parameters, such as solubility vs. pH profiles, biorelevant solubility measures, particle radii for solid oral formulations and more. In order to model more complex scenarios, such as dissolution, formulation effects and precipitation, it is essential to have such information.

Information related to biorelevant media, dissolution rate, particle size and participation rate is highly relevant for appropriate characterisation of oral drug absorption. The level of missingness for this information was surprising especially considering that regulatory agencies typically require information on dissolution for example as a part of bioavailability (BA) and bioequivalence (BE) studies. However, it is possible that the database contains drugs that never went through BA/BE studies for the purposes of regulatory submissions. The stage of development for the drugs included in the database is not clear, though the majority

appear to be first in human, dose escalation studies. Further investigation is required to determine the reason for the missing information.

The focus of the OrBiTo project is on oral bioavailability and biopharmaceutics, thus the key interest of the simulation exercise is the predictive ability of the *in silico* methods to predict release, dissolution, permeation, and intestinal first-pass. In the evaluation of the predictive success of these processes, the availability of clinical i.v. data is important, as without this it is difficult to untangle these processes from systemic elimination and distribution. Only 27 of the 83 APIs in the database contained human i.v. study arms, 15 of which had i.v. and oral study arms in the same study. However, the missingness of i.v. study arms is not surprising for a dataset intended to focus on oral biopharmaceutics and formulation properties, as compounds that are ultimately intended for oral administration may never have an i.v. formulation developed. However, many pharmaceutical companies have recently made note of the importance of clearance, and have started including micro-dosing studies in their drug development plans for oral compounds (Rowland, 2012). Despite the sparseness of APIs with i.v. data in the OrBiTo database, APIs without i.v. data had multiple oral formulations available for simulation thus allowing the testing of relative bioavailability between formulations.

While the focus of the OrBiTo project was on oral biopharmaceutics, analysis of the oral formulations in the database revealed a sparseness of nonstandard orally administered formulations (e.g. nanoparticles, microspheres, solid dispersions, and self-emulsifying delivery systems) that would have been of great relevance to the goals of the OrBiTo initiative. The large majority of oral formulations administered were immediate release (87.6%), with only 7.1% prolonged release formulations, 1.6% delayed release, and 0.5% colon targeted formulations. However, the high proportion of immediate release formulations in these numbers could be related to their overrepresentation in certain study designs, such as

ascending dose studies. To provide a better context, 16.9% of the APIs in the database were administered as at least one of a prolonged, delayed, or colon target formulation, which gives a better picture of the availability of these relevant oral formulations in the database.

The comparison of the OrBiTo database to other previously published large drug datasets concluded the OrBiTo dataset to be a representative example of available drugs. Any differences may be related to the fact that the OrBiTo database reflects the research and development (R&D) portfolio, including legacy compounds and terminated projects, while, for example, the WHO list represents what is available on the market. One advantage of the OrBiTo database was its good representation of solubility-limited compounds (BCS class II), the inclusion of which has the advantage of allowing the testing of PBPK absorption models in the saturated range of solubility where one would expect more complex dissolution behaviour, and thus the possibility of poor prediction results. It also had a large representation of basic compounds, which when combined with solubility limitations have the potential to be subject to precipitation upon entering the small intestine.

Possibly one of the greatest strengths of the OrBiTo database is the multitude of APIs associated with study designs relevant to oral biopharmaceutics development such as bioavailability studies (18% of APIs), ascending dose studies (78.3%), formulation finding, (54.2%) and fasted vs. fed studies (51.8%), a focus that was not present in previous databases. Another key advantage of the database was to provide a representative sample of the R&D profile, including legacy compounds, while still enabling a completely prospective analysis of available PBPK software via the blinded nature of the database. Further details of this exercise can be found in our companion papers (Margolskee et al. – Part 2 – Submitted; Darwich et al. – Part 3 – Submitted).

5. Conclusion

The OrBiTo database provides a unique opportunity to perform a large scale evaluation of the PBPK approach to predicting oral drug bioavailability and formulation effects in human. A similar effort to test the ability of predicting biopharmaceutics has not been seen up to date. The database was found to be largely representative of previously published pharmaceutical compound datasets and reflects the API portfolio in industry R&D. Criteria for the APIs to be included in the simulation exercise were selected partly based on the parameters required for running the software programs to be tested. These criteria do not represent an exhaustive list of input parameters necessary for accurate PBPK predictions, and many APIs included in the simulation set were still suffering from significant missingness. Even in the presence of data, the quality of that information is not guaranteed. This lack of data richness has the potential to adversely affect the interpretability of the outcome of the simulation exercise. This highlights a clear need for utilising data rich examples in a systematic test of the effects of input parameters on PBPK predictions.

6. Acknowledgements

This work was performed under the OrBiTo Project, which has received support from the Innovative Medicines Joint Undertaking (http://www.imi.europa.eu) under Grant Agreement No. 115369. The authors would also like to acknowledge all participants who contributed to the OrBiTo database of APIs (Table 6; Table 7) and to the simulation exercise (Table 8), especially those who could not be named in the author list.

Table 6: List of participants that contributed to the design and implementation of the OrBiTo database architecture

Name	Affiliation(s)	Contribution(s)

Xavier Pepin	Sanofi/AstraZeneca	
Kristin Lacy-Jones	SimCYP	Scientist's training, establishment of web platform, database architecture, database maintenance and technical support
Philip Hayward	SimCYP	Excel plugin design, development, maintenance of web platform & database and training/support
Steve Andrews	SimCYP	Database architecture, establishment of web platform, database maintenance and technical support
Susan Burkhill	SimCYP	Grant co-ordinator and financial accounting for Simcyp

Table 7: List of participants that contributed to the OrBiTo database of APIs.

Name	**Affiliation(s)**	**Contribution(s)**
Jonas Angstenberger	AbbVie	Collected compound data from company's internal databases, entered compound data into API plugin, uploaded API data to the database
Franziska Graf	AbbVie	Collected compound data from company's internal databases, entered compound data into API plugin, uploaded API data to the database
Loic Laplanche	AbbVie	Collected compound data from company's internal databases, PI for AbbVie (contributing to Abbvie API selection; Get clearance from the IP and legal approval from the company);
Thomas Müller	AbbVie	Collected compound data from company's internal databases, entered compound data into API plugin, uploaded API data to the database
Sara Carlert	AstraZeneca	Collected compound data from company's internal databases, entered compound data into API plugin, uploaded API data to the database
Pankaj Daga	AstraZeneca	Collected compound data from company's internal databases, entered compound data into API plugin,
Donal Murphy	AstraZeneca	Collected compound data from company's internal databases, entered compound data into API plugin,
Christer Tannergren	AstraZeneca	Collected compound data from company's internal databases, entered compound data into API plugin. Leader of AZ activities (allocating scientists etc)
Mohammed Yashin	AstraZeneca	Collected compound data from company's internal databases, entered compound data into API plugin,
Susanne Greschat-Schade	Bayer Pharma AG	Contributed to API selection considering overall Orbito goals. Collected compound data (animal PK data) from company's internal databases, entered compound data into API plugin
Wolfgang Mück	Bayer Pharma AG	Head of Clinical PK Dpmt. (contributed to API selection considering overall Orbito goals)
Uwe Muenster	Bayer Pharma AG	Contributed to API selection considering overall Orbito goals. Collected compound data (physchem data API & formulations) from company's internal databases, entered compound data into API plugin, uploaded API data to the database; PI (Orbito Lead Scientist / Representative for Bayer)
Andreas Ohm	Bayer Pharma AG	Head of Formulations Dpmt. (contributed to API selection considering overall Orbito goals)
Dorina van der Mey	Bayer Pharma AG	Contributed to API selection considering overall Orbito goals. Collected compound data (human PK data) from company's internal databases, entered compound data into

		API plugin
Kerstin Julia Frank	Boehringer Ingelheim Pharma GmbH & Co KG	Collection of data, entering into plugin, uploading to database, PI for Boehringer Ingelheim
Alexander Staab	Boehringer Ingelheim Pharma GmbH & Co KG	Data collection
Peter Stopfer	Boehringer Ingelheim Pharma GmbH & Co KG	Data collection
Peter Sieger	Boehringer Ingelheim Pharma GmbH & Co KG	Data collection
Jeannine Fleth-James	Boehringer Ingelheim Pharma GmbH & Co KG	Data collection
Richard Lloyd	GlaxoSmithKline	Collected compound data from company's internal databases, entered compound data into API plugin, uploaded API data to the database
Lieve Adriaenssen	Janssen	data collected, data entered into API plugin
Jan Bevernage	Janssen	data collected, data entered into API plugin, uploaded data to database, point of contact for data base
Loeckie De Zwart	Janssen	data collected
Dominique Swerts	Janssen	data collected, data entered into API plugin
Christophe Tistaert	Janssen	data collected, data entered into API plugin
An Van Den Bergh	Janssen	data collected
Achiel Van Peer	Janssen	data collected, data entered into API plugin
Stefania Beato	Novartis	Selected API to be uploaded in the database. Get clearance from the IP and legal approval from the company. Collected compound data from company's internal databases, entered compound data into API plugin.
Anh-Thu Nguyen-Trung	Novartis	Collected compound data from company's internal databases, entered compound data into API plugin, uploaded API data to the database. Answered questions from the modeler and updated data in database.
Joanne Bennett	Pfizer	Collected compound data from company's internal databases, entered compound data into API plugin, uploaded API data to the database
Mark McAllister	Pfizer	PI for Pfizer, responsible for compound selection and internal data approval and release
Mei Wong	Pfizer	Contributed to data collation and selection of compounds for submission
Patricia Zane	Sanofi	Collected compound data from company's internal databases, entered compound data into API plugin, uploaded API data to the database
Céline Ollier	Sanofi	Collected compound data from company's internal databases, entered compound data into API plugin,
Pascale Vicat	Sanofi	Collected compound data from company's internal databases, entered compound data into API plugin
Markus Kolhmann	Sanofi	Collected compound data from company's internal databases, entered compound data into API plugin
Alexander Marker	Sanofi	Collected permeability data from company's internal databases, entered compound data into API plugin
Priscilla Brun	Sanofi	Collected permeability compound data from company's internal databases, entered compound data into API plugin
Florent Mazuir	Sanofi	Collected permeability compound data from company's

		internal databases, entered compound data into API plugin
Stéphane Beilles	Sanofi	Collected permeability compound data from company's internal databases, entered compound data into API plugin
Marta Venczel	Sanofi	Collected permeability compound data from company's internal databases, entered compound data into API plugin
Xavier Boulenc	Sanofi	Collected permeability compound data from company's internal databases, entered compound data into API plugin
Petra Loos	Sanofi	Collected permeability compound data from company's internal databases, entered compound data into API plugin
Xavier Pepin	Sanofi	Collected permeability compound data from company's internal databases, entered compound data into API plugin, uploaded API data to the database

Table 8: List of participants who contributed to the OrBiTo simulation exercise

Name	Affiliation(s)	Contributions(s)
Leon Aarons	University of Manchester	PI for University of Manchester
Adam S. Darwich	University of Manchester	Performed gap analysis on database, performed simulations, analysed results
Aleksandra Galetin	University of Manchester	PI for University of Manchester
Alison Margolskee	University of Manchester	Performed gap analysis on database, performed simulations, analysed results
Amin Rostami-Hodjegan	University of Manchester/SimCYP	Work package co-leader, PI for University of Manchester, PI for SimCYP
Sara Carlert	AstraZeneca	Performed simulations, analysed results
Maria Hammarberg	AstraZeneca	Performed simulations
Constanze Hilgendorf	AstraZeneca	Performed simulations
Pernilla Johansson	AstraZeneca	Performed simulations
Eva Karlsson	AstraZeneca	Performed simulations
Donal Murphy	AstraZeneca	Performed simulations
Christer Tannergren	AstraZeneca	Performed simulations
Helena Thörn	AstraZeneca	Performed simulations
Mohammed Yasin	AstraZeneca	Performed simulations
Florent Mazuir	Sanofi	Performed simulations
Olivier Nicolas	Sanofi	Performed simulations, analysed results
Xavier Pepin	Sanofi/AstraZeneca	Work package co-leader, PI for Sanofi until March 2015, performed gap analysis on database, performed simulations, analysed results
Sergej Ramusovic	Sanofi	Performed simulations
Christine Xu	Sanofi	Performed simulations
Shriram M. Pathak	SimCYP	Performed gap analysis on database, performed simulations, analysed results
Timo Korjamo	Orion Pharma	Performed simulations, analysed results
Johanna Laru	Orion Pharma	Performed simulations
Jussi Malkki	Orion Pharma	Performed simulations, analysed results

Sari Pappinen	Orion Pharma	Analysed results
Johanna Tuunainen	Orion Pharma	Analysed results
Jennifer Dressman	Goethe University	PI for Goethe University
Carmen Gött	Goethe University	Analysed results
Simone Hansmann	Goethe University	Performed simulations, analysed results
Edmund Kostewicz	Goethe University	PI for Goethe University
Handan He	Novartis	Performed simulations, analysed results
Tycho Heimbach	Novartis	Performed simulations, analysed results
Fan Wu	Novartis	Performed simulations, analysed results
Carolin Hoft	AbbVie	Performed simulations
Loic Laplanche	AbbVie	PI for Abbvie, analysed results
Yan Pang	AbbVie	Performed simulations
Michael B. Bolger	Simulations Plus	PI for Simulations Plus, lead for analysis of impact of solubility and dissolution
John DiBella	Simulations Plus	Financial and time accounting for Simulations Plus
Eva Huehn	Simulations Plus	Performed gap analysis on database, performed simulations
Viera Lukacova	Simulations Plus	Co-PI for Simulations Plus
James M. Mullin	Simulations Plus	Performed gap analysis on database, performed simulations
Ke X. Szeto	Simulations Plus	Performed gap analysis on database, performed simulations
Chester Costales	Pfizer	Performed simulations
Jian Lin	Pfizer	Performed simulations
Mark McAllister	Pfizer	Performed simulations
Sweta Modi	Pfizer	Performed simulations
Charles Rotter	Pfizer	Performed simulations
Manthena Varma	Pfizer	Performed simulations
Mei Wong	Pfizer	Performed simulations
Amitava Mitra	Merck Sharp & Dohme (MSD)	Performed simulations, analysed results
Jan Bevernage	Janssen	Performed simulations
Jeike Biewenga	Janssen	Performed simulations
Achiel Van Peer	Janssen	Performed simulations
Richard Lloyd	GlaxoSmithKline	Performed simulations, analysed results
Carole Shardlow	GlaxoSmithKline	Performed simulations, analysed results
Peter Langguth	University of Mainz	PI for University of Mainz
Irina Mishenzon	University of Mainz	Performed simulations
Mai Anh Nguyen	University of Mainz	Performed simulations
Jonathan Brown	Bristol-Myers Squibb	Performed simulations

Appendix

For grouping based on BCS class, BCS class reported in the API data file was used. If no BCS class was given then estimations of fraction absorbed (f_a) and dose number (D_o) were used to assign classification, with estimated $f_a \geq 0.9$ signifying highly permeable (BCS 1 & 2) and $D_o \leq 1$ highly soluble (BCS 1 & 3) compounds. *In vitro* permeability measured in Caco-2 cell monolayers was scaled to P_{eff} using a power-model fit to lab-specific Caco-2 apparent permeability (P_{app}) of reference compounds and their associated P_{eff} values measured via Loc-I-Gut (Lennernas et al., 1997).

$$P_{eff} = a \cdot (P_{app})^b$$

Equation 1

These calibrated P_{eff} values were then converted to estimated f_a (**Error! Reference source not found.**), and grouping of APIs by f_a was carried out according to the BCS f_a cut-off point of 0.9 (Amidon et al., 1995; Yu and Amidon, 1999).

$$f_a = 1 - \left(1 + 0.54 \cdot P_{eff}\right)^{-7}$$

Equation 2

Calculations of D_o were based on the maximum reported dose (M_o) in the clinical data set of the corresponding API file. The solubility (C_s) was informed by the minimum reported aqueous solubility over the physiological pH range and the concomitant fluid intake (V_o) was assumed to be 250 mL (**Error! Reference source not found.**; (Amidon et al., 1995).

$$D_o = \frac{M_o / V_o}{C_s}$$

Equation 3

In cases where temperature of the solubility measure and melting point were given, solubility was corrected to estimated solubility at 37°C using Equation 4, where λ is the nonideality of system, x_A is the mole fraction of the solute along the saturation line (sat); h is the enthalpic factor, T is the temperature in Kelvin (K) and T_m is the melting point.

$$ln[1 + \lambda \cdot (1 - x_A)/x_A]_{sat} = \lambda \cdot h \cdot (T^{-1} - T_m^{-1})$$

Equation 4

Assuming an ideal system (λ=1) **Error! Reference source not found.** can be rearranged to give the solubility x_s at a given temperature T_s, where x_{ref} is a reference solubility measured at temperature T_{ref} (**Error! Reference source not found.**; (Tannergren et al., 2003).

$$ln(x_s) = ln(x_{ref}) \cdot \frac{T_{ref} \cdot (T_m - T_s)}{T_s \cdot (T_m - T_{ref})}$$

Equation 5

References

Amidon GL, Lennernas H, Shah VP, and Crison JR (1995) A theoretical basis for a biopharmaceutic drug classification: the correlation of in vitro drug product dissolution and in vivo bioavailability. *Pharm Res* 12:413-420.

Benet LZ, Broccatelli F, and Oprea TI (2011) BDDCS Applied to Over 900 Drugs. *AAPS J* 13:519-547.

Bu HZ (2006) A literature review of enzyme kinetic parameters for CYP3A4-mediated metabolic reactions of 113 drugs in human liver microsomes: structure-kinetics relationship assessment. *Curr Drug Metab* 7:231-249.

Chang WW and Leblond CP (1971) Renewal of the epithelium in the descending colon of the mouse. I. Presence of three cell populations: vacuolated-columnar, mucous and argentaffin. *Am J Anat* 131:73-99.

De Buck SS, Sinha VK, Fenu LA, Nijsen MJ, Mackie CE, and Gilissen RA (2007) Prediction of human pharmacokinetics using physiologically based modeling: a retrospective analysis of 26 clinically tested drugs. *Drug Metab Dispos* 35:1766-1780.

FDA (2000) Guidance for Industry: Waiver of In Vivo Bioavailability and Bioequivalence Studies for Immediate-Release Solid Oral Dosage Forms Based on a Biopharmaceutics Classification System.

Futagami S, Yamawaki H, Izumi N, Shimpuku M, Kodaka Y, Wakabayashi T, Nagoya H, Shindo T, Kawagoe T, Gudis K, Itoh T, and Sakamoto C (2013) Impact of sleep disorders in Japanese patients with functional dyspepsia (FD): nizatidine improves

clinical symptoms, gastric emptying and sleep disorders in FD patients. *J Gastroenterol Hepatol* 28:1314-1320.

Hosea NA, Collard WT, Cole S, Maurer TS, Fang RX, Jones H, Kakar SM, Nakai Y, Smith BJ, Webster R, and Beaumont K (2009) Prediction of human pharmacokinetics from preclinical information: comparative accuracy of quantitative prediction approaches. *J Clin Pharmacol* 49:513-533.

Huang SM, Abernethy DR, Wang Y, Zhao P, and Zineh I (2013) The utility of modeling and simulation in drug development and regulatory review. *J Pharm Sci* 102:2912-2923.

Jones HM, Gardner IB, Collard WT, Stanley PJ, Oxley P, Hosea NA, Plowchalk D, Gernhardt S, Lin J, Dickins M, Rahavendran SR, Jones BC, Watson KJ, Pertinez H, Kumar V, and Cole S (2011) Simulation of human intravenous and oral pharmacokinetics of 21 diverse compounds using physiologically based pharmacokinetic modelling. *Clin Pharmacokinet* 50:331-347.

Jones HM, Parrott N, Jorga K, and Lave T (2006) A novel strategy for physiologically based predictions of human pharmacokinetics. *Clin Pharmacokinet* 45:511-542.

Kasim NA, Whitehouse M, Ramachandran C, Bermejo M, Lennernas H, Hussain AS, Junginger HE, Stavchansky SA, Midha KK, Shah VP, and Amidon GL (2004) Molecular properties of WHO essential drugs and provisional biopharmaceutical classification. *Mol Pharm* 1:85-96.

Kostewicz ES, Aarons L, Bergstrand M, Bolger MB, Galetin A, Hatley O, Jamei M, Lloyd R, Pepin X, Rostami-Hodjegan A, Sjogren E, Tannergren C, Turner DB, Wagner C, Weitschies W, and Dressman J (2013) PBPK models for the prediction of in vivo performance of oral dosage forms. *Eur J Pharm Sci*.

Lennernas H, Aarons L, Augustijns P, Beato S, Bolger M, Box K, Brewster M, Butler J, Dressman J, Holm R, Julia Frank K, Kendall R, Langguth P, Sydor J, Lindahl A, McAllister M, Muenster U, Mullertz A, Ojala K, Pepin X, Reppas C, Rostami-Hodjegan A, Verwei M, Weitschies W, Wilson C, Karlsson C, and Abrahamsson B (2013) Oral biopharmaceutics tools - Time for a new initiative - An introduction to the IMI project OrBiTo. *Eur J Pharm Sci*.

Lennernas H, Nylander S, and Ungell AL (1997) Jejunal permeability: a comparison between the ussing chamber technique and the single-pass perfusion in humans. *Pharm Res* 14:667-671.

Obach RS, Lombardo F, and Waters NJ (2008) Trend analysis of a database of intravenous pharmacokinetic parameters in humans for 670 drug compounds. *Drug Metab Dispos* 36:1385-1405.

Parrott N and Lave T (2002) Prediction of intestinal absorption: comparative assessment of GASTROPLUS and IDEA. *Eur J Pharm Sci* 17:51-61.

Patel N, Polak S, Jamei M, Rostami-Hodjegan A, and Turner DB (2014) Quantitative prediction of formulation-specific food effects and their population variability from in vitro data with the physiologically-based ADAM model: a case study using the BCS/BDDCS Class II drug nifedipine. *Eur J Pharm Sci* 57:240-249.

Poulin P, Jones HM, Jones RD, Yates JW, Gibson CR, Chien JY, Ring BJ, Adkison KK, He H, Vuppugalla R, Marathe P, Fischer V, Dutta S, Sinha VK, Bjornsson T, Lave T, and Ku MS (2011) PhRMA CPCDC initiative on predictive models of human pharmacokinetics, part 1: Goals, properties of the PhRMA dataset, and comparison with literature datasets. *J Pharm Sci*.

Rowland M (2012) Microdosing: a critical assessment of human data. *J Pharm Sci* 101:4067-4074.

Sjogren E, Westergren J, Grant I, Hanisch G, Lindfors L, Lennernas H, Abrahamsson B, and Tannergren C (2013) In silico predictions of gastrointestinal drug absorption in pharmaceutical product development: application of the mechanistic absorption model GI-Sim. *Eur J Pharm Sci* 49:679-698.

Tannergren C, Knutson T, Knutson L, and Lennernas H (2003) The effect of ketoconazole on the in vivo intestinal permeability of fexofenadine using a regional perfusion technique. *Br J Clin Pharmacol* 55:182-190.

Thondre PS, Shafat A, and Clegg ME (2013) Molecular weight of barley beta-glucan influences energy expenditure, gastric emptying and glycaemic response in human subjects. *The British journal of nutrition* 110:2173-2179.

Wu T, Bound MJ, Zhao BR, Standfield SD, Bellon M, Jones KL, Horowitz M, and Rayner CK (2013) Effects of a D-xylose preload with or without sitagliptin on gastric emptying, glucagon-like peptide-1, and postprandial glycemia in type 2 diabetes. *Diabetes Care* 36:1913-1918.

Yu LX and Amidon GL (1999) A compartmental absorption and transit model for estimating oral drug absorption. *Int J Pharm* 186:119-125.