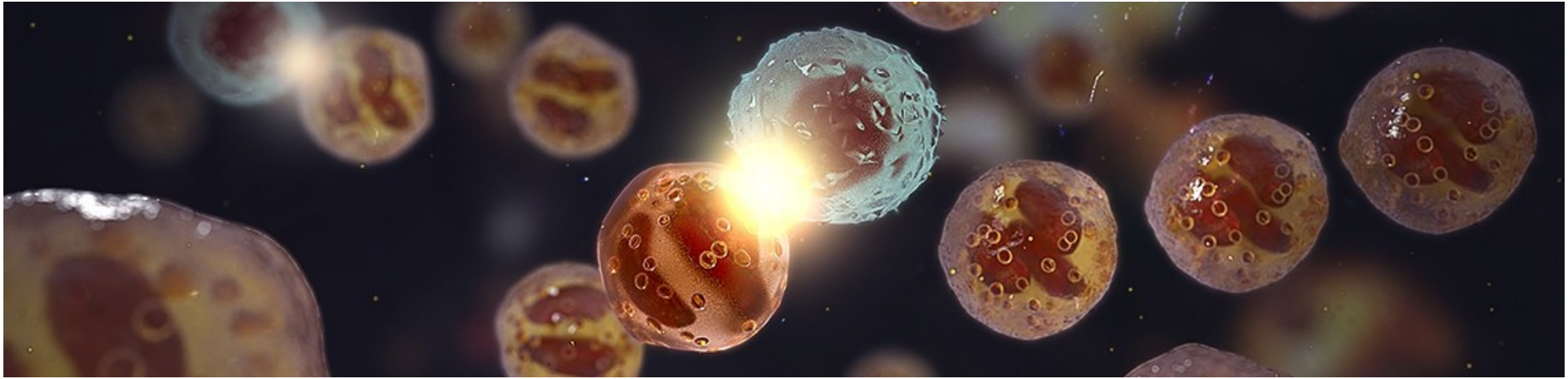
In silico models to estimate a drug's brain penetration likeliness

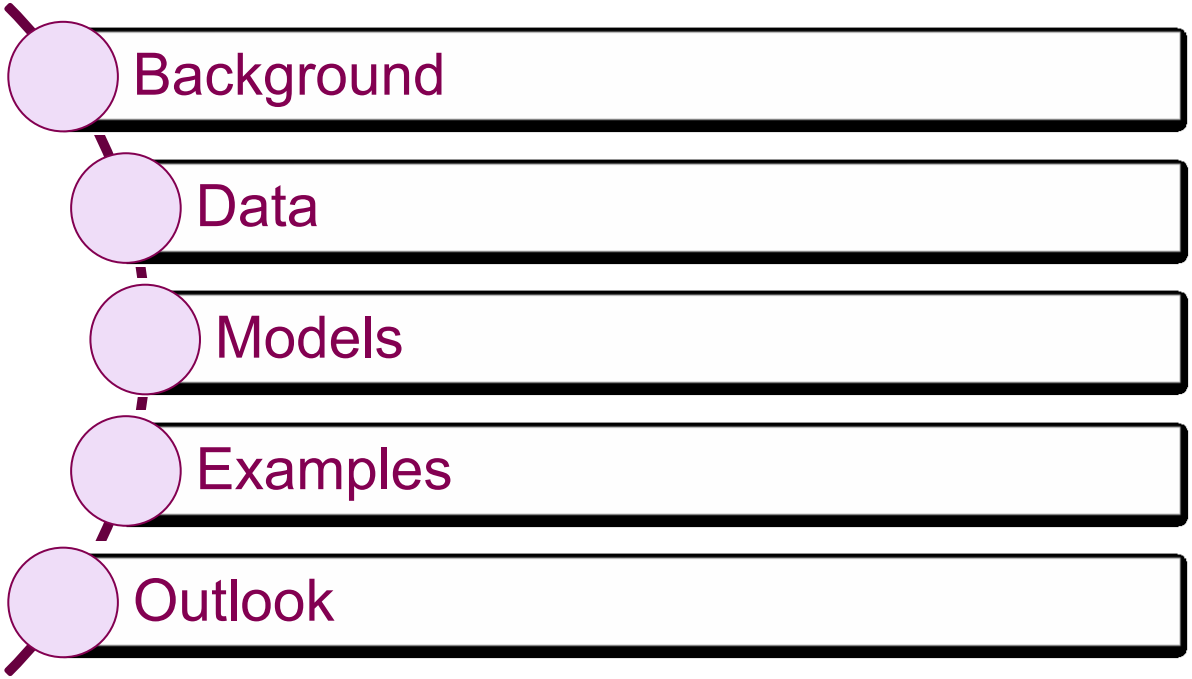
Susanne Winiwarter

UK QSAR meeting, Stevenage

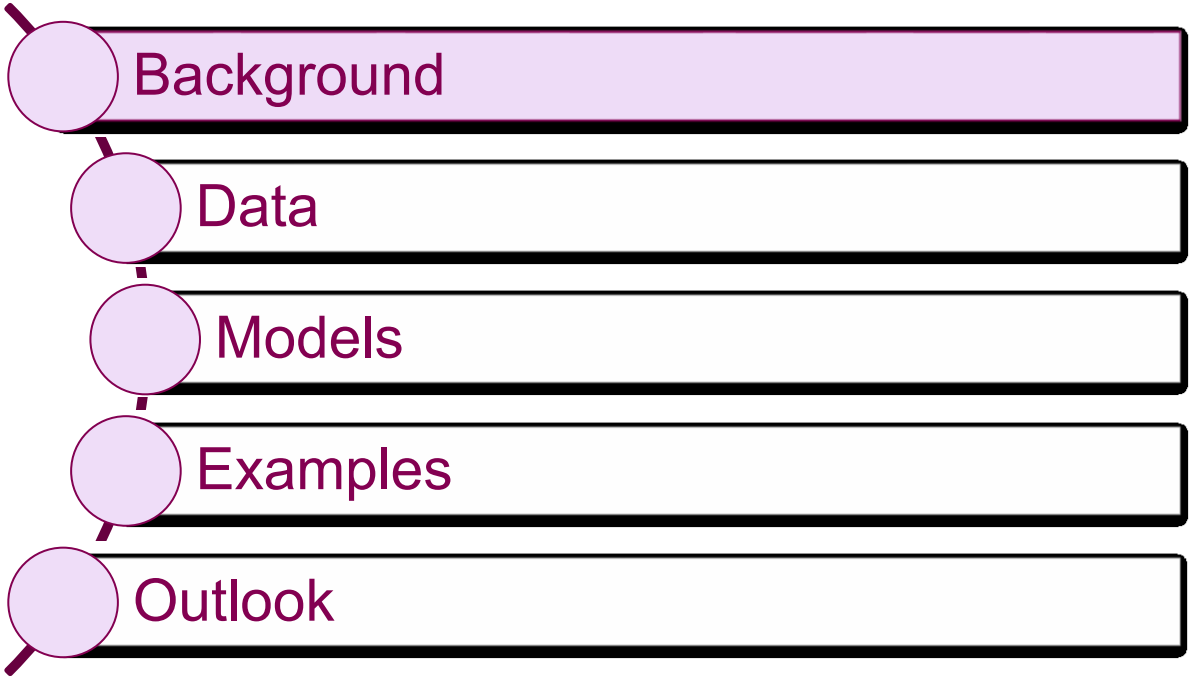
2016-03-16



Outline



Outline



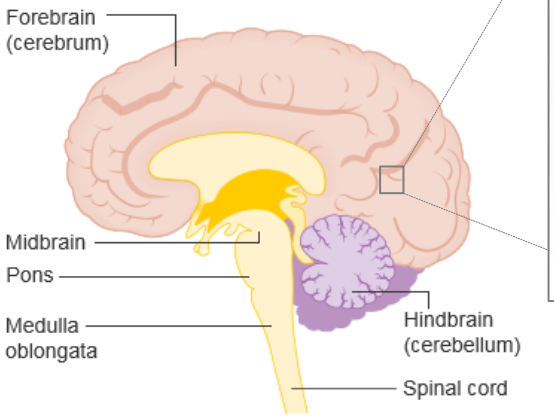
Importance in drug discovery

Two aspects

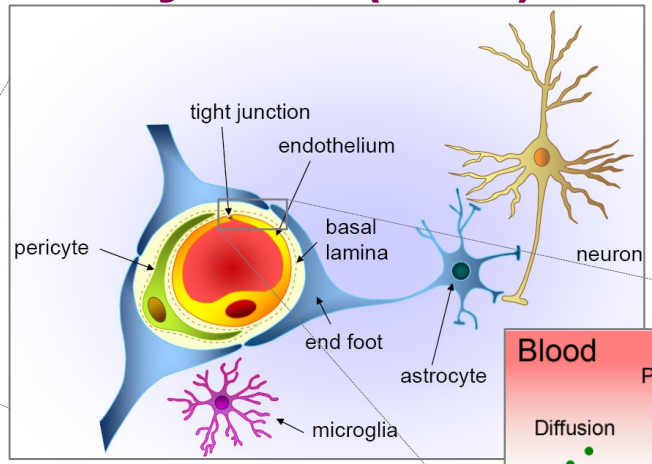
1. Drugs that target a receptor in the brain
   - Need to be able to cross the blood brain barrier, ie need sufficient permeability
     → often compounds were made more lipophilic to meet the requirement

2. Drugs that target a peripheral receptor
   - May be required to stay outside to avoid potential central sideeffects
     eg, antihistamines with quarternary amines were found to have less central, sedative effects

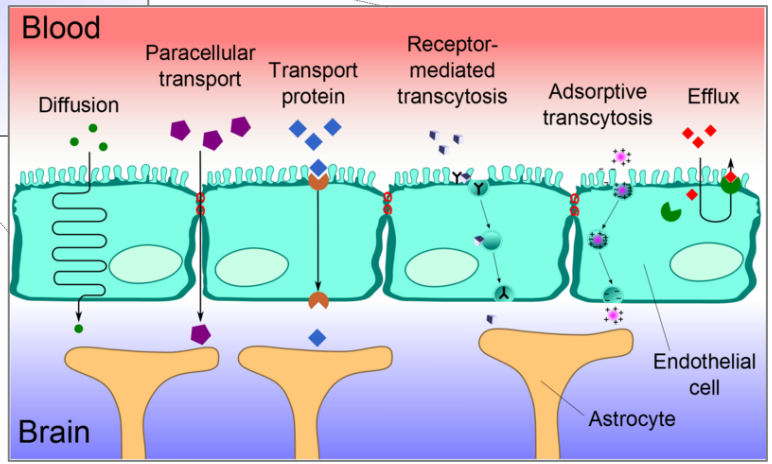
The Central Nervous System (CNS)



Schematic overview
(Source: https://commons.wikimedia.org/; by Cancer Research UK; original email from CRUK)



Cross section of brain blood vessel
(Source: Abbott&Yusof: Structure and function of the blood brain barrier)

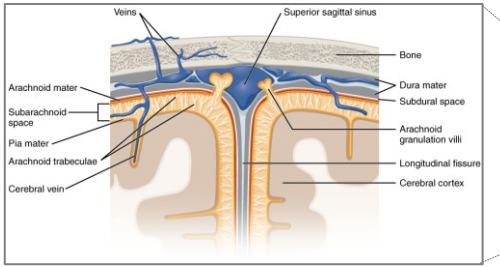


Transport types at the blood-brain barrier
(Source: https://commons.wikimedia.org/; A.Kübelbeck)

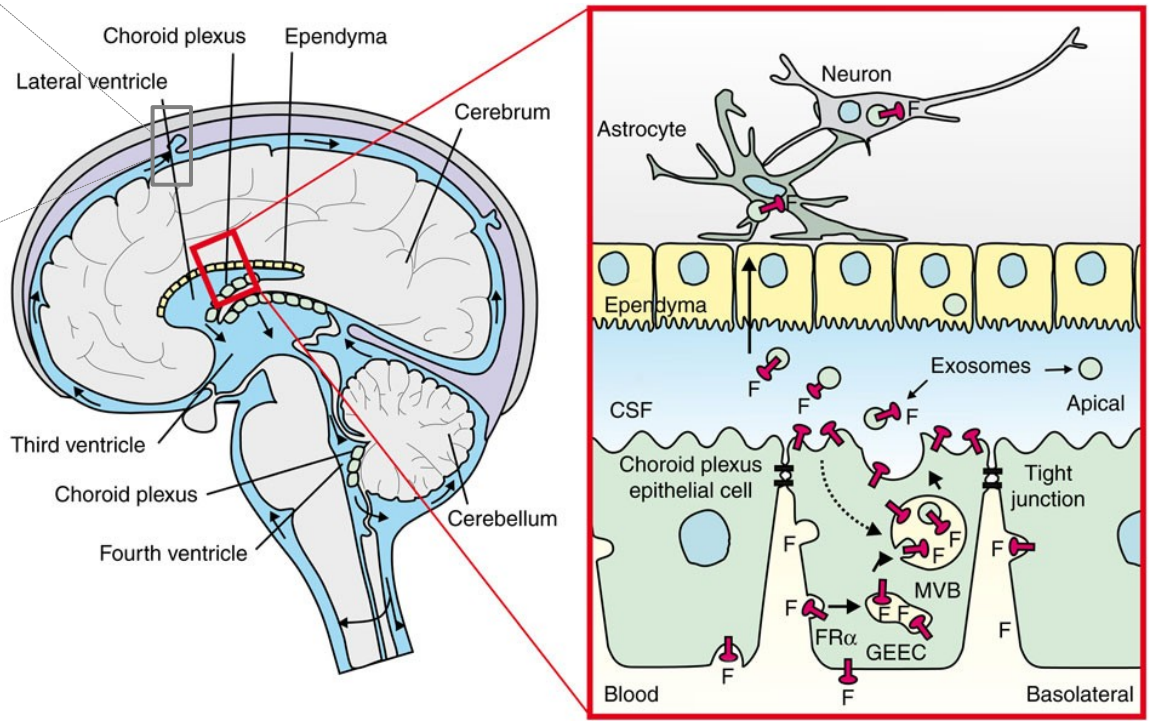
Blood Brain Barrier (BBB)

- protects and regulates the brain microenvironment
- consists of tightly joined endothelial cells
  (tight junction proteins, transporters, metabolism)
- pericytes and astrocytes required for BBB function – neurovascular unit

Blood-CSF and CSF-brain barriers



CSF clearance through arachnoidal villi (Source: https://commons.wikimedia.org/; from Anatomy & Physiology, Connexions Web site)

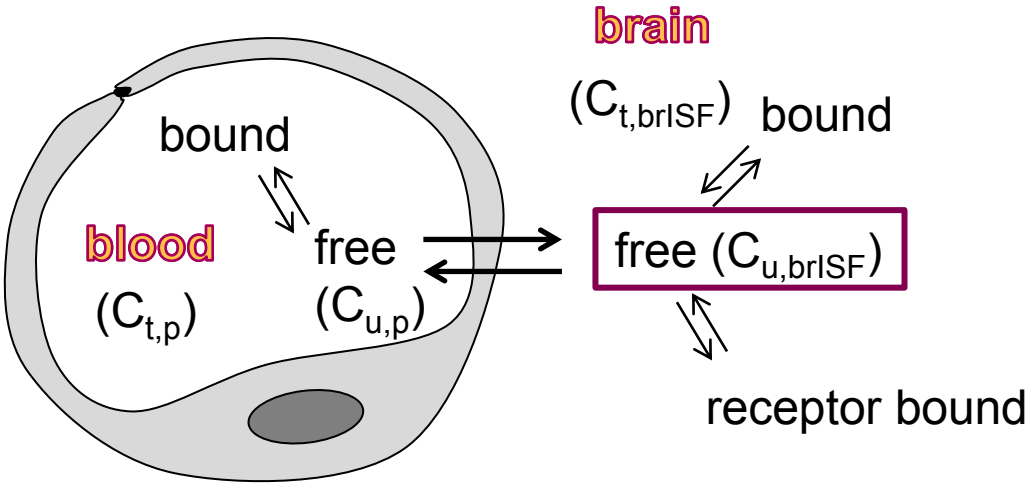


ISF bulk flow: 0.3 µL/g brain
CSF production: 2-5 µL/g brain

Abbreviations: CSF, cerebrospinal fluid; ISF, interstitial fluid

Source: Grapp et al Nature Communications Volume: 4, Article number: 2123; DOI:doi:10.1038/ncomms3123 http://www.nature.com/ncomms/2013/130705/ncomms3123/fig_tab/ncomms3123_F7.html

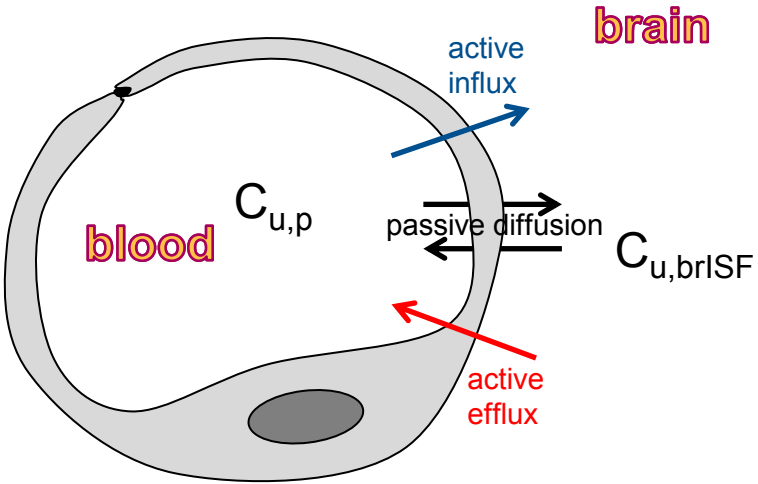
Consider free drug hypothesis



free brain-to-blood ratio:

$$\frac{C_{u,brISF}}{C_{u,p}} = K_{p,uu,br}$$

Consider free drug hypothesis

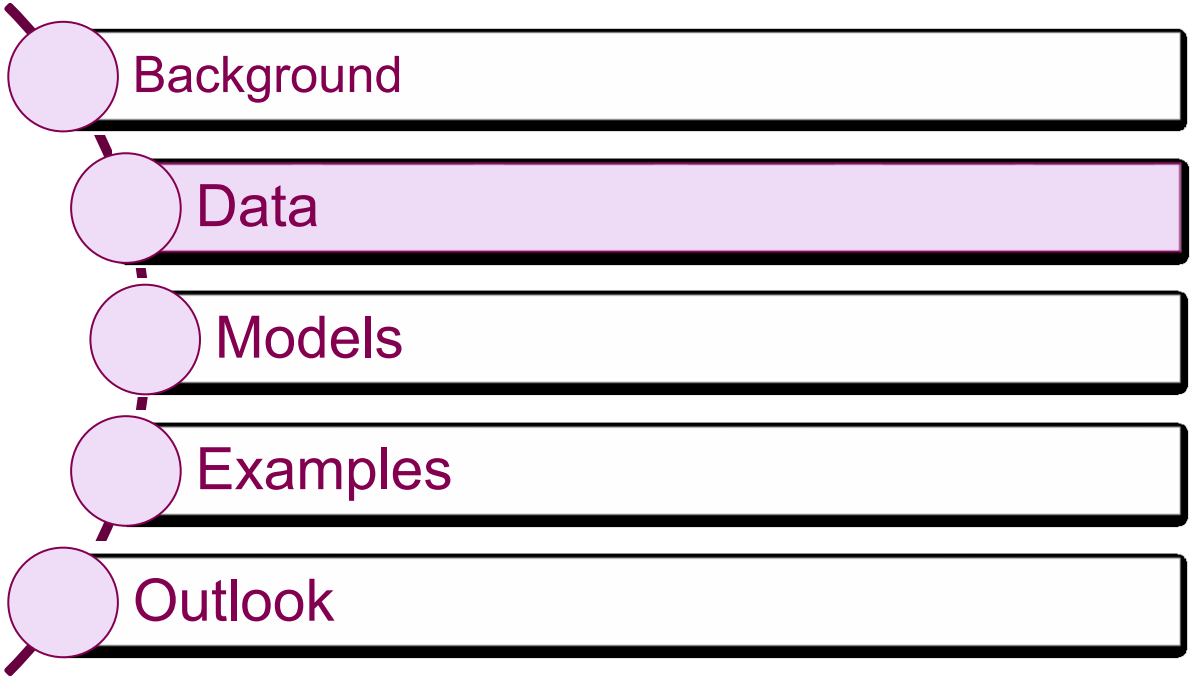


$$K_{p,uu,br} = \frac{C_{u,brISF}}{C_{u,p}} = \frac{CL^{BBB}_{passive} + CL^{BBB}_{influx}}{CL^{BBB}_{passive} + CL^{BBB}_{efflux}}$$

passive: $K_{p,uu,br} \sim 1$
active influx: $K_{p,uu,br} > 1$
active efflux: $K_{p,uu,br} < 1$

→ **Transport mechanism** determines free brain-to-blood ratio

Outline



Data to understand brain penetration

in vivo data – clinical

- PET data (receptor occupancy)
- CSF data
- Classification depending on known central pharmacological effect

→ free brain conc not really accessible

in vivo data – preclinical

- Free brain conc accessible via microdialysis
- PET, CSF
- $C_{br,ss}$ → $K_{p,br}$
- $K_{p,uu,br}$ (in combination with *in vitro* studies of binding)
- PS (*in situ* brain perfusion, measures penetration rate)

in vitro data – BBB models

brain microvascular endothelial cells (BMEC)

primary cells or immortalized cell lines

various species, eg bovine

co-cultures, eg with astrocytes

in vitro data – surrogates

transporter interaction models, eg MDCK-MDR1 efflux measuring P-gp interaction

passive permeability, eg Caco or PAMPA

in vitro data – supporting

measurement of unspecific binding in plasma (f_u) and brain tissue ($f_{u,br}$ or $V_{u,br}$)

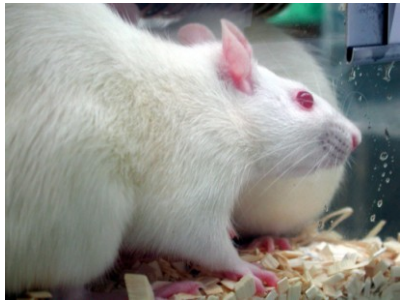
$K_{p,uu,br}$ – assessment in rat

$$K_{p,uu} = \frac{C_{u,br,ISF}}{C_{u,p}}$$

$$K_{p,uu} = \frac{A_{br}}{V_{u,br} * C_p * f_{u,p}}$$

1) *in vivo* assessment of $K_{p,br}$

- Cannulation of femoral vein
- 4-hour constant rate i.v. infusion (steady-state assumed)
- Terminal sampling of blood and brain tissue



Measured by HPLC/MSMS:

- Amount of drug in brain tissue (total brain concentration) $\mathbf{A_{brain}}$ (µmol/kg_{brain})
- Total plasma concentration $\mathbf{C_p}$ (µmol/L)

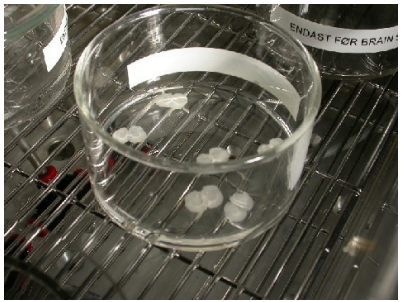
$$C_{u,br,ISF} = \frac{A_{br}}{V_{u,br}}$$

$$C_{u,p} = C_p * f_{u,p}$$

2) *in vitro* assessment of drug binding

Brain slice method:

- preparation of fresh 300 µm slices from drug naïve rats
- incubation with drug in buffer to uptake equilibrium (5 hrs)



- at equilibrium $C_{buffer} = C_{u,sliceISF}$ then $V_{u,br} = \frac{A_{slice}}{C_{u,buffer}}$

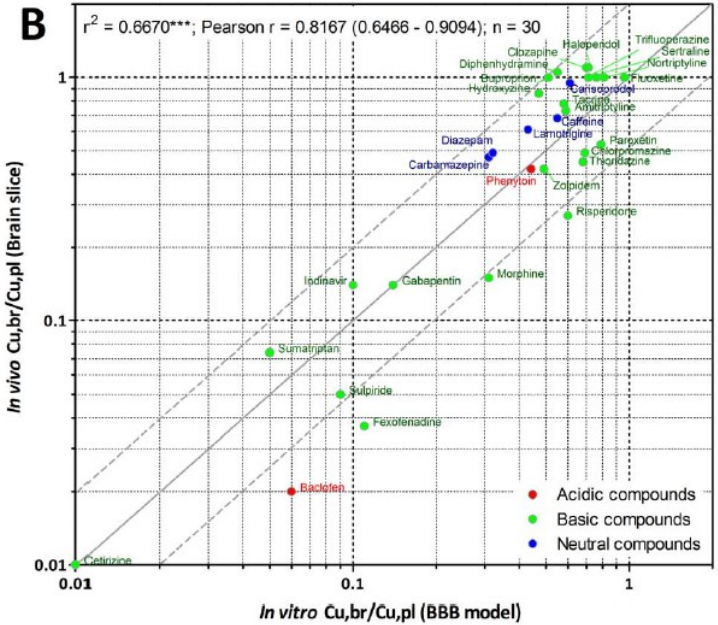
Plasma protein binding determined via equilibrium dialysis (*in vitro* or *ex vivo*)

in vitro BBB models – recent progress:

A Simple Method for Assessing Free Brain/Free Plasma Ratios Using an *In Vitro* Model of the Blood Brain Barrier

Maxime Culot[1*], Anaëlle Fabulas - da Costa[1], Emmanuel Sevin[1], Erica Szorath[2], Stefan Martinsson[2], Mila Renftel[2], Yan Hongmei[2], Romeo Cecchelli[1], Stefan Lundquist[2]

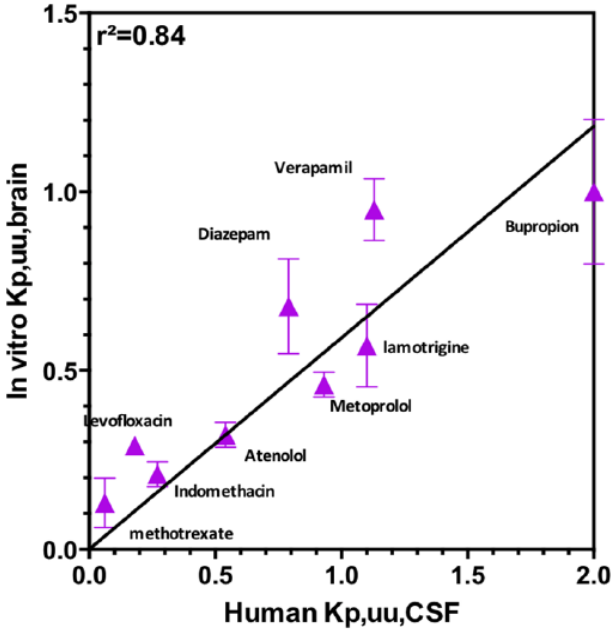
PloS one (2013), 8(12) e80634, doi:10.1371/journal.pone.0080634



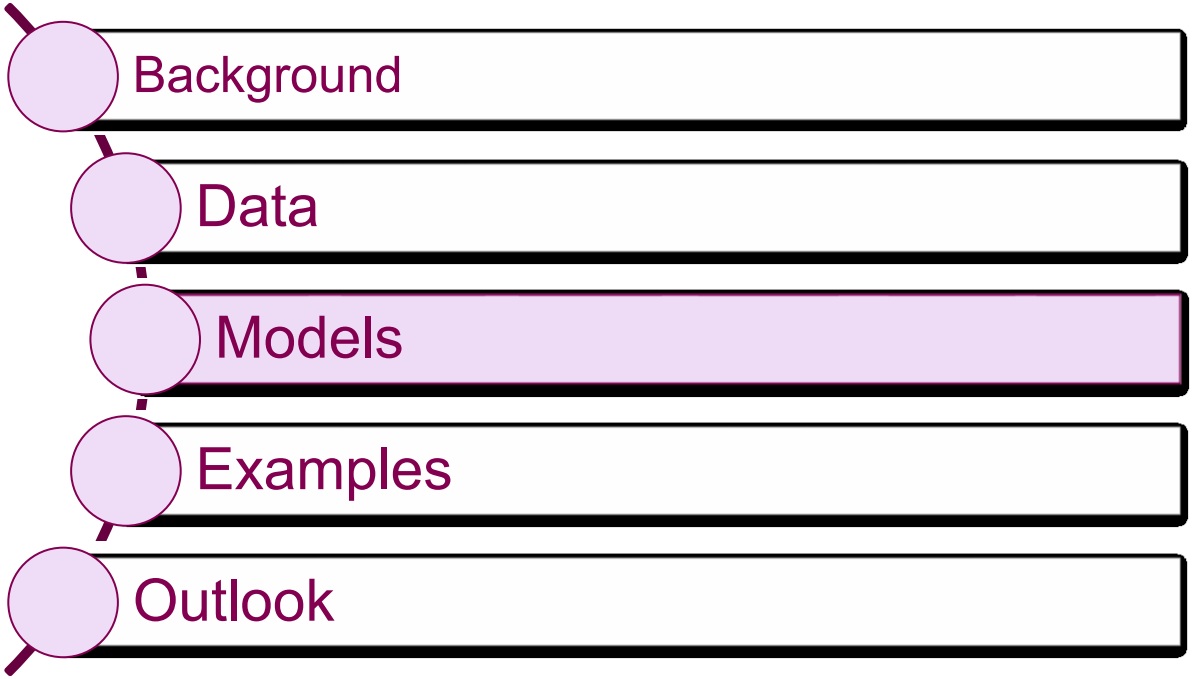
A Stable and Reproducible Human Blood-Brain Barrier Model Derived from Hematopoietic Stem Cells

Romeo Cecchelli[1], Sezin Aday[2,3], Emmanuel Sevin[1], Catarina Almeida[2,3,4], Maxime Culot[1], Lucie Dehouck[1], Caroline Coisne[5], Britta Engelhardt[5], Marie-Pierre Dehouck[1], Lino Ferreira[2,3]*

PloS one (2014), 9(6), e99733, doi:10.1371/journal.pone.0099733



Outline



$K_{p,uu}$ model 1

Journal of Medicinal Chemistry Article

J. Med. Chem. **2009,** *52,* 6233–6243 **6233**
DOI: 10.1021/jm901036q

Structure−Brain Exposure Relationships in Rat and Human Using a Novel Data Set of Unbound Drug Concentrations in Brain Interstitial and Cerebrospinal Fluids

Markus Fridén,*[,†,‡] Susanne Winiwarter,[†] Gunilla Jerndal,[†] Ola Bengtsson,[†,‡] Hong Wan,[§] Ulf Bredberg,[†] Margareta Hammarlund-Udenaes,[‡] and Madeleine Antonsson[†]

[†]*Discovery DMPK, AstraZeneca R&D Mölndal, SE-431 83 Mölndal, Sweden,* [‡]*Department of Pharmaceutical Biosciences, Division of Pharmacokinetics and Drug Therapy, Uppsala University, Box 591, SE-751 24 Uppsala, Sweden, and* [§]*Lead Generation DMPK and Physical Chemistry, AstraZeneca R&D Mölndal, Mölndal, Sweden*

Received July 13, 2009

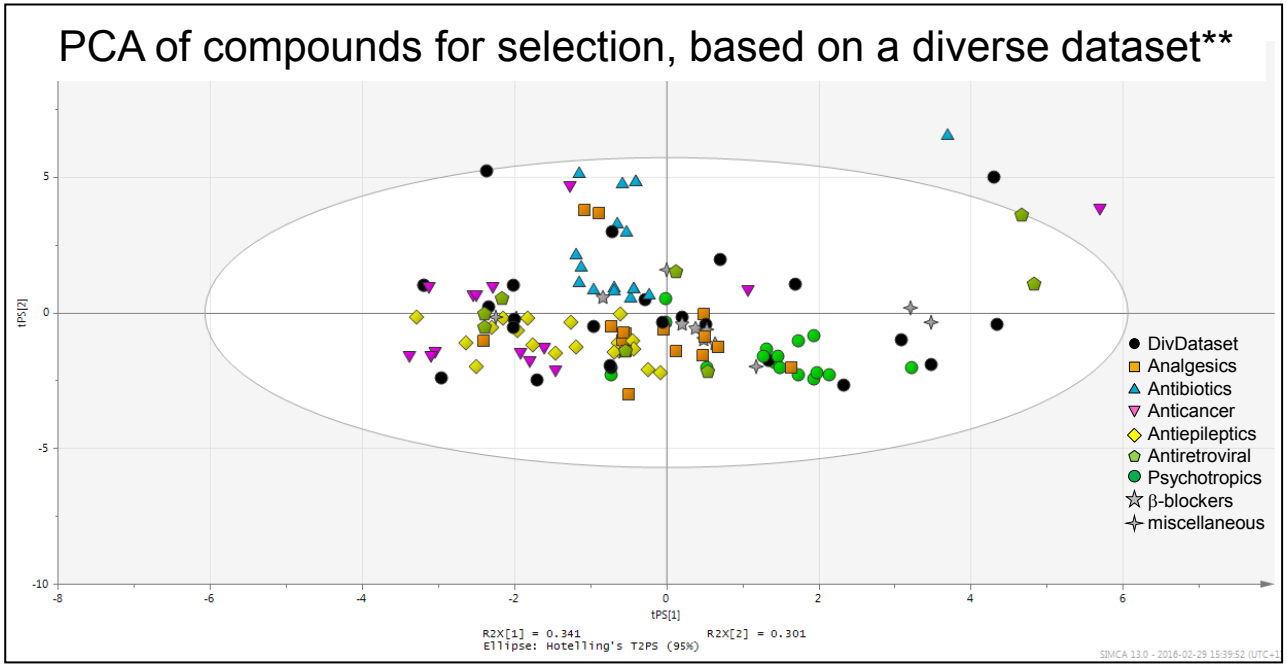
New experimental methodologies were applied to measure the unbound brain-to-plasma concentration ratio ($K_{p,uu,brain}$) and the unbound CSF-to-plasma concentration ratio ($K_{p,uu,CSF}$) in rats for 43 structurally diverse drugs. The relationship between chemical structure and $K_{p,uu,brain}$ was dominated by hydrogen bonding. Contrary to popular understanding based on the total brain-to-plasma concentration ratio (logBB), lipophilicity was not a determinant of unbound brain exposure. Although changing the number of hydrogen bond acceptors is a useful design strategy for optimizing $K_{p,uu,brain}$, future improvement of in silico prediction models is dependent on the accommodation of active drug transport. The structure−brain exposure relationships found in the rat also hold for humans, since the rank order of the drugs was similar for human and rat $K_{p,uu,CSF}$. This cross-species comparison was supported by $K_{p,uu,CSF}$ being within 3-fold of $K_{p,uu,brain}$ in the rat for 33 of 39 drugs. It was, however, also observed that $K_{p,uu,CSF}$ overpredicts $K_{p,uu,brain}$ for highly effluxed drugs, indicating lower efflux capacity of the blood−cerebrospinal fluid barrier compared to the blood−brain barrier.

Three main results:

- Structure brain exposure relationships explored
- $K_{p,uu,br}$ compared to $K_{p,uu,CSF}$ (rat)
- Rat $K_{p,uu,CSF}$ compared to human $K_{p,uu,CSF}$

$K_{p,uu}$ model 1 – dataset selection

Selection based on Shen dataset* and diversity;
Additional model drugs and β-blockers included



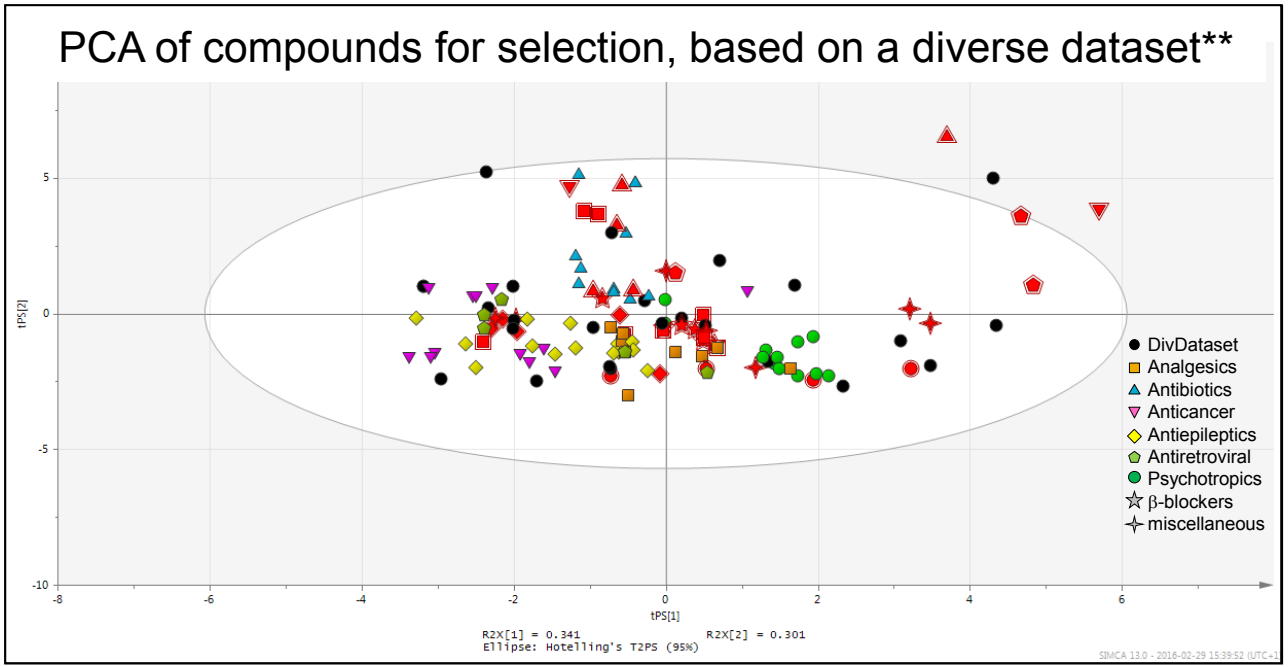
* ..... Shen et al. Principles and applicability of CSF sampling for the assessment of CNS drug delivery and pharmacodynamics; Advanced Drug Delivery Reviews 56 (2004) 1825– 1857
** … Sköld et al. Presentation of a structurally diverse and commercially available drug data set for correlation and benchmarking studies. J. Med. Chem. 2006, 49, 6660–6671.

$K_{p,uu}$ model 1 – dataset selection

Selection based on Shen dataset* and diversity;
Additional model drugs and β-blockers included

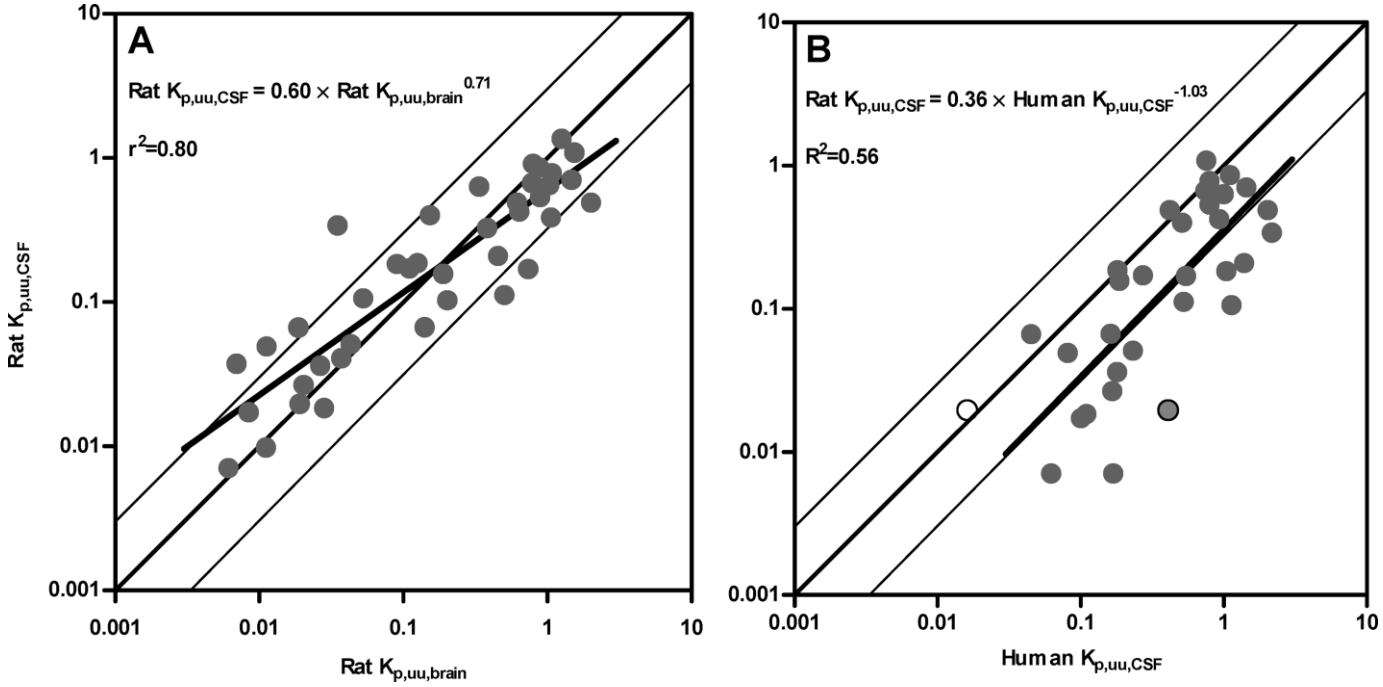
→ 43 compounds

= Fridén dataset



* ..... Shen et al. Principles and applicability of CSF sampling for the assessment of CNS drug delivery and pharmacodynamics; Advanced Drug Delivery Reviews 56 (2004) 1825– 1857
** … Sköld et al. Presentation of a structurally diverse and commercially available drug data set for correlation and benchmarking studies. J. Med. Chem. 2006, 49, 6660–6671.

Comparison of experimental data

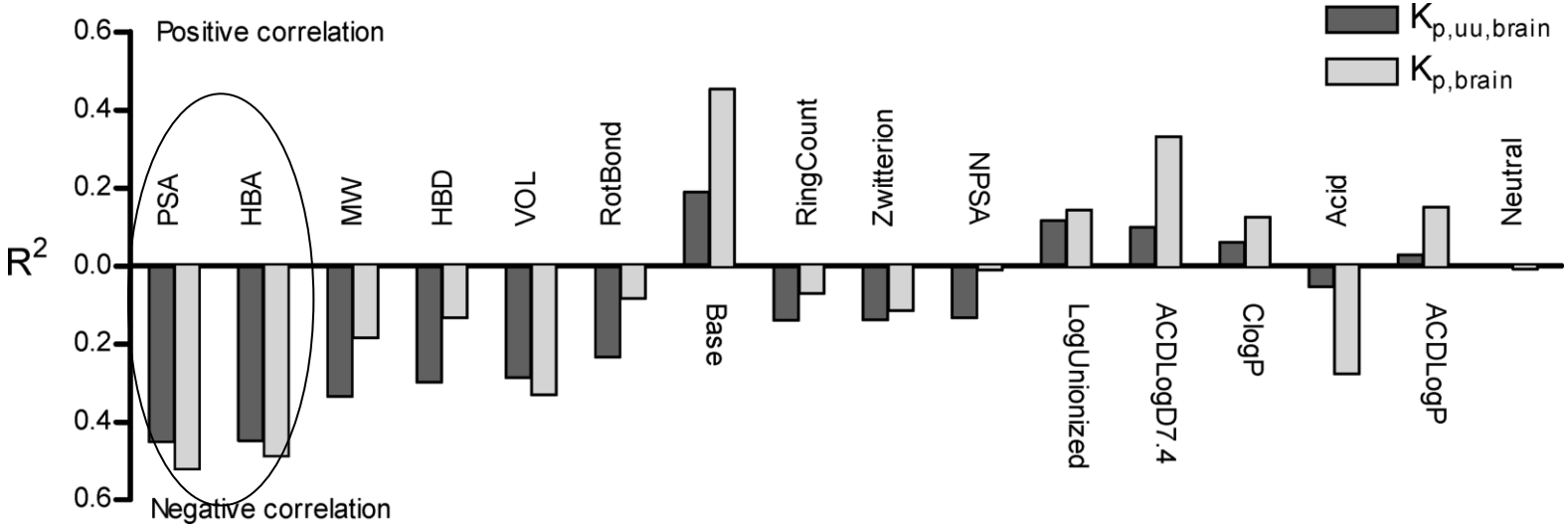


Relationship between Kp,uu,CSF and Kp,uu,brain in the rat, n=39, (A) and the agreement between rat and human Kp,uu,CSF, n=31* (B), average values shown.

* … moxalactam was included twice (healthy volunteers, open circle, vs patient with bacterial meningities, filled circle)

$K_{p,uu}$ model 1 – results

Correlations to simple physchem descriptors



→ hydrogen bond (acceptor) descriptors important, lipophilicity of less importance for $K_{p,uu,br}$

$K_{p,uu}$ model 2 – use of all in house data

Journal of Molecular Graphics and Modelling 29 (2011) 985–995

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

In silico prediction of unbound brain-to-plasma concentration ratio using machine learning algorithms

Hongming Chen[a,*], Susanne Winiwarter[b], Markus Fridén[b,c], Madeleine Antonsson[b], Ola Engkvist[a]

[a] DECS GCS Computational Chemistry, AstraZeneca R&D Mölndal, SE-43183 Mölndal, Sweden
[b] Discovery DMPK, AstraZeneca R&D Mölndal, SE-43183 Mölndal, Sweden
[c] Department of Pharmaceutical Bioscience, Division of Pharmacokinetics and Drug Therapy, Uppsala University, Box 591, SE-75124 Uppsala, Sweden

ARTICLE INFO

Article history:
Received 21 February 2011
Received in revised form 8 April 2011
Accepted 12 April 2011
Available online 27 April 2011

Keywords:
Quantitative structure–activity relationship (QSAR)
Unbound brain-to-plasma concentration ratio
Total brain-to-plasma concentration ratio
Random forest (RF)
Support vector machine (SVM)
Partial least squares (PLS)

ABSTRACT

Distribution over the blood–brain barrier (BBB) is an important parameter to consider for compounds that will be synthesized in a drug discovery project. Drugs that aim at targets in the central nervous system (CNS) must pass the BBB. In contrast, drugs that act peripherally are often optimised to minimize the risk of CNS side effects by restricting their potential to reach the brain. Historically, most prediction methods have focused on the total compound distribution between the blood plasma and the brain. However, recently it has been proposed that the unbound brain-to-plasma concentration ratio ($K_{p,uu,brain}$) is more relevant. In the current study, quantitative $K_{p,uu,brain}$ prediction models have been built on a set of 173 in-house compounds by using various machine learning algorithms. The best model was shown to be reasonably predictive for the test set of 73 compounds ($R^2 = 0.58$). When used for qualitative prediction the model shows an accuracy of 0.85 (Kappa = 0.68). An additional external test set containing 111 marketed CNS active drugs was also classified with the model and 89% of these drugs were correctly predicted as having high brain exposure.

© 2011 Elsevier Inc. All rights reserved.

- in house $K_{p,uu}$ data for 246 compounds (73 as test set; log $K_{p,uu}$: -2.5 – 1; mean -1.1)
- different machine learning algorithms used (RF, PLS, SVM)
- direct and indirect models
- 196 2D and 3D physico-chemical descriptors (’AZdescriptors’)
- accuracy of qualitative model 85% (89% on external test set)
- important descriptors related to hydrogen bonding

$K_{p,uu}$ model 2 – revisited

RESEARCH ARTICLE – *Pharmacokinetics, Pharmacodynamics and Drug Transport and Metabolism*

Exploring *In Silico* Prediction of the Unbound Brain-to-Plasma Drug Concentration Ratio: Model Validation, Renewal, and Interpretation

SRINIDHI VARADHARAJAN,[1,2] SUSANNE WINIWARTER,[3] LARS CARLSSON,[3] OLA ENGKVIST,[2] AJAY ANANTHA,[2,4] THIERRY KOGEJ,[2] MARKUS FRIDÉN,[5] JONNA STÅLRING,[3] HONGMING CHEN[2]

[1]Department of Biology, Lund University, Lund SE-22100, Sweden
[2]Chemistry Innovation Centre, Discovery Sciences, AstraZeneca R&D Mölndal, Mölndal SE-43183, Sweden
[3]Computational ADME and Safety, Drug Safety and Metabolism, AstraZeneca R&D Mölndal, Mölndal SE-43183, Sweden
[4]Department of Chemistry and Molecular Biology, University of Gothenburg, Göteborg SE-40530, Sweden
[5]RIA Innovative Medicines, AstraZeneca R&D Mölndal, Mölndal SE-43183, Sweden

Received 7 October 2014; revised 14 November 2014; accepted 18 November 2014

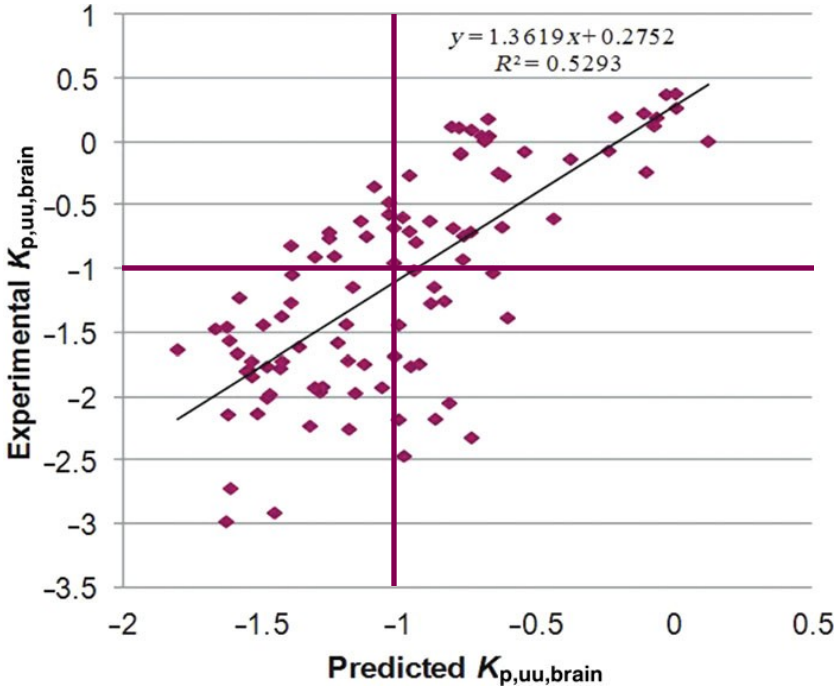
Published online 24 December 2014 in Wiley Online Library (wileyonlinelibrary.com). DOI 10.1002/jps.24301

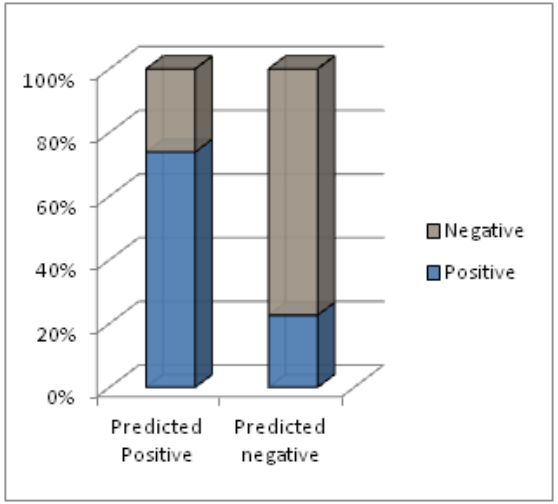
ABSTRACT: Recently, we built an *in silico* model to predict the unbound brain-to-plasma concentration ratio ($K_{p,uu,brain}$), a measure of the distribution of a compound between the blood plasma and the brain. Here, we validate the previous model with new additional data points expanding the chemical space and use that data also to renew the model. The model building process was similar to our previous approach; however, a new set of descriptors, molecular signatures, was included to facilitate the model interpretation from a structure perspective. The best consensus model shows better predictive power than the previous model ($R^2 = 0.6$ vs. $R^2 = 0.53$, when the same 99 compounds were used as test set). The two-class classification accuracy increased from 76% using the previous model to 81%. Furthermore, the atom-summarized gradient based on molecular signature descriptors was proposed as an interesting new approach to interpret the $K_{p,uu,brain}$ machine learning model and scrutinize structure $K_{p,uu,brain}$ relationships for investigated compounds. © 2014 Wiley Periodicals, Inc. and the American Pharmacists Association J Pharm Sci 104:1197–1206, 2015

Keywords: ADME; blood–brain barrier; computational ADME; drug transport; unbound brain-to-plasma concentration ratio; machine learning; in silico modeling

- Validation of previous model with additional, temporal data (accuracy 76%, somewhat lower than for original test set)
- New model built, similar methods, but including new set of descriptors (signatures)
- Accuracy of 2 class model ~80% for temporal test set
- Potential to interpret model for individual compounds through new descriptors

Validation of $K_{p,uu}$ model 2 (three years later)





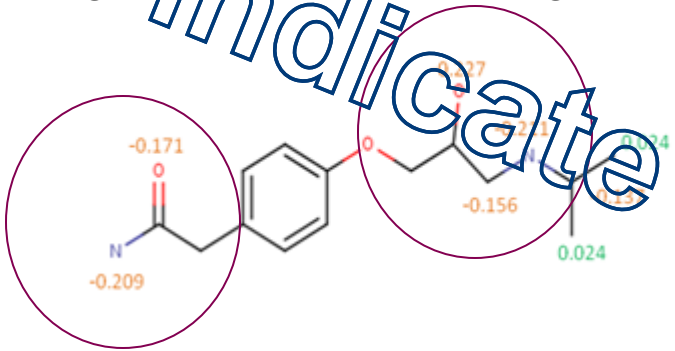
log $K_{p,uu,br}$ limit -1 ($K_{p,uu,br}$ 0.1)

n=99, temporal test set

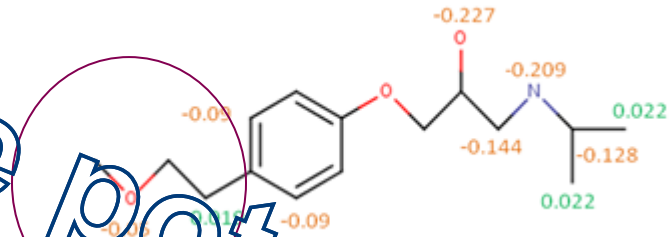
Identify substructures for modification

through atom summarised gradient values (SVM model using signature descriptors)

to indicate the potential...



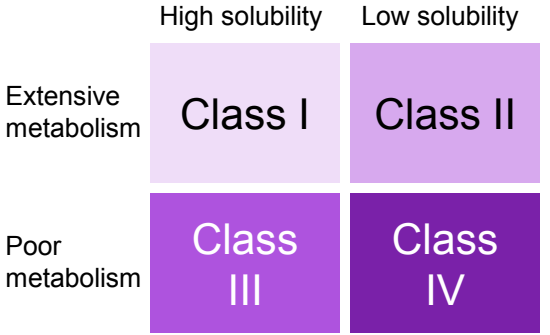
Atenolol,
Exp. log $K_{p,uu,br}$ = -1.59, pred -1.06



Metoprolol,
Exp. log $K_{p,uu,br}$ = -0.2, pred -0.67

Prediction using CNS+/- data

Broccatelli et al. Adv. Drug Del. Rev. 64 (2012) 95-109:
Improving the prediction of the brain disposition for orally administered drugs using BDDCS



BDDCS
(=Biopharmaceutics drug disposition classification system)

Identified 153 drugs with BDDCS and Pgp interaction information
(~80% CNS+!)
CNS+ classification based on: marketed CNS drugs; $K_{p,uu}$>0.1 (hu if available)

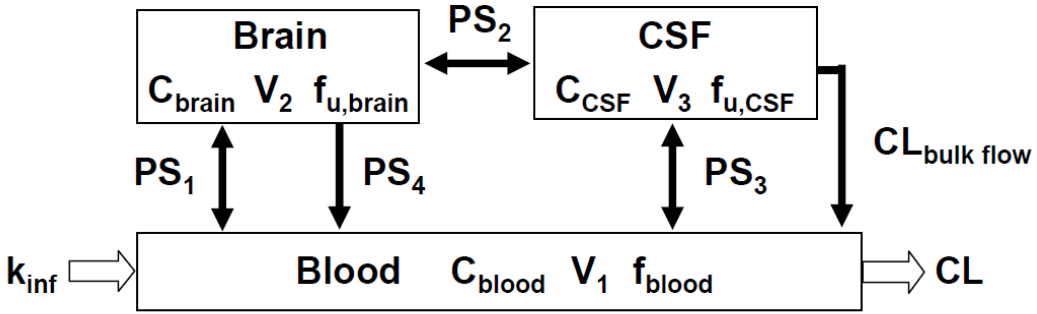
Suggested decision tree:

1. low (predicted) permeability → likely CNS-
2. low P-gp ER (exp) → likely CNS+
3. BDDCS class I → likely CNS+

NOTE: about 98% of BDDCS class I cmpds CNS+
(>50% of cmpds in class I)

Pharmacokinetic model for $K_{p,uu,br}$ and $K_{p,uu,CSF}$

Kodaira et al, DMD 42:983-989, 2014



PS1 – passive permeability at BBB estimated from physchem descr
PS3 – passive permeability at BCSFB assumed to be proportional to PS1
PS2 – passive diffusion at ependyma inversely proportional to sqrt(MW)
PS4 – active efflux
in vitro Pgp and BCRP activities considered

Resulting equations fitted (4 parameters) to $K_{p,uu,br}$ and $K_{p,uu,CSF}$ data for 19 compounds (three BCRP substrates, five P-gp substrates, one substrate for both)

Model could explain parameters within 3 fold for most compounds

in silico models for *in vitro* endpoints

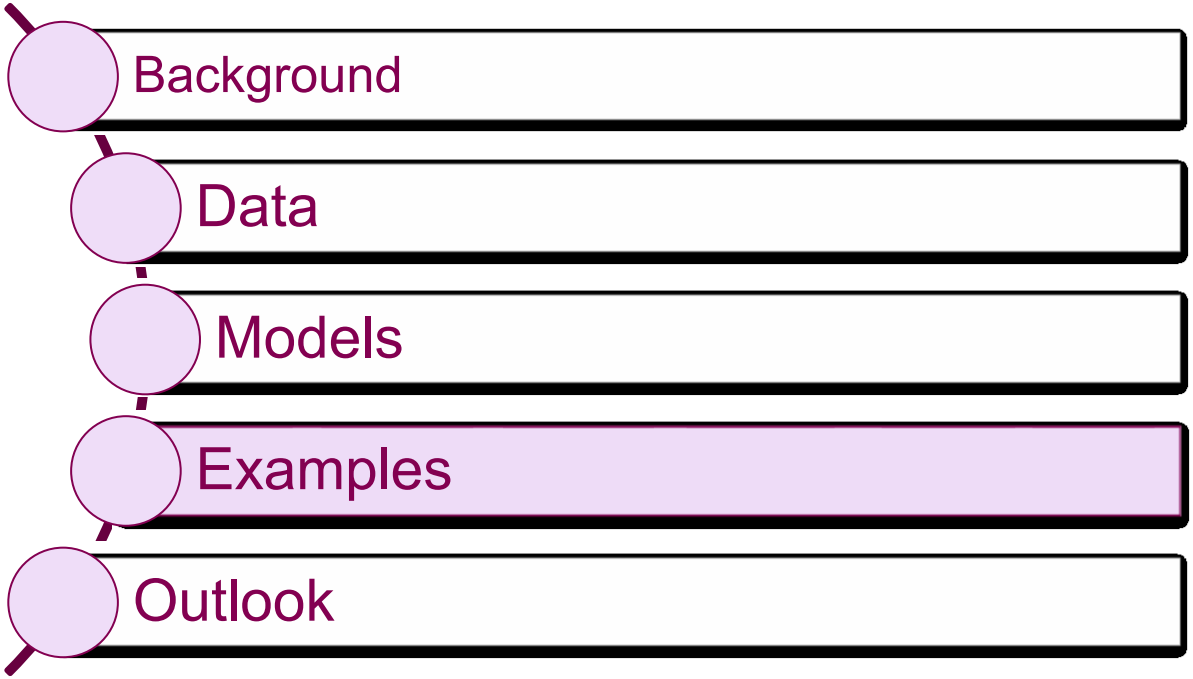
Models of interest: P-gp interaction (eg MDCK-MDR1 efflux), BCRP interaction, combinations (eg Caco-2 efflux), passive permeability
Quite a few available in literature

In house models available for MDCK-MDR1 efflux, Caco-2 efflux and passive permeability (measured in Caco-2 cells)

Model	# train	method	# test	RMSEP	accuracy
MDCK-MDR1 ER	>1700	RF	190	0.46	0.89
Caco-2 ER (class)	>13000	RF	1500	-	0.82
Caco-2 passive	>2500	RF	280	0.45	0.87

Hydrogen bonding descriptors important for all models

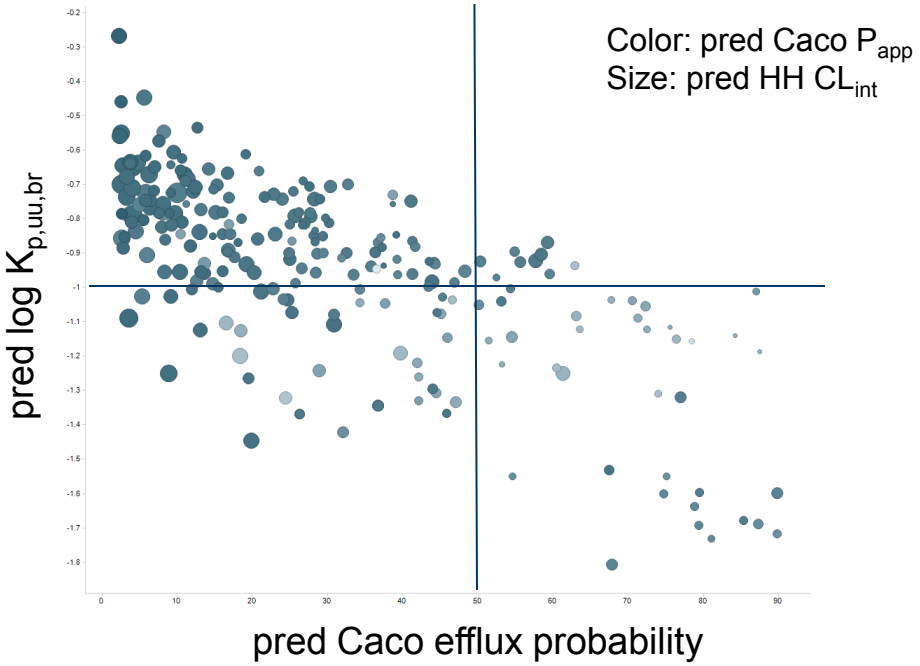
Outline



Early discovery – select compounds/series

Usually not too concerned about BBB, unless targeting CNS

Combination of models used to select compounds to be made/series to be forwarded



Model usage for prioritizing – eg lead optimization

Target in CNS, receptor occupancy can be measured, but not for all compounds

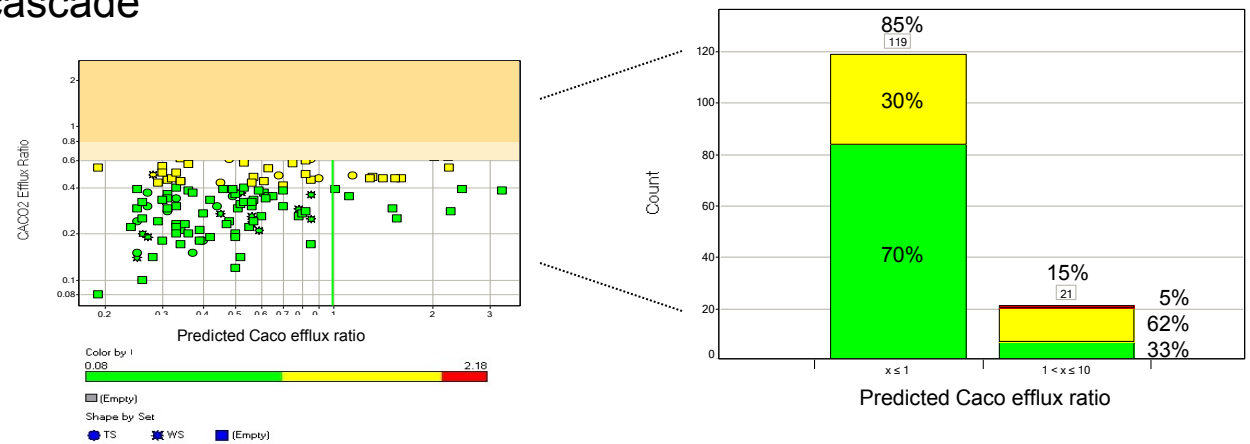
Efflux in Caco-2 shown to predict receptor occupancy

in silico model for Caco-2 efflux could be utilized as pre-filter in screening cascade

Most compounds low efflux → likely to reach CNS

Only compounds with predicted high efflux needed measurement immediately

Other compounds measured for efflux, when progressed



Judge brain penetration using all data I

Are there differences in risk of potential CNS side effects for compound and its metabolite?

	Caco-2 P_{app}	logD	pred P_{app}	pred logD	Caco-2 efflux prob (%)	MDCK-MDR1 ER (pred)	pred $K_{p,uu}$ class	pred log $K_{p,uu}$
cmpd X	**high**	**4.2**	16	3.8	43%	6.2	POS	-0.69
metabolite	**high**	**3.1**	9	3.1*	86%	5.8	NEG	-1.08

* Compound in model training set

Interpretation:

- Cmpd X **likely** to reach CNS (high P_{app} and logD, efflux risk <50%, $K_{p,uu}$ POS)
- Metabolite **less likely** to reach CNS (P_{app} still high, efflux risk >50%; $K_{p,uu}$ NEG)

Judge brain penetration using all data II

Judge a new compound in comparison to a closely related one:

	$K_{p,uu}$	Caco-2 P_{app}	Caco-2 ER	MDCK-MDR1-ER	pred P_{app}	Caco-2 efflux prob (%)	MDCK-MDR1 ER (pred)	pred $K_{p,uu}$ class	pred log $K_{p,uu}$
cmpd a	**0.05**		**2.5**		84	95%	2.5	POS	-0.46
cmpd b		**47**	**3.2**	**1.4**	47*	77%	1.9	POS	-0.71

* Compound in model training set

cmpd a: low risk to enter the brain according to experiment (exp $K_{p,uu}$=0.05, Caco efflux);
pred $K_{p,uu}$=0.35 (wrong), high Caco-permeability, high efflux risk → efflux to be considered in series

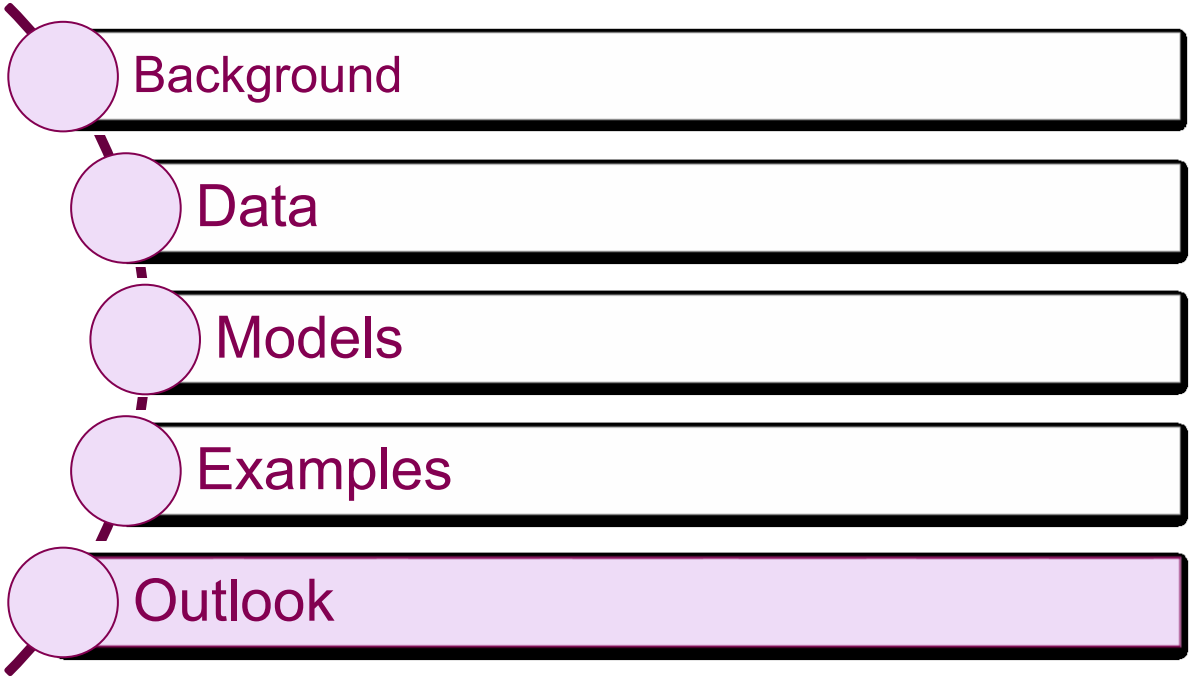
cmpd b: exp: good permeability, but efflux in Caco (higher than cmpd a), no efflux in MDCK-MDR1;
predicted: permeability (in train set) and efflux data in line with experiment; $K_{p,uu}$ as pos, but value lower than cmpd c; → risk for brain availability **lower** than for cmpd a

→ $K_{p,uu}$ model not to be used by itself, but in conjunction with known data and other relevant predictions

Summary – examples and usage

- Models for unbound brain to plasma ratio exist and can be used to estimate the likelihood of new chemical entities to enter the brain
- BUT: Even low ratios can lead to CNS effects when compound is potent enough (or needs really high plasma concentrations)
- Judgement to be based on **all** available data – including potency (or secondary pharmacology)
  - consider permeability and efflux data (and predictions)
  - consider how well close compounds are predicted
  - consider model applicability (eg distance to training set)
- **Key question**: Which model(s) help a project forward?

Outline



Outlook

1. Consensus that free drug concentration is important
   → $K_{p,uu,br}$ is a useful parameter to understand brain penetration
   → more data will become available for model building

2. Efforts to improve *in vitro* models for brain penetration – eg BBB models based on human stem cells
   → will enable human relevant values
   → will enable even more data to become available

3. Improvement of usability of *in silico* or pharmacokinetic models
   - utilisation of descriptors that can explain structural requirements
   - more ways to include transporter interaction parameters
   - enable modeling for different species/populations

Acknowledgements

DMPK
Markus Fridén
Madeleine Antonsson
Gunilla Jerndal
Hong Wan
Ulf Bredberg

Uppsala University
Ola Bengtsson
Margareta Hammarlund-Udenaes

Computational Chemistry/Toxicology
Hongming Chen
Ola Engkvist
Srinidhi Varadharajan
Lars Carlsson
Jonna Stålring
Ernst Ahlberg

Confidentiality Notice

This file is private and may contain confidential and proprietary information. If you have received this file in error, please notify us and remove it from your system and note that you must not copy, distribute or take any action in reliance on it. Any unauthorized use or disclosure of the contents of this file is not permitted and may be unlawful. AstraZeneca PLC, 2 Kingdom Street, London, W2 6BD, UK, T: +44(0)20 7604 8000, F: +44 (0)20 7604 8151, www.astrazeneca.com

BACK-UP

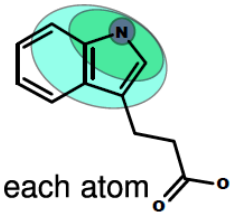
Descriptors

AZdescriptors

Set of 196 physico-chemical 2D and 3D descriptors,
eg MW, clogP, V, PSA, NPSA, HBA, HBD, charge related, topology indices, atom counts, …

Wood et al Mol.Inf. 2011, 30, 960-972

The Signature Descriptor



Signatures are

- chemical substructures, originating from each atom in a compound
- built layer by layer, see Figure, generating a canonical representation of the substructure.
- The number of added layers is said to be the height of the signature, where height zero represents the origin atom only.
- Each node can be arbitrarily labeled but here only atom-types are used

Courtesy E.Ahlberg