Supplemental Materials for

Instant integrated ultradeep quantitative-structural membrane proteomics discovered PTM signatures for human Cys-loop receptor subunit bias

by Xi Zhang, Ph.D.

1. Pdf of Supplemental Figures and Notes

Supplemental Fig. S1. PSM distribution against the sequences of $GABA_AR$ subunits and HPLC MS chromatography from three runs, supplementing **Fig. 3A.**

Supplemental Fig. S2. Ox-TMT2 PSM distribution against the sequences of $GABA_AR$ subunits, HPLC MS chromatography from three samples that combined multiple independent digestions, and representative MS and HCD MS/MS spectra, supplementing **Fig. 3G.**

Supplemental Fig. S3. HCD MS/MS spectra for selected Cys-PTM identifications (peptide FDR<1% by PD1.3 SEQUEST) discussed in **Fig. 2.** More annotated MS/MS spectra for tentative PTM identifications are shown in supplemental Data 1 and Data 2.

Supplemental Fig. S4. Bases for Cys-loop hLGIC PTM signature modeling, supplementing **Fig. 5**.

Supplemental Fig. S5. ECD sequence alignment of $hGABA_AR$ subunits with 3RHW and 2QC1, and sequence details of the four major mid-ECD NXS/T coding regions (A1, B, C1/C2) across Cys-loop hLGICs, supplementing **Fig. 5A** and **5E**.

Supplemental Fig. S6. Sequence alignments of ICL2 of human Cl^- and cation channels showed high diversity in sequence and in K/R and D/E distribution, supplementing **Fig. 5A.**

Supplemental Fig. S7. Applicability to direct Ox-TMT structural mapping of $GABA_AR$.

Supplemental Note 1. Calculation of column-residence time.

Supplemental Note 2. A catalytic site-occupancy model for digestion to overcome bias and enhance efficiency.

2. Excel of all Supplemental Tables

<Please see excel for all Supplemental Tables>

Supplemental Table S1. Peptides identified in a representative direct analysis of deglycosylated $GABA_AR$ digest (Expt. 6 in **Fig.2A**), supplementing **Fig. 2A** and **Fig. 4A**.

Supplemental Table S2. Full list of proteins identified in $GABA_AR$ digest by searching against the human proteome and filtered with CRAPome (version 1.1, 411 entries), supplementing **Fig. 2C**.

Supplemental Table S3. Peptides for PTM label-free peptide-centric quantitation of % site occupancy (N-glycosylation/deamidation and Mox), supplementing **Fig. 4BCDE**.

Supplemental Table S4. Peptides for representative K-Me1/Me2/Me3/Ac screening, supplementing **Fig. 5A**.

Supplemental Table S5. Basal oxidation (ox, diox) screening for multiple resides, using raw data of Expt. 4-6 in **Fig. 2A**.

Supplemental Table S6. Peptides for identification of H2O2-oxidized residues (ox, diox) in the apo7'/apo 4' Ox-TMT2 sample, for supplemental **Fig. S7A**.

Supplemental Table S7. MS/MS TMT2 127/126 ratios for peptides identified in H_2O_2-oxidized GABA 7'/apo 7' sample, for supplemental **Fig. S7B**.

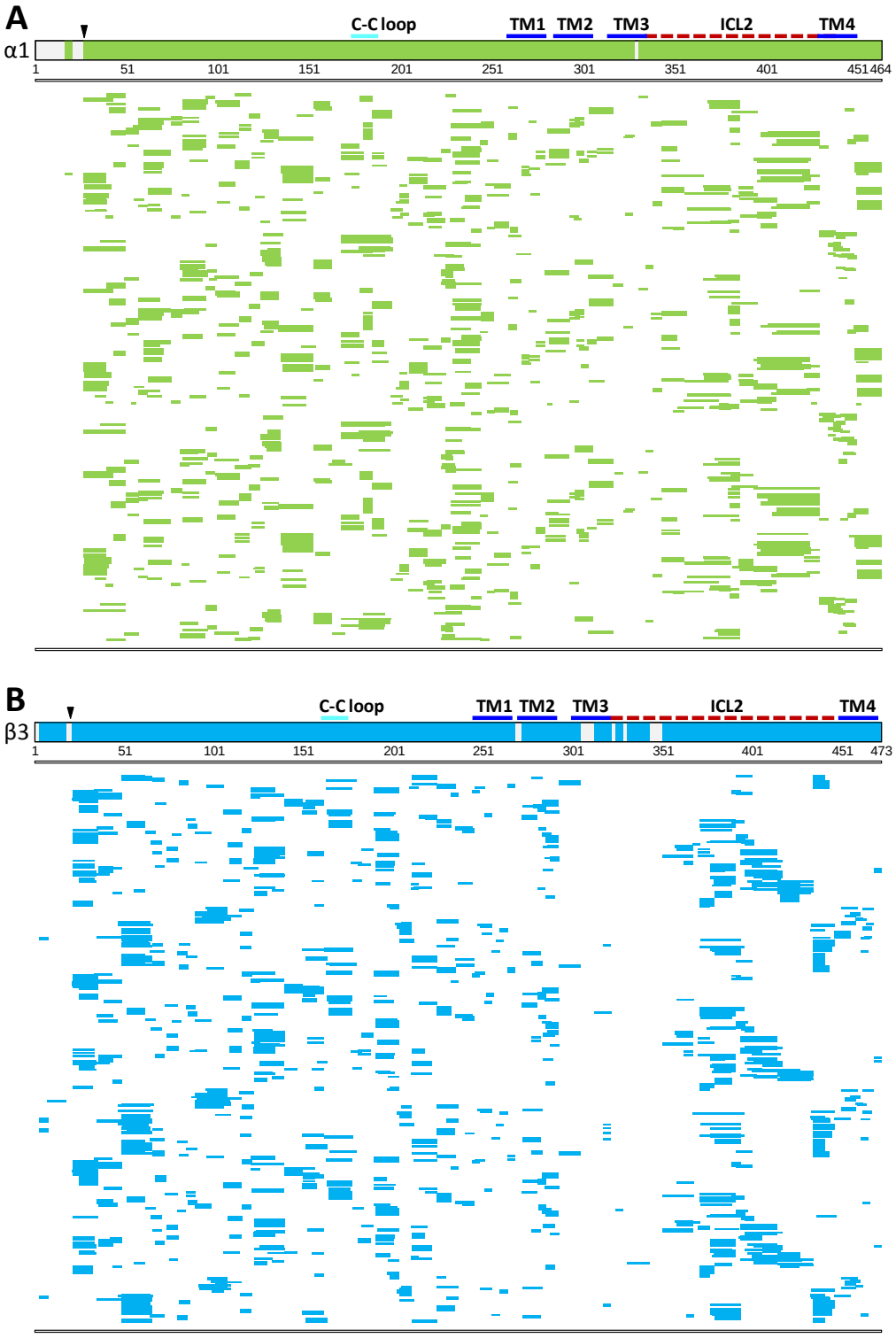
3. Excel of Supplemental Data: Annotated MS/MS Spectra for PTM Identification

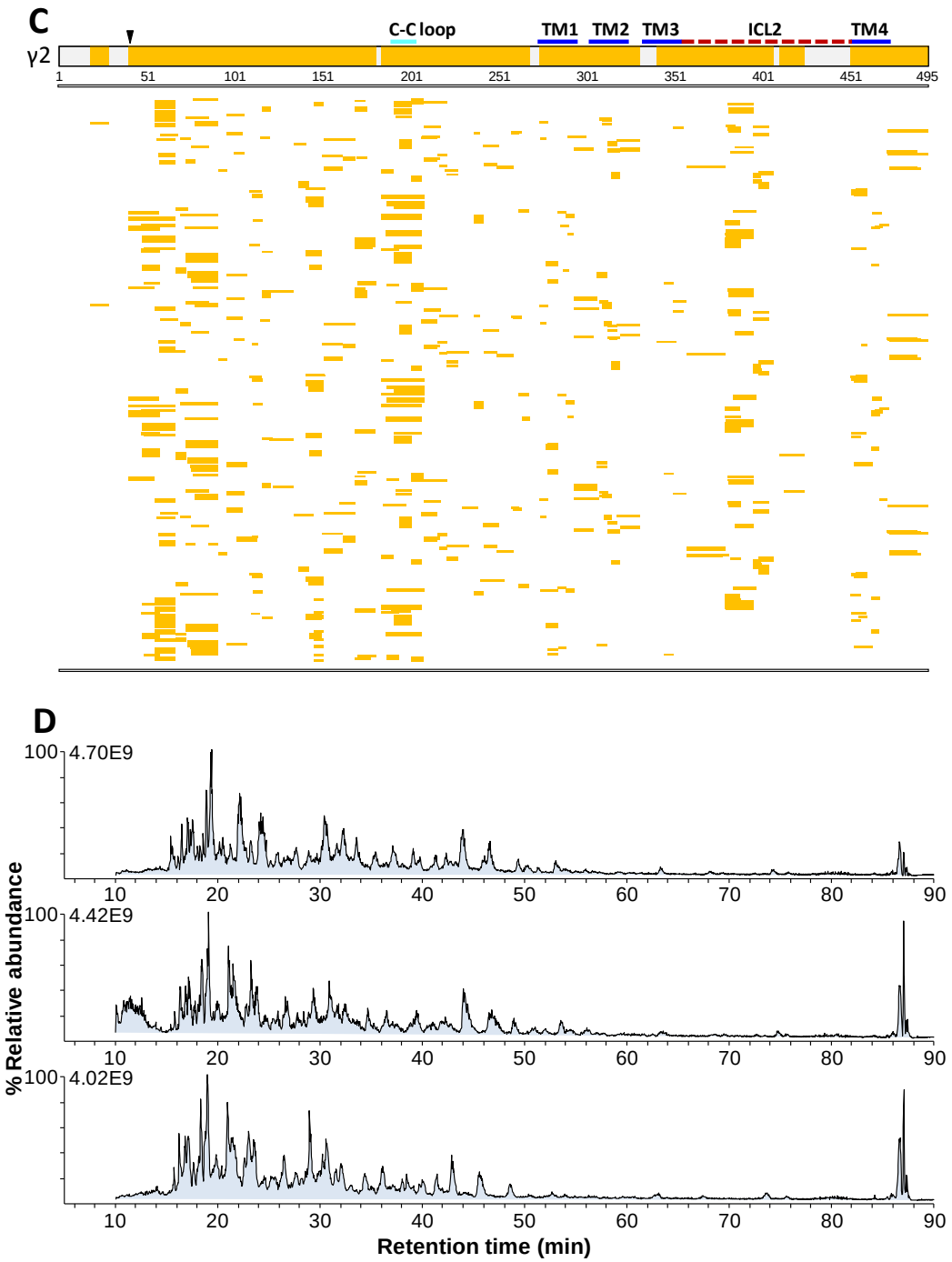
<Please see excel for all supplemental Data>

Supplemental Data 1. Representative annotated MS/MS spectra for N-deamidation, Mox and C-Cb PTM-bearing peptides shown in **supplemental Table S3**. These PTMs were identified by searching for dynamic Mox, C-Cb, deglyco and STY-P against the overexpressed $GABA_AR$ sequences (SEQUEST-HT PD1.4, peptide FDR<1% "high", using spectra of Expt. 4-6 in **Fig. 2A**). Both modified and non-modified N- and M-peptides are shown.

Supplemental Data 2. Representative annotated MS/MS spectra for K-PTM peptides shown in **supplemental Table S4**. Representative peptide K-Me1/Me2/Me3/Ac screening was done by searching for dynamic CK-Cb, KRC-Me, KR-Me2, KR-Me3, KC-Ac and deglyco against the overexpressed $GABA_AR$ sequence (SEQUEST-HT PD1.4, peptide FDR<1% "high", using raw data of Expt. 4-6 in **Fig. 2A**). Both modified and non-modified K-peptides are shown.

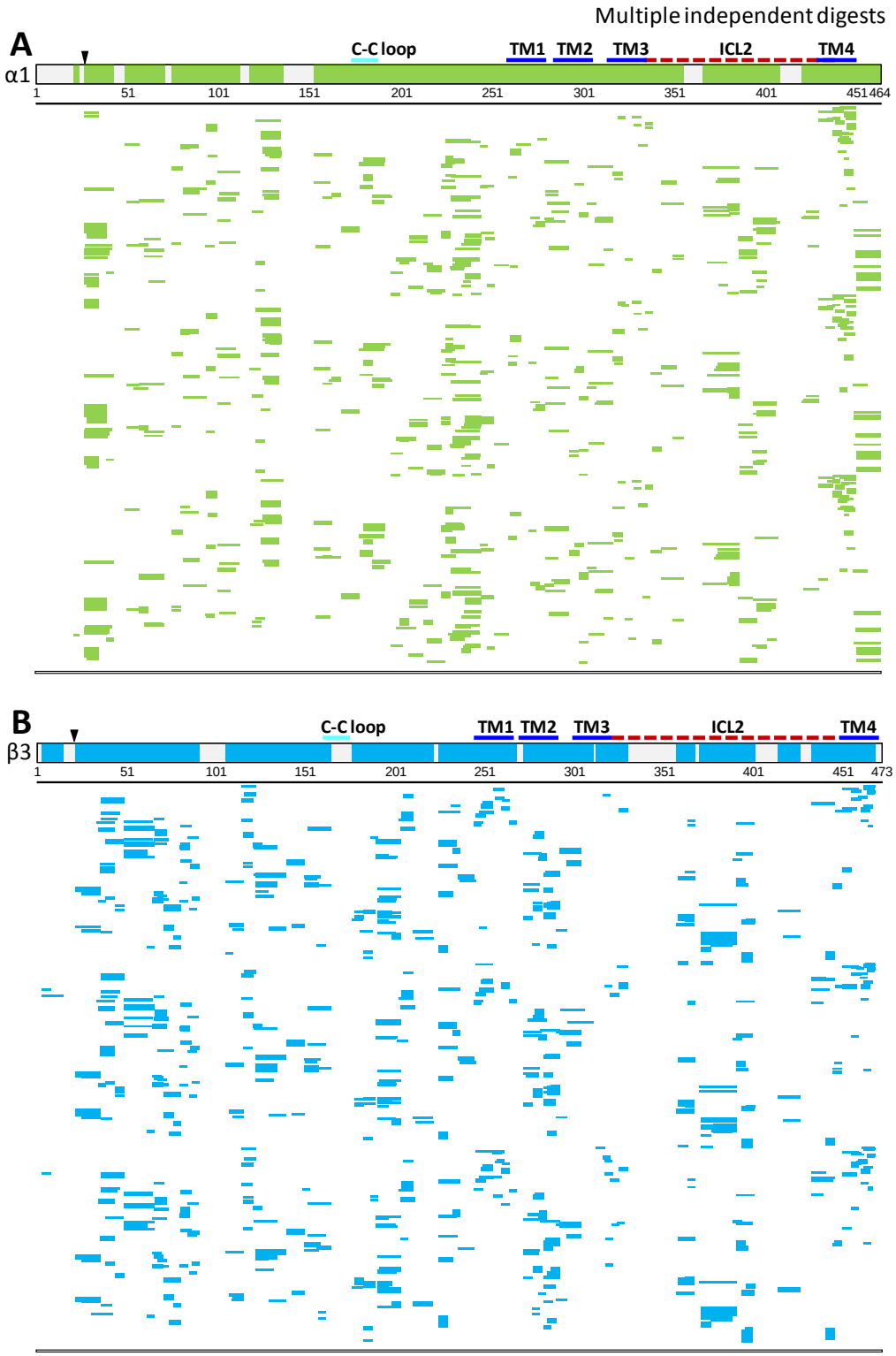
Supplemental Figure S1.

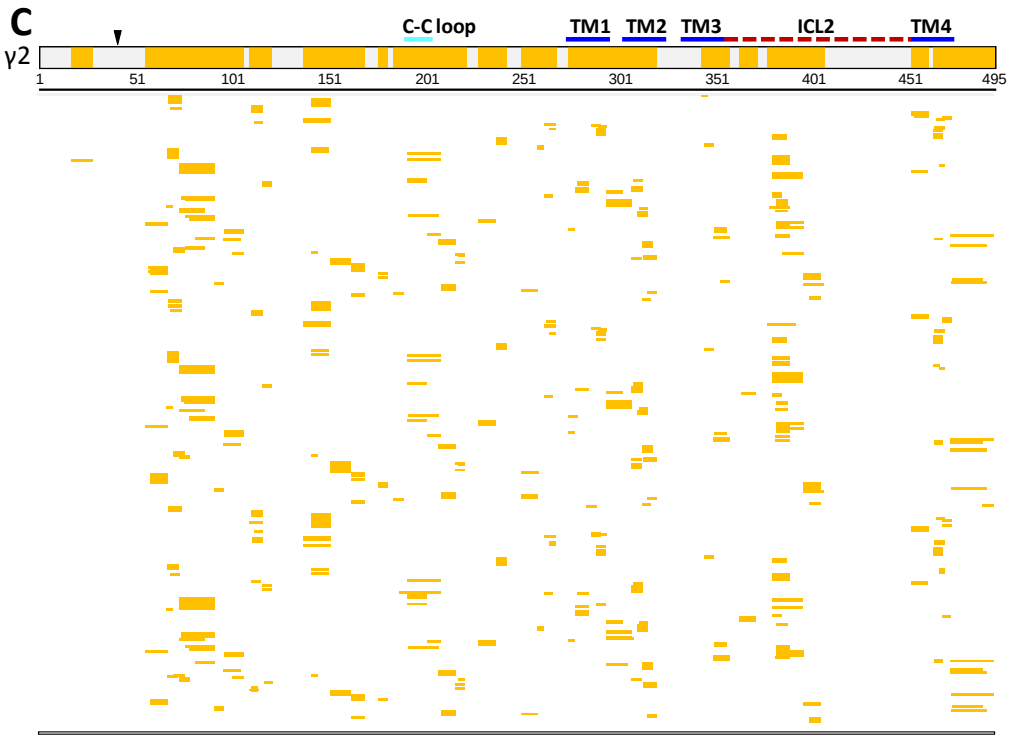


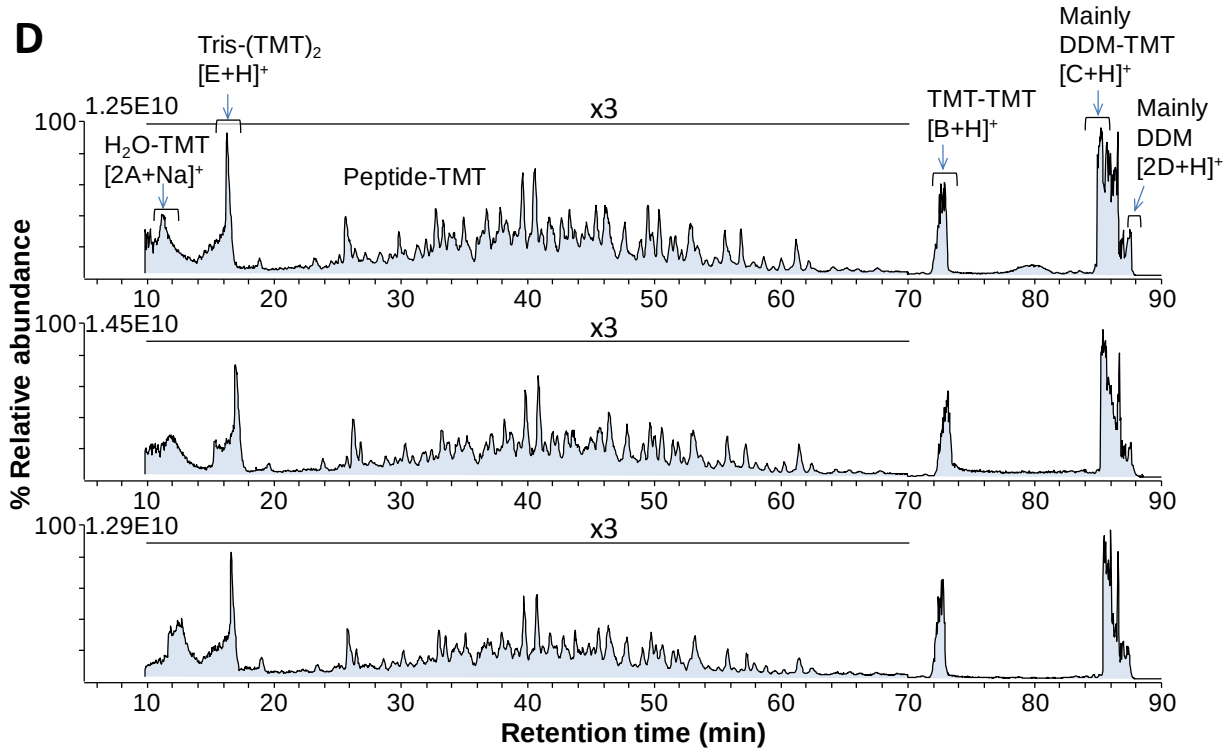


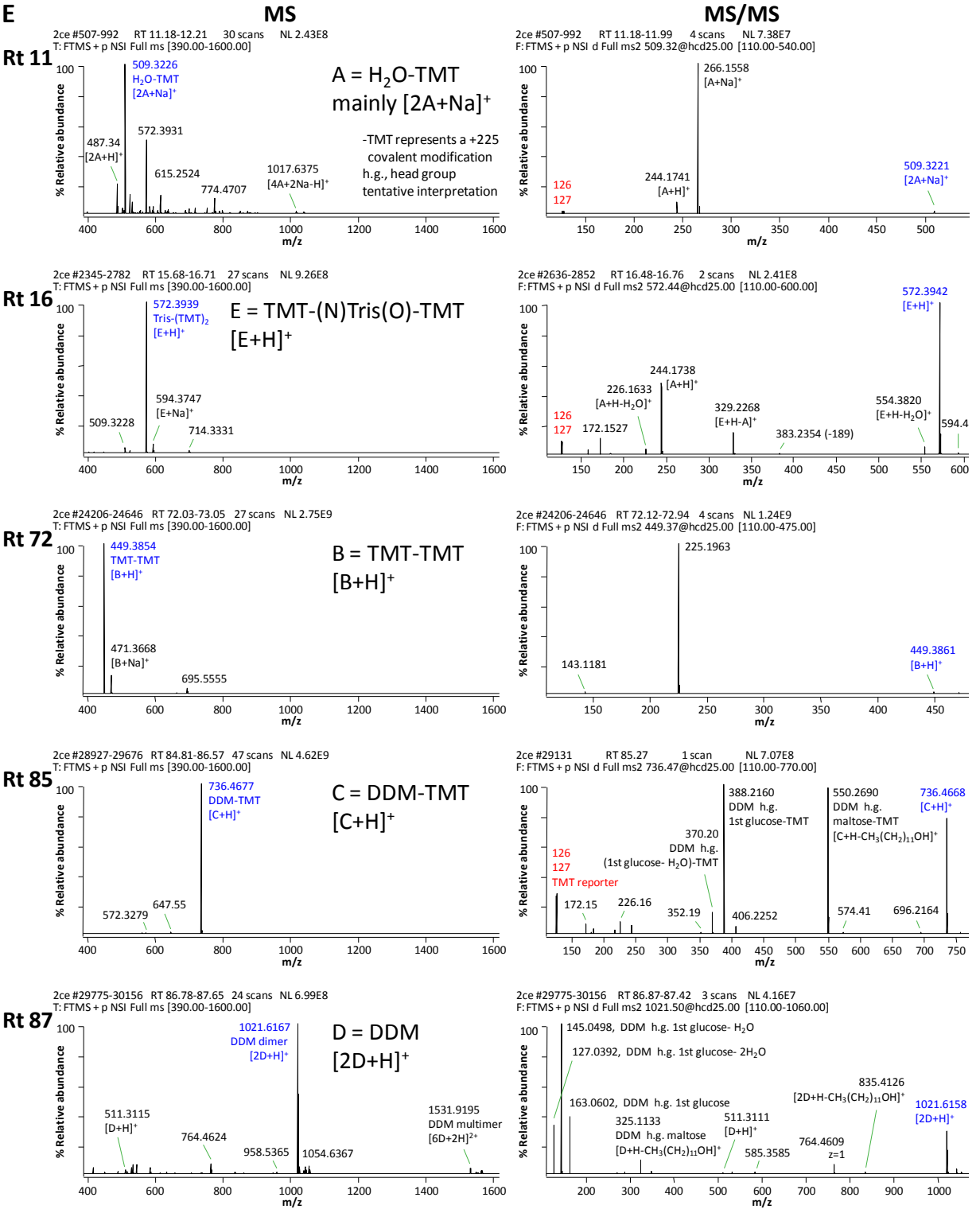
Supplemental Figure S1. PSM distribution against the sequences of $GABA_AR$ subunits (**ABC**) supplementing **Fig. 3A**, and HPLC MS chromatography in total ion counts from runs 4, 5 and 6 in **Fig. 2** (**D**). Peptides were identified at FDR<1% by PD1.3 SEQUEST using b and y ions, and showed high coverage and basal level of run-to-run peptide reproducibility.

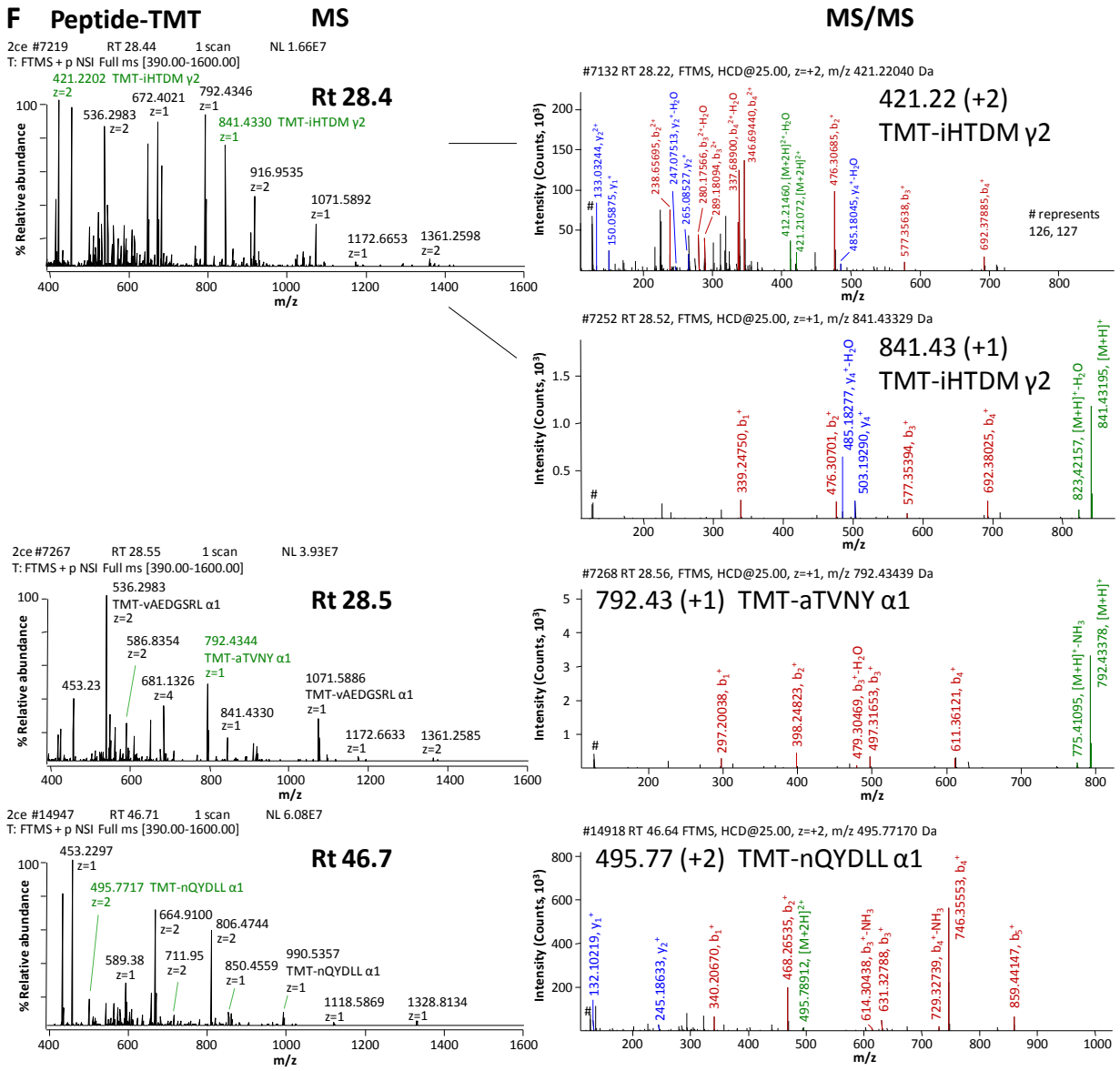
Supplemental Figure S2.




C
γ2
C-C loop
TM1
TM2
TM3
ICL2
TM4
1
51
101
151
201
251
301
351
401
451
495

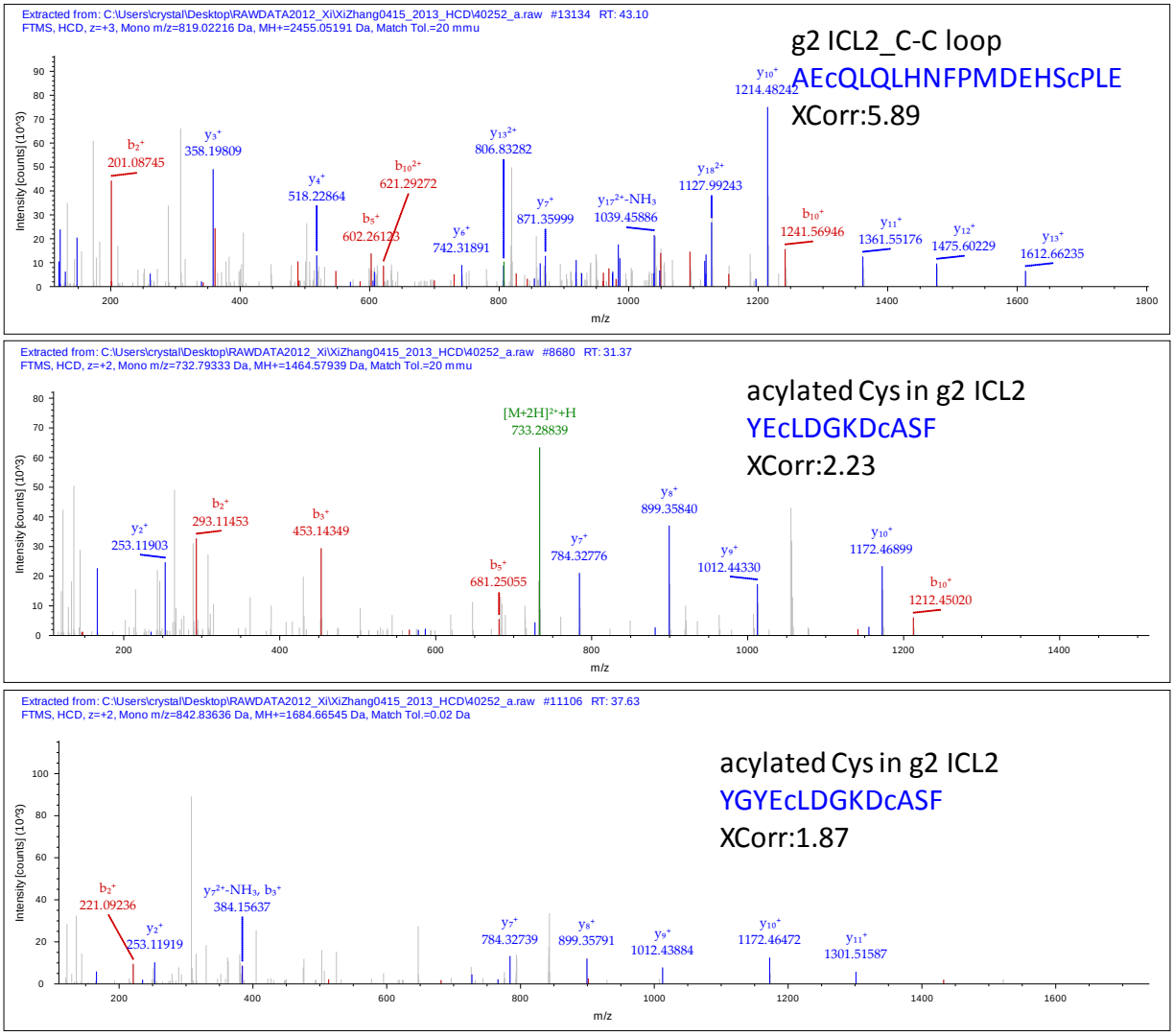

D
Tris-(TMT)2
[E+H]+
Mainly
DDM-TMT
[C+H]+
TMT-TMT
[B+H]+
Mainly
DDM
[2D+H]+
H2O-TMT
[2A+Na]+
Peptide-TMT
1.25E10
1.45E10
1.29E10
x3
100
% Relative abundance
10
20
30
40
50
60
70
80
90
Retention time (min)


E
MS
MS/MS
Rt 11
2ce #507-992 RT 11.18-12.21 30 scans NL 2.43E8
T: FTMS + p NSI Full ms [390.00-1600.00]
509.3226
H2O-TMT
[2A+Na]+
572.3931
487.34
[2A+H]+
615.2524
774.4707
1017.6375
[4A+2Na-H]+
A = H2O-TMT
mainly [2A+Na]+
-TMT represents a +225
covalent modification
h.g., head group
tentative interpretation
2ce #507-992 RT 11.18-11.99 4 scans NL 7.38E7
F: FTMS + p NSI d Full ms2 509.32@hcd25.00 [110.00-540.00]
266.1558
[A+Na]+
244.1741
[A+H]+
126
127
509.3221
[2A+Na]+
Rt 16
2ce #2345-2782 RT 15.68-16.71 27 scans NL 9.26E8
T: FTMS + p NSI Full ms [390.00-1600.00]
572.3939
Tris-(TMT)2
[E+H]+
E = TMT-(N)Tris(O)-TMT
[E+H]+
594.3747
[E+Na]+
509.3228
714.3331
2ce #2636-2852 RT 16.48-16.76 2 scans NL 2.41E8
F: FTMS + p NSI d Full ms2 572.44@hcd25.00 [110.00-600.00]
572.3942
[E+H]+
244.1738
[A+H]+
226.1633
[A+H-H2O]+
329.2268
[E+H-A]+
554.3820
[E+H-H2O]+
126
127
172.1527
383.2354 (-189)
594.4
Rt 72
2ce #24206-24646 RT 72.03-73.05 27 scans NL 2.75E9
T: FTMS + p NSI Full ms [390.00-1600.00]
449.3854
TMT-TMT
[B+H]+
B = TMT-TMT
[B+H]+
471.3668
[B+Na]+
695.5555
2ce #24206-24646 RT 72.12-72.94 4 scans NL 1.24E9
F: FTMS + p NSI d Full ms2 449.37@hcd25.00 [110.00-475.00]
225.1963
143.1181
449.3861
[B+H]+
Rt 85
2ce #28927-29676 RT 84.81-86.57 47 scans NL 4.62E9
T: FTMS + p NSI Full ms [390.00-1600.00]
736.4677
DDM-TMT
[C+H]+
C = DDM-TMT
[C+H]+
572.3279
647.55
2ce #29131 RT 85.27 1 scan NL 7.07E8
F: FTMS + p NSI d Full ms2 736.47@hcd25.00 [110.00-770.00]
388.2160
DDM h.g.
1st glucose-TMT
550.2690
DDM h.g.
maltose-TMT
[C+H-CH3(CH2)11OH]+
736.4668
[C+H]+
370.20
DDM h.g.
(1st glucose- H2O)-TMT
126
127
TMT reporter
172.15
226.16
352.19
406.2252
574.41
696.2164
Rt 87
2ce #29775-30156 RT 86.78-87.65 24 scans NL 6.99E8
T: FTMS + p NSI Full ms [390.00-1600.00]
1021.6167
DDM dimer
[2D+H]+
D = DDM
[2D+H]+
511.3115
[D+H]+
764.4624
958.5365
1054.6367
1531.9195
DDM multimer
[6D+2H]2+
2ce #29775-30156 RT 86.87-87.42 3 scans NL 4.16E7
F: FTMS + p NSI d Full ms2 1021.50@hcd25.00 [110.00-1060.00]
145.0498, DDM h.g. 1st glucose- H2O
127.0392, DDM h.g. 1st glucose- 2H2O
835.4126
[2D+H-CH3(CH2)11OH]+
163.0602, DDM h.g. 1st glucose
1021.6158
[2D+H]+
325.1133
DDM h.g. maltose
[D+H-CH3(CH2)11OH]+
511.3111
[D+H]+
585.3585
764.4609
z=1
% Relative abundance
m/z



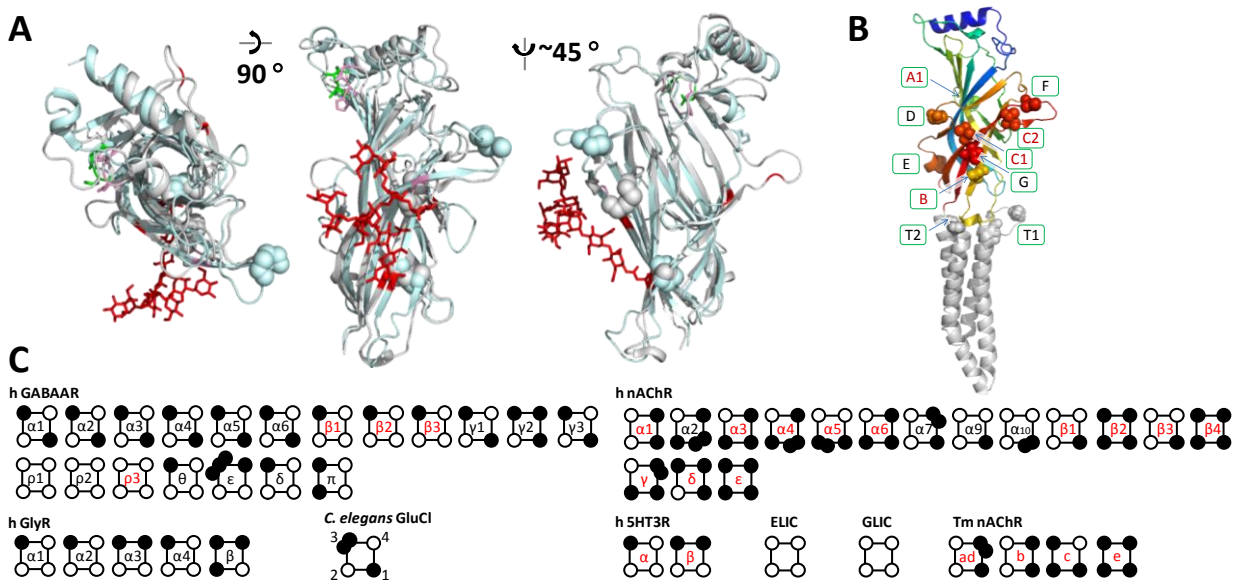
Supplemental Figure S2. Ox-TMT2 PSM distribution against the sequences of $GABA_AR$ subunits (**ABC**), supplementing **Fig. 3G**, HPLC MS chromatography in total ion counts from three Ox-TMT2 samples that combined multiple independent digestions, showing high peptide reproducibility (**D**), and representative MS and HCD MS/MS spectra for non-peptide components (**E**) and for typical TMT-peptides (**F**) at distinct elution ranges, demonstrating method tolerance. Peptides were identified at peptide FDR<1% by PD1.3 SEQUEST using b and y ions. Precursor ions were marked in blue (**E**) or green (**F**). Rt, retention time.

Supplemental Figure S3.



Supplemental Figure S3. HCD MS/MS spectra for selected Cys-PTM identifications (peptide FDR<1% by PD1.3 SEQUEST) discussed in **Fig. 2**. More annotated MS/MS spectra for tentative PTM identifications are shown in **supplemental Data**.

Supplemental Figure S4.



Supplemental Figure S4. Bases for Cys-loop hLGIC PTM signature modeling, supplementing **Fig. 5**. (**A**) 3D locations of N-glycans were estimated by sequence alignment with resolved structures of both C. elegans GluCl/Sf9 (3RHW, silver) and mouse α1 AChR-ECD/yeast (2QC1, cyan). Red cartoon, residues corresponding to high-occupancy N-glycosylated sites identified in α1β3γ2 hGABA$_A$R. (**B**) Locations (sphere) of 8 mid-ECD N-glycosylation coding sites and two potential TMD-originated ECD-TMD interface sites, based on sequence alignment with 3RHW. ECD sequence was colored from N- (blue) to C-terminus (red). (**C**) TM-Cys are common to eukaryotic Cys-loop LGICs, and vary by subunits in distribution among 4 TM helices, suggesting a third layer of subunit signature—again missing in bacteria. All three TM-Cys residues in *C. elegans* GluCl were resolved as free in ivermectin-opened state (3RHW)(8), but traditional mutagenesis and Cys-crosslinking found inter-helix C-C bridges in rat α1β1γ2S GABA$_A$R in closed but not in GABA-bound state*. *Ref: Jansen, M. & Akabas, M.H. *J Neurosci* **26**, 4492-4499 (2006).

Supplemental Figure S5.

A ECD domain

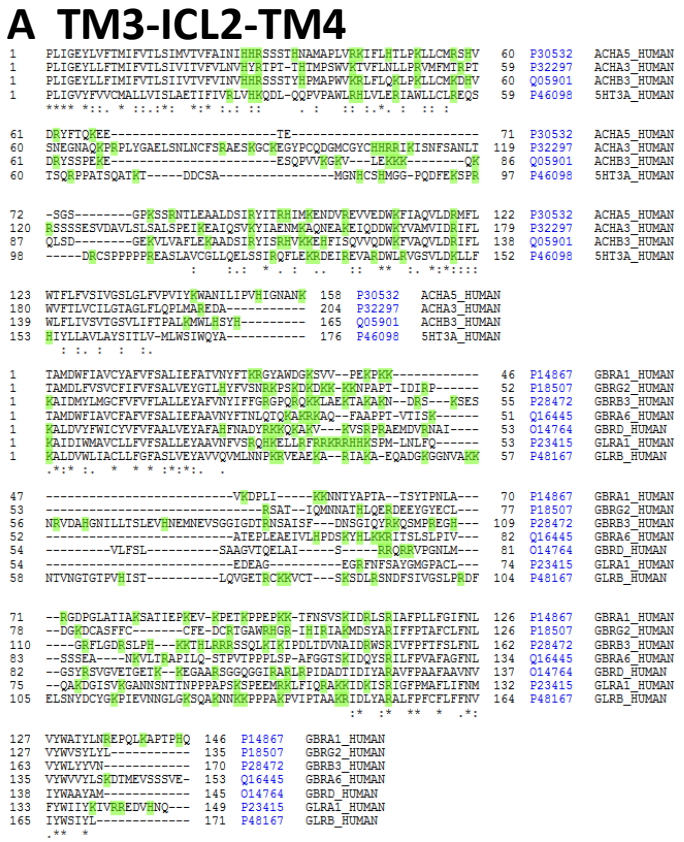
```
1    ------------------------------------------------------------    0  2QC1:B|PDBID|CHAIN|SEQUENCE
1    ------------------------------------------------------------    0  3RHW:A|PDBID|CHAIN|SEQUENCE
1    -------------MRKSPGLSDCLWAWILLLSTLT--GRSYGQPSLQD------------   33  P14867   GBRA1_HUMAN
1    MCSGL--------LEL---LLPIWLSWTLGTRG------S-EPRSVNDPGNMS-------   35  P28472-2 GBRB3_HUMAN
1    -MSSPNIWSTGSSVYSTPVFSQKMTVWILLLLSLY--PGFTSQKSDDDYEDYASNKTWVL   57  P18507-2 GBRG2_HUMAN
1    -MDAP--------ARL---LAPLLLLCAQQLRG------T---RAMNDIGDYVGS-----   34  O14764   GBRD_HUMAN
1    -MLAV--------PNMRFGIFLLWWGWVLATESRMHWPGR-EVHEMSKKGRPQRQRREVH   50  P24046   GBRR1_HUMAN


1    ------------------KSEHETRLEAKLFEDYSSVVRPVEDHREIVQVTVGLQLIQLI   42  2QC1:B|PDBID|CHAIN|SEQUENCE
1    ------------------SD--SKILAHLFTSGYDFRVRPPTDNGGPVVVSVNMLLRTIS   40  3RHW:A|PDBID|CHAIN|SEQUENCE
34   ----------ELK--DNTTVFTRILD--RLLDGYDNRLRPGL-GERVTEVKTDIFVTSFG   78  P14867   GBRA1_HUMAN
36   -------------------FVKETVD--KLLKGYDIRLRPDF-GGPPVCVGMNIDIASID   73  P28472-2 GBRB3_HUMAN
58   ----------TPK--VPEGDVTVILN--NLLEGYDNKLRPDI-GVKPTLIHTDMYVNSIG  102  P18507-2 GBRG2_HUMAN
35   ---------------NLEISWLPNLD--GLIAGYARNFRPGI-GGPPVNVALALEVASID   76  O14764   GBRD_HUMAN
51   EDAHKQVSPILRRSPDITKSPLTKSEQLLRIDDHDFSMRPGF-GGPAIPVGVDVQVESLD  109  P24046   GBRR1_HUMAN
                                     :   .**         :   : :  :


43   NVDEVNQIVTTNVRLKQQWVDYNLKWNPDDYGGVKKI-HIPSEKIWRPDVVLYNNADGDF  101  2QC1:B|PDBID|CHAIN|SEQUENCE
41   KIDVVNMEYSAQLTLRESWIDKRLSYGVKGDGQPDFVILTVGHQIWMPDTFFPNEKQAYK  100  3RHW:A|PDBID|CHAIN|SEQUENCE
79   PVSDHDMEYTIDVFFRQSWKDERLKFKGP-MTVLRLN-NLMASKIWTPDTFFHNGKKSVA  136  P14867   GBRA1_HUMAN
74   MVSEVNMDYTLTMYFQQYWRDKRLAYSGI-PLNLTLD-NRVADQLWVPDTYFLNDKKSFV  131  P28472-2 GBRB3_HUMAN
103  PVNAINMEYTIDIFFAQTWYDRRLKFNST-IKVLRLN-SNMVGKIWIPDTFFRNSKKADA  160  P18507-2 GBRG2_HUMAN
77   HISEANMEYTMTVFLHQSWRDSRLSYNHT-NETLGLD-SRFVDKLWLPDTFIVNAKSAWF  134  O14764   GBRD_HUMAN
110  SISEVDMDFTMTLYLRHYWKDERLSFPSTNNLSMTFD-GRLVKKIWVPDMFFVHSKRSFI  168  P24046   GBRR1_HUMAN
      :.  :   :  : : . * * .* :                ::* **  : .   .


102  AIVKFTKVLL--DYTGHITWTPPAIFKSYCEIIVTHFPFDEQNCSMKLGTRTYDGSAVAI  159  2QC1:B|PDBID|CHAIN|SEQUENCE
101  HTIDKPNVLIRIHNDGTVLYSVRISLVLSCPMYLQYYPMDVQQCSIDLASYAYTTKDIEY  160  3RHW:A|PDBID|CHAIN|SEQUENCE
137  HNMTMPNKLLRITEDGTLLYTMRLTVRAECPMHLEDFPMDAHACPLKFGSYAYTRAEVVY  196  P14867   GBRA1_HUMAN
132  HGVTVKNRMIRLHPDGTVLYGLRITTTAACMMDLRRYPLDEQNCTLEIESYGYTTDDIEF  191  P28472-2 GBRB3_HUMAN
161  HWITTPNRMLRIWNDGRVLYTLRLTIDAECQLQLHNFPMDEHSCPLEFSSYGYPREEIVY  220  P18507-2 GBRG2_HUMAN
135  HDVTVENKLIRLQPDGVILYSIRITSTVACDMDLAKYPMDEQECMLDLESYGYSSEDIVY  194  O14764   GBRD_HUMAN
169  HDTTTDNVMLRVQPDGKVLYSLRVTVTAMCNMDFSRFPLDTQTCSLEIESYAYTEDDLML  228  P24046   GBRR1_HUMAN
           : ::     * : :         * : .  :*:* : * :.: :  *    :


160  NPESDQPDLS----NFMESGEWVIKEARGWKHWVFYSCCPTTPYLDITYHFVMQRLP-  212  2QC1:B|PDBID|CHAIN|SEQUENCE
161  LWKEHSPLQLKV-GLSSSLPSFQLTNTSTTY-CTSVT--NTGIYSCLRTTIQLKREFS  214  3RHW:A|PDBID|CHAIN|SEQUENCE
197  EWTREPARSVVVAEDGSRLNQYDLLGQTVDSGIV-QS--STGEYVVMTTHFHLKRKIG  251  P14867   GBRA1_HUMAN
192  YWRGGDKAV--TGVERIELPQFSIVEHRLVSRNVVF---ATGAYPRLSLSFRLKRNIG  244  P28472-2 GBRB3_HUMAN
221  QWKRSSVEV--GDTRSWRLYQFSFVGLRNTTEVV-KT--TSGDYVVMSVYFDLSRRMG  273  P18507-2 GBRG2_HUMAN
195  YWSESQEHI--HGLDKLQLAQFTITSYRFTTELMNFK--SAGQFPRLSLHFHLRRNRG  248  O14764   GBRD_HUMAN
229  YWKKGNDSL--KTDERISLSQFLIQEFHTTTKLAFYS--STGWYNRLYINFTLRRHIF  282  P24046   GBRR1_HUMAN
```

B

receptor/channel	subunit	A1 = inner surface b4-b5	length	B = C-C loop (##.C-C.##)	length	C1/C2 = mid/upper-b9 pre-loopC (highly diverse)	length	A2 = b5-b6 linker before loopE (A2 NGT is back to top surface)	length
Cl_GBR_h	α1	KKSVAHNMTMPNKL	14	AECPMHLEDFPMDAHACPL	19	-----GQTVDSGIVQ	10	LRITEDGTLL	10
Cl_GBR_h	α2	KKSVAHNMTMPNKL	14	AECPMHLEDFPMDAHSCPL	19	-----GQSIGKETIK	10	LRIQDDGTLL	10
Cl_GBR_h	α3	KKSVAHNMTTPNKL	14	AECPMHLEDFPMDVHACPL	19	-----GHVVGTEIIR	10	LRLVDNGTLL	10
Cl_GBR_h	α4	KKSVSHNMTAPNKL	14	AECPMRLVDFPMDGHACPL	19	-----GQTVSSETIK	10	FRIMRNGTIL	10
Cl_GBR_h	α5	KKSIAHNMTTPNKL	14	AECPMQLEDFPMDAHACPL	19	-----GQTVGTENIS	10	LRLEDDGTLL	10
Cl_GBR_h	α6	KKSIAHNMTTPNKL	14	ADCPMRLVNFPMDGHACPL	19	-----GQTVSSETIK	10	FRIMQNGTIL	10
Cl_GBR_h	β1	KKSFVHGVTVKNRM	14	AACMMDLRRYPLDEQNCTL	19	DYKMVSKKVE-----	10	IRLHPDGTVL	10
Cl_GBR_h	β2	KKSFVHGVTVKNRM	14	AACMMDLRRYPLDEQNCTL	19	DYKLITKKVV-----	10	IRLHPDGTVL	10
Cl_GBR_h	β3	KKSFVHGVTVKNRM	14	AACMMDLRRYPLDEQNCTL	19	EHRLVSRNVV-----	10	IRLHPDGTVL	10
Cl_GBR_h	β3-isof2	KKSFVHGVTVKNRM	14	AACMMDLRRYPLDEQNCTL	19	EHRLVSRNVV-----	10	IRLHPDGTVL	10
Cl_GBR_h	γ1	RKSDAHWITTPNRL	14	AECYLQLHNFPMDEHSCPL	19	GLRNSTEITH-----	10	LRIWNDGRVL	10
Cl_GBR_h	γ2	KKADAHWITTPNRM	14	AECQLQLHNFPMDEHSCPL	19	GLRNTTEVVK-----	10	LRIWNDGRVL	10
Cl_GBR_h	γ2L	KKADAHWITTPNRM	14	AECQLQLHNFPMDEHSCPL	19	GLRNTTEVVK-----	10	LRIWNDGRVL	10
Cl_GBR_h	γ3	KTAEAHWITTPNQL	14	AECQLQLHNFPMDEHSCPL	19	GLRNTTEIVT-----	10	LRIWNDGKIL	10
Cl_GBR_h	ρ1	KRSFIHDTTTDNVM	14	AMCNMDFSRFPLDTQTCSL	19	EFHTTTKLAF-----	10	LRVQPDGKVL	10
Cl_GBR_h	ρ2	KRSFTHDTTTDNIM	14	AMCNMDFSHFPLDSQTCSL	19	KFHTTSRLAF-----	10	LRVFPDGHVL	10
Cl_GBR_h	ρ3	KRSFIHDTTMENIM	14	AMCFMDFSRFPLDTQNCSL	19	DFSASSGLAF-----	10	LRVHPDGNVL	10
Cl_GBR_h	δ	KSAWFHDVTVENKL	14	VACDMDLAKYPMDEQECML	19	SYRFTTELMN-----	10	IRLQPDGVIL	10
Cl_GBR_h	ε	KRTHEHEITMPNQM	14	AGCSLHMLRFPMDSHSCPL	19	GVSNKTEIIT-----	10	VRIYKDGKVL	10
Cl_GBR_h	θ	KDAFVHDVTVENRV	14	AACSLDLHKFPMDKQACNL	19	-----GRTITSKEVY	10	FQLHPDGTVR	10
Cl_GBR_h	π	KKSFLHEVTVGNRL	14	VACNMDLSKYPMDTQTCKL	19	RYFTLVTRSQ	10	IRLFSNGTVL	10
Cl_GluCl_Celeg	homomer	KQAYKHTIDKPNVL	14	LSCPMYLQYYPMDVQQCSI	19	-NTSTTYCTSV	10	IRIHNDGTVL	10
Cl_GlyR_h	α1	KGAHFHEITTDNKL	14	LACPMDLKNFPMDVQTCIM	19	EEKDLRYCTK	10	LRISRNGNVL	10
Cl_GlyR_h	α2	KGANFHDVTTDNKL	14	LSCPMDLKNFPMDVQTCTM	19	EEKELGYCTK	10	LRISKNGKVL	10
Cl_GlyR_h	α3	KGANFHEVTTDNKL	14	LSCPMDLKNFPMDVQTCIM	19	EEKDLRYCTK	10	LRIFKNGNVL	10
Cl_GlyR_h	α4	KGANFHEVTTDNKL	14	LSCLMDLKNFPMDIQTCTM	19	DEKDLGCCTK	10	LRIFKNGNVL	10
Cl_GlyR_h	β	KSANFHDVTQENIL	14	LSCPLDLTLFPMDTQRCKM	19	KEDIEYGNCTK	11	LFIFRDGDVL	10
NaKCa_5HT3R_h	α	FVDVGKSPNIPY--	12(--)	TACSLDIYNFPFDVQNCSL	19	GVLPYFREFS	10	VYIRHQGEVQ	10
NaKCa_5HT3R_h	β	FVDIERYPDLPY--	12(--)	SACSLETYAFPFDVQNCSL	19	SVSSTYSILQ	10	VYVNSSGTIE	10
NaKCa_nAChR_h	α1	ADGDFAIVKFTK--	12(--)	SYCEIIVTHFPFDEQNCSM	19	ESRGWKHSVT	10	VLLQYTGHIT	10
NaKCa_nAChR_h	α2	ADGEFAVTHMTK--	12(--)	SSCSIDVTFFPFDQQNCKM	19	NATGTYNSKK	10	AHLFSTGTVH	10
NaKCa_nAChR_h	α3	AVGDFQVDDKTK--	12(--)	SSCKIDVTYFPFDYQNCTM	19	KAPGYKHDIK	10	ALLKYTGEVT	10
NaKCa_nAChR_h	α4	ADGDFAVTHLTK--	12(--)	SSCSIDVTFFPFDQQNCTM	19	DAVGTYNTRK	10	AHLFHDGRVQ	10
NaKCa_nAChR_h	α5	ADGRFEGTS-TK--	12(--)	SSCTIDVTFFPFDLQNCSM	19	SATGSKGNRT	10	TVIRYNGTVT	10
NaKCa_nAChR_h	α6	AVGDFQVEGKTK--	12(--)	SSCPMDITFFPFDHQNCSL	19	DASGYKHDIK	10	ALLKYNGMIT	10
NaKCa_nAChR_h	α7	ADERFDATFHTN--	12(--)	SSCYIDVRWFPFDVQHCKL	19	GIPGKRSERF	10	VLVNSSGHCQ	10
NaKCa_nAChR_h	α9	ADDESSEPVNTN--	12(--)	SSCVVDVTYFPFDNQQCNL	19	GMPAVKNVIS	10	VVLRYDGLIT	10
NaKCa_nAChR_h	α10	ADAQPPGSASTN--	12(--)	SSCRVDVAAFPFDAQHCGL	19	GMPARRRVLT	10	VVLRHDGAVR	10
NaKCa_nAChR_h	β1	NDGNFDVALDIS--	12(--)	SSCSIQVTYFPFDWQNCTM	19	HKPSRLIQPP	10	VVVSSDGSVR	10
NaKCa_nAChR_h	β2	ADGMYEVSFYSN--	12(--)	SACKIEVKHFPFDQQNCTM	19	ALPGRRNENP	10	AVVSYDGSIF	10
NaKCa_nAChR_h	β3	ADGRFEGSLMTK--	12(--)	SSCTMDVTFFPFDRQNCSM	19	NAKGMKGNRR	10	VIVKSNGTVV	10
NaKCa_nAChR_h	β4	ADGTYEVSVYTN--	12(--)	SACKIEVKYFPFDQQNCTL	19	ALPGRRTVNP	10	LIVRSNGSVL	10
NaKCa_nAChR_h	γ	VDGVFEVALYCN--	12(--)	SACSISVTYFPFDWQNCSL	19	HRPAKMLLDP	10	VLVSPDGCIY	10
NaKCa_nAChR_h	δ	NDGSFQISYSCN--	12(--)	SSCPISVTYFPFDWQNCSL	19	HRPARVNVDP	10	VLVYHYGFVY	10
NaKCa_nAChR_h	ε	IDGQFGVAYDAN--	12(--)	SVCAVEVTYFPFDWQNCSL	19	FCPGVIRRHH	10	VLVYEGGSVT	10
NaKCa_nAChR_Tm	a=d	ADGDFAIVHMTK--	12(--)	SYCEIIVTHFPFDQQNCTM	19	DYRGWKHWVY	10	LLLDYTGKIM	10
NaKCa_nAChR_Tm	b	NDGSFEITLHVN--	12(--)	SSCTIKVMYFPFDWQNCTM	19	HKPSRKNWRS	10	VLVQHTGAVS	10
NaKCa_nAChR_Tm	c	NDGQYNVAYFCN--	12(--)	SSCPINVLYFPFDWQNCSL	19	HKPAKKNIYG	10	VLVRPNGYVT	10
NaKCa_nAChR_Tm	e	VDGQFEVAYYAN--	12(--)	STCPIAVTYFPFDWQNCSL	19	HRPAKKNYNW	10	VLVYNDGSMY	10
NaKCa_nAChBP_Ac	AChBP	TRPVQV---LSPQI	11(---)	FMCDPTGV-DSEEGVTCAV	18	SATQTRQVQH	10	AVVTHDGSVM	10
NaKCa_nAChR_m	a1_AChBP c	ADGDFAIVKFTK--	12(--)	SYCEIIVTHFPFDEQNCSM	19	EARGWKHWVF	10	VLLDYTGHIT	10
NaKCa_ELIC_xtl	homomer	VGSPD----TGNKR	10(----)	FSNDMDFRLFPFDRQQFVL	19	KASTHISDIR	10	LMLFPDGRVI	10
NaKCa_GLIC_xtl	homomer	ENARD----ADVVD	10(----)	VLSPLDFRRYPFDSQTLHI	19	SFTAVVKPAN	10	ISVSPDGTVQ	10

Supplemental Figure S5. ECD sequence alignment of hGABA$_A$R (GBR) subunits with 3RHW and 2QC1 (**A**), and detailed sequences of the four major coding regions (**A1, B, C1/C2**) pinpointed by screening for mid-ECD NXS/T motif across Cys-loop hLGICs (**B**), supplementing **Fig. 5A** and **5E**.

Supplemental Figure S6.



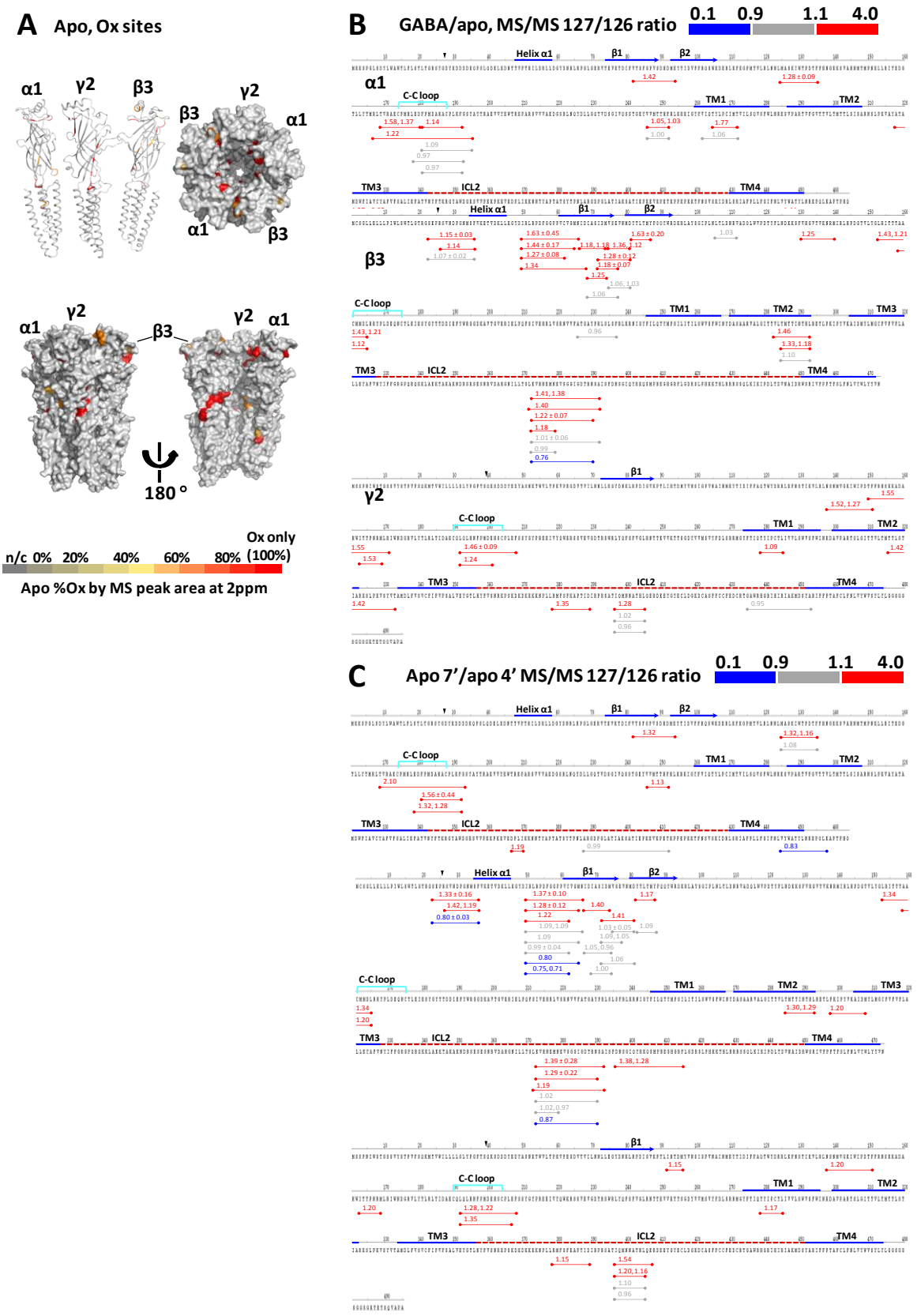
B

neuronal cation channels TM3-ICL2-TM4	estimated pI	pH 7.5 charge	KR-DE	sum DE	sum KR
5HT3R α	7.8	**0.7**	1	17	18
nAChR α3	8.0	**1.5**	2	18	20
nAChR α5	9.3	**4.3**	4	13	17
nAChR β3 *	9.7	**9.3**	9	14	23
Cl⁻ channels					
GlyR α1	10.0	**10.9**	11	13	24
GlyR β	9.4	**9.3**	10	14	24
GABAAR α1	9.7	**8.7**	9	12	21
GABAAR β3	9.9	**10.1**	10	16	26
GABAAR δ	10.0	**9.7**	10	12	22
GABAAR γ2 †	8.2	**2.4**	3	15	18
GABAAR α6 #	9.0	**2.8**	3	11	14
* nAChR β3_1st half	9.5	**5**	5	7	12
* nAChR β3_2nd half	9.6	**4.1**	4	7	11
† GABAAR γ2_1st half	9.7	**4.8**	5	6	11
† GABAAR γ2_2nd half	5.8	**-2.6**	-2	9	7
# GABAAR α6_1st half	9.2	**2.8**	3	6	9
# GABAAR α6_2nd half	6.6	**-0.3**	0	5	5

Sequence pI and charge state at pH 7.5 were estimated by using program at http://www.scripps.edu/cgi-bin/cdputnam/protcalc3

Supplemental Figure S6. Sequence alignments of the ICL2 of human Cl^- and cation channels showed high diversity (sequences showing TM3-ICL2-TM4) (**A**), and counting of K/R and D/E residues in TM3-ICL2-TM4 highlighted their differences in ICL2 K/R and D/E distribution (**B**), supplementing **Fig. 5A**. The subunits' TM3 and TM4 helices each contains only one or zero of each of the K, R, D and E residues.

Supplemental Figure S7.



Supplemental Figure S7. Applicability to direct Ox-TMT structural mapping of $GABA_AR$. (**A**) H_2O_2-oxidized residues pinpointed in apo $GABA_AR$ (labeled by H_2O_2 for 7 and 4 min, combined), marked on 3D model (3RHW). Oxidation occupancy was estimated by MS peak area (PD1.4). (**B**) Tentative MS/MS 127/126 ratios in GABA-bound/apo $GABA_AR$ (each H_2O_2-labeled for 7 min) mapped to protein sequence, showing only H_2O_2-oxidized peptides. (**C**) MS/MS 127/126 ratios for apo 7'/apo 4' $GABA_AR$. Full list of GABA/apo MS/MS 127/126 peptides was shown in **supplemental Table S7**.

Supplemental Note S1. Calculation of column-residence time.

We defined enzyme-protein overall contact time as the residence time of protein solution in the enzyme column operated as a plug-flow reactor (**Fig. 1C**).

POROS beads (enzyme-immobilized): spheres, diameter d = 20 μm.

Cylindrical cartridge: internal diameter D = 1 mm (50 fold as large as bead diameter), length L = 20 mm, so empty cartridge capacity C is,

$C = L (\pi D^2 / 4) = 20 (3.14 \times 1 / 4) = 15.7\ mm^3 = 15.7\ \mu L$

Assuming porosity of packed column ρ = 26 % (cylindrical cartridge tightly packed with small spheres at size ratio 50 : 1), the volume for solution in packed column v is,

$v = \rho C = \rho L(\pi D^2 / 4) = 26\% \times 20 (3.14 \times 1 / 4) = 26\% \times 15.7 = 4.1\ mm^3 = 4.1\ \mu L.$

When flow rate through column u1 is 25 μL/min, time of column residence for protein t1 is,

$t1 = v / u1 = 4.1\ \mu L / (25\ \mu L/min) = 4.1\ \mu L / (0.42\ \mu L/s) = 9.8\ s$, sub-10 seconds.

When flow rate through column u2 is 50 μL/min, time of column residence for protein t2 is,

$t2 = v / u2 = 4.1\ \mu L / (50\ \mu L/min) = 4.1\ \mu L / (0.82\ \mu L/s) = 4.9\ s.$

For the 2 mm i.d. x 20 mm PNGase F column applied at flow rate (u°) 2 μL/min,

$v° = \rho C° = \rho L(\pi D°^2 / 4) = 26\% \times 20 (3.14 \times 2^2 / 4) = 26\% \times 62.8 = 16.3\ mm^3 = 16.3\ \mu L.$

$t° = v° / u° = 16.3\ \mu L / (2\ \mu L/min) = 16.3\ \mu L / (0.03\ \mu L/s) = 490.0\ s = 8.2\ min.$

Supplemental Note S2. A catalytic site-occupancy model for digestion to overcome bias and enhance efficiency.

We defined protein abundance biases on digestion at two levels: **Type I**) $(\sum \mathrm{H_{pept}} / \sum \mathrm{L_{pept}}) \neq (\mathrm{H_{prot}}/\mathrm{L_{prot}})$, the abundance ratio of total peptide products (all forms) for the high (H)- and low (L)-abundance species doesn't maintain the ratio of their original protein substrates, and **Type II**) $(\mathrm{H^{i}_{pept}}/\mathrm{H_{prot}}) \neq (\mathrm{L^{i}_{pept}}/\mathrm{L_{prot}})$, the transmission of a given peptide (harboring residue of interest i, i-peptide) from its original protein substrate varies with protein's relative abundance (fraction of total proteins), meaning when the protein's relative abundance changes between samples, the distribution of the form and quantity of its peptide products also varies. Type I bias concerns quantitation by PSMs, and type II bias may skew peptide-centric quantitation by peak area or height. Further bias from protease preference sites can be explained by reaction kinetics.

Digestion model assuming completion of reactions

To address both types of protein abundance biases, we proposed a catalytic site-occupancy model to describe protein digestion in pepsin column **(Fig. 1D)**. We started with a simplified model assuming each reaction reaches completion, and considered kinetics afterwards. We defined digestion as effective catalytic site occupation: an accessible cleavage site in a protein or peptide substrate occupies an effective catalytic site on pepsin column. For the tightly packed pepsin column (20 µm diameter spherical beads in 1 mm i.d. x20 mm column, 25 µL/min), we assumed the protein solution flows though the column in an ideal plug-flow mode (**Fig. 1C**), where each starting species travels the same length (column length) at the same flow

rate. As a slice of protein solution moves along the column, proteins undergo sequential micro-digestion events, and meet a fresh slice of pepsin sites for each new event.

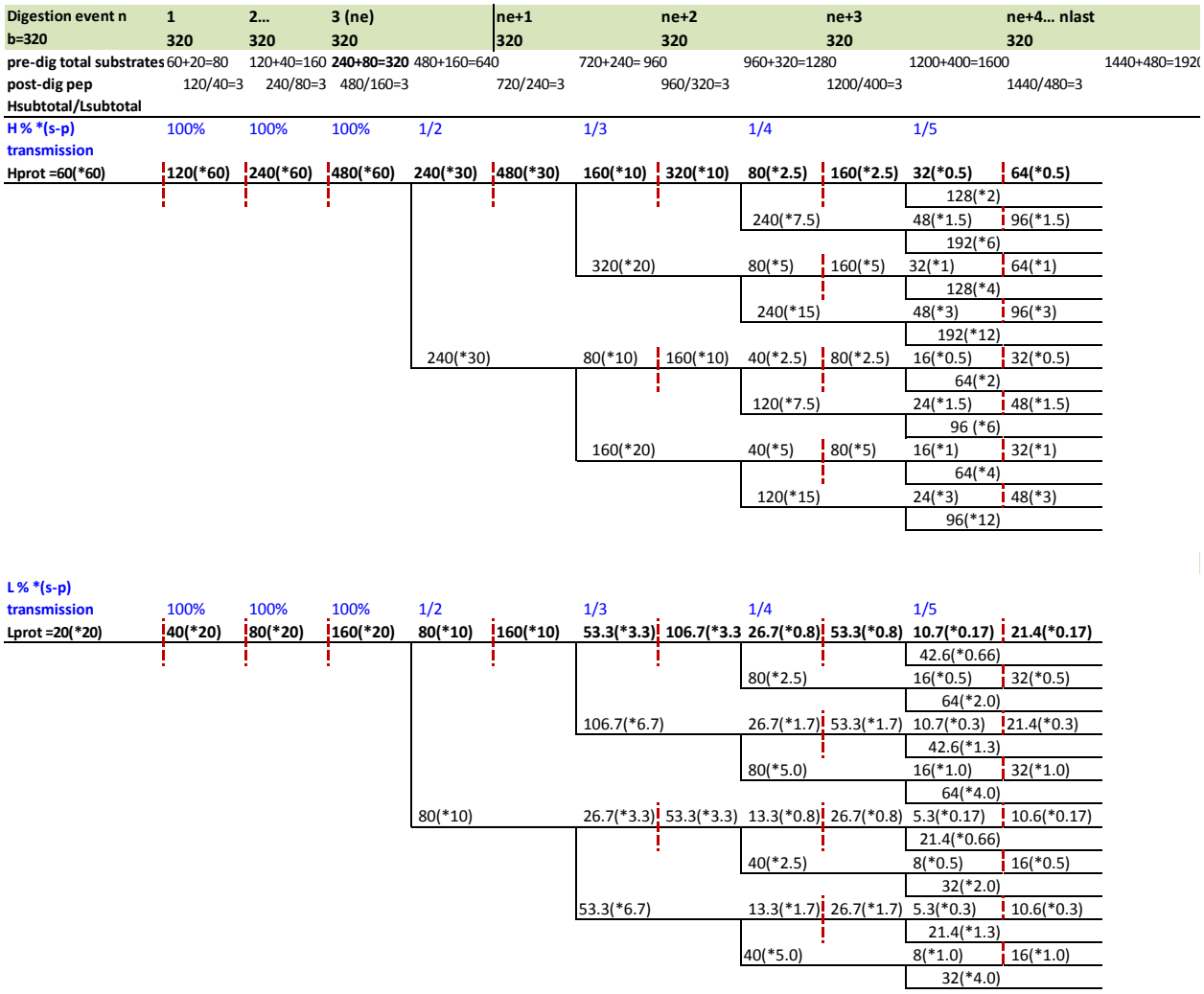


Figure SN1. Product formation in the catalytic site-occupancy model for digestion to understand biases, assuming instant completion of each reaction. For illustration, total effective enzyme catalytic site surface concentration (b) is 320, and starting protein concentration is 60 and 20 for high(H)- and low(L)-abundance species respectively. Red dash line, digestion. n_e, n_{equal}, the theoretical digestion event number when total substrates (a) equal b, here $n_e = 3$. *, protein or peptide containing a given residue of interest. S-P, substrate to peptide.

The starting model assumed: 1) Reaction rates of site adsorption/desorption >> bond cleavage >> site residence time (~10 s), thus each digestion event reaches completion, and site occupancy determines digestion occurrence. 2) A substrate prefers taking an empty catalytic site to competing for an occupied one. 3) Only up to one site per substrate can get bound and cleaved per event; each cleavage generates two different peptides, which can serve as substrates for the next digestion. 4) All starting and intermediate substrates can be treated equal. 5) Flow and DDM facilitation minimizes effects of slow substrate/product diffusion.

Abundance refers to the number of molecules, or concentration in molarity.

H, high-abundance species; H_{prot}, amount of original protein; H_{pept}, amount of peptide products.

L, low-abundance species; L_{prot}, amount of original protein; L_{pept}, amount of peptide products.

S (a) the total accessible substrates (original protein or intermediate peptide substrates) before a given digestion event: S_H, for a high-abundance species (H); S_L, for a low-abundance species (L).

b, the total number of effective catalytic sites on enzymes in a given digestion event.

n, theoretical number of digestion events, n=1, 2, 3… $n_{last,}$ and n controls peptide product length by $aa_{pept} = \frac{aa_{prot}}{2^n}$.

n_e, n_{equal}, the digestion event when S(a) = b.

t, the transmission efficiency from substrate to peptide (S-P) harboring a given residue in a digestion event.

Type I bias is defined as $\frac{\sum \mathrm{H_{pept}}}{\sum \mathrm{L_{pept}}} \neq \frac{\mathrm{H_{prot}}}{\mathrm{L_{prot}}}$, no-bias is defined as $\frac{\sum \mathrm{H_{pept}}}{\sum \mathrm{L_{pept}}} = \frac{\mathrm{H_{prot}}}{\mathrm{L_{prot}}}$.

Type II bias for a given peptide form (i) is defined as $\frac{\mathrm{H}^{\mathrm{i}}_{\mathrm{pept}}}{\mathrm{H}_{\mathrm{prot}}} \neq \frac{\mathrm{L}^{\mathrm{i}}_{\mathrm{pept}}}{\mathrm{L}_{\mathrm{prot}}}$, no-bias is defined as $\frac{\mathrm{H}^{\mathrm{i}}_{\mathrm{pept}}}{\mathrm{H}_{\mathrm{prot}}} = \frac{\mathrm{L}^{\mathrm{i}}_{\mathrm{pept}}}{\mathrm{L}_{\mathrm{prot}}}$.

1) When a ≤ b ($n \leq n_e$), digestion is not limited by b, and each substrate molecule has a 100% chance to occupy an available catalytic site to complete the digestion event.

After n digestion events, the total numbers of peptide products for H and L species are,

$$\sum \mathrm{H}_{\mathrm{pept}} = \mathrm{H}_{\mathrm{prot}} \, (2 \times 100\%)^n = \mathrm{H}_{\mathrm{prot}} \, 2^n \quad (1)$$

$$\sum \mathrm{L}_{\mathrm{pept}} = \mathrm{L}_{\mathrm{prot}} \, (2 \times 100\%)^n = \mathrm{L}_{\mathrm{prot}} \, 2^n \quad (2)$$

Peptide product abundance ratio $\frac{\sum \mathrm{H}_{\mathrm{pept}}}{\sum \mathrm{L}_{\mathrm{pept}}}$ is,

$$\frac{\sum \mathrm{H}_{\mathrm{pept}}}{\sum \mathrm{L}_{\mathrm{pept}}} = \frac{\mathrm{H}_{\mathrm{prot}} \, 2^n}{\mathrm{L}_{\mathrm{prot}} \, 2^n} = \frac{\mathrm{H}_{\mathrm{prot}}}{\mathrm{L}_{\mathrm{prot}}} \quad (3)$$

Thus there is no type I bias over abundance.

After n digestion events, for a given n-level i-peptide in the products, each preceding intermediate was generated at 100% transmission efficiency, thus each one starting protein generates one copy of i-peptide, regardless of H or L.

$$\mathrm{H}^{\mathrm{i}}_{\mathrm{pept}} = \mathrm{H}_{\mathrm{prot}} \times 100\%^{\mathrm{n}} = \mathrm{H}_{\mathrm{prot}} \quad (4)$$

$$\frac{\mathrm{H}^{\mathrm{i}}_{\mathrm{pept}}}{\mathrm{H}_{\mathrm{prot}}} = 1$$

Likewise, $\mathrm{L}^{\mathrm{i}}_{\mathrm{pept}} = \mathrm{L}_{\mathrm{prot}} \times 100\%^{\mathrm{n}} = \mathrm{L}_{\mathrm{prot}}$ (5)

$$\frac{\mathrm{H}^{\mathrm{i}}_{\mathrm{pept}}}{\mathrm{H}_{\mathrm{prot}}} = \frac{\mathrm{L}^{\mathrm{i}}_{\mathrm{pept}}}{\mathrm{L}_{\mathrm{prot}}} = 1,$$

Therefore, there is no type II bias over abundance.

Upon completion of n_e, $\mathrm{H}_{\mathrm{prot}} \, 2^{n_e} + \mathrm{H}_{\mathrm{prot}} \, 2^{n_e} = \mathrm{b}$ (6)

2) When a > b (n > n_e), digestion is limited by b—the total number of substrates that can get cleaved by digestion event n is only b. The chance for a species' substrates to occupy catalytic sites is less than 100%, but is proportional to its fraction in the total substrate abundance.

Prior to digestion n_e+1, substrates are n_e peptides with uniform length.

Total H or L substrates are: $\sum \mathrm{S_{H(n_e+1)}} = 2^{\mathrm{n_e}}\mathrm{H_{prot}}$, $\sum \mathrm{S_{L(n_e+1)}} = 2^{\mathrm{n_e}}\mathrm{L_{prot}}$

Total of all substrates is $\sum \mathrm{S_{n_e+1}} = 2^{\mathrm{n_e}}\mathrm{H_{prot}} + 2^{\mathrm{n_e}}\mathrm{L_{prot}}$ (7)

After digestion event n_e+1, b substrates are consumed, and 2b new products are formed. The distribution of these new products between H and L species is,

$$\frac{2b\sum \mathrm{S_{H(n_e+1)}}}{\sum \mathrm{S_{n_e+1}}} = \frac{2b2^{n_e}H_{prot}}{2^{n_e}(H_{prot}+L_{prot})} = \frac{2bH_{prot}}{H_{prot}+L_{prot}} \quad (8)$$

$$\frac{2b\sum \mathrm{S_{L(n_e+1)}}}{\sum \mathrm{S_{n_e+1}}} = \frac{2b2^{n_e}L_{prot}}{2^{n_e}(H_{prot}+L_{prot})} = \frac{2bL_{prot}}{H_{prot}+L_{prot}} \quad (9)$$

The total peptide products after digestion n_e+1, now containing mixed peptide lengths, are also the total substrates prior to the next n_e+2 digestion,

$$\sum \mathrm{H_{pept}} = 2^{\mathrm{n_e}}\mathrm{H_{prot}} + \frac{bH_{prot}}{H_{prot}+L_{prot}} = \frac{H_{prot}}{H_{prot}+L_{prot}}(2^{\mathrm{n_e}}\mathrm{H_{prot}} + 2^{\mathrm{n_e}}\mathrm{L_{prot}} + \mathrm{b}) = \sum \mathrm{S_{H(n_e+2)}} \quad (10)$$

$$\sum \mathrm{L_{pept}} = 2^{\mathrm{n_e}}\mathrm{L_{prot}} + \frac{bL_{prot}}{H_{prot}+L_{prot}} = \frac{L_{prot}}{H_{prot}+L_{prot}}(2^{\mathrm{n_e}}\mathrm{H_{prot}} + 2^{\mathrm{n_e}}\mathrm{L_{prot}} + \mathrm{b}) = \sum \mathrm{S_{L(n_e+2)}} \quad (11)$$

$$\sum \mathrm{H_{pept}} + \sum \mathrm{L_{pept}} = 2^{\mathrm{n_e}}\mathrm{H_{prot}} + 2^{\mathrm{n_e}}\mathrm{L_{prot}} + \mathrm{b} = \sum \mathrm{S_{n_e+2}} \quad (12)$$

The abundance ratio of total H or L peptide products is therefore,

$\frac{\sum \mathrm{H_{pept}}}{\sum \mathrm{L_{pept}}} = \frac{2^{\mathrm{n_e}}\mathrm{H_{prot}} + \frac{bH_{prot}}{H_{prot}+L_{prot}}}{2^{\mathrm{n_e}}\mathrm{L_{prot}} + \frac{bL_{prot}}{H_{prot}+L_{prot}}} = \frac{\mathrm{H_{prot}}}{\mathrm{L_{prot}}}$, thus no type I bias.

The amount of (n_e+1)-level peptides bearing site i is,

$$\mathrm{H_{pept}^{i}} = (\mathrm{H_{prot}} \times 100\%^{\mathrm{n_e}}) \times \frac{b}{\sum \mathrm{H_{pept}} + \sum \mathrm{L_{pept}}} = \frac{b\mathrm{H_{prot}}}{2^{\mathrm{n_e}}\mathrm{H_{prot}} + 2^{\mathrm{n_e}}\mathrm{L_{prot}} + \mathrm{b}} \quad (13)$$

$$\mathrm{L^{i}_{pept}} = \frac{b\mathrm{L_{prot}}}{2^{\mathrm{n_e}}\mathrm{H_{prot}}+2^{\mathrm{n_e}}\mathrm{L_{prot}}+\mathrm{b}} \quad (14)$$

Thus the abundance ratio of i-peptide to its original protein, after applying eq (6), is,

$$\frac{\mathrm{H^{i}_{pept}}}{\mathrm{H_{prot}}} = \frac{\frac{b\mathrm{H_{prot}}}{2^{\mathrm{n_e}}\mathrm{H_{prot}}+2^{\mathrm{n_e}}\mathrm{L_{prot}}+\mathrm{b}}}{\mathrm{H_{prot}}} = \frac{b}{2^{\mathrm{n_e}}\mathrm{H_{prot}}+2^{\mathrm{n_e}}\mathrm{L_{prot}}+\mathrm{b}} = \frac{1}{2} \quad (15)$$

$$\frac{\mathrm{L^{i}_{pept}}}{\mathrm{L_{prot}}} = \frac{\frac{b\mathrm{L_{prot}}}{2^{\mathrm{n_e}}\mathrm{H_{prot}}+2^{\mathrm{n_e}}\mathrm{L_{prot}}+\mathrm{b}}}{\mathrm{L_{prot}}} = \frac{1}{2} \quad (16)$$

Thus $\frac{\mathrm{H^{i}_{pept}}}{\mathrm{H_{prot}}} = \frac{\mathrm{L^{i}_{pept}}}{\mathrm{L_{prot}}} = \frac{1}{2}$, no type II bias when n = n_e+1.

This means that after n_e+1, the abundance of i-peptide with aimed length starts to under-represent its original protein abundance by a factor of 50%, regardless of protein relative abundance. Nonetheless, peptide-centric relative quantitation of a protein between cell states is not skewed by protein's abundance change.

Likewise, products can be calculated for digestion event n_e+2.

Before n_e+2, total substrates (containing mixed peptide lengths) are shown by eq (12),

$$\sum \mathrm{S_{n_e+2}} = 2^{\mathrm{n_e}}\mathrm{H_{prot}} + 2^{\mathrm{n_e}}\mathrm{L_{prot}} + \mathrm{b}$$

After digestion event n_e+2, b substrates are consumed, and 2b new products are formed. The distribution of these new products among H and L species is,

$$\frac{2b\sum \mathrm{S_{H(n_e+2)}}}{\sum \mathrm{S_{n_e+2}}} = 2b\frac{\frac{H_{prot}}{H_{prot}+L_{prot}}(2^{\mathrm{n_e}}\mathrm{H_{prot}}+2^{\mathrm{n_e}}\mathrm{L_{prot}}+\mathrm{b})}{2^{\mathrm{n_e}}\mathrm{H_{prot}}+2^{\mathrm{n_e}}\mathrm{L_{prot}}+\mathrm{b}} = \frac{2bH_{prot}}{H_{prot}+L_{prot}} \quad (17)$$

$$\frac{2b\sum \mathrm{S_{L(n_e+2)}}}{\sum \mathrm{S_{n_e+2}}} = 2b\frac{\frac{L_{prot}}{H_{prot}+L_{prot}}(2^{\mathrm{n_e}}\mathrm{H_{prot}}+2^{\mathrm{n_e}}\mathrm{L_{prot}}+\mathrm{b})}{2^{\mathrm{n_e}}\mathrm{H_{prot}}+2^{\mathrm{n_e}}\mathrm{L_{prot}}+\mathrm{b}} = \frac{2bL_{prot}}{H_{prot}+L_{prot}} \quad (18)$$

$$\sum \mathrm{H_{pept}} = \frac{H_{prot}}{H_{prot}+L_{prot}}\left(2^{\mathrm{n_e}}\mathrm{H_{prot}} + 2^{\mathrm{n_e}}\mathrm{L_{prot}} + \mathrm{b}\right) + \frac{bH_{prot}}{H_{prot}+L_{prot}}$$

$$= \frac{H_{prot}}{H_{prot}+L_{prot}}(2^{\mathrm{n_e}}\mathrm{H_{prot}} + 2^{\mathrm{n_e}}\mathrm{L_{prot}} + 2\mathrm{b}) = \sum \mathrm{S_{H(n_e+3)}} \quad (19)$$

$$\sum \mathrm{L_{pept}} = \frac{L_{prot}}{H_{prot}+L_{prot}}\left(2^{\mathrm{n_e}}\mathrm{H_{prot}} + 2^{\mathrm{n_e}}\mathrm{L_{prot}} + \mathrm{b}\right) + \frac{bL_{prot}}{H_{prot}+L_{prot}}$$

$$= \frac{L_{prot}}{H_{prot}+L_{prot}}\left(2^{\mathrm{n_e}}\mathrm{H_{prot}} + 2^{\mathrm{n_e}}\mathrm{L_{prot}} + 2\mathrm{b}\right) = \sum \mathrm{S_{L(n_e+3)}} \quad (20)$$

$$\sum \mathrm{H_{pept}} + \sum \mathrm{L_{pept}} = 2^{\mathrm{n_e}}\mathrm{H_{prot}} + 2^{\mathrm{n_e}}\mathrm{L_{prot}} + 2\mathrm{b} = \sum \mathrm{S_{n_e+3}} \quad (21)$$

The abundance ratio of total H or L peptide products is therefore,

$$\frac{\sum \mathrm{H_{pept}}}{\sum \mathrm{L_{pept}}} = \frac{\frac{H_{prot}}{H_{prot}+L_{prot}}\left(2^{\mathrm{n_e}}\mathrm{H_{prot}}+2^{\mathrm{n_e}}\mathrm{L_{prot}}+2\mathrm{b}\right)}{\frac{L_{prot}}{H_{prot}+L_{prot}}\left(2^{\mathrm{n_e}}\mathrm{H_{prot}}+2^{\mathrm{n_e}}\mathrm{L_{prot}}+2\mathrm{b}\right)} = \frac{\mathrm{H_{prot}}}{\mathrm{L_{prot}}}$$, thus no type I bias.

The amount of (n_e+2)-level i-peptide is,

$$\mathrm{H^i_{pept}} = \frac{b\mathrm{H_{prot}}}{2^{\mathrm{n_e}}\mathrm{H_{prot}}+2^{\mathrm{n_e}}\mathrm{L_{prot}}+\mathrm{b}} \times \frac{b}{2^{\mathrm{n_e}}\mathrm{H_{prot}}+2^{\mathrm{n_e}}\mathrm{L_{prot}}+2\mathrm{b}} \quad (22)$$

$$\mathrm{L^i_{pept}} = \frac{b\mathrm{L_{prot}}}{2^{\mathrm{n_e}}\mathrm{H_{prot}}+2^{\mathrm{n_e}}\mathrm{L_{prot}}+\mathrm{b}} \times \frac{b}{2^{\mathrm{n_e}}\mathrm{H_{prot}}+2^{\mathrm{n_e}}\mathrm{L_{prot}}+2\mathrm{b}} \quad (23)$$

Thus the abundance ratio of peptide i to its original protein, after applying eq (6), is,

$$\frac{\mathrm{H^i_{pept}}}{\mathrm{H_{prot}}} = \frac{b}{2^{\mathrm{n_e}}\mathrm{H_{prot}}+2^{\mathrm{n_e}}\mathrm{L_{prot}}+\mathrm{b}} \times \frac{b}{2^{\mathrm{n_e}}\mathrm{H_{prot}}+2^{\mathrm{n_e}}\mathrm{L_{prot}}+2\mathrm{b}} = \frac{1}{2}\times\frac{1}{3} = \frac{1}{6} \quad (24)$$

$$\frac{\mathrm{L^i_{pept}}}{\mathrm{L_{prot}}} = \frac{b}{2^{\mathrm{n_e}}\mathrm{H_{prot}}+2^{\mathrm{n_e}}\mathrm{L_{prot}}+\mathrm{b}} \times \frac{b}{2^{\mathrm{n_e}}\mathrm{H_{prot}}+2^{\mathrm{n_e}}\mathrm{L_{prot}}+2\mathrm{b}} = \frac{1}{6} \quad (25)$$

Thus $\frac{\mathrm{H^i_{pept}}}{\mathrm{H_{prot}}} = \frac{\mathrm{L^i_{pept}}}{\mathrm{L_{prot}}} = \frac{1}{6}$, no type II bias between H and L when n = n_e+2, despite under-representation of protein abundance by peptide within individual H or L species.

In summary, after n digestion events,

For type I bias, $\frac{\sum \mathrm{H_{pept}}}{\sum \mathrm{L_{pept}}} = \frac{\mathrm{H_{prot}}}{\mathrm{L_{prot}}}$ (26)

For type II bias,

When n ≤ n_e (a ≤ b),

$$\frac{\mathrm{H^i_{pept}}}{\mathrm{H_{prot}}} = \frac{\mathrm{L^i_{pept}}}{\mathrm{L_{prot}}} = 1^{\mathrm{n}} = 1$$

When $n > n_e$ ($a > b$),

$$\frac{H^i_{pept}}{H_{prot}} = \frac{L^i_{pept}}{L_{prot}} = 1^{n_e}\prod_{j=1}^{n-n_e}\frac{b}{2^{n_e}H_{prot}+2^{n_e}L_{prot}+jb} = \prod_{j=1}^{n-n_e}\frac{1}{j+1} < 1$$

Thus for type II bias, $$\frac{H^i_{pept}}{H_{prot}} = \frac{L^i_{pept}}{L_{prot}} = \begin{cases} 1 & n \le n_e \ (a \le b) \\ \prod_{j=1}^{n-n_e}\frac{1}{j+1} & n > n_e \ \ (a > b) \end{cases} \quad (27)$$

In conclusion, pepsin column doesn't incur type I or type II abundance-biases throughout the protein processing, according to this digestion model assuming fast completion of reactions. But peptide reproducibility and the proportion of peptides with useful/target length start to deteriorate and propagate rapidly after total substrates exceed catalytic sites ($n > n_e$, $a > b$). Thus operating digestion within the $n \le n_e$ region is desirable. The ratio of total protein amount—but not fractions for individual species—over total catalytic sites, determines whether a digestion experiment incurs low efficiency. Therefore, an optimal digestion method should provide effective catalytic sites that far exceed protein concentration, so that sufficient number of digestion events can occur before exceeding the n_e boundary.

Determination of n_e

We defined n_e as the theoretical digestion event number upon which a = b, the total of accessible substrate cleavage sites (a) equals that of effective catalytic sites available (b). n_e can be applied as an evaluation parameter for pepsin column operation. Upon the completion of n_e ($a \le b$), the total number of substrates has increased exponentially after each event for each species, therefore

$$a = 2^{n_e}H_{prot} + 2^{n_e}L_{prot} = 2^{n_e}(H_{prot} + L_{prot}) = b$$

$$n_e = ln\frac{b}{H_{prot}+L_{prot}}/ln2 = ln\frac{pepsin\ effective\ surface\ concentration}{total\ starting\ protien\ concentration}/0.693 \quad (28)$$

Thus n_e is unaffected by protein relative abundance, but can be controlled by adjusting the total protein concentration in samples—through dilution. In a pepsin column with effective surface concentration at over 1 mM, for a 1 μM protein solution,

$$n_e = ln\frac{1000\ \mu M}{1\ \mu M}/0.693 = 6.91/0.693 = 10$$

To determine how many theoretical digestion events are needed to obtain desired peptide length, we assumed each digestion cuts at the mid-point of a substrate: if a protein length is aa_{prot}, after m digestion events, the product peptide length aa_{pept} is,

$$aa_{pept} = \frac{aa_{prot}}{2^m}$$

Thus m can be determined by $m = ln\frac{aa_{prot}}{aa_{pept}}/0.693$ (29)

For a 600-residue protein, to obtain peptides 10-residue long, the number of digestion events needed is,

$$m = ln\frac{aa_{prot}}{aa_{pept}}/0.693 = ln\frac{600}{10}/0.693 = 5.9 < n_e$$

Therefore n_e far exceeds the number of digestions needed to get desired peptide length for most proteins, allowing pepsin column to operate well within the ideal $n < n_e$ region throughout protein processing.

Considering reaction kinetics

We then analyzed reaction kinetics by combining models of competitive-substrate Langmuir adsorption and classical enzymatic reaction, for each digestion event. We considered two protein species, H and L, that compete for digestion at surface catalytic sites (*). Given the fast diffusion of substrates and products in the plug-flow pepsin reactor, we assumed steady-state approximation in the reaction: fast equilibrium of adsorption and desorption, k_a, $k_d >> k_{cat}$, thus

surface concentration of ES rapidly reaches a steady level. Reaction rate constant k_{cat} includes both bond cleavage and product dissociation from the surface site. This model suggests that Type I abundance bias is not a concern, but pepsin preference can cause site/domain-bias in the rate to complete reactions.

$$E^* + S_H \overset{k_a^H,\, k_d^H}{\longleftrightarrow} ES_H^* \overset{k_{cat}^H}{\longrightarrow} 2P_H + E^*$$

$$E^* + S_L \overset{k_a^L,\, k_d^L}{\longleftrightarrow} ES_L^* \overset{k_{cat}^L}{\longrightarrow} 2P_L + E^*$$

$k_a, k_d \gg k_{cat}$, for both H and L

The total of surface coverage (Γ) for free, H and L substrate-occupied catalytic sites is constant b. The fraction occupancy of each species is represented by θ: $\theta_E = \Gamma_{E^*}/b$, and $\theta_{ES} = \Gamma_{ES^*}/b$.

$$\Gamma_{E^*} + \Gamma_{ES_H^*} + \Gamma_{ES_L^*} = b \quad (1)$$

$$\theta_E + \theta_{ES}^H + \theta_{ES}^L = 1 \quad (2)$$

By steady-state approximation,

$\frac{d\theta_{ES}^H}{dt} = 0$ and $\frac{d\theta_{ES}^L}{dt} = 0$

$$\frac{d\theta_{ES}^H}{dt} = k_a^H \theta_E [S_H] - k_d^H \theta_{ES}^H - k_{cat}^H \theta_{ES}^H = 0 \quad (3)$$

$$\frac{d\theta_{ES}^L}{dt} = k_a^L \theta_E [S_L] - k_d^L \theta_{ES}^L - k_{cat}^L \theta_{ES}^L = 0 \quad (4)$$

Solving eq (2)-(4) for θ_E leads to,

$$\theta_E = \frac{1}{1+\frac{k_a^H[S_H]}{k_d^H+k_{cat}^H}+\frac{k_a^L[S_L]}{k_d^L+k_{cat}^L}} = \frac{1}{1+\frac{[S_H]}{K_M^H}+\frac{[S_L]}{K_M^L}} = \frac{K_M^H K_M^L}{K_M^H K_M^L + K_M^L[S_H] + K_M^H[S_L]} \quad (5)$$

Where K_M^H and K_M^L are defined as,

$$K_M^H = \frac{k_d^H + k_{ca}^H}{k_a^H}, \qquad K_M^L = \frac{k_d^L + k_{cat}^L}{k_a^L} \quad (6)$$

Eq (3)-(5) lead to,

$$\theta_{ES}^{H} = \frac{\theta_E[S_H]}{K_M^H} = \frac{K_M^L[S_H]}{K_M^H K_M^L + K_M^L[S_H] + K_M^H[S_L]} \quad (7)$$

$$\theta_{ES}^{L} = \frac{\theta_E[S_L]}{K_M^L} = \frac{K_M^H[S_L]}{K_M^H K_M^L + K_M^L[S_H] + K_M^H[S_L]} \quad (8)$$

Product generation rates for H and L species are,

$$\frac{dP_H}{dt} = 2k_{cat}^H(\theta_{ES}^H b) = \frac{2bk_{cat}^H K_M^L[S_H]}{K_M^H K_M^L + K_M^L[S_H] + K_M^H[S_L]} \quad (9)$$

$$\frac{dP_L}{dt} = 2k_{cat}^L(\theta_{ES}^L b) = \frac{2bk_{cat}^L K_M^H[S_L]}{K_M^H K_M^L + K_M^L[S_H] + K_M^H[S_L]} \quad (10)$$

Therefore,

$$\frac{\frac{dP_H}{dt}}{\frac{dP_L}{dt}} = \frac{k_{cat}^H K_M^L[S_H]}{k_{cat}^L K_M^H[S_L]} = \frac{[S_H]}{[S_L]}\left(\frac{k_{cat}^H K_M^L}{k_{cat}^L K_M^H}\right) = \frac{[S_H]}{[S_L]}\left(\frac{\frac{k_{cat}^H}{K_M^H}}{\frac{k_{cat}^L}{K_M^L}}\right) \quad (11)$$

We define $\frac{k_{cat}^H}{K_M^H} \Big/ \frac{k_{cat}^L}{K_M^L}$ as the site-bias factor.

When $\frac{k_{cat}}{K_M}$ is the same for H and L species,

$\frac{\frac{dP_H}{dt}}{\frac{dP_L}{dt}} = \frac{[S_H]}{[S_L]}$, thus peptide generation doesn't suffer type I bias, even considering kinetics.

But $\frac{k_{cat}}{K_M}$ more likely differ with residues' protease preference, than with protein abundance, thus eq (11) may also explain domain-bias within the same protein for pepsin: peptides from accessible higher-preference residues (such as FWLMC (39)) will get generated faster in the digestion events.

Implications for digestion methods

Immobilizing pepsin on POROS bead surface enhances pepsin effective surface concentration to over-mM scale (33), able to reach over thousands fold higher than the 1 μM concentration of

total protein applied herein. This allows the pepsin column to operate within the abundance-unbiased high-efficiency region ($a \leq b$) for multiple digestion events. Unrestricted by sample size, as the flow instantly removes products from catalytic sites preventing inhibition, pepsin column digestion can process large amounts of protein sample—in the optimal region by diluting to sub-1 μM, and reproducibly form peptides between states where protein relative abundances may vary. Considering reaction kinetics, the theoretical number of digestion events (peptide length) can be controlled by flow rate, size and temperature of pepsin column, to obtain optimal peptide products.

By contrast, traditional in-solution trypsin digestion demands low concentration ratio of protease : protein substrate (such as 1 : 50, m/m) to reduce product contamination by protease, thereby placing the near-entire digestion in the low-efficiency region ($a >> b$), until substrate sites approach exhaustion (such as below 2% of starting substrates) which requires longer incubation time. Still, missed cleavages are routinely observed. Digestion efficiency is further compromised by partial trypsin deactivation from the frequently applied 2 M urea. Furthermore, substrate and product diffusion to/from the catalytic site, though often excluded by kinetic models, is slow during gentle agitation, and unpredictably decreases catalytic sites available. Thus traditional urea-trypsin in-solution digestion methods suffer low efficiency caused by intrinsic catalytic site-deficiency and slow diffusion—which FDD overcomes.