Structure modeling of KLF1 zinc-finger domain

1. Search of pertinent templates

Structural templates have been searched using PSI-BLAST software[1] from the websites of NCBI (www.ncbi.nlm.nih.gov/BLAST/), EBI (http://www.ebi.ac.uk/Tools/psiblast/) and NPSA (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_psiblast.html). Protein structures have been retrieved from the Protein Data Bank (PDB) website (http://www.pdb.org/pdb/home/home.do). Visualization of 3D structures was done with PyMOL (http://www.pymol.org/).

The most interesting template was the series of structures of the zinc-finger domain of Wilms' tumor suppressor protein 1 (WT1) bound to DNA, obtained by X-ray crystallography and NMR spectroscopy (*i.e.* PDB files 2JP9, 2JPA and 2PRT)[2]. The identity between the zinc-finger domain of KLF1 and WT1 is 56% (Figure S1).

```
KLF1    AHTCAHPGCGKSYTKSSHLKAHLRTHTGEKPYACTWEGCGWRFARSDELTRHYRKHTGQR
WT1     PFMCAYPGCNKRYFKLSHLQMHSRKHTGEKPYQCDFKDCERRFSRSDQLKRHQRRHTGVK
        .. **:***.* * * ***: * *.******* * ::.*  **:***:*.** *:*** :

KLF1    PFRCQLCPRAFSRSDHLALHMKRHL
WT1     PFQCKTCQRKFSRSDHLKTHTRTHT
        **:*: * * *******  * : *
```

Figure S1. Alignment of the zinc-finger domain of KLF1 with WT1. The residues implicated into KLF1 ZF2 are in bold while residue 325 is in red.

For comparison, the currently used structural models of KLF1 are based on the structure of the zinc-finger domain of Zif268 bound to DNA (PDB files 1AAY)[3]. The identity between the zinc-finger domain of KLF1 and Zif268 is 43% and their alignment requires a two amino-acid gap in KLF1 ZF2, changing the local topology (Figure S2).

```
KLF1    ---AHTCAHPGCGKSYTKSSHLKAHLRTHTGEKPYACTWEGCGWRFARSDELTRHYRKHTGQR
Zif268  MERPYACPVESCDRRFSRSDELTRHIRIHTGQKPFQCR--ICMRNFSRSDHLTTHIRTHTGEK
           .::*.  .*.: :::*..*. *:* ***:**: *    *  .*:***.** * *.***::

KLF1    PFFRCQLCPRAFSRSDHLALHMKRHL----
Zif268  PFFACDICGRKFARSDERKRHTKIHLRQKD
        *** *::* * *:***.   * * **
```

Figure S2. Alignment of the zinc-finger domain of KLF1 with Zif268. The residues implicated into KLF1 ZF2 are in bold while residue 325 is in red.

2. Analysis of structural templates

The structural quality of 2PRT (WT1 zinc-finger domain) had been assessed by ProCheck[4] and MolProbity[5] softwares. No unusual local conformations or missing atoms have been observed. An important local constraint is the succession of three C_2H_2-type zinc fingers (ZF1, ZF2 and ZF3). The distances between the cysteine and histidine residues contacting Zn^{2+} are well conserved and were used to select the most pertinent structural model (Figure S3). Noteworthy, the ZF2 region is the best resolved region.

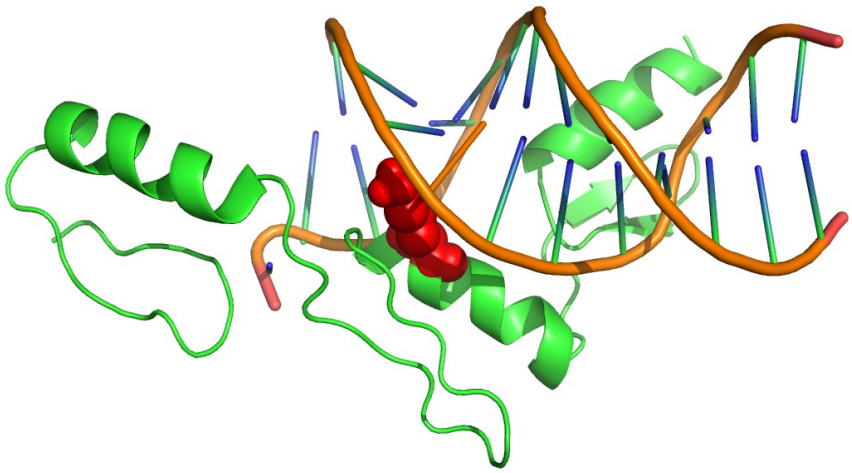

Figure S3. Structure of 2PRT (WT1 zinc-finger domain) shown using PyMOL software. Glutamine 329 (corresponding to glutamate 325 in KLF1) is represented in red spheres; it is part of a helix and is oriented toward the DNA backbone.

3. Comparative modeling and assessment of the models

Five hundreds independent structural models have been generated. The average root mean square deviation (RMSd) between the structural models equals to 0.7 Å with the template (computation done with ProFit software, http://www.bioinf.org.uk/software/profit/). Figure S4 shows the distributions of the different structural models according to their objective and discrete optimized protein energy (DOPE) function scores[6]. Some models have slightly higher objective and DOPE function scores than the selected model, but the latter has the lowest RMSd with the template. Furthermore, the models with the lowest RMSd are very similar, and only show slight structural orientation changes at the protein ends.

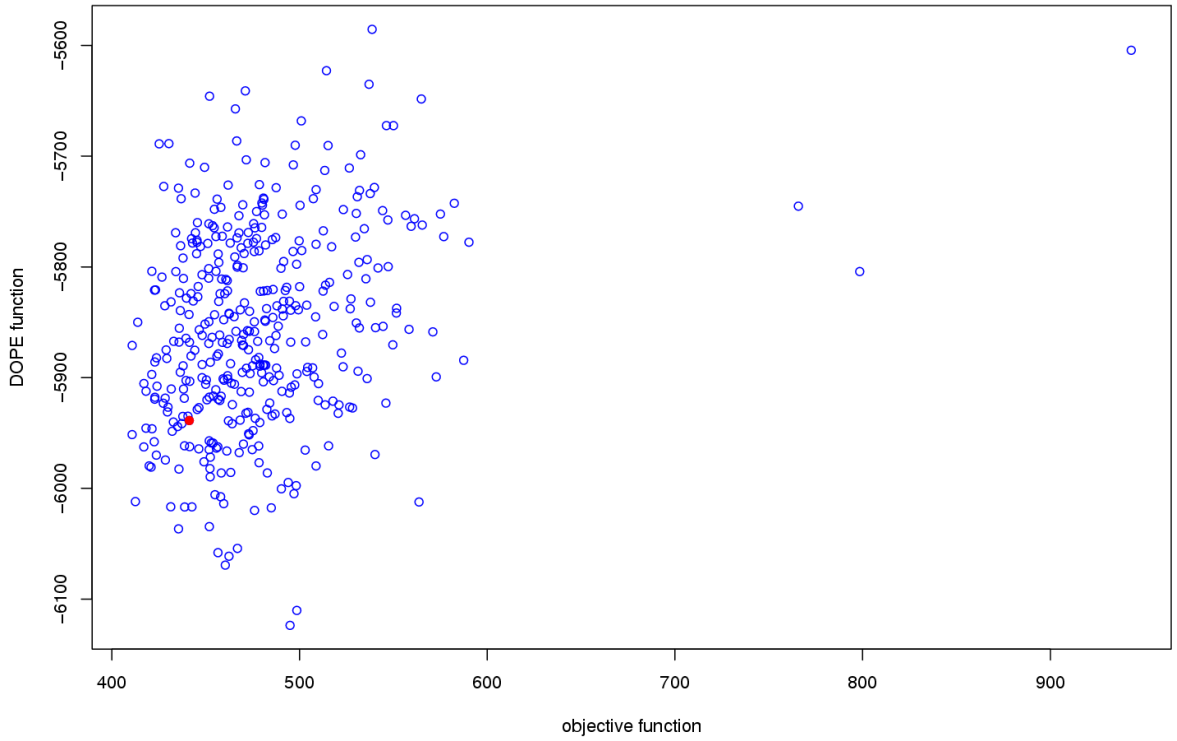

Figure S4. Analysis of structural models according to their objective (x-axis) and their DOPE (y-axis) function scores. The selected model is indicated in red.

The side-chain position of the structural model was then refined using SCWRL4 software[7]. Assessment of the quality of the structural model has been done with ProCheck[4], MolProbity[5] and Verify3D[8,9] softwares. In terms of geometry and atomic distances, the refined model is slightly better than the original template, underlying the quality of the approach. Moreover, the important binding region is considered as well defined.

4. Analysis of structural models of KLF1 wild type and mutant E325K

Glutamate 325 (E325) in KLF1 has a direct orientation toward the DNA (Figure S5a) as glutamine 369 in the template (Figure S3). The point mutation E325K has been introduced by using SCWRL4 software[7]. The single amino acid change does not impact the rest of the protein, but mainly the distance between the side-chain extremity and the DNA backbone (Figure S5); the latter is significantly reduced in the mutant E325K (Figure S5b).

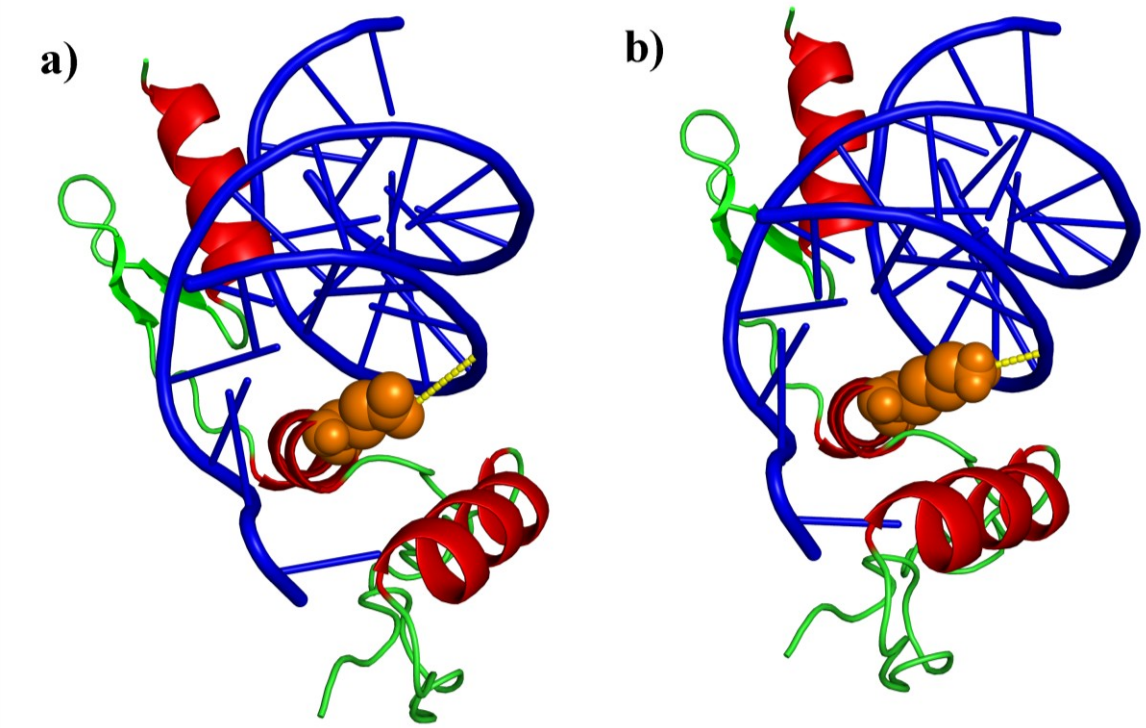

Figure S5. Representation of KLF1 wild type (a) and mutant E325K (b) using PyMOL software. The residue 325, a glutamate in KLF1 wild type and a lysine in mutant E325K, is highlighted with orange spheres. The yellow blocks symbolize the distance between the side-chain extremity of residue 325 and the DNA backbone.

Electrostatic energy has been computed using PyMOL software (Figure S6). In the wild type protein, the local charge of residue 325 is negative (Figure S6a) whereas it is positive in the mutant E325K (Figure S6b).

Hence, E325K mutation (i) extends the side-chain length of residue 325 and (ii) changes the charge of the residue from negative to positive one. The local charge is then very positive while the DNA backbone is negatively charged, so that the electrostatic interaction between KLF1 and the DNA is enhanced. Actually, a hydrogen bond is likely present between lysine 325 and the DNA backbone phosphate.

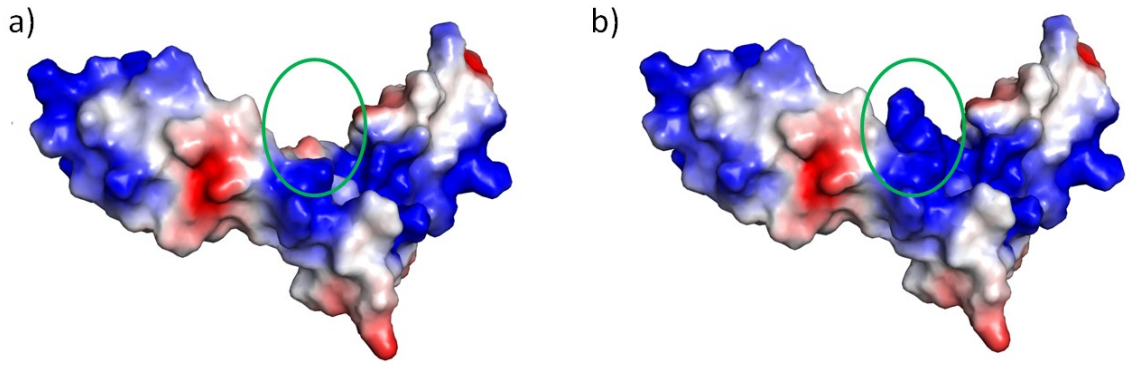

Figure S6. Electrostatics of structural models of KLF1 wild type (a) and mutant E325K (b) using PyMOL software. Blue and red colors indicate positive and negative charges, respectively. The residue 325, a glutamate in KLF1 wild type and a lysine in mutant E325K, is circled in green.

The structural models of KLF1 wild type and mutant E325K can be found at http://www.dsimb.inserm.fr/~debrevern/KLF1

References

1. Altschul, S.F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389-402 (1997).
2. Stoll, R. et al. Structure of the Wilms tumor suppressor protein zinc finger domain bound to DNA. *J Mol Biol* **372**, 1227-45 (2007).
3. Elrod-Erickson, M., Rould, M.A., Nekludova, L. & Pabo, C.O. Zif268 protein-DNA complex refined at 1.6 A: a model system for understanding zinc finger-DNA interactions. *Structure* **4**, 1171-80 (1996).
4. Laskowski, R.A., W., M.M., Moss, D.S. & Thornton, J.M. PROCHECK - a program to check the stereochemical quality of protein structures. *J App Cryst* **26**, 283-291 (1993).
5. Chen, V.B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* (in press, PMID: 20057044).
6. Eswar, N. et al. Comparative protein structure modeling using MODELLER. *Curr Protoc Protein Sci* **Chapter 2**, Unit 2 9 (2007).
7. Wang, Q., Canutescu, A.A. & Dunbrack, R.L., Jr. SCWRL and MolIDE: computer programs for side-chain conformation prediction and homology modeling. *Nat Protoc* **3**, 1832-47 (2008).
8. Bowie, J.U., Luthy, R. & Eisenberg, D. A method to identify protein sequences that fold into a known three-dimensional structure. *Science* **253**, 164-70 (1991).
9. Luthy, R., Bowie, J.U. & Eisenberg, D. Assessment of protein models with three-dimensional profiles. *Nature* **356**, 83-5 (1992).