Journal of the Korean Chemical Society
2015, *Vol. 59, No. 2*
Printed in the Republic of Korea
http://dx.doi.org/10.5012/jkcs.2015.59.2.188

단신 (Notes)

Identification of Undifferentiated Embryonic Cell Transcription Factor 1 as a Potential Substrate of Carboxyl-Terminal Domain Small Phosphatases

Jimoo Hong[†], Hackyoung Kim[†], Chanin Park[‡], Minky Son[‡], Keun Woo Lee[‡], and Young Jun Kim[†,*]

[†]*Department of Applied Biochemistry, Konkuk University, 268 Chungwondaero, Chungju 380-701, Korea.*
[*]*E-mail: ykim@kku.ac.kr*
[‡]*Division of Applied Life Science, Systems and Synthetic Agrobiotech Center, Plant Molecular Biology and Biotechnology Research Center, Research Institute of Natural Science, Gyeongsang National University, 501 Jinju-daero, Jinju 660-701, Korea*

(Received January 24, 2015; Accepted February 26, 2015)

Key words: Protein phosphatase, Stem cell regulation, Protein dephosphorylation, Enzyme kinetics, Molecular docking

Protein phosphorylation and dephosphorylation play important roles in intracellular communication.[1] C-terminal domain (CTD) small phosphatase 1 (CTDSP1), previously known as small CTD phosphatase 1 (SCP1), is an enzyme that preferentially dephosphorylates serine residues in the CTD of RNA polymerase II (RNAPII).[2] CTDSP1 catalyzes the dephosphorylation of the fifth phosphorylated serine of $Y_1S_2P_3T_4S_5P_6S_7$ in the RNAPII CTD.[3] CTD small phosphatase 2 (CTDSP2), previously known as small CTD phosphatase 2 (SCP2), and CTD small phosphatase like 1 (CTDSPL1), previously known as small CTD phosphatase 3 (SCP3), have a similar sequence and structure to CTDSP1.[4] CTD small phosphatases (CTDSPs) have been documented as neuronal development regulators by silencing neuronal genes.[5] Recent studies suggest that cell division cycle associated 3 (CdcA3)[6] and receptor-regulated SMADs (R-SMADs)[7,8] can be dephosphorylated by CTDSPs, and that CTDSPs are involved in cell cycle regulation and differentiation.

Undifferentiated embryonic cell transcription factor 1 (Utf1), which is specifically expressed in two pluripotent cell lines (mouse embryonic carcinoma cells and mouse embryonic stem cells), is a key player in embryonic cell development and cell fate determination.[9] Interestingly, Utf1 is a Eutherian-specific pluripotency marker.[10] Utf1 is a chromatin-associated protein with repressor activity and is also involved embryonic stem (ES) cell differentiation.[11] A recent study proposed that Utf1 prevents excessive inhibition of bivalent genes by blocking polycomb repressive complex 2 (PRC2) binding and subsequent silencing via Histone 3 (H3) lysine 27 trimethylation. The same study also proposes that Utf1 fine-tunes bivalent gene expression by tagging newly transcribed mRNAs in the nucleus for cytoplasmic degradation. Therefore, Utf1 acts as an epigenetic and translation-modulating factor, and contributes to regulation of pluripotency.[12] In recent studies, the role of Utf1 in cervical carcinoma and carcinogenesis was defined.[13] Two phosphoproteomics studies of human embryonic stem cells reported that Utf1 could be phosphorylated during their differentiation.[14,15] Utf1 has five phosphorylated serine and threonine residues: S18, T35, S42, S54, and S245. The roles of these phosphorylations might be related to regulation of Utf1-binding to target proteins or nucleic acids. Additional investigation to reveal the biological roles of these phosphorylations is necessary to understand Utf1's role in ES cell differentiation. Furthermore, studies describing how these phosphorylated residues are dephosphorylated are indispensable to ascertain how Utf1 is regulated.

The 3-dimensional structures of CTDSP1 (e.g., PDB ID: 2GHT, 2GHQ, 1TAO),[3,16] CTDSP2 (PDB ID: 2Q5E),[17] and CTDSPL1 (PDB ID: 2HHL)[17] have been solved by X-ray crystallography. The X-ray crystal structures of a dominant-negative form of human CTDSP1 bound to mono- and di-phosphorylated peptides (PDB ID: 2GHQ and 2GHT) encompassing the CTD heptad repeat in RNAPII ($Y_1S_2P_3T_4S_5P_6S_7$) identified the residues in CTDSP1 involved in CTD binding and its preferential dephosphorylation of p.Ser5 of the CTD heptad repeat. Based upon this crystallization study, the PX(pS/T)P sequence was selected as a specific substrate motif for catalysis by CTDSP1. We have used this motif to search protein sequence databases. We selected Utf1 as a candidate because both CTDSP1 and Utf1 localize to the nucleus. In this study, we examined Utf1 as a potential substrate of CTDSP1 by using steady-state kinetics, molecular docking studies, and immunoprecipitation pull-down assays.

Initially, we searched proteins homologous to Utf1 in the National Center for Biotechnology Information (NCBI)

database. We found five sequences of Utf1 proteins in Euarchontoglires (accession numbers in supplementary information) and used these sequences to explore conservation of the phosphorylated serine or threonine residues among Utf1 homologs. We aligned the five sequences using T-Coffee. The multiple sequence alignment showed that two regions of Utf1 are highly conserved in Euarchontoglires (*Fig*. S1). Most of the phosphorylated serine or threonine residues of Utf1 are also conserved among the homologs. The sequences near S18 are similar between *Homo sapiens* Utf1 and *Macaca mulatta* Utf1-like, but show little similarity to *Mus musculus* Utf1, *Rattus norvegicus* Utf1, and *Eptesicus fuscus* Utf1-like. The sequence similarity is observed in the sequences near T35, S42, and S245. Interestingly, the sequences near S54 are quite similar among five Utf1 homologs. Therefore, we hypothesized that the biological function of phosphorylation and dephosphorylation of S54 might be conserved in Euarchontoglires.

We analyzed the protein sequence of human Utf1 using data from previous reports.[9,18,19] Utf1 has 341 amino acids and contains a Myb/SANT-like DNA-binding domain (amino acids 94-134) and a leucine zipper motif (amino acids 279-310) (*Fig*. 1A). The DNA-binding domain controls the interaction of Utf1 with DNA,[9] and leucine-zipper is required for activation of activating transcription factor 2 (ATF2) through its binding to the N-terminal residues of ATF2.[9,18] We checked the Human Protein Reference Database (HPRD) for reports of phosphorylated Utf1 residues, and found that it has five phosphorylated residues (S18, T35, S42, S54, and S245) (*Fig*. 1A). Phosphorylation and dephosphorylation of these sites might be involved in DNA binding and in the activation of ATF2 by human Utf1. These phosphorylation sites could be located within unstructured regions of human Utf1.[18,19] We noted that the five-residue sequence in Utf1 is similar to the phosphorylated RNAPII CTD sequence. They both contain the PX(pS/T)P motif (*Fig*. 1B), although there are different amino acids positions 1, 2, 4, and 7 of the RNAPII CTD heptad repeat. Therefore, we decided to characterize the steady-state kinetics of CTDSP1 against these five phosphopeptides derived from human Utf1.

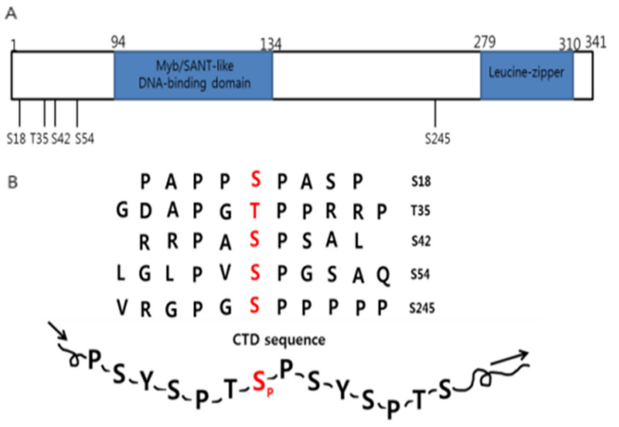

***Figure* 1.** (A) A schematic diagram of human Utf1 with delimited domains. The selected sites of phosphorylation are marked. (B) Comparison of synthesized phosphopeptides selected from human Utf1 with CTD sequence.

The five Utf1-derived phosphopeptides were synthesized by a commercial supplier. They are named S18, T35, S42, S54, S245, and CTD. Their sequences are summarized in *Table* S1. We characterized steady-state kinetics of CTDSP1 against these five synthesized phosphopeptides (*Fig*. S2). Initially, we optimized the reaction conditions for steady-state kinetics. The CTD phosphopeptide (YSPTSPSYSPTpSPS) was used as a positive control. The kinetic values of phosphopeptides are summarized in *Table* 1.

Among five phosphopeptides, the phosphopeptide S54 had the lowest K_m value (0.197 ± 0.010 mM) and highest k_{cat} value (5.539 ± 0.173 s^{-1}), which were comparable to the values (K_m, 0.222 ± 0.021 mM; k_{cat}, 4.128 ± 0.345 s^{-1}) of the CTD phosphopeptide. The phosphopeptide S245 had a similar K_m value (0.228 ± 0.031 mM) to that of CTD phosphopeptide, but a lower k_{cat} value (2.246 ± 0.094 s^{-1}) than that of CTD phosphopeptide. The phosphopeptides S42 and S18 had much higher K_m values (3.675 ± 0.857 mM, 1.058 ± 0.176 mM, respectively) and much lower k_{cat} values (0.898 ± 0.051 s^{-1}, 0.443 ± 0.058 s^{-1}, respectively) than did CTD phosphopeptide. Interestingly, we were unable to measure the activity of CTDSP1 against the phosphopeptide T35 under steady-state

***Table* 1.** Steady-state kinetic characterization of CTDSP1 against Utf1-derived phosphopeptides

	K_m (mM)[a]	k_{cat} (s^{-1})[a]	k_{cat}/K_m ($s^{-1}mM^{-1}$)[a]
S18	1.058 ± 0.176	0.443 ± 0.058	0.420 ± 0.016
T35	ND[b]	ND[b]	ND[b]
S42	3.675 ± 0.857	0.898 ± 0.051	0.254 ± 0.023
S54	0.197 ± 0.010	5.539 ± 0.173	28.113 ± 0.759
S245	0.228 ± 0.031	2.246 ± 0.094	9.946 ± 1.019
CTD	0.222 ± 0.021	4.128 ± 0.345	18.603 ± 0.965

[a]The data represent the average value of three independent measurements ± S.D. [b]ND, not detectable.

conditions, even though it has the PX(pS/T)P motif. The phosphopeptide S54 had the highest k_{cat}/K_m value (28.113 ± 0.759 s^{-1} mM^{-1}). The specificity of CTDSP1 for the S54 phosphopeptide was comparable to that of CTD phosphopeptide (18.603 ± 0.965 $s^{-1}mM^{-1}$) and ~7 times greater than that of the next most reactive phosphopeptide, S245 (9.946 ± 1.019 $s^{-1}mM^{-1}$). These results suggest that CTDSP1 might dephosphorylate at least one of phosphorylated residues of human Utf1, and that the most likely phosphorylated residue is S54.

We examined binding modes to understand different activities between six phosphopeptides from structural aspects through molecular docking protocols. We selected final poses of each phosphopeptide based on the binding conformation of the peptide in the crystal structure (2GHQ).[3] Considering the core interaction and the direction of the phosphopeptide chain (N-terminal to C-terminal), the best poses were chosen from SwissDock for CTD, S18, S54, and S245 and from GOLD for T35 and S42 (*Fig*. 2). The core interactions between all the phosphopeptides and the CTDSP1 binding groove show similar patterns when compared to the previously reported structures,[3] except for T35. The binding of the N- and C-terminal regions of the phosphopeptides indicated the presence of variable conformations. These observations agree with a previous study[3] that reported disorder in the end regions due to high flexibility. The coordination of the Pro_3 of CTD phosphopeptide in a hydrophobic pocket formed by Phe106, Val118, Ile120, Val127, and Leu155 of CTDSP1 is important for phospho-CTD recognition of CTDSP1.[3] We carefully observed the coordination of Pro in the PX(S/T)P sequence of the phosphopeptides S18, T35, S42, S54, and S245 in the hydrophobic pocket of CTDSP1. We found that Pro of S54 was located in the hydrophobic pocket, but the Pro of S18, T35, S42, and S245 was outside the hydrophobic pocket. This observation agrees with our steady-state kinetic data. However, the molecular docking study did not explain the order of specificity constants of S18, T35, S42, and S245. We acknowledge that

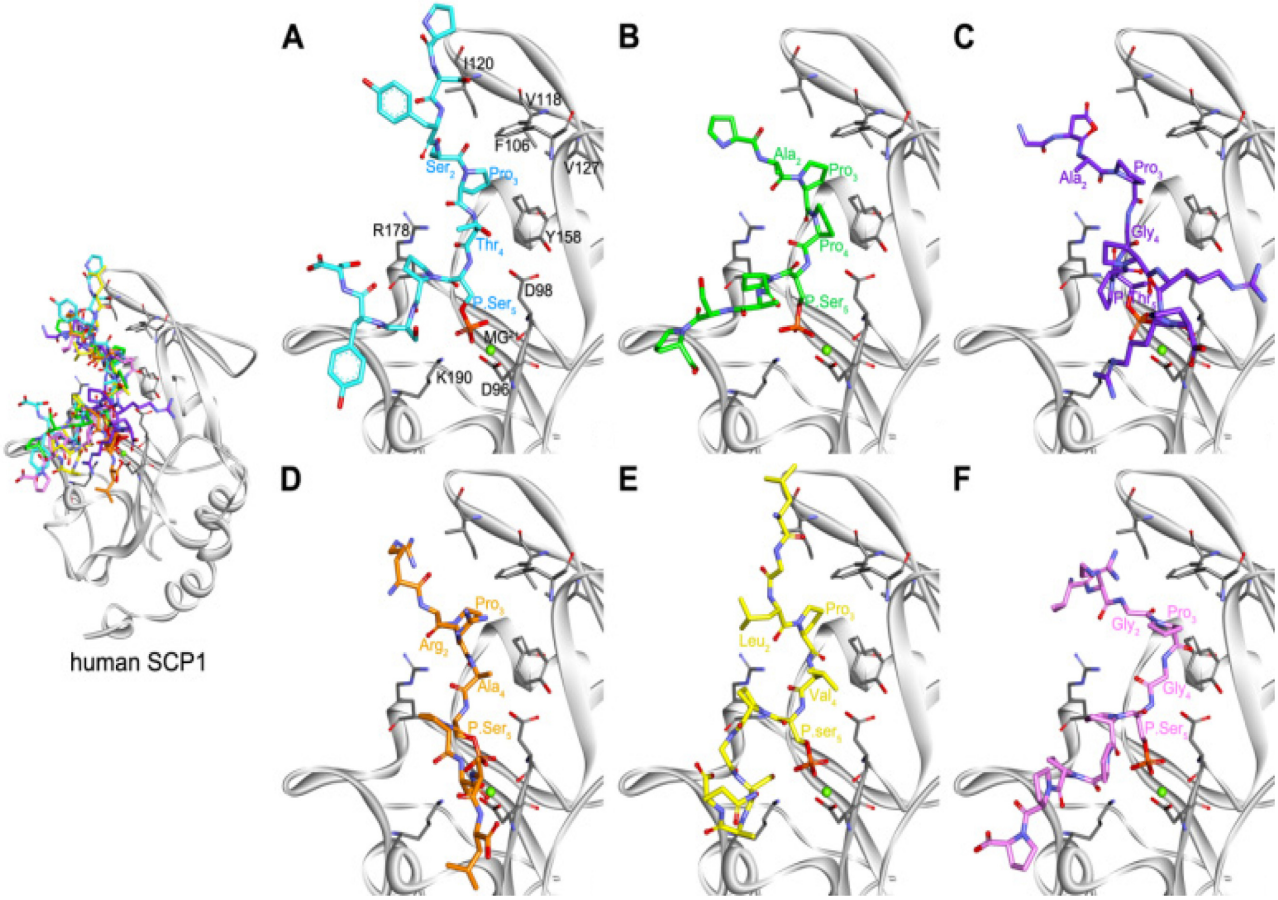

***Figure* 2.** The putative binding conformations of the six phosphopeptides at the active site of CTDSP1 (SCP1). (A) Control (CTD), (B) S18, (C) T35, (D) S42, (E) S54, and (F) S245 are represented as cyan, green, purple, orange, yellow, and pink stick models, respectively. Key residues and Mg^{2+} ion are shown as gray-stick and green-sphere models.

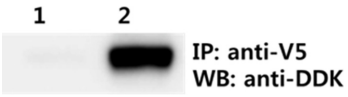

***Figure* 3.** Coimmunoprecipitation with anti-V5 antibody using total lysates of co-transfected HeLa cells with pCDNA-CTDSP1-V5-His and pCMV-Utf1-Myc-DDK. 1. Immunoprecipitate with control HeLa cell, 2. Immunoprecipitate with co-transfected HeLa cell.

structural investigation is necessary to address this. Thus, we are performing a structural investigation to further define the interaction between the Utf1-dervied phosphopeptides and CTDSP1.

We next checked the interaction of human CTDSP1 and Utf1 in cells by immunoprecipitation pull-down assay. After co-transfection of CTDSP1 and Utf1 in HeLa cells, we attempted to detect an interaction. We found that Utf1 was present in the immunoprecipitates of CTDSP1 (*Fig*. 3), but that CTDSP1 was not identified in immunoprecipitates of Utf1. Thus, we are planning to observe cellular interaction of CTDSP1 and Utf1 using active site mutants of CTDSP1, because the active site mutants of a protein phosphatase might be useful for the interaction study. We acknowledge that cell biological observation is necessary to confirm the dephosphorylation of Utf1 by CTDSPs *in vivo*. We are planning to observe the dephosphorylation through several approaches, including a knock-down assay.

In this study, we investigated if Utf1 is a substrate of CTDSPs by sequence alignment, steady-state kinetics, molecular docking, and co-immunoprecipitation. Here, we have documented conservation of the phosphorylated site at S54 among Utf1 proteins in Euarchontoglires, kinetic specificity of CTDSP1 toward the phosphorylated site at S54, and proper coordination of the Pro residue corresponding to Pro_3 in CTD heptad repeat in phosphopeptide of S54. Therefore, one of CTDSPs could dephosphorylate one (S54) of the phosphorylated sites in Utf1 proteins. Taken together, our data suggest that Utf1 is a potential substrate of CTDSPs.

EXPERIMENTAL

Preparation and Purification of Human Recombinant CTDSP1

An *E. coli* plasmid containing the human *CTDSP1* gene spanning residues 76–261 was subcloned into the *E. coli* expression vector pET 21a(+). The pET 21a(+)/ΔN CTDSP1$^{76–261}$/His vector was introduced into the *E. coli* Rosetta 2 (DE3) strain (EMD Bioscience, Darmstadt, Germany). After an OD_{600} value of 0.6 was reached, the *E. coli* culture was transferred to a pre-cooled incubator at 16 °C, and expression of the recombinant protein was induced with 0.4 mM isopropyl-β-D-thiogalactopyranoside (IPTG; Sigma Korea, Seoul, Korea). The human recombinant CTDSP1 was expressed and purified by following a previously described method.[3]

Synthesis of Phosphopeptides

The phosphopeptides generated from human Utf1 based on the phosphorylation position reported in the Human Protein Reference Database (HPRD) were synthesized by FMOC solid phase peptide synthesis with ASP48S and purified by reverse-phase high-performance liquid chromatography (HPLC) on the Vydac Everest C_{18} Column (Peptron Inc., Daejeon, Korea). The sequences of the synthetic phosphopeptides are summarized in *Fig*. 1B.

Malachite Green Assay

CTDSP1-catalyzed dephosphorylation of the phosphorylated substrate was performed as previously described[3] with slight modifications. The assays were conducted at 37 °C in a buffer composed of 50 mM sodium acetate (pH 5.5), 20 mM $MgCl_2$, 5 μM–1 mM phosphopeptides, and 5–50 ng of the wild-type CTDSP1. Phosphate release was quantified by a malachite green-based colorimetric assay for inorganic phosphate by measuring the absorbance at 620 nm. The malachite green solution and inorganic phosphate standards were prepared as previously described.[20] To derive the K_M and k_{cat} values, the data were fitted by a nonlinear regression to the Michaelis-Menten equation by using PRISM software.

Molecular Docking

Molecular docking calculations were performed to predict the binding modes of the six phosphopeptides CTD, S18, T35, S42, S45, and S245. We selected a structure of the holo form containing a Mg^{2+} ion (PDB ID: 3PGL, B chain),[21] which is one of the seven X-ray crystal structures for human CTDSP1 available in the Protein Data Bank (PDB). The structures of the phosphopeptides were drawn and then subjected to energy minimization with the CHARMm force field by using the Discovery Studio 3.5 (Accelrys Software Inc., San Diego, CA). The phosphopeptides were docked into the active site of the CTDSP1 by using GOLD 5.2 software[22] and SwissDock web-server.[23,24] GOLD uses a genetic algorithm (GA) to explore the ligand conformational space in the protein binding site. The residues for docking calculation were selected within a radius of 15 Å from a coordinate that is defined from the center of mass of the co-crystal ligand in the CTDSP1-phosphopeptide complex struc-

ture (PDB ID: 2GHQ).[3] The number of GA runs was set to 150. All other parameters were set as their default values. An additional docking was also performed using the web-based docking server SwissDock, which is based on the docking algorithm EADock DSS.

Immunoprecipitation Pull-Down Assay

A plasmid containing the human *CTDSP1* gene was subcloned into the mammalian expression vector pCDNA-V5-His (Invitrogen, San Diego, CA, USA), and a pCMV-Myc-DDK vector containing the human *Utf1* gene was purchased from Origene (Rockville, MD, USA). The pull-down assay was performed using immunocomplexes that were immunoprecipitated from the total lysates of co-transfected HeLa cells with pCDNA-CTDSP1-V5-His and pCMV-Utf1-Myc-DDK. Co-transfected cells were lysed with a lysis buffer [50 mM Tris-HCl (pH 7.4), 150 mM NaCl, 1 mM ethylenediaminetetraacetic acid (EDTA), and 1% TritonX-100 containing protease and phosphatase inhibitor cocktails (Roche, Manheim, Germany)] or radioimmunoprecipitation assay (RIPA) buffer containing protease and phosphatase inhibitors. The procedures for immunoprecipitation assays were essentially performed as previously described.[25] Primary antibodies used were as follows: mouse anti-V5 (Invitrogen) and mouse anti-DDK (Origene).

Acknowledgments. This work was supported by the Basic Science Research Program (NRF-2013R1A1A2007315 to Y.J.K.) by the Management of Climate Change Program (2010-0029084 to K.L.) through the National Research Foundation of Korea (NRF) funded by the Ministry of Education and by the Next-Generation BioGreen 21 Program (PJ009486 to K.L.) from Rural Development Administration (RDA) in Republic of Korea. This paper was written as part of Konkuk University's research support program for its faculty on sabbatical leave in 2014.

Supporting Information. Figure S1 showing sequence alignment and accession numbers of Utf1 proteins in Euarchontoglires, Figure S2 of kinetic characterization of CTDSP1, and the sequences of synthesized phosphopeptides.

REFERENCES

1. Alonso, A.; Sasin, J.; Bottini, N.; Friedberg, I.; Osterman, A.; Godzik, A.; Hunter, T.; Dixon, J.; Mustelin, T. *Cell* **2004**, *117*, 699.
2. Yeo, M.; Lin, P. S.; Dahmus, M. E.; Gill, G. N. *J. Biol. Chem.* **2003**, *278*, 26078.
3. Zhang, Y.; Kim, Y.; Genoud, N.; Gao, J.; Kelly, J. W.; Pfaff, S. L.; Gill, G. N.; Dixon, J. E.; Noel, J. P. *Mol. Cell* **2006**, *24*, 759.
4. Yeo, M.; Lin, P. S. *Methods Mol. Biol.* **2007**, *365*, 335.
5. Yeo, M.; Lee, S. K.; Lee, B.; Ruiz, E. C.; Pfaff, S. L.; Gill, G. N. *Science* **2005**, *307*, 596.
6. Kim, Y. J.; Bahk, Y. Y. *Biochem. Biophys. Res. Commun.* **2014**, *448*, 189.
7. Wrighton, K. H.; Willis, D.; Long, J.; Liu, F.; Lin, X.; Feng, X. H. *J. Biol. Chem.* **2006**, *281*, 38365.
8. R, H. R.; Kim, H.; Noh, K.; Kim, Y. J. *BMB. Rep.* **2014**, *47*, 192.
9. Okuda, A.; Fukushima, A.; Nishimoto, M.; Orimo, A.; Yamagishi, T.; Nabeshima, Y.; Kuro-o, M.; Nabeshima, Y.; Boon, K.; Keaveney, M.; Stunnenberg, H. G.; Muramatsu, M. *EMBO. J.* **1998**, *17*, 2019.
10. Nishimoto, M.; Katano, M.; Yamagishi, T.; Hishida, T.; Kamon, M.; Suzuki, A.; Hirasaki, M.; Nabeshima, Y.; Katsura, Y.; Satta, Y.; Deakin, J. E.; Graves, J. A.; Kuroki, Y.; Ono, R.; Ishino, F.; Ema, M.; Takahashi, S.; Kato, H.; Okuda, A. *PLoS. One* **2013**, *8*, e68119.
11. Kooistra, S. M.; Thummer, R. P.; Eggen, B. J. *Stem Cell Res.* **2009**, *2*, 211.
12. Gifford, C. A.; Ziller, M. J.; Gu, H.; Trapnell, C.; Donaghey, J.; Tsankov, A.; Shalek, A. K.; Kelley, D. R.; Shishkin, A. A.; Issner, R.; Zhang, X.; Coyne, M.; Fostel, J. L.; Holmes, L.; Meldrim, J.; Guttman, M.; Epstein, C.; Park, H.; Kohlbacher, O.; Rinn, J.; Gnirke, A.; Lander, E. S.; Bernstein, B. E.; Meissner, A. *Cell* **2013**, *153*, 1149.
13. Wu, X. L.; Zheng, P. S. *Carcinogenesis* **2013**, *34*, 1660.
14. Williamson, A. J.; Smith, D. L.; Blinco, D.; Unwin, R. D.; Pearson, S.; Wilson, C.; Miller, C.; Lancashire, L.; Lacaud, G.; Kouskoff, V.; Whetton, A. D. *Mol. Cell. Proteomics* **2008**, *7*, 459.
15. Van Hoof, D.; Munoz, J.; Braam, S. R.; Pinkse, M. W.; Linding, R.; Heck, A. J.; Mummery, C. L.; Krijgsveld, J. *Cell Stem Cell* **2009**, *5*, 214.
16. Kamenski, T.; Heilmeier, S.; Meinhart, A.; Cramer, P. *Mol. Cell* **2004**, *15*, 399.
17. Almo, S. C.; Bonanno, J. B.; Sauder, J. M.; Emtage, S.; Dilorenzo, T. P.; Malashkevich, V.; Wasserman, S. R.; Swaminathan, S.; Eswaramoorthy, S.; Agarwal, R.; Kumaran, D.; Madegowda, M.; Ragumani, S.; Patskovsky, Y.; Alvarado, J.; Ramagopal, U. A.; Faber-Barata, J.; Chance, M. R.; Sali, A.; Fiser, A.; Zhang, Z. Y.; Lawrence, D. S.; Burley, S. K. *J. Struct. Funct. Genomics* **2007**, *8*, 121.
18. Laskowski, A. I.; Knoepfler, P. S. *Biochem. Biophys. Res. Commun.* **2013**, *435*, 551.
19. Laskowski, A. I.; Knoepfler, P. S. *Cell Stem Cell* **2012**, *11*, 732.
20. Taylor, G. S.; Dixon, J. E. *Methods Mol. Biol.* **2004**, *284*, 217.
21. Zhang, M.; Cho, E. J.; Burstein, G.; Siegel, D.; Zhang, Y. *ACS Chemical Biology* **2012**, *6*, 511.
22. Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C.

W.; Taylor, R. D. *Proteins* **2003**, *52*, 609.
23. Grosdidier, A.; Zoete, V.; Michielin, O. *J. Comput. Chem.* **2011**, May 3. doi: 10.1002/jcc.21797.
24. Grosdidier, A.; Zoete, V.; Michielin, O. *Nucleic Acids Res.* **2011**, *39*, 270.
25. Kim, Y.; Gentry, M. S.; Harris, T. E.; Wiley, S. E.; Lawrence, J. C.; Dixon, J. E. *Proc. Natl. Acad. Sci.* **2007**, *104*, 6596.