Supplemental Data

THE STRUCTURE OF NSD1 REVEALS AN AUTOREGULATORY MECHANISM UNDERLYING HISTONE H3K36 METHYLATION

Qi Qiao, Yan Li, Zhi Chen, Mingzhu Wang, Danny Reinberg, and Rui-Ming Xu

Supplemental Experimental Procedures

Molecular dynamics simulation

MD simulations were performed using the GROMACS program package, version 3.2.1 (1) with a GROMOS96 force field (2). The molecular topology of SAM was generated by PRODRG (3). Partial atomic charges of SAM and zinc ions were determined by CHelpG (4) using Gaussian98 (Gaussian Inc.) with a level of HF/6-31G*. For MD simulations, the NSD1-SAM complex structure was solvated using the simple point charge (SPC) water model (5). A Berendsen thermostat was applied using a coupling time of 0.1 ps to maintain the system at a constant temperature of 310 K. The pressure was maintained by coupling to a reference pressure of 1 bar. All bond lengths including hydrogen atoms were constrained by the LINCS algorithm (6). Electrostatic interactions between charge groups at a distance less than 10 Å were calculated explicitly, while long-range electrostatic interactions were calculated using the particle-mesh Ewald (PME) method (7) with a grid width of 1.2 Å and a fourth-order spline interpolation. A cutoff distance of 10 Å was applied for Lennard-Jones interactions. For each system, the simulation cell was a rectangular periodic box, with the minimum distance between the protein and the box walls set to more than 20 Å to avoid direct interactions between molecules in different periodic cells. Na^+ and Cl^- ions were added to neutralize the modeled system, at an ionic strength of ~0.1 M NaCl. The cysteines involved in zinc binding were deprotonated, and the distances between the zinc ions and sulphur atoms of cysteine were restrained between 2.1 to 3.0 Å. An initial round of energy minimization was performed on the whole system but with the protein mainchain and the C_α atoms fixed, and was followed by three equilibration rounds of 20 ps MD simulations: in the first round the solvent molecules and ions were heated to 310 K and the protein and ligands were fixed; the main protein chain and ligands were fixed in the second round; and the whole system was relaxed except for the protein C_α atoms and ligands in the third round. Finally, a 2 ns MD simulation was performed on the whole system.

H3K36 and nucleosome docking

The conformation of the NSD1-SAM complex at 276 ps in the molecular dynamics (MD) simulation was extracted to carry out the docking procedure. The hydrogen atoms and partial charges of the protein were assigned with the UCSF Chimera package (http://www.cgl.ucsf.edu/chimera/) (8). The hydrogen atoms and Gasteiger charges (9) were assigned to each atom of SAM with Chimera. The mono-, di-, tri-methylated and non-methylated lysine molecules were prepared in the same way as SAM, and then docked into the NSD1-SAM complex structure with GOLD (10). To model NSD1-nucleosome interactions, we allowed the first 37 residues of histone H3 to move freely, while the rest of the NCP (PDB code: 1KX5) remained as a rigid body (11). The latter was treated as a receptor in the ZDOCK program encoded in Discovery Studio 2.0 (Accelrys Inc.), and the NSD1 model at 276 ps in the MD simulation was treated as a ligand. As Pro38 of histone H3 is next to Lys36, it was chosen as a receptor binding site residue to filter hits from the search, and the distance cutoff was set at 12.0 Å. An angular step of 15 degrees was used. The top 2000 conformations were clustered, and the docking conformations in which Pro38 was close to the lysine binding channel were selected. The histone H3 tail of the first 37 residues was separately docked onto the NSD1 model, with K36 occupying the lysine binding channel in a manner similar to that from the results of the GOLD docking protocol described above. Finally, the separately docked histone H3 tail and the rigid part of NCP were combined by splicing the two H3 fragments, and the resultant model of the NSD1-NCP complex was subject to a round of energy minimization.

Supplemental References

1. Van Der Spoel D*, et al.* (2005) GROMACS: fast, flexible, and free. *J Comput Chem* 26:1701-1718.
2. van Gunsteren W*, et al.* (1996) Biomolecular simulation: the GROMOS96 manual and user guide. *Zürich, Switzerland: Vdf Hochschulverlag AG an der ETH Zürich*:p. 1-1042.
3. Schuttelkopf AW & van Aalten DM (2004) PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr D Biol Crystallogr* 60:1355-1363.
4. Breneman CM & Wiberg KB (1990) Determining Atom-Centered Monopoles from Molecular Electrostatic Potentials - the Need for High Sampling Density in Formamide Conformational-Analysis. *Journal of Computational Chemistry* 11:361-373.

5. Berendsen HJC, Postma JPM, Gunsteren WFv, & Hermans J (1981) Interaction models for water in relation to protein hydration. In Intermolecular Forces. *Eds. D. Reidel Publishing Company: Dordrecht, The Netherlands*.
6. Hess B, Bekker H, Berendsen HJC, & Fraaije JGEM (1997) LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry* 18:1463-1472.
7. Darden T, York D, & Pedersen L (1993) Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. *Journal of Chemical Physics* 98:10089-10092.
8. Pettersen EF*, et al.* (2004) UCSF chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry* 25:1605-1612.
9. Wang JM, Wang W, Kollman PA, & Case DA (2006) Automatic atom type and bond type perception in molecular mechanical calculations. *Journal of Molecular Graphics & Modelling* 25:247-260.
10. Verdonk ML, Cole JC, Hartshorn MJ, Murray CW, & Taylor RD (2003) Improved protein-ligand docking using GOLD. *Proteins* 52:609-623.
11. Davey CA, Sargent DF, Luger K, Maeder AW, & Richmond TJ (2002) Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 a resolution. *J Mol Biol* 319:1097-1113.