Supporting information

From Alphafold to site-selective inhibition of DNA methylation in glioblastoma multiforme

Jean-Maxime Besson,[a] Alain Chavanieu,[a] Arulraj Nadaradjane,[b c d] Pierre-François Cartron,[b c d] and Marie Lopez* [a]

[a] Institut des Biomolécules Max Mousseron (IBMM), UMR 5247, CNRS-Université de Montpellier-ENSCM, 1919 route de Mende, 34296 Montpellier cedex 5, France

[b] Nantes Université, Inserm UMR 1307, CNRS UMR 6075, Université d'Angers, CRCI2NA, F-44000 Nantes – France

[c] Institut de Cancérologie de l'Ouest, F-44805 Saint-Herblain, France

[d] EpICO – Plateforme Epigénétique/Epitranscriptomique, Institut de Cancérologie de l'Ouest, F-44805, Saint-Herblain, France

I. Protein visualisation and AlphaFold multimer prediction

AlphaFold Prediction

AlphaFold2 complex prediction. Protein-protein complex structure prediction was performed using the public available ColabFold (Mirdita *et al.* 2022; Jumper *et al.* 2021; Evans *et al.* 2021) https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb#scrollTo=mbaIO9pWjaN0). Amino acid sequences for human DNMT1 (UniProt : AF-P26358-F1, 1-1616), DMAP1 (UniProtKB: AF-Q9NPF5-F1, 1-467), USP7 (UniProt: AF-Q9009-F1, 1-1102), CFP1 (UniProt, AF-Q9P0U4-F1,1-656), UHRF1 (UniProt: AF-Q96T88-F1, 1-793), Sp1 (UniProt, AF-P08047-F1,1-785) and Stat3 (UniProt:AF-P40763-F1, 1-770) were retrieved from UniProt. Previously proposed in literature interactions sequences were pasted to colab_batch connected by ":" using the fasta file from Uniprot. The default parameters for MSA generations and the number of recycles (3) were used. For relaxed prediction, only the first model was relaxed using num_relax: 1. The ColabFold was executed using T4 GPU under ColabFold pro account. Five models were generated for each prediction with the pLDDT (predicted Local Distance Difference Test) score (Mariani *et al.* 2013), a per-residue confidence score between 0 and 100. Generally, pLDDT scores above 90 are considered ''very confident,'' pLDDT scores between 70 and 90 are considered ''confident,'' pLDDT scores between 50 and 70 are ''low confidence'' and scores below 50 are considered ''very low confidence'' and may be unstructured. pLDDT score of each predicted model was visualized in Pymol by using the following github https://github.com/cbalbin-bio/pymol-color-alphafold . and by writing on the command prompt the following code:

```
run https://raw.githubusercontent.com/cbalbin-bio/pymol-color-alphafold/master/coloraf.py
from pymol import cmd
def coloraf(selection="all"):
   coloraf sele
   sele (all)
   cmd.color("blue", f"({selection}) and b > 90")
   cmd.color("cyan", f"({selection}) and b < 90 and b > 70")
   cmd.color("yellow", f"({selection}) and b < 70 and b > 50")
   cmd.color("orange", f"({selection}) and b < 50")
cmd.extend("coloraf", coloraf)
cmd.auto_arg[0]["coloraf"] = [cmd.object_sc, "object", ""]
```

ColabFold ranks the five models based on pLDDT scores, pTM (predicted Template Modelling), and ipTM (interface predicted Template Modelling). A pTM score above 0.5 indicates good overall folding of the complex, while a score below 0.5 suggests an incorrect structure. For the ipTM score, values above 0.8 denote high-confidence, high-quality predictions. Scores between 0.6 and 0.8 should be interpreted with caution, as the predictions may be accurate or incorrect. Scores below 0.6 likely indicate a failed prediction. To evaluate how a prediction is reliable or not, a model confidence score could be calculated as follow:

$$Model\ confidence = (0{,}2 \times pTM + 0{,}8 \times ipTM) \times 100$$

A heat-map representation of the MSA indicates all sequences mapped. The colour scale indicates the identity score. White regions are not covered. The black line identifies the relative coverage of the sequence with respect to the total number of aligned sequences. Top-ranked model from the generated PDB file was used for further molecular visualization using Pymol. Each protein was distinguished by coloration. Protein-protein interaction was determined using Modify/Around/Atoms within 4 Å. Synthetised peptides were represented using stick. Pymol picture was downloaded using Ray (slow) at 300 DPI.

Figure SI-1. *FASTA alignments of the protein sequence from PDB structure (PDB code: 4wxx, 351-1600) align with full sequence of AF model (UniProt : AF-P26358-F1, 1-1616) of DNMT1.*

4wxx_model_1

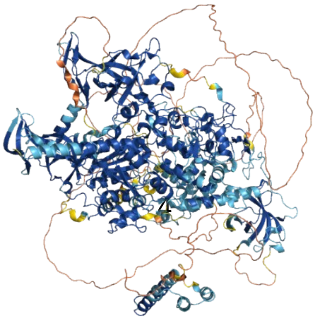
```
-----------------------------------------------------------------------------------------------------------------------------

-------------------------------------------------------------------------------------------------------------------------------------------------------------
----------------------------------------------------------------------------------PKCIQCGQYLDDPDLKYGQHPPDAVDEPQMLTNE----
-----------
KALPQHKLTCFSVYCKHGHLCPIDTGLIEKNIELFFSGSAKPIYDDDPSLEGGVNGKNLGPINEWWITGFDGGEKALI
GFSTSFAEYILMDPSPEYAPIFGLMQEKIYISKIVVEFLQSNSDSTYEDLINKIETTVPPSGLNLNRFTEDSLLRHAQFV
VEQVESYDEAGDSDEQPIFLTPCMRDLIKLAGVTLGQRRAQARRQT-------------KATTTKLVYQIFDTFFAEQIEK------
---
KRRRCGVCEVCQQPECGKCKACKDMVKFGGSGRSKQACQERRCPNMAMKEADDDEEVDDNIPEMPSPKKMHQG
KKKKQNKNRISWVGEAVKTDKKSYYKKVCIDAETLEVGDCVSVIPDDSSKPLYLARVTALWEDSSNGQMFHAHW
FCAGTDTVLGATSDPLELFLVDECEDMQLSYIHSKVKVIYKAPSENWAMEGGMDPESLLEGDDGKTYFYQLWYDQ
DYARFESPPKTQPTEDNKFKFCVSCARLAEMRQKEIPRVLEQLEDLDSRVLYYSATKNGILYRVGDGVYLPPEAFTF
NI---------RKEPVDEDLYPEHYRKY--------
LDAPEPYRIGRIKEIFCPKKSNGRPNETDIKIRVNKFYRPENTHKSTPASYHADINLLYWSDEEAVVDFKAVQGRCTV
EYGEDLPECVQVYSMGGPNRFYFLEAYNAKSKSFEDPPNHARS----------------------------
KLPKLRTLDVFSGCGGLSEGFHQAGISDTLWAIEMWDPAAQAFRLNNPGSTVFTEDCNILLKLVMAGETTNSRGQR
LPQKGDVEMLCGGPPCQGFSGMNRFNSRTYSKFKNSLVVSFLSYCDYYRPRFFLLENVRNFVSFKRSMVLKLTLRC
LVRMGYQCTFGVLQAGQYGVAQTRRRAIILAAAPGEKLPLFPEPLHVFAPRACQLSVVVDDKKFVSNITRLSSGPFR
TITVRDTMSDLPEVRNGASALEISYNGEPQSWFQRQLRGAQYQPILRDHICKDMSALVAARMRHIPLAPGSDWRDL
PNIEVRLSDGTMARKLRYTHHDRKNGRSSSGALRGVCSCVEAGKACDPAARQFNTLIPWCLPHTGNRHNHWAGLY
GRLEWDGFFSTTVTNPEPMGKQGRVLHPEQHRVVSVRECARSQGFPDTYRLFGNILDKHRQVGNAVPPPLAKAIGL
EIKLCMLAKALTRVW-----------
```

sp|P26358|DNMT1_HUMAN DNA (cytosine-5)-methyltransferase 1 OS=Homo sapiens OX=9606 GN=DNMT1 PE=1 SV=2

```
MPARTAPARVPTLAVPAISLPDDVRRRLKDLERDSLTEKECVKEKLNLLHEFLQTEIKNQLCDLETKLRKEELSEEG
YLAKVKSLLNKDLSLENGAHAYNREVNGRLENGNQARSEARRVGMADANSPPKPLSKPRTPRRSKSDGEAKPEPSP
SPRITRKSTRQTTITSHFAKGPAKRKPQEESERAKSDESIKEEDKDQDEKRRRVTSRERVARPLPAEEPERAKSGTRTE
KEEERDEKEEKRLRSQTKEPTPKQKLKEEPDREARAGVQADEDEDGDEKDEKKHRSQPKDLAAKRRPEEKEPEKV
NPQISDEKDEDEKEEKRRKTTPKEPTEKKMARAKTVMNSKTHPPKCIQCGQYLDDPDLKYGQHPPDAVDEPQMLT
NEKLSIFDANESGFESYEALPQHKLTCFSVYCKHGHLCPIDTGLIEKNIELFFSGSAKPIYDDDPSLEGGVNGKNLGPI
NEWWITGFDGGEKALIGFSTSFAEYILMDPSPEYAPIFGLMQEKIYISKIVVEFLQSNSDSTYEDLINKIETTVPPSGLN
LNRFTEDSLLRHAQFVVEQVESYDEAGDSDEQPIFLTPCMRDLIKLAGVTLGQRRAQARRQTIRHSTREKDRGPTKA
TTTKLVYQIFDTFFAEQIEKDDREDKENAFKRRRCGVCEVCQQPECGKCKACKDMVKFGGSGRSKQACQERRCPN
MAMKEADDDEEVDDNIPEMPSPKKMHQGKKKKQNKNRISWVGEAVKTDGKKSYYKKVCIDAETLEVGDCVSVIP
DDSSKPLYLARVTALWEDSSNGQMFHAHWFCAGTDTVLGATSDPLELFLVDECEDMQLSYIHSKVKVIYKAPSEN
WAMEGGMDPESLLEGDDGKTYFYQLWYDQDYARFESPPKTQPTEDNKFKFCVSCARLAEMRQKEIPRVLEQLEDL
DSRVLYYSATKNGILYRVGDGVYLPPEAFTFNIKLSSPVKRPRKEPVDEDLYPEHYRKYSDYIKGSNLDAPEPYRIGR
IKEIFCPKKSNGRPNETDIKIRVNKFYRPENTHKSTPASYHADINLLYWSDEEAVVDFKAVQGRCTVEYGEDLPECV
QVYSMGGPNRFYFLEAYNAKSKSFEDPPNHARSPGNKGKGKGKGKGKPKSQACEPSEPEIEIKLPKLRTLDVFSGCG
GLSEGFHQAGISDTLWAIEMWDPAAQAFRLNNPGSTVFTEDCNILLKLVMAGETTNSRGQRLPQKGDVEMLCGGPP
CQGFSGMNRFNSRTYSKFKNSLVVSFLSYCDYYRPRFFLLENVRNFVSFKRSMVLKLTLRCLVRMGYQCTFGVLQA
GQYGVAQTRRRAIILAAAPGEKLPLFPEPLHVFAPRACQLSVVVDDKKFVSNITRLSSGPFRTITVRDTMSDLPEVRN
GASALEISYNGEPQSWFQRQLRGAQYQPILRDHICKDMSALVAARMRHIPLAPGSDWRDLPNIEVRLSDGTMARKL
RYTHHDRKNGRSSSGALRGVCSCVEAGKACDPAARQFNTLIPWCLPHTGNRHNHWAGLYGRLEWDGFFSTTVTNP
EPMGKQGRVLHPEQHRVVSVRECARSQGFPDTYRLFGNILDKHRQVGNAVPPPLAKAIGLEIKLCMLAKARESASA
KIKEEEAAKD
```

Predicted unstructured regions list on Alphafold model missing in PDB structure of DNMT1 : 1-19; 93-347 ; 389-399 ; 603-614 ; 639-648 ; 952-959 ; 978-984 ; 1106-1134

Figure SI-2. *Visualization of* *DNMT1 pLDDT score using the*



adapted plugin obtained from the following URL: https://github.com/APAJanssen/Alphafold2import.

***Figure SI-3**. Details of USP7 AF (UniProt: AF-Q9009-F1, 1-1102) model highlighting in purple area the identified interacting region with DNMT1 (740-765) as previously reported (Xu et al., 2015). The extended region 656-791 represented in orange and purple was used to construct the complex model with identified DNMT1 interaction.*

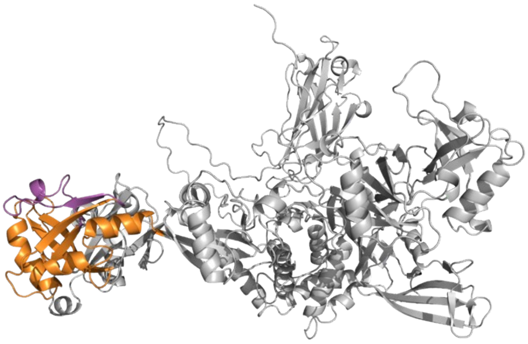


Figure SI-4. *Prediction of USP7(676-793) / DNMT1(1102-1137) complex using ColabFold.*

A. Heat map representing Multiple Sequence Alignment (MSA) using MMseq. The sequence coverage is shown along the Y-axis indicating the number of sequences under study. The X-axis represents amino acid positions. The color gradient from blue to red signifies the identity score of each studied sequence, where blue represents the highest identity scores (1.0) and red indicates the lowest identity scores (0.0). Each color line represents one sequence from MSA. **B.** pLDDT scores for each position across a protein sequence, as determined by five predictive models (Rank 1 to 5). The Y-axis quantifies the predicted IDDT score, and the X-axis displays the sequence positions. The black line represents the limit between both protein sequences.

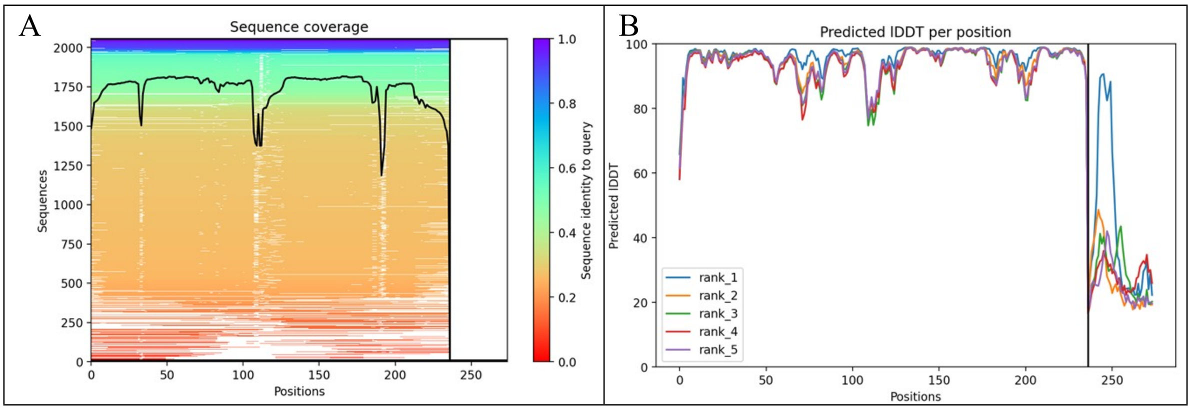


Figure SI-5. *Prediction of USP7(676-793) / DNMT1(20-95; 349-1600) complex using ColabFold.*

A. Heat map representing Multiple Sequence Alignment (MSA) using MMseq. The sequence coverage is shown along the Y-axis indicating the number of sequences under study. The X-axis represents amino acid positions. The color gradient from blue to red signifies the identity score of each studied sequence, where blue represents the highest identity scores (1.0) and red indicates the lowest identity scores (0.0). Each color line represents one sequence from MSA. **B.** pLDDT scores for each position across a protein sequence, as determined by five predictive models (Rank 1 to 5). The Y-axis quantifies the predicted IDDT score, and the X-axis displays the sequence positions. The black line represents the limit between both protein sequences.

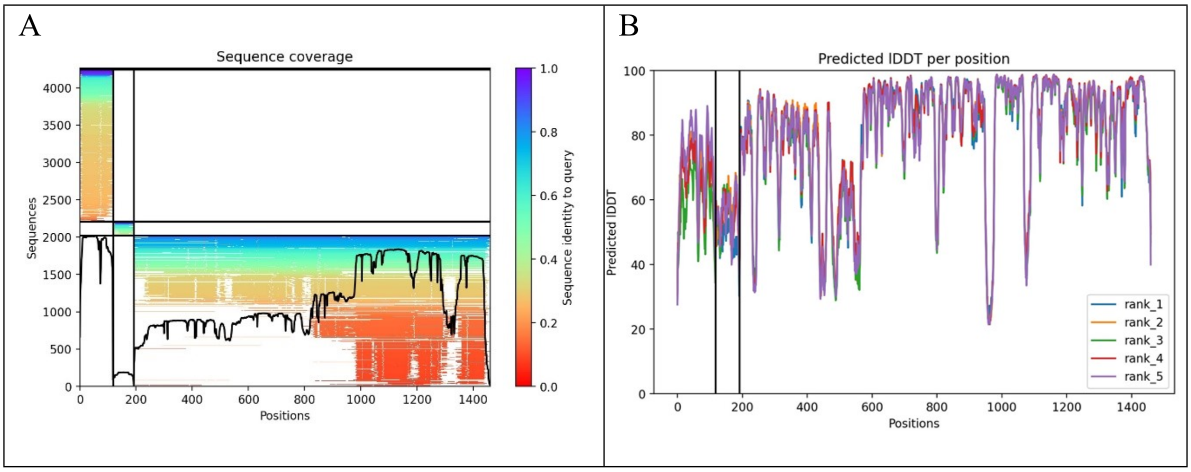


Figure SI-6. *Details of UHRF1 AF (UniProt: AF-Q96T88-F1, 1-793) model highlighting in purple area the identified interacting region with DNMT1 (435-586) as previously reported (Achour et al., 2008). The extended region represented in cyan and purple was used to construct the complex model with identified DNMT1 interaction.*

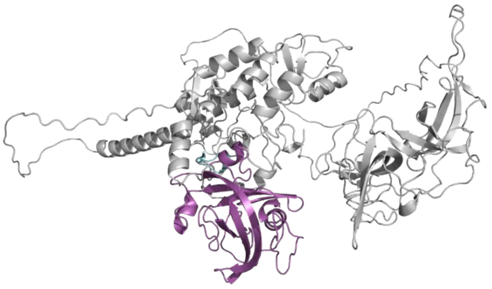


Figure SI-7. *Prediction of UHRF1(426-586) / DNMT1(375-594) complex using ColabFold.*

A. Heat map representing Multiple Sequence Alignment (MSA) using MMseq. The sequence coverage is shown along the Y-axis indicating the number of sequences under study. The X-axis represents amino acid positions. The color gradient from blue to red signifies the identity score of each studied sequence, where blue represents the highest identity scores (1.0) and red indicates the lowest identity scores (0.0). Each color line represents one sequence from MSA.

B. pLDDT scores for each position across a protein sequence, as determined by five predictive models (Rank 1 to 5). The Y-axis quantifies the predicted IDDT score, and the X-axis displays the sequence positions. The black line represents the limit between both protein sequences.

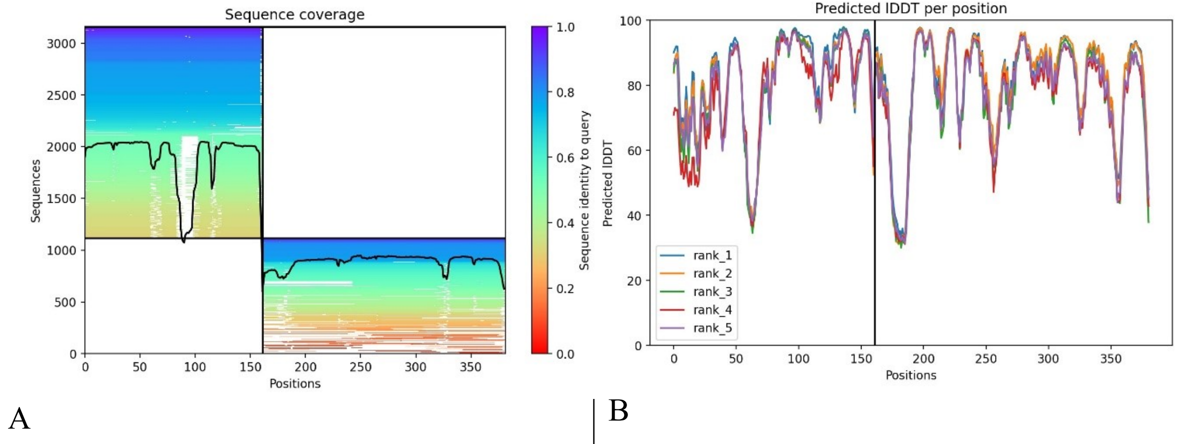


Figure SI-8. *Details of Sp1 AF (UniProt, AF-P08047-F1,1-785) model highlighting in purple area the identified interacting region with DNMT1 (56-67) as previously reported (Estève et al., 2006). The extended region 54-70 represented in forest green and purple was used to construct the complex model with identified DNMT1 interaction.*

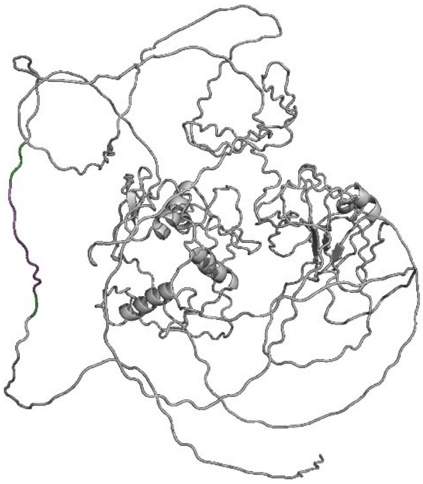


Figure SI-9. *Prediction of Sp1(54-70) / DNMT1(730-907) complex using ColabFold.*

A. Heat map representing Multiple Sequence Alignment (MSA) using MMseq. The sequence coverage is shown along the Y-axis indicating the number of sequences under study. The X-axis represents amino acid positions. The color gradient from blue to red signifies the identity score of each studied sequence, where blue represents the highest identity scores (1.0) and red indicates the lowest identity scores (0.0). Each color line represents one sequence from MSA.
B. pLDDT scores for each position across a protein sequence, as determined by five predictive models (Rank 1 to 5). The Y-axis quantifies the predicted IDDT score, and the X-axis displays the sequence positions. The black line represents the limit between both protein sequences.

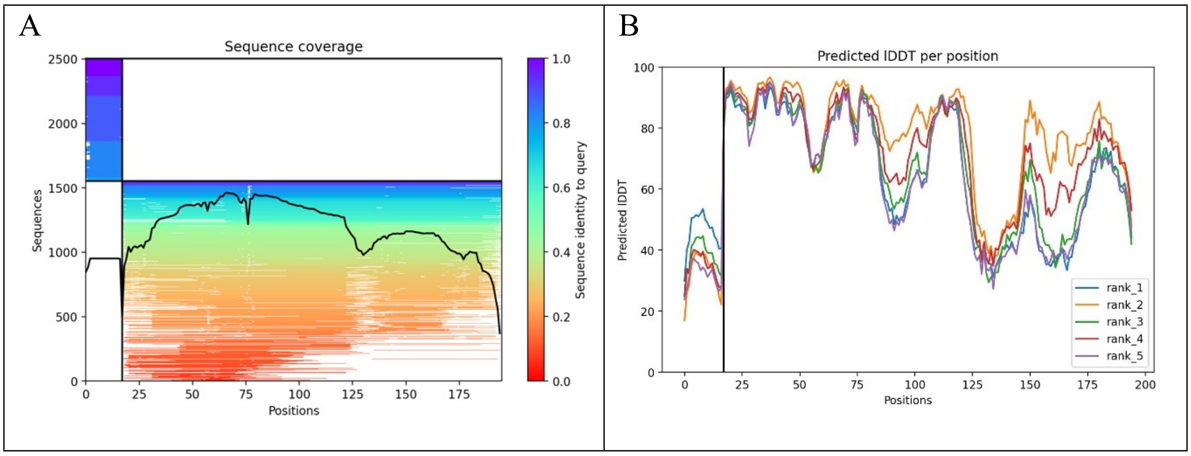


Figure SI-10. *Details of CFP1 AF (UniProt, AF-Q9P0U4-F1,1-656) model highlighting in purple area the identified interacting region with DNMT1 (163-213) as previously reported (Hashimoto et al., 2010). The extended region 165-215 represented in salmon and purple was used to construct the complex model with identified DNMT1 interaction.*

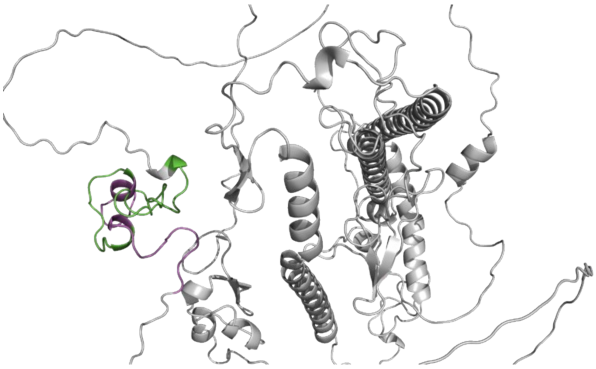


Figure SI-11. *Prediction of CFP1(165-215) / DNMT1(349-1600) complex using ColabFold.*

A. Heat map representing Multiple Sequence Alignment (MSA) using MMseq. The sequence coverage is shown along the Y-axis indicating the number of sequences under study. The X-axis represents amino acid positions. The color gradient from blue to red signifies the identity score of each studied sequence, where blue represents the highest identity scores (1.0) and red indicates the lowest identity scores (0.0). Each color line represents one sequence from MSA.

B. pLDDT scores for each position across a protein sequence, as determined by five predictive models (Rank 1 to 5). The Y-axis quantifies the predicted IDDT score, and the X-axis displays the sequence positions. The black line represents the limit between both protein sequences.

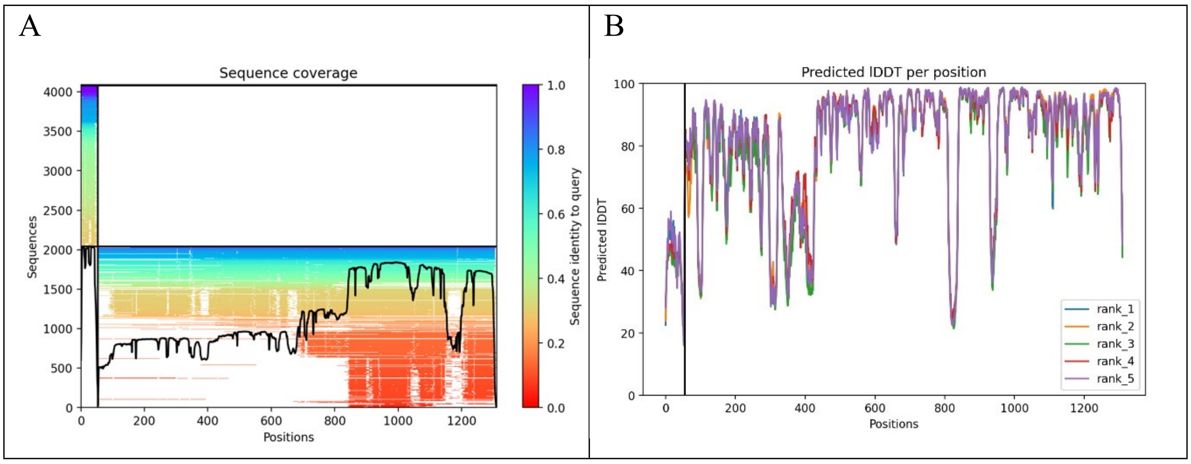


Figure SI-12. *Details of Stat3 AF (UniProt:AF-P40763-F1, 1-770) model highlighting in purple area a peptide enables to disrupt the interaction with DNMT1 (677-693) as previously reported (Cheray et al., 2014).*

Figure SI-13. *Prediction of Stat3 (498-770) / Stat3 (498-770) complex using ColabFold and relaxation mode for rank 1.*

A. Heat map representing Multiple Sequence Alignment (MSA) using MMseq. The sequence coverage is shown along the Y-axis indicating the number of sequences under study. The X-axis represents amino acid positions. The color gradient from blue to red signifies the identity score of each studied sequence, where blue represents the highest identity scores (1.0) and red indicates the lowest identity scores (0.0). Each color line represents one sequence from MSA.

B. pLDDT scores for each position across a protein sequence, as determined by five predictive models (Rank 1 to 5). The Y-axis quantifies the predicted IDDT score, and the X-axis displays the sequence positions. The black line represents the limit between both protein sequences.

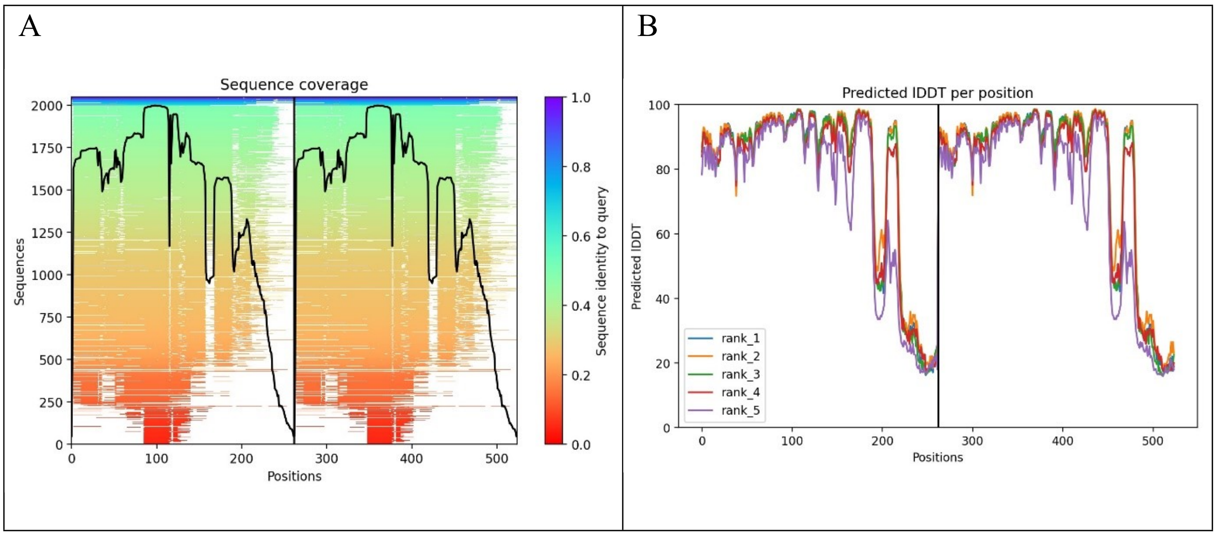


Figure SI-14. *Dimeric complex of Stat3, in chocolate, predicted by AF2.*

A. The model highlighting in hot pink area a peptide enables to disrupt the interaction with DNMT1 (677-693) as previously reported (Cheray et *al.*, 2014).
B. Focus on PPST2213 conformation. The stick representation highlights amino acids close to P^{678} and P^{689} likely to be involved in conformational changes due to K^{685} acetylation, essential for DNMT1 interaction as previously reported (REF). The AF model was superimposed with X-Ray structure (PDB code: 1BG1) and presents a RMSD of 2,06 A.

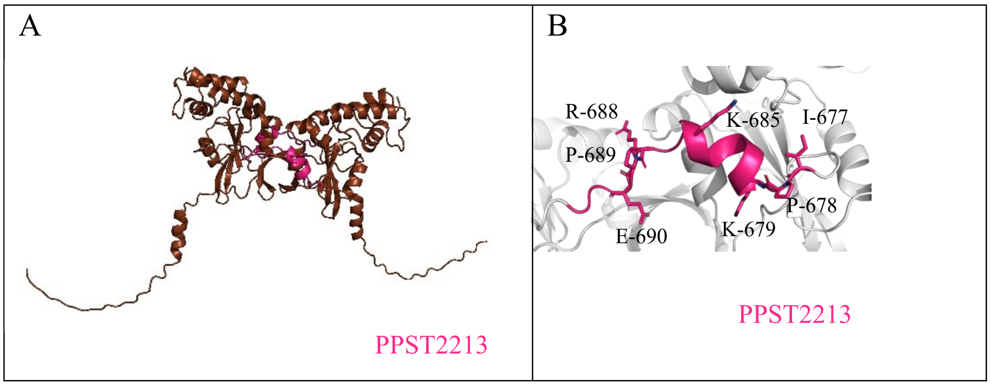


Figure SI-15. *Prediction of Stat32(498-720) / DNMT1(349-1616) complex using ColabFold and relaxation mode for rank 1.*

A. Heat map representing Multiple Sequence Alignment (MSA) using MMseq. The sequence coverage is shown along the Y-axis indicating the number of sequences under study. The X-axis represents amino acid positions. The color gradient from blue to red signifies the identity score of each studied sequence, where blue represents the highest identity scores (1.0) and red indicates the lowest identity scores (0.0). Each color line represents one sequence from MSA.
B. pLDDT scores for each position across a protein sequence, as determined by five predictive models (Rank 1 to 5). The Y-axis quantifies the predicted IDDT score, and the X-axis displays the sequence positions. The black line represents the limit between both protein sequences.

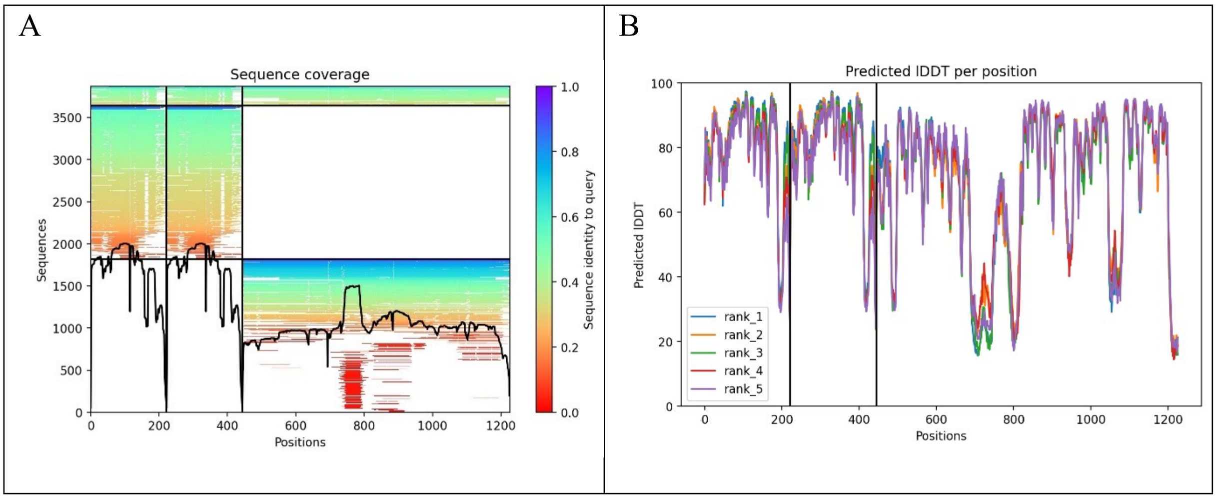


Figure SI-16. *Details of DMAP1 AF (UniProtKB: AF-Q9NPF5-F1, 1-467) model highlighting in purple area the identified interacting region with DNMT1 (212-272) as previously reported (Rountree et al., 2010). The extended region 140-290 represented in green and purple was used to construct the complex model with identified DNMT1 interaction.*

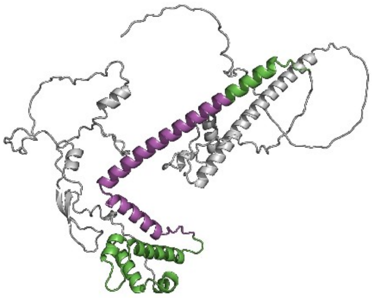


Figure SI-17. *Prediction of DNMT1(20-95) / DMAP1(140-290) complex using ColabFold.*

A. Heat map representing Multiple Sequence Alignment (MSA) using MMseq. The sequence coverage is shown along the Y-axis indicating the number of sequences under study. The X-axis represents amino acid positions. The color gradient from blue to red signifies the identity score of each studied sequence, where blue represents the highest identity scores (1.0) and red indicates the lowest identity scores (0.0). Each color line represents one sequence from MSA.
B. pLDDT scores for each position across a protein sequence, as determined by five predictive models (Rank 1 to 5). The Y-axis quantifies the predicted IDDT score, and the X-axis displays the sequence positions. The black line represents the limit between both protein sequences.

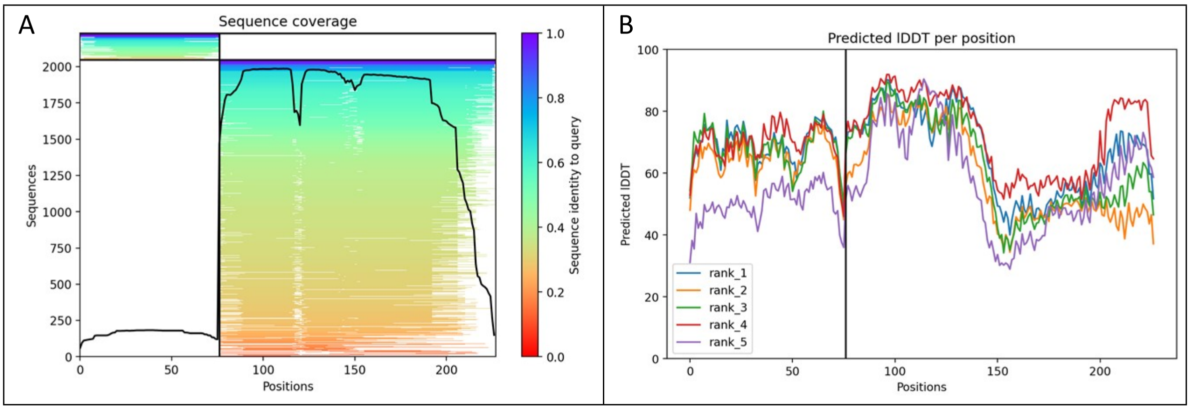
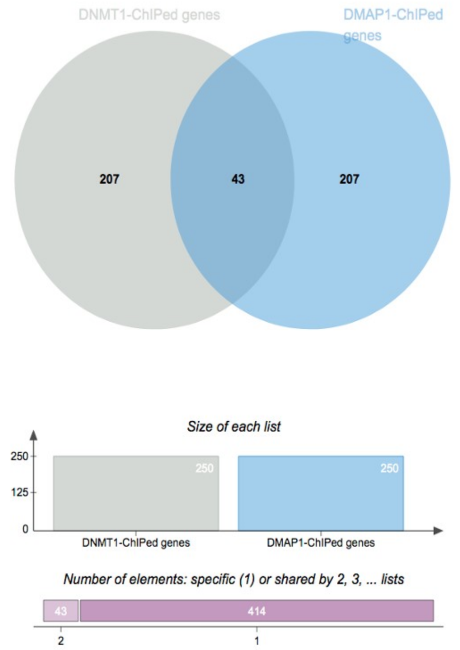
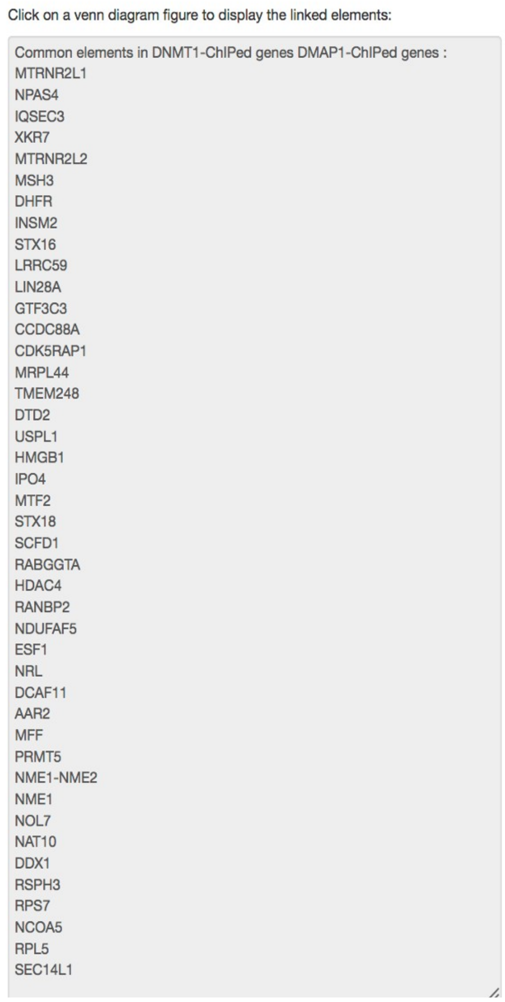


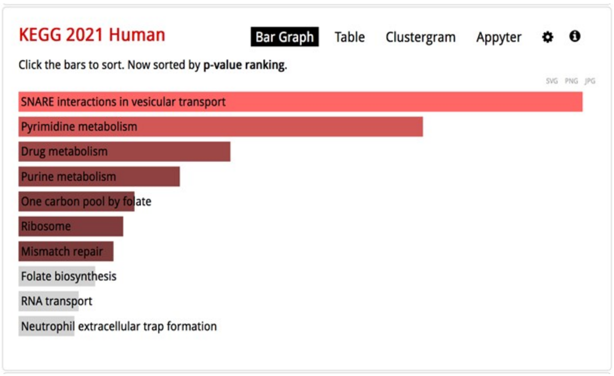
Figure SI-18. *Computational identification of methylation specific site mediated by DNMT1/DMAP1.*





Click on a venn diagram figure to display the linked elements:

Common elements in DNMT1-ChIPed genes DMAP1-ChIPed genes :
MTRNR2L1
NPAS4
IQSEC3
XKR7
MTRNR2L2
MSH3
DHFR
INSM2
STX16
LRRC59
LIN28A
GTF3C3
CCDC88A
CDK5RAP1
MRPL44
TMEM248
DTD2
USPL1
HMGB1
IPO4
MTF2
STX18
SCFD1
RABGGTA
HDAC4
RANBP2
NDUFAF5
ESF1
NRL
DCAF11
AAR2
MFF
PRMT5
NME1-NME2
NME1
NOL7
NAT10
DDX1
RSPH3
RPS7
NCOA5
RPL5
SEC14L1



II. Peptide characterisation

PPUS2202 (-CONH$_2$, USP7, 755-765, NH$_2$-Ala^1-Leu^2-Asp^3-Glu^4-Leu^5-Met^6-Asp^7-Asp^8-Gly^9-Asp^{10}-Ile^{11}-CONH$_2$). **PPUS2202** was synthetised using general method B and purified *via* method A (28 mg, 23 %). **HPLC**: $t_R = 3.65$ min, $\%_X = 99$ %; **HPLC-MS (APCI)**: calcd for $[C_{49}H_{79}N_{11}O_{22}S + H]^+$ m/z= 1205.51, found m/z= 1206.30, calcd for $[M+2H]^{2+}$ m/z= 603.76, found m/z= 603.90, $t_R = 2.52$ min.

PPUS2212 (-CONH$_2$, USP7, 735-745, NH$_2$-Tyr^1-Glu^2-Glu^3-Val^4-Lys(Alloc)^5-Pro^6-Asn^7-Leu^8-Thr^9-Glu^{10}-Arg^{11}-CONH$_2$) **PPUS2212** was synthetised using general method A and purified *via* method A (90 mg, 62 %). **HPLC**: t_R = 3.53 min; $\%_X$ = 99 %; **HPLC-MS (APCI)**: calcd for $[C_{64}H_{101}N_{17}O_{22}+H]^+$ m/z=1460.74, found m/z= 1461.10, calcd for $[C_{64}H_{101}N_{17}O_{22}+2H]^{2+}$ m/z= 730.87, found m/z= 731.45, t_R = 2.23 min.

PPUH2206 (-CONH$_2$, UHRF1, 734-741, NH$_2$-Pro^1-Val^2-Arg^3-Val^4-Val^5-Arg^6-Asn^7-Val^8-CONH$_2$). **PPUH2206** was synthetised using general method A and purified *via* method A (87 mg, 75 %). **HPLC**: t_R = 2.99 min, $\%_X$ = 75 %; **HPLC-MS (APCI)**: calcd for $[C_{41}H_{76}N_{16}O_9+H]^+$ m/z= 937.61, found m/z= 937.50, calcd for $[C_{41}H_{76}N_{16}O_9+2H]^{2+}$ m/z= 469.30, found m/z= 469.50, t_R = 1.51 min.

PPSP2201 (-CONH$_2$, Sp1, 56-67, NH$_2$-Ser^1-Gln^2-Pro^3-Ser^4-Pro^5-Ala^6-Leu^7-Leu^8-Ala^9-Ala^{10}-Thr^{11}-CONH$_2$). **PPSP2201** was synthetised using general method A and purified *via* method A (90 mg, 77%). **HPLC**: t_R = 3.95 min, $\%_X$ = 99 %; **HPLC-MS (APCI)**: calcd for $[C_{52}H_{90}N_{14}O_{16}+H]^+$ m/z= 1167.67, found m/z= 1167.45, calcd for $[C_{52}H_{90}N_{14}O_{16}+2H]^{2+}$ m/z= 584.34, found m/z= 584.55, t_R = 2.69 min.

PPCF2209 (-CONH$_2$, CFP1, 205-215, NH$_2$-Leu^1-Arg^2-Gln^3-Cys^4-Gln^5-Leu^6-Arg^7-Ala^8-Arg^9-Glu^{10}-Ser^{11}-CONH$_2$). **PPCF2209** was synthetised using general method A and purified *via* method B (9 mg, 10 %). **HPLC** t_R = 2.99 min, $\%_X$ = 95 %; **HPLC-MS (APCI)**: calcd for $[C_{54}H_{99}N_{23}O_{16}S+H]^+$ m/z= 1358.94, found m/z= 1358.80, $[C_{54}H_{99}N_{23}O_{16}S +2H]^{2+}$ m/z= 679.87, found m/z= 680.35, t_R = 1.4 min.

PPCF2214 (-CONH$_2$, CFP1, 195-205, NH$_2$-Gly^1-Pro^2-Asn^3-Lys^4-Ile^5-Arg^6-Gn^7-Lys^8-Cys^9-Arg^{10}-Leu^{11}-CONH$_2$). **PPCF2214** was synthetised using general method A and purified *via* method A (18 mg, 13 %). **HPLC** t_R = 3.96 min, $\%_X$ = 95 %; **HPLC-MS (APCI)**: calcd for $[C_{63}H_{110}N_{22}O_{17}S +2H]^{2+}$ m/z= 740.41, found m/z= 740.95, calcd for $[C_{63}H_{110}N_{22}O_{17}S +3H]^{3+}$ m/z= 493.94 found m/z= 494.30, t_R = 2.42 min.

PPSt2213 (-CONH$_2$, Stat3, 677-693, NH$_2$-Ile^1-Pro^2-Lys(Alloc)^3-Glu^4-Glu^5-Ala^6-Phe^7-Gly^8-Lys (Alloc)^9-Tyr^{10}-Cys^{11}-Arg^{12}-Pro^{13}-Glu^{14}-Ser^{15}-Gln^{16}-Glu^{17}-CONH$_2$). **PPSP2213** was synthetised using general method A and purified *via* method A (28 mg, 10 %). **HPLC-MS (APCI)**: calcd for $[C_{96}H_{144}N_{24}O_{32}S+2H]^{2+}$ m/z= 1089.51, found m/z= 1090.25, calcd for $[C_{96}H_{144}N_{24}O_{32}S+3H]^{3+}$ m/z= 726.67, found m/z= 727.05, t_R = 2.63 min; **HPLC** t_R = 3.92 min, $\%_X$ = 99 %.

PPDM2203 (-CONH$_2$, DMAP1, 235-245, NH$_2$-Asn^1-Arg^2-Thr^3-Pro^4-Glu^5-Gln^6-Val^7-Ala^8-Glu^9-Glu^{10}-Glu^{11}-CONH$_2$). **PPDM2203** was synthetised using general method A and purified *via* method A (53 mg, 41 %). **HPLC**: t_R = 2.96 min, $\%_X$ = 95 %; **HPLC-MS (APCI)**: calcd for $[C_{52}H_{84}N_{16}O_{23}+H]^+$ m/z= 1301.60, found m/z= 1301.80, calcd for $[C_{52}H_{84}N_{16}O_{23}+2H]^{2+}$ m/z= 651.30, found m/z= 651.85, t_R = 1.69 min.

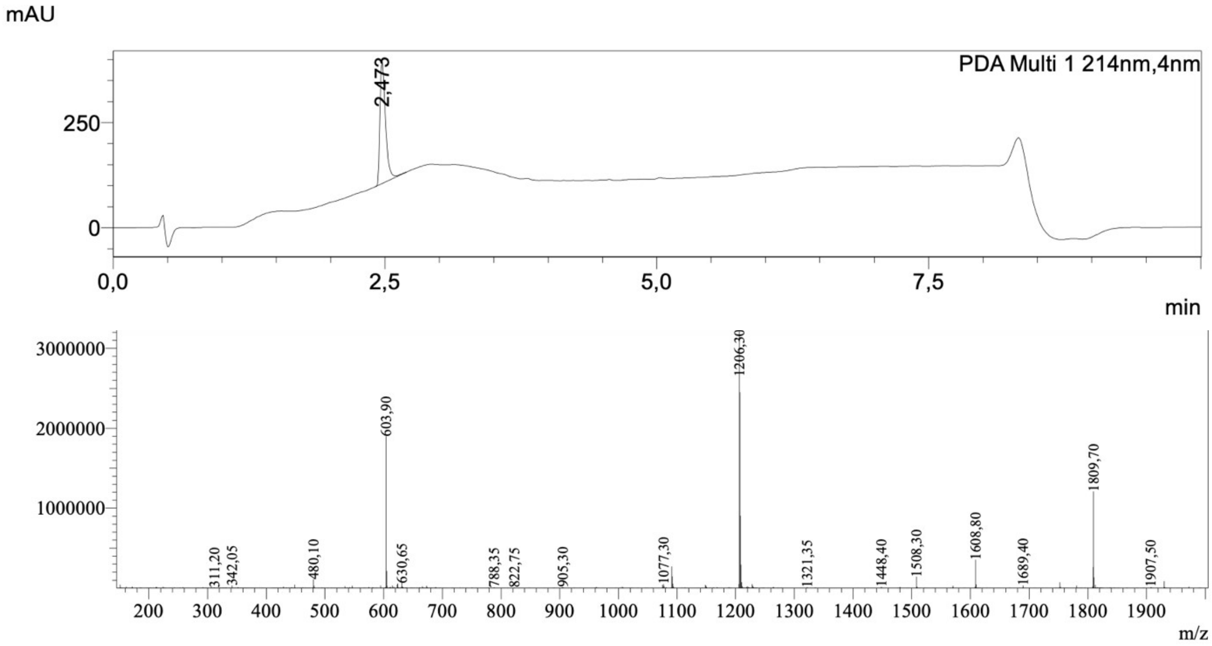
PPDM2211 (-COOH, DMAP1, 235-245, NH_2-Asn^1-Arg^2-Thr^3-Pro^4-Glu^5-Gln^6-Val^7-Ala^8-Glu^9-Glu^{10}-Glu^{11}-COOH). **PPDM2211** was synthetised using general method B and purified *via* method A (17 mg, 15 %). **HPLC**: t_R = 2.99 min, $\%_X$ = 99 %; **HPLC-MS (APCI)**: calcd for $[C_{52}H_{85}N_{17}O_{22}+H]^+$ m/z= 1300.61, found m/z= 1301.25, calcd for $[C_{52}H_{85}N_{17}O_{22}+2H]^{2+}$ m/z= 650.81, found m/z= 651.40, t_R = 1.72 min.

PPDM2215 (-$CONH_2$, DMAP1, 255-265, NH_2-Glu^1-Ala^2-Arg^3-Lys^4-Lys^5-Glu^6-Arg^7-Glu^8-Lys^9-Arg^{10}-Ser^{11}-$CONH_2$). **PPDM2215** was synthetised using general method A and purified *via* method B (43 mg, 26 %). **HPLC**: t_R = 3.48 min, $\%_X$ = 96 %; **HPLC-MS (APCI)**: calcd for $[C_{69}H_{118}N_{24}O_{24}+H]^+$ m/z= 1168.88, found m/z= 1669.50, calcd for $[C_{69}H_{119}N_{24}O_{24}+2H]^{2+}$ m/z= 834.45, found m/z= 834.50, t_R = 1.93 min.

III. Peptide Spectra

Figure SI-19. PPUS2202 *peptide spectra*

LC-MS analysis



HPLC analysis

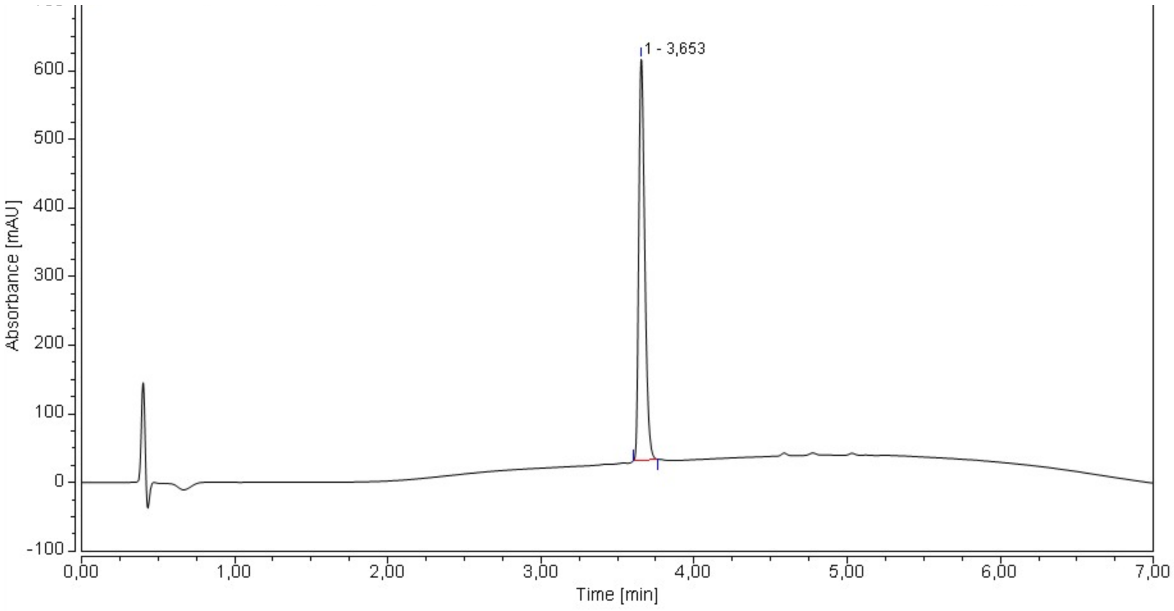
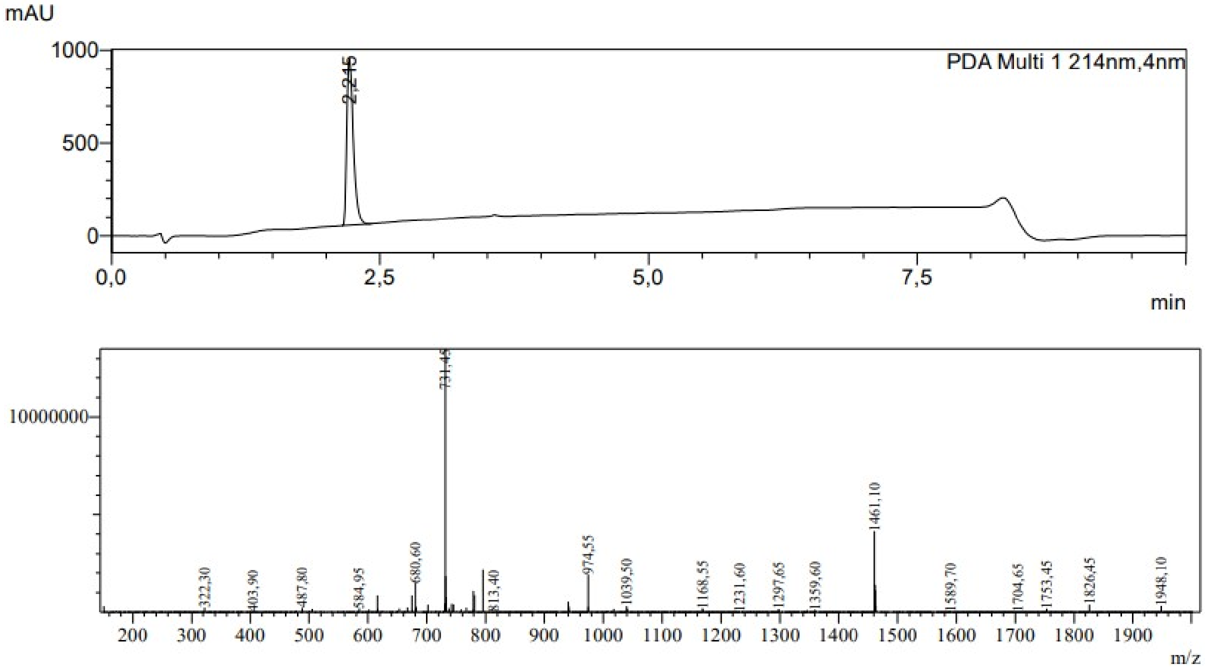


***Figure SI-20. PPUS2212** peptide spectra*



LC-MS analysis

HPLC analysis

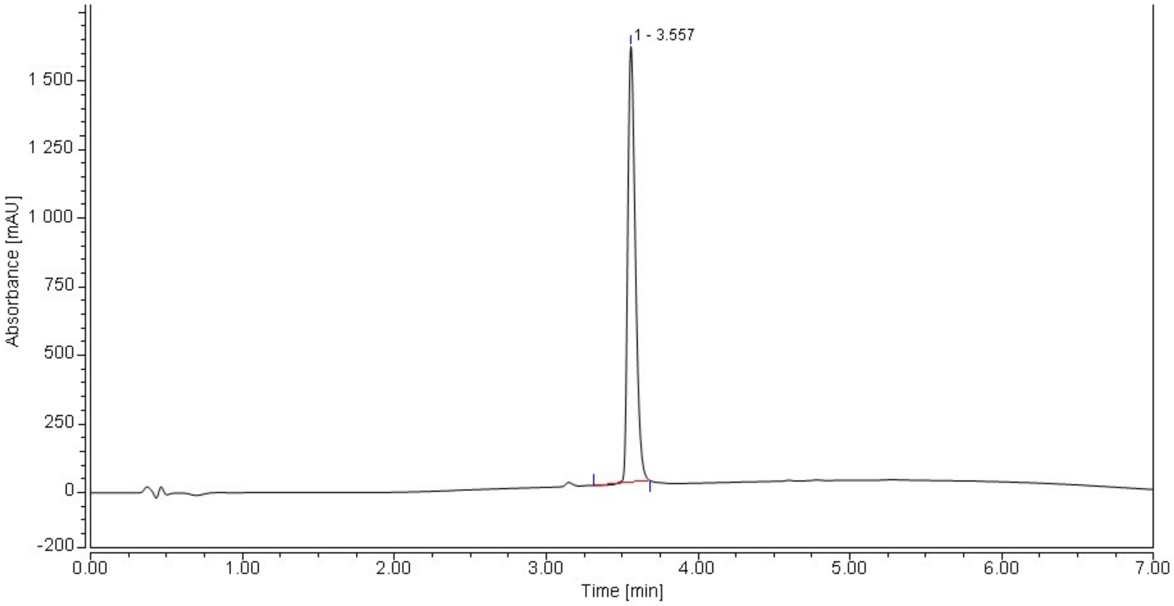
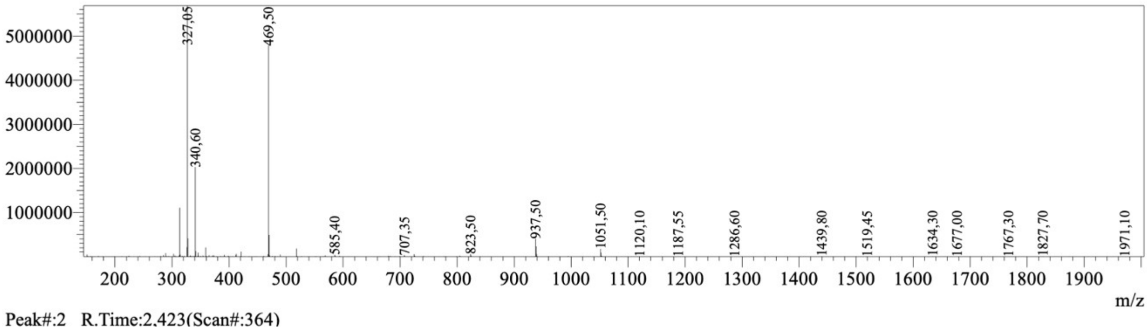


***Figure SI-21. PPUH2206** peptide spectra*

LCMS analysis



HPLC analysis

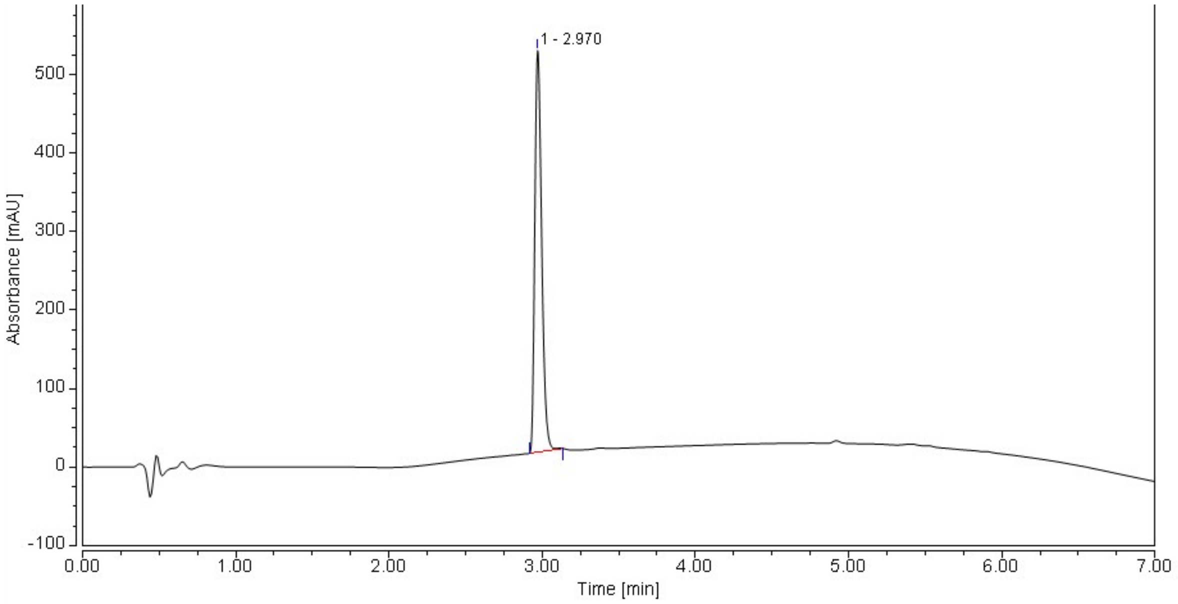
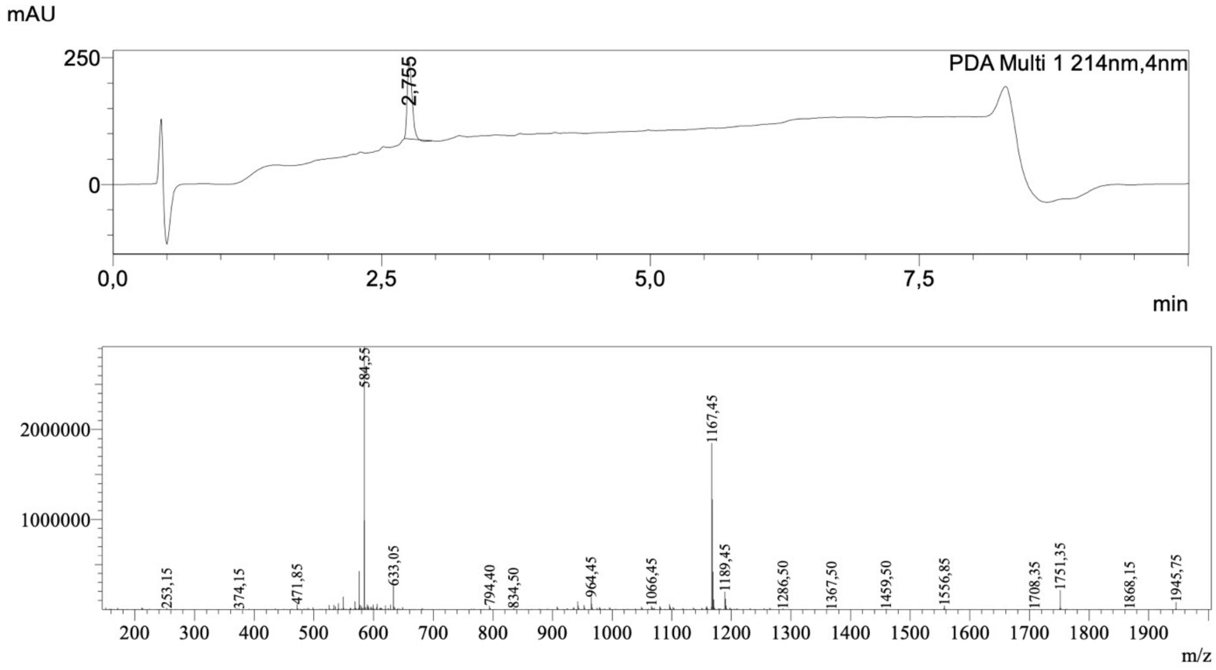


Figure SI-22. PPSP2201 *peptide spectra*

LCMS analysis



HPLC analysis

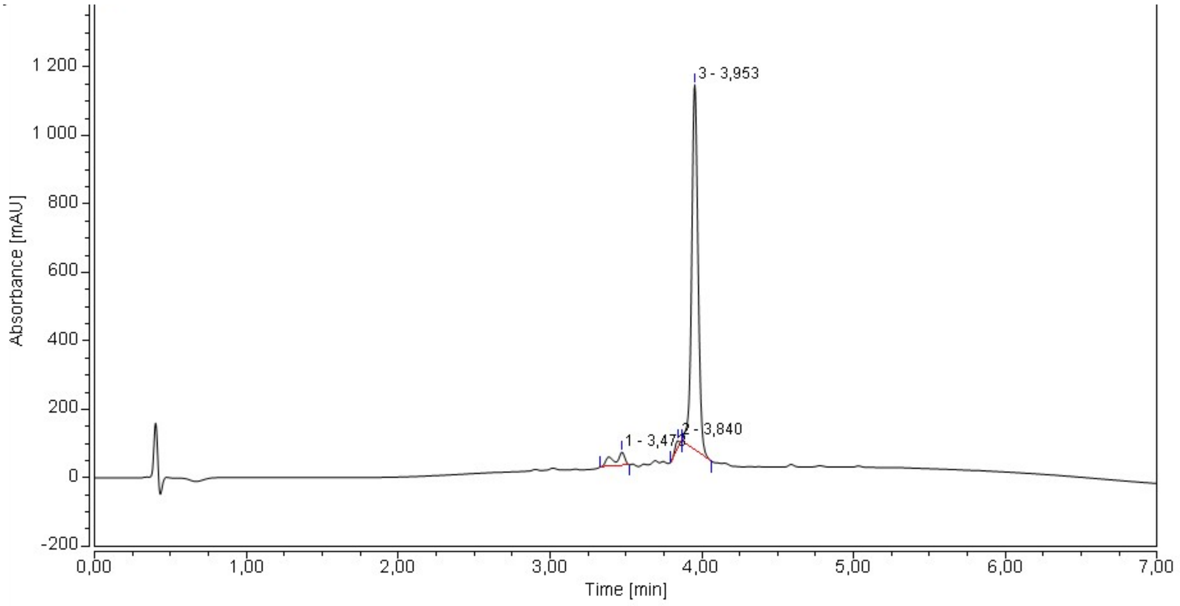
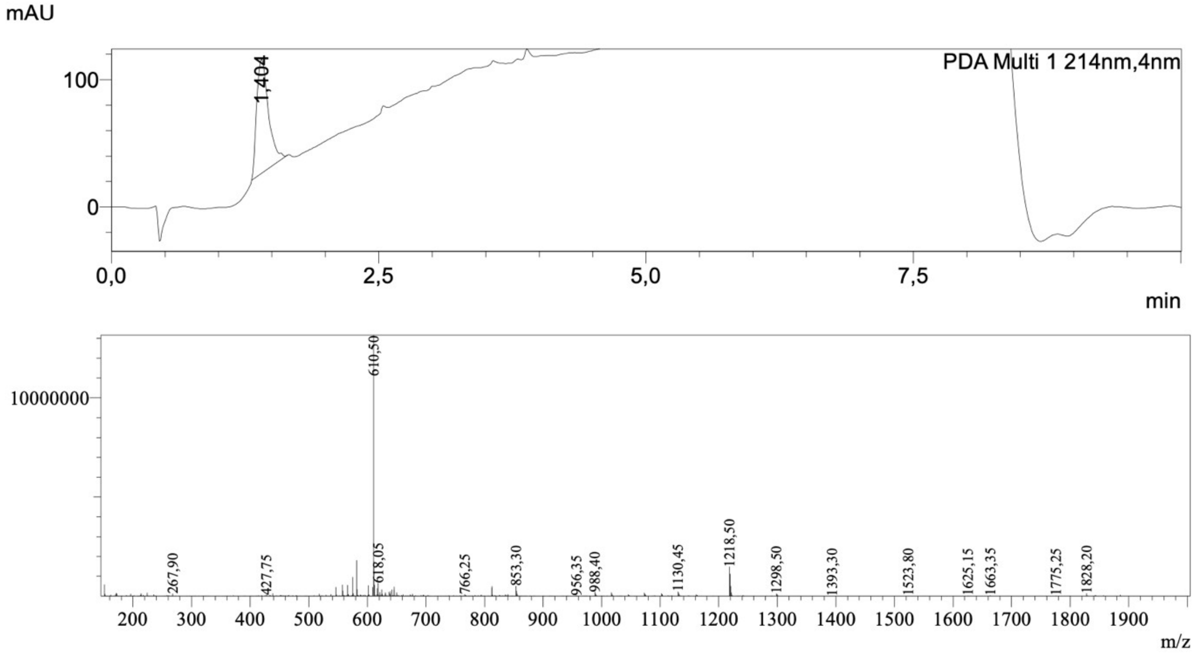


Figure SI-23. PPCF2209 *peptide spectra*



LCMS analysis

HPLC analysis

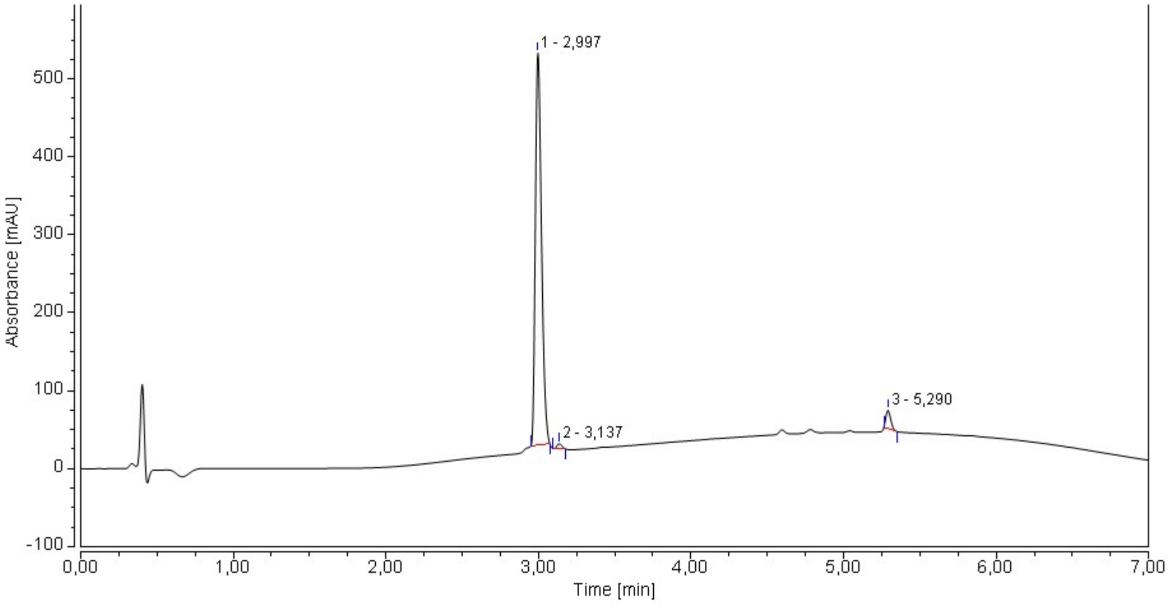
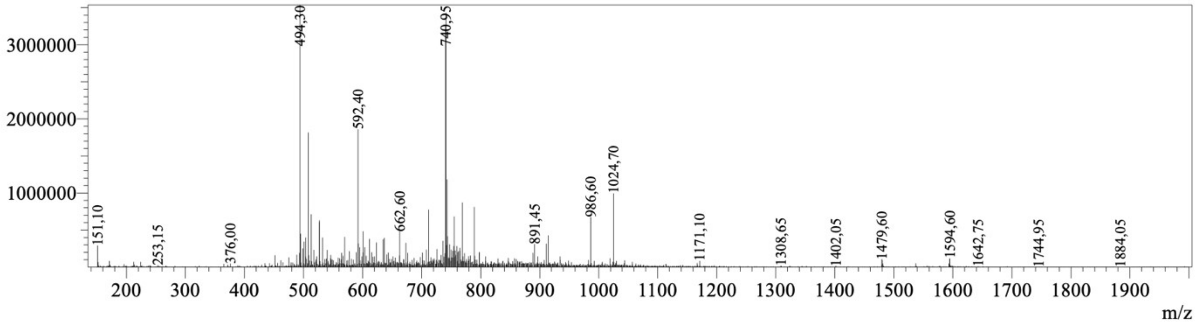


Figure SI-24. PPCF2214 *peptide spectra*

LCMS analysis



HPLC analysis

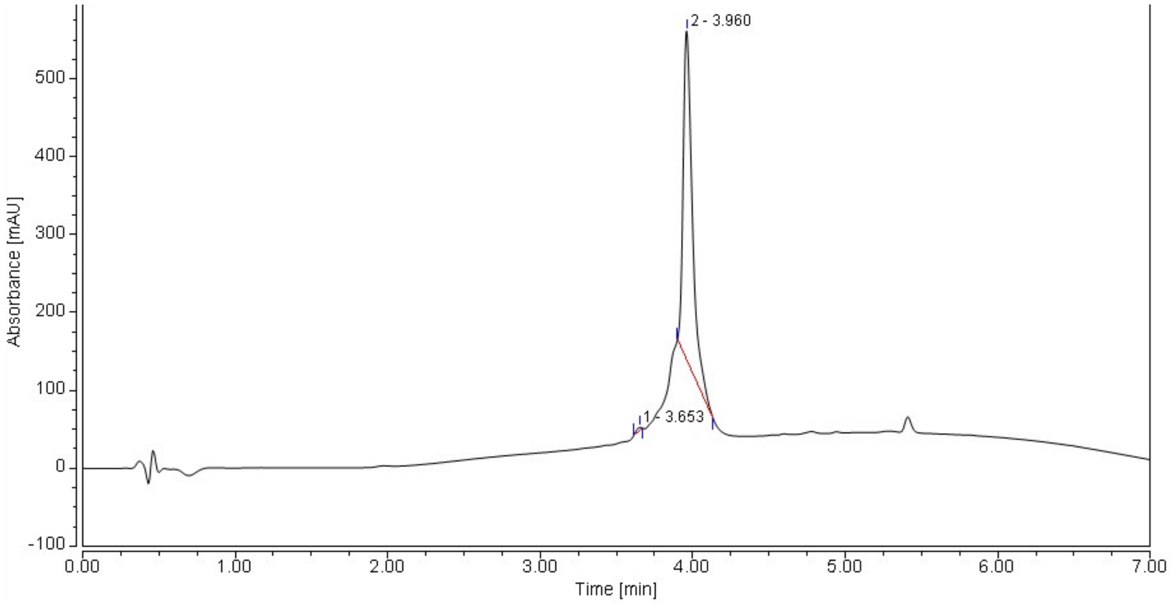
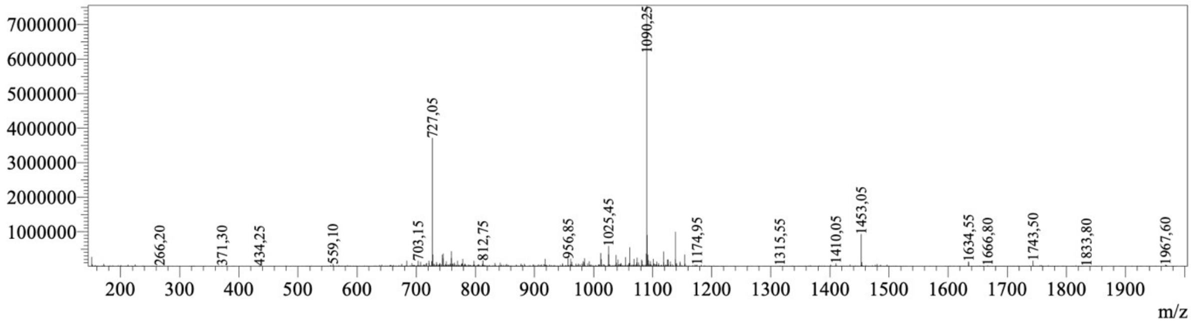


***Figure SI-25**. **PPSt2213** peptide spectra*

LCMS analysis



HPLC analysis

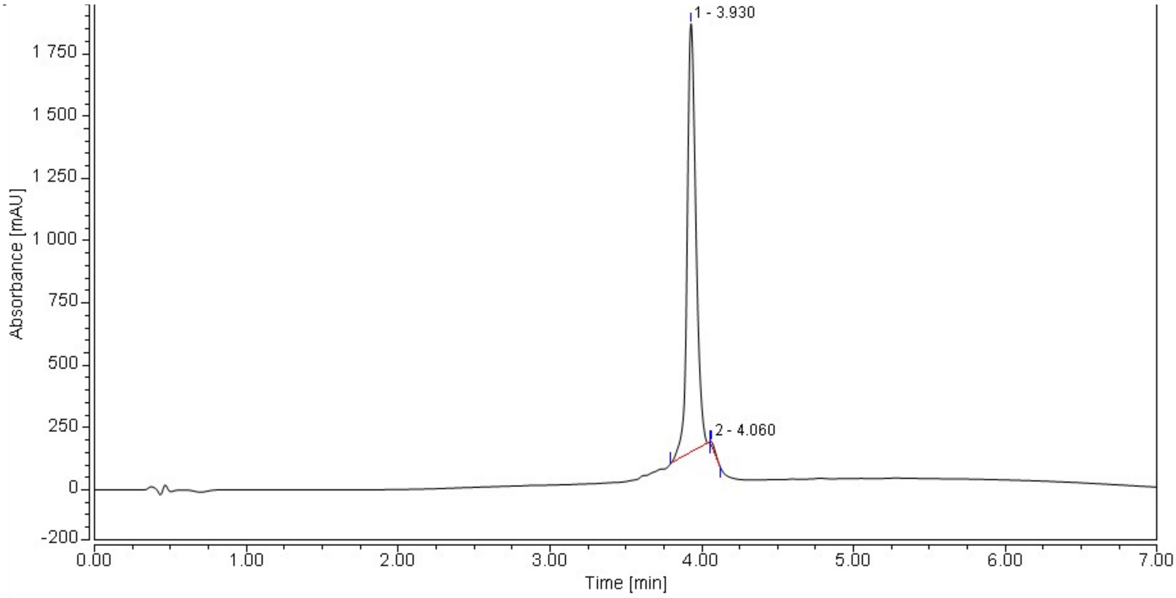
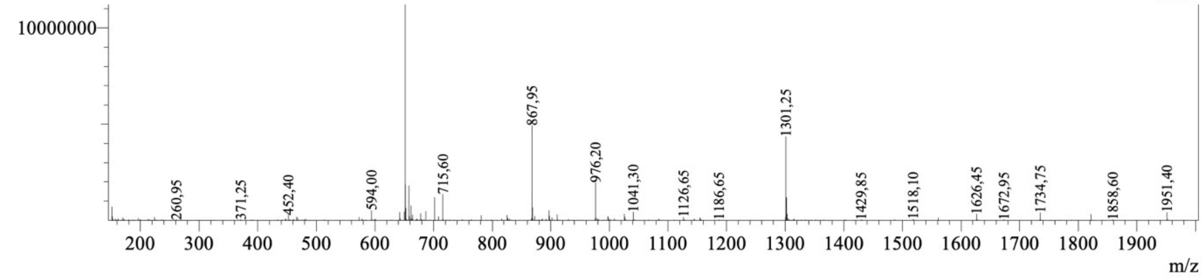


Figure SI-26. PPDM2203 *peptide spectra*

LCMS analysis



HPLC analysis

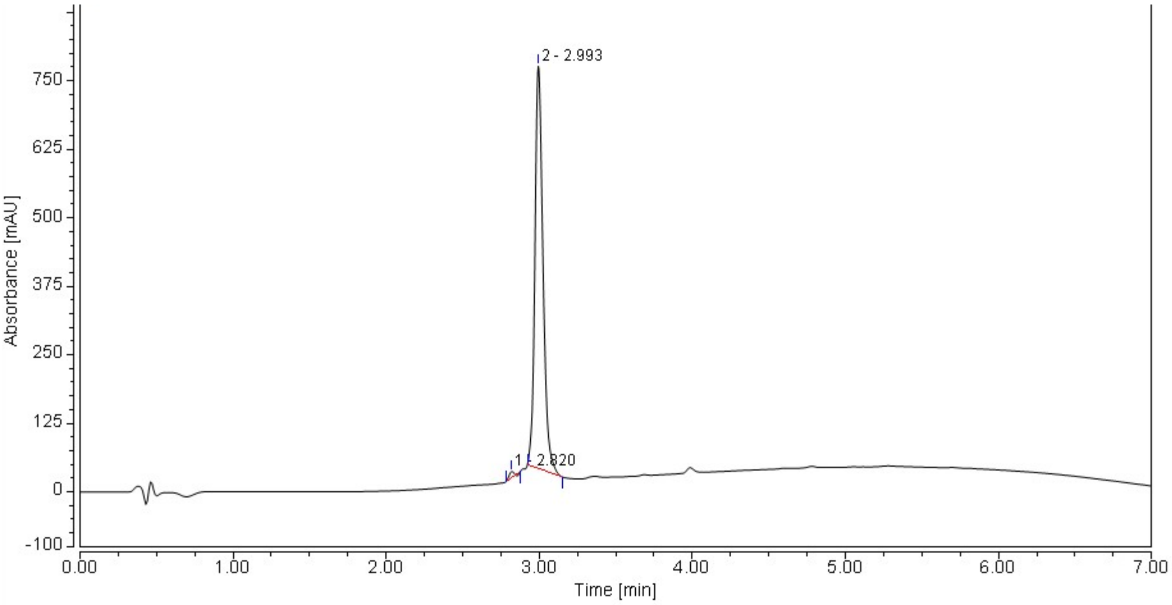
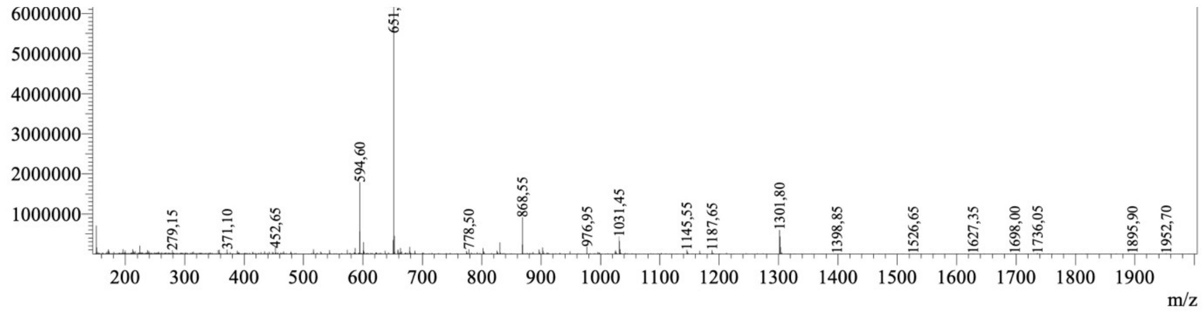


***Figure SI-27. PPDM2211** peptide spectra*

LCMS analysis



HPLC analysis

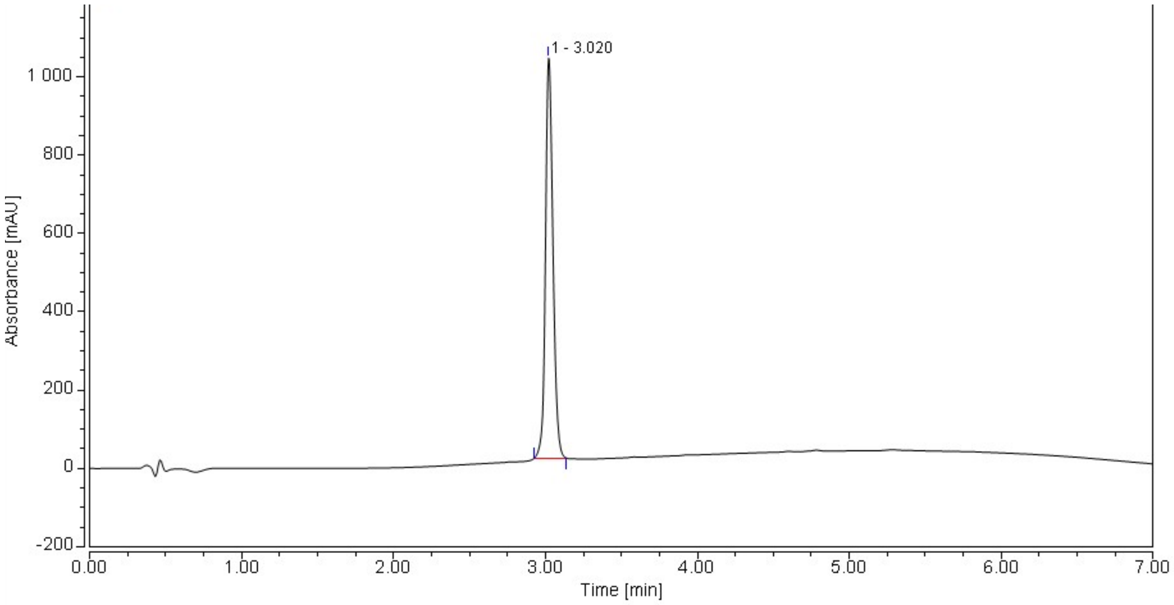
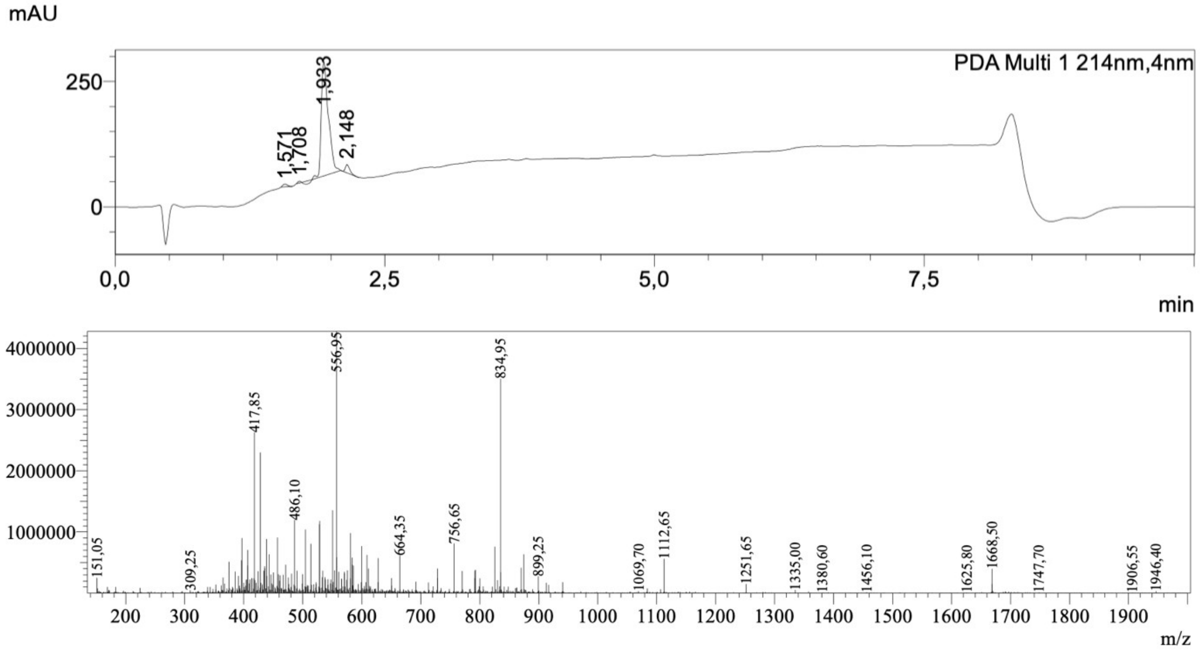
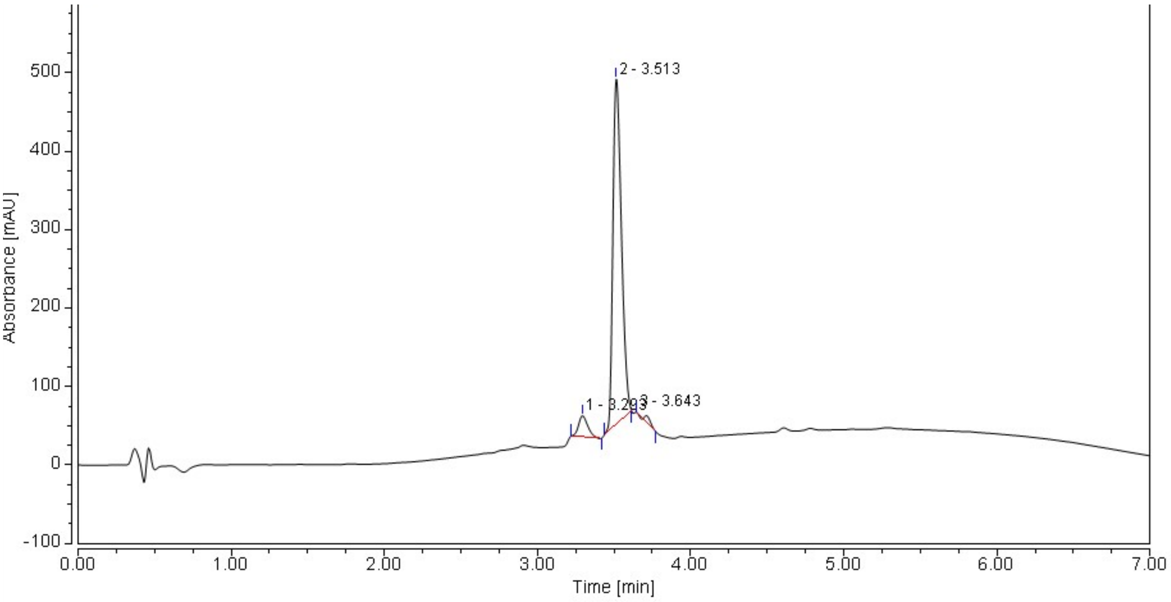


Figure SI-28. PPDM2215 *peptide spectra*

LCMS analysis

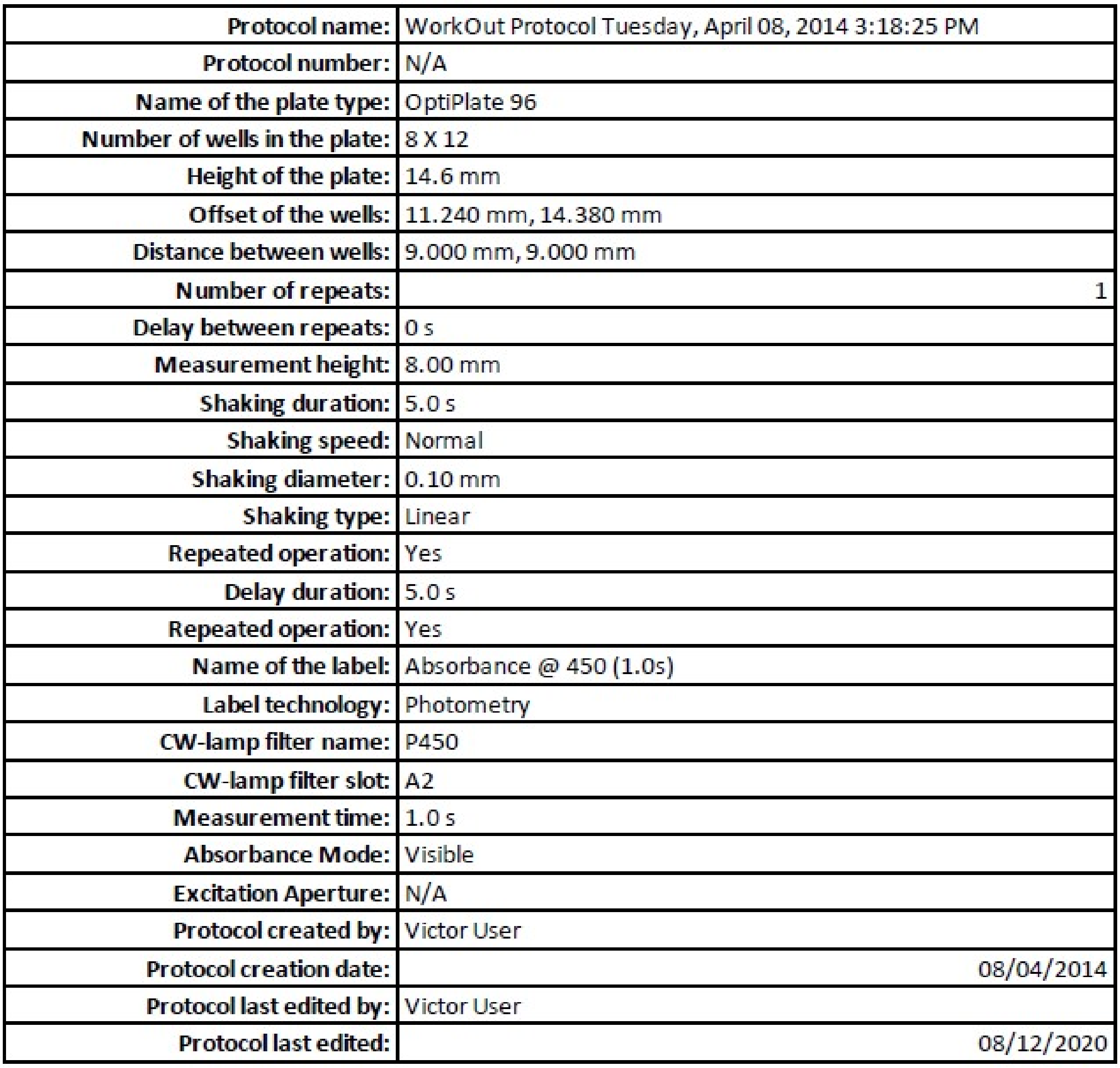


HPLC analysis



IV. Biological Evaluation

ELISA experimental conditions

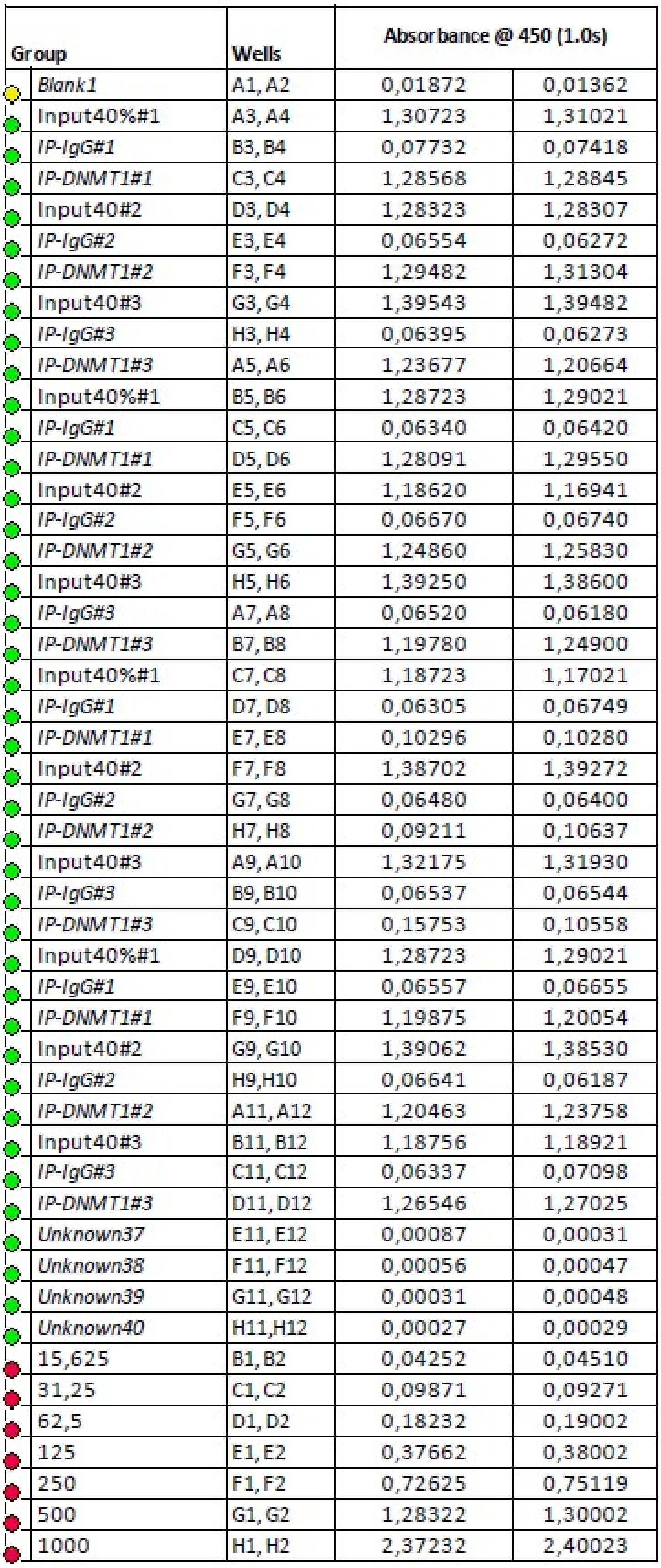


Protocol name:	WorkOut Protocol Tuesday, April 08, 2014 3:18:25 PM
Protocol number:	N/A
Name of the plate type:	OptiPlate 96
Number of wells in the plate:	8 X 12
Height of the plate:	14.6 mm
Offset of the wells:	11.240 mm, 14.380 mm
Distance between wells:	9.000 mm, 9.000 mm
Number of repeats:	1
Delay between repeats:	0 s
Measurement height:	8.00 mm
Shaking duration:	5.0 s
Shaking speed:	Normal
Shaking diameter:	0.10 mm
Shaking type:	Linear
Repeated operation:	Yes
Delay duration:	5.0 s
Repeated operation:	Yes
Name of the label:	Absorbance @ 450 (1.0s)
Label technology:	Photometry
CW-lamp filter name:	P450
CW-lamp filter slot:	A2
Measurement time:	1.0 s
Absorbance Mode:	Visible
Excitation Aperture:	N/A
Protocol created by:	Victor User
Protocol creation date:	08/04/2014
Protocol last edited by:	Victor User
Protocol last edited:	08/12/2020

ELISA quantification results for DNMT1-DMAP1 co-immunoprecipitation

Group			Wells	Absorbance @ 450 (1.0s)	
●		*Blank1*	A1, A2	0,00782	0,00873
●		Input40%#1	A3, A4	0,67272	0,66527
●		*IP-IgG#1*	B3, B4	0,05272	0,05177
●		*IP-DNMT1#1*	C3, C4	0,31254	0,30052
●		Input40#2	D3, D4	0,62828	0,62828
●		*IP-IgG#2*	E3, E4	0,04929	0,05182
●		*IP-DNMT1#2*	F3, F4	0,27623	0,28002
●		Input40#3	G3, G4	0,64522	0,63998
●		*IP-IgG#3*	H3, H4	0,05887	0,04872
●		*IP-DNMT1#3*	A5, A6	0,27652	0,27544
●		Input40%#1	B5, B6	0,62623	0,64522
●		*IP-IgG#1*	C5, C6	0,05877	0,05688
●		*IP-DNMT1#1*	D5, D6	0,36621	0,36882
●		Input40#2	E5, E6	0,66562	0,65982
●		*IP-IgG#2*	F5, F6	0,05762	0,05882
●		*IP-DNMT1#2*	G5, G6	0,40636	0,41998
●		Input40#3	H5, H6	0,61721	0,62012
●		*IP-IgG#3*	A7, A8	0,06002	0,05926
●		*IP-DNMT1#3*	B7, B8	0,40622	0,41002
●		Input40%#1	C7, C8	0,61723	0,62009
●		*IP-IgG#1*	D7, D8	0,05982	0,05872
●		*IP-DNMT1#1*	E7, E8	0,80032	0,80392
●		Input40#2	F7, F8	0,59823	0,59772
●		*IP-IgG#2*	G7, G8	0,06182	0,06202
●		*IP-DNMT1#2*	H7, H8	0,83229	0,83823
●		Input40#3	A9, A10	0,60082	0,60119
●		*IP-IgG#3*	B9, B10	0,05782	0,05972
●		*IP-DNMT1#3*	C9, C10	0,84723	0,84113
●		Input40%#1	D9, D10	0,62732	0,62882
●		*IP-IgG#1*	E9, E10	0,06272	0,06182
●		*IP-DNMT1#1*	F9, F10	0,81923	0,82009
●		Input40#2	G9, G10	0,59837	0,58927
●		*IP-IgG#2*	H9,H10	0,06188	0,06172
●		*IP-DNMT1#2*	A11, A12	0,79082	0,79882
●		Input40#3	B11, B12	0,59122	0,59817
●		*IP-IgG#3*	C11, C12	0,06002	0,05987
●		*IP-DNMT1#3*	D11, D12	0,78223	0,77982
●		*Unknown37*	E11, E12	0,00035	0,00044
●		*Unknown38*	F11, F12	0,00047	0,00018
●		*Unknown39*	G11, G12	0,00034	0,00036
●		*Unknown40*	H11,H12	0,00015	0,00012
●		78,125	B1, B2	0,03877	0,03698
●		156,25	C1, C2	0,08252	0,08352
●		312,5	D1, D2	0,18762	0,17897
●		625	E1, E2	0,37621	0,37821
●		1250	F1, F2	0,76721	0,75751
●		2500	G1, G2	1,32721	1,30922
●		5000	H1, H2	2,52421	2,49982

ELISA quantification results for DNMT1-UHRF1 co-immunoprecipitation



Group	Wells	Absorbance @ 450 (1.0s)	
🟡 *Blank1*	A1, A2	0,01872	0,01362
🟢 Input40%#1	A3, A4	1,30723	1,31021
🟢 *IP-IgG#1*	B3, B4	0,07732	0,07418
🟢 *IP-DNMT1#1*	C3, C4	1,28568	1,28845
🟢 Input40#2	D3, D4	1,28323	1,28307
🟢 *IP-IgG#2*	E3, E4	0,06554	0,06272
🟢 *IP-DNMT1#2*	F3, F4	1,29482	1,31304
🟢 Input40#3	G3, G4	1,39543	1,39482
🟢 *IP-IgG#3*	H3, H4	0,06395	0,06273
🟢 *IP-DNMT1#3*	A5, A6	1,23677	1,20664
🟢 Input40%#1	B5, B6	1,28723	1,29021
🟢 *IP-IgG#1*	C5, C6	0,06340	0,06420
🟢 *IP-DNMT1#1*	D5, D6	1,28091	1,29550
🟢 Input40#2	E5, E6	1,18620	1,16941
🟢 *IP-IgG#2*	F5, F6	0,06670	0,06740
🟢 *IP-DNMT1#2*	G5, G6	1,24860	1,25830
🟢 Input40#3	H5, H6	1,39250	1,38600
🟢 *IP-IgG#3*	A7, A8	0,06520	0,06180
🟢 *IP-DNMT1#3*	B7, B8	1,19780	1,24900
🟢 Input40%#1	C7, C8	1,18723	1,17021
🟢 *IP-IgG#1*	D7, D8	0,06305	0,06749
🟢 *IP-DNMT1#1*	E7, E8	0,10296	0,10280
🟢 Input40#2	F7, F8	1,38702	1,39272
🟢 *IP-IgG#2*	G7, G8	0,06480	0,06400
🟢 *IP-DNMT1#2*	H7, H8	0,09211	0,10637
🟢 Input40#3	A9, A10	1,32175	1,31930
🟢 *IP-IgG#3*	B9, B10	0,06537	0,06544
🟢 *IP-DNMT1#3*	C9, C10	0,15753	0,10558
🟢 Input40%#1	D9, D10	1,28723	1,29021
🟢 *IP-IgG#1*	E9, E10	0,06557	0,06655
🟢 *IP-DNMT1#1*	F9, F10	1,19875	1,20054
🟢 Input40#2	G9, G10	1,39062	1,38530
🟢 *IP-IgG#2*	H9,H10	0,06641	0,06187
🟢 *IP-DNMT1#2*	A11, A12	1,20463	1,23758
🟢 Input40#3	B11, B12	1,18756	1,18921
🟢 *IP-IgG#3*	C11, C12	0,06337	0,07098
🟢 *IP-DNMT1#3*	D11, D12	1,26546	1,27025
🟢 *Unknown37*	E11, E12	0,00087	0,00031
🟢 *Unknown38*	F11, F12	0,00056	0,00047
🟢 *Unknown39*	G11, G12	0,00031	0,00048
🟢 *Unknown40*	H11,H12	0,00027	0,00029
🔴 15,625	B1, B2	0,04252	0,04510
🔴 31,25	C1, C2	0,09871	0,09271
🔴 62,5	D1, D2	0,18232	0,19002
🔴 125	E1, E2	0,37662	0,38002
🔴 250	F1, F2	0,72625	0,75119
🔴 500	G1, G2	1,28322	1,30002
🔴 1000	H1, H2	2,37232	2,40023

ELISA quantification results for DNMT1-DNMT1 co-immunoprecipitation

Group		Wells	Absorbance @ 450 (1.0s)	
🟡	*Blank1*	A1, A2	0,00291	0,03092
🟢	Input40%#1	A3, A4	0,51423	0,52982
🟢	*IP-IgG#1*	B3, B4	0,00403	0,00387
🟢	*IP-DNMT1#1*	C3, C4	0,63872	0,62187
🟢	Input40#2	D3, D4	0,50092	0,48982
🟢	*IP-IgG#2*	E3, E4	0,00365	0,03920
🟢	*IP-DNMT1#2*	F3, F4	0,62982	0,63823
🟢	Input40#3	G3, G4	0,48739	0,49003
🟢	*IP-IgG#3*	H3, H4	0,03282	0,03172
🟢	*IP-DNMT1#3*	A5, A6	0,65422	0,65002
🟢	Input40%#1	B5, B6	0,50032	0,50102
🟢	*IP-IgG#1*	C5, C6	0,00287	0,00382
🟢	*IP-DNMT1#1*	D5, D6	0,62773	0,63098
🟢	Input40#2	E5, E6	0,47883	0,48039
🟢	*IP-IgG#2*	F5, F6	0,00278	0,00276
🟢	*IP-DNMT1#2*	G5, G6	0,64832	0,65172
🟢	Input40#3	H5, H6	0,49382	0,48983
🟢	*IP-IgG#3*	A7, A8	0,00276	0,00300
🟢	*IP-DNMT1#3*	B7, B8	0,62723	0,63098
🟢	Input40%#1	C7, C8	0,51723	0,51082
🟢	*IP-IgG#1*	D7, D8	0,00302	0,00287
🟢	*IP-DNMT1#1*	E7, E8	0,63982	0,64992
🟢	Input40#2	F7, F8	0,48739	0,48372
🟢	*IP-IgG#2*	G7, G8	0,00382	0,00318
🟢	*IP-DNMT1#2*	H7, H8	0,63921	0,63723
🟢	Input40#3	A9, A10	0,49930	0,49291
🟢	*IP-IgG#3*	B9, B10	0,00382	0,00403
🟢	*IP-DNMT1#3*	C9, C10	0,62823	0,63981
🟢	Input40%#1	D9, D10	0,52432	0,51998
🟢	*IP-IgG#1*	E9, E10	0,00327	0,00312
🟢	*IP-DNMT1#1*	F9, F10	0,64029	0,64001
🟢	Input40#2	G9, G10	0,50292	0,50119
🟢	*IP-IgG#2*	H9,H10	0,00287	0,00298
🟢	*IP-DNMT1#2*	A11, A12	0,63118	0,63422
🟢	Input40#3	B11, B12	0,48732	0,48873
🟢	*IP-IgG#3*	C11, C12	0,00302	0,00318
🟢	*IP-DNMT1#3*	D11, D12	0,61262	0,62009
🟢	*Unknown37*	E11, E12	0,00066	0,00061
🟢	*Unknown38*	F11, F12	0,00049	0,00045
🟢	*Unknown39*	G11, G12	0,00058	?00028
🟢	*Unknown40*	H11,H12	0,00046	0,00039
🔴	*1*	B1, B2	0,07862	0,08102
🔴	*2*	C1, C2	0,18762	0,17652
🔴	*4*	D1, D2	0,30723	0,31098
🔴	*10*	E1, E2	0,43423	0,42316
🔴	*20*	F1, F2	0,61512	0,60092
🔴	*40*	G1, G2	0,75353	0,76522
🔴	*Standard7*	H1, H2	0,00052	0,00047