Welcome to the Target Protein Database
June, 2016 Download PDF copy

The TPDB is an attempt to catalog in a convenient, searchable fashion all publicly available information about the identities of mammalian proteins that become covalently adducted by chemically reactive metabolites (CRMs) derived from xenobiotic agents including drugs. Reasons for interest in this information are elaborated below. We hope you find the TPDB to be useful. We would appreciate your comments as well as suggestions for improvement - send them to rhanzlik@ku.edu.

Contents:
1. Chemically reactive metabolites (CRMs) and cytotoxicity
2. The Reactive Metabolite Target Protein Database (TPDB)
3. Searching the TPDB
4. References Cited

1. Chemically reactive metabolites (CRMs) and cytotoxicity

The toxicity of numerous small organic molecules including drugs, agricultural- and industrial chemicals has been correlated with the extent to which they undergo biotransformation to chemically-reactive metabolites (CRMs), and in particular with the extent to which the latter express their reactivity in situ by covalently modifying endogenous cellular proteins. The process of CRM formation is broad in scope but well understood in terms of mechanistic organic chemistry and enzymology. Many excellent reviews of the subject are available (1-5)

The vast majority of protein-modifying CRMs can be classified as either electrophilic alkylating agents or electrophilic acylating agents. The latter group includes glucuronides of carboxylic acids, acyl- and thioacyl halides, and S-oxidized metabolites of thioamides and thioureas. The former include epoxides, solvolytically active sulfate ester or glucuronide derivatives of benzylic or allylic alcohols or hydroxylamines, diazoalkanes, alpha-halo ketones, or beta-haloethylthio compounds (i.e., sulfur mustards), and Michael acceptors including quinones, unsaturated esters and certain products of lipid peroxidation (see below).

Electrophilic metabolites are thought to react spontaneously with nucleophilic target sites in cellular proteins. These are most often found on the side chains of the amino acids cysteine (S), lysine (N) and histidine (N), but examples of adduct formation via methionine (S), tryptophan (C), tyrosine (C), serine (O), threonine (O) and carboxylate (O) are known.

In its simplest form protein adduction is a bimolecular chemical reaction and is therefore expected to follow ordinary second-order kinetics. However, the relative nucleophilicity of protein side chains varies widely as a function of the nucleophilic center (e.g. cysteine vs. serine), ionization state (e.g., Lys-NH$_2$ vs. Lys-NH$_3^+$), steric accessibility (surface vs. buried), intramolecular hydrogen bonding or ion pairing, and fortuitous positioning near a hydrophobic pocket that can hold the electrophile non-covalently and thus favor nearby nucleophiles over others.

By definition, CRMs do not accumulate as such; if they are not detoxified by enzymes or by chemical reaction with water or other abundant small nucleophiles such as glutathione,
protein adduction becomes more prominent. On the other hand, *protein adducts do accumulate over time*. Repair of adducted proteins, *per se*, is rare, a possible exception being acetylation, which in some instances can be physiological. The usual case is that damaged proteins are degraded and replaced by re-synthesis. Thus, the time course for adduct formation and disappearance is governed by the flux of reactive metabolite generation, the competition between adduct formation vs. other routes of CRM removal, and the rate of removal of adducted proteins.

Under normal conditions the turnover rates of cellular proteins vary widely, but little is known about whether xenobiotic adduction of a given protein specifically accelerates its degradation. In the few cases in which the extended time course of protein adduction has been measured following a single dose of CRM precursor in vivo, there is often an initial lag time of 3-6 hr (usually for glutathione depletion). This is then followed by an increase in adduct accumulation (almost always measured as total C-14) leading to a peak in net CVB at 12-24 hr, and finally a gradual decline in adducts over several days (in those animals or cells that survive).

Perhaps the least understood aspects of CRM toxicity are 1) the mechanism(s) by which covalent modification of cellular proteins is detected or recognized by a living cell, and 2) the mechanism(s) by which adduct formation leads to toxic consequences. Early mechanistic studies focused on identifying individual protein targets using classical chromatographic approaches to purify adduct-bearing (i.e. C-14 labeled) proteins and identifying them using Edman N-terminal sequencing. From 1978 through 1996 a total of five protein targets were identified in this way. The application of 2D gel electrophoresis, autoradiography, in-gel tryptic digestion and peptide mass-mapping to identify 23 targets of acetaminophen metabolites in mouse liver (6) literally revolutionized the fledgling field of proteomics, and especially the field of CRM target proteomics. This is illustrated in Figures 1a and 1b below, which show the cumulative number of target proteins identified and reported from 1976 through 2014 (note the difference in ordinate scales).

![Figure 1a (left) and 1b (right). Cumulative number of CRM target proteins identified and reported by year. Note the differences in scale.](image)

Advancing from peptide mass mapping by MS to peptide sequencing by LC-MS/MS greatly enhanced the certainty of protein identification, but generally proved insufficient to facilitate the direct observation of adducted peptides from the target proteins. Two major reasons for this are that 1) even in the most favorable cases the absolute levels of adduction are quite low, and 2) the adducts themselves may be distributed statistically across multiple nucleophilic sites in any given protein. For example, protein adduction in the livers of rodents
treated with acetaminophen or bromobenzene are in the order of 1-3 nmol-equivalents per milligram of total protein; for many other drugs the levels are 1-2 orders of magnitude lower (7). For a "typical" 50 kDa protein a CVB level of 2 nmol/mg (relatively high) corresponds to a mole fraction of adduction of only 0.10. This means that only one protein molecule in ten carries a single adduct that is statistically distributed across numerous potential nucleophilic sites in the protein. These factors make structure elucidation of individual adducts extremely challenging. Indeed, while many protein targets are now known, it is quite rare to know their specific site(s) of adduction.

Many CRM target proteins are metabolic enzymes. However, the very low levels of adduction suggests that even if adduction was inhibitory (which is usually not even known with certainty), enzyme inhibition per se is unlikely to be a major cause of cytotoxicity because the fractional loss of enzyme activity would be so small (8). Likewise, sorting target proteins into functional groupings or biochemical pathways has not significantly aided the understanding of CRM toxicity (9, 10). For a low level of adduction to exert powerful cytotoxic effects, it would seem logical that some sort of amplification mechanism would be required.

Such mechanisms often occur in intracellular signaling pathways that propagate via sequential protein-protein interactions and culminate in the activation of transcription factors. Many of these pathways are regulated physiologically by covalent but reversible post-translational modifications such as acetylation and phosphorylation. In some cases irreversible covalent modifications can also trigger amplification mechanisms. One example involves the protein Keap1, a protein that has multiple sulfhydryl groups and appears to function as a detector for electrophilic stress due to CRMs. Alkylation of -SH groups in Keap1 releases the transcription factor Nrf2 and initiates cytoprotective responses via gene expression (11-13). Like Keap1, the ion channel protein TRPA1 found on sensory neurons has multiple sulfhydryl groups that undergo alkylation by "irritant" molecules, most of which are electrophilic compounds (14, 15). In this case, modification of a small number of protein molecules can lead to a profound cellular response (i.e. depolarization).

Protein-protein interactions underlie many important cellular processes. Disrupting or initiating PPIs by CRM adduction could conceivably be a general way for cells to detect and initiate responses to the presence of CRMs. While direct information on this potential mechanism is lacking, this possibility was explored by building a list of all proteins known to partner with 302 CRM target proteins listed in this database. These two lists (targets and partners) were then merged and the result sorted into KEGG and GO categories. The populated categories were then ranked on the basis of statistics and likely toxicological relevance. This analysis highlighted many intracellular signaling pathways that are not detected as significant when only target proteins are considered. This result lends some credence to the notion that it is PPIs rather than proteins per se that are the real targets of many CRMs (16).

2. The Reactive Metabolite Target Protein Database (TPDB)

As noted above, the TPDB is a conveniently searchable catalog containing all publicly available information about the identities of mammalian proteins that become covalently adducted by CRMs derived from xenobiotic agents including drugs. At present all entries in the TPDB are well-identified proteins that became adducted by CRMs following defined exposures of living animals or cells in culture to drugs or chemicals as CRM precursors (i.e., "in-life" studies only). In most cases the exposures used were known to cause specific dose-dependent toxic responses.
We have not included results from studies that used only in vitro enzymatic or chemical model systems. For the most part we have also excluded results from studies which exposed living cells to intrinsically-reactive chemicals that do not require metabolic activation to react covalently with proteins. Such experiments often use un-physiologically high concentrations of the test agent, resulting in extensive adduction of large numbers of proteins, many of which may not be significantly adducted by lower fluxes of bone fide CRMs generated intracellularly in real time. Finally, we have also excluded information about the adduction of proteins by endogenous, lipid-derived electrophiles generated under conditions of "oxidative stress" (i.e., the chemical processes of lipid peroxidation or auto-oxidation). For examples of these electrophiles see (17-19).

3. Searching the TPDB

The TPDB provides search functions that are relatively simple and intuitive, so most users can go directly to the Search page and begin searching. Additional explanations, examples and references may be found on the Help page. Publications about the TPDB itself may be found in references (8, 16, 20, 21).

4. References Cited


