



In silico characterization of human cytochrome P450 monooxygenases

K. Syed Ibrahim, Anuradha Laishram, Esther Lalnunmawii, N. Senthil Kumar*

Department of Biotechnology, Mizoram University, Aizawl 796004, India

Received 20 July 2015 | Revised 26 October 2015 | Accepted 30 October 2015

ABSTRACT

Cytochrome P450 monooxygenases (CYPs) represent a large group and diverse family of enzymes involved in the myriad of biological processes in humans. In the present study, a total of 57 protein sequences of human CYPs retrieved from UniprotKB have been characterized for various physicochemical properties, homology search, motif and super family search and phylogenetic relationship. Physicochemical analysis showed that the isoelectric point values and GRAVY index ranged from 5.84 to 9.47 and 0.018 to -0.367, respectively. Many proteins (50 members, 87.8%) were in basic form, while few (7 members, 12.2%) were of acidic nature. Moreover, GRAVY index revealed that only CYP26C1 as hydrophobic while all others as hydrophilic. Phylogenetic analysis revealed that P450 proteins basically fall into two main clades and are divided into five subgroups. Motif analysis with MEME indicated presence, absence and even shuffling of motifs within clades. Clustering using the maximum likelihood analysis was also in accordance with P450s central roles in drug and xenobiotic metabolism as well as steroid hormone synthesis, fat-soluble vitamin metabolism, and the conversion of polyunsaturated fatty acids to biologically active molecules. Motif conservation within clusters showed the evolutionary pressure for maintenance of the structural and functional organization between different groups of protein. These results will help in the context of understanding the characteristics of the cytochrome P450 monooxygenase isoforms.

Key words: Cytochrome P450; *in silico*; maximum likelihood; motif analysis; phylogenetic analysis.

INTRODUCTION

Cytochrome P450 (CYP450) is a large family of heme-thiolate proteins that plays key role in xenobiotic metabolism, metabolizing most of the drugs and chemicals of toxicological im-

portance. Besides that they also play major roles in diverse physiological processes including steroid and cholesterol biosynthesis, fatty acid metabolism (prostacyclin, thromboxane) and the maintenance of calcium homeostasis.¹ These highly conserved genes are found distributed in different life forms, including prokaryotes (archaea, bacteria), unicellular eukaryotes (protists, fungi) and multicellular eukaryotes (plants and animals).² The human genome en-

Corresponding author: Senthil Kumar
 Phone: +91-9436352574
 E-mail: nskmzu@gmail.com

codes 57 P450 genes³ grouped into 18 mammalian families and 44 subfamilies. They are basically membrane-associated proteins⁴ located on the smooth endoplasmic reticulum of cells ubiquitously in the body, but predominantly expressed in the liver.⁵ Any imbalance in enzyme availability or its malfunction, e.g., due to a genetic mutation, may lead to a disease state in humans.^{6,7} In humans, CYP450 are known for their central role in phase I drug metabolism where they are of critical importance to two of the most significant problems in clinical pharmacology: drug interactions and inter-individual variability in drug metabolism.⁸ Besides detoxification, they also synthesize biologically active compounds such as steroids, prostaglandins, and arachidonate metabolites. CYP proteins have been identified in life forms like animals, plants, fungi, protists, bacteria, archaea, and even in viruses,⁹ but not in *Escherichia coli*.⁸

P450s are usually divided into gene families. Besides, they are also divided into four different classes depending on the redox partner required. They have a varied primary, secondary, and tertiary structures. Each P450, has a preferred set of substrates, must have a unique method of substrate recognition and an active site they all appear to have a similar structural core. Thus, the charge distribution and topography of the 'redox-partner binding region' of each class of P450 vary to accommodate the various types of redox partners there by providing a great variety of phylogenetically distributed isoform activities. Though much studied, still more remains poorly understood.

In the present study, we performed an *in silico* analysis on human P450 protein sequences by analyzing their biochemical features, homology, motif patterns, cluster and superfamily distribution to understand their functional evolution.

METHOD

The amino acid sequences of the human P450 in FASTA format were downloaded from the UniProt Knowledgebase (UniProtKB) (Table 1). Physicochemical analysis were per-

formed by methods explained by Brindha *et al.*¹⁰ Data on physicochemical properties were generated using tools like ProtParam, Protein calculator, Compute pI/Mw, ProtScale from Expert Protein Analysis System (EXPASY) proteomic server from the protein sequences.¹¹ The molecular weights (kilo dalton) were calculated by the addition of average isotopic masses of amino acid in the protein and deducting the average isotopic mass of one water molecule. The pI was calculated using pK values of amino acid according to Bjellqvist *et al.*¹² The atomic composition, extinction coefficients and aliphatic index was derived using the ProtParam tool, available at ExPASy. The Instability Index which predicts regional instability by calculating the weighted sum of dipeptides that occur more frequently in unstable proteins when compared to stable proteins was calculated using the approach of Guruprasad *et al.*¹³ The Grand average hydropathy (GRAVY) was calculated by adding the hydropathy value for each residue and dividing by the length of the sequence,¹⁴ respectively. Aliphatic index was calculated using the formula $x^* (\text{ALA}) + a^*x (\text{VAL}) + b^*x (\text{LEU}) + b^*x (\text{ILE})$ where $a = 2.9$ and $b = 3.9$ are constants.¹⁵ Mega 6.01¹⁶ was used for phylogenetic analysis by maximum likelihood method (ML). An optimal model of evolution for the aligned dataset was determined. Sequence divergences were then calculated using the best model which was identified by Mega 6.01. This model choice was guided by the need to avoid under parameterisation.¹⁷ Domain analysis was performed using Pfam 27.0.¹⁸ Motif analysis was done using MEME¹⁹ using the expectation maximization approach.

RESULTS AND DISCUSSION

Physicochemical characterization of P450 proteins

Humans have 57 CYP450 genes,³ which are subdivided to 18 families and 44 subfamilies. There are more than 9000 known CYP450 sequences.²⁰ Owing to its importance, many ani-

Table 1. Physicochemical parameters of the P450 proteins.

Group	Clade	Gene/ amino acids	Accession no.	Mw	pI	Extinction coefficient	Instability index	Aliphatic index	GRAVY	Domain predicted by pfam	Function	
Major group 1	Sub clade 1	CYP3A4/503	P08684	57343.1	8.27	46215	41.19	95.47	-0.04	442(C).		
		CYP3A5/502	P20815	57108.6	8.86	53080	32.31	96.08	-0.116	441(C).	Drug/xenobiotic metabolism	
		CYP3A7/503	P24462	57525.5	9.21	53205	29.92	97.59	-0.07	442(K).		
		CYP3A43/503	Q9HB55	57669.7	8.27	50225	33.92	98.03	-0.036	442(C).		
		CYP4A11/519	Q02928	59347.8	8.96	85870	43.76	98.46	-0.186	457(C), 321(E).		
		CYP4A22/519	Q5TCH4	59245.8	9.21	74870	42.4	96.2	-0.208	457(C), 321(E).		
		CYP4B1/511	P13584	58991.1	8.47	86330	44.19	85.11	-0.204	453(C), 315(E).		
		CYP4F2/520	P78329	59853.4	6.6	99515	57.86	94.71	-0.118	468(C), 328(E).		
		CYP4F3/520	Q08477	59846.7	7.59	102745	54.23	95.29	-0.074	468(C), 328(E).		
		CYP4F8/520	P98187	59994.6	8.73	88640	52.23	94.88	-0.16	468(C).	Arachadonic acid and fatty acid metabolism	
		CYP4F11/524	Q9HB16	60145.7	6.26	95630	55.43	94.18	-0.134	468(C), 328(E).		
		CYP4F12/524	Q9HCS2	60269.8	7.02	92650	52.58	94.39	-0.126	468(C).		
		CYP4F22/531	Q6NT55	61958.2	8.95	83935	52.52	96.97	-0.201	475(C), 335(E).		
		CYP4V2/525	Q6ZWL3	60724	7.19	102830	49.69	90.11	-0.284	467(C), 329(E).		
		CYP4X1/509	Q8N118	58875.2	8.74	84060	43.32	88.55	-0.21	454(C).		
		CYP4Z1/505	Q86W10	59085.9	9.29	85995	52.4	90.16	-0.196	452(C).		
		CYP5A1/533	P24557	60518.3	7.56	47620	44.89	90.75	-0.059	479(C).	Thromboxane A2 synthesis	
		Sub clade 2	CYP11A1/521	P05108	60102	8.89	87445	34.54	89.06	-0.245	462(C).	
			CYP11B1/503	P15538	57572.9	9.4	69245	47.96	96.96	-0.1	450(C).	Key steps in steroid biosynthesis
CYP11B2/521	P19099		57560.1	9.47	67755	42.31	99.11	-0.073	450(C).			
CYP24A1/514	Q07973		58875.4	8.94	80955	49.31	88.04	-0.367	428(C).	Vitamin D metabolism/inactivation		
CYP27A1/531	Q02318		60234.7	9.05	78185	49.52	90.23	-0.185	476(C).			
CYP27B1/508	O15528		56504	9.34	66390	56.21	89.31	-0.179	455(C).	Bile acid biosynthesis, vitamin D activation		

Sub clade	CYP26A1/497	O43174	56198.6	8.96	51755	50.27	95.15	-0.081	442(C).	
3	CYP26B1/512	Q9NR63	57512.6	8.68	49765	35.09	98.44	-0.079	441(C).	Retinoic acid metabolism/inactivation
	CYP26C1/522	Q6V0L0	57111	9.24	63535	47.86	97.43	0.018	459(C).	
	CYP46A1/500	Q9Y6A2	56821	9.15	51255	52.18	87.92	-0.212	437(C).	Cholesterol metabolism
	CYP51A1/503	Q16850	56803.7	8.72	64665	42.99	92.5	-0.151	449(C).	Cholesterol biosynthesis
Sub clade	CYP1A1/512	P04798	58167.3	8.61	63870	40.85	92.29	-0.084	224(F).	
4	CYP1A2/515	P05177	58294.3	9.18	64775	41.43	89.73	-0.161	226(F).	
	CYP1B1/543	Q16678	60845	9.18	52410	37.69	87.83	-0.134	470(C).	
	CYP2A6/494	P11509	56510.3	9.25	26025	40.66	80.85	-0.227	439(C), 107(F), 297(N).	
	CYP2A7/494	P20853	56425	7.69	24660	45.67	81.66	-0.18	439(R).	
	CYP2A13/494	Q16696	56687.5	9.31	27390	38.14	77.73	-0.275	297(N).	
	CYP2B6/491	P20813	56278.3	8.43	29590	37.42	93.14	-0.099	436(L).	
	CYP2C8/490	P10632	55824.7	8.8	33765	41.5	88.45	-0.167	435(C), 100(S), 204(N), 241(R).	
	CYP2C9/490	P11712	55627.8	8.13	35130	45.12	94.65	-0.073	435(R)	Drug/xenobiotic metabolism
	CYP2C18/490	P33260	55710.6	6.83	38235	45.76	89.49	-0.116	435(C).	
	CYP2C19/490	P33261	55931	7.11	33640	43.68	94.84	-0.098	435(K).	
	CYP2D6/497	P10635	55769.4	6.77	47815	44.63	95.15	-0.031	443(C), 301(D).	
	CYP2E1/493	P05181	56848.9	8.28	54820	33.65	91.14	-0.199	437(R).	
	CYP2F1/491	P24903	55501.2	6.93	31775	46.79	97.52	-0.047	436(C).	
	CYP2J2/502	P51589	57610.6	8.76	64650	33.98	86.1	-0.225	448(C).	
	CYP2R1/501	Q6VVX0	57359.3	7.25	67185	45.02	88.94	-0.094	448(G), 250(A).	
	CYP2S1/504	Q96SQ9	55816.6	8.8	42650	32.53	93.67	-0.024	440(G).	
	CYP2U1/544	Q7Z449	61987.1	8.63	76360	55.49	94.96	-0.078	490(C).	
	CYP2W1/490	Q8TAV3	53843.8	9.03	49430	45.43	104.12	-0.099	433(C).	
	CYP17A1/508	P05093	57370.5	8.72	63160	33.18	99.8	-0.154	442(C).	Testosterone and oestrogen biosynthesis
	CYP19A1/503	P11511	57882.9	7.2	58830	38.23	98.81	-0.003	437(C), 309(D), 374(M).	Oestrogen biosynthesis (aromatase)
	CYP21A2/494	P08686	55887.3	7.71	77265	49.27	106.19	-0.057	428(C).	Steroid biosynthesis

Sub clade	CYP7A1/504	P22680	57660.7	8.48	60445	44.61	88.71	-0.21	444(C).	Rate-limiting step of bile acid biosynthesis (cholesterol elimination)
5	CYP7B1/506	O75881	58255.9	8.2	56310	42.78	100	-0.041	449(C).	Prostacyclin and bile acid biosynthesis
Major group 2	CYP8A1/500	Q16647	57104	6.8	80455	45.69	99.9	-0.13	440(C).	Cholesterol metabolism
	CYP8B1/501	Q9UNU6	58068.4	8.78	90090	33.45	85.39	-0.165	440(C).	
	CYP39A1/469	Q9NYL5	54115.7	8.85	77975	38.29	102.45	-0.022	414(C).	

Table 2. Motifs observed in the human P450.

Clade	Motif	Width	Regular expression
Sub clade 1	Motif 1	50	[GH]KALSD[EL]DE[IL]RA[EQ]AVS[D]IIT[IF]MIF[EA]G[H]Y[DE]TT[AS]SG[L]S[W]F[IV]LYELA[TK]HP[ED][YHV]Q[EQ][RK][CL][RQ][EQ]E[IV][DQ]
	Motif 2	41	[HSD]P[YL]A[FY]I[IL]PF[SG]AGPRNC[IG]M[RA]M[L]A[EN]I[M]L[K]V[L]A[V]L[A][LR]I[TV]L[LQ][RH]N[F]RS[IF]L[L]K[IP][DC]
	Motif 3	50	[PG]W[F]L[GK]D[G]A[LI][LS][SLA]E[GD][DE]KE[W]K[S]RQ[HRR][L]M[LT]S[P]A[T]F[H]T[IF]S[IND]I[K]K[PE]I[VM]I[M]I[KP]I[FMI][AN]I[EQ]S[VAG][NDR][IV][ML][LHV][DAR]K[W]L[EQR]
	Motif 4	41	[PI][DN]GR[V]FS[IL]P[KA]G[V]V[C]L[IS]PIYAG[L]H[H]ND[PKA]I[V]YWP[DEN]PEV[FY]DPI[E]F[RF]SD][PK][EK]N
	Motif 5	41	[DEN]I[PG]K[A]K[TE]P[S]I[TE][W]Y[DD]LAQ[L]M[PE]I[VF]I[LT]I[TD]M[CV]I[IV]K[N]E[ST]LRL[Y]F[H]P[IV]A[PR]IV[S]R[X]C[L]T[S]K[K]Q[J]D[P]I[V]I
Sub clade 2	Motif 1	41	[IKY][AG]N[IV]S[TM]E[LM]I[LT]A[L][AG][GS]VDDT[SA][FN][TP]L[LS][WMM][TA]L[YF]EL[AS]R[N]H[PEV]CQAL[RH]QE[ISV]
	Motif 2	39	[HR][NP]F[AG][HS][LV]PFGF[IV]K[M]R[Q]S[C]L[G]RR[L]A[E]L[A]E[IM]L][HLQ][L]L[A]L[AH][HQ][ILV]L[KQ][HK][FY]E[V]I[EQ]
	Motif 3	50	PLL[KR]A[AV]L[KET]RLYP[V]G[P]L[FG]T[ILN][ES]RVLDS][K]S[D][LV]L[V][GQ][DGN]Y[HL]P[KA][GN][TLQ][V]LQ[L]V][CF][HL]Y[ASV][L]T][G]SJR
	Motif 4	50	[NR]F[IL]HDM[A]I[LE]KY[IV]QTM[F]M][K]H[S]I[ST]TFS[IV]G[Q]PR[IL]M][L]M[L]F[NV][ML]T[IP]RPV][SDE]L[SFH][RK][W]L[S][FIL]T][SNR][PT]K[V]T
	Motif 5	50	W[K]Q][DE]H[F]V][EAL]AWD[CTV]I[F]QKS][YKS][GAV][DK][NA]I][CY]I[IT][MSV][CY][V]L][AM][LS]PE[DL]M][E]KLQ[RQ][QV]E[D][GS][L]P][HY]PERMS[IL]EPW[V]K][AE][Y]HJR[QD][HY]R[GQ][HR]K[CY]G[V]L][FL][L]T][L]ENJG[E]P[E]W
Sub clade 3	Motif 1	48	[YL][LV]IDG][CRI]V[C]I[V]K[E]T[V]I[M]R[L]F[L]N[R]I[PT]I[V]I[S]M[P]I[G]I[GM]Y[F]M[R][T]M[V][AV][ILR][I]K[R]T][TP][FQ][ET][LV]DAN][G]YF][Q]T][P][K]P][GLW]H][SNQ][V]M[C]I][Y]V[S]I[P]I[R]C]T][D]V][T]N][I]H]Q
	Motif 2	50	[ST]M[G]F[W]P[F]L][F]I][G]E]T[L]G][H]Q][W]M][LV][LV]Q][GR][SR][GKR]F[HLQ][S]Q][SM][R]K[R]E[R][KRY]G[F]N[T]I][V]I][FY]K[TH]L][L]F][G]R[P][L]T][V]I][V]R]V]I[M]S]T][G]A][E]D]N]V
	Motif 3	39	[RK]I[F]H[AST]I][Y]F][I]F]L]V]P]F][GS][GAL]G[A]H[L]R][R]H][S]R]T]C][IL]V]G]K]Q]E][E]H]N]Q][F]L]A][K]Q][Y][A]I][L]M]V][E]E]L]Q]V][L]I]V]I][K]Q][V]I]L]T][I]L]F]V]I][A]M]T]W]I]V]A]S][E]K]T][L]M][A]L]V][I]R]Q]S][T]H]L]R][A]C]L]S]Y][E]R]D][I]F]W][E]D]Q]R]L
	Motif 4	50	GEH[IR]H[L]V][SR]I[S]T]V][E]H]Q]W]P][AQR]S]A]T]V][I]R]H][I]M]T][L]I]L]G]S]P][G]H]N]I]T]C][I]L]V][S]L][N]G][A]L]S][H]I]V][G]D][D]E]S][I]P]S]H][R]K][N]Q]R][I]R]K][R]K][V][F]L][A]M]S][I]R]K][A]I]V]F]S][I]R]H][E]A][L]E][C]R]S]Y
	Motif 5	45	DAL[IQ]L[L]I][E]H][S]I][S]A][KRW]E[HLR]G[E]H[K]E[R][L]M]P][D]S]T]M]Q]E]A]L]K]D]E]Q][S]G][A]S]T][L]T]V]E]L]L]I]F]F]A]G][A]G][I]F]H]Y][A]E]F]T]T]A]S]A][S]A]T]S]L][I]V]

Sub clade 4	Motif 1	50	R[YW][GA][L][F][L][Y][L][M][L][M][K][L][H][Y][P][E][D][I][V][I][Q][E][T][A][E][K][R][V][Q][H][E][E][L][D][E][R][V][I][V][G][R][P][N][R][Q][S][P][C][S][L][M][I][E][Q][D][R][A][S][I][H][K][I][M][L][P][Y][T][M][I][E][D][J][A][V][I][V][L][H][E][I][V][I][Q]
	Motif 2	32	F[M][L][P][F][S][G][A][I][G][K][R][V][C][L][G][E][G][L][A][R][M][E][L][F][F][L][T][L][L][I][L][M][Q][N][R][F]
	Motif 3	50	G[L][V][P][H][R][I][A][R][I][V][T][R][D][T][K][S][I][F][L][R][G][Y][F][L][I][P][K][G][T][E][T][I][V][I][F][I][P][T][N][L][T][W][S][V][L][H][I][H][R][I][D][I][P][E][K][E][F][W][I][P][E][I][N][P][E][E][F][I][N][D][I][P][E][H][R][F][L][D]
	Motif 4	41	[S][G][I][L][E][D][I][R][I][V][Q][E][E][A][R][C][F][L][V][I][E][D][I][A][E][L][R][K][T][K][G][A][I][A][S][I][P][F][C][I][D][P][T][F][L][I][L][I][G][S][I][C][R][A][I][V][P][S][C][I][N][V][I][C][S][I][S]
	Motif 5	29	G[P][R][P][R][W][V][L][H][G][Y][E][D][A][V][I][K][R][E][A][L][I][V][I][D][Q][A][I][E][D][I][F][S][A][I][G][D][I][R][G][P]
Sub clade 5	Motif 1	50	[P][L][T][G][R][W][I][E][R][P][P][I][C][I][D][I][E][K][L][N][I][G][W][L][S][T][I][L][V][P][I][W][Y][L][I][G][I][V][C][H][Y][I][A][G][V][I][L][F][M][I][A][D][E][N][Q][I][F][L][I][G][R][I][K][A][I][D][N][A][I][P][A][M][I][L][A][F][I][E][R][S][F][L][I][M][I][K][E][R][T][I][R][A][K][T][I][M][A][L][N][I][Q][R][K][I][E][I][K][R][T][I][K][Q][I][H][Y][G][D][H][P][I][V][T][T][V][C][I][L][F][K][Q][I][L][A][V][I][G][M][I][K][N][Q][R][I][Y][R][I][V][F][I][M][I][T][H][I][F][V]
	Motif 2	41	[D][N][G][I][K][S][I][E][K][R][T][I][K][T][K][I][D][T][F][F][Y][I][K][C][I][D][N][R][T][G][I][K][I][K][R][I][L][I][K][H][I][C][H][N][Y][Y][L][N][T][Y][M][P][I][F][W][G][I][S][A][T][G][I][A][H][T][V][I][S][N][T][I][I][H][K][I][C][I][P][L][G][R][F][L][S][I][F][Y][A][I][L][V][I][H][M][N][S][I][E][S][I][V][I][K][I][Q][L]
	Motif 3	42	L[W][A][S][T][I][Q][L][V][I][A][G][S][I][N][T][A][M][I][G][I][V][P][A][T][V][I][A][M][S][T][F][W][I][A][L][S][T][I][L][M][I][L][A][F][Y][I][Y][F][Q][I][L][M][V][I][L][I][K][R][S][I][H][N][P][I][E][D][I][A][I][M][H][I][L][I][A][K][R][I][A][V][A][I][R][M][T][I][E][D][G][I][E][G][I][I][A][L][V][I][D][E][K][S][T][I][R][S][Q][I][V][I][L][T][I][L][F][I][G][E][Q][W][I][E][K][N][Q][S][I][A][T][I][G][E][R][I][Q][K][L]
	Motif 4	15	[D][N][I][P][E][D][K][I][Y][I][F][Y][H][I][P][E][T][I][D][E][A][I][P][E][L][T][I][V][E][L][T][F][I][K][R][I][Y][P][I][D][E][N][I][R]
	Motif 5	41	L[I][D][N][Q][I][S][D][H][I][L][T][I][P][I][I][V][C][I][L][D][E][I][S][I][M][S][I][L][V][I][E][F][K][S][I][E][S][A][T][I][R][L][R][I][S][R][T][I][A][S][I][A][Y][I][P][S][I][T][F][L][I][L][N][T][I][L][T][R][I][E][F][L][I][V][A][I][E][H][K][V][I][E][V][I][D][I][L][F][Y][I][T][A][I][L][M][I][H][K][P][S][I][M][L][S][I][E][A][S][I][D][S][T][I][G][I][D][Q][R][S][I][E][Y][I][C][F][N][Y]

mals possess more CYP genes than humans like mice have genes for 101 CYPs, and sea urchins have even more (perhaps as many as 120 genes).²¹ Apart from humans, CYPs from insects, including CYP6G1 from *Drosophila melanogaster*²² and CYP6Z1 from *Anopheles gambiae*²³ have also been heavily studied so as to understand pesticide resistance. The amino acid composition and physicochemical features of protein is essential for understanding the evolutionarily conserved structure and function.²⁴ In the present study these 57 p450 protein sequences from human were analyzed by using bioinformatics tools. Physicochemical analysis showed that *pI* values 5.84 to 9.47. Isoelectric point (*pI*) is a pH value where net charge of protein is zero and shows whether protein character is acidic or basic. It was observed that most of proteins (50 members, 87.8%) were considered as basic character ($pI \geq 7$), while 7 protein sequences (12.2%) were considered as acidic ($pI \leq 7$). The most basic protein was found as CYP11B2 while the most acidic protein was found as CYP20A1. The GRAVY index ranged from 0.018 to -0.367. Positive GRAVY value corresponds to hydrophobic protein structure while negative GRAVY value means hydrophilic protein structure. GRAVY index revealed that except CYP26C1, all others as hydrophilic in nature. P450s are designed to break down environmentally toxic compounds such as polycyclic aromatic hydrocarbons and fluorocarbons, but the substrates of these proteins are predominately hydrophobic compounds and made more water soluble by monooxygenation.²⁵ The relative volume of valine and leucine/isoleucine side chains in comparison to the side chains of alanine is demonstrated by Aliphatic index.¹⁵ The dipeptides and tripeptides composition and their frequency of occurrence have been associated with solubility and folding of over-expressed proteins.²⁶ In general, proteins are mostly in the range of 42 to 62 kDa. The smallest protein was CYP27C1 containing 372 amino acids, had a molecular weight of 42.6 kDa. Proteins CYP4F22 with 531 amino acids and CYP2U1 with 544 amino acids had approximately were found to be the heaviest,

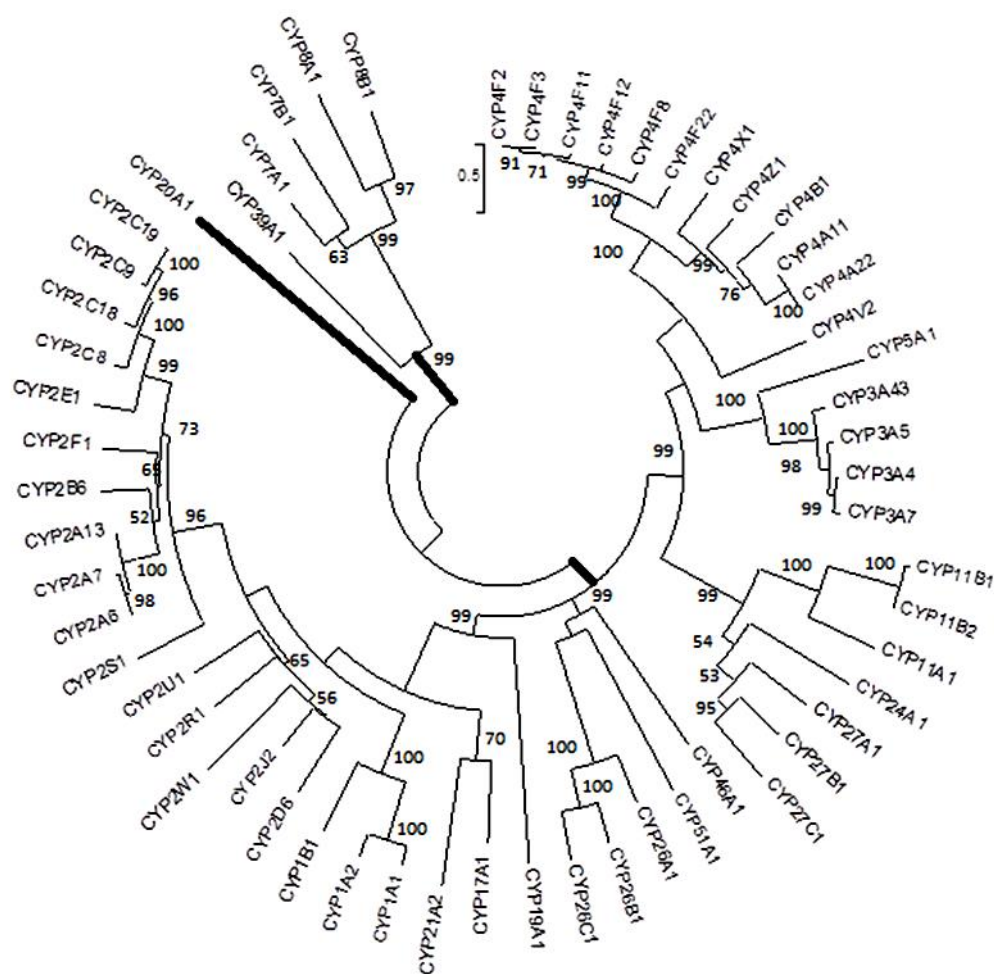


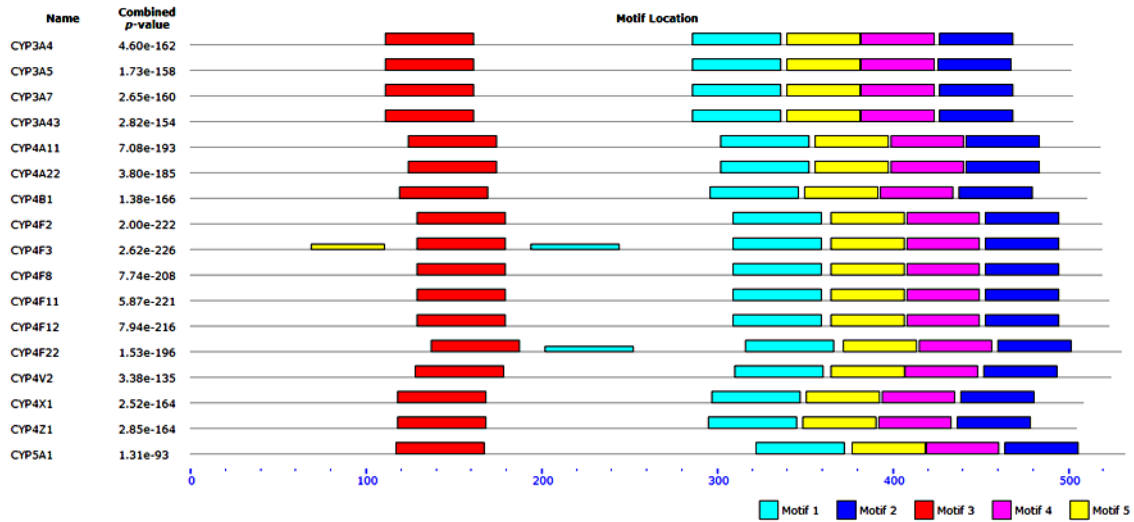
Figure 1. Radial representation of the Maximum Likelihood tree of the human P450 proteins.

sharing a molecular weight of 61.9 kDa. Total amino acid composition analysis revealed that leucine is the major composition in all P450 proteins, and up to five motifs were determined by using MEME tool from proteins of every clade. The intron-exon organization of P450 genes exhibits a diversity of gene structure clearly indicating that multiple gains and losses of introns have occurred during the evolution of P450 genes in diverse species, with little conservation of intron positions among divergent P450 families.^{27,28}

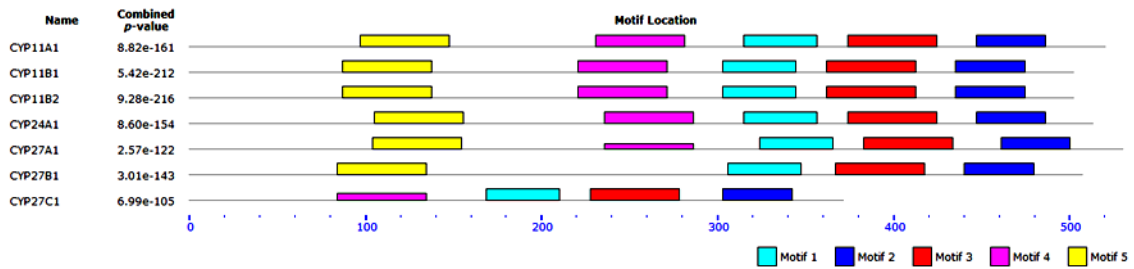
Phylogenetic relationship of P450 proteins

The phylogenetic tree was constructed using 57 P450 protein sequences maximum likelihood method (ML). LG+G+I was used as the model. Two major groups were observed on the phylogenetic tree, the smaller one had CYP7, CYP8 and CYP39 group of proteins while the larger clade, contained the rest of the proteins. Apart from the major clades, CYP20A1 was found to be alone. Clustering of the proteins was mainly on the identity. If an isoform has an identity of

a)



b)



c)



Figure 2. Combined block diagrams of the conserved protein motifs in P450 proteins using MEME server. (a) Sub clade 1 consisting families 3, 4 and 5. (b) Sub clade 2 consisting families 11, 24 and 27. (c) Sub clade 3 consisting families 26, 46 and 51.

d)



e)

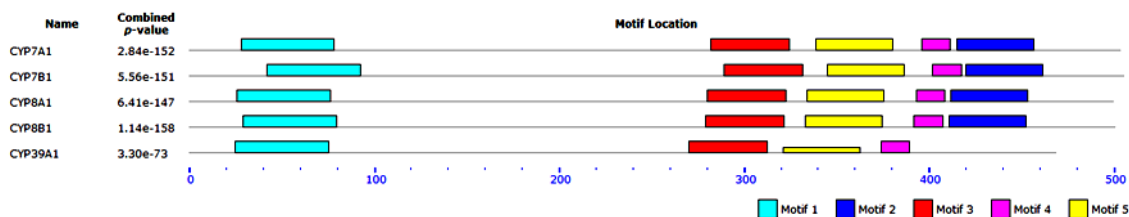


Figure 2. (d) Sub-clade 4 consisting families 1, 2, 17, 19 and 21. (e) 7, 8 and 39.

more than 40%, it belongs to the same CYP gene family and if approximately 55-60%, then they belong to the same subfamily such as CYP1A, CYP2B, etc.^{20,29}

The bigger group had four sub-clades (Fig. 1). Each cluster of proteins was subjected to motif analysis. The motif and their regular expression are represented in Table 1. Variations among various P450enzymes in terms of physiochemical parameters are given in Table 2.

Major clade 1 comprises the first biggest group of proteins containing about 24 protein sequences with two sub groups. One represents families 3, 4 and 5 with functions related to drug/xenobiotic and fatty acid metabolism, while the other with families 11, 24 and 27 with

functions related to bile acid biosynthesis and vitamin D activation. The pI of the proteins ranges from 6.26 to 9.47. Motif analysis showed absence of motif 3 in CYP11B2 and CYP27C1 (Fig. 2a).

Sub-clade 1 consists of families 3, 4 and 5 with functions related to drug/xenobiotic and fatty acid metabolism. The pI of the proteins ranges from 6.26 to 9.29.

Sub-clade 2 comprised of families 11, 24 and 27 with functions related to bile acid biosynthesis and vitamin D activation. The pI of the proteins ranges from 8.89 to 9.47. Motif analysis showed absence of motif 4 in CYP27B and motif 5 in CYP27C1 (Fig. 2b). All the proteins in this clade are of mitochondrial.

Sub clade 3 represent 5 protein sequences belonging to families 26, 46 and 51. Cholesterol metabolism and retinoic acid metabolism were observed as the main functions of this group. The pI ranged from 8.68 to 9.24. The only protein (CYP26C1) that is hydrophobic in nature falls in this cluster. Domain analysis showed that as cysteine as the key residue of all the proteins in this segment. Motif analysis showed absence of motif 2 in CYP46A1 and motif 4 in CYP51A1 (Fig. 2c).

Sub clade 4, represent 22 protein sequences belonging to families 1, 2, 17, 19 and 21. These had functions related to drug/xenobiotic and steroid metabolism. The pI ranged from 6.77 to 9.31. Motif analysis showed absence of motif 4 in CYP1B1 and motif 5 in CYP19A1 (Fig. 2d). CYP1A5 and CYP3A37 in turkeys were found to be very similar to the human CYP1A2 and CYP3A4 respectively, in terms of their kinetic properties as well as in the metabolism of aflatoxin B1.³⁰

Clade 5 represent 5 protein sequences belonging to families 7, 8 and 39. These had functions related to cholesterol metabolism, prostacyclin and bile acid biosynthesis. The pI ranged from 6.8 to 8.85. Domain analysis showed that cysteine as the key residue. Motif analysis showed absence of motif 2 in CYP39A1 (Fig. 2e).

The protein CYP20A1 function of this protein is still not attributed. This was separated from the other two major clades. A similar protein was found in *Rattus norvegicus* (Q6P7D4) with slightly different physiochemical properties.

Computational analyses indicate that the human P450 proteins have wide sequence diversity, physicochemical properties and distinct phylogenetic tree topology. Cluster analysis of all retrieved sequences indicates the evolutionary history of P450 proteins. Conserved motifs within clusters showed the evolutionary pressure to maintain important residues for the structural and functional organization of different groups of protein. The data represented here can also be important for better understanding of physiological properties and roles of P450 family pro-

teins in humans.

ACKNOWLEDGEMENTS

The authors wish to acknowledge the Department of Biotechnology (DBT), Government of India, New Delhi, for the financial support in the form of Bioinformatics Infrastructure Facility (BT/BI/12/060/2012 (NERBIF-MUA)). KSI and AL thank the State Biotech Hub (BT/04/NE/2009) sponsored by DBT, New Delhi for their fellowship.

REFERENCES

1. McKinnon RA, Sorich MJ & Ward MB (2008). Cytochrome P450 Part 1: multiplicity and function. *J Pharm Pract Res*, **38**, 55–7.
2. Omura T (1999). Forty years of cytochrome P450. *Biochem Biophys Res Commun*, **266**, 690–698.
3. Nielsen KA & Moller BL (2005). Cytochrome P450s in plants. In: *Cytochrome P450: Structure, Mechanism, and Biochemistry* (Ortiz de Montellano PR ed), Kluwer Academic/Plenum Publishers, New York, pp. 553–583.
4. Berka K, Hendrychova T, Anzenbacher P & Otyepka M (2011). Membrane position of ibuprofen agrees with suggested access path entrance to cytochrome P450 2C9 active site. *J Phys Chem A*, **115**, 11248–11255.
5. Buck ML (1997). The cytochrome P450 enzyme system and its effect on drug metabolism. *Pediatr Pharmacother*, **3**, 211–216.
6. Fechter K & Porollo A (2014). MutaCYP: Classification of missense mutations in human cytochromes P450. *BMC Med Genomics*, **30**, 7:47.
7. Nebert DW, Wikvall K & Miller WL (2013). Human cytochromes P450 in health and disease. *Philos Trans R Soc Lond B Biol Sci*, **368**, 20120431.
8. Danielson PB (2002). The cytochrome P450 superfamily: biochemistry, evolution and drug metabolism in humans. *Curr Drug Metab*, **3**, 561–597.
9. Lamb DC, Lei L, Warrilow AG, Lepesheva GI, Mullins JG, Waterman MR & Kelly SL (2009). The first virally encoded cytochrome P450. *J Virol*, **83**, 8266–8269.
10. Brindha SK, Zothansanga R, Ajit Kumar P, Vineet Kumar M, Biku Moni C, Bhim Pratap S, Gurusubramanian G & Senthil Kumar N (2013). Characterization of *Bacillus thuringiensis* Cry1 class proteins in relation to their insecticidal action. *Interdiscip Sci Comput Life Sci*, **5**, 127–135.
11. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins

In silico characterization of human cytochrome P450 monooxygenases

- MR, Appel RD & Bairoch A (2008). Protein identification and analysis tools on the EXPASY Server. In: *The Proteomics Protocols Handbook* (John MW ed), Humana Press, pp. 571–607.
12. Bjellqvist B, Hughes GJ, Pasquali C, Paquet N, Ravier F, Sanchez JC, Frutiger S & Hochstrasser D (1993). The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences. *Electrophoresis*, **14**, 1023–1031.
13. Guruprasad K, Reddy BVB & Pandit MW (1990). Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. *Protein Engg*, **4**, 155–161.
14. Kyte J & Doolittle RF, (1982). A simple method for displaying the hydropathic character of a protein. *J Mol Biol*, **157**, 105–132.
15. Atsushi I (1980). Thermostability and aliphatic index of globular proteins. *J Biochem*, **88**, 1895–1898.
16. Tamura K, Stecher G, Peterson D, Filipski A & Kumar S (2013). MEGA6: Molecular evolutionary Genetics Analysis Version 6.0. *Mol Biol Evol*, **30**, 2725–2729.
17. Huelsenbeck JP & Rannala B (2004). Frequentist properties of phylogenetic trees under simple and complex substitution models. *Syst Biol*, **53**, 904–913.
18. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer EL, Tate J & Punta M (2014). Pfam: the protein families database. *Nucleic Acids Res*, **42** (Database issue), D222–230.
19. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW & Noble WS (2008). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res*, **37** (suppl 2), W202–W208.
20. Nelson DR 2009. The cytochrome p450 homepage. *Hum Genomics*, **4**, 59–65.
21. Goldstone JV, Hamdoun A, Cole BJ, Howard-Ashby M, Nebert DW, Scally M, Dean M, Epel D, Hahn ME & Stegeman JJ (2006). The chemical defensible: Environmental sensing and response genes in the *Strongylocentrotus purpuratus* genome. *Dev Biol*, **300**, 366–384.
22. McCart C & French-Constant RH (2008). Dissecting the insecticide-resistance-associated cytochrome P450 gene Cyp6g1. *Pest Manag Sci*, **64**, 639–45.
23. Chiu TL, Wen Z, Rupasinghe SG & Schuler MA (2008). Comparative molecular modeling of *Anopheles gambiae* CYP6Z1, a mosquito P450 capable of metabolizing DDT. *Proc Natl Acad Sci USA*, **105**, 8855–8860.
24. Brindha SK, Sangzuala S, Gurusubramanian G & Senthil Kumar N (2012). Prot-Prop: J-tool to predict the subcellular location of proteins based on physicochemical characterization. *Interdiscip Sci Comput Life Sci*, **4**, 296–301.
25. Graham SE & Peterson JA (1999). How similar are P450s and what can their differences teach us? *Arch Biochem Biophys*, **369**, 24–29.
26. Chan HS & Dill KA (1994). Transition states and folding dynamics of proteins and heteropolymers. *J Chem Phys*, **100**, 9238.
27. Gotoh O (1998). Divergent structures of *Caenorhabditis elegans* cytochrome P450 genes suggest the frequent loss and gain of introns during the evolution of nematodes. *Mol Biol Evol*, **15**, 1447–1459.
28. Tijet N, Helvig C & Feyereisen R (2001). The cytochrome P450 gene superfamily in *Drosophila melanogaster*: annotation, intron-exon organization and phylogeny. *Gene*, **262**, 189–198.
29. Eddershaw P & Dickins M (2004). Phase I metabolism. In: *A Handbook of Bioanalysis and Drug Metabolism* (Evans G, ed), CRC press, Florida, USA, pp. 208–221.
30. Rawal S & Coulombe RA Jr (2011). Metabolism of aflatoxin B1 in turkey liver microsomes: the relative roles of cytochromes P450 1A5 and 3A37. *Toxicol Appl Pharmacol*, **254**, 349–354.