GENERAL RULES FOR DATA IMPORT MUTEIN-DB

Generally it's essential to register as much information as possible to obtain a meaningful and reliable data base. Mostly literature contains data for more than one mutein, therefore we created an excel file to facilitate data entry (available at http://www.MuteinDB.org or MuteinDB@genome.tugraz.at) In principle its possibly to enter different enzymes in one excel file, however for more convenient data handling each enzyme should be entered in its own file! These files are finally imported to mutein database by authorized persons (e.g. Gerhard Thallinger, IGB, TU-Graz)

EXCEL FILE:

Do not change the template (only add data)

Delete all empty rows between data rows!

The Excel file is separated in different areas (first row):

- Basic data
- Signal sequence, pH, Temperature, Stability
- Reaction
- Activity

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Basic Data												
2	UserName	Date	MuteinName	WildtypeName	GenBankID	UniProtID	PDBID	Kingdom	O-Source	Tissue	Localization	PTMs	Mutations
3	Weinhandl	20090824		CYP3A4	NP_059488	P08684	1TQN	Mammalian	Homo sapiens	human liver	microsome (ER)		
4	Weinhandl	20090824	CYP3A4-L293P	CYP3A4				Mammalian	Homo sapiens	human liver	microsome (ER)		L293P
5	Braun	20100119		CYP3A4	NP_059488	P08684	1TQN	Mammalian	Homo sapiens	human liver	microsome (ER)		
6	Braun	20100119	CYP3A4-L211F/D214E	CYP3A4				Mammalian	Homo sapiens	human liver	microsome (ER)		L211F, D214E
7	Braun	20100120		CYP3A4	NP_059488	P08684	1TQN	Mammalian	Homo sapiens	human liver	microsome (ER)		
8	Braun	20100120	CYP3A4.53x	CYP3A4				Mammalian	Homo sapiens	human liver	microsome (ER)		G56D, T185S, S222P, L293P
9	Braun	20100120		CYP3A4	NP_059488	P08684	1TQN	Mammalian	Homo sapiens	human liver	microsome (ER)		
10	Braun	20100120	CYP3A4.53-1	CYP3A4				Mammalian	Homo sapiens	human liver	microsome (ER)		G56-D57insLQS
11	Braun	20100205		CYP3A4	NP_059488	P08684	1TQN	Mammalian	Homo sapiens	human liver	microsome (ER)		
12	Braun	20100205	CYP3A4-NF25	CYP3A4	AAA35744			Mammalian	Homo sapiens	human liver	microsome (ER)		V392W, I431T
13	Braun	20100205	CYP3A4-NF10	CYP3A4				Mammalian	Homo sapiens	human liver	microsome (ER)		T224I, V225del, V392W, I431T

Fig.1.: Area "Basic data"

Following issues should be considered to obtain a consistent appearance:

For each mutant/ wild type activity please start a new row!

Always enter wild type activities first (required for the data import procedure)

BASIC DATA:

- Username: = Last Name
- Date: Date of entry (format: YYYYMMDD e.g.: 20101024)
- MuteinName: consists of wild type-Name, hyphen, mutated position (e.g.: CYP2D6-R440H); if the mutein contains more mutations, use slash (e.g. CYP2D6-R440H/S486T); naturally existing mutants have own names (e.g. CYP2D6.31); Sometimes authors worked with muteins containing more then three mutations and gave them individual names, in this case assume that name out of literature. (if necessary add wild type name in front e.g.: literature 132-10 → CYP2D6-132-10);
-

Example	**Entry**
one Mutation (R440H)	CYP2D6-R440H
more Mutations (R440H and S486T)	CYP2D6-R440H/S486T
naturally occurring mutant (due to polymorphism)	CYP2D6.31
individual names (132-10)	CYP2D6-132-10
Fusion Protein (CYP2D6 fused with CPR)	CYP2D6-CPR

Tab.1.: Entry of mutein names

If a wild type is entered, leave "MuteinName" empty

- Wild type Name: for muteins always specify corresponding wild type (e.g. CYP2D6)
- GenBankID: (accession number, wild type: mandatory because of sequence import; muteins: only if explicit for the mutein)
- UniProtID (accession number, wild type: mandatory because of linkage; muteins: only if explicit for the mutein)
- PDB-ID (wild type: if available; muteins: only if explicit for the mutein) if there are several PDB entries the first should be the one containing cofactors but no ligands (natural form), add the others separated by comma and space (e.g. 1rg5, 6fs9)
- Kingdom = organism kingdom (use dropdown menu e.g. mammalian)

- O-Source = name of the origin organism (“complete name”, eg. Homo sapiens; if a strain designation is known: “complete name” hyphen “strain designation” e.g. Saccharomyces cerevisiae-AH22)
- Tissue, Localization = natural occurrence
  (Tissue: e.g. humane liver; Localization: use dropdown menu e.g. microsomes (ER))
- PTMs = Post translational modifications of the enzyme (Multiple PTMS are separated by comma and space eg: phosphorylation, glycosylation)
- Mutations = mutated position: = **mandatory for muteins**!
  Numbering: Sequence includes the start Methionine (M = Number 1)
  (several mutations are listed using comma and space e.g. R44H, D101F;
  muteins, that contain one or more changes, separate by using coma and space e.g. R44H, T97-W98insLQS)

Mutation	Example	Description
Substitution	R440H	R440 is substituted to H
Deletion* (del)	T97-C102del	T97 – C102 is deleted
	T97del	T at position 97 is deleted
Insertion* (ins)	T97-W98insLQS	Between T97 and W98 L, Q and S are inserted
fusion (fus)	CfusHCPR NfusHCPR	fusion of the enzyme HCPR (humane cytochrome P450 reductase) at the C-terminus / N-terminus
truncation (tru)	Ctru10 Ntru10	10 AS are cut of at the C-terminus N-terminus

Tab.2.: Entry of mutations

*Source: JT den Dunnen and SE Antonarakis, 2000, Human mutation 15: 7-12

SIGNAL SEQUENCE, PH, TEMPERATURE, STABILITY

- Signal Sequences: N- and C-terminal sequences (amino acids; e.g. PPLLLALV) (Tag's, e.g. His-Tag, should be stated in the comment column of the reaction section)
- pH: conditions of pH tests (buffer, temperature, pH min and max, resulting optimum pH)
- temperature: conditions of temperature tests (buffer, temp min and max, resulting optimum temperature) – unit of temperature is always °C
- Stability (e.g.stable or unstable)

REACTION

- Substrate ID Product ID → (if unknown type unknown)
  preferred CAS-number (look at www.scifinder.at or http://ctd.mdibl.org); e.g. 54340-62-4,
  alternatively CID (look at http://pubchem.ncbi.nlm.nih.gov/); e.g. CID 1615
  !Remember: important for Structure Search function
- S-Name = Substrate, e.g. bufuralol
- P-Name = Product, e.g. 1-hydroxybufuralol
  !Attention: chemical names start with lower case character

For molecules, that contain alpha, beta etc., please don't use special signs, write out the whole name (e.g. 16alpha-hydroxyprogesterone)!

If products (or reaction types) are unknown, register them as "unknown" (field for Product ID = unknown, don't enter characters like "-", "ND" etc.)
If the structure is given and just named with a number (e.g. M1, 2a,….), enter the chemical name, otherwise register them as "unknown".

Although some conversions give several products in one reaction, nevertheless create an own row for each product.

If a Protein is inactive towards a certain substrate, create a normal entry (observed reaction, e.g. Substrate, Unit etc.).
For Product enter “no product” and enter “0” for Activity Value.

- EC-Number (if unknown take a look at http://www.brenda-enzymes.info/)
- Reaction type: use dropdown-menu, if the reaction type is not there; add the reaction type in the “dropdown-table” (e.g. hydroxylation, no position, except (N-demethylation, O-demethylation, etc.)

AE	AF	AG	AH	AI	AJ
Reaction					
Substrat-ID	**S-Name**	**Product-ID**	**P-Name**	**EC**	**ReactionType**
2921-88-2	chlorpyrifos	5598-15-2	chlorpyrifos oxon	1.14.14.1	S-oxidation
2921-88-2	chlorpyrifos	5598-15-2	chlorpyrifos oxon	1.14.14.1	S-oxidation
439-14-6	diazepam	1088-11-5	nordiazepam	1.14.14.1	N-demethylation
439-14-6	diazepam	1088-11-5	nordiazepam	1.14.14.1	N-demethylation
59467-70-8	midazolam	59468-90-5	1-hydroxymidazolam	1.14.14.1	hydroxylation
59467-70-8	midazolam	59468-90-5	1-hydroxymidazolam	1.14.14.1	hydroxylation
21829-25-4	nifedipine	CID 128753	nifedipine M (dehydro)	1.14.14.1	oxidation
21829-25-4	nifedipine	CID 128753	nifedipine M (dehydro)	1.14.14.1	oxidation
137-58-6	lidocaine	7728-40-7	nonoethylglycinexylidid	1.14.14.1	N-deethylation
137-58-6	lidocaine	7728-40-7	nonoethylglycinexylidid	1.14.14.1	N-deethylation
137-58-6	lidocaine	7728-40-7	nonoethylglycinexylidid	1.14.14.1	N-deethylation

Fig.2.: Area “Reaction”

ACTIVITY

- Entry Type: Type of entered data (use dropdown menu e.g. activity, inhibition, enantiomeric excess, etc.)
- Inhibitor ID: (if not known type unknown)
  CAS-number (look at www.scifinder.at or http://ctd.mdibl.org); e.g. 56-54-2,
  CID (look at http://pubchem.ncbi.nlm.nih.gov/); e.g. CID 1615
- I-Name = Inhibitor, e.g. quinidine
  !Attention: chemical names start with lower case character
- K-Value: enter K_M, K_I always use **µM**! (no special sign’s, e.g. “>”) or IC_{50} (IC_{50}-Values always with the prefix “IC50” and “space” eg. IC50 35)

- Value + Unit: enter activity value (no special sign's, e.g. >, "0" for inactive enzymes); use dropdown menu to choose unit, for activity preferentially use the unit **pmol/min/pmol**!!!!; (to convert mg Protein to pmol we recommend to use http://www.molbiol.ru/eng/scripts/01_04.html for the calculation)
  if there is no activity per Protein given, also "nmol/min/mg total protein", "nmol/min/g CDW" are allowed;
  enantiomeric excess in % e.e; E-Value has no Unit (empty field)
  → leave empty for entry type "inhibition" (if activity data for the inhibition reaction is given, add them to the comment filed)
- Inactive enzymes: give the "Entry type", and the Unit and put as value = 0
- Conversion % and Time:
  only in combination with enantionmeric excess or E-value
- Active sites: number of active sites (for CYP2D6 = 1)

AK	AL	AM	AN	AO	AP	AQ	AR	AS	AT	AU	AV	AW
											Activity	
Entry type	Inhibitor-ID	I-Name	K-Value [µM]	Value	Unit	Conversion [%]	Time [min]	ActiveSites	Method	Temp [°C]	Buffer	Solvent
inhibition	65277-42-1	ketoconazole	IC50 0.4					1	HPLC	37	50 mM HEPES (pH 7.4)	unknown
inhibition	65277-42-1	ketoconazole	IC50 4.31					1	HPLC	37	50 mM HEPES (pH 7.4)	unknown
activity			135	11	pmol/min/pmol			1	HPLC	37	50 mM HEPES (pH 7.6)	unknown
activity			129	73	pmol/min/pmol			1	HPLC	37	50 mM HEPES (pH 7.6)	unknown
enantiomeric excess				53	% e.e.	90	120	1	HPLC	37	100 mM potassium phosphate (pH 7.4)	methanol/water
enantiomeric excess				60	% e.e.	90	120	1	HPLC	37	100 mM potassium phosphate (pH 7.4)	methanol/water
activity				53	pmol/min/pmol			1	HPLC	37	100 mM potassium phosphate (pH 7.4)	methanol/water
activity				0	pmol/min/pmol			1	HPLC	37	100 mM potassium phosphate (pH 7.4)	methanol/water
activity				1.3	pmol/min/pmol			1	HPLC	37	50 mM Tris/HCl (pH 7.4)	water/acetonitrile
activity				2	pmol/min/pmol			1	HPLC	37	50 mM Tris/HCl (pH 7.4)	water/acetonitrile
activity				0	pmol/min/pmol			1	HPLC	37	50 mM Tris/HCl (pH 7.4)	water/acetonitrile

Fig.3.: Area „Activity"

- Reaction conditions → Method (e.g. HPLC), Temp.(always in °C), Buffer: use dropdown menu, if the buffer is not there, add the buffer in the "dropdown-table" (e.g. 50 mM potassium phosphate (pH 7.2))
- Solvent for Substrate (e.g. methanol)
  (if unknown type "unknown", for Solvent mixtures separate the solvents with a "slash" e.g. methanol/water, do not enter specific data like concentration etc.)
  !Attention: chemical names start with lower case character
- Comment: special information about the reaction, that can't be registered anywhere else (e.g. if values where estimated from graphics, special reaction conditions, Tag's e.g. His-Tag, …..)

- CoFactors and –proteins: present during the described reaction (e.g. for CYP2D6: Cofactor = NADPH, heme, iron; Coprotein = human P450 reductase → in parenthesis: if this protein is co-expressed (ce) or added to reaction (atr)). (Multiple Cofactors – Proteins are separated by comma and space eg. NADPH, FAD)
- Expression host: organism, in which recombinant DNA was expressed use dropdown menu, if the host is not there, add the host in the "dropdown-table" ("complete name" hyphen "strain designation"; e.g. Escherichia coli-JM109)
- Reference ID: please give information about the literature source; either PubmedID or DOI-ID should appear for each entry (**preferred ID = PubMed,** DOI with prefix and "space" e.g. DOI 10.1248/jhs.50.503)!
  Why PubMed ID: a link to the PubMed entry will be created during the import process

		Wildtype	Mutein	Comments
Basic Data	UserName	mandatory	mandatory	
	Date	mandatory	mandatory	
	MuteinName	**leave empty**	mandatory	
	WildtypeName	mandatory	mandatory	
	GenBankID	mandatory		
	UniProtID	mandatory		
	PDB-ID			
	Kingdom	mandatory	mandatory	
	O-Source	mandatory	mandatory	
	Tissue			
	Localization			
	PTMs			
	Mutations		mandatory	sequence includes start methionine (M gets the number = 1)
SignalSequences	N-Terminal			
	C-Terminal			
pHRange	pHMin			
	pHMax			
	OptimumpH			
	Temp [°C]			
	Buffer			
	Comment			
Temp	TempMin [°C]			
	TempMax [°C]			
	OptimumTemp [°C]			
	pH			
	Buffer			
	Comment			
Storage Stability	RT			
	+4			
	-20			
Reaction	Substrat-ID (CAS or CID)	mandatory	mandatory	
	S-Name	mandatory	mandatory	„**unknown**“ if unknown
	Product-ID (CAS or CID)	mandatory	mandatory	
	P-Name	mandatory	mandatory	
	EC	mandatory	mandatory	
	ReactionType	mandatory	mandatory	„**unknown**“ if unknown

Activity	Entry Type	mandatory	mandatory	
	Inhibitor-ID	mandatory	mandatory	in case of inhibition
	I-Name	mandatory	mandatory	in case of inhibition
	K-Value [µM]			inactive value = 0
	Value			
	Unit			
	Conversion %			
	Time [min]			
	ActiveSites	mandatory	mandatory	in most cases „1“
	Method			
	Temp			
	Buffer			
	Solvent			
	Comment			
	CoFactor			
	CoProtein			
	ExpreHost	mandatory	mandatory	
	Reference ID (PubMedID, DOI)	mandatory	mandatory	use PubMedID or DOI

Please send your actual file to Gerhard Thallinger.

Gerhard Thallinger
Institute for Genomics and Bioinformatics
Graz University of Technology
Tel: +43-316-873-5343
Fax: +43-316-873-5340
Mail: MuteinDB@genome.tugraz.at

ENTRY IN MUTEIN-DATABASE

If you want to register just one or a few entries, it's easier to use the „Create"-application at http://www.MuteinDB.org
(Please request accession data at Gerhard Thallinger, only for registered users)

The general entry rules are the same like for the Excel File, but the surface and the basic areas are a bit different:

1. Create → Choose, if you want to register Mutein or Wild type data
2. **Basic information**: Give Name, IDs (as described in area „Basic data" in the Excel file), choose a kingdom, Tissue, Localization, O-Source (= natural habitat) and the expression host in the drop down menu or create a new one („New button") but please avoid double entries! Please add your user ID in „Availability/Contact".
3. **Properties**: as described in the area pH, Temperature and Stability of the Excel File; for new Inhibitors: „Add inhibitor";
4. **Substrates**: choose or add Substrates and Products (don't forget CAS- or CID-numbers!), activity in pmol/min/pmol, enantiomeric excess (ee%), Km in µM; choose or add "Cofactor", "Coprotein"; or Expression host
   Literature source,
5. **Sequence**; area for mutations (AA mutation, codon mutation), signal sequences

Summary of entry format:

Topic	Format/unit	example
Date	year month day	20101024
K-value (K_M, K_I, IC_{50})	µM	53, IC50 53
Activity Units	[pmol/min/pmol] only if not otherwise possible [nmol/min/mg total Protein], [nmol/min/g CDW]	-
Buffer	mM buffer name (pH)	100 mM potassium phosphate (pH 7.4)
Reference ID	PubMed ID (alternative = DOI with prefix)	16269134 DOI 10.1248/jhs.50.503
CAS	alternative = CID with prefix	33817-09-3 CID 1615
Expression Host	"complete name" hyphen "strain designation"	Escherichia coli-DH5alpha