Supplementary Information

Bioinformatics and molecular dynamics analyses

of Ser/Thr motifs in transmembrane proteins

Coral del Val, Stephen H. White, and Ana-Nicoleta Bondar

Supplementary Information contains:

- Figure legends
- Figures S1-S10
- Supplementary Information References

Figure legends

Figure S1. Examples of molecular transporters from data-α set that have significant number of Ser/Thr Signatures. (A) The AMPA subtype ionotropic glutamate receptor (3kg2, Sobolevsky et al, 2009). Chain A is shown as orange ribbons, and chains B-D as transparent ribbons. CA atoms of the Ser/Thr amino acids from chain A are shown as orange van der Waals spheres. (B) The full-length potassium channel KcsA (1f6g, Cortes et al, 2001). Chain A is shown in red, and chains B-D transparent pink. CA atoms of Ser/Thr in chain A are shown as green van der Waals spheres. In panels C-I, the protein is shown is orange, with CA atoms of the Ser/Thr groups as green van der Waals spheres. (C) The AHA2 P-type ATPase proton pump (3b8c, Pederson et al 2007). (D) The P-type ATP-ase Ca2+ pump (2agv_a, Obara et al, 2005). (E) The *Neurospora* proton pump (1mhs; Kühlbrandt et al 2002). (F) The ATP-gated P2X4 ion channel (3h9v, chain A, Kawate et al. 2009). (G) The KirBac3.4 potassium channel (Gulbis et al (pdb entry 1XL4). (H). The Wza translocon ((2j58_A; Dong et al 2006). (I) Mouse P-glycoprotein ABCB1 (3g5u, chain A; Aller et al, 2009). (J) The arginine-agmatine antiporter (Fang et al 2009). The sequence used in the analysis was extracted from the 3hqk_A pdb structure. To prepare the image, we used pdb 3ncy, which superseded 3hqk. (K) Aquaporin AQP1 (1J4N chainA; Sui et al 2001). (L) Bovine rhodopsin (3c9l, Stenkamp 2008). (M) The rat muscarinic M3 receptor (2amk; Li et al, 2005). (N) The ShuA transporter (3fhh, Brillet et al). (O) The hemophore receptor HasR (3csl, Krieg et al, 2009). (P) The outer membrane porin (2f1c, Subbarao et al 2006). (Q) The colicin I receptor from *E. coli* (Buchanan et al. 2007). Ser/Thr sidechains are shown as van der Waals spheres with carbon atoms colored cyan, oxygen – red, and nitrogen – blue. (R) *N. europaea* Rh50 protein from Lupo et al. 2007. In the left-hand image, all Ser/Thr sidechains are depicted; the right-hand panel highlights three transmembrane helices (shown in orange) with their Ser/Thr content. (S) Ser/Thr amino acids of bacteriorhodopsin. Ser/Thr sidechains other than the three pairs of consecutive Thr groups (Thr46/Thr47, Thr89/Thr90, and Thr141/Thr142) are shown as van der Waals spheres colored in green. The Asp96 sidechain is depicted as purple van der Waals spheres, and retinal with the Lys216 sidechain as black bonds. The structure of bacteriorhodopsin was taken from Belrhali et al, 1999. In left-hand panel, helix B is highlighted in dark blue; in the right-hand panel, three TM helices (shown in orange) and their Ser/Thr content are highlighted, with remaining TM helices and loops shown in transparent gray.

Figure S2. Percentage data-α sequences that contain specific Signatures. The specific signature assessed is indicated for each of the Panels A-F. At least one motif is present in 20-30% of the data set. The presence of 2 Signatures can also be significant: 17% of the proteins have two SxxT motifs (panel B). Detailed analyses of the data-α Signatures is given in Figures S3-S8.

Figure S3. Data-β sequences that contain specific Signatures. 21-25% of the sequences have at least one Signature. See Figures S9-S14 for detailed analyses of data-β Signatures.

Figure S4. Summary of the statistical analyses of the sequences in data-α and data-β.

Figure S5. Summary of the Shapiro test.

Figures S6. Number of Signatures per chain in data-α. In Panels A-S, the histogram gives the number and type of a specific Signature present in each chain of data-α. Only chains with at least one Signature are included in the histogram. The insets give the color scheme used for each Signature subset.

Figures S7. Number of Signatures per chain in data-β. See legend of Figure S6 above for a brief description of how the histograms from panels A-F were computed.

Figures S8-S9. Hierarchichal clustering of Signatures counts per sequence. Columns correspond to the signatures analyzed (TxxT, SxxxS, TT, ST, SxxT, SS, TxxxT, SxxS). All counts were normalized between 0 and 1 for each Signature (see main Text). The clustering method used was complete linkage using the Euclidean distance as similarity measure. The colors indicate the cluster to which the sequences belong according to the clustering procedure described in the main text. The number of records used is 291 in Figure S8 (data-α), and 48 in Figure S9 (data-β). The sequences are identified according to the pdb ID and the chain used in the analysis. The numbers under each Signature type indicates the number of Signatures for that particular sequence. For example, sequences 1a91_A and 1afo_A have no Signatures.

Figure S10. Cα-rmsd profiles for peptides in Sim1 (panel A), Sim2 (panel B), Sim3 (panel C), and Sim4 (panel D). For the peptide in Sim1, the total rmsd profile is decomposed into the rmsd for the transmembrane part, and the N- and C-termini (amino acid residues 1-4 and 21-27, respectively. All rmsd values are reported in Å and were computed relative to the starting coordinates of the peptides.

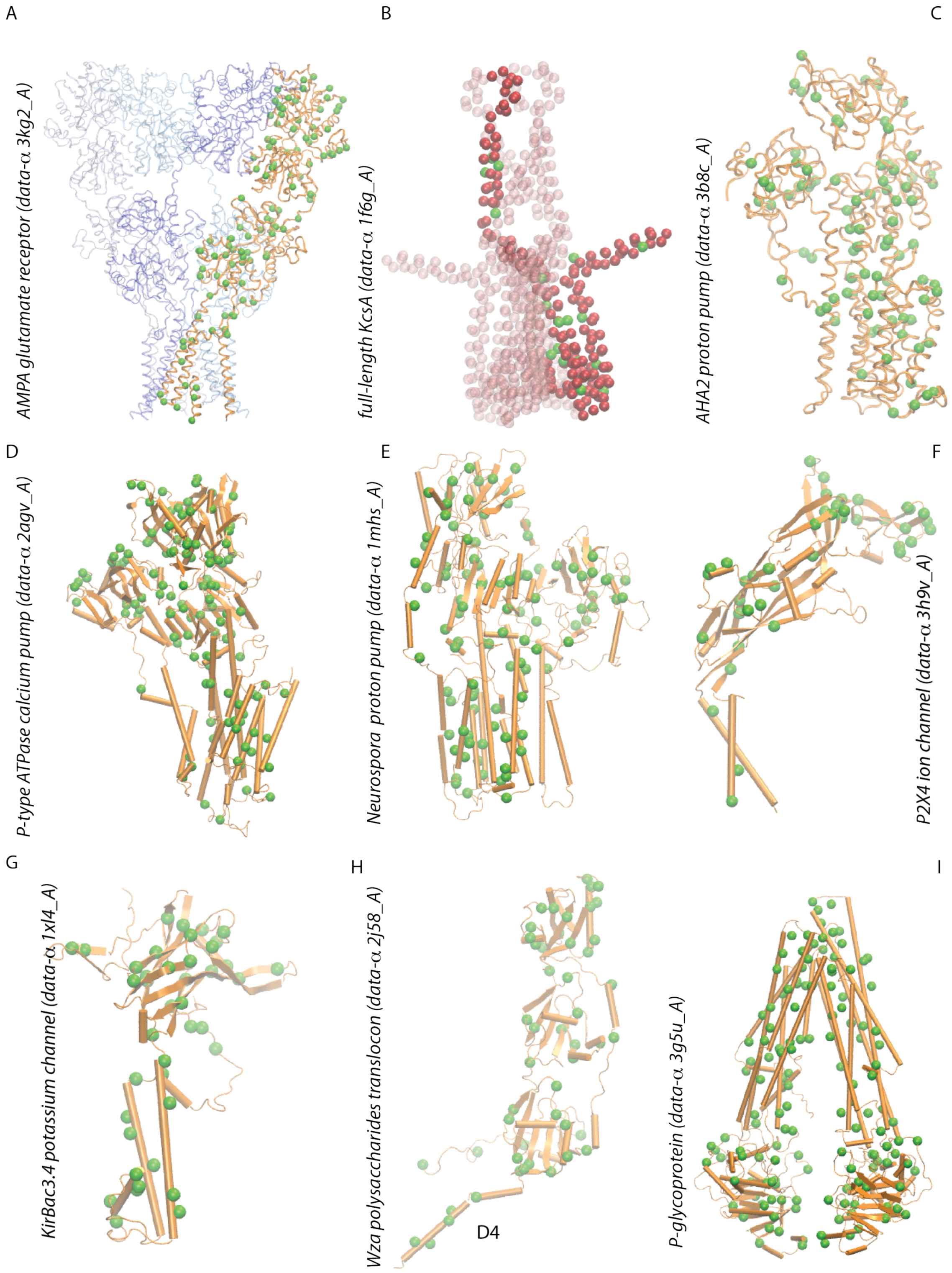


Figure S1- part1, panels A-I

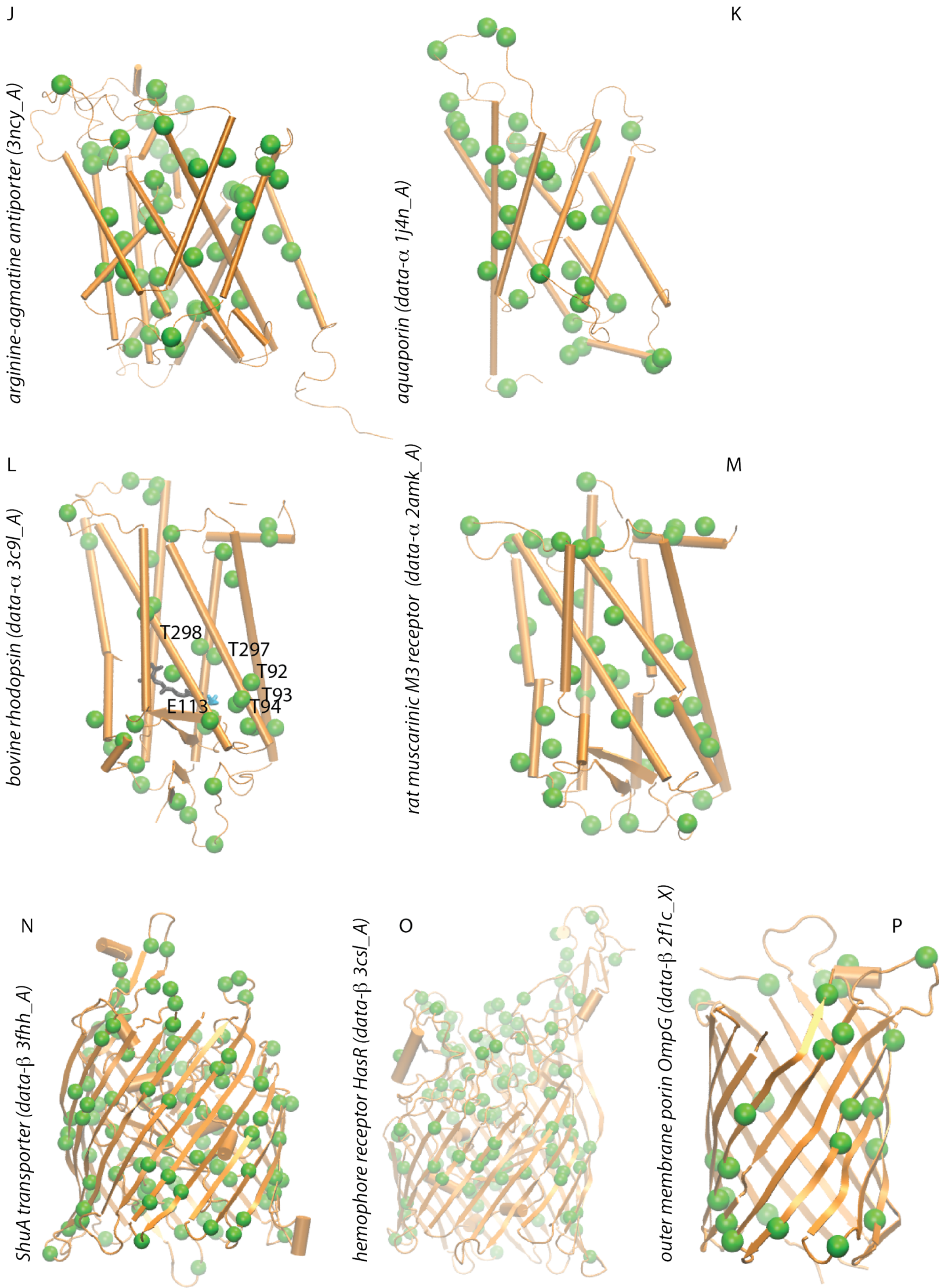
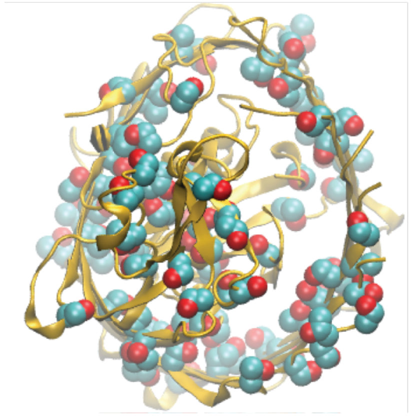
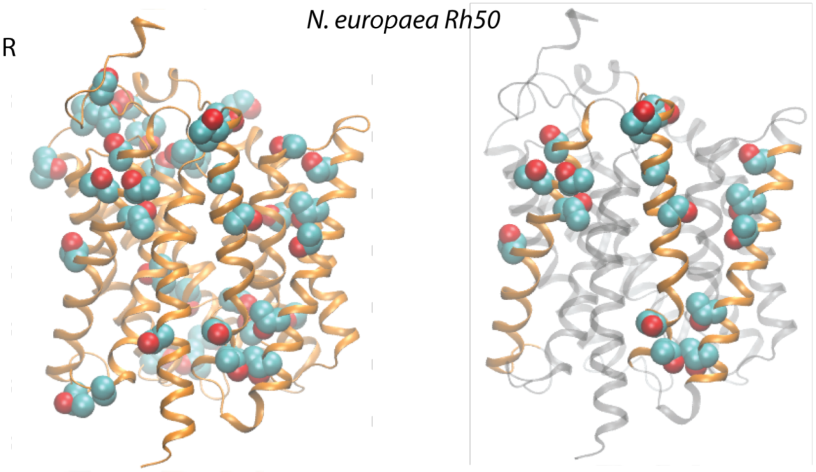


Figure S1- part 2, panels J-P

Q *colicin I receptor*



R *N. europaea Rh50*



S *bacteriorhodopsin*

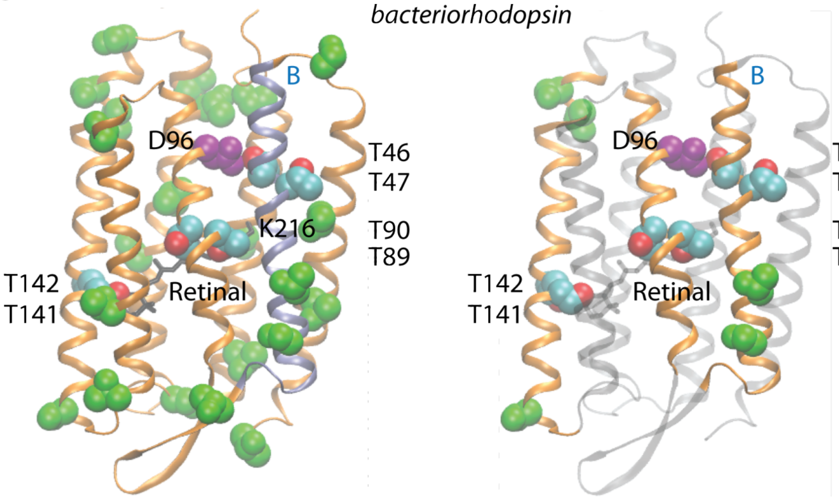


Figure S1 – part 3, panels Q-S

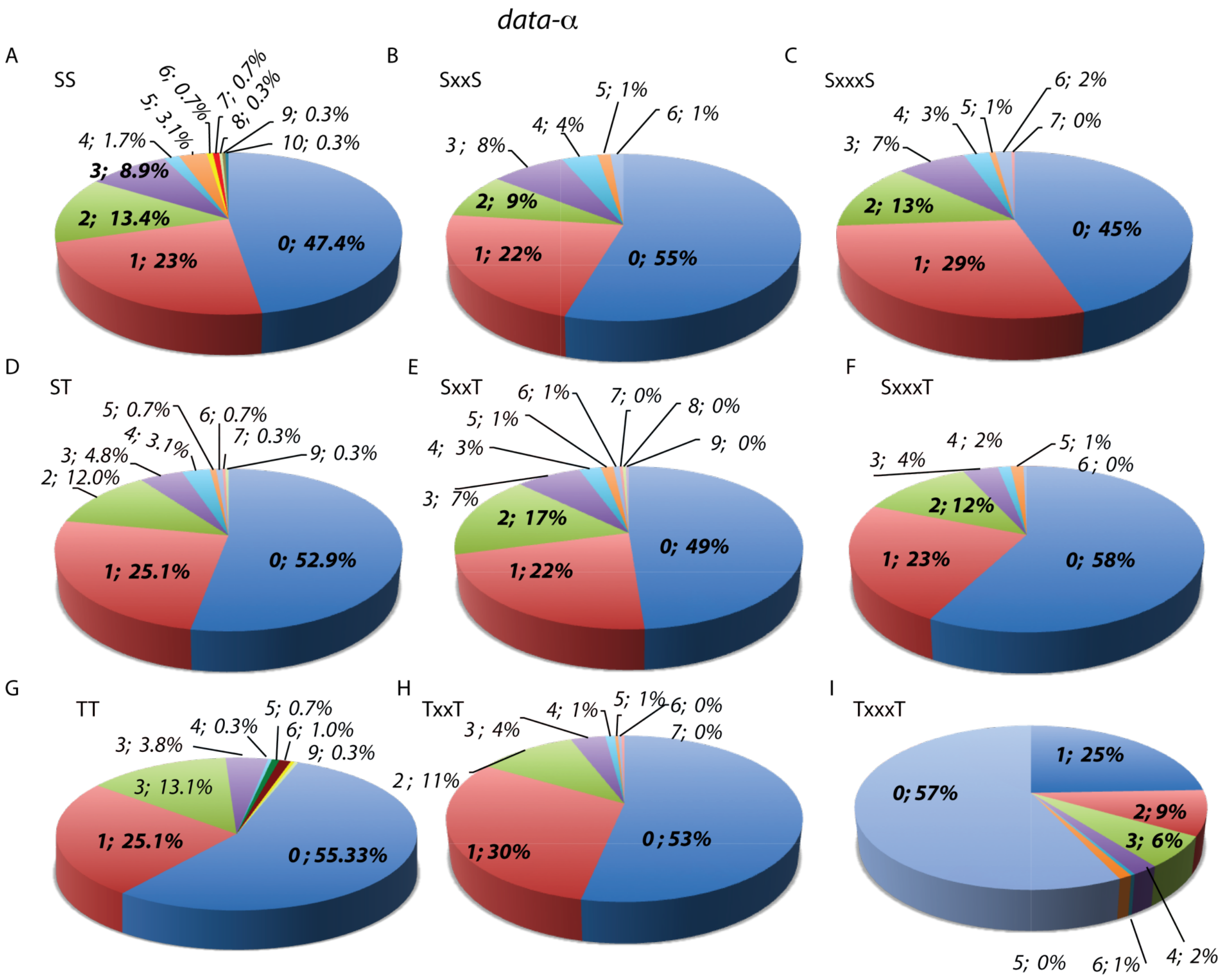

data-α
A
SS
6; 0.7%
7; 0.7%
5; 3.1%
8; 0.3%
9; 0.3%
4; 1.7%
10; 0.3%
3; 8.9%
2; 13.4%
0; 47.4%
1; 23%
B
SxxS
5; 1%
6; 1%
4; 4%
3 ; 8%
2; 9%
1; 22%
0; 55%
C
SxxxS
6; 2%
4; 3%
5; 1%
7; 0%
3; 7%
2; 13%
0; 45%
1; 29%
D
ST
5; 0.7%
6; 0.7%
4; 3.1%
7; 0.3%
3; 4.8%
9; 0.3%
2; 12.0%
0; 52.9%
1; 25.1%
E
SxxT
6; 1%
7; 0%
8; 0%
5; 1%
9; 0%
4; 3%
3; 7%
2; 17%
0; 49%
1; 22%
F
SxxxT
4 ; 2%
5; 1%
3; 4%
6 ; 0%
2; 12%
1; 23%
0; 58%
G
TT
5; 0.7%
4; 0.3%
6; 1.0%
3; 3.8%
9; 0.3%
3; 13.1%
1; 25.1%
0 ; 55.33%
H
TxxT
5; 1%
6; 0%
4; 1%
3 ; 4%
7; 0%
2 ; 11%
0 ; 53%
1; 30%
I
TxxxT
1; 25%
0; 57%
2; 9%
3; 6%
5; 0%
6; 1%
4; 2%

Figure S2

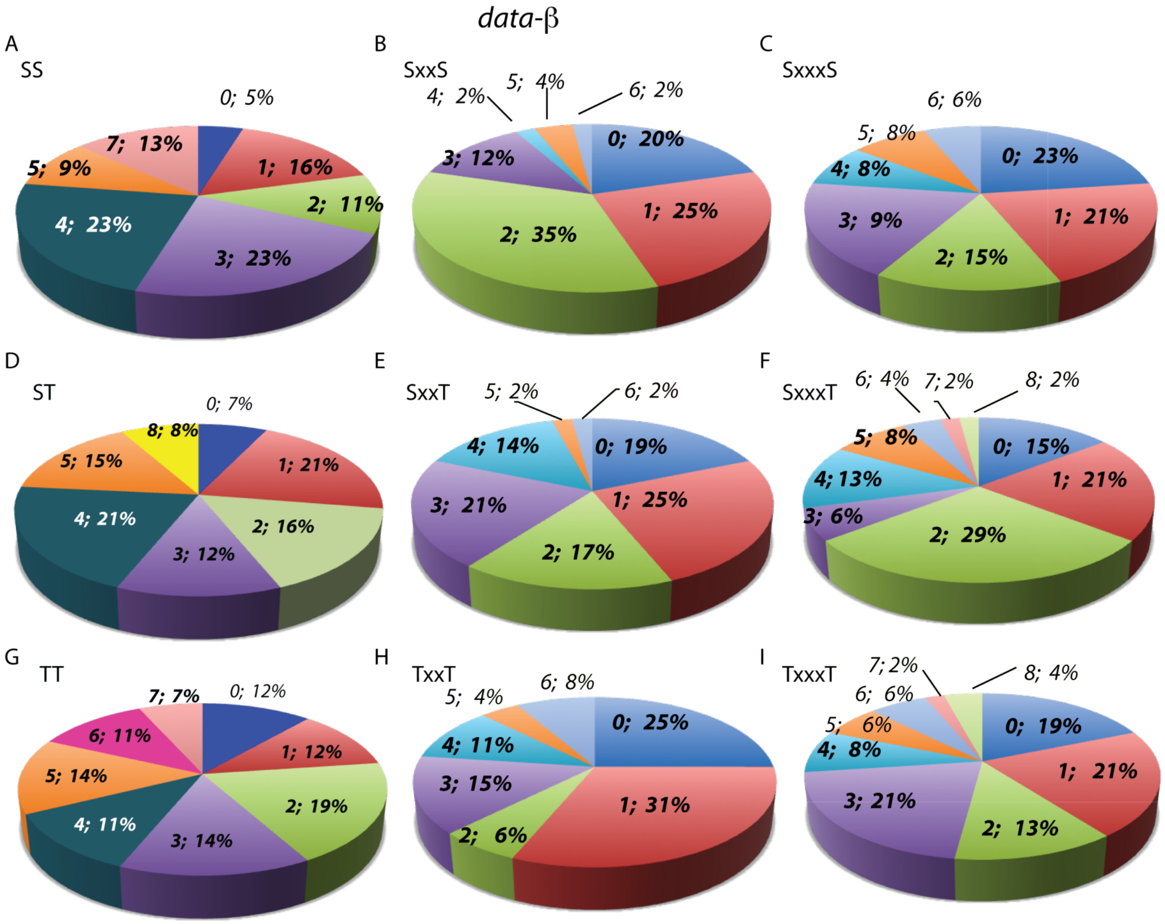

data-β
A
SS
0; 5%
1; 16%
2; 11%
3; 23%
4; 23%
5; 9%
7; 13%
B
SxxS
4; 2%
5; 4%
6; 2%
0; 20%
1; 25%
2; 35%
3; 12%
C
SxxxS
6; 6%
5; 8%
4; 8%
0; 23%
1; 21%
2; 15%
3; 9%
D
ST
0; 7%
8; 8%
1; 21%
5; 15%
4; 21%
2; 16%
3; 12%
E
SxxT
5; 2%
6; 2%
4; 14%
0; 19%
1; 25%
3; 21%
2; 17%
F
SxxxT
6; 4%
7; 2%
8; 2%
5; 8%
0; 15%
4; 13%
1; 21%
3; 6%
2; 29%
G
TT
7; 7%
0; 12%
6; 11%
1; 12%
5; 14%
2; 19%
4; 11%
3; 14%
H
TxxT
6; 8%
5; 4%
0; 25%
4; 11%
3; 15%
1; 31%
2; 6%
I
TxxxT
7; 2%
8; 4%
6; 6%
5; 6%
0; 19%
4; 8%
1; 21%
3; 21%
2; 13%

Figure S3

Figure S4-A

Alpha value (for confidence interval) 0.02

Variable #1 (TT)

Count	291	*Skewness*	2.28
Mean	0.76	*Skewness Standard Error*	0.14
Mean LCL	0.6	*Kurtosis*	10.01
Mean UCL	0.92	*Kurtosis Standard Error*	0.28
Variance	1.31	*Alternative Skewness (Fisher's)*	2.29
Standard Deviation	1.14	*Alternative Kurtosis (Fisher's)*	7.15
Mean Standard Error	0.07	*Coefficient of Variation*	1.5
Minimum	0.E+0	*Mean Deviation*	0.85
Maximum	7.	*Second Moment*	1.31
Range	7.	*Third Moment*	3.4
Sum	221.	*Fourth Moment*	17.06
Sum Standard Error	19.49	*Median*	0.E+0
Total Sum Squares	547.	*Median Error*	0.
Adjusted Sum Squares	378.58	*Percentile 25% (Q1)*	0.E+0
Geometric Mean	1.19	*Percentile 75% (Q2)*	1.
Harmonic Mean	2.99	*IQR*	1.
Mode	0.E+0	*MAD*	0.E+0

Alpha value (for confidence interval) 0.02

Variable #1 (TT)

Count	48	*Skewness*	0.97
Mean	1.94	*Skewness Standard Error*	0.34
Mean LCL	1.29	*Kurtosis*	3.16
Mean UCL	2.58	*Kurtosis Standard Error*	0.63
Variance	3.42	*Alternative Skewness (Fisher's)*	1.
Standard Deviation	1.85	*Alternative Kurtosis (Fisher's)*	0.32
Mean Standard Error	0.27	*Coefficient of Variation*	0.95
Minimum	0.E+0	*Mean Deviation*	1.44
Maximum	7.	*Second Moment*	3.35
Range	7.	*Third Moment*	5.94
Sum	93.	*Fourth Moment*	35.51
Sum Standard Error	12.82	*Median*	1.5
Total Sum Squares	341.	*Median Error*	0.05
Adjusted Sum Squares	160.81	*Percentile 25% (Q1)*	1.
Geometric Mean	1.75	*Percentile 75% (Q2)*	3.
Harmonic Mean	2.34	*IQR*	2.
Mode	#N/A	*MAD*	1.5

Alpha value (for confidence interval) 0.02

Variable #1 (ST)

Count	291	*Skewness*	2.27
Mean	0.89	*Skewness Standard Error*	0.14
Mean LCL	0.71	*Kurtosis*	10.2
Mean UCL	1.07	*Kurtosis Standard Error*	0.28
Variance	1.74	*Alternative Skewness (Fisher's)*	2.28
Standard Deviation	1.32	*Alternative Kurtosis (Fisher's)*	7.35
Mean Standard Error	0.08	*Coefficient of Variation*	1.48
Minimum	0.E+0	*Mean Deviation*	0.94
Maximum	9.	*Second Moment*	1.73
Range	9.	*Third Moment*	5.18
Sum	259.	*Fourth Moment*	30.66
Sum Standard Error	22.5	*Median*	0.E+0
Total Sum Squares	735.	*Median Error*	0.01
Adjusted Sum Squares	504.48	*Percentile 25% (Q1)*	0.E+0
Geometric Mean	1.24	*Percentile 75% (Q2)*	1.
Harmonic Mean	2.96	*IQR*	1.
Mode	0.E+0	*MAD*	0.E+0

Alpha value (for confidence interval) 0.02

Variable #1 (ST)

Count	48	*Skewness*	1.36
Mean	1.9	*Skewness Standard Error*	0.34
Mean LCL	1.31	*Kurtosis*	4.97
Mean UCL	2.48	*Kurtosis Standard Error*	0.63
Variance	2.86	*Alternative Skewness (Fisher's)*	1.41
Standard Deviation	1.69	*Alternative Kurtosis (Fisher's)*	2.33
Mean Standard Error	0.24	*Coefficient of Variation*	0.89
Minimum	0.E+0	*Mean Deviation*	1.3
Maximum	8.	*Second Moment*	2.8
Range	8.	*Third Moment*	6.4
Sum	91.	*Fourth Moment*	39.05
Sum Standard Error	11.72	*Median*	1.
Total Sum Squares	307.	*Median Error*	0.04
Adjusted Sum Squares	134.48	*Percentile 25% (Q1)*	1.
Geometric Mean	1.64	*Percentile 75% (Q2)*	3.
Harmonic Mean	1.76	*IQR*	2.
Mode	1.	*MAD*	1.

Alpha value (for confidence interval) 0.02

Variable #1 (SS)

Count	291	*Skewness*	2.11
Mean	1.17	*Skewness Standard Error*	0.14
Mean LCL	0.95	*Kurtosis*	8.77
Mean UCL	1.4	*Kurtosis Standard Error*	0.28
Variance	2.68	*Alternative Skewness (Fisher's)*	2.12
Standard Deviation	1.64	*Alternative Kurtosis (Fisher's)*	5.89
Mean Standard Error	0.1	*Coefficient of Variation*	1.4
Minimum	0.E+0	*Mean Deviation*	1.19
Maximum	10.	*Second Moment*	2.67
Range	10.	*Third Moment*	9.22
Sum	341.	*Fourth Moment*	62.61
Sum Standard Error	27.93	*Median*	1.
Total Sum Squares	1,177.	*Median Error*	0.01
Adjusted Sum Squares	777.41	*Percentile 25% (Q1)*	0.E+0
Geometric Mean	1.37	*Percentile 75% (Q2)*	2.
Harmonic Mean	2.93	*IQR*	2.
Mode	0.E+0	*MAD*	1.

Alpha value (for confidence interval) 0.02

Variable #1 (SS)

Count	48	*Skewness*	1.75
Mean	2.33	*Skewness Standard Error*	0.34
Mean LCL	1.59	*Kurtosis*	7.13
Mean UCL	3.08	*Kurtosis Standard Error*	0.63
Variance	4.61	*Alternative Skewness (Fisher's)*	1.81
Standard Deviation	2.15	*Alternative Kurtosis (Fisher's)*	4.73
Mean Standard Error	0.31	*Coefficient of Variation*	0.92
Minimum	0.E+0	*Mean Deviation*	1.61
Maximum	11.	*Second Moment*	4.51
Range	11.	*Third Moment*	16.78
Sum	112.	*Fourth Moment*	145.23
Sum Standard Error	14.88	*Median*	2.
Total Sum Squares	478.	*Median Error*	0.06
Adjusted Sum Squares	216.67	*Percentile 25% (Q1)*	1.
Geometric Mean	1.9	*Percentile 75% (Q2)*	3.
Harmonic Mean	1.92	*IQR*	2.
Mode	1.	*MAD*	1.

Figure S4-B

ALPHA Chains

Variable #1 (Count SxxxS)			
Count	291	*Skewness*	1.81268
Mean	1.05498	*Skewness Standard Error*	0.14237
Mean LCL	0.86905	*Kurtosis*	7.07353
Mean UCL	1.24091	*Kurtosis Standard Error*	0.28183
Variance	1.83835	*Alternative Skewness (Fisher's)*	1.82209
Standard Deviation	1.35586	*Alternative Kurtosis (Fisher's)*	4.1653
Mean Standard Error	0.07948	*Coefficient of Variation*	1.28519
Minimum	0.E+0	*Mean Deviation*	0.97472
Maximum	8.	*Second Moment*	1.83203
Range	8.	*Third Moment*	4.49489
Sum	307.	*Fourth Moment*	23.74108
Sum Standard Error	23.12917	*Median*	1.
Total Sum Squares	857.	*Median Error*	0.00584
Adjusted Sum Squares	533.12027	*Percentile 25% (Q1)*	0.E+0
Geometric Mean	1.29781	*Percentile 75% (Q2)*	2.
Harmonic Mean	2.53724	*IQR*	2.
Mode	0.E+0	*MAD*	1.

Variable #2 (Count SxxS)			
Count	291	*Skewness*	1.65303
Mean	0.93471	*Skewness Standard Error*	0.14237
Mean LCL	0.74797	*Kurtosis*	5.3019
Mean UCL	1.12144	*Kurtosis Standard Error*	0.28183
Variance	1.85434	*Alternative Skewness (Fisher's)*	1.66161
Standard Deviation	1.36174	*Alternative Kurtosis (Fisher's)*	2.36286
Mean Standard Error	0.07983	*Coefficient of Variation*	1.45686
Minimum	0.E+0	*Mean Deviation*	1.02143
Maximum	6.	*Second Moment*	1.84797
Range	6.	*Third Moment*	4.15263
Sum	272.	*Fourth Moment*	18.10598
Sum Standard Error	23.22959	*Median*	0.E+0
Total Sum Squares	792.	*Median Error*	0.00586
Adjusted Sum Squares	537.75945	*Percentile 25% (Q1)*	0.E+0
Geometric Mean	1.27831	*Percentile 75% (Q2)*	1.
Harmonic Mean	3.25564	*IQR*	1.
Mode	0.E+0	*MAD*	0.E+0

Variable #3 (Count SxxT)			
Count	291	*Skewness*	2.02538
Mean	1.05498	*Skewness Standard Error*	0.14237
Mean LCL	0.85788	*Kurtosis*	8.72826
Mean UCL	1.25209	*Kurtosis Standard Error*	0.28183
Variance	2.06593	*Alternative Skewness (Fisher's)*	2.03589
Standard Deviation	1.43734	*Alternative Kurtosis (Fisher's)*	5.84882
Mean Standard Error	0.08426	*Coefficient of Variation*	1.36243
Minimum	0.E+0	*Mean Deviation*	1.05379
Maximum	9.	*Second Moment*	2.05883
Range	9.	*Third Moment*	5.98326
Sum	307.	*Fourth Moment*	36.99726
Sum Standard Error	24.5191	*Median*	1.
Total Sum Squares	923.	*Median Error*	0.00619
Adjusted Sum Squares	599.12027	*Percentile 25% (Q1)*	0.E+0
Geometric Mean	1.32682	*Percentile 75% (Q2)*	2.
Harmonic Mean	2.96146	*IQR*	2.
Mode	0.E+0	*MAD*	1.

Variable #4 (Count SxxxT)			
Count	291	*Skewness*	2.04607
Mean	0.7354	*Skewness Standard Error*	0.14237
Mean LCL	0.58205	*Kurtosis*	8.28952
Mean UCL	0.88874	*Kurtosis Standard Error*	0.28183
Variance	1.25043	*Alternative Skewness (Fisher's)*	2.05668
Standard Deviation	1.11823	*Alternative Kurtosis (Fisher's)*	5.40245
Mean Standard Error	0.06555	*Coefficient of Variation*	1.52058
Minimum	0.E+0	*Mean Deviation*	0.84912
Maximum	7.	*Second Moment*	1.24614
Range	7.	*Third Moment*	2.84621
Sum	214.	*Fourth Moment*	12.87241
Sum Standard Error	19.07553	*Median*	0.E+0
Total Sum Squares	520.	*Median Error*	0.00482
Adjusted Sum Squares	362.62543	*Percentile 25% (Q1)*	0.E+0
Geometric Mean	1.18858	*Percentile 75% (Q2)*	1.
Harmonic Mean	3.19396	*IQR*	1.
Mode	0.E+0	*MAD*	0.E+0

BETA CHAINS

Variable #1 (Count SxxxS)			
Count	48	*Skewness*	0.51381
Mean	2.1875	*Skewness Standard Error*	0.33592
Mean LCL	1.5435	*Kurtosis*	2.21899
Mean UCL	2.8315	*Kurtosis Standard Error*	0.63196
Variance	3.43218	*Alternative Skewness (Fisher's)*	0.53054
Standard Deviation	1.85261	*Alternative Kurtosis (Fisher's)*	-0.73269
Mean Standard Error	0.2674	*Coefficient of Variation*	0.84691
Minimum	0.E+0	*Mean Deviation*	1.55208
Maximum	6.	*Second Moment*	3.36068
Range	6.	*Third Moment*	3.16553
Sum	105.	*Fourth Moment*	25.06156
Sum Standard Error	12.83529	*Median*	2.
Total Sum Squares	391.	*Median Error*	0.04837
Adjusted Sum Squares	161.3125	*Percentile 25% (Q1)*	1.
Geometric Mean	1.9517	*Percentile 75% (Q2)*	3.
Harmonic Mean	2.55319	*IQR*	2.
Mode	0.E+0	*MAD*	1.

Variable #2 (Count SxxS)			
Count	48	*Skewness*	1.61804
Mean	1.79167	*Skewness Standard Error*	0.33592
Mean LCL	1.26007	*Kurtosis*	7.27672
Mean UCL	2.32326	*Kurtosis Standard Error*	0.63196
Variance	2.33865	*Alternative Skewness (Fisher's)*	1.67071
Standard Deviation	1.52927	*Alternative Kurtosis (Fisher's)*	4.89434
Mean Standard Error	0.22073	*Coefficient of Variation*	0.85354
Minimum	0.E+0	*Mean Deviation*	1.06771
Maximum	8.	*Second Moment*	2.28993
Range	8.	*Third Moment*	5.60692
Sum	86.	*Fourth Moment*	38.15754
Sum Standard Error	10.59506	*Median*	2.
Total Sum Squares	264.	*Median Error*	0.03993
Adjusted Sum Squares	109.91667	*Percentile 25% (Q1)*	1.
Geometric Mean	1.68553	*Percentile 75% (Q2)*	2.
Harmonic Mean	2.0623	*IQR*	1.
Mode	2.	*MAD*	1.

Variable #3 (Count SxxT)			
Count	48	*Skewness*	0.39149
Mean	2.02083	*Skewness Standard Error*	0.33592
Mean LCL	1.48668	*Kurtosis*	2.37423
Mean UCL	2.55499	*Kurtosis Standard Error*	0.63196
Variance	2.36126	*Alternative Skewness (Fisher's)*	0.40423
Standard Deviation	1.53664	*Alternative Kurtosis (Fisher's)*	-0.55998
Mean Standard Error	0.22179	*Coefficient of Variation*	0.7604
Minimum	0.E+0	*Mean Deviation*	1.27517
Maximum	6.	*Second Moment*	2.31207
Range	6.	*Third Moment*	1.37632
Sum	97.	*Fourth Moment*	12.69179
Sum Standard Error	10.64615	*Median*	2.
Total Sum Squares	307.	*Median Error*	0.04012
Adjusted Sum Squares	110.97917	*Percentile 25% (Q1)*	1.
Geometric Mean	1.85415	*Percentile 75% (Q2)*	3.
Harmonic Mean	2.23776	*IQR*	2.
Mode	1.	*MAD*	1.

Variable #4 (Count SxxxT)			
Count	48	*Skewness*	0.81957
Mean	2.4375	*Skewness Standard Error*	0.33592
Mean LCL	1.74262	*Kurtosis*	3.05187
Mean UCL	3.13238	*Kurtosis Standard Error*	0.63196
Variance	3.99601	*Alternative Skewness (Fisher's)*	0.84625
Standard Deviation	1.999	*Alternative Kurtosis (Fisher's)*	0.19395
Mean Standard Error	0.28853	*Coefficient of Variation*	0.8201
Minimum	0.E+0	*Mean Deviation*	1.60677
Maximum	8.	*Second Moment*	3.91276
Range	8.	*Third Moment*	6.34326
Sum	117.	*Fourth Moment*	46.72326
Sum Standard Error	13.84949	*Median*	2.
Total Sum Squares	473.	*Median Error*	0.0522
Adjusted Sum Squares	187.8125	*Percentile 25% (Q1)*	1.
Geometric Mean	2.08915	*Percentile 75% (Q2)*	4.
Harmonic Mean	2.41192	*IQR*	3.
Mode	2.	*MAD*	1.

Figure S4-C

Variable #5 (Count TxxxT)			
Count	291	*Skewness*	1.92348
Mean	0.76976	*Skewness Standard Error*	0.14237
Mean LCL	0.61001	*Kurtosis*	7.10462
Mean UCL	0.92951	*Kurtosis Standard Error*	0.28183
Variance	1.35715	*Alternative Skewness (Fisher's)*	1.93346
Standard Deviation	1.16497	*Alternative Kurtosis (Fisher's)*	4.19694
Mean Standard Error	0.06829	*Coefficient of Variation*	1.51342
Minimum	0.E+0	*Mean Deviation*	0.87821
Maximum	6.	*Second Moment*	1.35249
Range	6.	*Third Moment*	3.02544
Sum	224.	*Fourth Moment*	12.99594
Sum Standard Error	19.87287	*Median*	0.E+0
Total Sum Squares	566.	*Median Error*	0.00502
Adjusted Sum Squares	393.57388	*Percentile 25% (Q1)*	0.E+0
Geometric Mean	1.20018	*Percentile 75% (Q2)*	1.
Harmonic Mean	3.15618	*IQR*	1.
Mode	0.E+0	*MAD*	0.E+0

Variable #5 (Count TxxxT)			
Count	48	*Skewness*	0.77054
Mean	2.58333	*Skewness Standard Error*	0.33592
Mean LCL	1.81827	*Kurtosis*	2.8654
Mean UCL	3.3484	*Kurtosis Standard Error*	0.63196
Variance	4.84397	*Alternative Skewness (Fisher's)*	0.79563
Standard Deviation	2.2009	*Alternative Kurtosis (Fisher's)*	-0.01352
Mean Standard Error	0.31767	*Coefficient of Variation*	0.85196
Minimum	0.E+0	*Mean Deviation*	1.77431
Maximum	8.	*Second Moment*	4.74306
Range	8.	*Third Moment*	7.95949
Sum	124.	*Fourth Moment*	64.46166
Sum Standard Error	15.2483	*Median*	2.
Total Sum Squares	548.	*Median Error*	0.05747
Adjusted Sum Squares	227.66667	*Percentile 25% (Q1)*	1.
Geometric Mean	2.1615	*Percentile 75% (Q2)*	4.
Harmonic Mean	2.54964	*IQR*	3.
Mode	#N/A	*MAD*	1.

Variable #6 (Count TxxT)			
Count	291	*Skewness*	2.17813
Mean	0.73196	*Skewness Standard Error*	0.14237
Mean LCL	0.58947	*Kurtosis*	10.10485
Mean UCL	0.87444	*Kurtosis Standard Error*	0.28183
Variance	1.07963	*Alternative Skewness (Fisher's)*	2.18943
Standard Deviation	1.03905	*Alternative Kurtosis (Fisher's)*	7.24936
Mean Standard Error	0.06091	*Coefficient of Variation*	1.41955
Minimum	0.E+0	*Mean Deviation*	0.77975
Maximum	7.	*Second Moment*	1.07592
Range	7.	*Third Moment*	2.43082
Sum	213.	*Fourth Moment*	11.69741
Sum Standard Error	17.72491	*Median*	0.E+0
Total Sum Squares	469.	*Median Error*	0.00448
Adjusted Sum Squares	313.09278	*Percentile 25% (Q1)*	0.E+0
Geometric Mean	1.15955	*Percentile 75% (Q2)*	1.
Harmonic Mean	2.68385	*IQR*	1.
Mode	0.E+0	*MAD*	0.E+0

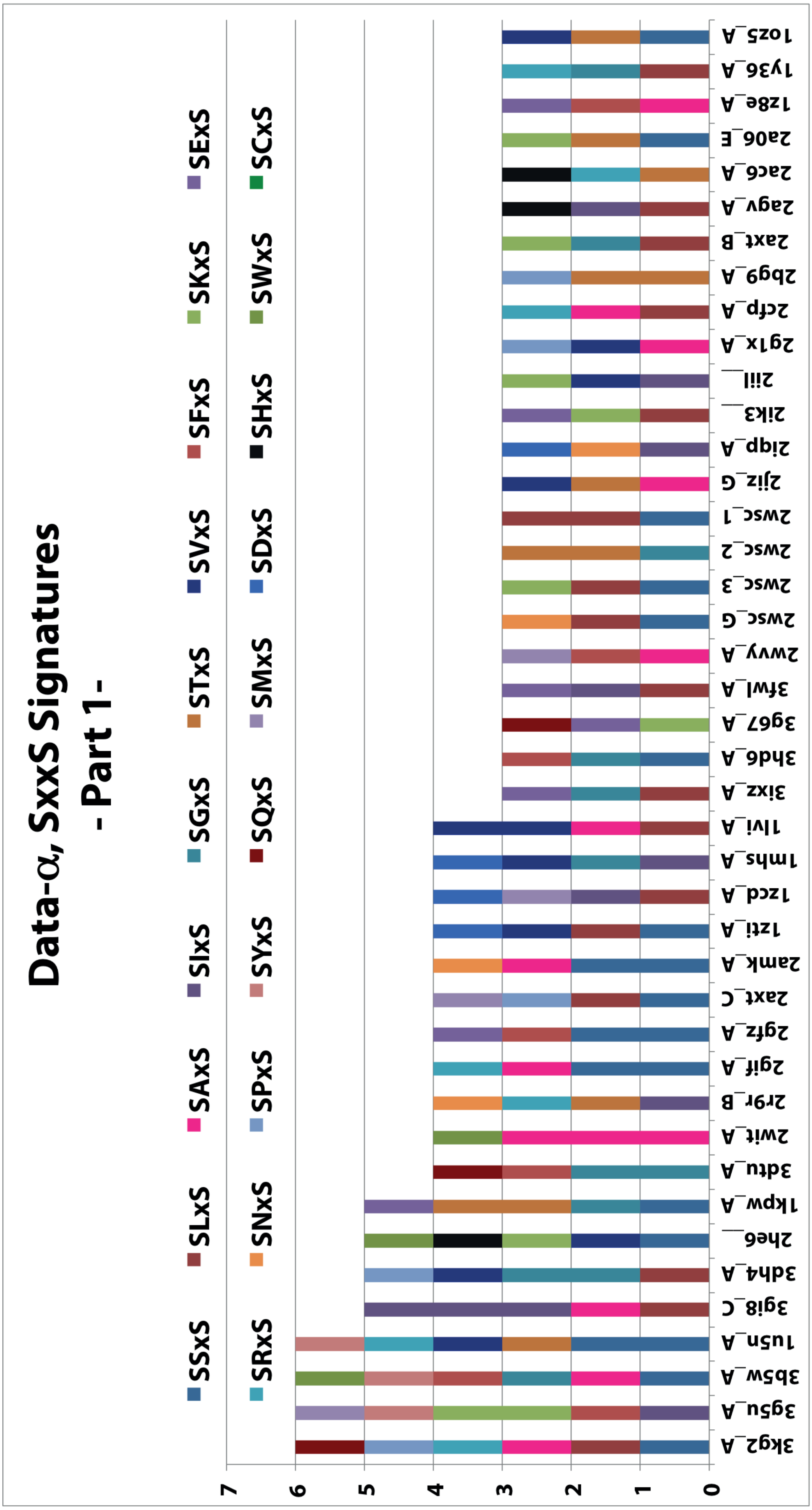
Variable #6 (Count TxxT)			
Count	48	*Skewness*	0.75555
Mean	2.	*Skewness Standard Error*	0.33592
Mean LCL	1.33501	*Kurtosis*	2.4013
Mean UCL	2.66499	*Kurtosis Standard Error*	0.63196
Variance	3.65957	*Alternative Skewness (Fisher's)*	0.78014
Standard Deviation	1.913	*Alternative Kurtosis (Fisher's)*	-0.52986
Mean Standard Error	0.27612	*Coefficient of Variation*	0.9565
Minimum	0.E+0	*Mean Deviation*	1.625
Maximum	6.	*Second Moment*	3.58333
Range	6.	*Third Moment*	5.125
Sum	96.	*Fourth Moment*	30.83333
Sum Standard Error	13.25366	*Median*	1.
Total Sum Squares	364.	*Median Error*	0.04995
Adjusted Sum Squares	172.	*Percentile 25% (Q1)*	1.
Geometric Mean	1.75824	*Percentile 75% (Q2)*	3.
Harmonic Mean	2.2695	*IQR*	2.
Mode	1.	*MAD*	1.

Figure S5

Signature	**Alpha chains**			Beta chains		
	W	**P**	**Normal Distribution**	**W**	**P**	**Normal Distribution**
SxxxS	0.7575	2.20E-16	No	0.9046	0.0008876	No
SxxS	0.7193	2.20E-16	No	0.8342	8.48E-06	No
SxxT	0.7368	2.20E-16	No	0.9215	0.003362	No
SxxxT	0.6908	2.20E-16	No	0.9062	0.001001	No
TxxxT	0.6951	2.20E-16	No	0.9068	0.001051	No
TxxT	0.703	2.20E-16	No	0.8591	3.83E-05	No

if the p-value is less than the chosen alpha level (0.05), then the null hypothesis is rejected

Figure S6-A


Data-α, SxxS Signatures
- Part 1-
SSxS
SLxS
SAxS
SIxS
SGxS
STxS
SVxS
SFxS
SKxS
SExS
SRxS
SNxS
SPxS
SYxS
SQxS
SMxS
SDxS
SHxS
SWxS
SCxS
0
1
2
3
4
5
6
7
3kg2_A
3g5u_A
3b5w_A
1u5n_A
3gi8_C
3dh4_A
2he6_
1kpw_A
3dtu_A
2wit_A
2r9r_B
2gif_A
2gfz_A
2axt_C
2amk_A
1zti_A
1zcd_A
1mhs_A
1lvi_A
3ixz_A
3hd6_A
3g67_A
3fwl_A
2wvy_A
2wsc_G
2wsc_3
2wsc_2
2wsc_1
2jiz_G
2iqp_A
2ik3_
2iil_
2g1x_A
2cfp_A
2bg9_A
2axt_B
2agv_A
2ac6_A
2a06_E
1z8e_A
1y36_A
1oz5_A

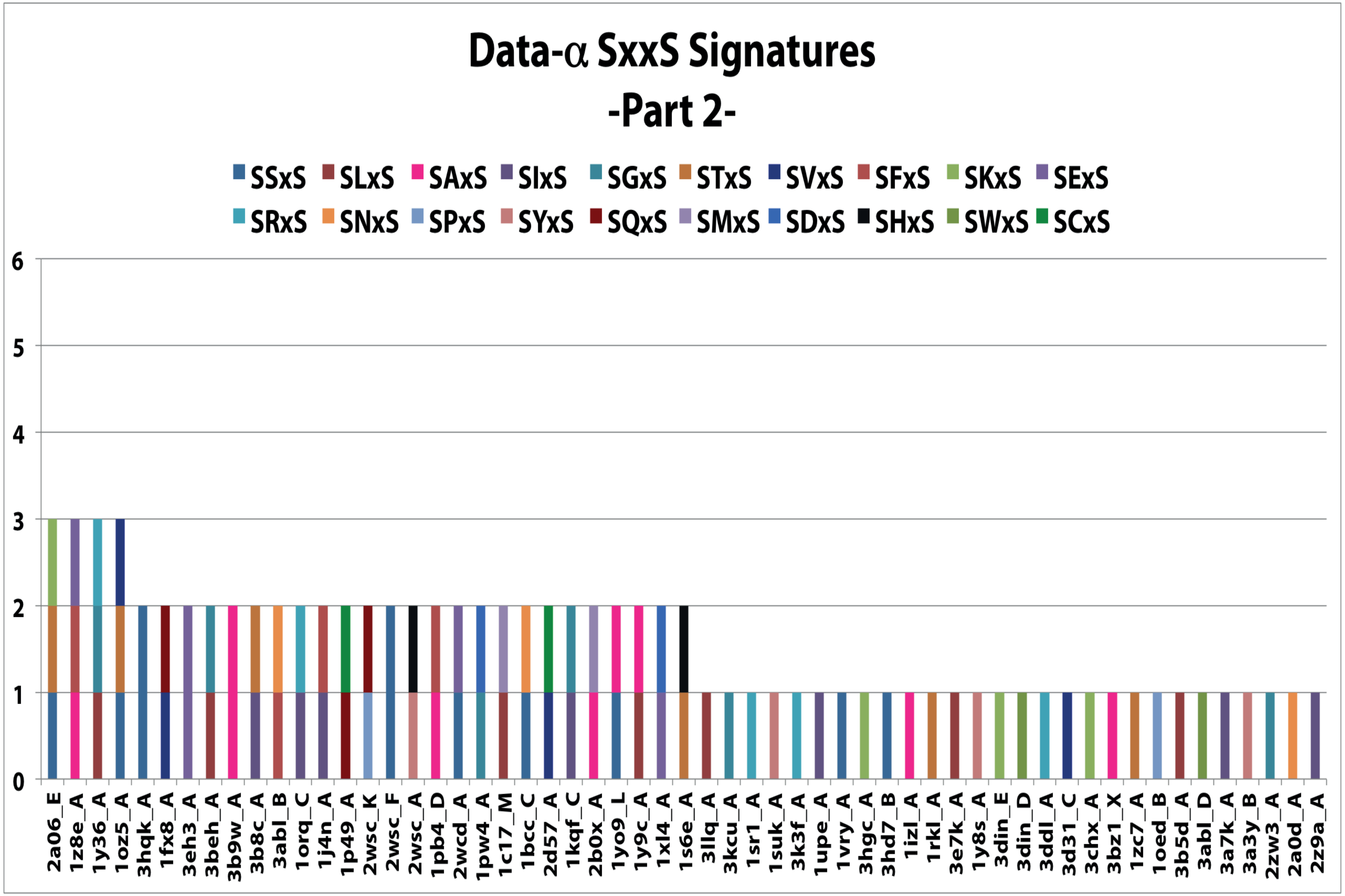

Data-α SxxS Signatures
-Part 2-

Figure S6-B

Figure S6-C

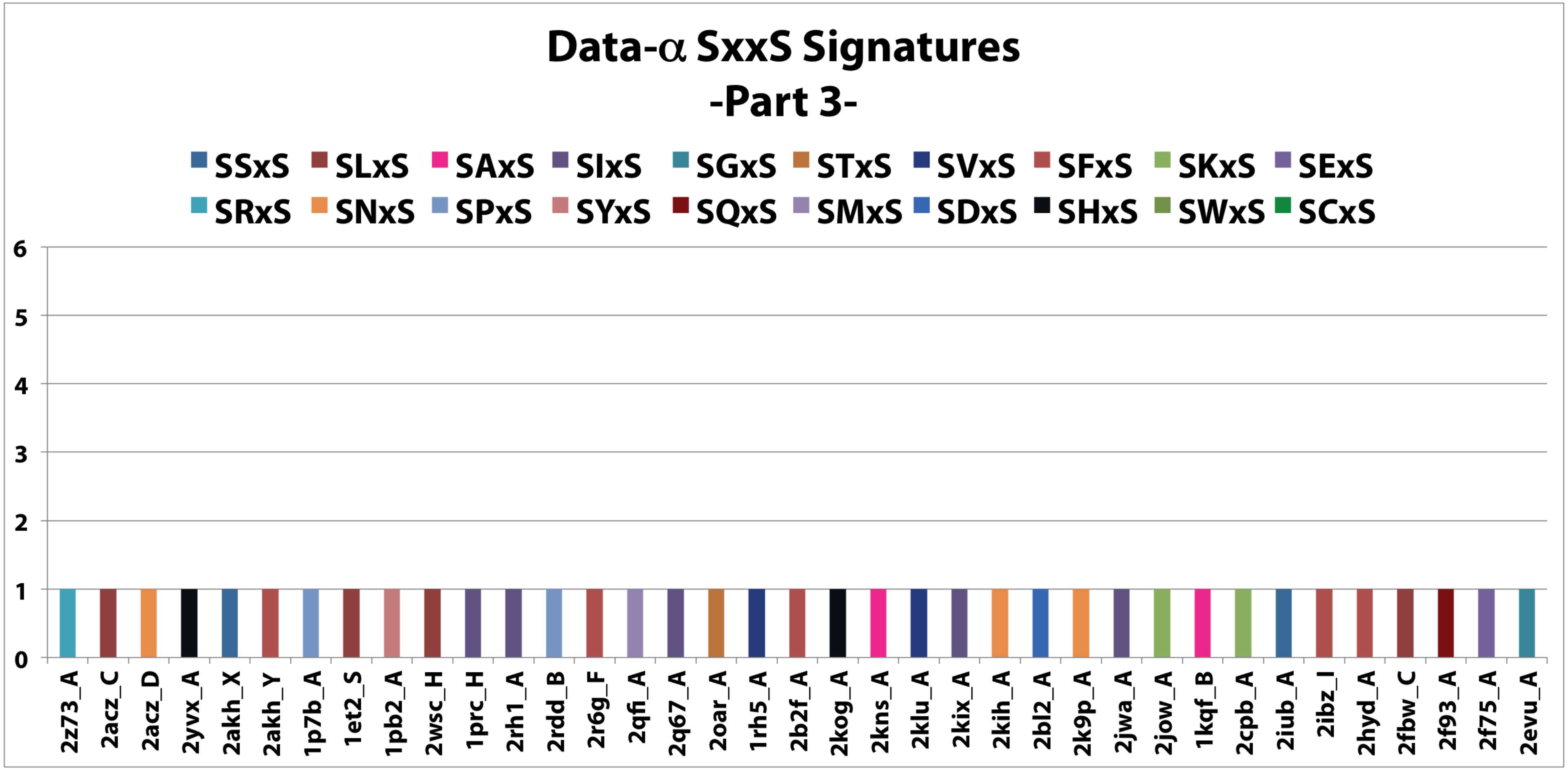
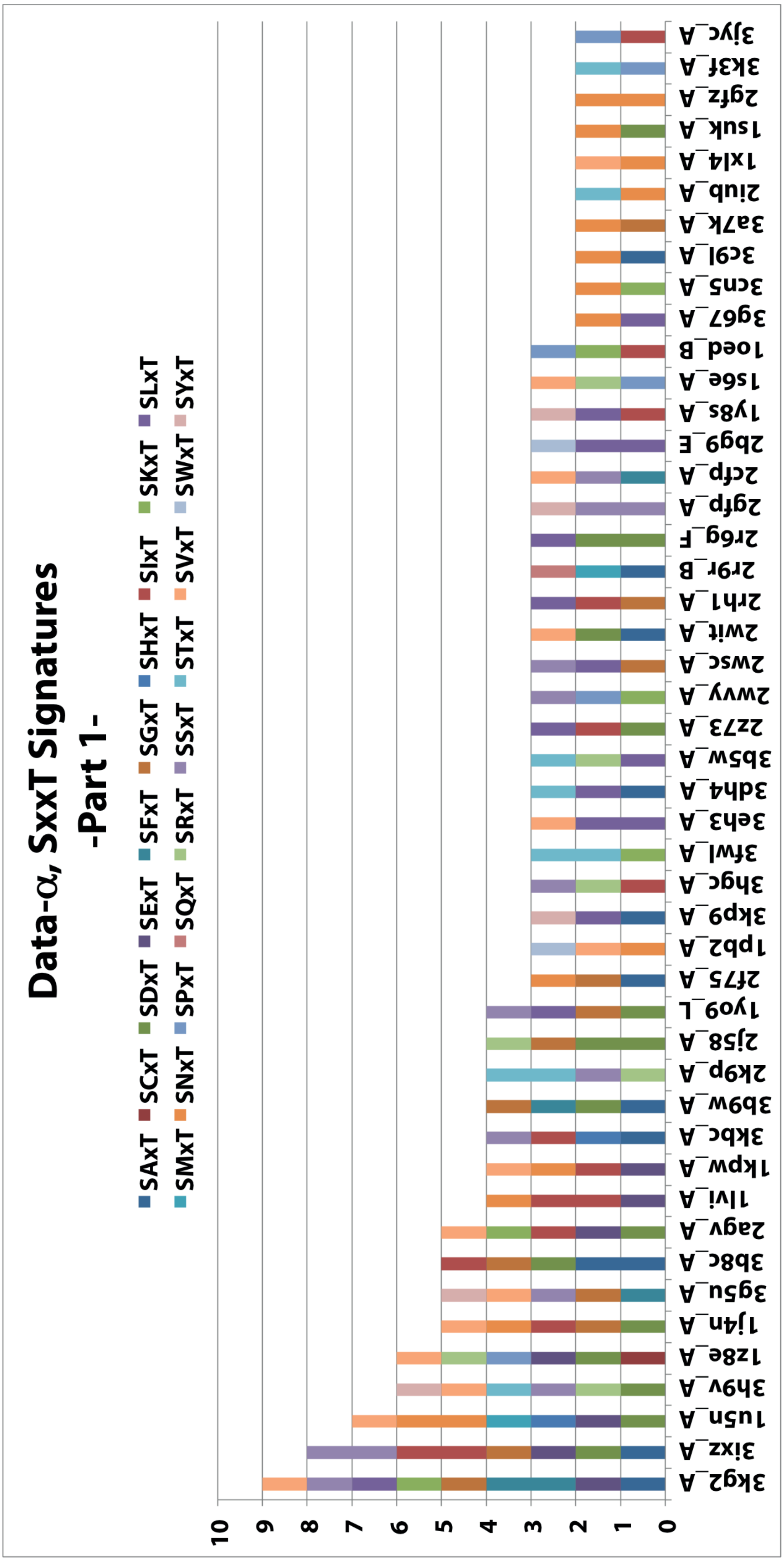

Data-α SxxS Signatures
-Part 3-
SSxS
SLxS
SAxS
SIxS
SGxS
STxS
SVxS
SFxS
SKxS
SExS
SRxS
SNxS
SPxS
SYxS
SQxS
SMxS
SDxS
SHxS
SWxS
SCxS
6
5
4
3
2
1
0
2z73_A
2acz_C
2acz_D
2yvx_A
2akh_X
2akh_Y
1p7b_A
1et2_S
1pb2_A
2wsc_H
1prc_H
2rh1_A
2rdd_B
2r6g_F
2qfi_A
2q67_A
2oar_A
1rh5_A
2b2f_A
2kog_A
2kns_A
2klu_A
2kix_A
2kih_A
2bl2_A
2k9p_A
2jwa_A
2jow_A
1kqf_B
2cpb_A
2iub_A
2ibz_I
2hyd_A
2fbw_C
2f93_A
2f75_A
2evu_A

Figure S6-D


Data-α, SxxT Signatures
-Part 1-
SAxT
SCxT
SDxT
SExT
SFxT
SGxT
SHxT
SIxT
SKxT
SLxT
SMxT
SNxT
SPxT
SQxT
SRxT
SSxT
STxT
SVxT
SWxT
SYxT
0
1
2
3
4
5
6
7
8
9
10
3kg2_A
3ixz_A
1usn_A
3h9v_A
1z8e_A
1j4n_A
3g5u_A
3b8c_A
2agy_A
1lvi_A
1kpw_A
3kbc_A
3b9w_A
2k9p_A
2js8_A
1yo9_L
2f75_A
1pb2_A
3kp9_A
3hgc_A
3fwl_A
3eh3_A
3dh4_A
3b5w_A
2z73_A
2wvy_A
2wsc_A
2wit_A
2rh1_A
2r9r_B
2r6g_F
2gfp_A
2cfp_A
2bg9_E
1y8s_A
1s6e_A
1oed_B
3g67_A
3cn5_A
3c9l_A
3a7k_A
2iub_A
1xl4_A
1suk_A
2gfz_A
3k3f_A
3jyc_A

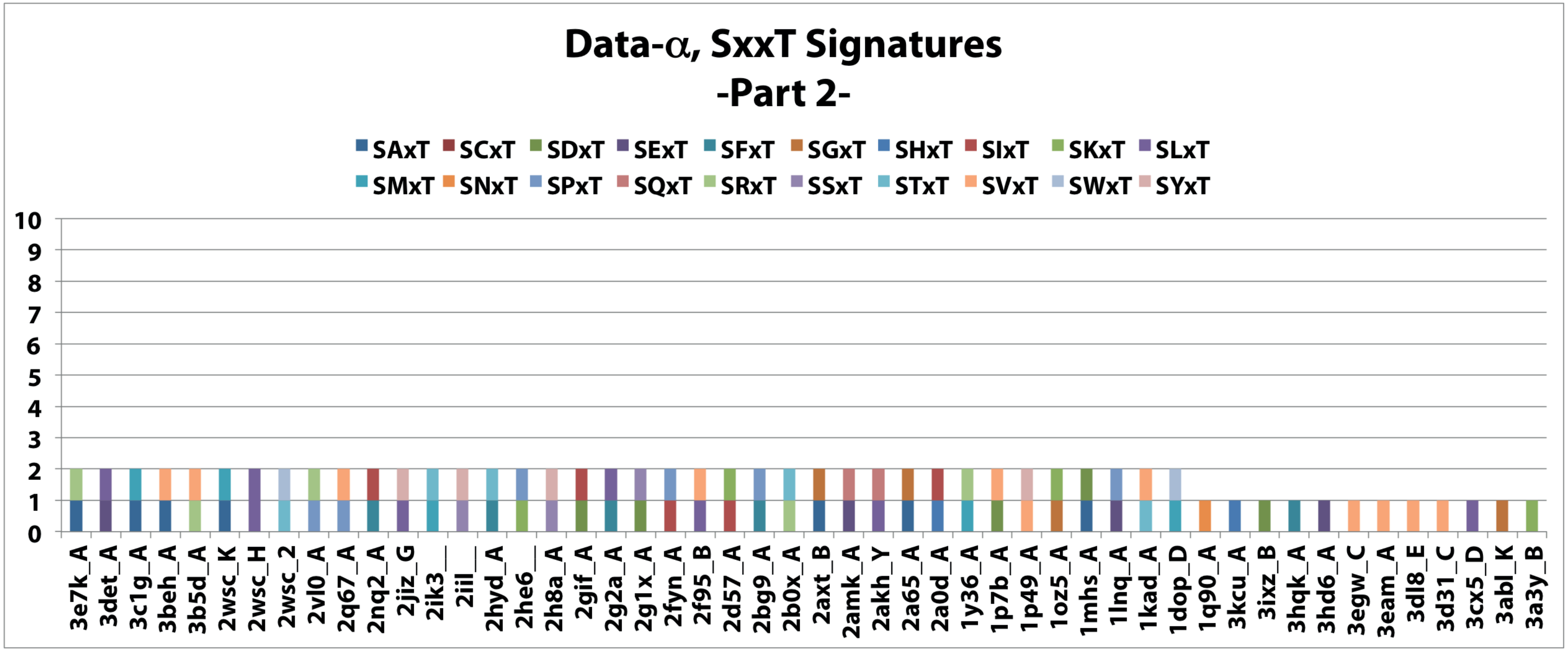

Data-α, SxxT Signatures
-Part 2-
SAxT
SCxT
SDxT
SExT
SFxT
SGxT
SHxT
SIxT
SKxT
SLxT
SMxT
SNxT
SPxT
SQxT
SRxT
SSxT
STxT
SVxT
SWxT
SYxT
10
9
8
7
6
5
4
3
2
1
0
3e7k_A
3det_A
3c1g_A
3beh_A
3b5d_A
2wsc_K
2wsc_H
2wsc_2
2vl0_A
2q67_A
2nq2_A
2jiz_G
2ik3__
2iil__
2hyd_A
2he6__
2h8a_A
2gif_A
2g2a_A
2g1x_A
2fyn_A
2f95_B
2d57_A
2bg9_A
2b0x_A
2axt_B
2amk_A
2akh_Y
2a65_A
2a0d_A
1y36_A
1p7b_A
1p49_A
1oz5_A
1mhs_A
1lnq_A
1kad_A
1dop_D
1q90_A
3kcu_A
3ixz_B
3hqk_A
3hd6_A
3egw_C
3eam_A
3dl8_E
3d31_C
3cx5_D
3abl_K
3a3y_B

Figure S6-E

Figure S6-F

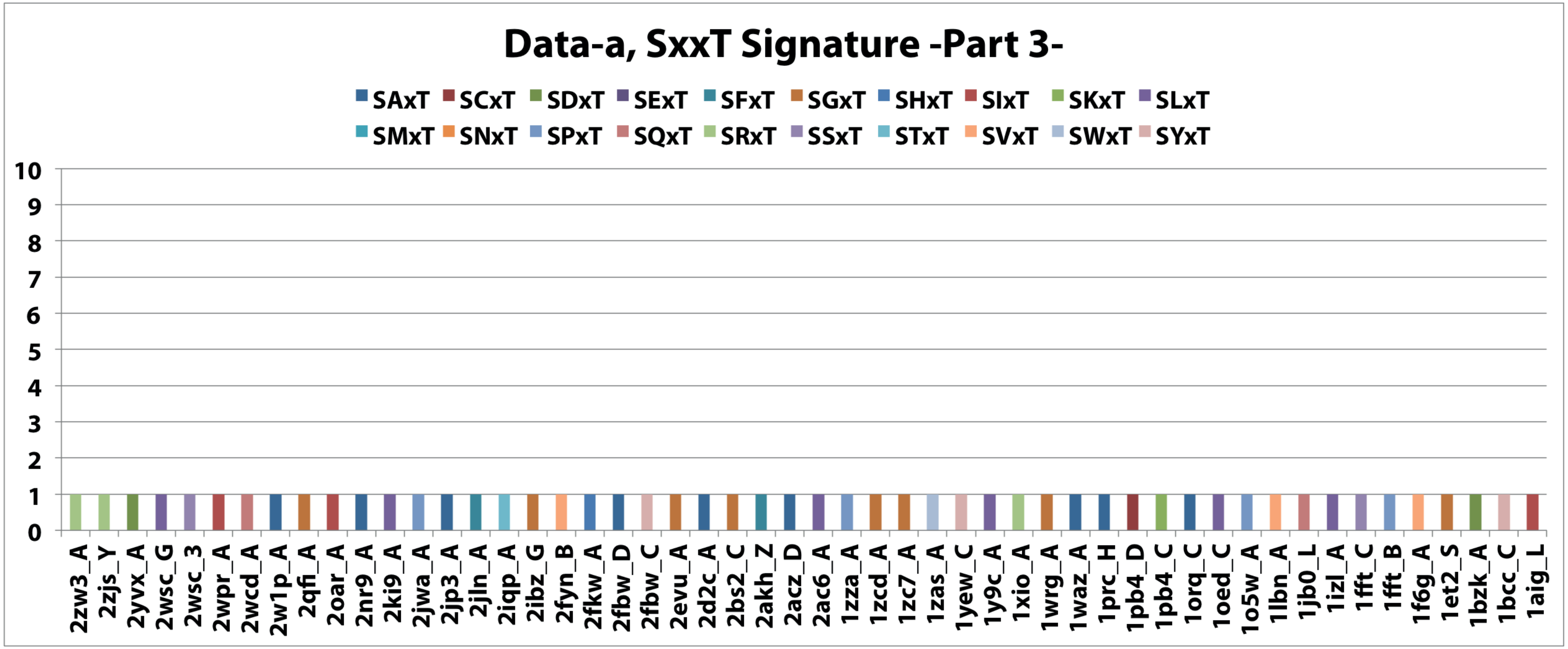

Data-a, SxxT Signature -Part 3-
SAxT
SCxT
SDxT
SExT
SFxT
SGxT
SHxT
SIxT
SKxT
SLxT
SMxT
SNxT
SPxT
SQxT
SRxT
SSxT
STxT
SVxT
SWxT
SYxT
10
9
8
7
6
5
4
3
2
1
0
2zw3_A
2zjs_Y
2yvx_A
2wsc_G
2wsc_3
2wpr_A
2wcd_A
2w1p_A
2qfi_A
2oar_A
2nr9_A
2ki9_A
2jwa_A
2jp3_A
2jln_A
2iqp_A
2ibz_G
2fyn_B
2fkw_A
2fbw_D
2fbw_C
2evu_A
2d2c_A
2bs2_C
2akh_Z
2acz_D
2ac6_A
1zza_A
1zcd_A
1zc7_A
1zas_A
1yew_C
1y9c_A
1xio_A
1wrg_A
1waz_A
1prc_H
1pb4_D
1pb4_C
1orq_C
1oed_C
1o5w_A
1lbn_A
1jb0_L
1izl_A
1fft_C
1fft_B
1f6g_A
1et2_S
1bzk_A
1bcc_C
1aig_L

Figure S6-G

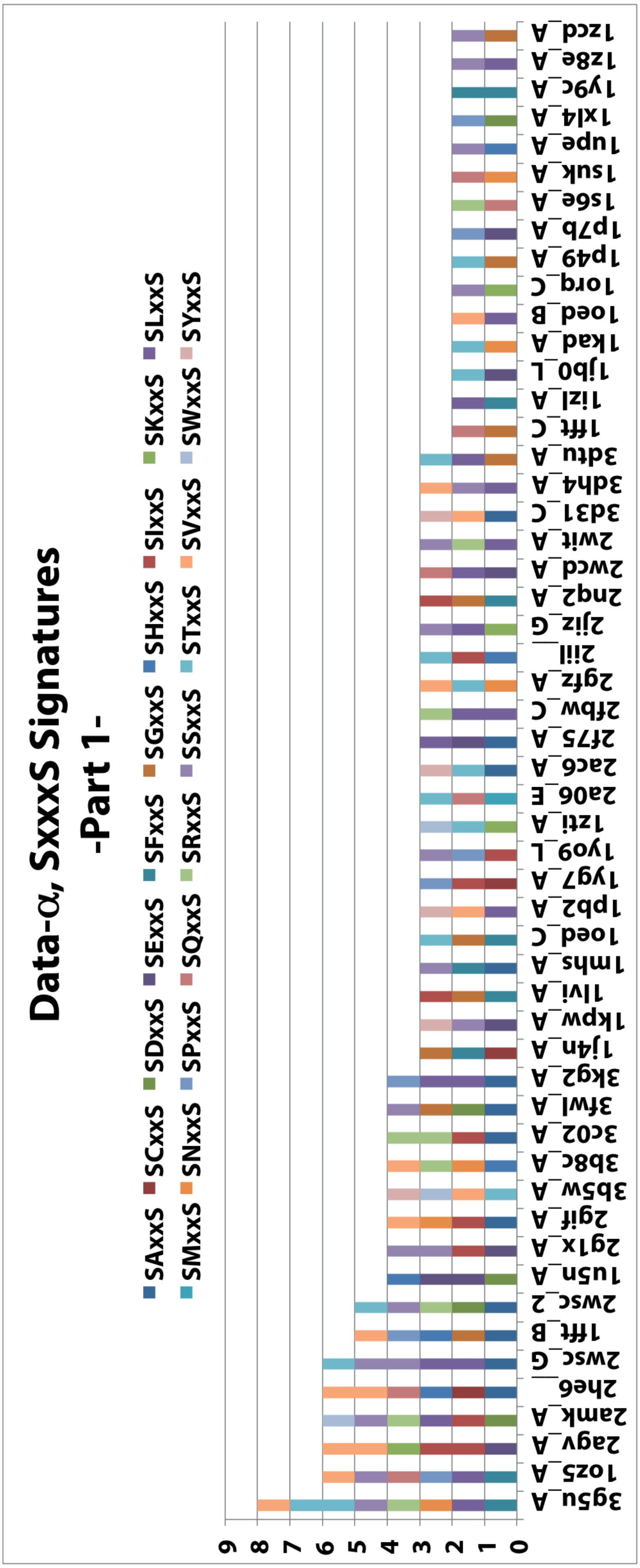

Data-α, SxxxS Signatures
-Part 1-
SAxxS
SCxxS
SDxxS
SExxS
SFxxS
SGxxS
SHxxS
SIxxS
SKxxS
SLxxS
SMxxS
SNxxS
SPxxS
SQxxS
SRxxS
SSxxS
STxxS
SVxxS
SWxxS
SYxxS
0
1
2
3
4
5
6
7
8
9
3g5u_A
1oz5_A
2agv_A
2amk_A
2he6_
2wsc_G
1fft_B
2wsc_2
1u5n_A
2g1x_A
2gif_A
3b5w_A
3b8c_A
3c02_A
3fwl_A
3kg2_A
1j4n_A
1kpw_A
1lvi_A
1mhs_A
1oed_C
1pb2_A
1yg7_A
1yo9_L
1zti_A
2a06_E
2ac6_A
2f75_A
2fbw_C
2gfz_A
2iil_
2jiz_G
2nq2_A
2wcd_A
2wit_A
3d31_C
3dh4_A
3dtu_A
1fft_C
1izi_A
1jb0_L
1kad_A
1oed_B
1org_C
1p49_A
1p7b_A
1s6e_A
1suk_A
1upe_A
1xl4_A
1y9c_A
1z8e_A
1zcd_A

Figure S6-H

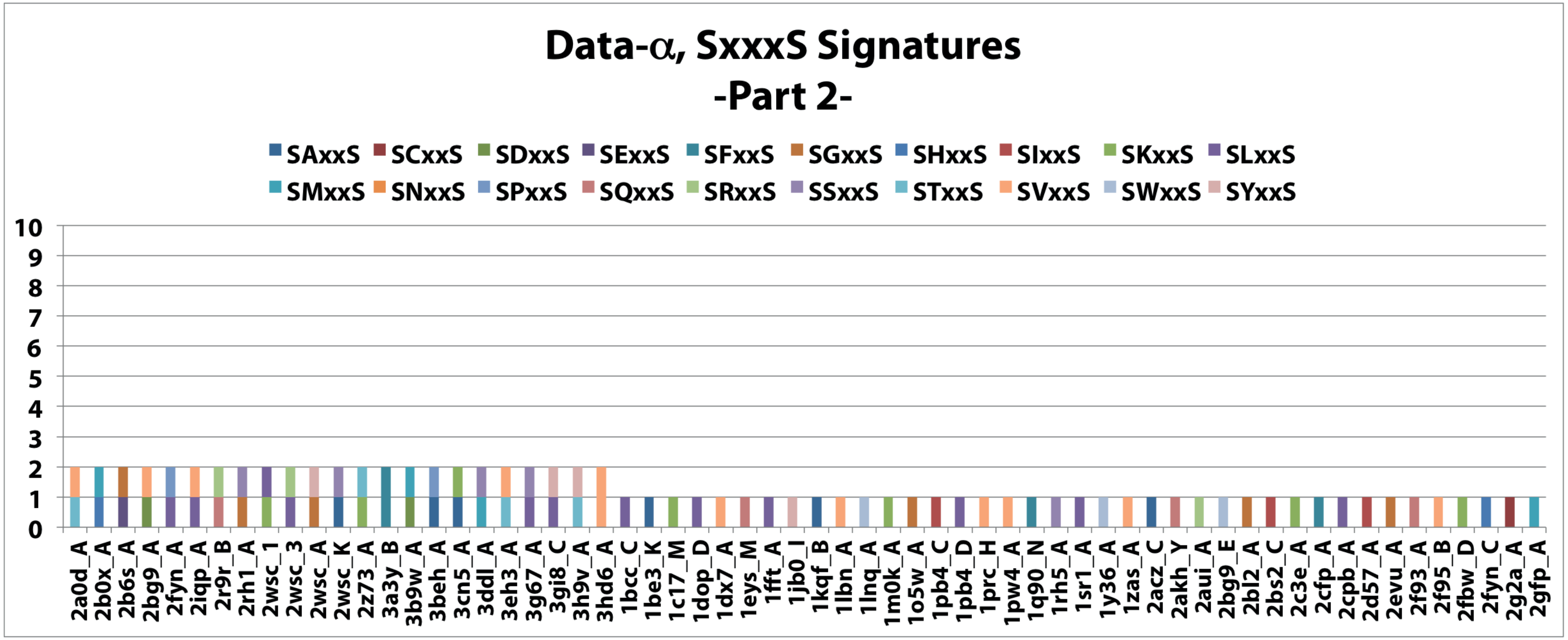

Data-α, SxxxS Signatures
-Part 2-
SAxxS
SCxxS
SDxxS
SExxS
SFxxS
SGxxS
SHxxS
SIxxS
SKxxS
SLxxS
SMxxS
SNxxS
SPxxS
SQxxS
SRxxS
SSxxS
STxxS
SVxxS
SWxxS
SYxxS
10
9
8
7
6
5
4
3
2
1
0
2a0d_A
2b0x_A
2b6s_A
2bg9_A
2fyn_A
2iqp_A
2r9r_B
2rh1_A
2wsc_1
2wsc_3
2wsc_A
2wsc_K
2z73_A
3a3y_B
3b9w_A
3beh_A
3cn5_A
3ddl_A
3eh3_A
3g67_A
3gi8_C
3h9v_A
3hd6_A
1bcc_C
1be3_K
1c17_M
1dop_D
1dx7_A
1eys_M
1fft_A
1jb0_I
1kqf_B
1lbn_A
1lnq_A
1m0k_A
1o5w_A
1pb4_C
1pb4_D
1prc_H
1pw4_A
1q90_N
1rh5_A
1sr1_A
1y36_A
1zas_A
2acz_C
2akh_Y
2aui_A
2bg9_E
2bl2_A
2bs2_C
2c3e_A
2cfp_A
2cpb_A
2d57_A
2evu_A
2f93_A
2f95_B
2fbw_D
2fyn_C
2g2a_A
2gfp_A

Figure S6-I

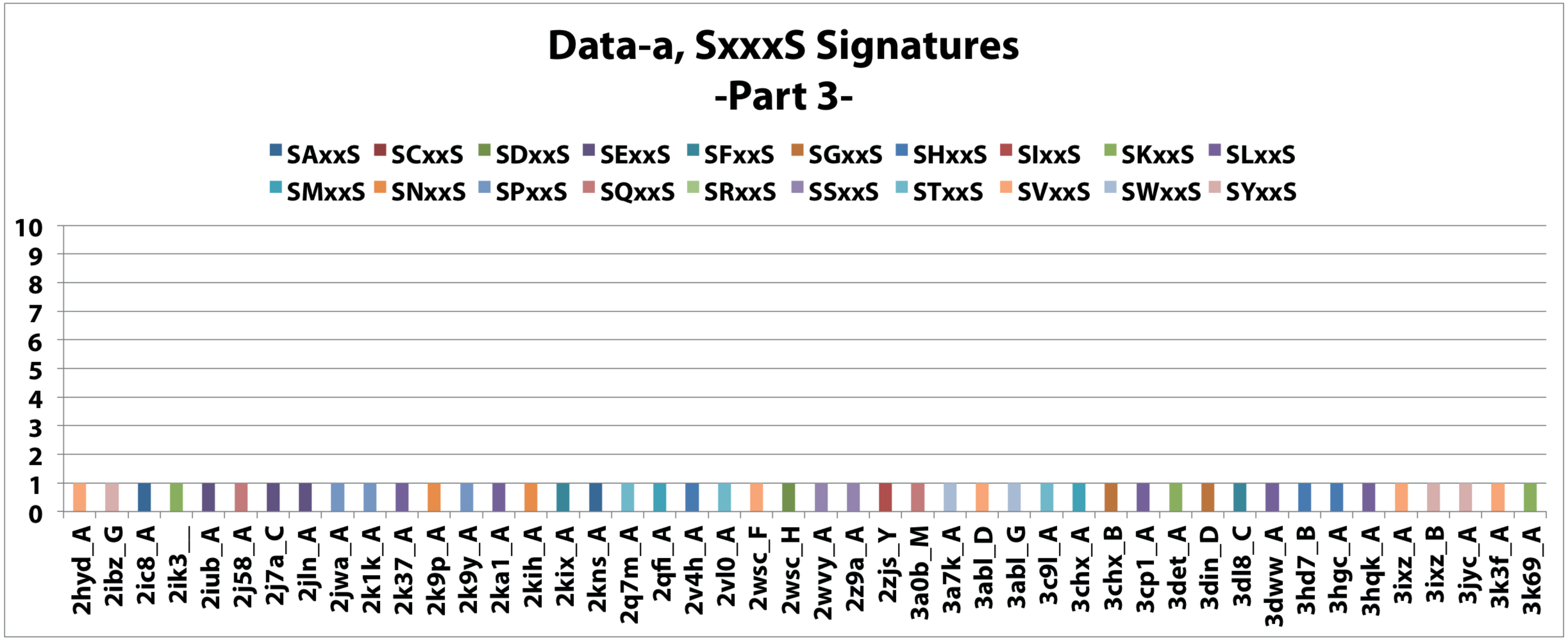

Data-a, SxxxS Signatures
-Part 3-
SAxxS
SCxxS
SDxxS
SExxS
SFxxS
SGxxS
SHxxS
SIxxS
SKxxS
SLxxS
SMxxS
SNxxS
SPxxS
SQxxS
SRxxS
SSxxS
STxxS
SVxxS
SWxxS
SYxxS
10
9
8
7
6
5
4
3
2
1
0
2hyd_A
2ibz_G
2ic8_A
2ik3__
2iub_A
2j58_A
2j7a_C
2jln_A
2jwa_A
2k1k_A
2k37_A
2k9p_A
2k9y_A
2ka1_A
2kih_A
2kix_A
2kns_A
2q7m_A
2qfi_A
2v4h_A
2vl0_A
2wsc_F
2wsc_H
2wvy_A
2z9a_A
2zjs_Y
3a0b_M
3a7k_A
3abl_D
3abl_G
3c9l_A
3chx_A
3chx_B
3cp1_A
3det_A
3din_D
3dl8_C
3dww_A
3hd7_B
3hgc_A
3hqk_A
3ixz_A
3ixz_B
3jyc_A
3k3f_A
3k69_A

Figure S6-J

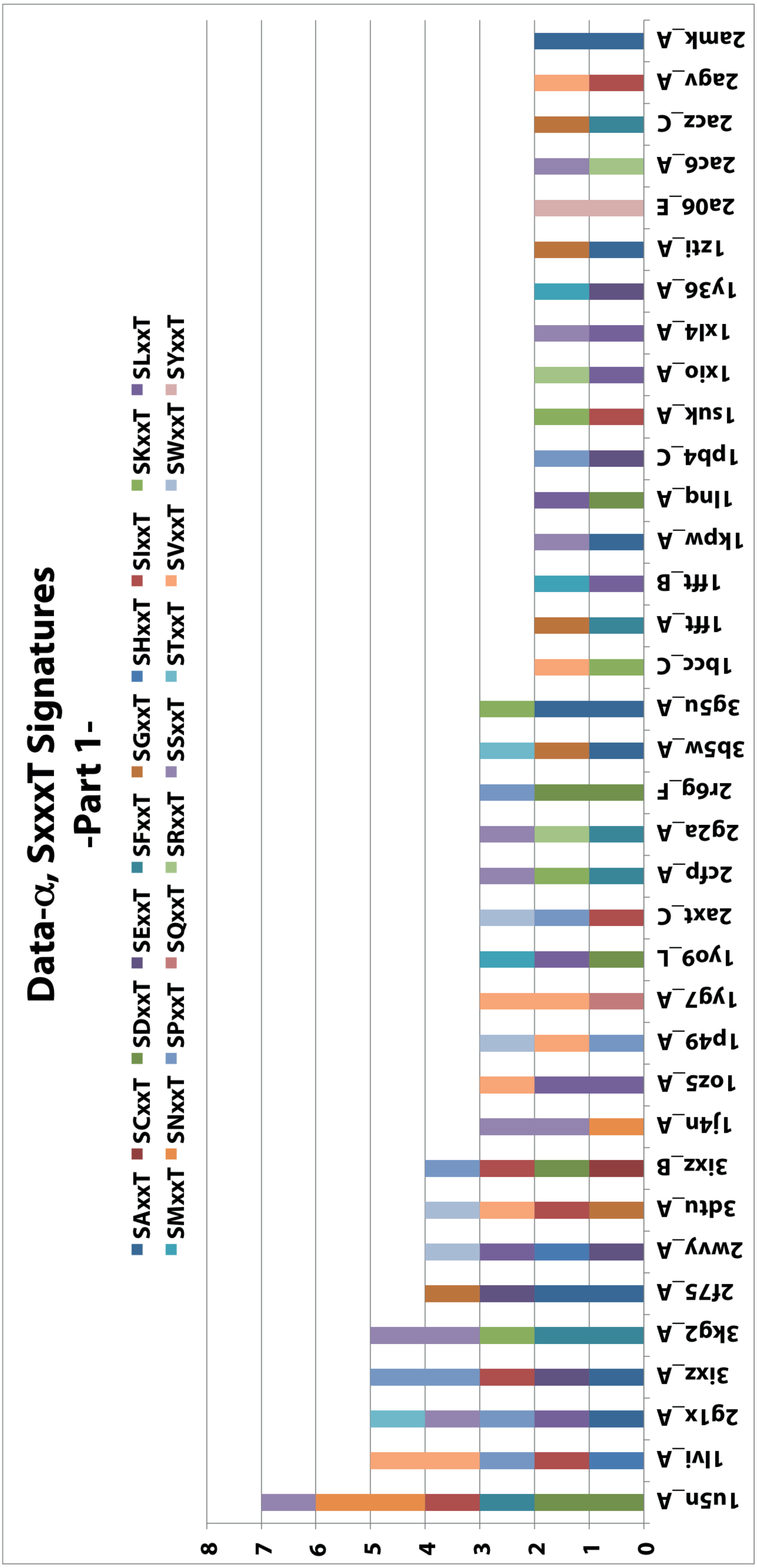
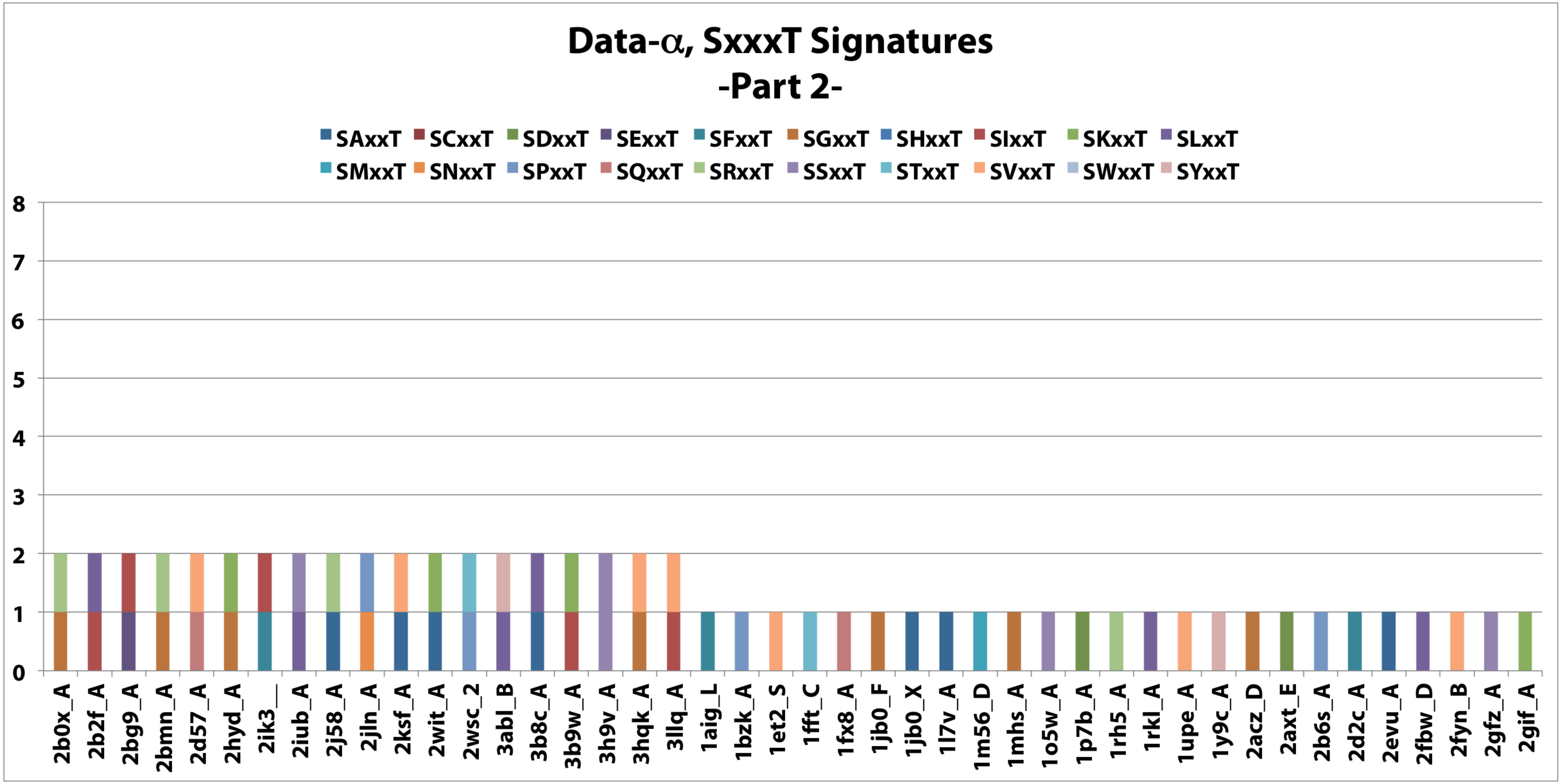

Data-α, SxxxT Signatures
-Part 1-
SAxxT
SCxxT
SDxxT
SExxT
SFxxT
SGxxT
SHxxT
SIxxT
SKxxT
SLxxT
SMxxT
SNxxT
SPxxT
SQxxT
SRxxT
SSxxT
STxxT
SVxxT
SWxxT
SYxxT
0
1
2
3
4
5
6
7
8
1u5n_A
1lvi_A
2g1x_A
3ixz_A
3kg2_A
2f75_A
2wvy_A
3dtu_A
3ixz_B
1j4n_A
1oz5_A
1p49_A
1yg7_A
1yo9_L
2axt_C
2cfp_A
2g2a_A
2r6g_F
3b5w_A
3g5u_A
1bcc_C
1fft_A
1fft_B
1kpw_A
1lnq_A
1pb4_C
1suk_A
1xio_A
1xl4_A
1y36_A
1zti_A
2a06_E
2ac6_A
2acz_C
2agy_A
2amk_A

Figure S6-K


Data-α, SxxxT Signatures
-Part 2-
SAxxT
SCxxT
SDxxT
SExxT
SFxxT
SGxxT
SHxxT
SIxxT
SKxxT
SLxxT
SMxxT
SNxxT
SPxxT
SQxxT
SRxxT
SSxxT
STxxT
SVxxT
SWxxT
SYxxT
8
7
6
5
4
3
2
1
0
2b0x_A
2b2f_A
2bg9_A
2bmn_A
2d57_A
2hyd_A
2ik3__
2iub_A
2j58_A
2jln_A
2ksf_A
2wit_A
2wsc_2
3abl_B
3b8c_A
3b9w_A
3h9v_A
3hqk_A
3llq_A
1aig_L
1bzk_A
1et2_S
1fft_C
1fx8_A
1jb0_F
1jb0_X
1l7v_A
1m56_D
1mhs_A
1o5w_A
1p7b_A
1rh5_A
1rkl_A
1upe_A
1y9c_A
2acz_D
2axt_E
2b6s_A
2d2c_A
2evu_A
2fbw_D
2fyn_B
2gfz_A
2gif_A

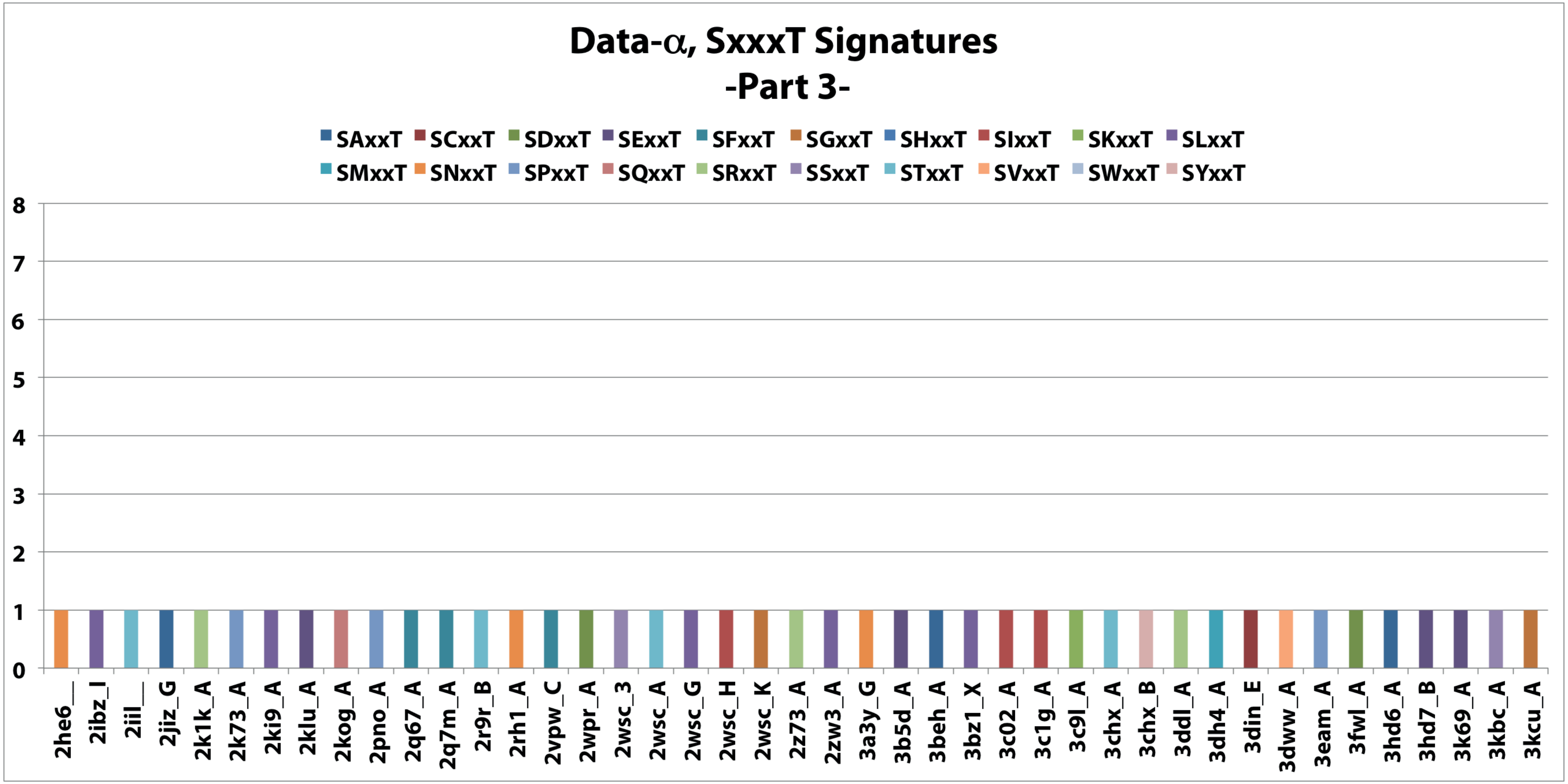

Data-α, SxxxT Signatures
-Part 3-
SAxxT
SCxxT
SDxxT
SExxT
SFxxT
SGxxT
SHxxT
SIxxT
SKxxT
SLxxT
SMxxT
SNxxT
SPxxT
SQxxT
SRxxT
SSxxT
STxxT
SVxxT
SWxxT
SYxxT
8
7
6
5
4
3
2
1
0
2he6__
2ibz_I
2iil__
2jiz_G
2k1k_A
2k73_A
2ki9_A
2klu_A
2kog_A
2pno_A
2q67_A
2q7m_A
2r9r_B
2rh1_A
2vpw_C
2wpr_A
2wsc_3
2wsc_A
2wsc_G
2wsc_H
2wsc_K
2z73_A
2zw3_A
3a3y_G
3b5d_A
3beh_A
3bz1_X
3c02_A
3c1g_A
3c9l_A
3chx_A
3chx_B
3ddl_A
3dh4_A
3din_E
3dww_A
3eam_A
3fwl_A
3hd6_A
3hd7_B
3k69_A
3kbc_A
3kcu_A

Figure S6-L

Figure S6-M

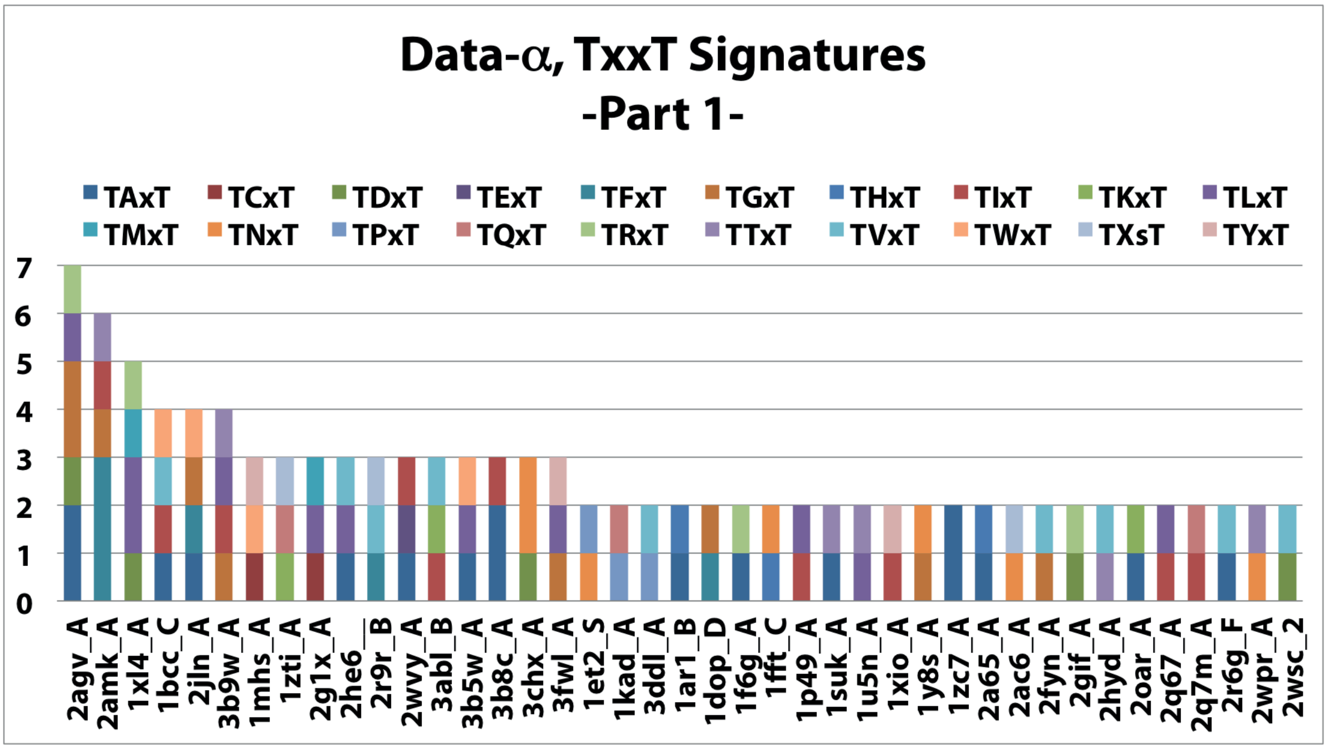
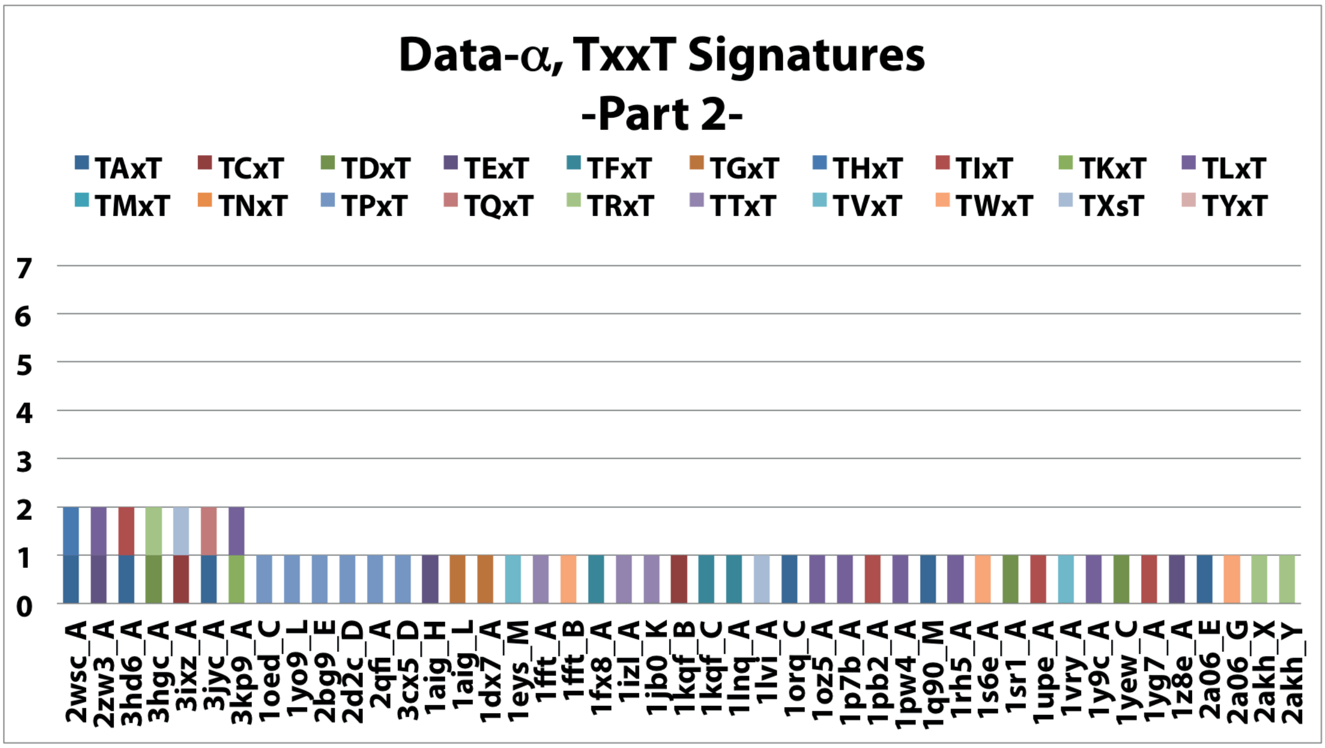


Figure S6-N



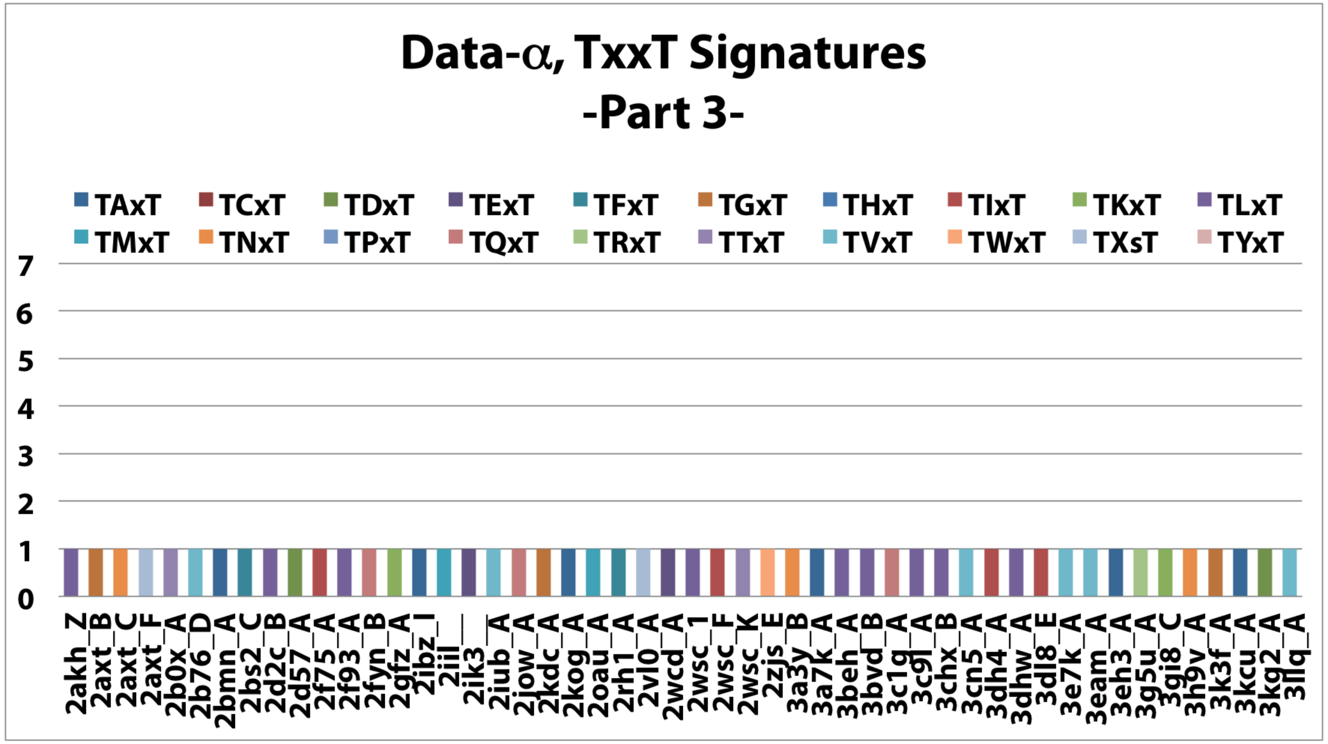

Data-α, TxxT Signatures
-Part 3-
TAxT
TCxT
TDxT
TExT
TFxT
TGxT
THxT
TIxT
TKxT
TLxT
TMxT
TNxT
TPxT
TQxT
TRxT
TTxT
TVxT
TWxT
TXsT
TYxT
7
6
5
4
3
2
1
0
2akh_Z
2axt_B
2axt_C
2axt_F
2b0x_A
2b76_D
2bmn_A
2bs2_C
2d2c_B
2d57_A
2f75_A
2f93_A
2fyn_B
2gfz_A
2ibz_I
2iil_
2ik3_
2iub_A
2jow_A
2kdc_A
2kog_A
2oau_A
2rh1_A
2vl0_A
2wcd_A
2wsc_1
2wsc_F
2wsc_K
2zjs_E
3a3y_B
3a7k_A
3beh_A
3bvd_B
3c1g_A
3c9l_A
3chx_B
3cn5_A
3dh4_A
3dhw_A
3dl8_E
3e7k_A
3eam_A
3eh3_A
3g5u_A
3gi8_C
3h9v_A
3k3f_A
3kcu_A
3kg2_A
3llq_A

Figure S6-O

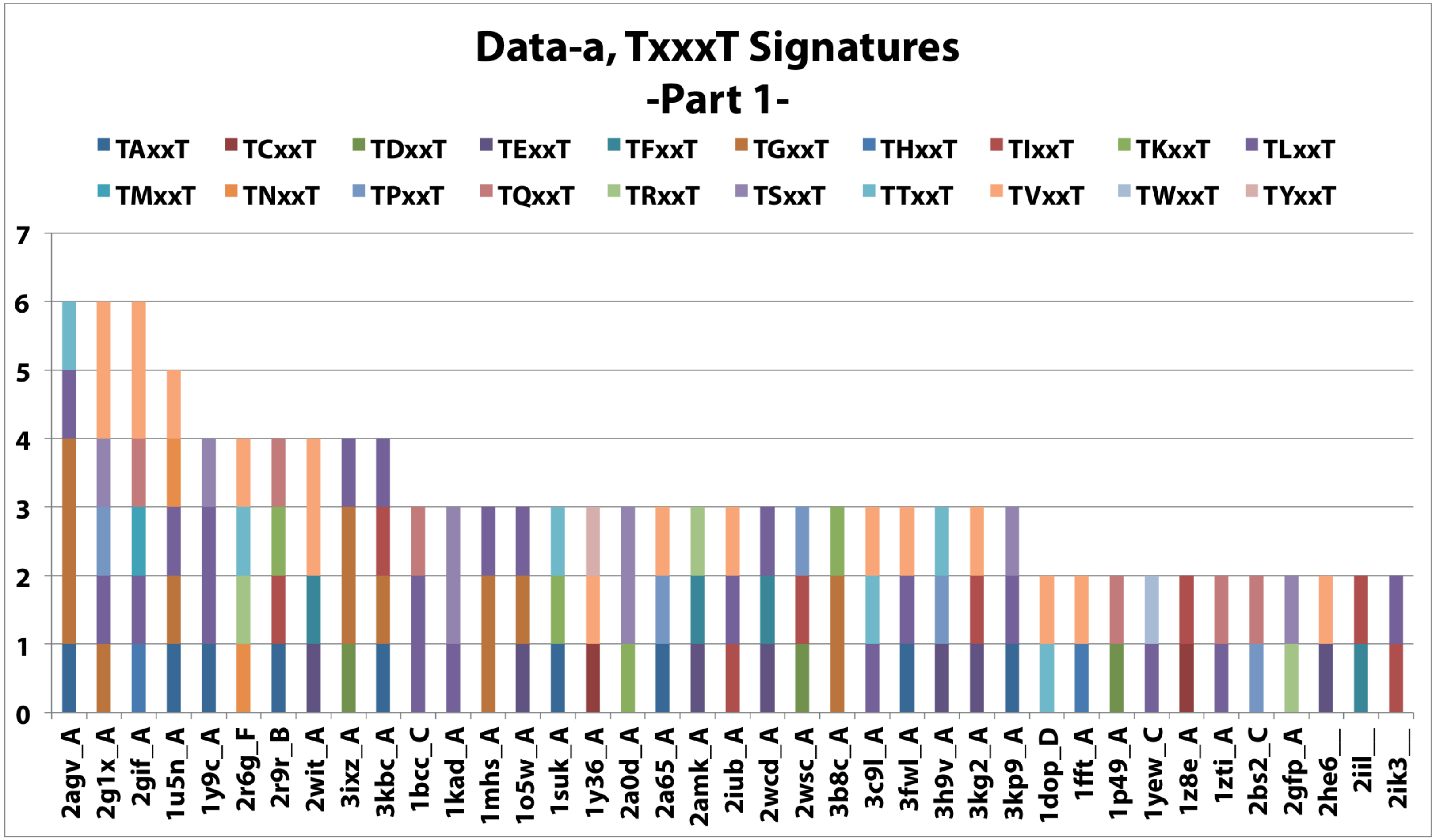
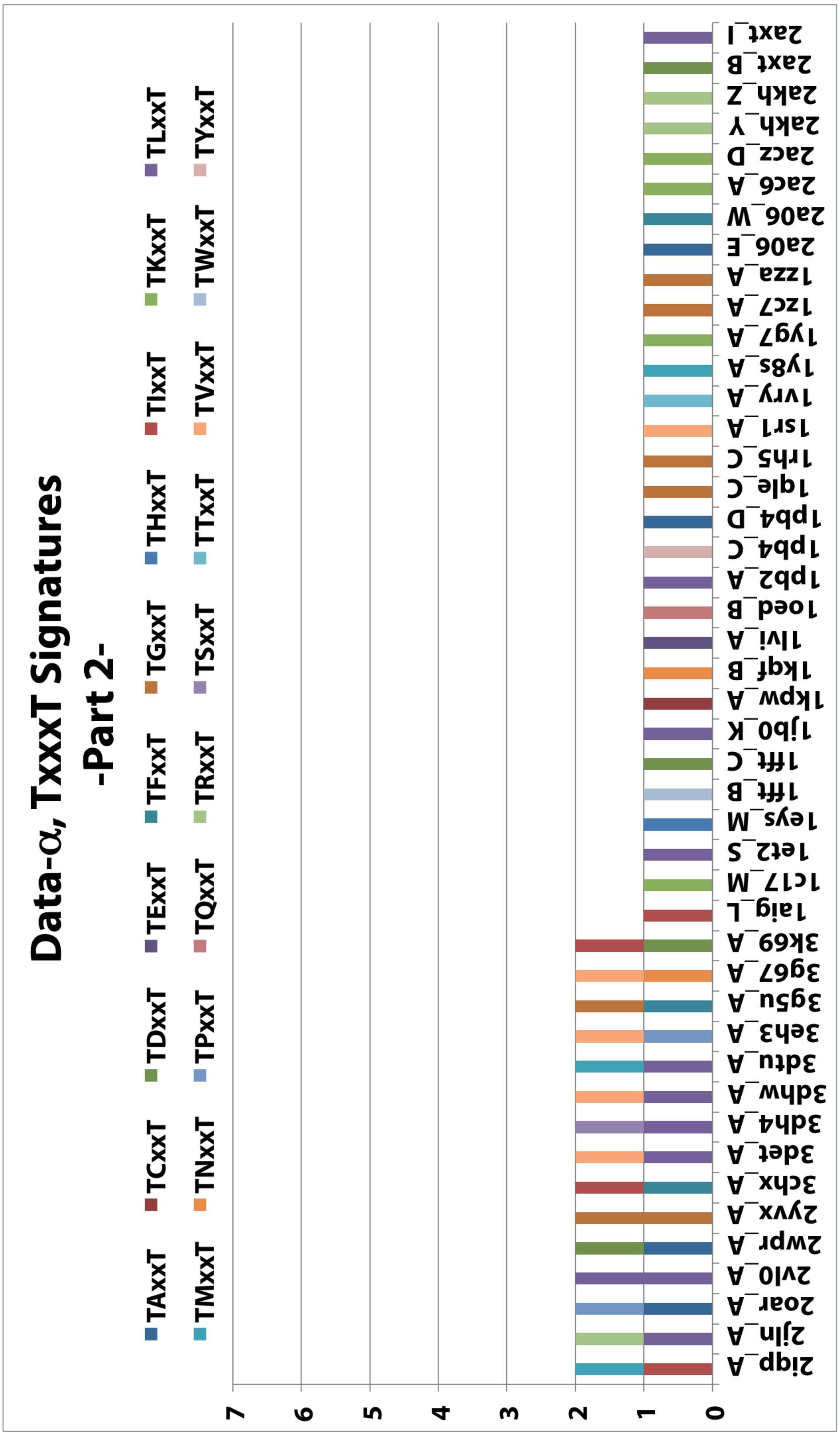

Data-a, TxxxT Signatures
-Part 1-
TAxxT
TCxxT
TDxxT
TExxT
TFxxT
TGxxT
THxxT
TIxxT
TKxxT
TLxxT
TMxxT
TNxxT
TPxxT
TQxxT
TRxxT
TSxxT
TTxxT
TVxxT
TWxxT
TYxxT
7
6
5
4
3
2
1
0
2agv_A
2g1x_A
2gif_A
1u5n_A
1y9c_A
2r6g_F
2r9r_B
2wit_A
3ixz_A
3kbc_A
1bcc_C
1kad_A
1mhs_A
1o5w_A
1suk_A
1y36_A
2a0d_A
2a65_A
2amk_A
2iub_A
2wcd_A
2wsc_A
3b8c_A
3c9l_A
3fwl_A
3h9v_A
3kg2_A
3kp9_A
1dop_D
1fft_A
1p49_A
1yew_C
1z8e_A
1zti_A
2bs2_C
2gfp_A
2he6__
2iil__
2ik3__

Figure S6-P

Figure S6-R


Data-α, TxxxT Signatures
-Part 2-
TAxxT
TCxxT
TDxxT
TExxT
TFxxT
TGxxT
THxxT
TIxxT
TKxxT
TLxxT
TMxxT
TNxxT
TPxxT
TQxxT
TRxxT
TSxxT
TTxxT
TVxxT
TWxxT
TYxxT
0
1
2
3
4
5
6
7
2iqp_A
2jln_A
2oar_A
2vl0_A
2wpr_A
2yvx_A
3chx_A
3det_A
3dh4_A
3dhw_A
3dtu_A
3eh3_A
3g5u_A
3g67_A
3k69_A
1aig_L
1c17_M
1et2_S
1eys_M
1fft_B
1fft_C
1jb0_K
1kpw_A
1kqf_B
1lvi_A
1oed_B
1pb2_A
1pb4_C
1pb4_D
1qle_C
1rh5_C
1sr1_A
1vry_A
1y8s_A
1yg7_A
1zc7_A
1zza_A
2a06_E
2a06_W
2ac6_A
2acz_D
2akh_Y
2akh_Z
2axt_B
2axt_I

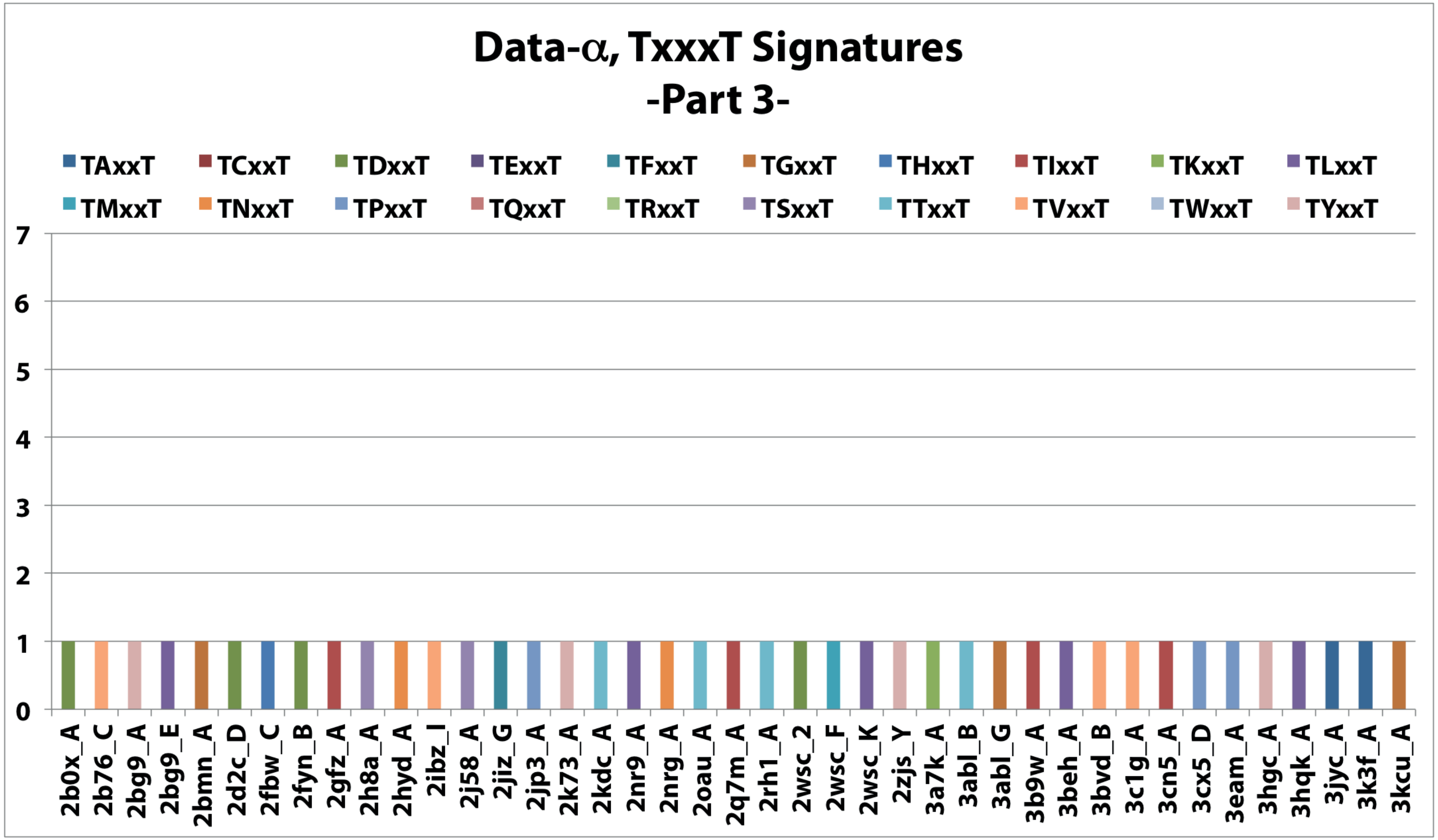

Data-α, TxxxT Signatures
-Part 3-
TAxxT
TCxxT
TDxxT
TExxT
TFxxT
TGxxT
THxxT
TIxxT
TKxxT
TLxxT
TMxxT
TNxxT
TPxxT
TQxxT
TRxxT
TSxxT
TTxxT
TVxxT
TWxxT
TYxxT
7
6
5
4
3
2
1
0
2b0x_A
2b76_C
2bg9_A
2bg9_E
2bmn_A
2d2c_D
2fbw_C
2fyn_B
2gfz_A
2h8a_A
2hyd_A
2ibz_I
2j58_A
2jiz_G
2jp3_A
2k73_A
2kdc_A
2nr9_A
2nrg_A
2oau_A
2q7m_A
2rh1_A
2wsc_2
2wsc_F
2wsc_K
2zjs_Y
3a7k_A
3abl_B
3abl_G
3b9w_A
3beh_A
3bvd_B
3c1g_A
3cn5_A
3cx5_D
3eam_A
3hgc_A
3hqk_A
3jyc_A
3k3f_A
3kcu_A

Figure S7-A

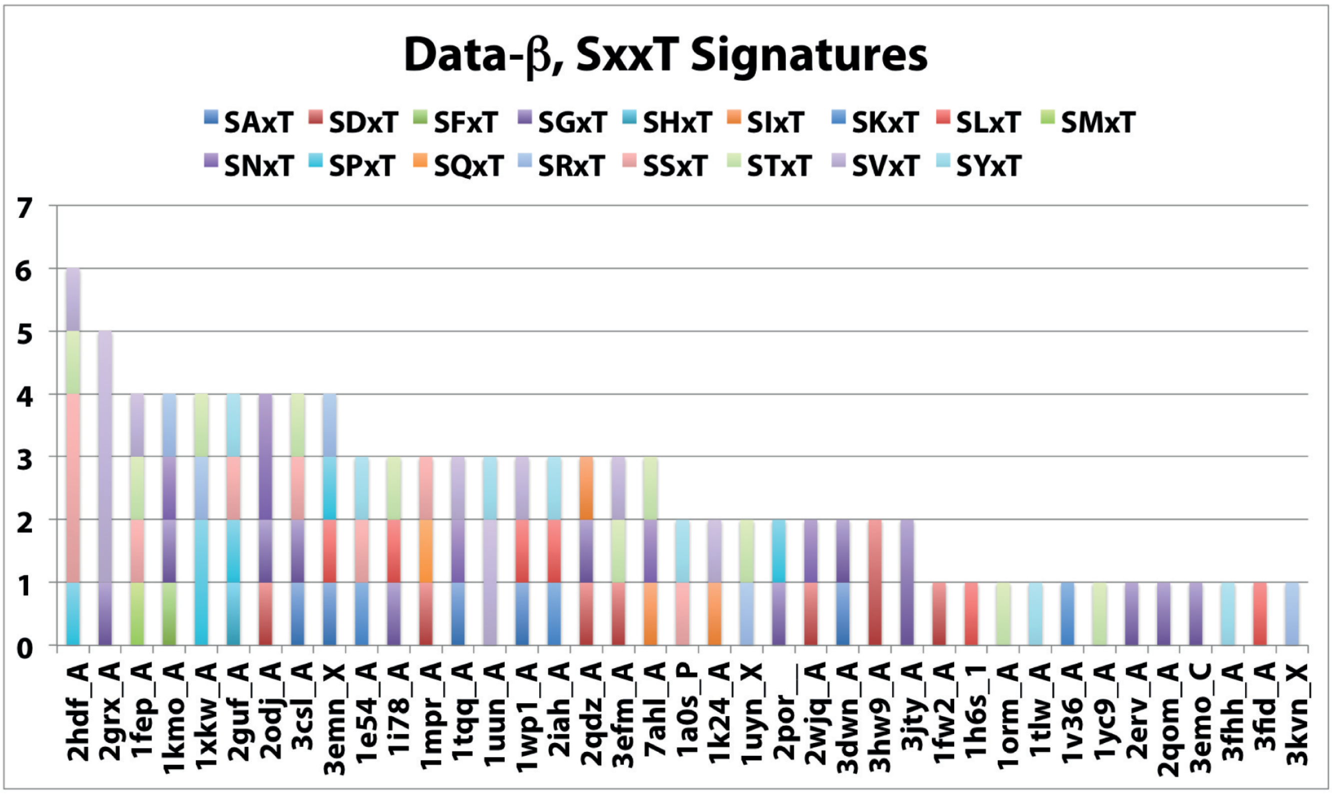

Data-β, SxxT Signatures
SAxT
SDxT
SFxT
SGxT
SHxT
SIxT
SKxT
SLxT
SMxT
SNxT
SPxT
SQxT
SRxT
SSxT
STxT
SVxT
SYxT
0
1
2
3
4
5
6
7
2hdf_A
2grx_A
1fep_A
1kmo_A
1xkw_A
2guf_A
2odj_A
3csl_A
3emn_X
1e54_A
1i78_A
1mpr_A
1tqq_A
1uun_A
1wp1_A
2iah_A
2qdz_A
3efm_A
7ahl_A
1a0s_P
1k24_A
1uyn_X
2por__A
2wjq_A
3dwn_A
3hw9_A
3jty_A
1fw2_A
1h6s_1
1orm_A
1tlw_A
1v36_A
1yc9_A
2erv_A
2qom_A
3emo_C
3fhh_A
3fid_A
3kvn_X

Figure S7-B

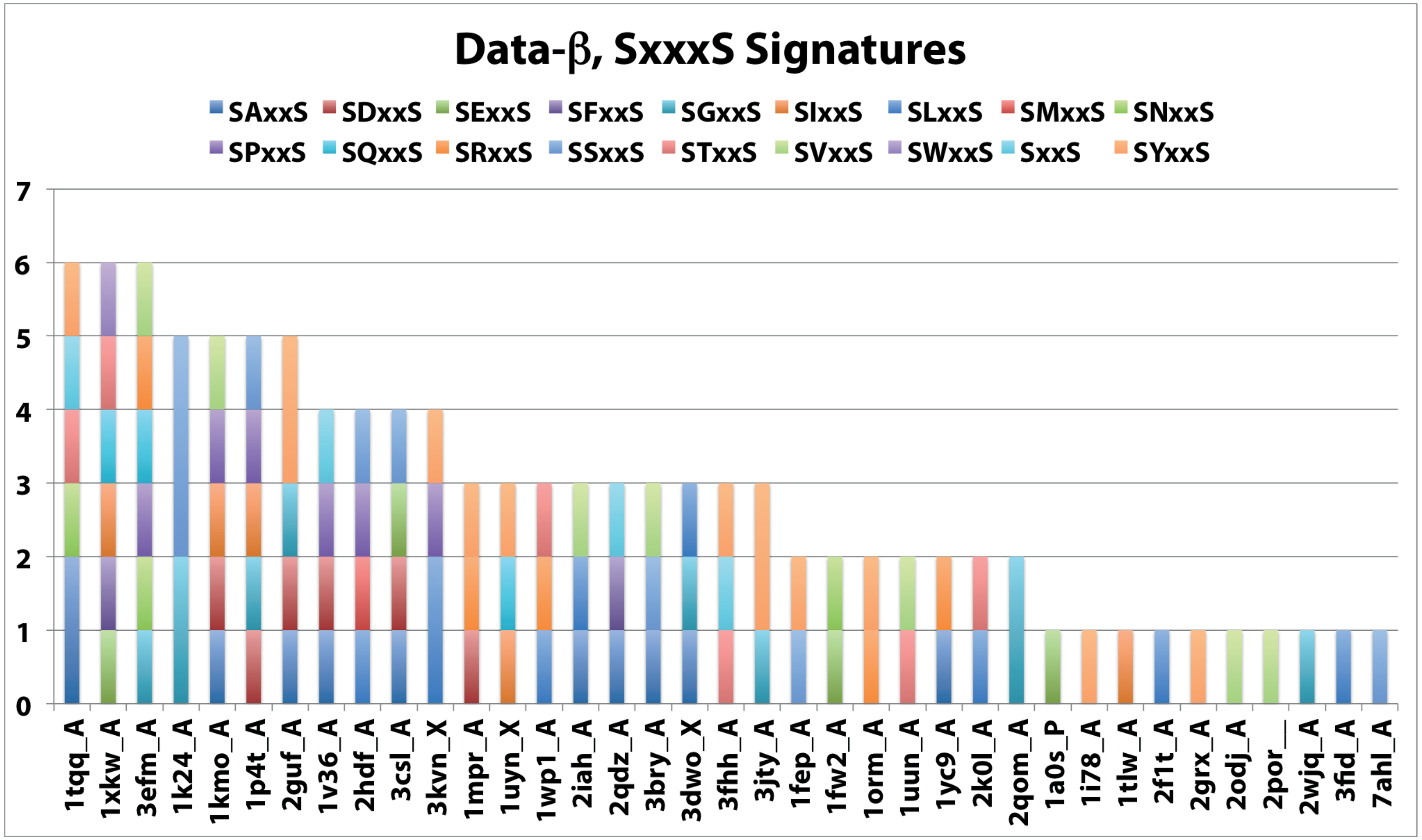

Data-β, SxxxS Signatures
SAxxS
SDxxS
SExxS
SFxxS
SGxxS
SIxxS
SLxxS
SMxxS
SNxxS
SPxxS
SQxxS
SRxxS
SSxxS
STxxS
SVxxS
SWxxS
SxxS
SYxxS
0
1
2
3
4
5
6
7
1tqq_A
1xkw_A
3efm_A
1k24_A
1kmo_A
1p4t_A
2guf_A
1v36_A
2hdf_A
3csl_A
3kvn_X
1mpr_A
1uyn_X
1wp1_A
2iah_A
2qdz_A
3bry_A
3dwo_X
3fhh_A
3jty_A
1fep_A
1fw2_A
1orm_A
1uun_A
1yc9_A
2k0l_A
2qom_A
1a0s_P
1i78_A
1tlw_A
2f1t_A
2grx_A
2odj_A
2por__
2wjq_A
3fid_A
7ahl_A

Figure S7-C

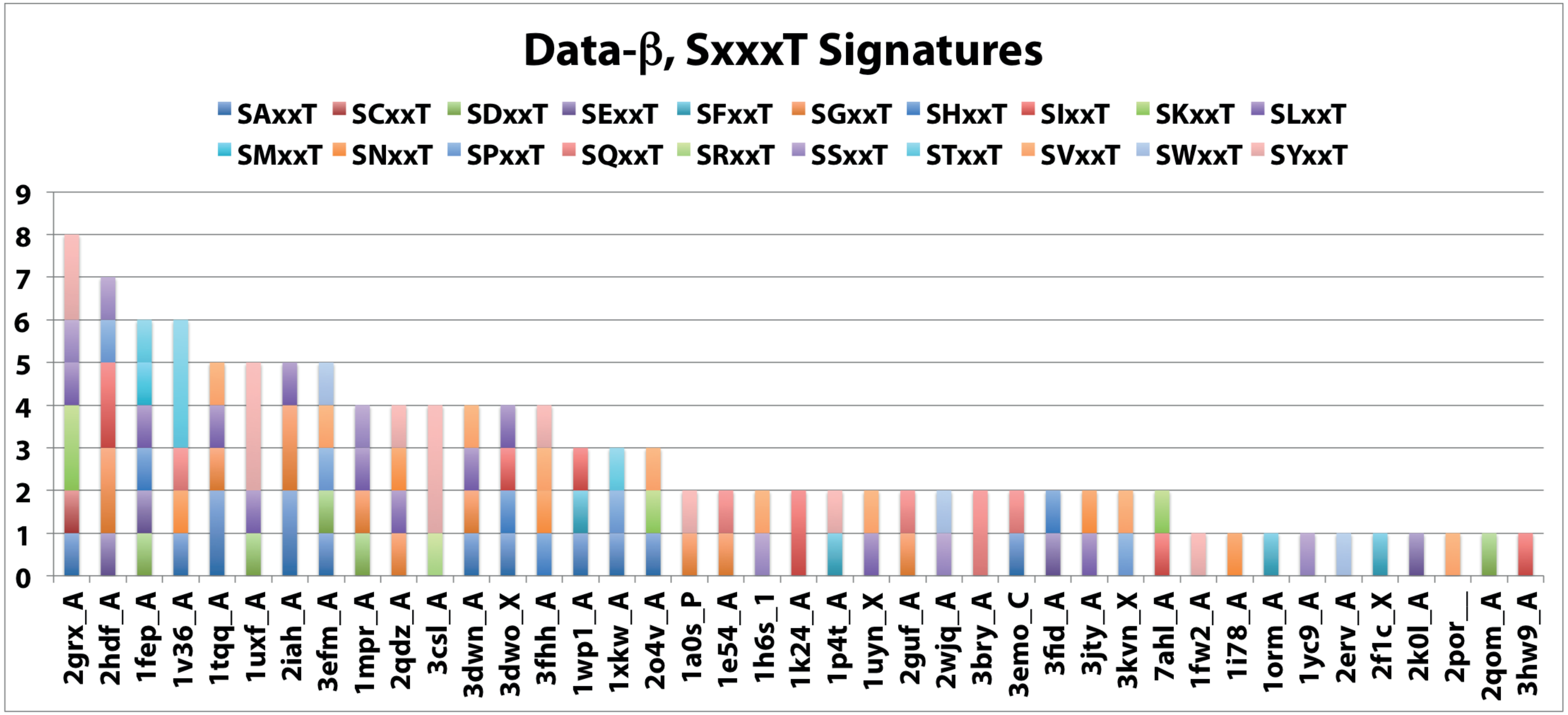
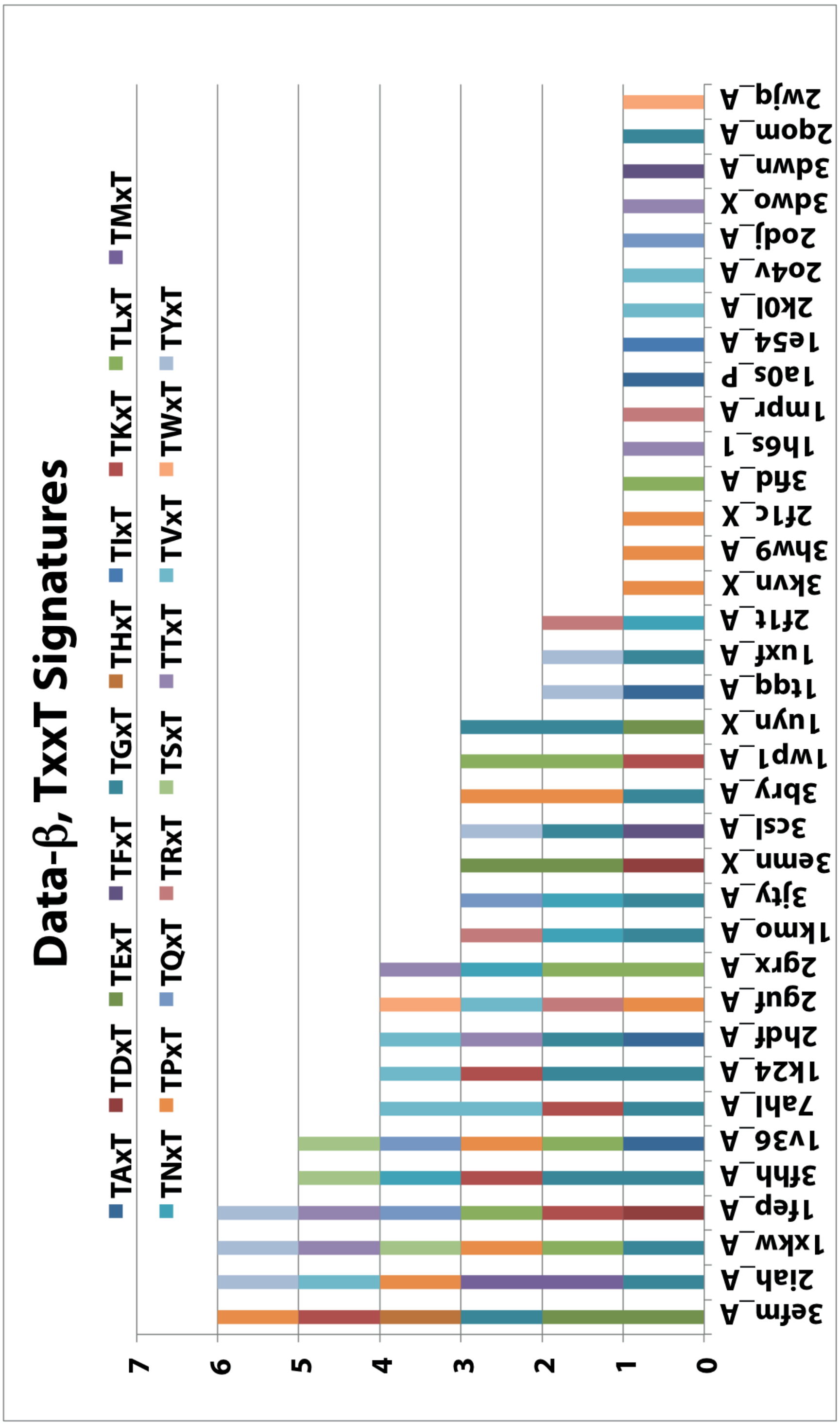

Data-β, SxxxT Signatures
SAxxT
SCxxT
SDxxT
SExxT
SFxxT
SGxxT
SHxxT
SIxxT
SKxxT
SLxxT
SMxxT
SNxxT
SPxxT
SQxxT
SRxxT
SSxxT
STxxT
SVxxT
SWxxT
SYxxT
0
1
2
3
4
5
6
7
8
9
2grx_A
2hdf_A
1fep_A
1v36_A
1tqq_A
1uxf_A
2iah_A
3efm_A
1mpr_A
2qdz_A
3csl_A
3dwn_A
3dwo_X
3fhh_A
1wp1_A
1xkw_A
2o4v_A
1a0s_P
1e54_A
1h6s_1
1k24_A
1p4t_A
1uyn_X
2guf_A
2wjq_A
3bry_A
3emo_C
3fid_A
3jty_A
3kvn_X
7ahl_A
1fw2_A
1i78_A
1orm_A
1yc9_A
2erv_A
2f1c_X
2k0l_A
2por__
2qom_A
3hw9_A

Figure S7-D


Data-β, TxxT Signatures
TAxT
TDxT
TExT
TFxT
TGxT
THxT
TIxT
TKxT
TLxT
TMxT
TNxT
TPxT
TQxT
TRxT
TSxT
TTxT
TVxT
TWxT
TYxT
0
1
2
3
4
5
6
7
3efm_A
2iah_A
1xkw_A
1fep_A
3fhh_A
1v36_A
7ahl_A
1k24_A
2hdf_A
2guf_A
2grx_A
1kmo_A
3jty_A
3emn_X
3csl_A
3bry_A
1wp1_A
1uyn_X
1tgq_A
1uxf_A
2f1t_A
3kvn_X
3hw9_A
2f1c_X
3fid_A
1h6s_1
1mpr_A
1a0s_P
1e54_A
2k0l_A
2o4v_A
2odj_A
3dwo_X
3dwn_A
2qom_A
2wjq_A

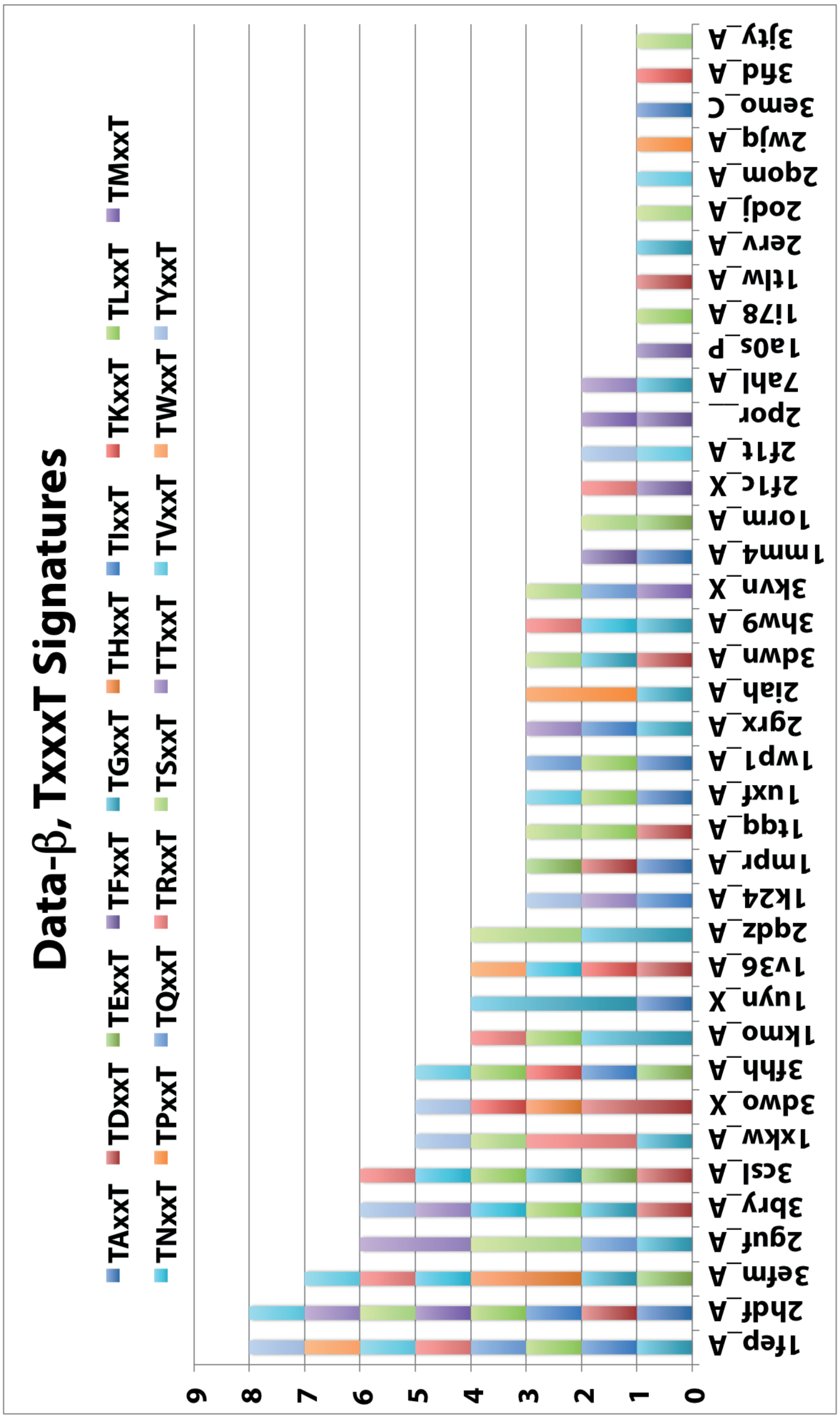

Data-β, TxxxT Signatures
TAxxT
TDxxT
TExxT
TFxxT
TGxxT
THxxT
TIxxT
TKxxT
TLxxT
TMxxT
TNxxT
TPxxT
TQxxT
TRxxT
TSxxT
TTxxT
TVxxT
TWxxT
TYxxT
0
1
2
3
4
5
6
7
8
9
1fep_A
2hdf_A
3efm_A
2guf_A
3bry_A
3csl_A
1xkw_A
3dwo_X
3fhh_A
1kmo_A
1uyn_X
1v36_A
2qdz_A
1k24_A
1mpr_A
1tqq_A
1uxf_A
1wp1_A
2grx_A
2iah_A
3dwn_A
3hw9_A
3kvn_X
1mm4_A
1orm_A
2f1c_X
2f1t_A
2por_
7ahl_A
1a0s_P
1i78_A
1tlw_A
2erv_A
2odj_A
2qom_A
2wjq_A
3emo_C
3fid_A
3jty_A

Figure S7-F

Data-β, SxxS Signatures

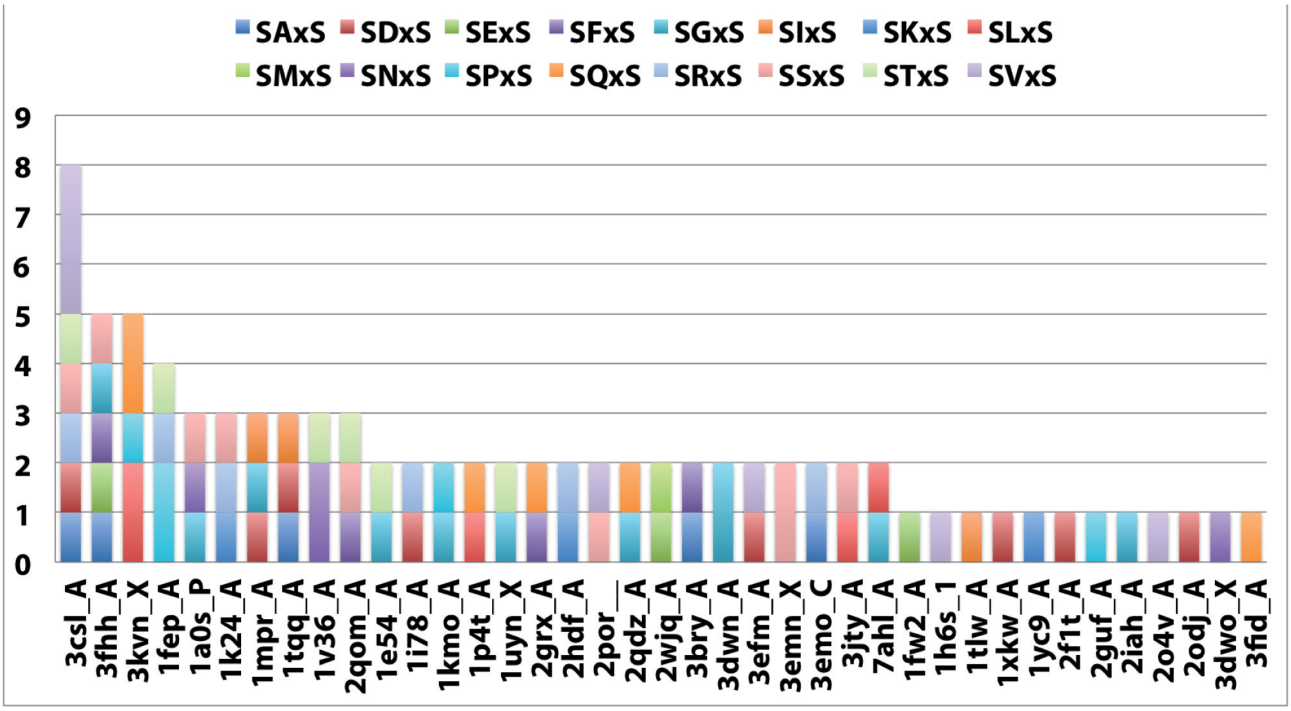


Figure S8-A

Sequences	Cluster	TT	ST	SS	TxxT	TxxxT	SxxxT	SxxxS	SxxT	SxxS
1a91_A	1	0	0	0	0	0	0	0	0	0
1afo_A	1	0	0	0	0	0	0	0	0	0
1aig_H	1	0	1	1	1	0	0	0	0	0
1aig_L	1	0	0	0	1	1	1	0	1	0
1ar1_B	1	0	0	1	4	0	0	0	0	0
1b9u_A	1	0	0	0	0	0	0	0	0	0
1ba4_A	1	0	0	0	0	0	0	0	0	0
1be3_K	1	0	0	0	0	0	0	1	0	0
1bzk_A	1	0	0	0	0	0	1	0	1	0
1c17_M	1	0	0	0	0	1	0	1	0	2
1c51_B	1	0	0	0	0	0	0	0	0	0
1dx7_A	1	0	0	0	1	0	0	1	0	0
1et2_S	1	0	1	1	4	1	1	0	1	1
1eys_M	1	0	0	0	1	1	0	1	0	0
1f6g_A	1	2	0	0	4	0	0	0	1	0
1fft_C	1	0	1	1	4	1	1	4	1	0
1fx8_A	1	2	2	0	1	0	1	0	0	2
1ijd_B	1	0	0	0	0	0	0	0	0	0
1izl_A	1	3	2	2	1	0	0	4	1	1
1jb0_F	1	1	0	0	0	0	1	0	0	0
1jb0_I	1	0	0	0	0	0	0	1	0	0
1jb0_K	1	2	0	0	1	1	0	0	0	0
1jb0_L	1	0	1	2	0	0	0	4	1	0
1jb0_M	1	0	1	0	0	0	0	0	0	0
1jb0_X	1	0	0	0	0	0	1	0	0	0
1jdm_A	1	0	0	0	0	0	0	0	0	0
1kqf_B	1	0	1	0	1	1	0	1	0	1
1kqf_C	1	0	0	0	1	0	0	0	0	2
1kyk_A	1	0	0	1	0	0	0	0	0	0

Figure S8-B

1l7v_A	1	0	2	0	0	0	1	0	0	0
1lbn_A	1	0	0	0	0	0	0	1	1	0
1lgh_A	1	0	1	0	0	0	0	0	0	0
1m0k_A	1	2	2	0	0	0	0	1	0	0
1m56_D	1	0	0	0	0	0	1	0	0	0
1nw0_A	1	0	2	0	0	0	0	0	0	0
1oed_B	1	1	0	0	0	1	0	4	3	1
1oed_C	1	0	4	2	1	0	0	9	1	0
1orq_C	1	1	0	1	1	0	0	4	1	2
1p7b_A	1	1	1	3	1	0	1	4	2	1
1pb2_A	1	2	1	1	1	1	0	9	3	1
1pb4_D	1	2	1	1	0	1	0	1	1	2
1prc_H	1	0	0	0	0	0	0	1	1	1
1pw4_A	1	0	0	2	1	0	0	1	0	2
1q90_A	1	0	0	0	0	0	0	0	1	0
1q90_L	1	0	0	0	0	0	0	0	0	0
1q90_M	1	0	0	0	1	0	0	0	0	0
1q90_N	1	0	0	0	0	0	0	1	0	0
1qle_C	1	1	0	0	0	1	0	0	0	0
1rh5_A	1	1	1	2	1	0	1	1	0	1
1rh5_B	1	1	0	0	0	0	0	0	0	0
1rh5_C	1	0	0	0	0	1	0	0	0	0
1rkl_A	1	0	1	0	0	0	1	0	0	1
1s6e_A	1	0	2	0	1	0	0	4	3	2
1spf_A	1	0	0	0	0	0	0	0	0	0
1sr1_A	1	1	0	1	1	1	0	1	0	1
1upe_A	1	0	2	2	1	0	1	4	0	1
1vry_A	1	2	0	1	1	1	0	0	0	1
1waz_A	1	0	0	0	0	0	0	0	1	0
1wrg_A	1	0	0	1	0	0	0	0	1	0

Figure S8-C

1xio_A	1	2	1	1	4	0	2	0	1	0
1y8s_A	1	0	3	1	4	1	0	0	3	1
1yew_C	1	1	0	0	1	2	0	0	1	0
1zas_A	1	0	0	0	0	0	0	1	1	0
1zc7_A	1	0	2	1	4	1	0	0	1	1
1zza_A	1	1	0	0	0	1	0	0	1	0
2a06_E	1	2	1	3	1	1	2	9	0	3
2a06_G	1	0	0	0	1	0	0	0	0	0
2a06_W	1	0	1	0	0	1	0	0	0	0
2acz_C	1	0	0	1	0	0	2	1	0	1
2acz_D	1	0	0	0	0	1	1	0	1	1
2akh_X	1	0	0	1	1	0	0	0	0	1
2akh_Y	1	1	1	2	1	1	0	1	2	1
2akh_Z	1	2	0	0	1	1	0	0	1	0
2aui_A	1	0	0	0	0	0	0	1	0	0
2axt_E	1	1	1	0	0	0	1	0	0	0
2axt_F	1	0	0	0	1	0	0	0	0	0
2axt_H	1	1	1	0	0	0	0	0	0	0
2axt_I	1	0	0	0	0	1	0	0	0	0
2axt_K	1	0	0	0	0	0	0	0	0	0
2axt_T	1	0	0	0	0	0	0	0	0	0
2b0x_A	1	1	2	2	1	1	2	4	2	2
2b2f_A	1	1	2	3	0	0	2	0	0	1
2b6s_A	1	0	2	1	0	0	1	4	0	0
2b76_C	1	1	1	0	0	1	0	0	0	0
2b76_D	1	0	0	0	1	0	0	0	0	0
2bg9_E	1	1	2	1	1	1	0	1	3	0
2bl2_A	1	1	0	0	0	0	0	1	0	1
2c3e_A	1	0	0	0	0	0	0	1	0	0
2cpb_A	1	0	0	0	0	0	0	1	0	1

Figure S8-D

2d2c_A	1	0	0	1	0	0	1	0	1	0
2d2c_B	1	1	0	0	1	0	0	0	0	0
2d2c_D	1	1	0	0	1	1	0	0	0	0
2d2c_E	1	0	0	0	0	0	0	0	0	0
2d2c_F	1	0	0	0	0	0	0	0	0	0
2d2c_G	1	0	0	0	0	0	0	0	0	0
2d57_A	1	1	1	2	1	0	2	1	2	2
2dyr_I	1	0	1	0	0	0	0	0	0	0
2evu_A	1	1	0	1	0	0	1	1	1	1
2f93_A	1	2	0	1	1	0	0	1	0	1
2f95_B	1	0	1	0	0	0	0	1	2	0
2fbw_C	1	1	0	2	0	1	0	9	1	1
2fbw_D	1	0	0	1	0	0	1	1	1	0
2fkw_A	1	1	0	0	0	0	0	0	1	0
2fyn_A	1	1	1	0	4	0	0	4	2	0
2fyn_B	1	1	0	0	1	1	1	0	1	0
2fyn_C	1	1	0	1	0	0	0	1	0	0
2h8a_A	1	0	1	1	0	1	0	0	2	0
2hac_A	1	0	0	0	0	0	0	0	0	0
2ibz_G	1	0	0	0	0	0	0	1	1	0
2ibz_I	1	0	0	1	1	1	1	0	0	1
2ic8_A	1	0	0	0	0	0	0	1	0	0
2j7a_C	1	0	1	0	0	0	0	1	0	0
2jo1_A	1	0	1	1	0	0	0	0	0	0
2jow_A	1	0	0	0	1	0	0	0	0	1
2jp3_A	1	0	1	1	0	1	0	0	1	0
2jwa_A	1	0	0	0	0	0	0	1	1	1
2k1k_A	1	0	0	0	0	0	1	1	0	0
2k37_A	1	0	0	0	0	0	0	1	0	0
2k73_A	1	0	0	0	0	1	1	0	0	0

Figure S8-E

2k9j_B	1	0	0	0	0	0	0	0	0	0
2k9p_A	1	0	2	1	0	0	0	1	4	1
2k9y_A	1	0	0	0	0	0	0	1	0	0
2ka1_A	1	0	0	0	0	0	0	1	0	0
2kb7_P	1	0	1	0	0	0	0	0	0	0
2kdc_A	1	1	0	1	1	1	0	0	0	0
2ki9_A	1	1	0	0	0	0	1	0	1	0
2kih_A	1	0	0	1	0	0	0	1	0	1
2kix_A	1	0	0	1	0	0	0	1	0	1
2klu_A	1	0	0	0	0	0	1	0	0	1
2knc_A	1	0	0	0	0	0	0	0	0	0
2kns_A	1	0	0	3	0	0	0	1	0	1
2kog_A	1	0	1	0	1	0	1	0	0	1
2ksd_A	1	0	0	0	0	0	0	0	0	0
2kse_A	1	1	1	1	0	0	0	0	0	0
2ksf_A	1	0	0	0	0	0	2	0	0	0
2nq2_A	1	0	1	1	0	0	0	9	2	0
2nr9_A	1	0	0	0	0	1	0	0	1	0
2nrg_A	1	0	0	0	0	1	0	0	0	0
2oar_A	1	0	1	1	4	2	0	0	1	1
2oau_A	1	1	1	3	1	1	0	0	0	0
2pno_A	1	1	0	0	0	0	1	0	0	0
2q67_A	1	1	1	0	4	0	1	0	2	1
2q7m_A	1	1	2	0	4	1	1	1	0	0
2qfi_A	1	0	0	0	1	0	0	1	1	1
2rdd_B	1	0	0	0	0	0	0	0	0	1
2rh1_A	1	1	0	3	1	1	1	4	3	1
2v4h_A	1	0	0	2	0	0	0	1	0	0
2vpw_C	1	0	0	0	0	0	1	0	0	0
2w1p_A	1	1	0	0	0	0	0	0	1	0

Figure S8-F

2w83_C	1	0	0	0	0	0	0	0	0	0
2wgm_A	1	0	2	0	0	0	0	0	0	0
2wsc_1	1	0	0	2	1	0	0	4	0	3
2wsc_H	1	2	0	1	0	0	1	1	2	1
2wsc_I	1	0	1	0	0	0	0	0	0	0
2wsc_J	1	0	1	0	0	0	0	0	0	0
2wsc_K	1	3	1	2	1	1	1	4	2	2
2yvx_A	1	0	0	3	0	2	0	0	1	1
2z73_A	1	1	1	1	0	0	1	4	3	1
2z9a_A	1	0	0	1	0	0	0	1	0	1
2zjs_E	1	0	0	0	1	0	0	0	0	0
2zjs_Y	1	0	0	0	0	1	0	1	1	0
2zw3_A	1	0	2	1	4	0	1	0	1	1
3a0b_J	1	0	0	1	0	0	0	0	0	0
3a0b_L	1	0	0	1	0	0	0	0	0	0
3a0b_M	1	0	0	1	0	0	0	1	0	0
3a0b_Z	1	0	0	0	0	0	0	0	0	0
3a3y_B	1	2	0	2	1	0	0	4	1	1
3a3y_G	1	0	0	0	0	0	1	0	0	0
3a7k_A	1	1	3	3	1	1	0	1	2	1
3abl_B	1	2	1	3	9	1	2	0	0	2
3abl_D	1	0	1	1	0	0	0	1	0	1
3abl_G	1	0	0	0	0	1	0	1	0	0
3abl_J	1	0	0	0	0	0	0	0	0	0
3abl_K	1	0	0	0	0	0	0	0	1	0
3abl_L	1	0	0	0	0	0	0	0	0	0
3abl_M	1	0	0	1	0	0	0	0	0	0
3b4r_A	1	0	1	0	0	0	0	0	0	0
3b5d_A	1	1	1	0	0	0	1	0	2	1
3beh_A	1	3	2	2	1	1	1	4	2	2

Figure S8-G

3bvd_B	1	1	1	0	1	1	0	0	0	0
3bvd_C	1	0	0	0	0	0	0	0	0	0
3bz1_X	1	0	0	1	0	0	1	0	0	1
3bz1_y	1	0	0	0	0	0	0	0	0	0
3c02_A	1	1	1	2	0	0	1	16	0	0
3c1g_A	1	0	0	1	1	1	1	0	2	0
3chx_B	1	2	1	0	1	0	1	1	0	0
3cn5_A	1	0	0	1	1	1	0	4	2	0
3cp1_A	1	0	0	0	0	0	0	1	0	0
3cx5_D	1	2	0	1	1	1	0	0	1	0
3d31_C	1	0	0	2	0	0	0	9	1	1
3ddl_A	1	0	2	2	4	0	1	4	0	1
3dhw_A	1	0	0	0	1	2	0	0	0	0
3din_D	1	0	0	0	0	0	0	1	0	1
3din_E	1	0	0	0	0	0	1	0	0	1
3dl8_C	1	0	0	0	0	0	0	1	0	0
3dl8_E	1	1	1	0	1	0	0	0	1	0
3dww_A	1	0	1	1	0	0	1	1	0	0
3e7k_A	1	0	0	0	1	0	0	0	2	1
3eam_A	1	0	2	0	1	1	1	0	1	0
3egw_C	1	2	0	1	0	0	0	0	1	0
3eh3_A	1	0	2	0	1	2	0	4	3	2
3fh6_G	1	1	1	1	0	0	0	0	0	0
3hd7_B	1	0	1	1	0	0	1	1	0	1
3hgc_A	1	1	1	2	4	1	0	1	3	1
3jyc_A	1	3	1	1	4	1	0	1	2	0
3k3f_A	1	2	2	0	1	1	0	1	2	1
3k69_A	1	0	0	0	0	2	1	1	0	0
3kcu_A	1	1	1	0	1	1	1	0	1	1
3llq_A	1	0	2	1	1	0	2	0	0	1

Figure S8-H

1dop_D	2	1	1	1	4	2	0	1	2	0
1fft_A	2	2	1	1	1	2	2	1	0	0
1kad_A	2	1	4	0	4	3	0	4	2	0
1o5w_A	2	0	1	1	0	3	1	1	1	0
1suk_A	2	3	1	1	4	3	2	4	2	1
2a0d_A	2	2	4	5	0	3	0	4	2	1
2a65_A	2	1	0	1	4	3	0	0	2	0
2bs2_C	2	1	1	2	1	2	0	1	1	0
2gfp_A	2	1	1	2	0	2	0	1	3	0
2hyd_A	2	3	2	0	4	1	2	1	2	1
2iub_A	2	1	1	3	1	3	2	1	2	1
2j58_A	2	3	1	0	0	1	2	1	4	0
2jln_A	2	1	4	1	16	2	2	1	1	0
2r6g_F	2	6	1	3	4	4	3	0	3	1
2vl0_A	2	1	1	1	1	2	0	1	2	0
2wpr_A	2	4	0	0	4	2	1	0	1	0
2wsc_A	2	6	1	3	4	3	1	4	3	2
3c9l_A	2	5	1	0	1	3	1	1	2	0
3chx_A	2	2	2	1	9	2	1	1	0	1
3det_A	2	1	1	1	0	2	0	1	2	0
3h9v_A	2	2	1	3	1	3	2	4	6	0
3kbc_A	2	0	0	2	0	4	1	0	4	0
3kp9_A	2	2	0	1	4	3	0	0	3	0
1fft_B	3	1	0	0	1	1	2	25	1	0
1j4n_A	3	2	0	3	0	0	3	9	5	2
1lnq_A	3	0	3	0	1	0	2	1	2	0
1lvi_A	3	0	2	2	1	1	5	9	4	4
1p49_A	3	1	3	3	4	2	3	4	2	2
1pb4_C	3	1	0	1	0	1	2	1	1	0
1xl4_A	3	1	0	2	25	0	2	4	2	2

Figure S8-I

1y36_A	3	2	2	3	0	3	2	1	2	3
1yg7_A	3	0	0	1	1	1	3	9	0	0
1yo9_L	3	1	0	5	1	0	3	9	4	2
1zti_A	3	1	2	4	9	2	2	9	0	4
2ac6_A	3	0	3	4	4	1	2	9	1	3
2bg9_A	3	0	3	3	0	1	2	4	2	3
2bmn_A	3	1	0	1	1	1	2	0	0	0
2cfp_A	3	1	0	2	0	0	3	1	3	3
2f75_A	3	1	1	5	1	0	4	9	3	1
2g2a_A	3	0	0	5	0	0	3	1	2	0
2iil__	3	1	3	2	1	2	1	9	2	3
2ik3__	3	2	3	1	1	2	2	1	2	3
2wsc_2	3	2	5	3	4	1	2	25	2	3
2wvy_A	3	0	1	3	9	0	4	1	3	3
3b9w_A	3	3	0	0	16	1	2	4	4	2
3dtu_A	3	2	3	2	0	2	4	9	0	4
3hqk_A	3	1	2	4	0	1	2	1	1	2
3ixz_B	3	1	2	0	0	0	4	1	1	0
1bcc_C	4	0	0	2	16	3	2	1	1	2
1mhs_A	4	3	2	4	9	3	1	9	2	4
1y9c_A	4	1	1	0	1	4	1	4	1	2
1z8e_A	4	2	3	2	1	2	0	4	6	3
1zcd_A	4	1	0	2	0	0	0	4	1	4
2axt_B	4	2	0	3	1	1	0	0	2	3
2axt_C	4	2	0	5	1	0	3	0	0	4
2gfz_A	4	0	3	2	1	1	1	9	2	4
2iqp_A	4	0	1	2	0	2	0	4	1	3
2jiz_G	4	1	1	3	0	1	1	9	2	3
2r9r_B	4	2	4	5	9	4	1	4	3	4
2wcd_A	4	3	0	2	1	3	0	9	1	2

Figure S8-J

2wit_A	4	2	4	2	0	4	2	9	3	4
2wsc_3	4	0	0	3	0	0	1	4	1	3
2wsc_F	4	0	0	6	1	1	0	1	0	2
3b8c_A	4	1	3	2	9	3	2	16	5	2
3dh4_A	4	2	2	2	1	2	1	9	3	5
3fwl_A	4	2	2	2	9	3	1	16	3	3
3g67_A	4	0	0	2	0	2	0	4	2	3
3gi8_C	4	0	1	2	1	0	0	4	0	5
3hd6_A	4	0	2	3	4	0	1	4	1	3
1kpw_A	5	0	4	5	0	1	2	9	4	5
1oz5_A	5	2	3	8	1	0	3	36	2	3
2he6__	5	1	1	5	9	2	1	36	2	5
2wsc_G	5	0	2	7	0	0	1	36	1	3
3b5w_A	5	1	4	6	9	0	3	16	3	6
2g1x_A	6	1	3	3	9	6	5	16	2	3
2gif_A	6	3	7	7	4	6	1	16	2	4
3ixz_A	7	2	4	3	4	4	5	1	8	3
3kg2_A	7	2	6	3	1	3	5	16	9	6
2agv_A	8	7	5	4	49	6	2	36	5	3
2amk_A	8	5	2	9	36	3	2	36	2	4
3g5u_A	9	6	9	5	1	2	3	64	5	6
1u5n_A	10	9	6	10	4	5	7	16	7	6

Figure S9

Data-b	Cluster	SS	ST	TT	SxxxS	SxxS	SxxT	TxxxT	TxxT	SxxxT
1mm4_A	1	0	1	1	0	0	0	2	0	0
2f1c_X	1	0	1	1	0	0	0	2	1	1
2k0l_A	1	0	1	0	2	0	0	0	1	1
3hw9_A	1	0	1	0	0	0	2	3	1	1
1fw2_A	1	1	0	0	2	1	1	0	0	1
1orm_A	1	1	1	1	2	0	1	2	0	1
1tlw_A	1	1	2	0	1	1	1	1	0	0
2erv_A	1	1	1	0	0	0	1	1	0	1
2f1t_A	1	1	0	2	1	1	0	2	2	0
2qom_A	1	1	1	0	2	3	1	1	1	1
2wjq_A	1	1	0	1	1	2	2	1	1	2
3emo_C	1	1	2	1	0	2	1	1	0	2
3fid_A	1	1	0	0	1	1	1	1	1	2
1yc9_A	1	2	2	1	2	1	1	0	0	1
2o4v_A	1	3	1	2	0	1	0	0	1	3
3dwn_A	1	3	1	3	0	2	2	3	1	4
1h6s_1	1	4	1	3	0	1	1	0	1	2
1uxf_A	1	4	2	1	0	0	0	3	2	5
1uun_A	2	0	1	2	2	0	3	0	0	0
1e54_A	2	1	3	0	0	2	3		1	2
2por__	2	1	1	2	1	2	2	2	0	1
1a0s_P	2	2	3	2	1	3	2	1	1	2
2odj_A	2	2	0	3	1	1	4	1	1	0
1i78_A	2	3	1	2	1	2	3	1	0	1
3emn_X	2	3		1	0	2	4	0	3	0
7ahl_A	2	3	2	2	1	2	3	2	4	2
3dwo_X	3	2	1	1	3	1	0	5	1	4
2qdz_A	3	1	2	0	3	2	3	4	0	4
3bry_A	3	4	1	1	3	2	0	6	3	2
1mpr_A	3	3	5	2	3	3	3	3	1	4
3kvn_X	3	1	4	0	4	5	1	3	1	2
1p4t_A	3	2	1	0	5	2	0	0	0	2
1tqq_A	4	3	2	5	6	3	3	3	2	5
2guf_A	4	3	3	4	5	1	4	6	4	2
1k24_A	4	4	1	2	5	3	2	3	4	2
1kmo_A	4	4	4	2	5	2	4	4	3	0
1wp1_A	4	1	1	3	3	0	3	3	3	3
1uyn_X	4	0	4	1	3	2	2	4	3	2
2iah_A	4	4	3	1	3	1	3	3	6	5
3jty_A	4	1	1	0	3	2	2	1	3	2
2grx_A	5	5	0	4	1	2	5	3	4	8
2hdf_A	5	7	1	4	4	2	6	8	4	7
1fep_A	6	2	5	6	2	4	4	8	6	6
1xkw_A	6	1	4	7	6	1	4	5	6	3
3efm_A	6	1	2	5	6	2	3	7	6	5
1v36_A	7	5	8	3	4	3	1	4	5	6
3csl_A	8	7	4	5	4	8	4	6	3	4
3fhh_A	8	11	5	6	3	5	1	5	5	4

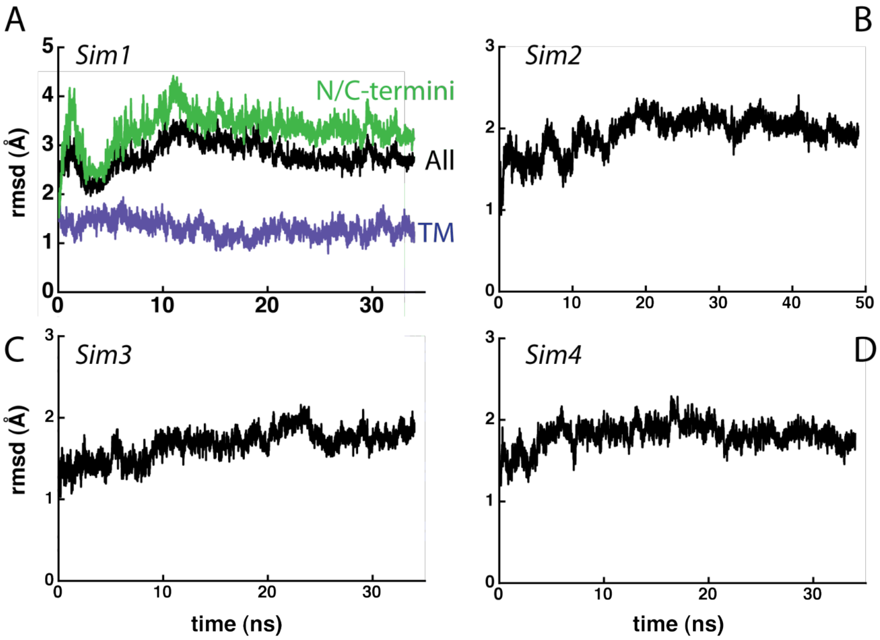

A
Sim1
N/C-termini
All
TM
rmsd (Å)
0
1
2
3
4
5
0
10
20
30
B
Sim2
0
1
2
3
0
10
20
30
40
50
C
Sim3
rmsd (Å)
0
1
2
3
0
10
20
30
time (ns)
D
Sim4
0
1
2
3
0
10
20
30
time (ns)

Figure S10

Supplementary Information References

Aller SG, Yu J, Ward A, Weng Y, Chittaboina S, Zhuo R, Harrel PM, Trinh YT, Zhang Q, Urbatsch IL, Chang G (2009) Structure of P-glycoprotein revelas a molecular basis for poly-specific drug binding. *Science* 323:1718-1722.

Belrhali H, Nollert P, Royant A, Menzel C, Rosenbusch JP, Landau EM, Pebay-Peyroula E (1999) Protein, lipid and water organization in bacteriorhodopsin crystals: a molecular view of the purple membrane at 1.9 Å resolution. *Structure* 7:909–917.

Brillet K, Meksem A, Cobessi D. Crystal structure of the heme/hemoglobin outer membrane transporter ShuA from Shigella dysenteriae. DOI:10.2210/pdb3fhh/pdb.

Buchanan S, Lukacik P, Grizot S, Ghirlando R, Ali MM, Barnard TJ, Jakes KS, Kienker PK, Esser L (2007) Structure of colicin I receptor bound to the R domain of colicin Ia: implications for protein import. *EMBO J* 26:2594-2604.

Cortes, DM, Cuello, LG, Perozo, E (2001) Molecular architecture of full-length KcsA: role of cytoplasmic domains in ion permeation and activation gating. *J Gen Physiol* 117:165-180.

Dong C, Beis K, Nesper J, Brunkan-LaMontagne AL, Clarke BR, Whitfiled C, Naismith JH (2006) Wza the translocon for E. coli capsular polysaccharides defines a new class of membrane protein. *Nature* 444:226-229.

Fang Y, Jayaram H, Shane T, Kolmakova-Partensky L, Wu F, Williams C, Xiong Y, Miller C (2009) Structure of a prokaryotic virtual proton pump at 3.2 Å resolution. *Nature* 460:1040-1043.

Gulbis JM, Kuo A, Smith B, Doyle DA, Edwards A, Arrowsmith C, Sundstrom M. Intermediate gating structure 1 of the inwardly rectifying K+ channel KirBac3.1. DOI:10.2210/pdb1xl4/pdb

Kawate T, Michel JC, Birdsong WT, Gouaux E (2009) Crystal structure of the ATP-gated $P2X_4$ ion channel in the closed state. *Nature* 460: 592-598.

Krieg S, Huche F, Diederichs K, Izadi-Pruneyre N, Lecroisey A, Wandersman C, Delepelaire P, Welte W (2009). Heme uptake across the outer membrane as revealed by crystal structures of the receptor-hemophore complex. *Proc Natl Acad Sci USA* 106:1045-1050.

Kühlbrandt W, Zeelen J, Dietrich J (2002) Structure, mechanism, and regulation of the Neurosphora plasma membrane H^+-ATPase. *Science* 297:1692-1696.

Li B, Nowak NM, Kim SK, Jacobson KA, Bagheri A, Schmidt C, Wess J, Theoretical model of rat M3 muscarinic rat receptor. http://www.pdb.org/pdb pdb ID 2amk.

Lupo D, Li X-D, Durand A, Tomizaki T, Cherif-Zahar B, Matassi G, Merrik M, Winkler FK (2007) The 1.3 Å resolution structure of *Nitrosomonas europaea* Rh50 and mechanistic implications for NH3 transport by Rhesus family proteins. *Proc Natl Acad Sci USA* 104:19303-19308.

Obara K, Miyashita N, Xu C, Toyoshima I, Sugita Y, Inesi G, Toyoshima C (2005) Structural role of countertransport revealed in Ca^{2+} pump crystal structure in the absence of Ca^{2+}. *Proc Natl Acad Sci USA* 102:14489-14496.

Pederson BP, Buch-Pedersen MJ, Morth JP, Palmgren MG, Nissen P (2007) Crystal structure of the plasma membrane proton pump. *Nature* 450:1111-1114.

Sobolevsky AI, Rosconi MP, Gouaux E (2009) X-ray structure, symmetry and mechanism of an AMPA-subtype glutamate receptor. *Nature* 462:745-756.

Stenkamp RE (2008) Alternative models for two crystal structures of bovine rhodopsin. Acta Crystallogr D64:902-904

Subbarao GV, van der Berg B (2006) Crystal structure of the monomeric porin OmpG. J Mol Biol 360:750-759.

Sui H, Han B-G, Lee JK, Wallan P, Jap BK (2001) Structural basis of water-specific transport through the AQP1 water channel. *Nature* 414:872-878.