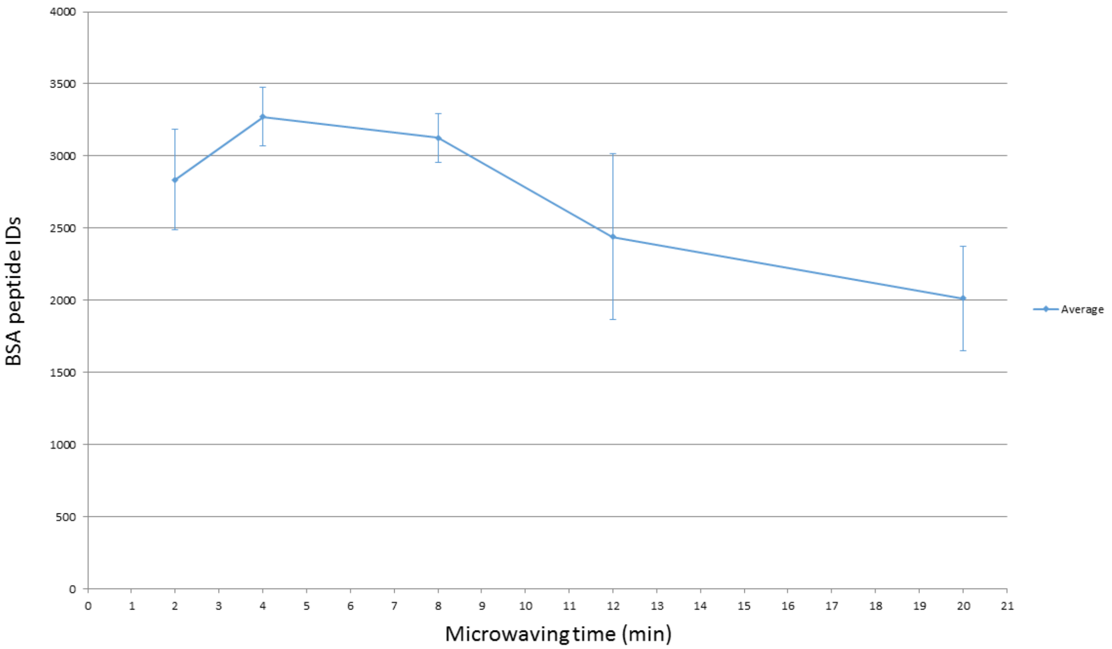
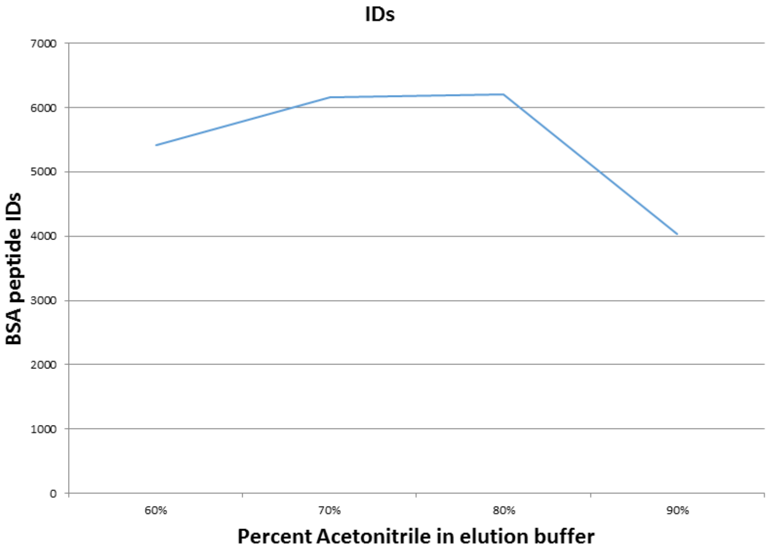
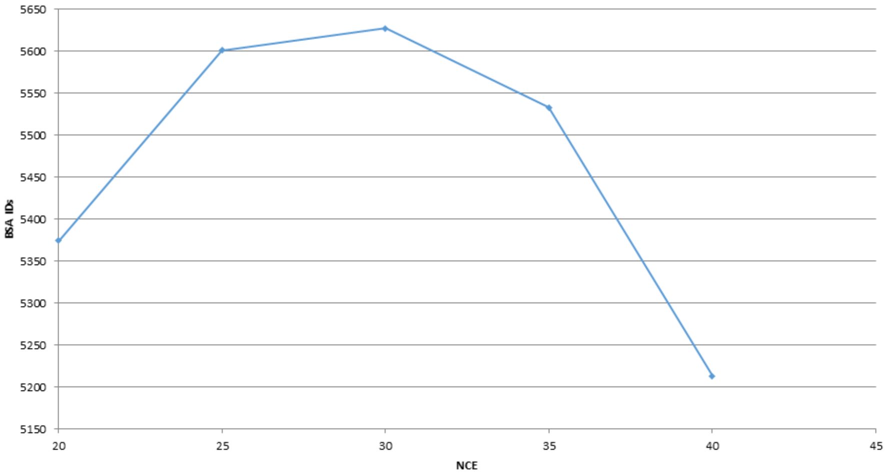


Supplementary Figure S1- Optimization of microwave assisted acid hydrolysis time. A stock solution of 0.25 µg\µl reduced and alkylated BSA in 3 M HCl were prepared and divided into 15 tubes, each containing 10 µg BSA (40 µl). Tubes were microwaved together and at 2, 4, 8, 12 and 20 min, microwaving was briefly paused. Three tubes were transferred at that time point and placed on ice. Once all tubes were collected, they were subjected to identical solid phase extraction, LC-MS/MS analysis and database searching against the BSA sequence and common lab contaminants. The average number of BSA peptides identified in the database search of the 3 technical replicates was plotted against the microwaving time.



Supplementary Figure S2- Optimization of solid phase extraction. A stock solution of BSA peptides produced by 4 min MAAH was divided into 5 tubes. Each sample was subject to solid phase extraction by loading onto Oasis HLB sorbent in 96-well extraction plate, according to the manufacturer's instructions. Elution was performed with 60%, 70%, 80%, or 90% acetonitrile with 0.1% formic acid. Resulting elution was subjected to LC-MS/MS analysis and database searching against the BSA sequence and common lab contaminants. The number of BSA peptides identified in the database search was plotted against the acetonitrile percentage in the elution buffer.



Supplementary Figure S3- Optimization of normalized collision energy (NCE). A sample of BSA peptides produced by 4 min MAAH and solid phase extraction (elution with 60% acetonitrile) was analyzed by LC-MS/MS using 20, 25, 30, 35 or 40 NCE, with separate injections for each analysis. Resulting spectra were subjected to database searching against the BSA sequence and common lab contaminants. The number of BSA peptides identified in the database search is plotted against the NCE.

pTA_BSA_Rep1_DiPS

Parameters

acid input files:
H:\Proteomics\Projects\RnD\Hydrolysis\manuscript\Assembly
\BSA\BSA_Rep1_DiPS_de-Novo_Peaks-output.csv
trypsin input files:

kmer size: 7
kmer min overlap: 5
unite min overlap: 5
unite min extension: 7
merge min quality: 0.7

Final Contig1

Sequence with potential ambiguities/variants

🟦 - Replacements of N->D and Q->E [acid]

🟩 - Potential variants

E* - Pyroglutamate (Post-translational modification of N-terminal Q or E)

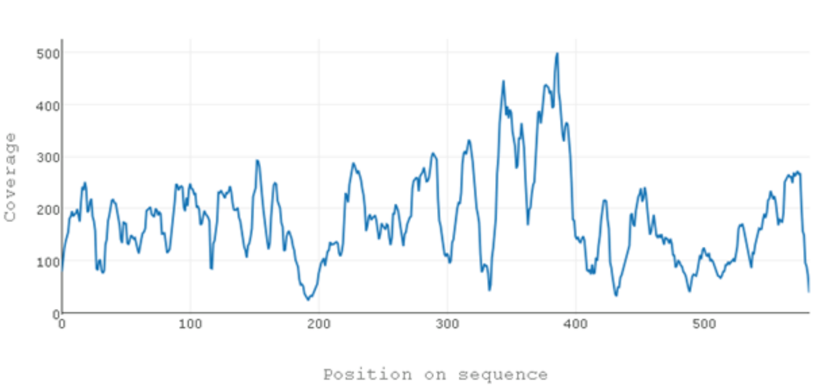
```
DTHKSELAHRFK[D:198,N:1]LGEEHFKGLVLLAFS[E:81,Q:1]YLE[E:75,Q:1
]CPFDEHVKLV[D:168,N:28]ELTEFAKTCVADESHAGCEKSLHTLFGD[E:183,Q:
1]LCKVASLR[E:133,Q:2]TYGDMADCCEK[E:240,Q:3][E:240,Q:1]P[E:18
8,Q:7]R[D:171,N:34][E:230,Q:8]CFLSHK[D:201,N:4][D:195,N:3]SP
DLPKLK[D:86,P:76][P:85,D:63]P[D:113,N:26]TLCDEFKAD[E:226,Q:3
]KKFWGKYLYELARRHPYFYAP[E:239,Q:1]LLYYA[D:160,N:38]KY[D:121,N
:16]GVF[E:212,Q:2][E:244,Q:1]CC[E:209,Q:6]AEDKGACLLPKLETMR[E
:62,Q:2]KVLTSSAR[E:32,Q:1]RLRCASL[E:95,Q:1]KFG[E:101,Q:2]RAL
KAWSVARLS[E:130,Q:3]KFPKA[E:261,Q:2]FV[E:281,Q:2]VTKLVTDLTKV
HKECCHGDLLECA[D:168,N:2][D:165,N:3]RA[D:139,N:4]LAKYLCD[D:11
9,N:18][E:127,Q:1]DTLSSKLKECC[D:232,N:1]KPLLEKSHCLAEVEKDALPE
[D:102,N:6]LPPLTADFA[E:209,Q:2][D:208,N:1]K[D:285,N:1]VCK[D:
252,N:71]Y[E:325,Q:2]EAKDAFLGSFLY[E:87,Q:3]YSRRHP[E:180,Q:1]
YAVSVLLRLAKEYEATL[E:334,Q:2]ECCAK[D:216,N:1][D:167,N:1]PHACY
STVF[D:362,N:1]KLKHLV[D:418,N:2][E:425,Q:1]P[E:389,Q:7][D:37
6,N:84]LLK[E:398,Q:7][D:316,N:58]C[D:325,N:4]EFEKLG[E:177,Q:
2]YGF[E:144,Q:2][D:117,N:20]ALLVRYTRKVP[E:76,Q:1]VSTPTLVEVSR
SLGKVGTRCCTKPESERMPCTE[D:222,N:1]YLSLLL[D:165,N:33]RLCVLHEKT
PVSEKVTKCCT[E:109,Q:1]SLV[D:81,N:17]RRPCFSALTPDETYVPKAFDEKLF
TFHADLCTLPDTEKELKKETALVELLKHKPKATEE[E:140,Q:1]LKTVM[E:188,Q:
2][D:152,N:30]FVAFVDKCCAA[D:176,N:1][D:172,N:1]KEACFAVEGPKLV
VST[E:145,Q:4]TALA
```

Final consensus sequence

Position: 0   Weak Strong

```
  0 DTHKSELAHRFKDLGEEHFKGLVLLAFSEYLEECPFDEHVKLVNELTEFAKTCVADESHA 59
 60 GCEKSLHTLFGDELCKVASLRETYGDMADCCEKEEPQRNQCFLSHKDDSPDLPKLKDPPN 119
120 TLCDEFKADEKKFWGKYLYELARRHPYFYAPELLYYANKYNGVFEECCQAEDKGACLLPK 179
180 LETMRQKVLTSSARQRLRCASLEKFGERALKAWSVARLSEKFPKAEFVEVTKLVTDLTKV 239
240 HKECCHGDLLECADDRADLAKYLCDNEDTLSSKLKECCDKPLLEKSHCLAEVEKDALPED 299
300 LPPLTADFAEDKDVCKNYEEAKDAFLGSFLYQYSRRHPEYAVSVLLRLAKEYEATLEECC 359
360 AKDDPHACYSTVFDKLKHLVDEPENLLKENCDEFEKLGEYGFENALLVRYTRKVPEVSTP 419
420 TLVEVSRSLGKVGTRCCTKPESERMPCTEDYLSLLLNRLCVLHEKTPVSEKVTKCCTESL 479
480 VNRRPCFSALTPDETYVPKAFDEKLFTFHADLCTLPDTEKELKKETALVELLKHKPKATE 539
540 EELKTVMENFVAFVDKCCAADDKEACFAVEGPKLVVSTQTALA                  582
```

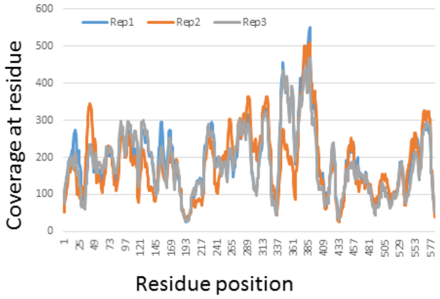
Coverage Graph



Supplementary Figure S4- Example of pTA final report. pTA reports the final assembled contigs. For each reported contig, there are 3 panels displayed: Top panel- "Sequence with potential ambiguities/variants" reports the assembled sequence including both optional residues and their coverage at positions where there is evidence supporting different residues. Middle panel- "Final consensus sequence" reports the final sequence decided upon where the selected residue in case of non-conclusive assignments is color coded for confidence in assignment. If proteolytic digestion data is also used, in the top panel the ratios are reported separately for MAAH and proteolysis in order to separate MAAH-derived chemical modifications from other causes of sequence variation. Bottom panel- "Coverage Graph" displays the coverage (number of peptide tags covering the position) at each position. In this example (BSA replicate #1), the final sequence decided upon contained a "swap" of two residues ('DP' instead of 'PD') at positions 117-118 as evident from alignment to the known BSA sequence (see Figure 3 in the main text of the manuscript). These positions are colored red (low confidence) in the "Final consensus sequence" panel, and show about equal amount of evidence for both options ("PD" or "DP") when examined in the top panel.

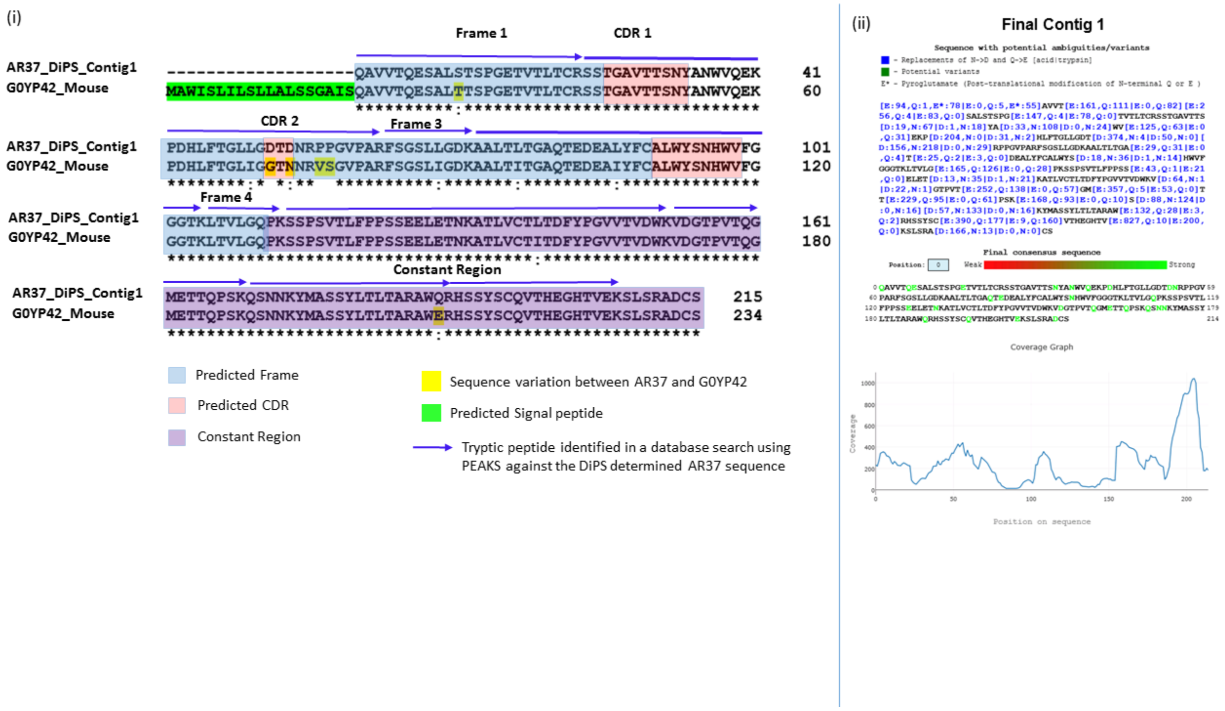
b) Bovine serum albumin (583 AA long, excluding signal peptide)

Replicate	Coverage	Accuracy	# contigs	Incorrect Q/E assignment without trypsin data	Incorrect N/D assignment without trypsin data	Incorrect Q/E assignment with trypsin data	Incorrect N/D assignment with trypsin data
1	100%	100%	1	21	1	3	0
2	100%	100%	1	18	2	5	0
3	100%	100%	1	17	0	4	0

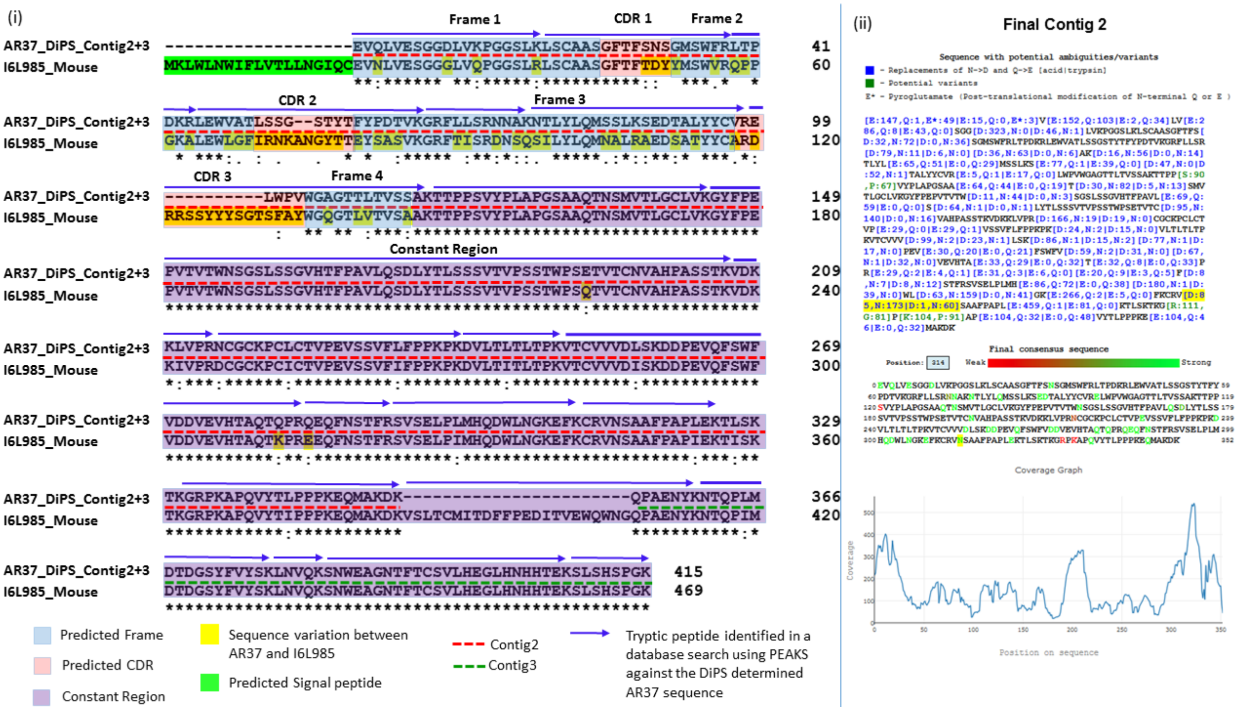


```
P02769|BSA      MKWVTFISLLLLFSSAYSRGVFRRDTHKSEIAHRFKDLGEEHFKGLVLIAFSQYLQQCPFDEHVKLVNELTEFAKTCVADESHAGCEKSLHTLFGDELCKVASLRETYGDMADCCEKQEP
BSA_NGPS_Rep1   ------------------------DTHKSELAHRFKDLGEEHFKGLVLLAFSQYLQQCPFDEHVKLVNELTEFAKTCVADESHAGCEKSLHTLFGDELCKVASLRETYGDMADCCEKQEP
BSA_NGPS_Rep2   ------------------------DTHKSELAHRFKDLGEEHFKGLVLLAFSQYLQQCPFDEHVKLVNELTEFAKTCVADESHAGCEKSLHTLFGDELCKVASLRETYGDMADCCEKQEP
BSA_NGPS_Rep3   ------------------------DTHKSELAHRFKDLGEEHFKGLVLLAFSQYLQQCPFDEHVKLVNELTEFAKTCVADESHAGCEKSLHTLFGDELCKVASLRETYGDMADCCEKQEP
                                        ******:*****************:***********************************************************************

P02769|BSA      ERNECFLSHKDDSPDLPKLKPDPNTLCDEFKADEKKFWGKYLYEIARRHPYFYAPELLYYANKYNGVFQECCQAEDKGACLLPKIETMREKVLTSSARQRLRCASIQKFGERALKAWSVA
BSA_NGPS_Rep1   ERNQCFLSHKDDSPDLPKLKPDPNTLCDEFKADEKKFWGKYLYELARRHPYFYAPELLYYANKYNGVFQECCQAEDKGACLLPKLETMRQKVLTSSARQRLRCASLQKFGERALKAWSVA
BSA_NGPS_Rep2   ERNECFLSHKDDSPDLPKLKPDPNTLCDEFKADEKKFWGKYLYELARRHPYFYAPELLYYANKYNGVFQECCQAEDKGACLLPKLETMRQKVLTSSARQRLRCASLQKFGERALKAWSVA
BSA_NGPS_Rep3   ERNQCFLSHKDDSPDLPKLKPDPNTLCDEFKADEKKFWGKYLYELARRHPYFYAPELLYYANKYNGVFQECCQAEDKGACLLPKLETMRQKVLTSSARERLRCASLQKFGERALKAWSVA
                ***:****************************************:***************************************:****:********:******:**************

P02769|BSA      RLSQKFPKAEFVEVTKLVTDLTKVHKECCHGDLLECADDRADLAKYICDNQDTISSKLKECCDKPLLEKSHCIAEVEKDAIPENLPPLTADFAEDKDVCKNYQEAKDAFLGSFLYEYSRR
BSA_NGPS_Rep1   RLSQKFPKAEFVEVTKLVTDLTKVHKECCHGDLLECADDRADLAKYLCDNQDTLSSKLKECCDKPLLEKSHCLAEVEKDALPENLPPLTADFAEDKDVCKNYEEAKDAFLGSFLYEYSRR
BSA_NGPS_Rep2   RLSQKFPKAEFVEVTKLVTDLTKVHKECCHGDLLECADDRADLAKYLCDNQDTLSSKLKECCDKPLLEKSHCLAEVEKDALPENLPPLTADFAEDKDVCKNYEEAKDAFLGSFLYEYSRR
BSA_NGPS_Rep3   RLSQKFPKAEFVEVTKLVTDLTKVHKECCHGDLLECADDRADLAKYLCDNQDTLSSKLKECCDKPLLEKSHCLAEVEKDALPQNLPPLTADFAEDKDVCKNYQEAKDAFLGSFLYEYSRR
                **********************************************:******:******************:*******:*:*******************:*****************

P02769|BSA      HPEYAVSVLLRLAKEYEATLEECCAKDDPHACYSTVFDKLKHLVDEPQNLIKQNCDQFEKLGEYGFQNALIVRYTRKVPQVSTPTLVEVSRSLGKVGTRCCTKPESERMPCTEDYLSLIL
BSA_NGPS_Rep1   HPEYAVSVLLRLAKEYEATLEECCAKDDPHACYSTVFDKLKHLVDEPQNLLKQNCDQFEKLGEYGFQNALLVRYTRKVPQVSTPTLVEVSRSLGKVGTRCCTKPESERMPCTEDYLSLLL
BSA_NGPS_Rep2   HPEYAVSVLLRLAKEYEATLEECCAKDDPHACYSTVFDKLKHLVDEPQNLLKQNCDQFEKLGEYGFQNALLVRYTRKVPQVSTPTLVEVSRSLGKVGTRCCTKPESERMPCTEDYLSLLL
BSA_NGPS_Rep3   HPEYAVSVLLRLAKEYEATLEECCAKDDPHACYSTVFDKLKHLVDEPQNLLKQNCDQFEKLGEYGFQNALLVRYTRKVPQVSTPTLVEVSRSLGKVGTRCCTKPESERMPCTEDYLSLLL
                **************************************************:*******************:***********************************************:*

P02769|BSA      NRLCVLHEKTPVSEKVTKCCTESLVNRRPCFSALTPDETYVPKAFDEKLFTFHADICTLPDTEKQIKKQTALVELLKHKPKATEEQLKTVMENFVAFVDKCCAADDKEACFAVEGPKLVV
BSA_NGPS_Rep1   NRLCVLHEKTPVSEKVTKCCTESLVNRRPCFSALTPDETYVPKAFDEKLFTFHADLCTLPDTEKELKKQTALVELLKHKPKATEEQLKTVMENFVAFVDKCCAADDKEACFAVEGPKLVV
BSA_NGPS_Rep2   NRLCVLHEKTPVSEKVTKCCTESLVNRRPCFSALTPDETYVPKAFDEKLFTFHADLCTLPDTEKELKKQTALVELLKHKPKATEEQLKTVMENFVAFVDKCCAADDKEACFAVEGPKLVV
BSA_NGPS_Rep3   NRLCVLHEKTPVSEKVTKCCTESLVNRRPCFSALTPDETYVPKAFDEKLFTFHADLCTLPDTEKELKKQTALVELLKHKPKATEEQLKTVMENFVAFVDKCCAADDKEACFAVEGPKLVV
                *******************************************************:********::******************************************************

P02769|BSA      STQTALA
BSA_NGPS_Rep1   STQTALA
BSA_NGPS_Rep2   STETALA
BSA_NGPS_Rep3   STQTALA
                **:****
```

Supplementary Figure S5 - DiPS results of BSA using MAAH and trypsin digestion data. BSA was cleaved using MAAH in triplicates and analyzed by nanoLC-MS/MS. A single BSA tryptic digest was also analyzed by nanoLC-MS/MS. Tryptic peptide tags were used as input for pTA in addition to each of the MAAH replicates. pTA's output sequences were aligned to the known sequence of each protein using Clustal-Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/). Alignment mismatches that are the result of isobaric I/L, deamidated-Q/E, or deamidated-N/D ambigouities were not counted as sequencing mistakes in the accuracy calculations. Yellow – Q/E mismatches N/D mismatches. Grey – signal peptide or N terminal methionine.


(i)
Frame 1
CDR 1
AR37_DiPS_Contig1 -------------------QAVVTQESALSTSPGETVTLTCRSSTGAVTTSNYANWVQEK 41
G0YP42_Mouse MAWISLILSLLALSSGAISQAVVTQESALTTSPGETVTLTCRSSTGAVTTSNYANWVQEK 60
CDR 2
Frame 3
AR37_DiPS_Contig1 PDHLFTGLLGDTDNRPPGVPARFSGSLLGDKAALTLTGAQTEDEALYFCALWYSNHWVFG 101
G0YP42_Mouse PDHLFTGLIGGTNNRVSGVPARFSGSLIGDKAALTITGAQTEDEAIYFCALWYSNHWVFG 120
Frame 4
AR37_DiPS_Contig1 GGTKLTVLGQPKSSPSVTLFPPSSEELETNKATLVCTLTDFYPGVVTVDWKVDGTPVTQG 161
G0YP42_Mouse GGTKLTVLGQPKSSPSVTLFPPSSEELETNKATLVCTITDFYPGVVTVDWKVDGTPVTQG 180
Constant Region
AR37_DiPS_Contig1 METTQPSKQSNNKYMASSYLTLTARAWQRHSSYSCQVTHEGHTVEKSLSRADCS 215
G0YP42_Mouse METTQPSKQSNNKYMASSYLTLTARAWERHSSYSCQVTHEGHTVEKSLSRADCS 234
Predicted Frame
Predicted CDR
Constant Region
Sequence variation between AR37 and G0YP42
Predicted Signal peptide
Tryptic peptide identified in a database search using PEAKS against the DiPS determined AR37 sequence
(ii)
Final Contig 1
Sequence with potential ambiguities/variants
- Replacements of N->D and Q->E [acid|trypsin]
- Potential variants
E* - Pyroglutamate (Post-translational modification of N-terminal Q or E)
[E:94,Q:1,E*:78|E:0,Q:5,E*:55]AVVT[E:161,Q:111|E:0,Q:82][E:256,Q:4|E:83,Q:0]SALSTSPG[E:147,Q:4|E:78,Q:0]TVTLTCRSSTGAVTTS[D:19,N:67|D:1,N:18]YA[D:33,N:108|D:0,N:24]WV[E:125,Q:63|E:0,Q:31]EKP[D:204,N:0|D:31,N:2]HLFTGLLGDT[D:374,N:4|D:50,N:0][D:156,N:218|D:0,N:29]RPPGVPARFSGSLLGDKAALTLTGA[E:29,Q:31|E:0,Q:4]T[E:25,Q:2|E:3,Q:0]DEALYFCALWYS[D:18,N:36|D:1,N:14]HWVFGGGTKLTVLG[E:165,Q:126|E:0,Q:28]PKSSPSVTLFPPSS[E:43,Q:1|E:21,Q:0]ELET[D:13,N:35|D:1,N:21]KATLVCTLTDFYPGVVTVDWKV[D:64,N:1|D:22,N:1]GTPVT[E:252,Q:138|E:0,Q:57]GM[E:357,Q:5|E:53,Q:0]TT[E:229,Q:95|E:0,Q:61]PSK[E:168,Q:93|E:0,Q:10]S[D:88,N:124|D:0,N:16][D:57,N:133|D:0,N:16]KYMASSYLTLTARAW[E:132,Q:28|E:3,Q:2]RHSSYSC[E:390,Q:177|E:9,Q:160]VTHEGHTV[E:827,Q:10|E:200,Q:0]KSLSRA[D:166,N:13|D:0,N:0]CS
Final consensus sequence
Position: 0
Weak
Strong
0 QAVVTQESALSTSPGETVTLTCRSSTGAVTTSNYANWVQEKPDHLFTGLLGDTDNRPPGV 59
60 PARFSGSLLGDKAALTLTGAQTEDEALYFCALWYSNHWVFGGGTKLTVLGQPKSSPSVTL 119
120 FPPSSEELETNKATLVCTLTDFYPGVVTVDWKVDGTPVTQGMETTQPSKQSNNKYMASSY 179
180 LTLTARAWQRHSSYSCQVTHEGHTVEKSLSRADCS 214
Coverage Graph
Coverage
1000
800
600
400
200
0
0
50
100
150
200
Position on sequence

Supplementary Figure S6- DiPS results of AR37 antibody light chain. AR37 was subjected to DiPS using two LC-MS/MS experiments of a single MAAH preparation and one experiment of a tryptic digest. Parameters for *de novo* analysis of MS/MS spectra included glutamine and glutamate conversion to pyroglutamate as variable modifications. (i) The resulting pTA assembled contig 1 was aligned to G0YP42 (its homolog identified by BLAST search). Domains (Frames 1-4, CDRs 1-3, and constant region) and tryptic peptides identified by a database search of the tryptic digest against DiPS determined sequence are marked on the sequence. (ii) pTA output web report for contig 1. Pyroglutamate is denoted by E* in the top panel.



```
AR37_DiPS_Contig2+3  -------------------EVQLVESGGDLVKPGGSLKLSCAASGFTFSNSGMSWFRLTP   41
I6L985_Mouse         MKLWLNWIFLVTLLNGIQCEVNLVESGGGLVQPGGSLRLSCAASGFTFTDYYMSWVRQPP   60
                                        **:****** **:*****:***********::  ***.*  *

AR37_DiPS_Contig2+3  DKRLEWVATLSSG--STYTFYPDTVKGRFLLSRNNAKNTLYLQMSSLKSEDTALYYCVRE   99
I6L985_Mouse         GKALEWLGFIRNKANGYTTEYSASVKGRFTISRDNSQSILYLQMNALRAEDSATYYCARD  120
                      * ***:. :     .  * *  :***** :**:*::. *****.:*::**:* ***.*:

AR37_DiPS_Contig2+3  ----------LWPVWGAGTTLTVSSAKTTPPSVYPLAPGSAAQTNSMVTLGCLVKGYFPE  149
I6L985_Mouse         RRSSYYYSGTSFAYWGQGTLVTVSAAKTTPPSVYPLAPGSAAQTNSMVTLGCLVKGYFPE  180
                                : ** ** :***:***********************************

AR37_DiPS_Contig2+3  PVTVTWNSGSLSSGVHTFPAVLQSDLYTLSSSVTVPSSTWPSETVTCNVAHPASSTKVDK  209
I6L985_Mouse         PVTVTWNSGSLSSGVHTFPAVLQSDLYTLSSSVTVPSSTWPSQTVTCNVAHPASSTKVDK  240
                     ******************************************:*****************

AR37_DiPS_Contig2+3  KLVPRNCGCKPCLCTVPEVSSVFLFPPKPKDVLTLTLTPKVTCVVVDLSKDDPEVQFSWF  269
I6L985_Mouse         KIVPRDCGCKPCICTVPEVSSVFIFPPKPKDVLTITLTPKVTCVVVDISKDDPEVQFSWF  300
                     *:***:******:**********:**********:************:************

AR37_DiPS_Contig2+3  VDDVEVHTAQTQPRQEQFNSTFRSVSELPLMHQDWLNGKEFKCRVNSAAFPAPLEKTLSK  329
I6L985_Mouse         VDDVEVHTAQTKPREEQFNSTFRSVSELPIMHQDWLNGKEFKCRVNSAAFPAPIEKTISK  360
                     ***********:**:**************:***********************:***:**

AR37_DiPS_Contig2+3  TKGRPKAPQVYTLPPPKEQMAKDK-----------------------QPAENYKNTQPLM  366
I6L985_Mouse         TKGRPKAPQVYTIPPPKEQMAKDKVSLTCMITDFFPEDITVEWQWNGQPAENYKNTQPIM  420
                     ************:***********                       ***********:*

AR37_DiPS_Contig2+3  DTDGSYFVYSKLNVQKSNWEAGNTFTCSVLHEGLHNHHTEKSLSHSPGK  415
I6L985_Mouse         DTDGSYFVYSKLNVQKSNWEAGNTFTCSVLHEGLHNHHTEKSLSHSPGK  469
                     *************************************************
```

Sequence with potential ambiguities/variants

- 🟦 - Replacements of N->D and Q->E (acid|trypsin)
- 🟩 - Potential variants
- E* - Pyroglutamate (Post-translational modification of N-terminal Q or E)

```
[E:147,Q:1,E*:49|E:15,Q:0,E*:3]V[E:152,Q:103|E:2,Q:34]LV[E:2
86,Q:8|E:43,Q:0]SGG[D:323,N:0|D:46,N:1]LVKPGGSLKLSCAASGFTFS[
D:32,N:72|D:0,N:36]SGMSWFRLTPDKRLEWVATLSSGSTYTFYPDTVKGRFLLSR
[D:79,N:11|D:6,N:0][D:36,N:63|D:0,N:6]AK[D:16,N:56|D:0,N:14]
TLYL[E:65,Q:51|E:0,Q:29]MSSLKS[E:77,Q:1|E:39,Q:0][D:47,N:0|D
:52,N:1]TALYYCVR[E:5,Q:1|E:17,Q:0]LWPVWGAGTTLTVSSAKTTPP[S:90
,P:67]VYPLAPGSAA[E:64,Q:44|E:0,Q:19]T[D:30,N:82|D:5,N:13]SMV
TLGCLVKGYFPEPVTVTW[D:11,N:44|D:0,N:3]SGSLSSGVHTFPAVL[E:69,Q:
59|E:0,Q:0]S[D:64,N:1|D:0,N:1]LYTLSSSVTVPSSTWPSETVTC[D:95,N:
140|D:0,N:16]VAHPASSTKVDKKLVPR[D:166,N:19|D:19,N:0]CGCKPCLCT
VP[E:29,Q:0|E:29,Q:1]VSSVFLFPPKPK[D:24,N:2|D:15,N:0]VLTLTLTP
KVTCVVV[D:99,N:2|D:23,N:1]LSK[D:86,N:1|D:15,N:2][D:77,N:1|D:
17,N:0]PEV[E:30,Q:20|E:0,Q:21]FSWFV[D:59,N:2|D:31,N:0][D:67,
N:1|D:32,N:0]VEVHTA[E:33,Q:29|E:0,Q:32]T[E:32,Q:8|E:0,Q:33]P
R[E:29,Q:2|E:4,Q:1][E:31,Q:3|E:6,Q:0][E:20,Q:9|E:3,Q:5]F[D:8
,N:7|D:8,N:12]STFRSVSELPLMH[E:86,Q:72|E:0,Q:38][D:180,N:1|D:
39,N:0]WL[D:63,N:159|D:0,N:41]GK[E:266,Q:2|E:5,Q:0]FKCRV[D:8
5,N:173|D:1,N:60]SAAFPAPL[E:459,Q:1|E:81,Q:0]KTLSKTKG[R:111,
G:81]P[K:104,P:91]AP[E:104,Q:32|E:0,Q:48]VYTLPPPKE[E:104,Q:4
6|E:0,Q:32]MAKDK
```

Final consensus sequence

Position: 314 Weak Strong

```
  0 EVQLVESGGDLVKPGGSLKLSCAASGFTFSNSGMSWFRLTPDKRLEWVATLSSGSTYTFY  59
 60 PDTVKGRFLLSRNNAKNTLYLQMSSLKSEDTALYYCVRELWPVWGAGTTLTVSSAKTTPP 119
120 SVYPLAPGSAAQTNSMVTLGCLVKGYFPEPVTVTWNSGSLSSGVHTFPAVLQSDLYTLSS 179
180 SVTVPSSTWPSETVTCNVAHPASSTKVDKKLVPRNCGCKPCLCTVPEVSSVFLFPPKPKD 239
240 VLTLTLTPKVTCVVVDLSKDDPEVQFSWFVDDVEVHTAQTQPRQEQFNSTFRSVSELPLM 299
300 HQDWLNGKEFKCRVNSAAFPAPLEKTLSKTKGRPKAPQVYTLPPPKEQMAKDK        352
```

Supplementary Figure S7- DiPS results of AR37 antibody heavy chain. (i) The resulting pTA assembled contigs 2 and 3 were aligned to I6L985 (the homolog protein to contig 2 identified by BLAST search). Domains (Frames 1-4 and CDRs 1-3) and tryptic peptides identified by a database search of the tryptic digest against DiPS determined sequence are marked on the sequence. (ii) pTA output web report for contig 3. Pyroglutamate is denoted by E* in the top panel.