Supplementary Information

Mix-and-Match Evolution II

M. Gregor Madej[1] and H. Ronald Kaback[1,2,3]*

[1]Department of Physiology, [2]Department of Microbiology, Immunology & Molecular Genetics, [3]Molecular Biology Institute, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA 90095

Running title: Structure-Function Correlations in the MFS

Keywords: membrane transport | symport | MFS | sequence alignment | bioenergetics

*Corresponding author: H. Ronald Kaback, rkaback@mednet.ucla.ed

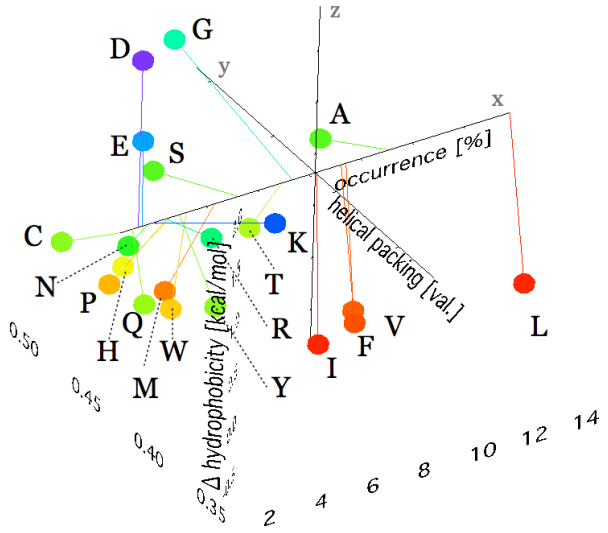

Figure S1: Clustering of amino acid properties (three-dimensional representation)**.** The coordinates are computed from the relative occurrence [%] in alpha-helical membrane proteins for x-coordinate (1), the packing value [val.] for y-coordinate (1) and Δ hydrophobicity [kcal/mol] for z-coordinate (2 and http://blanco.biomol.uci.edu/hydrophobicity_scales.html). The amino acid identity is indicated by single letter code and the positions are colored according to Δ hydrophobicity (from violet to red). Hydrophobic side chains I, F, V and L display similar helical packing and hydrophobicity properties paired with high abundance in in alpha-helical membrane proteins.

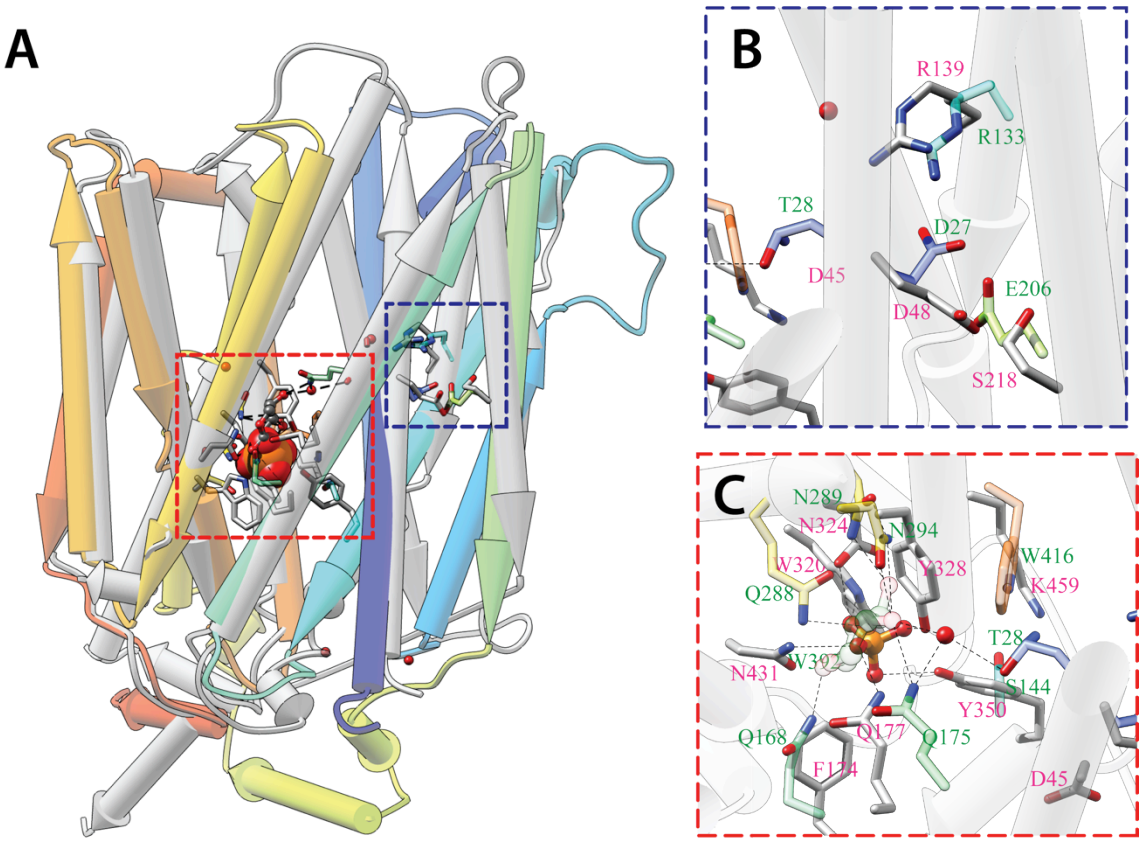

Figure S2: Structure superposition of PiPT and XylE. (A) Schematic representation of crystallographic structures of PiPT (PDBid 4J05; gray) and XylE (PDBid 4GBY; rainbow). A satisfying structure alignment is generated (rmsd achieved: 2.6 Å, number of residues reference (4GBY): 475, number of residues moving (4J05): 422, number of aligned residues: 346, sequence identity: 19.9%) where the putative H^+-binding-sites (B, blue square) and the substrate binding-sites (C, red square) superpose. In the detailed representations residue labels of XylE are shown in green color and residue labels of PiPT are shown in magenta color. Xylose is shown a transparent shape for better visually.

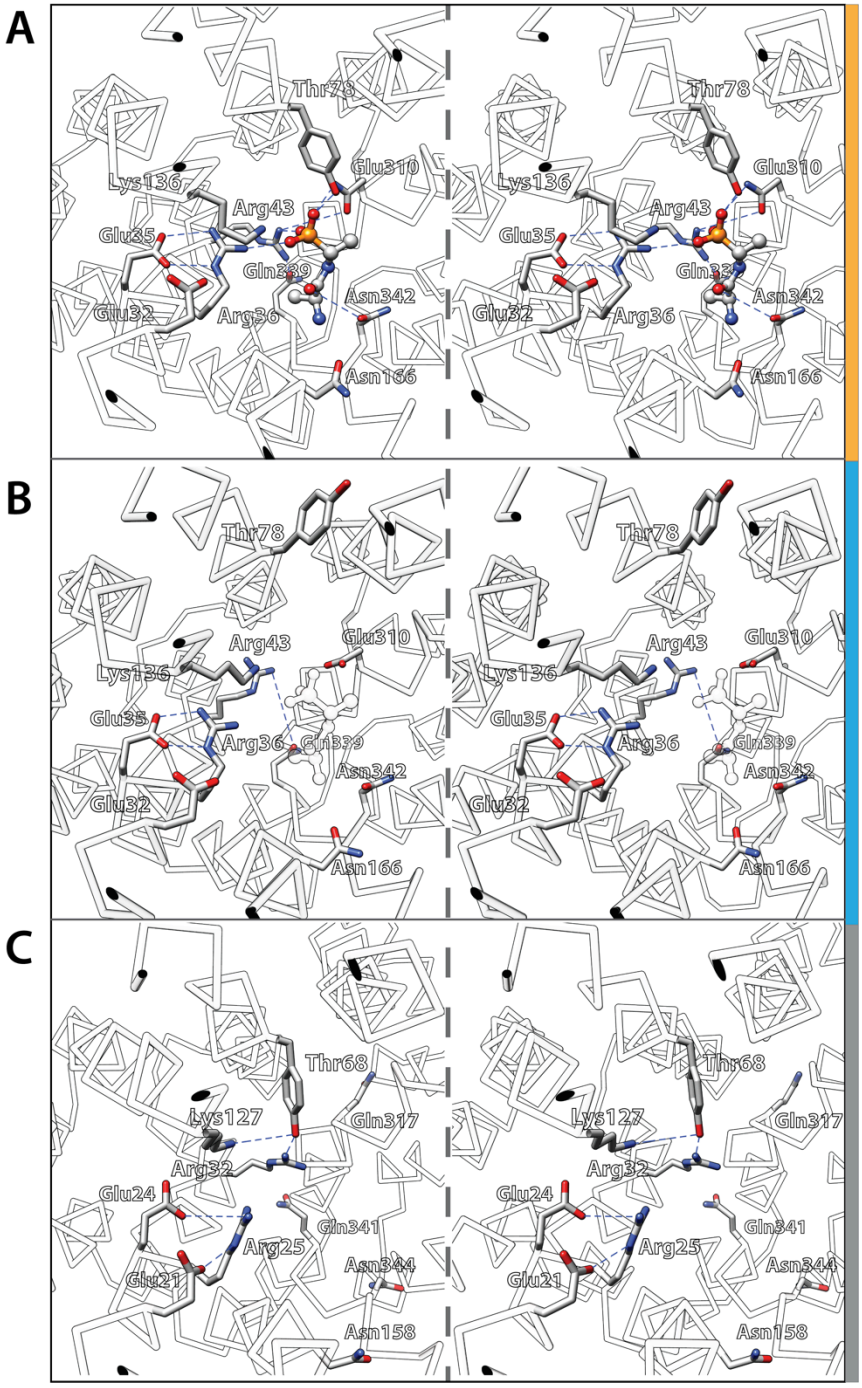

Figure S3: Conformational changes in the substrate binding site of PepT. (A) Inward-open, substrate-bound state of $PepT_{Gk}$ (pdbID: 4IKZ). (B) Inward-open, substrate-free state of $PepT_{Gk}$ (pdbID: 4IKV). The position of the substrate detected in the substrate-bound state is indicated as white profile. (C) Inward-occluded conformation of $PepT_{So}$ (pdbID: 2XUT). The colors of the bars on the right side indicate the mechanistic affiliation of the respective states with regard to the mechanistic model shown in Fig. 5.

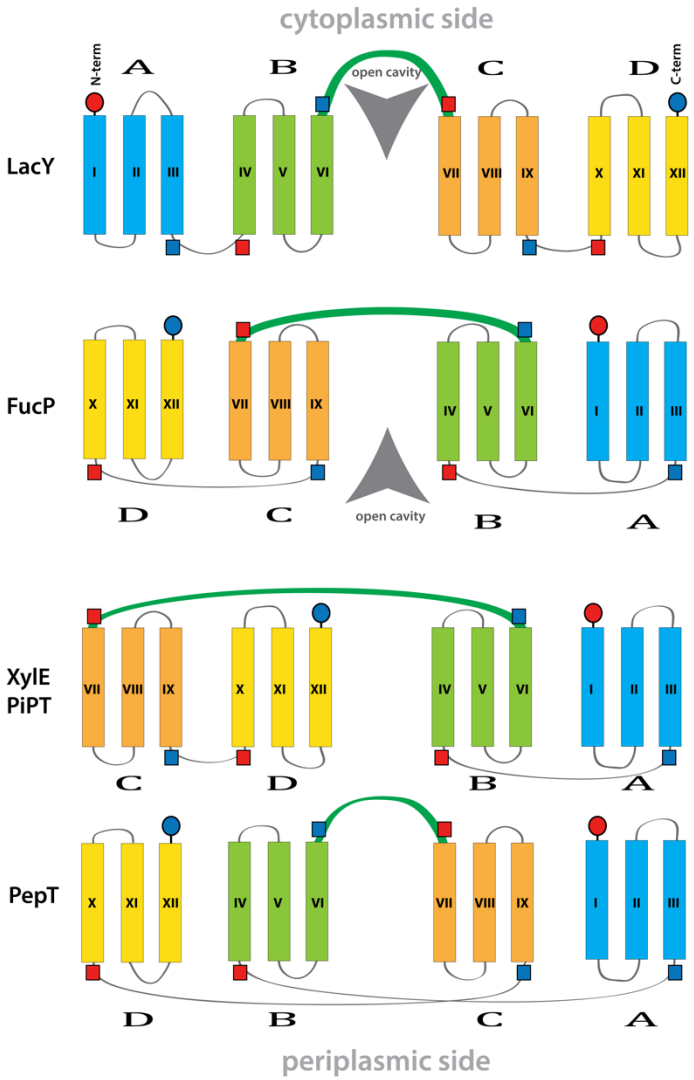

Figure S4: Orientation of helix-triplets. Helix-triplets from FucP , PiPT, XylE and PepT are aligned with LacY (helices 1–3, blue; helices 4–6, green; helices 7–9, orange; helices 10–12, yellow) according to Fig. 2, the order of the helices (roman numerals) is shown according to the alignment to LacY. The N- and the C-termini of the helix-triplets are indicated by red and blue rectangles respectively. A red dot marks the N-terminus of the protein and a blue dot marks the C-terminus. The central middle loop is shown as a green line. Although the orientation of homologous helix-triplets may be flipped relatively to LacY, The N-, the C-termini and the middle loop are always on the cytoplasmic side.

Table S1: Reduction matrix for sequence simplification. Grouping of residues based on reasonable simplification of protein sequence with no interlacing for different levels of reduction (modified from ref. 2). Letter grouped without spaces represent a reduction clade and are treated as similar in scoring of sequence alignments.

level	score	reduction alphabet	symbol
0.	20	L I V F Y W M G P A T S C N Q E D H R K	*
1.	10	LIV F YW M GP ATS C NQED H RK	!
2.	7	LIVF YW M GP ATSC NQED HRK	:
3.	4	LIVF YWM GPATSC NQEDHRK	.

Equation S1: Similarity score.

$$Q = \frac{\sum redValue}{(AA_{all} - AA_{LIVF}) \cdot 20}$$

Q: similarity score, $redValue$: score assigned from the reduction matrix (SI Appendix, Table S1), AA_{all}: number of all residues in the alignment, AA_{LIVF}: number of not compared side-chains. For 100% conservation, Q should equal 1.

Equation S2: Functional similarity score.

$$Q_{Act} = \left(\frac{sAA_{0-25\%} \cdot 5}{AA_{0-25\%}} + \frac{sAA_{25-50\%}}{AA_{25-50\%}} \right) \div 6$$

Q_{Act}: functional similarity score, $sAA_{0\text{-}25\%}$: number of aligned critical positions in the alignment, $AA_{0\text{-}25\%}$: number of all critical positions in the alignment, $sAA_{25\text{-}50\%}$: number of aligned functionally relevant positions in the alignment, $AA_{25\text{-}50\%}$: number of all functionally relevant positions in the alignment. Critical positions are positions where mutations to Cys cause greater than 50% and functionally relevant positions are positions where mutations to Cys cause 75% inhibition of the transport rate with Cys-less LacY (3, 4). For 100% conserved transporters with 100% conserved functional positions this value, Q_{Act}, should equal 1.

Sequence Alignments

LacY / PiPT

```
                                    ^     ^    ^       $      #   #$
                                                             =    -
LacY-A.pdb,_chain_B/1-97          NTNFWMFGLFFFFYFFIMGAYFPFF....PIWLHDINHIS.......... 36
PiPT-c.pdb,_chain_A/306-410       TWNHFRNLLGSMLGWFLVDIAFYGINLNQSVVLAQIGFAGKTGDVYDKLF 50
                                    *                              .     !   .

                                              ^    ^  #$     $#       ^       ^
                                                          -      =
LacY-A.pdb,_chain_B/1-97          .KSDTGIIFAAISLFSLLFQPLFGLLSDKLGLRKYLLWIITGMLVMFAPF 85
PiPT-c.pdb,_chain_A/306-410       QLATGNIIVTALGFLPGYYFTLF..LIDIVG.....RKKLQFMGFIMSGL 93
                                    ! .    !* .  .    .      *  *           *    !.

                                         $
                                         -
LacY-A.pdb,_chain_B/1-97          FIFIFGPLLQYN 97
PiPT-c.pdb,_chain_A/306-410       FLAILAGEIDHI 105
                                         !!  !

                                      $^  $    #     $# # #     ## $$ $  #^##      #$
                                      -              -= =       =  -       ==      =-
LacY-B.pdb,_chain_B/1-83          LVGSIVGGI.YLGFCFNAGAPAVEAFI..EKVSRRSNFEFGRARM..FGCV 46
PiPT-D.pdb,_chain_A/411-499       GKGPLLACFTFMQFFFNFGANTTTFIVAAELFPTRIRAS.AHGISAAAGKC 50
                                    *.  .         * ** .       *  . * .   .. !    *.

                                  ##  #    #                 #      ^^
                                  =   =
LacY-B.pdb,_chain_B/1-83          GWALCASIVGIMFTI.NNQFVFWLGSGCALILAVLLF.F 83
PiPT-D.pdb,_chain_A/411-499       GAILSSLVFNQLKAKIGTSAVLWIFFSTCILGFISTFLI 89
                                  *   :!                *   .

                                              #^ $     $#  #  ^$  ^    ^     $   $
                                                        =                    -
LacY-C.pdb,_chain_B/1-104         LFRQPKL.WFLSLYVIGVSCT.YDVFDQQFANFFTSFFATGEQGTRVFGYVT 50
PiPT-B.pdb,_chain_A/126-229       WDGNRVLTWITICRVFLGIGIGGDYPMSATVVSDR.ANI.HRRGTLLCFIFA 50
                                     !    *          .   *                 ..**      !

                                   $#   $       #   $  #$#  $^   #     #   $
                                   -            =       =  -          =   -
LacY-C.pdb,_chain_B/1-104        TMGE.LLNASIMFFAPLIINRIGGKN.ALLLAGTIMSVRIIGSSFATSALEV 100
PiPT-B.pdb,_chain_A/126-229      NQGWGSFVGSLVTIVTISGFKH..RLKSGHTHDVDKAWRILIGLSLIPAFGT 100
                                   *      .     .    !   !  !        ! **  .    .*
```

```
LacY-C.pdb,_chain_B/1-104        VILK 104
PiPT-B.pdb,_chain_A/126-229      LYQR 104
                                    !

                                        #  # ^# $^   $ ^#           $# #    ^   #
                                           =  =         =            =        =
LacY-D.pdb,_chain_B/1-89        SALEVVILKTLHMFEVPFLLVGCFKYITSQFEVRFSATIY.LVCFCFFKQ 49
PiPT-A.pdb,_chain_A/30-116      LVLLAGVGFFLDAYDL.FII...NQVAPMLAQVYFPKTGLPAQRQDLMKA 46
                                          .  !         .  .   !   . *          *

                                                $             ^        ^
                                                -
LacY-D.pdb,_chain_B/1-89        LAMIFMSVLAGNMYESIG......FQGAYLVLGLVALGFTLISVFT 89
PiPT-A.pdb,_chain_A/30-116      AANI.GCVVGQVMFGVLGDSFGRKFVYGKELILIIVATIFQM.... 87
                                 *    .  .  *    *         .
```

LacY / XylE

```
                                          ^     ^    ^       $   #   #$               ^
                                                               =   =
LacY-A.pdb,_chain_B/1-96                TNFWMFGLFFFFYFFIMGAYFPFF.PIWLHDINHISKSDTGII.FAAISLFS 50
XylE-C.pdb,_chain_A/276-365             GVIVIGVMLSIFQQFVGINVVLYYAPEVFKTLG.AS.TDIALLQTIIVGVIN 50
                                        .                        *   :     * !* .       .

                                        ^  #$     $#       ^       ^           $
                                                   =                          -
LacY-A.pdb,_chain_B/1-96                LLFQPLFGLLSDKLGLRKYLLWIITGMLVMFAPFFIFIFGPLLQYN 96
XylE-C.pdb,_chain_A/276-365             LTFTVLAIMTVDKFG.RKPL.QII.GALGMAIGMFSLGTAFYT... 90
                                                  ** * **       *   *  !      .

                                            $^  $   #     $# #  #   ## $$ $  #^##    #$  #
                                                   =      = =       =    -    ==    =
LacY-B.pdb,_chain_B/1-88                ILVGSIVGGIYLGFCF.NAGAPAVEAFIEKVSRRSNFEFGRARMFGCV.G 48
XylE-D.pdb,_chain_A/369-465             GIVALLSMLFYVAAFAMSWG.PVCWVLLSEIF..PNAIRGKALAIAVAAQ 47
                                           .      * .      * *       .    .*   *!*   .

                                        #  #    #                              #      ^^
                                        =  =    =
LacY-B.pdb,_chain_B/1-88                WALGASIVGIMFTIN...........NQFVFWL.GSGCALILAVLLFFAKTD 88
XylE-D.pdb,_chain_A/369-465             WL.ANYFVSWTFPMMDKNSWLVAHFHNGFSYWIYGCMGVLAALFMWKFV.PE 97
                                        *  .    .   .             *    *  *              .!
```

```
                                        #^ $    $#  #  ^$  ^   ^        $   $     $
                                        =             -                        -
LacY-C.pdb,_chain_B/1-91              KLWFLSLYVIGVSCTYDVFDQQFANFFTSFFAT..GEQGTRVFGYVTTMG 48
XylE-B-coot-0.pdb,_chain_A/125-221    YVPEFVIYRIIGGIGVGLASMLSPMYIAELAPAHIRGKLVSFNQFAIIFG 50
                                        *    . .        .   !   .!    .          *

                                      #  $       #   $       #$# $^   #     #   $
                                      =          =   -       = =
LacY-C.pdb,_chain_B/1-91              ELLNASIMFFAPLIINRI.....GGKNALLLAGTIMSVRIIGSSFATS 91
XylE-B-coot-0.pdb,_chain_A/125-221    QLLVYCVNYFIARSGDASWLNTDGWRYM.FASECIPALLFLMLLYTVP 97
                                      !          .   !       * !     !    !        ! .

                                                 #  #  ^# $^   $   ^#            $# #
                                                          = -    -    =            -=
LacY-D.pdb,_chain_B/1-101           GSSFATSALEVVILKTLHM.FEVPFLLVG.C.FKYIT.SQFEVRFSATIY 46
XylE-A-coot-0.pdb,_chain_A/5-110    NSSYIFSITLVATLGGLLFGYDTAVISGTVESLNTVFVAPQNLSESAANS 50
                                     **   *        .     ! .    . .  .    !  !   **!

                                       ^    #                 $       ^
                                                              -
LacY-D.pdb,_chain_B/1-101           LVCFCFF.KQLAMIFMSVLAGNMYESIGFQGAYLVLGLVALGFTLISVFT 95
XylE-A-coot-0.pdb,_chain_A/5-110    LLGFCVASALIGCIIGGALGGYCSNRFGRRDSLKIAAVLFFISGVGSAWP 100
                                        *      .    .  .*   !  * . !    .      .  *  .

                                    ^
LacY-D.pdb,_chain_B/1-101           LSGPGP 101
XylE-A-coot-0.pdb,_chain_A/5-110    ELGFTS 106
                                      * ..
```

LacY / PepT

```
                                      ^     ^    ^       $  #   #$               ^
                                                          =
LacY-A.pdb,_chain_B/1-96            TNFWMFGLFFFFYFFIMGAYFPFFPIWLHDINHIS.KSDTGIIFAAISLF 49
PepT_4IKZ_D.pdb,_chain_A/1-101      GGLVHPIWLVLSYFIVVLGELCLSPVGLSATTKLAPAAFSAQTMSLWFLS 50
                                             *     .     *       . !  ! !.   !

                                     ^  #$     $#        ^       ^           $
                                        =-      =                           -
LacY-A.pdb,_chain_B/1-96            SLLFQPLFGLLSD.KLGLRKYLLWIITGMLVMFAPFFIFIFGPLLQ...Y 95
PepT_4IKZ_D.pdb,_chain_A/1-101      NAAAQAINAQLVRFYTPENETAYFGTIGGAALVLGLILLAIAPRIGRLMK 100
                                        *.  .   .   ! ..       *      !      .*

LacY-A.pdb,_chain_B/1-96            N 96
PepT_4IKZ_D.pdb,_chain_A/1-101      G 101
```


LacY-B.pdb,_chain_B/1-83 LVGSIVGGIYLG...FCFNAGAPAVEAFIEKVSRRS...N.FE..... 36
PepT_4IKZ_B.pdb,_chain_A/1-100 MAGHIALAIPGGVAALFVSMALIVLGTGLLKPNVSSIVGDMYKPGDDRRD 50
LacY-B.pdb,_chain_B/1-83 ..FGRARMFGCVGWALGASIVGIMFTINNQFVFWLGSGCALILAVLLF.. 82
PepT_4IKZ_B.pdb,_chain_A/1-100 AGFSIFYMGINLGAFLAPLVVGTAGMKYNFHLGFGLAAVGMFLGLVVFVA 100
LacY-C.pdb,_chain_B/1-88 LWFLSLYVIGVSCTYDVFDQQFANFFTSF.FATGEQGTRVFGYVTTMGEL 49
PepT_4IKZ_C.pdb,_chain_A/1-102 RVIAYIPLFVASAMFWAIQQQGSTILANYADKRTQLDVAGIHLSPAWFQS 50
LacY-C.pdb,_chain_B/1-88 LNASIMFF.APLIINRIG.....G.KNA..LLLAGTIMSVRIIGSS 86
PepT_4IKZ_C.pdb,_chain_A/1-102 LNPLFIIILAPVFAWMWVKLGKRQPTIPQKFALGLLFAGLSFIVILVPGH 100
LacY-C.pdb,_chain_B/1-88 FA 88
PepT_4IKZ_C.pdb,_chain_A/1-102 LS 102
LacY-D.pdb,_chain_B/1-89 SALEVVILKTLHMFEVPFLLVGCFKYITSQ.....FEVRFSATIY 40
PepT_4IKZ_A.pdb,_chain_A/1-99 HPKGLFTLFFTEFWERFSYY.GM.RAILVYYMYYEVSKGGLGLDEHLALAIM 50
LacY-D.pdb,_chain_B/1-89 LVCFCFFKQLAMIFMSVLAGNMYESIGFQGAYLVLGLVALGFTLISVFT 89
PepT_4IKZ_A.pdb,_chain_A/1-99 SIYGALVYMSGIIGGWLADRVFGTSRAVFYGGLLIMAGHIALAIPGGVA 99

REFERENCES

1. Li T, Fan K, Wang J, & Wang W (2003) Reduction of protein sequence complexity by residue grouping. *Protein Eng* 16(5):323-330.
2. Eilers M, Shekar SC, Shieh T, Smith SO, & Fleming PJ (2000) Internal packing of helical membrane proteins. *Proc Natl Acad Sci U S A* 97(11):5796-5801.
3. Frillingos S, Sahin-Toth M, Wu J, & Kaback HR (1998) Cys-scanning mutagenesis: a novel approach to structure function relationships in polytopic membrane proteins. *Faseb J* 12(13):1281-1299.
4. Madej MG, Dang S, Yan N, & Kaback HR (2013) Evolutionary mix-and-match with MFS transporters. *Proc Natl Acad Sci U S A* 110(15):5870-5874.