

**Identification and Analysis of Conserved Sequence Motifs in Cytochrome P450 Family 2:
Functional and Structural Role of a Motif ¹⁸⁷RFDYKD¹⁹² in CYP2B Enzymes**

²Numan Oezguen[¶], ¹Santosh Kumar^{¶*}, ²Aditya Hindupur, ²Werner Braun, ³B. K. Muralidhara, and ¹James R. Halpert

¹Skaggs School of Pharmacy & Pharmaceutical Sciences, University of California San Diego, 9500 Gilman Drive, La Jolla, CA 92093

²Department of Biochemistry and Molecular Biology and The Sealy Center for Structural Biology and Molecular Biophysics, University of Texas Medical Branch, Galveston, TX 77555-1031

³Pfizer Inc., PhRD-Global Biologics, 700 Chesterfield Parkway West, Chesterfield, MO 63017

*Address correspondence to: Santosh Kumar, Ph.D., Skaggs School of Pharmacy & Pharmaceutical Sciences, Room 2231, University of California San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0703; Email: s8kumar@ucsd.edu

[¶] Authors with equal contributions

Running title: Investigation of Conserved Sequence Motifs in P450 Family 2

Table S1. Cytochromes P450 analyzed for conserved motifs from 175 P450 family 2 sequences.

CYP2A6_Human	Cyp2c55_Mouse	CYP2A2_Rat	CYP2X3_Fogu
CYP2A7v1_Human	Cyp2c65_Mouse	CYP2A3_Rat	CYP2Y1_Fogu
CYP2A13_Human	Cyp2c66_Mouse	CYP2B1_Rat	CYP2Y2_Fogu
CYP2B6_Human	Cyp2c67_Mouse	CYP2B2_Rat	CYP2Z1_Fogu
CYP2B7P1_Human	Cyp2c68_Mouse	CYP2B3_Rat	CYP2Z2_Fogu
CYP2C8_Human	Cyp2c69_Mouse	CYP2C6_Rat	CYP2K7_Zebrafish
CYP2C9_Human	Cyp2c70_Mouse	CYP2C7_Rat	CYP2K8_Zebrafish
CYP2C18_Human	Cyp2c73-ps_Mouse	CYP2C11_Rat	CYP2K16_Zebrafish
CYP2C19_Human	Cyp2ac1_Mouse	CYP2C12_Rat	CYP2K17_Zebrafish
CYP2D6_Human	Cyp2d22_Mouse	CYP2C13_Rat	CYP2K18_Zebrafish
CYP2D7P1_Human	Cyp2dww_Mouse	CYP2C22_Rat	CYP2K19_Zebrafish
CYP2D8P1_Human	Cyp2d11_Mouse	CYP2C23_Rat	CYP2K20_Zebrafish
CYP2E1_Human	Cyp2d10_Mouse	CYP2D2_Rat	CYP2K21_Zebrafish
CYP2F1_Human	Cyp2d9_Mouse	CYP2D3_Rat	CYP2N13_Zebrafish
CYP2G1P_Human	Cyp2d32-ps_Mouse	CYP2D4_Rat	CYP2P6_Zebrafish
CYP2G2P_Human	Cyp2d12_Mouse	CYP2D18_Rat	CYP2P7_Zebrafish
CYP2J2_Human	Cyp2d33-ps_Mouse	CYP2D5_Rat	CYP2P8_Zebrafish
CYP2R1_Human	Cyp2d34_Mouse	CYP2E1_Rat	CYP2P9_Zebrafish
CYP2S1_Human	Cyp2d13_Mouse	CYP2F4_Rat	CYP2P10_Zebrafish
CYP2T2P_Human	Cyp2d35-ps_Mouse	CYP2G1_Rat	CYP2R1_Zebrafish
CYP2T3P_Human	Cyp2d36-ps_Mouse	CYP2J3_Rat	CYP2U1_Zebrafish
CYP2U1_Human	Cyp2d37-ps_Mouse	CYP2A13_Dog	CYP2V1_Zebrafish
CYP2W1_Human	Cyp2d40_Mouse	CYP2B11_Dog	CYP2X6_Zebrafish
CYP2AB1P_Human	Cyp2d41-ps_Mouse	CYP2C21_Dog	CYP2X8_Zebrafish
CYP2AC1P_Human	Cyp2d26_Mouse	CYP2C41_Dog	CYP2X10_Zebrafish
Cyp2a4_Mouse	Cyp2e1_Mouse	CYP2D15_Dog	CYP2X11_Zebrafish
Cyp2a5_Mouse	Cyp2f2_Mouse	CYP2F_Dog	CYP2X1_Zebrafish
Cyp2a12_Mouse	Cyp2g1_Mouse	CYP2G1_Dog	CYP2Y3_Zebrafish
Cyp2a22_Mouse	Cyp2j5_Mouse	CYP2J_Dog	CYP2Y4_Zebrafish
Cyp2a21-ps_Mouse	Cyp2j9_Mouse	CYP2R1_Dog	CYP2AA1_Zebrafish
Cyp2b9_Mouse	Cyp2j6_Mouse	CYP2S1_Dog	CYP2AA2_Zebrafish
Cyp2b10_Mouse	Cyp2j8_Mouse	CYP2U1_Dog	CYP2AA3_Zebrafish
Cyp2b13_Mouse	Cyp2j11_Mouse	CYP2W1_Dog	CYP2AA5_Zebrafish
Cyp2b19_Mouse	Cyp2j7_Mouse	CYP2AC1_Dog	CYP2AA6_Zebrafish
Cyp2b23_Mouse	Cyp2j12_Mouse	CYP2K9_Fogu	CYP2AA7_Zebrafish
Cyp2c29_Mouse	Cyp2j13_Mouse	CYP2K10_Fogu	CYP2AA8_Zebrafish
Cyp2c37_Mouse	Cyp2r1_Mouse	CYP2K11_Fogu	CYP2AD2_Zebrafish
Cyp2c38_Mouse	Cyp2s1_Mouse	CYP2N9_Fogu	CYP2AD3_Zebrafish
Cyp2c39_Mouse	Cyp2t4_Mouse	CYP2N10_Fogu	CYP2AD6_Zebrafish
Cyp2c40_Mouse	Cyp2u1_Mouse	CYP2N11_Fogu	CYP2AE1_Zebrafish
Cyp2c44_Mouse	Cyp2w1_Mouse	CYP2N12_Fogu	
Cyp2c50_Mouse	Cyp2ab1_Mouse	CYP2P4_Fogu	
Cyp2c52-ps_Mouse	CYP2B4_Rabbit	CYP2R1_Fogu	
Cyp2c53-ps_Mouse	CYP2B5_Rabit	CYP2U1_Fogu	
Cyp2c54_Mouse	CYP2c5_Rabit	CYP2X2_Fogu	

Table S2. Oligonucleotides (forward and reverse primers) used in PCR for site-directed mutagenesis of CYP2B4dH/H226Y (H226Y).

Mutants	Oligonucleotides
K186A	5'-TCCATTGTCTTTGGAG C ACGCTTTGACTACAAG-3' 5'-CTTGTAGTCAAAGCGT G CTCCAAAGACAATGGA-3'
R187A	5'-ATTGTCTTTGGAAA A GCATTTGACTACAAGGACCCC-3' 5'-GGGGTCCTTGTAGTCAAAT G CTTTTCCAAAGACAAT-3'
R187K	5'-TCCATTGTCTTTGGAAAA A AGTTTGACTACAAGGACCCC-3' 5'-GGGGTCCTTGTAGTCAA A CTTTTTTCCAAAGACAATGGA-3'
F188A	5'-ATTGTCTTTGGAAA A CGCG C AGACTACAAGGACCCC-3' 5'-GGGGTCCTTGTAGT C TGCGGTTTTCCAAAGACAAT-3'
D189A	5'-TTTGGAAAACGCTTTGCCTACAAGGACCCCGTGTTTC-3' 5'-GAACACGGGGTCCTTGTAG G CAAAGCGTTTTCCAAA-3'
Y190A	5'-AAACGCTTTGAC G CCAAGGACCCCGTG-3' 5'-CACGGGGTCCTT G GCGTCAAAGCGTTT-3'
K191A	5'-AAACGCTTTGACTAC G CGGACCCCGTGTTTCCTG-3' 5'-CAGGAACACGGGGTCC G CGTAGTCAAAGCGTTT-3'
D192A	5'-TTTACTACAAGGCCCCCGTGTTTCCTG-3' 5'-CAGGAACACGGGG C CTTGTAGACAAA-3'
R187A ¹	5'-GTGTTTGGAGAGGCCTTTGACTACACA-3' 5'-TGTGTAGTCAAAG G CCTCTCCAAACAC-3'
D192A ¹	5'-TTTACTACACAGCCCGCCAGTTTCCTG-3' 5'-CAGGA A CTGGCG G CCTGTGTAGTCAA-3'

The nucleotides changed to make the desired mutation are in boldface. ¹The mutants were created in CYP2B1dH.

Table S3. PCPmer motifs of CYP2, CYP2A, CYP2B, CYP2C, CYP2D, and CYP2J enzymes generated from the sequences presented in Table S1.

CSM	CYP2 <i>CYP2B4_Rabbit</i>	CYP2A <i>CYP2A6_Human</i>	CYP2B <i>CYP2B4_Rabbit</i>	CYP2C <i>CYP2C8_Human</i>	CYP2D <i>CYP2D6_Human</i>	CYP2J <i>CYP2J2_Human</i>
1	³² PGPSPLPVLGN ⁴²	³⁴ PPGPTPLPFIG ⁴⁴	³¹ PPGPSPLPVLG ⁴¹	³⁰ PPGPTP ³⁵	³³ YPPGP ³⁷	⁴⁰ PKNYPPGP ⁴⁷
2	⁶² YGDVFTVYLG ⁷¹	⁶³ ERYGPFVFTIHLGPR ⁷⁶	⁶¹ KYGD ⁶⁴	⁵⁹ KVYGP ⁶³	⁶⁵ FGDVFS ⁷⁰	
3	⁹³ EAFSGRGKIA ¹⁰²	⁹⁹ SGRGEQA ¹⁰⁵	⁹⁶ SGRGKIA ¹⁰²			
4	¹¹⁸ GERWRALRRF ¹²⁷		¹¹¹ YGVIFANGERW ¹²¹	¹⁰⁹ GLGISSNGKRWK EIRRF ¹²⁶		
5	¹³³ RDFGMGK ¹³⁹		¹³⁶ GMGKRSVE ¹⁴³	¹³⁵ GMGK ¹³⁸	¹⁴³ GLGKKSLEQW ¹⁵²	¹⁴⁹ GLGKKSLE ¹⁵⁶
6	¹⁶² GALLDN ¹⁶⁷					
7	¹⁷⁴ ITSNIIC ¹⁸⁰			¹⁷² CAPCNVIC ¹⁷⁹	¹⁸⁰ KAVSN ¹⁸⁴	¹⁸⁷ AVSNIIC ¹⁹³
8	¹⁸⁷ RFDYKD ¹⁹²	¹⁹³ YKDKEFLS ²⁰⁰	¹⁷⁷ NIICSIVFGKRF YKDPVF ¹⁹⁵	¹⁸⁷ FDYKD ¹⁹¹	¹⁹² GRRFEYDDP ²⁰⁰	¹⁹⁸ GERFEYQDSWFQ Q ²¹⁰
9	²¹⁷ FELFSGF ²²³		²¹⁰ SSFSSQVFELFS ²²¹			
10	²²⁷ FPGTHRQI ²³⁴		²²⁶ HFPGTH ²³¹	²²⁵ CFPGTH ²³⁰		
11	²⁶¹ PRDFIDVY ²⁶⁸		²⁵⁷ DPSNPRDFIDVY ²⁶⁸	²⁶⁰ PRDFMDC ²⁶⁶		
12	²⁹⁵ LFFAGTET ³⁰²	³⁰¹ GGTETVS ³⁰⁷	²⁹⁸ AGTETTS ³⁰⁴	²⁹⁷ AGTETTS ³⁰³		³⁰⁷ DLFFAGTE ³¹⁴
13	³²⁵ QKEIEQVIG ³³³				³³⁴ EIDDVIG ³⁴⁰	³⁴⁰ EIDRVIG ³⁴⁶
14	³⁴⁷ PYTDAVIHE ³⁵⁵	³⁴⁹ MPYMEA ³⁵⁴	³⁴⁶ MPYTDA ³⁵¹	³⁴⁶ PYTDAVVH ³⁵³	³⁵³ MPYTDAVIHE ³⁶²	³⁵⁸ SMPYTNAVIH ³⁶⁷
15	³⁸⁰ YVIPKN ³⁸⁵		³⁷⁹ GYVIPKNTE ³⁸⁷	³⁷⁹ YLIPKG ³⁸⁴	³⁸⁶ GFRIPKG ³⁹²	
16	⁴⁰⁵ PNTFNPGHF ⁴¹³		⁴¹⁰ PGHFLDANGA ⁴¹⁹	⁴⁰⁸ DPGHFLDKNG ⁴¹⁷		⁴²² NPDHFLENGQFKK RE ⁴³⁶
17	⁴²⁸ PFSLG ⁴³²	⁴²⁸ AFVPFSIG ⁴³⁵	⁴²⁸ PFSLGKRICLG ²³⁸		⁴³² AFLPFSAG ⁴³⁹	⁴⁴⁰ PFSIGKRACLG ⁴⁵⁰
18	⁴³⁶ CLGEGIARTELF ⁴⁴⁷			⁴³⁵ CAGEGLA ⁴⁴¹	⁴⁴³ CLGEPLA ⁴⁴⁹	
19	⁴⁵¹ TTILQNF ⁴⁵⁷	⁴⁵⁸ QNFRLK ⁴⁶³	⁴⁵⁵ QNFSIASPVPP ⁴⁶⁵			
20	⁴⁸¹ PPSYQIR ⁴⁸⁷		⁴⁷⁶ GVGNVPPSY ⁴⁸⁴	⁴⁸⁰ PPSYQICFIP ⁴⁸⁹		
		Specific to CYP2A	Specific to CYP2B	Specific to CYP2C	Specific to CYP2D	Specific to CYP2J
		¹¹² KGYG ¹¹⁵	¹⁴⁹ EARCLVEELRKS G ¹⁶²	⁷ LVLCLSFMLLFS LW RQ ²²	⁷ VPLAVIVAIF ¹⁶	²¹ LLLGTVAF ²⁸
		¹¹⁹ SNGERAK ¹²⁵	²⁷² MEKDK ²⁷⁶	⁷⁸ HGYEAVKEALIDN G ⁹¹	⁵⁰ DFQNTPY ⁵⁶	⁵¹ PFLGNFFLVD ⁶⁰
		¹³⁹ GVGKRGIEE ¹⁴⁷	²⁸² EFHHQN ²⁸⁷	¹⁴⁷ EEAHC ¹⁵¹	⁸² NGLAAVRE ⁸⁹	⁷³ KKYGNLFS ⁸⁰
		¹⁵¹ EEAG ¹⁵⁴	³¹⁴ MLKYPH ³¹⁹	¹⁵⁸ KTKASPCDP ¹⁶⁶	⁹⁴ HGEDTADRPPVP ¹⁰⁵	⁹³ GLPLIKE ⁹⁹
		²⁰⁴ MMLGIFQF ²¹¹	³⁶¹ DLIPFGVPH ³⁶⁹	²¹¹ PWIQVCNNFP ²²⁰	¹¹¹ GFGPRSQ ¹¹⁸	¹⁰⁷ NFGNRP ¹¹²
		²¹⁷ GQLYEMFSSVM ²²⁷		²⁸¹ EFNIENL ²⁸⁷	¹²² ARYGPAW ¹²⁸	¹¹⁷ REHIFKKN ¹²⁵
		²⁴³ GLEDF ²⁴⁷	³⁹⁵ ALHDPRY ⁴⁰¹	³²¹ AKVQEE ³²⁶	¹⁵⁶ EAACLC ¹⁶¹	¹²⁹ SSGQAWKE ¹³⁶
		²⁶⁰ DPNSP ²⁶⁴		³³² GRHRSPCMQD ³⁴¹		¹⁷⁵ GQFPDPH ¹⁸¹
				⁴⁶⁵ KNLNTTA ⁴⁷¹	²²⁵ NAVSVLLHIPALAG	²¹⁶ DEVTYLEAS ²²⁴

²⁷⁵MQEEKPNTEFY
LKN²⁹⁰
³¹¹RYGF³¹⁴
³¹⁸MKHPEVEAKVHEE
ID³³²
³⁶⁷PMSLARRVKK³⁷⁶
³⁸⁰FRDFFLPKG³⁸⁸
⁴⁶⁸PKDIDVSP⁴⁷⁵
⁴⁸⁷YTMSFLP⁴⁹³

K ²³⁹	
²⁷⁴ AFLAEMEKAKGN ²⁸⁵	²³⁹ FLPGPHQ ²⁴⁵
³²¹ MILHPDVQ ³²⁸	²⁵¹ WKKLK ²⁵⁵
³⁴⁵ PEMGD ³⁴⁹	²⁶⁵ HRKDWNP ²⁷¹
³⁷⁵ THMTSRD ³⁸¹	²⁸⁴ EMSKH ²⁸⁸
⁴²² DAQGH ⁴²⁶	²⁹⁵ SFHEENLIC ³⁰³
⁴⁸⁷ PSPYELCA ⁴⁹⁴	³²¹ RWALLY ³²⁶
	³³¹ PEIQEK ³³⁶
	³⁷³ GNIPQNVPRE ³⁸³
	³⁹¹ AGYHLPKGTM ⁴⁰⁰
	⁴⁰⁹ HRDPTEWATP ⁴¹⁸
	⁴⁷⁴ PNNEKLS ⁴⁸⁰
	⁴⁹¹ PVSHRLC ⁴⁹⁷

The top panel represents the common motifs in subfamilies that are also present in the CYP2 family, whereas the bottom panel represents the motifs that are specific to each subfamily. Residue numbers in the motifs for the CYP2 family and each subfamily are based on the specific P450 indicated in the first row. The CSM8, which is the subject of the current study, is presented in bold. CSM 1, 8, and 14 are common in all subfamilies.