

SEQUENCE LISTING

SEQ ID NO:1 (UGT40087 version 1)

MDASDSSPLHIVIFPWLAFGHMLASLELAERLAARGHRVSFVSTPRNISRLRPVPPALAPLIDFVALPLP
RVDGLPDGAEATSDIPPCKTELHLKALDGLAAPFAAFLDAACADGSTNKVDWFLDNFQYWAAAAAADHK
IPCALNLTFAASTSAEYGVPRVEPPVDGSTASILQRFVLTLEKQCQFVIQRACFELEPEPLPLLSDFGKP
VIPYGLVPPCPPAEGHKREHGNAALSWLDKQQPESVLFIALGSEPVTVEQLHEIALGLELAGTTFLWAL
KKPNGLLLEADGDILPPGFEERTDRGLVAMGWVPQPIILAHSSVGAFLTHGGWASTIEGVMGHPMLFL
TFLDEQRINAQLIERKKAGLRVPRREKDGSYDRQGIAGAIRAVMCEEESKSVFAANAKMQEIVSDRNCQ
EKYIDELIQRLGSFEK

SEQ ID NO: 2 (sr.UGT_g252778)

MATNDDDRKQLHVAMFPWLAFGHILPFLELSKLIAQNGHKVSFLSTTRNIQRLPSHLTPLINLVKLTL
PRVQELPEDAEATTDIKHDDQDHLNADGLQPEVTRFLEEEESPDWIIFDYSYYWLPVAAELGISRA
FFMTFPTWTMALTRLPSDQLTAEDLMTLSKISFKKHEIVNLMYGTSTQGDLYRLTMACNGSDCILIRC
CYEFEPQWLTLLEKLLPVPVVPVGLLPPEIHGDEKDDDTWVSVKEWLDGQHKGHVVYVALGSEAMVSK
DELGELALGLELSGLPFFWALRKPPGSTESDSVELPDGFMERTRNRGVVWTSWAPQLRILSHESVCGF
LTHCGVSSIVEGLMFGHPLIMLPIFGDQIMNAQVLADKQVGIEIPRNEEDGWFTKESVAKSLRSVVVD
DEGEIYKANARELSKIFSDTDLGKKYISHFIDFLMMEIVKT*

SEQ ID NO: 3 (Bd_UGT10840)

MDNGSSSSSSPLHVVICPWLAFGHQLPCLDLAERLALRGHRVSFVSTPRIIARLPVVRPVAASLVDL
VALPLPRVDGLPEGAESTNDVPYEKFELHRKAFFDGLAVPFSEFLRAACAEKGKPDWIIIVDTFHHWAA
AAAEHKKVPCAMLMGLAAGLIVAWATQPSKHVTSEQQEQSAAEPPRFETERRKLATTQRASGMSIAER
CSVTLERCNLVAMRSCLEWEPESIPLATTIGGKQLVPLGLLPSPPEGGRGVSKEDATVRWLDAQPTKS
VVYVALGSEVPLGAKEVHELALGLELAGTRFLWSLRKPSGVSDADILPSGFEERTRGRGLVTMGWVPQ
ISVLAHGAVGAFLTHCGWNSIIIEGLQFGHPLVMLPIFGDQGPNARMMEGRKVGQVPRDESNGSFDRE
GVATTVRVAVVEEEGNRIFTANAKMQEIVADKGCHDKYVDKFIQKLSYME*

SEQ ID NO: 4 (Hv_UGT_V1)

MDGDGNSSSSSSPLHVVICPWLALGHLLPCLDLAERLASRGHRVSFVSTPRNIARLPPLRPVAVAPLVE
FVALPLPHVDGLPEGAESTNDVPYDKFELHRKAFFDGLAAPFSEFLRAACAEAGAGSRPDWLIIVDTFHHW
AAAAAVENKVPVCMMLLLGAATVIAGFARGVSEHAAAAGVKERPAAEAPSFETERRKLMTTQNASGMTV
AERYFLTMRSDLVAIRSCAEWEPESVAALTTLAGKPVVPLGLLPSPPEGGRGVSKEDAAVRWLDAQP
AKSVVYVALGSEVPLRAEQVHELALGLELSGARFLWALRKPTDAPDAAVLPPGFEERTRGRGLVVTGW
VPQIGVLAHGAVAAFLTHCGWNSTIEGLLFHGLPLIMLPISSDQGPNARLMEGRKVGVMQVPRDES DG SF
RREDVAATVRVAVEEDGRRVFTANAKMQEIVADGACHERCIDGFIQQLRSYKA*

SEQ ID NO: 5 (Bd_UGT10850)

MDAAGSSSPMHIVIFPWLASGHLLPCLDLAERLAARGHLVSFVSTPRNLARLPVSPALAPLVLDLVAL
PLPRVAGLPDGAESTADVPADKFDLHRQAFDGLAAPFAAFLDADVGKKKKPDWIVADVFHHWVAAAAQ
EREVPCAMLVPCAAAVAVLAGPPPEISNADERQVIVKVMDAAPRFEAEQAMEEFAAEDASGSSSGLS
VLSRFYMTLKRCKVVALRSCPELEPDAPFLTLRLYGKPAVPLGLLP PPPNGTRSRGMDDEAIIRWLNA
QPASSVVYVALGSEAPLRAELLRELAHGLELAGTRFLWALRKPVGVQDGDVLPDGFVERTSRRGLV
ARWVSQVSILAHGAVGAFLTHCGWGSVVEGLQFGRPLIMLP IAGDQGPNARLMEERKVGVSVPREK
GSFTRGGVAGAIRAVVVEEDGRLFAANAELKREIVASRECHERCIDGFIQHRLRCK

SEQ ID NO: 6 (Ob_UGT91B1_like)

MENGSSPLHVIFPWLAFGHLLPFLDLAERLAARGHRVSFVSTPRNLARLPVPRPALRGLVDLVALPL
PRVHGLPDGAEATSDVPFEKFELHRKAFFDGLAAPFSAFLDAACAGDKRPDWIPDFMHYWVAAAAQKR
GVPCAVLIPCSADVMALYGQPTETSTEQPEAIARSMAAEAPSF EAERNTEEYGTAGASGVSIMTRFSL
TLKWSKLVALRSCPELEPGVFTTLTRVYSKPVVPFGLLP PRRDGAHGVRKNGEDDGAIIRWLDEQPAK
SVVYVALGSEAPVSADLLRELAHGLELAGTRFLWALRRPAGVNDGDSILPNGFLERTGERGLVTTGWV
PQVSILAHAAVCAFLTHCGWGSVVEGLQFGHPLIMLP IIGDQGPNARFLEGRKVGAVVPRNHADGSFD
RSGVAGAVRAVAVEEEGKAFAANARKLQEIIVADRERDERCTDGFIIHLLTSWNELEA

SEQ ID NO: 7 (UGT91D_like3)

MYNVTYHQNSKAMATSDSIVDDRKLHVATFPWLAFGHILPYLQLSKLIAEKGHKVSFLSTTRNIQRL
SSHISPLINVVQLTLPRVQELPEDAEATTDVHPEDI PYLKKASDGLQPEVTRFLEQHSPDWIYDYTH
YWLPSIAASLGISRAHFSVTPWAIAYMGPSADAMINGS DGRTTVEDLTTPPKWFPFPTKVCWRKHD
ARLVYPYKAPGISDGYRMGLVLKGS DCLLSKCYHEFGTQWLPLETLHQVPVVPVGLLPPEIPGDEKDE
TWVSIKKWLDGKQKGSVVYVALGSEVLVSQTEVVELALGLELSGLPFVWAYRKPKGPAKSDSVELPDG
FVERTRDRGLVWTSWAPQLRILSHESVCGFLTHCGSGSIVEGLMFGHPLIMLP IFGDQPLNARLLEDK
QVGIEIPRNEEDGCLTKESVARSLRSVVVEKEGEIYKANARELSKIYNDTKVEKEYVSQFVDYLEKNA
RAVAIDHES*

SEQ ID NO: 8 (Os_UGT_91C1)

MDSGYSSSYAAAAGMHVVICPWLAFGHLLPCLDLAQRLASRGHRVSFVSTPRNISRLPPVRPALAPLV
AFVALPLPRVEGLPDGAESTNDVPHDRPDMVELHRRAFDGLAAPFSEFLGTACADWVIVDVFHAAAA
AALEHKVPCAMMLLGS AHMIAS IADRRLEAETES PAAAGQGRPAAAPT FEVARMKLIRTKGSSGMSL
AERFSLTISRSSLVGRSCVEFEPETVPLLSTLRGKPITFLGLMPPLHEGRREDGEDATVRWLDAQPA
KSVVYVALGSEVPLGVEKVHELALGLELAGTRFLWALRKPTGVSDADLLPAGFEERTRGRGVVATRWV
PQMSILAHAAVGAFLTHCGWNSTIEGLMFGHPLIMLP IFGDQGP NARLIEAKNAGLQVARNDGDGSFD
REGVAAAIRAVAVEEESSKVFQAKAKKLQEIVADMACHERYIDGFIQQLRSYKD*

SEQ ID NO: 9 (Sl_UGT_101249881)

MATNLRVLMFPWLAYGHISPFNLIAKQLADRGFLIYLCSTRINLESIIKKIPEKYADSIHLI
ELQLPELPELPPHYHTTNGLPPLNPTLHKALKMSKPNFSRILQNLKPDLLIYDVLPWAEH
VANEQNIPAGKLLTSCAAVFSYFFSFRKNPGVEFPFPAIHLPEVEKVKIREILAKEPEEGGR
LDEGNKQMMMLCTSRTIEAKYIDYCTELCNWKVVPVGPPFQDLITNDADNKEIDWLGTKEH
NSTVFVSFGSEYFLSKEDMEEVAFALELSNVNFIWVARFPKGEERNLEDALPKGFLEIRIGER
GRVLDKFAPQPRI LNHPSTGGFISHCGWNSAMESIDFGVP I IAMP I HNDQPINAKLMVELGV
AVEIVRDDDGIHRGEIAETLKS VVTGETGEILRAKVREISKNLKSIRDEEMDAVAEELIQ
CRNSNKS

SEQ ID NO: 10 (Sr.KAH)

MEASYLYISILLLLASYLFTTQLRRKSANLPPTVFPSP IPIIGHLYLLKKPLYRTLAKIAA
KYGPILQLQLGYRRVLVISSPSAAEECTNNDVIFANRPKTLFGKIVGGTSLGSLSYGDQ
WRNLRRVASIEILSVHRLNEFHDIRVDENRLLIRKLRSSSSPVTLITVFYALTINVIMRM
ISGKRYFDSGDRELEEEGKRFR EILDETL LLAGASNVGDYLPILNWLGVKSLEKKLIALQ
KKRDDFFQGLIEQVRKSRGAKVGKGRKTMIELLSLQSEPEYYTDAMIRS FVLGLLAAG
SDTSAGTMEWAMSLLVNHPHVLKKAQAEIDRVIGNNRLIDESDIGNIPYIGCIINETLRL
YPAGPLLPHESSADCVISGYNIPRGTM LIVNQWAIHHD PKVWDDPETFKPERFQGLEGT
RDGFKLMPFGSGRRGCPGEGLAIRLLGMTLGSVIQC F DWERVGDEMMDTEGLGVTLPKA
VPLVAKCKPRSEMTNLLSEL*

SEQ ID NO: 11 (UGT40087 version 2)

MDASSPLHIVIFPWLAFGHMLASLELAERLAARGHRVSFVSTPRNISRLRPVPPALAPLIDFVALPLP
RVDGLPDGAEATSDIPPGKTELHLKALDGLAAPFAAF LDAACADGSTNKVDWFLDNFQYAAAAAADHK
IPCALNLTFAASTSAEYGVPRVEPPVDGSTASILQRFVLTLEKCFVIQ RACFELEPEPLPLSDIFGKP
VIPYGLVPPC PPAEGHKREHGNAALSWLDKQQPESVLFIALGSEPPVTVEQLHEIALGLELAGTTFWLAL
KKPNGLLLEADGDILPPGFEERTDRGLVAMGWVPQPIILAHSSVGAFLTHGGWASTIEGVMSGHPMLFL
TFLDEQRINAQLIERKKAGLRVPRREKDGSYDRQGIAGAIRAVMCEEESKSVFAANAKMQEIVSDRNCQ
EKYIDELIQRLGSFEK

SEQ ID NO: 12 (Si91Dlike)

MDSSQSPPLHIAVFPWLAFGHLLPNLELAERLAARGHRVSFVSTPRIISRRLRPVPLALAPLIDFVALPLP
RVDGLPDGAEATSDIPPCKTDLHLKALDGLAAPFAAFDLDAACADGSTNKVDWLFVDSFQYWAAAAAADRK
IPCALILPFASSTLAIEFGVPRLEPPVEGSTASILQRVLTTFEKCQLVIHRACSELEPEHTPLLPGIFGKP
VIQYGLVPPCPPAQGHIEHDNAALSWLDKQQPESVLFIALGSEPPVTVEQLHEIALGLELAGTTFLWALK
KPNGLLLEADGILPPGFEEPTRDRGLVAMGWVPQLSILAHSSVGAFLLTHGGWSSTIEGAMSGHPMVFLT
FLDEQRINAQLIERKKAGLRVPRCEKDGSYDRQGIAGAIRAVMCEEESKSVFAANAKKMQEIINDRKQCE
RYIDELIQRLRSFEK

SEQ ID NO: 13 (UGT40087-1 codon optimized nucleic acid sequence)

ATGGATGCTTCCAGTAGTCCTTTACACATCGTTATCTTTCCATGGTTAGCTTTCGGTCATATGTTGGCTTCCTTG
GAATTGGCTGAGAGATTGGCTGCTCGTGGTCACAGAGTCTCCTTCGTTTCCACCCCTAGAAACATCTCTAGATTA
CGTCCAGTTCACACAGCTTTCAGTCCATTGATTGATTTGTGCGCTTTGCCATTGCCTAGAGTCGATGGTTTACCA
GATGGTGCCGAAGCTACCTCTGACATTCACACAGGTAAGACCGAATTACACTTGAAGGCTTTGGACGGTTTGGCT
GCTCCATTGCGCGCTTTTTTGGACGCTGCCTGTGCTGATGGTTCCACCAACAAGGTTGATTGGTTGTTTTTGGAC
AACTTCCAATACTGGGCTGCCGCTGCCGCTGCTGATCACAAAATTCCTTGCGCCTTAAACTTGACTTTTGGCGCT
TCCACCTCCGCTGAATACGGTGTTCCACGTGTTGAACCACAGTTGACGGTTCCACTGCCTCCATCTTACAAAGA
TTTGCTTTAACCTTAGAAAAATGTCAATTCGTTATCCAAAGAGCTTGTTTCGAATTGGAACCTGAACCATTGCCA
TTGTTGTCCGACATTTTCGGTAAGCCAGTCATCCCATACGGTTTAGTTCCTCCATGTCCACCAGCTGAAGGTCAC
AAAAGAGAACACGGTAACGCTGCTTTGTCTGGTTGGATAAGCAACAACCAGAATCTGTTTTGTTTCATCGCTTTG
GGTTCTGAACACCTGTTACCGTCGAACAATTGCACGAAATCGCTTTGGGTTTAGAATTGGCCGGTACCACCTTC
TTGTGGGCCTTGAAAAAGCCAAACGGTTTGTGTTTAGAAGCCGATGGTGATATTTTGCCACCAGGTTTCAAGAA
AGAAGTAGAGATAGAGGTTTAGTCGCTATGGGTTGGGTTCCACAACCAATTATCTTGGCCCATTCCTCTGTTGGT
GCCTTTTTGACTCACGGTGGTTGGGCTCCACTATTGAAGGTGTCATGTCCGGTCACCCTATGTTGTTCTTAACC
TTCTTGACGAACAACGTATCAACGCCCAATTGATCGAAAGAAAAAGGCTGGTTTAAGAGTCCCAAGAAGAGAA
AAGGATGGTTCTACGACAGACAAGGTATTGCTGGTGCTATTAGAGCCGTCATGTGTGAAGAAGAATCTAAGTCT
GTCTTCGCTGCTAACGCTAAGAAAATGCAAGAGATCGTTTCTGACAGAACTGTCAAGAAAAGTACATCGACGAA
TTGATTCAAAGATTGGGTCTTTTCGAAAAGTAA

SEQ ID NO: 14 (loop2 from Os_UGT_91C1)

EGLPDGAESTNDVPHDRPDMV

SEQ ID NO: 15 (loop3_1 from Os_UGT_91C1)

SEFLGTACAD

SEQ ID NO: 16 (loop3_2 from Os_UGT_91C1)

SEFLGTACADWVIVDVFHH

SEQ ID NO: 17 (loop4_1 from Os_UGT_91C1)

ADRRLERAETESPAAGQGRPAAPTFFEVARMKLIIRTKGSSGM

SEQ ID NO: 18 (loop4_2 from Os_UGT_91C1)

MMLLGSAHMIASIAARRLERAETESPAAAGQGRPAAAPTFEVARMKLIRTKGSSGM

SEQ ID NO: 19 (loop2 from UGT40087)

DGLPDGAEATSDIPPGKT

SEQ ID NO: 20 (loop3_1 from UGT40087)

AAFLDAACADGSTNKVD

SEQ ID NO: 21 (loop3_2 from UGT40087)

AAFLDAACADGSTNKVDWFLDNFQY

SEQ ID NO: 22 (loop4_1 from UGT40087)

GVPRVEPPVDGSTA

SEQ ID NO: 23 (loop4_2 from UGT40087)

LNLTFAASTSAEYGVPRVEPPVDGSTA

SEQ ID NO: 24 (UGT40087/Si91Dlike chimera)

MDASSSPLHIVIFPWLAFGHMLASLELAERLAARGHRVSFVSTPRNISRLRPVPPALAPLIDFVALPLPRVDGLP
DGAEATSDIPPGKTELHLKALDGLAAPFAAFLDAACADGSTNKVDWFLDNFQYWAAAAAADHKIPCALNLTFAA
STSAEYGVPRVEPPVDGSTASILQRFVLTLEKCFVIQRACFELEPEPLPLSDIFGKPVIPYGLVPPCPPAQGH
IEHDNAALSWLDKQQPESVLFIALGSEPPVTVEQLHEIALGLELAGTTFWLWALKKPNGLLLEADGDILPPGFEER
TRDRGLVAMGVWPQLSILAHSSVGAFETHGGWSSTIEGAMSGHPMVFLTFLDEQRINAQLIERKKAGLRVPRCEK
DGSYDRQGIAGAIRAVMCEEESKSVFAANAKMKQEIINDRKQCERYIDELIQRLRSFEK

SEQ ID NO: 25 (Os_UGT_91C1_loop4_1)

MDSGYSSSYAAAAGMHVVICPWLAFGHLLPCLDLAQLASRGHRVSFVSTPRNISRLPPVRPALAPLVAFVALPL
PRVEGLPDGAESTNDVPHDRPDMVELHRRAFDGLAAPFSEFLGTACADWVIVDVFHHWAAAAALEHKVPCAMLL
GSAHMIASIGVPRVEPPVDGSTASLAERFSLTLSSSLVVGSRSCVEFEPETVPLLSTLRGKPITFLGLMPPLHEG
RREDGEDATVRWLDAQPAKSVVYVALGSEVPLGVEKVHELALGLELAGTRFLWALRKPTGVSDADLLPAGFEERT
RGRGVVATRWVPQMSILAHAAVGAFETHCGWNSTIEGLMFGHPLIMLPIFGDQGPNARLIEAKNAGLQVARNDGD
GSFDREGVAAAIRAVAVEEESSKVFQAKAKKLQEIIVADMACHERYIDGFIQQLRSYKD

SEQ ID NO:26 (unoptimized nucleic acid sequence of UGT40087 having SEQ ID NO:1)

tcgtgacgca acagagcaac tctcgccggc accggtcgcc ccttcgcgag gcaggcagca
ggctcgcgcg catggacgcc tccgactcct ccccgctgca catcgtcac ttcccgggc
tcgcgttcgg ccacatgctc gccagcctgg agctcgccga gcgcctggcc gcgcgaggcc
accgcgtgtc cttcgtctcc accccgcgca acatcagccg cctccgccc gtcccgccc

cgctggcgcc gctcctcgac ttcgtggcgc tgccgtgcc gcgctcgac ggccctccccg
acggcgcgga ggccaccagc gacatccccg ccggcaagac cgagctccac ctcaaggccc
tagacggcct cgccgcgccc ttcgcagctt tcctcgacgc cgctcgccg gacgggagca
ccaacaaggt ggactggctc ttcctcgaca acttccaata ctgggcccgc gccgcccgtg
ccgaccataa gataccctgc gcgctgaacc tgacattcgc agcgtcgacg tcagcggagt
acgggtgtgcc acgcgttgag ccgcccgtgg atggctcaac agcctcaata ctccagcgat
ttgtgctaac cttggagaaa tgccagtttg tcatccaacg cgctgcttc gagctggagc
cggagcccc tgcctctcctg tcagacatct tcggcaagcc ggtgatcccc tacggcctag
tcccgcgctg tccccccgca gaaggtcaca aaagagagca cggcaacgca gctctgtcat
ggctcgacaa gcagcagccc gagtctgtcc tgttcattgc tctgggaagc gagcctccgg
tgaccgtcga acagctgcac gagatcgccg ttgggctgga gctcgccggg acgacattcc
tctgggctct gaagaagcct aacggcctcc tcctcgaggc ggacggcgac atcctgcccc
caggttttca ggagcggagc cgtgaccgtg ggctcgtggc catgggctgg gttcctcagc
ccatcatact ggctcacagc tccgtgggcg cgctcctgac gcacggcgga tgggcctcca
ccattgaagg ggttatgtcc gggcatccca tgctcttctt gacgttctta gatgaacaga
ggataaacgc gcaactgacg gagaggaaga aggcggggtt gcgagtgcc aaggcgtgaga
aggacggctc gtacgatcgc caaggcatcg ccggagcgat ccgggctgtc atgtgcgagg
aagaaagtaa gagcgtcttc gcggctaata ccaagaagat gcaggagatt gtgagcgaca
ggaattgccg ggagaagtac atcgacgagc ttatccagcg tctgggatcc ttcgagaagt
gaaataaggt gaaatatcct acaataaccg cctgttgatg gcttgatgca acgatgtagg
tgggcattcg cgctctgtat ctccatgttc cggcaataaa tccaccatat gttatggctc
tgacttactg aatttcctaa tatgtatgcc caaacacatg catagggttg tagttgcccc
tcgcgccggc attagcgata atgtcaccgc agtcgccagc acagggtgtag caatttgaca
t

SEQ ID NO: 27 (modified loop1 from Os_UGT_91C1 present in UGT40087_loop1)

TPRNISRLPPVPALAP

SEQ ID NO: 28 (modified loop1 from UGT40087 present in Os_UGT_91C1_loop1)

TPRNISRLRPVPALAP

SEQ ID NO: 29 (loop1 from Os_UGT_91C1 having SEQ ID NO:8)

TPRNISRLPPVPALAP

SEQ ID NO: 30 (loop1 from UGT40087 having SEQ ID NO:11)

TPRNISRLRPVPALAP