Supplemental figure 1

Ser16-Pro17, Ser33-Pro34, and Thr48-Pro49 motifs

```
                                1                                                  50
                                |                                                  |
CONSENSUS_OF_CONSENSUSES        PIVQNLQGQM VHQAISPRTL NAWVKVVEEK AFSPEVIPMF SALSEGATPQ
CONSENSUS_A1                    -----A---- ---SL----- ------I--- ---------- ----------
CONSENSUS_A2                    -----A---- ---------- ---------- ---------- T---------
CONSENSUS_B                     ---------- ---------- ---------- ---------- ----------
CONSENSUS_C                     ---------- ---------- ------I--- ---------- T---------
CONSENSUS_D                     ---------- ---------- ------I--- ---------- ----------
CONSENSUS_F1                    ---------- ---------- ------I--- ---------- ----------
CONSENSUS_G                     -----A---- ---------- ---------- ---------- ----------
CONSENSUS_H                     -----A---- ---------- ---------- ---------- ----------
CONSENSUS_K                     ---------- ----L----- ------I--- ---------- ----------
CONSENSUS_01_AE                 -----A---- ---PL----- ---------- G-N------- ----------
CONSENSUS_02_AG                 -----A---- T--SM----- ------I--- ---------- ----------
CONSENSUS_03_AB                 -----A---- T--SM----- ------I--- ---------- ----------
CONSENSUS_04_CPX                -----A---- ---S------ ------I--- ---------- ----------
CONSENSUS_06_CPX                -----A---- ---------- ------I--- ---------- ----------
CONSENSUS_07_BC                 ---------- ---P------ ---------- ---------- ----------
CONSENSUS_08_BC                 ---------- ---PL----- ---------- ---------- T---------
CONSENSUS_10_CD                 ---------- ---PL----- ------I--- ---------- ----------
CONSENSUS_11_CPX                -----A---- ---------- ---------- ---------- ----------
CONSENSUS_12_BF                 ---------- ----L----- ---------- ---------- ----------
CONSENSUS_14_BG                 -----A---- ---------- ---------- ---------- ----------
```

Ser146-Pro147 motif

```
                                131                         160
                                |                           |
CONSENSUS_OF_CONSENSUSES        IYKRWIILGLNK IVRMYSPVSI LDIRQGPKEP
CONSENSUS_A1                    ------------ ---------- ---K------
CONSENSUS_A2                    ------------ ---------- ----------
CONSENSUS_B                     ------------ -------T-- ----------
CONSENSUS_C                     ------------ ---------- ---K------
CONSENSUS_D                     ------------ ---------- ----------
CONSENSUS_F1                    ------------ ---------- ----------
CONSENSUS_G                     ------------ ---------- ----------
CONSENSUS_H                     ------------ ---------- ---K------
CONSENSUS_K                     ------------ ---------- ----------
CONSENSUS_01_AE                 ------------ ---------- ----------
CONSENSUS_02_AG                 ------V----- ---------- ----------
CONSENSUS_03_AB                 ------------ ---------- ----------
CONSENSUS_04_CPX                ------------ ---------- ----------
CONSENSUS_06_CPX                ------------ ---------- ----------
CONSENSUS_07_BC                 ------------ -------T-- ---K------
CONSENSUS_08_BC                 ------------ -------T-- ---K------
CONSENSUS_10_CD                 ------------ ---------- ----------
CONSENSUS_11_CPX                --R--------- ---------- ----------
CONSENSUS_12_BF                 ------------ ---------- ----------
CONSENSUS_14_BG                 ------------ ---------- ----------
```

Fig. S1. CA core contains three highly conserved Ser-Pro motifs (Ser_{16}-Pro_{17}, Ser_{33}-Pro_{34}, and Ser_{146}-Pro_{147}) and a concerved Thr-Pro motif (Thr_{48}-Pro_{49}). Sequence alignment of CA protein derived from consensus of each subtype, and M-group consensus-of-consensuses based on Consensus and Ancestral Sequence Alignments Current (Aug. 2004) (http://www.hiv.lanl.gov/content/hiv-db/CONSENSUS/M_GROUP/Consensus.html) in HIV database operated by the Division of AIDS of the National Institute of Allergy and Infectious Diseases (NIAID).

Supplemental figure 2

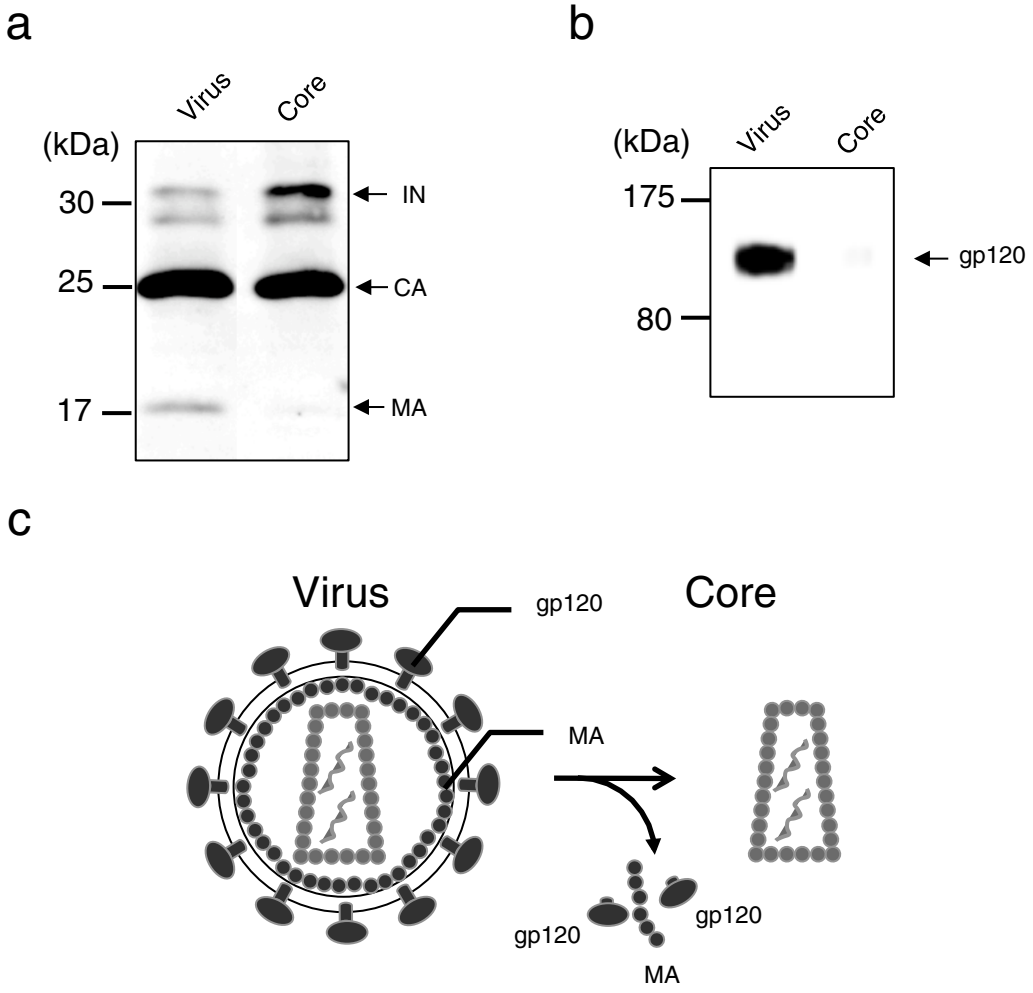

Fig. S2. Comparative protein analysis of HIV virion and core preparations. To determine the protein composition of HIV-1 cores, the extracts were separated by SDS-PAGE and separated proteins were analyzed by western immunoblotting with HIV- positive plasma (A) or anti-gp120 mAb (B). The amount of extract from HIV-1 virions or cores was adjusted to 2 ng of CA. (C) Schematic diagram of HIV core preparation. The HIV surface glycoprotein gp120 and the membrane-associated MA protein were substantially depleted in the core preparation.

Supplemental figure 3

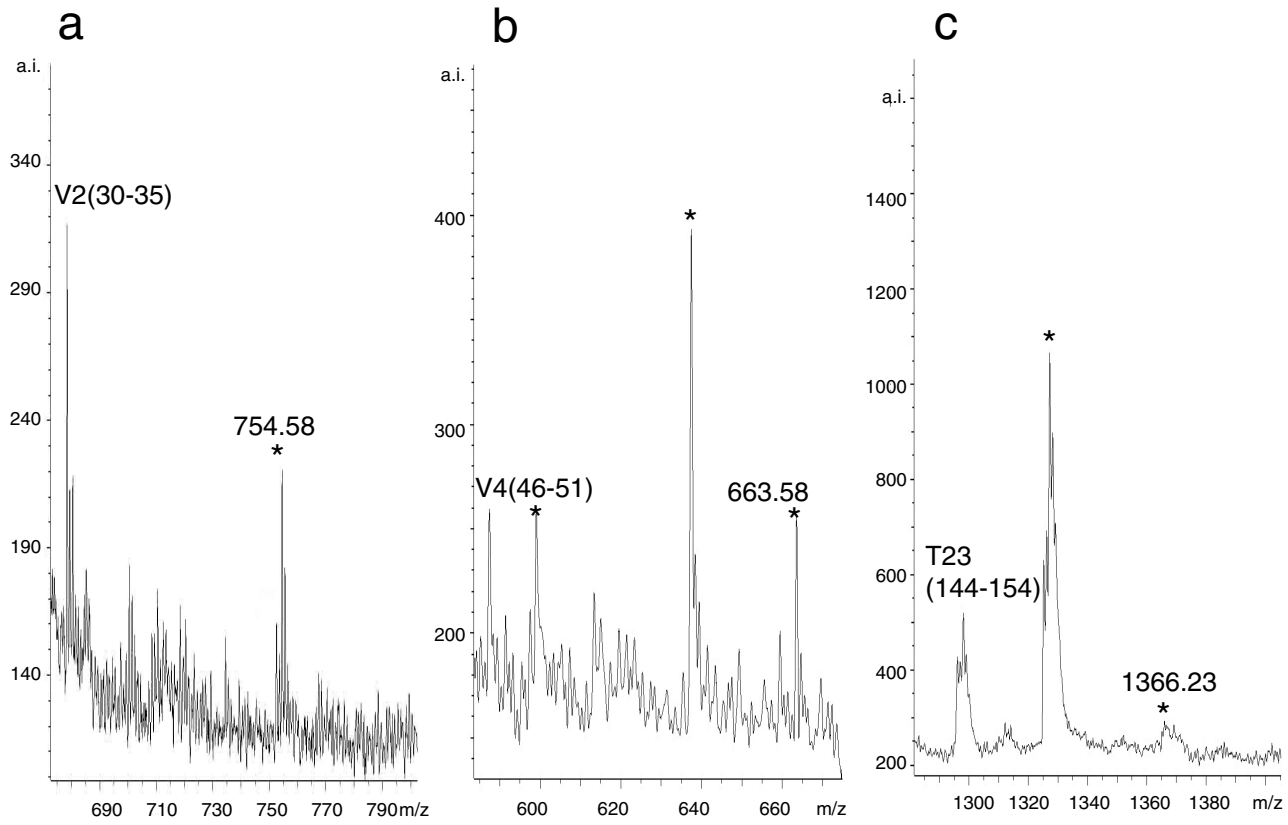

Fig. S3. To examine whether Ser33, Thr48, or Ser146 in CA core is phosphorylated, the mixture of spots derived from CA protein isoforms were subjected to proteome analysisa. Ser_{33} in the V8-digested peptide derived from the CA protein (V2(30-35), Lys_{30}-Glu_{35}) is not phosphorylated. We were not able to detect signals at m/z 758.35, which showed an 80 amu. delta-mass shift from the signals of the unmodified V2(30-35). b, Thr_{48} in the V8-digested peptide derived from the CA protein (V4(46-51), Gly_{46}-Asp_{51}) is not phosphorylated. We were not able to detect signals at m/z 668.26, which showed an 80 amu. delta-mass shift from the signals of the unmodified V4(46-51). c, Ser_{146} in the tryptic peptide derived from the CA protein (T23(144-154), Met_{144}-Arg_{154}) is not phosphorylated. We were not able to detect signals at m/z 1375.67, which showed an 80 amu. delta-mass shift from the signals of the unmodified T23(144-154). * Indicates an unassigned fragment from a sample.

Supplemental figure 4

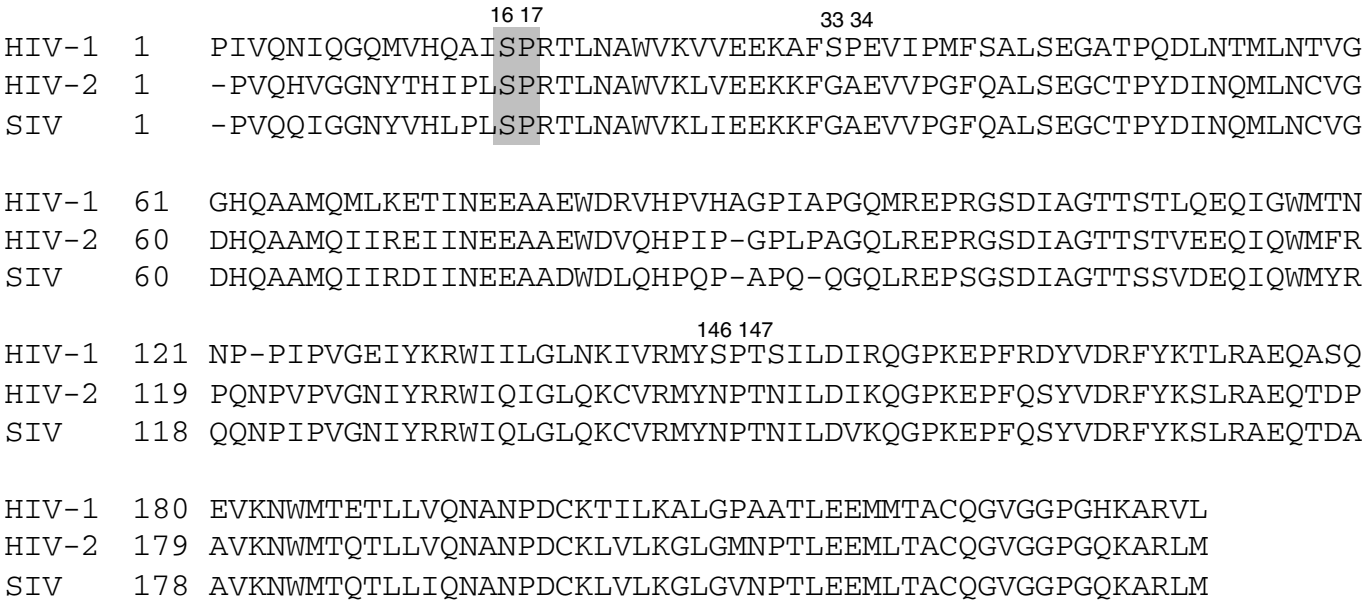

Fig. S4. Location of highly conserved Ser-Pro motifs in lentiviral CA protein. The CA protein sequences of HIV-1_{LAI} (HIV database accession no. A04321), HIV-2_{ROD} (HIV database accession no. X05291), and SIV_{mac239} (HIV database accession no. M33262) were obtained from the HIV database operated by the Division of AIDS of the National Institute of Allergy and Infectious Diseases (NIAID) and aligned with ClustalW (http://www.ebi.ac.uk/clustalw/). Numbers correspond to the positions of Ser-Pro motifs within the HIV-1 CA protein.

Supplemental figure 5

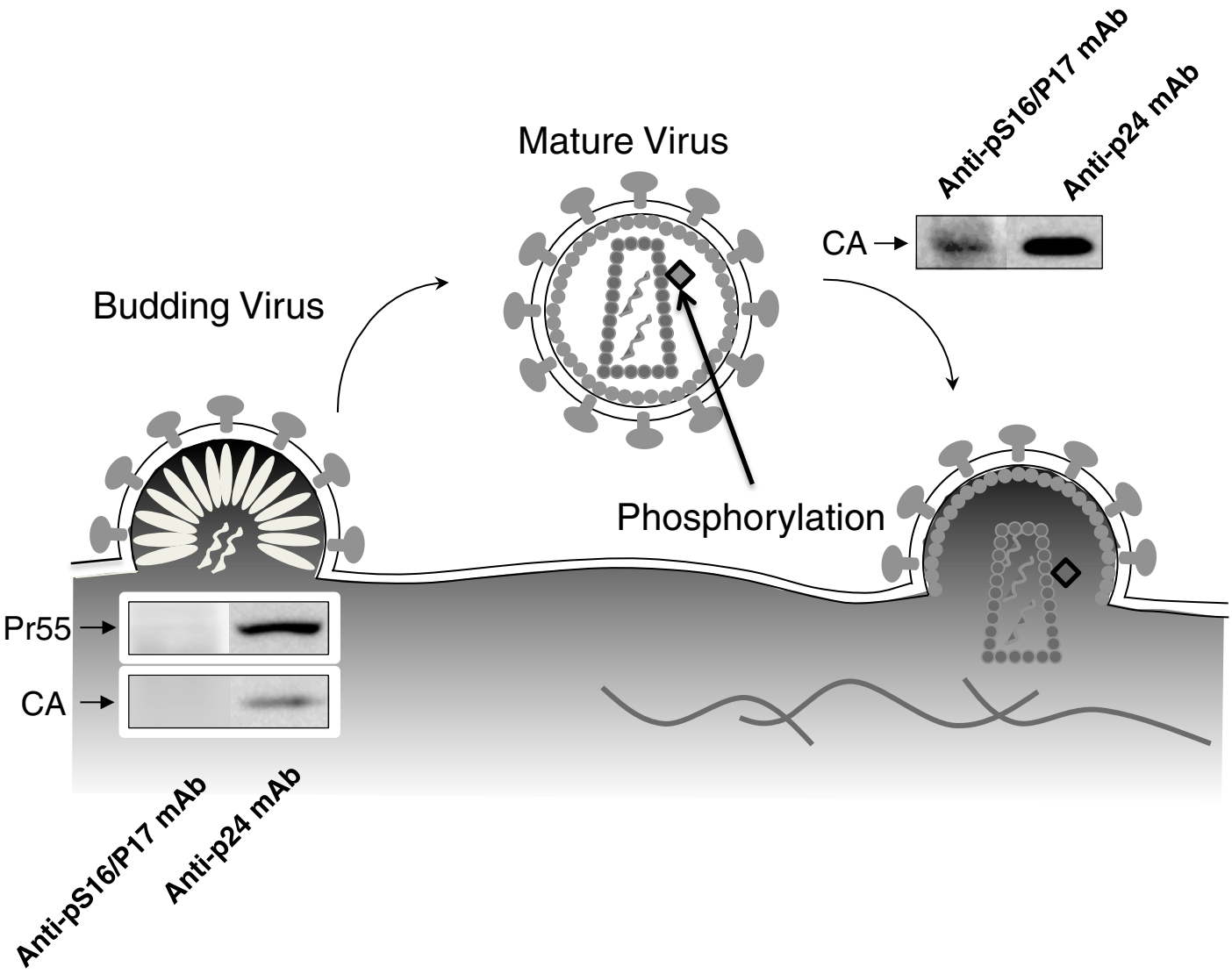

Fig. S5. Detection of phosphorylated Ser_{16} in CA protein. To investigate how Ser_{16} in the CA protein is phosphorylated in the HIV-1 life cycle, HIV-1_{LAV-1} virions and CEM/LAV-1 cells were lysed, and lysates were immunoprecipitated with HIV-1 positive plasma. The immunoprecipitates were analyzed by western blotting using monoclonal anti-p24 mAb or monoclonal anti-pS16/P17 mAb. These results demonstrated that Ser_{16} was phosphorylated in HIV-1_{LAV-1} virions but not in CEM/LAV-1 cells.

Supplemental Table I

Table SI. Comparison of Theoretical Masses and Observed Masses Derived from Tryptic Digests of CA Isoforms

Spot	Tryptic peptide number	Theoretical mass (m/z)	Observed mass (m/z)	Corresponding sequence	Number of missed cleavages
CA-a	T1(1-18)	2016.08	2016.46	PIVQNIQGQMVHQAISPR	0
	T1*(1-18)	2096.08	2096.55	PIVQNIQGQMVHQAIpSPR	0
	T3(19-25)	831.47	831.23	TLNAWVK	0
	T9(71-82)	1462.64	1462.64	ETINEEAAEWDR	0
	T11(83-97)	1566.83	1566.76	VHPVHAGPIAPGQMR	0
	T15(101-131)	3317.64	3317.72	GSDIAGTTSTLQEQIGWMTN NPPIPVGEIYK	0
	T18(132-140)	1112.69	1112.69	RWIILGLNK	1
	T19(133-140)	956.59	956.53	WIILGLNK	0
	T23(144-154)	1295.67	1295.61	MYSPTSILDIR	0
	T27(155-167)	1606.80	1606.90	QGPKEPFRDYVDR	2
	T28(159-162)	548.28	548.22	EPFR	0
	T29(159-167)	1196.57	1196.56	EPFRDYVDR	1
	T30(163-167)	667.30	667.25	DYVDR	0
CA-b	T1(1-18)	2016.08	2016.48	PIVQNIQGQMVHQAISPR	0
	T3(19-25)	831.47	831.37	TLNAWVK	0
	T9(71-82)	1462.64	1462.38	ETINEEAAEWDR	0
	T19(133-140)	956.59	956.46	WIILGLNK	0
	T23(144-154)	1295.67	1295.46	MYSPTSILDIR	0
	T26(155-162)	958.51	958.46	QGPKEPFR	1
	T27(155-167)	1606.80	1606.50	QGPKEPFRDYVDR	2
	T29(159-167)	1196.57	1196.41	EPFRDYVDR	1
	T35(171-182)	1359.72	1359.50	TLRAEQASQEVK	1
CA-c	T1(1-18)	2016.08	2016.46	PIVQNIQGQMVHQAISPR	0
	T3(19-25)	831.47	831.28	TLNAWVK	0
	T9(71-82)	1462.64	1462.63	ETINEEAAEWDR	0
	T10(71-97)	3010.46	3010.67	ETINEEAAEWDRVHPVHAGP IAPGQMR	1
	T11(83-97)	1566.83	1566.82	VHPVHAGPIAPGQMR	0
	T12(83-100)	1949.03	1949.06	VHPVHAGPIAPGQMREPR	1
	T15(101-131)	3317.64	3317.72	GSDIAGTTSTLQEQIGWMTN NPPIPVGEIYK	0
	T16(101-132)	3473.74	3474.07	GSDIAGTTSTLQEQIGWMTN NPPIPVGEIYKR	1
	T18(132-140)	1112.69	1112.69	RWIILGLNK	1
	T19(133-140)	956.59	956.57	WIILGLNK	0
	T23(144-154)	1295.67	1295.66	MYSPTSILDIR	0
	T27(155-167)	1606.80	1606.90	QGPKEPFRDYVDR	2
	T28(159-162)	548.28	548.22	EPFR	0
	T29(159-167)	1196.57	1196.53	EPFRDYVDR	1
	T30(163-167)	667.30	667.25	DYVDR	0
	T32(168-170)	457.24	457.38	FYK	0
	T34(171-173)	389.25	389.23	TLR	0
	T35(171-182)	1359.72	1359.85	TLRAEQASQEVK	1
CA-d	T1(1-18)	2016.08	2016.14	PIVQNIQGQMVHQAISPR	0
	T1**(1-18)	2044.08	2044.19	fPIVQNIQGQMVHQAISPR[a)]	0
	T3(19-25)	831.47	831.28	TLNAWVK	0
	T9(71-82)	1462.64	1462.63	ETINEEAAEWDR	0
	T10(71-97)	3010.46	3010.67	ETINEEAAEWDRVHPVHAGP IAPGQMR	1
	T11(83-97)	1566.83	1566.82	VHPVHAGPIAPGQMR	0
	T12(83-100)	1949.03	1949.06	VHPVHAGPIAPGQMREPR	1
	T15(101-131)	3317.64	3317.72	GSDIAGTTSTLQEQIGWMTN NPPIPVGEIYK	0
	T16(101-132)	3473.74	3474.07	GSDIAGTTSTLQEQIGWMTN NPPIPVGEIYKR	1
	T18(132-140)	1112.69	1112.69	RWIILGLNK	1
	T19(133-140)	956.59	956.57	WIILGLNK	0
	T23(144-154)	1295.67	1295.66	MYSPTSILDIR	0
	T27(155-167)	1606.80	1606.90	QGPKEPFRDYVDR	2
	T28(159-162)	548.28	548.22	EPFR	0
	T29(159-167)	1196.57	1196.53	EPFRDYVDR	1
	T30(163-167)	667.30	667.25	DYVDR	0
	T32(168-170)	457.24	457.38	FYK	0
	T34(171-173)	389.25	389.23	TLR	0
	T35(171-182)	1359.72	1359.85	TLRAEQASQEVK	1

a) fP indicates a formylated proline residue.