Chemico-Biological Interactions 390 (2024) 110876

Contents lists available at ScienceDirect

Chemico-Biological Interactions

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

Computational analysis of human medium-chain dehydrogenases/reductases revealing substrate- and coenzyme-binding characteristics

Linus J. Östberg [a,1], Jan-Olov Höög [b,*], Bengt Persson [c]

[a] Department of Medical Biochemistry and Biophysics, Science for Life Laboratory, Karolinska Institutet, SE-171 77 Stockholm, Sweden
[b] Department of Medical Biochemistry and Biophysics, Karolinska Institutet, SE-171 77 Stockholm, Sweden
[c] Department of Cell and Molecular Biology, Science for Life Laboratory, Uppsala University, SE-751 05 Uppsala, Sweden

ARTICLE INFO

Keywords:
MDR super-family
Coenzyme-binding
Substrate-binding
Functional prediction
Sequence comparison

ABSTRACT

The medium-chain dehydrogenase/reductase (MDR) superfamily has more than 600,000 members in UniProt as of March 2023. As the family has been growing, the proportion of functionally characterized proteins has been falling behind. The aim of this project was to investigate the binding pockets of nine different MDR protein families based on sequence conservation patterns and three-dimensional structures of members within the respective families.

A search and analysis methodology was developed. Using this, a total of 2000 eukaryotic MDR sequences belonging to nine different families were identified. The pairwise sequence identities within each of the families were 80–90 % for the mammalian sequences, like the levels observed for alcohol dehydrogenase, another MDR family. Twenty conserved residues were identified in the coenzyme part of the binding site by matching structural and conservation data of all nine protein families. The conserved residues in the substrate part of the binding pocket varied between the nine MDR families, implying divergent functions for the different families. Studying each family separately made it possible to identify multiple conserved residues that are expected to be important for substrate binding or catalysis of the enzymatic reaction.

By combining structural data with the conservation of the amino acid residues in each protein, important residues in the binding pocket were identified for each of the nine MDRs. The obtained results add new positions of interest for the binding and activity of the enzyme family as well as fit well to earlier published data. Three distinct types of binding pockets were identified, containing no, one, or two tyrosine residues.

1. Introduction

The advent of whole genome sequencing, RNA-seq, and other high-throughput methods have led to a massive amount of new genomic data. Along with the genomic data there has been an enormous increase in the number of known genes, and thus translated protein sequences, as shown by the UniProt database now containing almost 250 million protein sequences (March 2023) [1]. The influx of new data also means that while many proteins have well-known characteristics, even more have unknown functions, and need to be investigated.

One protein superfamily that has been steadily increasing in size is that of medium-chain dehydrogenases/reductases (MDR). In 2010, there were 15,000 MDR members in UniProt [2]. By March 2023, this number had grown by an order of magnitude, the superfamily now comprising more than 600,000 members, covering all major taxonomic branches, including both prokaryotes and eukaryotes.

The MDR superfamily can be divided into 86 unique clusters, each comprising a protein family, e.g. alcohol dehydrogenase (ADH, EC 1.1.1.1), sorbitol dehydrogenase (SDH, EC 1.1.1.14), and quinone oxidoreductase (QOR, EC 1.6.5.5) [2]. Some of the families have been thoroughly investigated, while others have not been confirmed at the protein level.

An MDR protein typically consists of about 350 amino acid residues, though there are exceptions, and while the families differ greatly at the protein sequence level, the enzymes with known structure all have a similar fold, containing two domains; one *C*-terminal coenzyme (NAD(H)/NADP(H)) binding domain, formed by the common Rossmann fold [3] with typically six β-strands forming a β-sheet with α-helices above and below, and one substrate-binding *N*-terminal domain consisting of

* Corresponding author.
E-mail addresses: linus.ostberg@conoa.se (L.J. Östberg), jan-olov.hoog@ki.se (J.-O. Höög), bpn@icm.uu.se (B. Persson).
[1] Present address: Conoa AB, PO 4061, SE-169 04 Solna, Sweden.

https://doi.org/10.1016/j.cbi.2024.110876
Received 1 November 2023; Received in revised form 29 December 2023; Accepted 15 January 2024
Available online 22 January 2024
0009-2797/© 2024 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

List of abbreviations

ADH	Alcohol dehydrogenase
HMM	Hidden Markov model
MDR	Medium-chain dehydrogenase/reductase
MECR	Enoyl-[acyl-carrier-protein] reductase, mitochondrial
PTGR1	Prostaglandin reductase 1
PTGR2	Prostaglandin reductase 2
PTGR3	Prostaglandin reductase 3
QOR	Quinone oxidoreductase
QORX	Quinone oxidoreductase PIG3
RT4I1	Reticulon-4-interacting protein 1, mitochondrial
SDH	Sorbitol dehydrogenase
VAT1	Synaptic vesicle membrane protein VAT-1 homolog
VAT1L	Synaptic vesicle membrane protein VAT-1 homolog-like

antiparallel β-strands with α-helices at the surface, similar to the GroES fold [4]. The binding pocket is located inside the cleft between the two domains. MDRs often form dimers, as is the case for mammalian ADHs, but both monomeric (e.g. mannitol dehydrogenase) and tetrameric (e.g. SDH) members are known to exist. The common reaction is alcohol + $NAD(P)^+$ ↔ aldehyde/ketone + NAD(P)H, or alkanal + $NAD(P)^+$ ↔ alkenal + NAD(P)H.

As of June 2023, 169 of the MDR members listed in UniProt had an X-ray crystal structure available in PDB [5], multiple orders of magnitude fewer than the number of MDR sequences in UniProt. The large gap between the number of known sequences and known structures is a well-known problem in the biochemical field. One way to address this is to use *in silico* methods to make models of the structures, e.g. by homology modelling.

This investigation focuses on the *in silico* analysis of nine MDR members, chosen from the 18 known human MDRs that are divided into twelve MDR families [2]. The seven ADHs (forming one family), SDH, and fatty acid synthase were not included as they are generally well-characterized, and fatty acid synthase is a complex enzyme with multiple domains [6]. The proteins analyzed were MECR, PTGR1, PTGR2, PTGR3, QOR, QORX, RT4I1, VAT1, and VAT1L (see Table 1).

2. Methods

An analysis strategy was developed, automating most of the steps described here. All functionality was implemented using Python 3 with pipeline management of the multiple scripts done using BASH 4. The code is available at https://github.com/talavis/mdr_analysis.

2.1. Estimating the total number of MDR members in UniProt and PDB

Hidden Markov Models (HMMs) of ADH_N (Alcohol dehydrogenase GroES-like domain; PF08240), ADH_N_2 (*N*-terminal domain of oxidoreductase; PF16884), and ADH_zinc_N (Zinc-binding dehydrogenase; PF00107) were obtained from the Pfam database [7]. HMMER version 3.1b2 [8] was used to find members in UniProt and the NCBI pdbaa BLAST database. The thresholds used to evaluate the hits were the official thresholds for each of the HMMs (21.0, 31.1, and 30.2). The number of hits was calculated by parsing the HMMER table output.

2.2. Identifying MDR members in UniProt

The amino acid sequences of the human forms of the nine MDR proteins; MECR, PTGR1, PTGR2, PTGR3, QOR, QORX, RT4I1, VAT1, and VAT1L (Table 1) were obtained from UniProt. The sequences were used for a blastp (version 2.4.0) [9,10] search against UniProt release 2023_03. The hits were filtered according to their assigned protein name, where all sequences better than the lowest scoring hit with the correct protein name were considered hits. The accuracy of this approach was confirmed by evaluation of multiple sequence alignments using the L–INS–i approach of MAFFT 7.304 b [11].

The sequences were further filtered, removing all sequences with more than 10 % unknown ("X") residues and sequences whose length deviated from that of the human sequence with more than 50 %. To avoid sequence bias, only one sequence per species, the one with the highest similarity to the human sequence, was retained. All non-eukaryotic sequences were removed using the taxonomy definitions from UniProt. Finally, the classification was confirmed by manual inspection of sequence alignments and phylogenetic trees (generated as described below).

2.3. Conservation

The residue identity was calculated at three different levels (mammals, vertebrates, and eukaryotes) for each of the nine proteins. The species filtering was performed using taxonomy definitions from UniProt, selecting the sequences that were part of the relevant branch of the taxonomy.

The sequences were aligned using the L–INS–i approach of MAFFT 7.304 b. The average sequence identity was calculated by performing pairwise comparisons of all sequences in the alignment, similar to what was previously performed for class V ADH [12]. All positions with gaps in any of the sequences were ignored for the calculation to compensate for e.g. trailing *N*-terminal sequences.

The positional conservation was also calculated for the alignments, focusing on the positions present in the human form of the protein (non-gap positions in the human protein sequence).

2.4. Phylogenetic trees

Phylogenetic trees were generated using PhyML [13] for each of the nine MDR families. Further, one tree was generated for all nine MDRs families together. In order to decrease the computational complexity to generate this tree, UCLUST 10.0.240 [14] was used to cluster the sequences at 90 % sequence identity, using one representative sequence from each cluster to generate the tree.

Table 1
Summary of the information about the nine human MDRs available in UniProt as of June 2023 [1].

Protein	Abbrev.	UniProt (human)	EC number	Substrate	PDB
Enoyl-[acyl-carrier-protein] reductase, mitochondrial	MECR	Q9BV79	1.3.1.104	*Trans*-2-hexenoyl-CoA	1ZSY
Prostaglandin reductase 1	PTGR1	Q14914	1.3.1.48/1.3.1.74	15-oxo-PGE1	2Y05
Prostaglandin reductase 2	PTGR2	Q8N8N7	1.3.1.48	15-keto-PGE2	2ZB4
Prostaglandin reductase 3	PTGR3	Q8N4Q0	1.3.1.48	15-keto-PGF2-alpha	2C0C
Quinone oxidoreductase (Zeta-Crystallin)	QOR	Q08257	1.3.1.27/1.6.5.5	1,2-naphthoquinone	1YB5
Quinone oxidoreductase PIG3	QORX	Q53FA7	1.6.5.5	1,2-beta-naphthoquinone	2J8Z
Reticulon-4-interacting protein 1, mitochondrial	RT4I1	Q8WWV3	–	Unknown	2VN8
Synaptic vesicle membrane protein VAT-1 homolog	VAT1	Q99536	1.-.-.-	Unknown	6K9Y
Synaptic vesicle membrane protein VAT-1 homolog-like	VAT1L	Q9HCJ6	1.-.-.-	Unknown	4A27

2.5. Identifying positions of importance for the binding pocket

The X-ray structures of the human forms of the proteins (MECR: 1 ZSY, PTGR1: 2Y05, PTGR2: 2ZB4 [15], PTGR3: 2C0C, QOR: 1YB5, QORX: 2J8Z [16], RT4I1: 2VN8, VAT1: 6K9Y [17], VAT1L: 4A27) were retrieved from the RCSB PDB [5]. The structures were imported into ICM version 3.8-6a and their binding pockets were mapped using the icm-PocketFinder method [18,19]. The list was mapped against the positional conservations, identifying highly conserved residues located inside the binding pocket.

The dataset was manually inspected and positions that both had a high level of conservation (more than 90 %) in at least five of the nine families and were present in the coenzyme part of the binding pocket were considered to be of importance.

The positions were then manually evaluated in the first subunit (identified as subunit A in the PDB files) of the structures using the PyMOL Molecular Graphics System, version 1.8.6 (Schrödinger, LLC, New York city, NY) and ICM version 3.8-6a. Further investigation of the substrate-binding part was performed by investigating the positions with a high level of conservation within each protein (rather than the conservation among all nine proteins).

2.6. Analyzing the GXGXXG motif

The positions corresponding to the GlyXaaGlyXaaXaaGly (GXGXXG) motif [20] (positions 150–155 in human PTGR1) involved in the binding of the coenzyme were extracted from the alignments of the eukaryotic sets of each of the nine proteins and a sequence logo was generated using Skylign [21]. The logos were analyzed and the most common residues at each of the six positions retrieved. The prevalence of the common residues was then calculated from the alignment.

3. Results and discussion

A total of 2000 eukaryotic MDR protein sequences belonging to nine MDR families were identified and investigated using the methodology developed in this study. Most of the proteins could be found in nearly all taxonomic branches among eukaryotes, except for PTGR2, which was only discovered in animals (metazoa). Of note is also that QORX was only discovered among euteleostomi at the vertebrate level, but also in non-vertebrate branches at the eukaryotic level. The discovered sequences were aligned to confirm that the protein family assignment was correct (to avoid false positives), yielding alignments with clear similarities at highly conserved positions.

The phylogenetic tree formed of the MDR protein sequences clearly results in the nine studied MDR families form distinct branches (Fig. 1). As expected, QOR and QORX, VAT1 and VAT1L, and the three PTGRs are all found in neighboring branches in the tree. The sequences of VAT1 and VAT1L are closely grouped, and several of the sequences could be identified by BLAST searches for both proteins. When all sequences were multiply aligned the VAT1 and VAT1L families could be separated, and in the tree, it corresponds to a distinct branching point (bootstrap value 1.0, Fig. 1), except for one VAT1L sequence that ended up in the VAT1 branch. VAT1 and VAT1L are thus closely related but form distinct protein families.

The intra-family sequence identity for all nine protein families showed a high value in mammals at least 80 %, with the highest, 89.3 %, observed for VAT1. This level of sequence conservation is comparable to that of class III ADH, another MDR that is known to be highly conserved (93.4 % in mammals) and involved in the elimination of formaldehyde [22]. Most of the other families have a level of conservation in the range of 80–88 %, like that of class I ADH (83.6 %), which is involved in the metabolism of ethanol to acetaldehyde, also of importance for many species [22]. Thus, the relatively high conservation levels imply that all nine families should be expected to have specific functions, especially in mammals. The pairwise sequence identities calculated for each of the

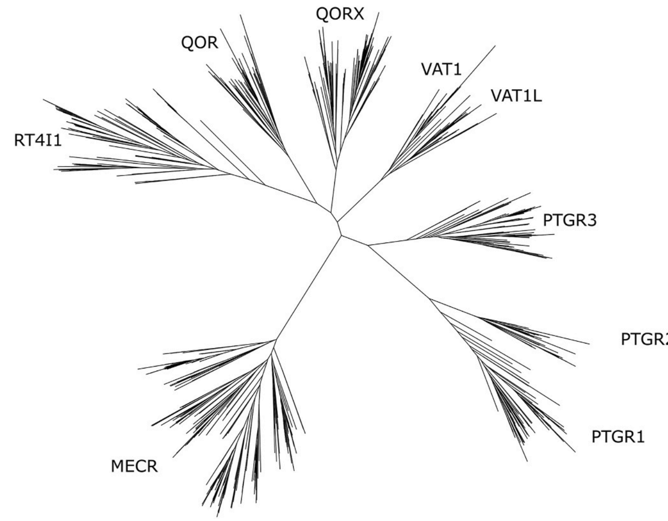

Fig. 1. Phylogenetic tree of the nine MDRs. Bootstrap values of 1.0 correspond to the branching points of VAT1, VAT1L and PTGR1/2.

families at three levels: mammals, vertebrates, and eukaryotic species resulted in the range 80–89 % for mammals (like the values for the ADHs), for the vertebrate level 70–92 %, and for the eukaryotic level 45–70 % (Table 2).

The X-ray structures were of varying quality for the intended analysis, some being high-resolution dimers with bound coenzymes (e.g. PTGR3: 2C0C; 1.45 Å), while others were monomeric, missing loops, and sometimes displaying unexpected structural properties, e.g. RT4I1: 2VN8, which formed a dimer using an α-helix, rather than the traditional β-sheet. The structures were mapped for potential binding sites using the ICM software, identifying the residues forming the binding pocket. The size of the identified binding pockets varied between the proteins, from 746.7 $Å^3$ (RT4I1; 2VN8) to 1,562 $Å^3$ (QOR; 1YB5). The variation can probably be explained by the substrate specificity of the protein, whether the protein conformation was "open" and containing substrates, how many chains were included, and the varying quality of the structures. Conserved residues near the binding pocket were identified by combining sequence and structure data. A total of 20 highly conserved residues were identified to be situated inside the coenzyme part of the binding pocket (Table 3). Eight of the 20 residues were highly conserved across the protein families (>70 % of the proteins had the same conserved residue). The identified residues are distributed over the whole coenzyme part of the binding pocket, which is easily seen when the conserved positions are visualized (Fig. 2).

Table 2
Conservation among the nine MDR families, are presented as values for mammals/vertebrates/eukaryotes.

Protein	Conservation[1]	Conservation rate[2]	Phylogenetic groups
MECR	81.7/70.3/43.8	Medium	Eukaryota
PTGR1	79.8/69.6/57.7	Low	Eukaryota
PTGR2	87.9/75.0/65.3	Medium	Metazoa
PTGR3	85.5/78.8/52.4	Medium	Eukaryota
QOR	82.3/72.5/58.1	Medium	Eukaryota
QORX	85.5/71.9/48.4	Medium	Eukaryota/Euteleostomi[3]
RT4I1	89.1/76.5/44.9	High	Eukaryota
VAT1	89.3/73.1/70.2	High	Eukaryota
VAT1L	88.5/91.6/67.2	High	Metazoa

[1] Average of pairwise amino acid residue identities, in percent.
[2] Conservation in mammals compared to that of alcohol dehydrogenase (ADH) as three levels (High: ADH3 (93.4), Medium: ADH1 (83.6), Low: ADH5 (77.8)) [12].
[3] Only found in euteleostomi among vertebrates, but also among other eukaryotic groups.

Table 3
Conserved positions localized inside the coenzyme-binding part of the binding site in each protein and the fraction of the proteins that have the most common residue.

Position[1]	Residues[2]	Prevalence[3]
47	NSD	70 %
48	PFRQA	50 %
124	(MV)EIRY	30 %
125	(NP)DAVS	30 %
128	TR	90 %
149	AGMDS	40 %
151	ASNGL	50 %
152	GS	80 %
153	GA	70 %
154	VCTF	70 %
155	G	100 %
172	TVAI	60 %
173	ACV	50 %
174	(SG)CR	40 %
178	KADGL	60 %
193	YVHER	60 %
239	YCVLI	40 %
240	GT	90 %
242	IMGVASR	30 %
272	LVN	70 %

[1] Positions are numbered according to the human form of PTGR1.
[2] The most conserved residue in each of the proteins, ordered by fraction of proteins that has the residue as the most conserved.
[3] Fraction of the proteins that has the first residue as the most common one.

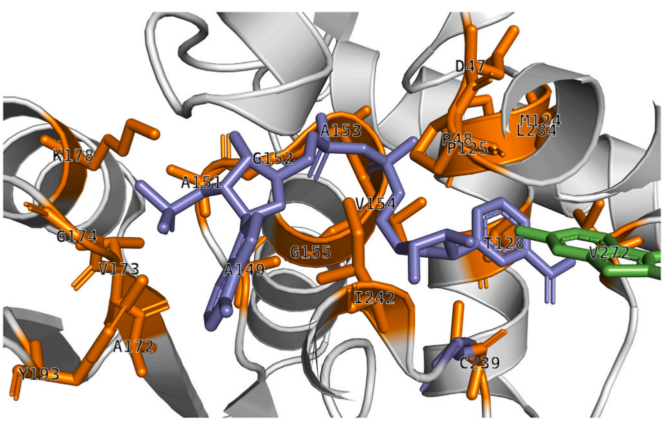

Fig. 2. Conserved amino acid residues located near NADP in the binding pocket (Table 3) visualized in the structure of PTGR1 (pdb id: 2Y05). Orange: the conserved amino acid residues; blue: NADP, green: raloxifene (substrate/inhibitor).

The coenzyme interacting GXGXXG motif [20,22], commonly found in ADH enzymes, was identified in the alignments of each of the nine proteins and the most common residues and their prevalence were calculated (Table 4). In the non-ADH MDRs, the motif can more accurately be described as AXGXXG, as Ala is the most common residue at position one in all but three of the nine MDR families. QORX has a GXSXXG motif, or even a GXSGXG motif, as the added Gly has more than 99 % conservation [16]. The GXXGXXG variant of the motif is also present in PTGR2 and RT4I1. All nine MDRs have a unique pattern of residues, and multiple positions have more than 90 % intra-family sequence identity. Largely, the motifs seem to hold true even at the MDR level, but, as can be seen in Table 4, the most common motif for MDRs is (A/G)X (S/G)XXG. A common theme is that the whole motif contains small residues, i.e. Gly, Ala, Ser, and Thr, as well as often Val before the final Gly. Interestingly, all nine of the investigated proteins had a unique pattern of conserved residues in this region. This shows the importance of this pattern in coenzyme interaction in this specific region, but also that it can be solved in different ways for the entire MDR superfamily.

Table 4
The GXGXXG motif among the nine proteins investigated in this study. Seven positions are shown to match the varying lengths observed in the nine protein families (i.e. adding one position ahead of the GXGXXG motif).

ADH[1]	X	G	X	G	X	X	G
MECR	X	A*	N*	S†	X	V†	G§
PTGR1	X	A†	A†	G§	A†	V†	G†
PTGR2	G†	A†	A†	G†	A*	C†	G§
PTGR3	A§	A§	A†	G§	G†	T†	G§
QOR	G†	A†	S†	G†	G*	V†	G§
QORX	A*	G†	X	S§	G§	V*	G*
RT4I1	G†	X	X	G§	G*	V*	G§
VAT1	M*	A§	A§	G§	G†	V†	G†
VAT1L	S†	A*	G†	G§	G†	V†	G§
Consensus	X	A/G	X	G/S	X	X	G

*residue found in at least 80 % of sequences
†residue found in at least 90 % of sequences
§residue found in least 99 % of sequences.
[1] General motif in ADHs

The residues forming the coenzyme part of the binding pocket were, as could be expected, easily identified by matching the mapped residues to sequence conservation. Most positions had a high level of variation among the different positions, but Asn47, Thr128, Gly152, Gly153, Val154, Gly155, Gly240, and Leu272 (PTGR1 numbering) were found in most of the nine protein families, implying functional importance for the interaction with the coenzyme. Of those, the most conserved was Gly155 which had a conservation rate greater than 98 % for all but two proteins (MECR and QORX), which instead had a Ser at the same conservation levels. The positions with conserved residues, but with a variety of residues between the proteins, also imply a functional importance, but probably as a combination of residues within each protein rather than specific residues.

The coenzyme binding positions identified here fit well with UniProt annotations. In some cases, the regions rather than positions are annotated as coenzyme-interacting, while the methodology used here focused on positions. Further, multiple positions were identified that are not listed as coenzyme-interacting in the databases. These residues probably do not interact with the coenzyme, but their localization and high level of conservation imply that they could be important for the structure of the coenzyme part of the binding pocket.

In the zinc containing MDRs families, ADHs and SDH, Asp223 (ADH1 numbering) is classified as determinant for the usage of NAD(H) as coenzyme within the MDR superfamily [2,22]. A corresponding Asp is not found in any sequence among the nine families investigated in this study, confirming both its accuracy and that all nine families use NADP (H) as coenzyme instead. Interestingly, this also means that eight (seven ADHs and SDH) out of eighteen human MDRs use NAD(H), while the remaining ten use NADP(H). At the family level however, only two out of twelve families use NAD(H), implying that the majority of all MDRs may use NADP(H) rather than NAD(H). In addition, the NAD (H)-dependent MDRs harbor zinc as part of the catalytic site, while a zinc-dependent catalysis has not been found for any NADP (H)-dependent eukaryotic MDR [2,22].

While the coenzyme part of the binding pocket had a high level of similarity among the nine families, the substrate part did not. In most cases, the positions that were conserved in most of the nine families were not located near the substrate part, and the conserved positions that were in the vicinity of the substrate part varied greatly between the protein families. To compensate for this, conservation was instead studied at a per-family level, rather than all nine families together. After mapping the conserved residues to the structure, 4–20 residues per protein were identified as potentially relevant for the substrate binding pockets (Table 5, Fig. 3). Most proteins had at least one conserved Tyr present at the active site in at least 90 % of the sequences, the exceptions

Table 5
Conserved positions in the substrate part of the binding pocket for each of the nine MDR families.

Res	M	V	E	Res	M	V	E	Res	M	V	E	Res	M	V	E	Res	M	V	E
MECR				Phe287	●	●	●	Pro69	●	●	●	Leu266	●	●	●	Pro150	●		
Asn88		●	●	Val289	○	●	●	Gly70	●	●	●	Arg267	●	●	○	Ile174	○	○	
Tyr94	●	●	●	Leu290		●	●	Thr98	●	●						Thr175	●	●	
Asn106	●	●						Ile99	●			**RT4I1**				Met178	●	●	
Glu107	●	●	●	**PTGR3**				Ser100	●	○		Ile86	●		○	Val179	○		
Ile129	●	○	○	Ser77	●	●	●	Gly101	●	●	●	Asn89	●			Asn320	●	●	●
Pro130		●		Asn80	●	●	●	Gly102	●	●		Tyr94	●	●	●	Arg321	○	○	○
Gly309	●	●	●	Arg85	●	●		Ile131	●	●	●	Ala98	●			Gly325	●	●	●
Phe310	●	●	●	Tyr86	●	●	●	Val245	●	●	●	Leu99	●	●		Leu328	●	●	●
Trp311	●	●	●	Phe98	●	●	●	Val246	●	●	●	Met101	●			Leu331	●		○
Ser313	●	○		Pro126	●			Gly247	●	●	●	Lys102	●						
Lys316	●	●		Ala140	●			Val269		●		Pro149	●	●		**VAT1L**			
Phe324	●			Val155	●	○		Leu271	●	●	●	Trp150	●	●		Leu80	○	○	○
				Phe271	●	●						Tyr183	●	●	●	Asn81	●	●	●
PTGR1				Tyr275	●	●	●	**QORX**				Leu300	●	●		Ile83	○	○	
Tyr49	●	●	●	Gly302	●	●	●	Asn40	●	●	●	Val301	●	●		Ile92	●	●	
Met124	●	●	●	Phe303	●	●	●	Arg41	●	●	●	Thr302	●	●		Pro96	●	●	●
Tyr245	●	●	●	Phe304	●	●		Ala42	●	●		Phe304	●	●		Phe104	●	●	
Phe270	●	●	●	Leu305	●	●	●	Leu44	○	○		Leu305	●	●		Asn131	●		
				Asn306	●			Tyr51	●	●	●	Met308	●			Tyr132	●		
PTGR2				Tyr312		●		Leu63	●	●	○	Ile313	○	○		Asn164	●	●	●
Tyr51	●	●						Glu123	●	●	●	Met317	●	●		Thr167	●		
Arg55	○	○	○	**QOR**				Thr127	●	●	●	Phe342	●	●		Tyr275	●	●	●
Tyr64	●	●	●	Val50	●	●	●	Leu237	●	○		Phe343	●	●		Gly276	●	●	●
Ile65	○	○	○	Glu51	○	○	○	Tyr238	●	●						Phe318	●	●	●
Phe99	●	●	●	Tyr53	●	●	●	Gly239	●	●		**VAT1**				Ser319	●	●	
Tyr100	●			Ile54	●	●	●	Leu240	●			Gln144	●			Leu321	●	●	●
Met135	●			Pro63	●	●	●	Thr263	●			Val147	●	○	○	Asn322	●	●	
Tyr259	●	●	●	Tyr67	●	●		Ser264	○	○		Thr148	●			Phe325	●		
Tyr265	●	●	●	Thr68	●	●		Leu265	○	●		Val149	●			Lys326	●		

M: mammals, V: vertebrates, E: eukaryotes.
● Residue conservation larger than 90 %.
○ Residue group (e.g. ILV) conservation larger than 90 %.

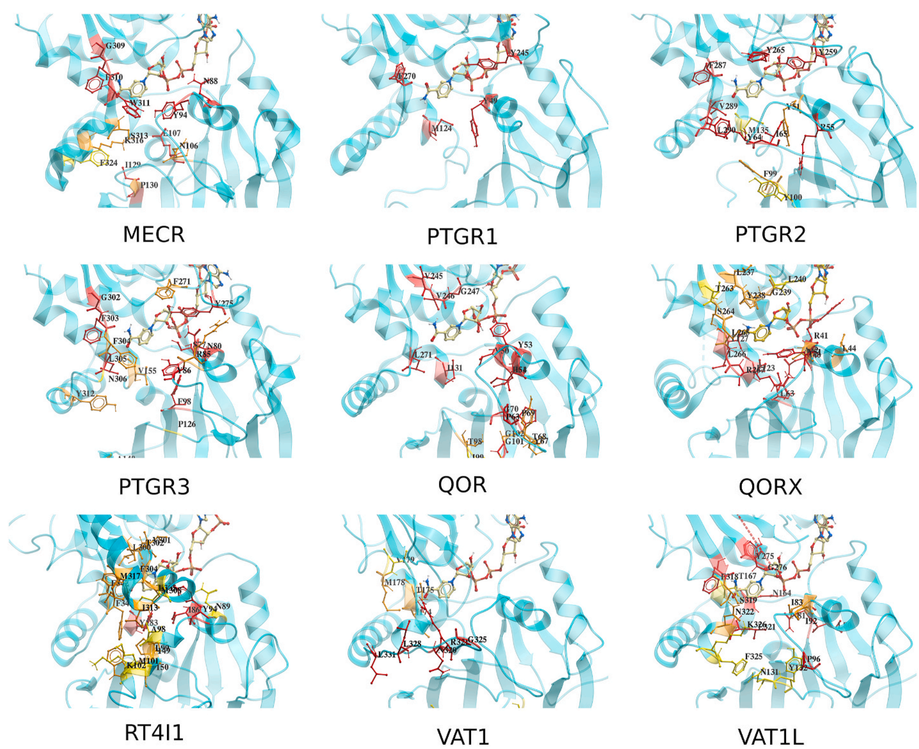

Fig. 3. Visualization of the conserved residues in the substrate part of the binding pockets of the nine MDRs discussed (as labelled). Conserved residues are visualized together with the coenzyme (NADPH). Red: conserved at the eukaryote level, orange: vertebrates, yellow: mammals.

being VAT1 and VAT1L. VAT1 had a Tyr in about 80 % of the complete sequences, while VAT1L instead had an Ile at the corresponding position in most sequences. This showed that Tyr49 (human PTGR1 numbering) was conserved in most of the families. This residue is located in the vicinity of the expected active site, though its exact location varied among the MDRs. In QORX, this residue seems to help with the coordination of the substrate, but mutation of the residue may even increase the catalytic efficiency [16], while in PTGR2 it is reported to be involved in the catalytic reaction [17], and in MECR mutation of the Tyr had severe effects on its function [23–25]. For the latter almost a complete loss of function. Thus, the importance of the residue varies greatly, even when its localization is nearly identical (as in the case of QORX and MECR).

In MECR, Tyr94, Glu107, Ile129, Gly165, Val166, Trp311, Lys316, and Phe324 have been suggested to be of importance for the function [23], with Tyr94 and Trp311 being critical to perform the function [23–25]. The analysis performed here confirms the importance of Tyr94, Glu107, and Trp311, which are all conserved even at the eukaryotic level. The group Ile/Leu/Val at position 129 is also conserved in eukaryotes, though Ile is only conserved at the mammalian level. The remaining suggested residues have varying levels of conservation, the most conserved being Lys316 which is conserved in vertebrates. Twelve conserved residues were identified in MECR at the substrate-interacting part of the binding pocket. Seven were conserved even at the eukaryotic level, four in vertebrates, and one only in mammals. Three of the residues conserved in eukaryotes, Asn88, Gly309, and Phe310 are located close to the coenzyme. Tyr94, Glu107, and Trp311 are conserved at the eukaryotic level, and located close to the active site. The remaining residues include Ile129 (Ile/Leu/Val in eukaryotes, hydrophobic environment), Lys316 (in vertebrates), and Phe324 (only in mammals).

Surprisingly, only four conserved residues were identified in the substrate-interacting part of the binding pocket of PTGR1. There are two conserved Tyr, Tyr49 and Tyr245, both of which may be involved in catalytic reactions. Tyr49 corresponds to the common Tyr observed in the other MDRs, while Tyr245 corresponds to a Tyr also seen in other PTGRs. These two Tyr residues have been shown by mutagenesis to be involved in the catalysis [26]. The remaining conserved residues, Met124 and Phe270, do not show any obvious way to perform substrate interaction, and may thus have structural importance instead.

The residues of interest that have earlier been identified in PTGR2 include Tyr51, Tyr64, Ile65, Phe99, Tyr100, Met135, Tyr259, Leu288, Val289, and Leu290 [15], where Tyr51 corresponds to Tyr49 in PTGR1. Out of these residues, only five are conserved in eukaryotes, including Tyr64 and Tyr259, believed to be directly interacting with the substrate [15]. An additional Tyr, Tyr100, is only conserved in mammals. There are also two more Tyr in the binding pocket: Tyr51 is conserved in vertebrates and Tyr265 in eukaryotes. The latter is positioned where it can interact with the coenzyme, but it is probably irrelevant for any substrate interactions. The most important residues seem to be Tyr64 and Tyr259, along with a hydrophobic region at position 287–290.

The structure of PTGR3 is like those of PTGR1 and PTGR2 with two Tyr, Tyr86 and Tyr275 (corresponding to PTGR1 Tyr49 and Tyr245) conserved at the eukaryotic level. Further, there are multiple other conserved residues in the vicinity of these Tyr. Ser77, Asn80, Arg85, Phe98, and Phe304 are all highly conserved, most even at the eukaryotic level. An additional Tyr, Tyr312, is identified at the entrance to the pocket, found in most vertebrate PTGR3s. These residues are all positioned where they may interact with the substrate and be involved in its coordination.

For QOR (Zeta-Crystallin) Val50, Tyr53, Tyr59, Ser71, Ile131, Thr135, Ser248, and Thr270 have all been suggested as important [27]. Out of these, Val50, Tyr53, and Ile131 are the only conserved residues. This means that Tyr59, corresponding to the Tyr conserved in most other MDRs, is not conserved. Instead, other residues in the same general area are conserved. Interestingly, mutation studies in combination with structural determinations have been shown that Tyr53 and Tyr59 are involved in different catalytic functions for the human QOR [27]. A negatively charged residue at position 51 and Ile54 are both conserved. There are also multiple conserved residues at an edge of the pocket, including e.g., Pro63, Tyr67, Pro69, and Gly70. The conserved prolines implying the importance of a certain fold in this region.

The residues previously confirmed to form the substrate part of the binding pocket in human QORX (p53-inducible quinone oxidoreductase PIG) are Asn40, Ala42, Tyr51, Leu63, Glu123, Thr127, Leu240, and Leu265 [16]. All are conserved though at varying levels, and Asn40, Tyr51, Leu63, Glu123, and Thr127 are conserved even at the eukaryotic level. Other highly conserved residues of interest include Arg41 and Arg267, located on opposite sides of Tyr51. Arg41 is directed away from Tyr51 in the 2J8Z structure, but it is located close enough to potentially interact.

The crystal structure of RT4I1 has, as noted above, an unusual means of dimer interaction which may cause artifacts in the results. There are only two residues conserved at the eukaryotic level, Tyr94 and Tyr183. Tyr94 corresponds to the Tyr common in many MDRs, while Tyr183 matches part of the definition of coenzyme-binding residues, though its positioning is unusual and makes it a good candidate for substrate interaction. Most of the remaining conserved residues are conserved at the mammal or vertebrate level, implying that the enzyme might be specific for different types of substrates depending on species. There are only three residues in the substrate part of the binding pocket of RT4I1 that are conserved in eukaryotes: Ile86, Tyr94 and Tyr183. All three are located close to the expected substrate binding site. Multiple residues are conserved at the vertebrate and/or mammalian levels. Positions 98–102 form an α-helix at the entrance of the pocket. Pro149 is probably coordinating the positioning of Trp150, located near the pocket entrance. Positions 300–302 are probably involved in coenzyme coordination.

The conserved Tyr common for the other MDRs was present in neither VAT1 nor VAT1L, not even at the mammalian level. It was present in about 60 % of the VAT1 sequences, and about 80 % of the VAT1L sequences covering the complete sequence. VAT1L has a highly conserved (98 %) hydrophobic residue (Ile/Leu/Val) at the corresponding position. In the human sequences, there was a Tyr in VAT1 and a Leu in VAT1L.

The residues conserved at the eukaryotic level in VAT1 included hydrophobic residues (Leu/Ile/Val) at position 147 and 331, Asn320, a charged positive residue at position 321 (Arg/Lys), Gly325, and Leu328. These residues are all located along the same side of the pocket [17].

VAT1L contains two conserved Tyr, Tyr132 (in mammals) and Tyr275 (in eukaryotes). Tyr275 is like Tyr265 in PTGR2 and Tyr238 in QORX, probable to interact with the coenzyme rather than a substrate. Tyr132 is located at the outer part of the binding pocket, potentially aiding with substrate coordination. There are seven residues in VAT1L that are conserved at the eukaryotic level. Most seem to be involved in coenzyme coordination (e.g., the Tyr275), or be of importance for the structure (e.g. Pro96, Gly276), though Asn81 may be involved in substrate interactions, a role that may be shared with Asn322 (conserved at the vertebrate level). There is also Ile92, located where the common Tyr is found in e.g., RT4I1. Many of the remaining residues seem to aid in the creation of a hydrophobic environment (e.g., Ile/Leu/Val80 and Ile/Leu/Val83). Finally, mammals have some unique conserved residues that may also be involved in substrate coordination, e.g., Lys326.

Based on the present study, some rough observations can be made about the presence of Tyr in the substrate part of the binding pocket. The prostaglandin reductases use a combination of Tyr49, located in an α-helix, and Tyr245 (PTGR1 numbering). The latter Tyr245 was strongly conserved at the eukaryotic level (>99 %), while Tyr49 had a lower level of conservation (95 %, 75 %, and 99 % for PTGR1, PTGR2, and PTGR3 respectively). Four of the remaining MDRs (MECR, QOR, QORX, RT4I1) have a single highly conserved Tyr (>95 % in eukaryotes), which is also present in VAT1 for some species. Finally, VAT1L avoids the use of Tyr at all. The exact involvement and coordination of the Tyr involved differs for every enzyme family. However, three overall distinct types of binding pockets could be identified, containing no, one, or two tyrosine residues (Fig. 3).

The use of Tyr in many MDRs is analogous to what is seen in many short chain dehydrogenases/reductases, where the combination of Ser139, Tyr152, and Lys156 forms a catalytic triad first seen in 3-alpha-(or 20-beta)-hydroxysteroid dehydrogenase (P19992) [28].

4. Conclusions

In this study, we have identified residues forming the binding pocket of nine different MDRs. Among the nine families being studied, most have a conserved Tyr close to the coenzyme which is probably involved in catalytic reactions. MECR, QOR, QORX, RT4I1, and many forms of VAT1 all seem to use one Tyr for their substrate binding, while PTGR1, PTGR2, and PTGR3 most likely use two Tyr. Finally, VAT1L seems to use another, unidentified, mechanism of substrate binding, and its substrate specificity is unknown, and it may have lost any catalytic activity. Meanwhile, the substrate-interacting part of the binding pocket was found to have a large variation, with 4–20 conserved residues present, depending on protein family. In opposite, the coenzyme part of the binding pocket was found to have similar conservation patterns among all the nine proteins.

Finally, it can be stated that most of the MDRs had a conserved Tyr present at the substrate part of their binding pockets, the PTGR enzymes even having two. The exceptions were VAT1 and VAT1L, which lacked any conserved Tyr in the region, though VAT1 did have a Tyr present in many species. VAT1L instead had a conserved Ile with a similar positioning as the Tyr.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and material

Code is available on Github: https://github.com/talavis/mdr_analysis.

Funding

The research presented here was supported by eSSENCE, Karolinska Institutet, and Uppsala University. They had no role in the design of the study, the collection, analysis, and interpretation of data, or in writing the manuscript.

CRediT authorship contribution statement

Linus J. Östberg: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Jan-Olov Höög:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Bengt Persson:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

The authors thank eSSENCE, Karolinska Institutet, and Uppsala University for funding this project.

References

[1] UniProt Consortium, UniProt: the universal protein knowledgebase in 2023, Nucleic Acids Res. 51 (2023) D523–D531, https://doi.org/10.1093/nar/gkac1052.

[2] J. Hedlund, H. Jörnvall, B. Persson, Subdivision of the MDR superfamily of medium-chain dehydrogenases/reductases through iterative hidden Markov model refinement, BMC Bioinf. 11 (2010) 534, https://doi.org/10.1186/1471-2105-11-534.

[3] M. Rossmann, D. Moras, K. Olsen, Chemical and biological evolution of a nucleotide-binding protein, Nature 250 (1974) 194–199, https://doi.org/10.1038/250194a0.

[4] B. Taneja, S.C. Mande, Conserved structural features and sequence patterns in the GroES fold family, Protein Eng. 12 (1999) 815–818, https://doi.org/10.1093/protein/12.10.815.

[5] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I. N. Shindyalov, P.E. Bourne, The protein data bank, Nucleic Acids Res. 28 (2000) 235–242, https://doi.org/10.1093/nar/28.1.235.

[6] A. Jayakumar, M.H. Tai, W.Y. Huang, W. al-Feel, M. Hsu, L. Abu-Elheiga, S. S. Chirala, S.J. Wakil, Human fatty acid synthase: properties and molecular cloning, Proc. Natl. Acad. Sci. U. S. A. 92 (1995) 8695–8699, https://doi.org/10.1073/pnas.92.19.8695.

[7] R.D. Finn, A. Bateman, J. Clements, P. Coggill, R.Y. Eberhardt, S.R. Eddy, A. Heger, K. Hetherington, L. Holm, J. Mistry, E.L. Sonnhammer, J. Tate, M. Punta, Pfam: the protein families database, Nucleic Acids Res. 42 (2014) D222–D230, https://doi.org/10.1093/nar/gkt1223.

[8] S.R. Eddy, Accelerated profile HMM searches, PLoS Comput. Biol. 7 (2011) e1002195, https://doi.org/10.1371/journal.pcbi.1002195.

[9] S.F. Altschul, T.L. Madden, A.A. Schäffer, J. Zhang, Z. Zhang, W. Miller, D. J. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, Nucleic Acids Res. 25 (1997) 3389–3402, https://doi.org/10.1093/nar/25.17.3389.

[10] C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, T. L. Madden, BLAST+: architecture and applications, BMC Bioinf. 10 (2009) 421, https://doi.org/10.1186/1471-2105-10-421.

[11] K. Katoh, D.M. Standley, MAFFT multiple sequence alignment software version 7: improvements in performance and usability, Mol. Biol. Evol. 30 (2013) 772–780, https://doi.org/10.1093/molbev/mst010.

[12] L.J. Östberg, B. Persson, J.-O. Höög, Computational studies of human class V alcohol dehydrogenase - the odd sibling, BMC Biochem. 17 (2016) 16, https://doi.org/10.1186/s12858-016-0072-y.

[13] S. Guindon, J.-F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, O. Gascuel, New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0, Syst. Biol. 59 (2010) 307–321, https://doi.org/10.1093/sysbio/syq010.

[14] R.C. Edgar, Search and clustering orders of magnitude faster than BLAST, Bioinformatics 26 (2010) 2460–2461, https://doi.org/10.1093/bioinformatics/btq461.

[15] Y.-H. Wu, T.-P. Ko, R.-T. Guo, S.-M. Hu, L.-M. Chuang, A.H.-J. Wang, Structural basis for catalytic and inhibitory mechanisms of human prostaglandin reductase PTGR2, Structure 16 (2008) 1714–1723, https://doi.org/10.1016/j.str.2008.09.007.

[16] S. Porté, E. Valencia, E.A. Yakovtseva, E. Borràs, N. Shafqat, J.É. Debreczeny, A.C. W. Pike, U. Oppermann, J. Farrés, I. Fita, X. Parés, Three-dimensional structure and enzymatic function of proapoptotic human p53-inducible quinone oxidoreductase PIG3, J. Biol. Chem. 284 (2009) 17194–17205, https://doi.org/10.1074/jbc.M109.001800.

[17] Y. Watanabe, Y. Tamura, C. Kakuta, S. Watanabe, T. Endo, Structural basis for interorganelle phospholipid transport mediated by VAT-1, J. Biol. Chem. 295 (2020) 3257–3268, https://doi.org/10.1074/jbc.

[18] J. An, M. Totrov, R. Abagyan, Pocketome via comprehensive identification and classification of ligand binding envelopes, Mol. Cell. Proteomics 4 (2005) 752–761, https://doi.org/10.1074/mcp.M400159-MCP200.

[19] R. Abagyan, I. Kufareva, The flexible pocketome engine for structural chemogenomics, Methods Mol. Biol. 575 (2009) 249–279, https://doi.org/10.1007/978-1-60761-274-2_11.

[20] R.K. Wierenga, P. Terpstra, W.G.J. Hol, Prediction of the occurrence of the ADP-binding βαβ-fold in proteins, using an amino acid sequence fingerprint, J. Mol. Biol. 187 (1986) 101–107, https://doi.org/10.1016/0022-2836(86)90409-2.

[21] T.J. Wheeler, J. Clements, R.D. Finn, Skylign: a tool for creating informative, interactive logos representing sequence alignments and profile hidden Markov models, BMC Bioinf. 15 (2014) 7, https://doi.org/10.1186/1471-2105-15-7.

[22] H. Jörnvall, J.O. Höög, B. Persson, X. Parés, Pharmacogenetics of the alcohol dehydrogenase system, Pharmacology 61 (2000) 184–191, https://doi.org/10.1159/000028399.

[23] Z.J. Chen, R. Pudas, S. Sharma, O.S. Smart, A.H. Juffer, J.K. Hiltunen, R. K. Wierenga, A.M. Haapalainen, Structural enzymological studies of 2-enoyl thioester reductase of the human mitochondrial FAS II pathway: new insights into its substrate recognition properties, J. Mol. Biol. 379 (2008) 830–844, https://doi.org/10.1016/j.jmb.2008.04.041.

[24] T.T. Airenne, J.M. Torkko, S. Van den plas, R.T. Sormunen, A.J. Kastaniotis, R. K. Wierenga, J.K. Hiltunen, Structure-function analysis of enoyl thioester reductase involved in mitochondrial maintenance, J. Mol. Biol. 327 (2003) 47–59, https://doi.org/10.1016/s0022-2836(03)00038-x.

[25] R.G. Rosenthal, B. Vögeli, N. Quade, G. Capitani, P. Kiefer, J.A. Vorholt, M. O. Ebert, T.J. Erb, The use of ene adducts to study and engineer enoyl-thioester reductases, Nat. Chem. Biol. 11 (2015) J398–J400, https://doi.org/10.1038/nchembio.1794.

[26] J. Mesa, C. Alsina, U. Oppermann, X. Parés, J. Farrés, S. Porté, Human prostaglandin reductase 1 (PGR1): substrate specificity, inhibitor analysis and site-directed mutagenesis, Chem. Biol. Interact. 234 (2015) 105–113, https://doi.org/10.1016/j.cbi.2015.01.021.

[27] S. Porté, A. Moeini, I. Reche, N. Shafqat, U. Oppermann, J. Farrés, X. Parés, Kinetic and structural evidence of the alkenal/one reductase specificity of human ζ-crystallin, Cell. Mol. Life Sci. 68 (2011) 1065–1077, https://doi.org/10.1007/s00018-010-0508-2.

[28] D. Ghosh, Z. Wawrzak, C.M. Weeks, W.L. Duax, M. Erman, The refined three-dimensional structure of 3 alpha,20 beta-hydroxysteroid dehydrogenase and possible roles of the residues conserved in short-chain dehydrogenases, Structure 2 (1994) 629–640, https://doi.org/10.1016/s0969-2126(00)00064-2.