

Theoretical Prediction System of Stabilization Complex Conformation of DNA Binding Protein and the Binding Affinity for Each DNA Sequence

Jun Koseki (✉ jkoseki@med.nagoya-u.ac.jp)

Nagoya University

Haruka Hirose

Nagoya University

Masamitsu Konno

Osaka University

Ayumu Asai

Osaka University

Takahito Ohshiro

Osaka University

Masateru Taniguchi

Osaka University

Teppei Shimamura

Nagoya University

Research Article

Keywords: DNA binding protein, P53, binding affinity, Thermodynamical effect, molecular mechanics, molecular dynamics

Posted Date: December 28th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-133722/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Theoretical methods such as molecular mechanics and molecular dynamics are very useful in understanding the fundamentals of life sciences. We have established a procedure to predict the structure of a complex of p53 with an arbitrary DNA complex, and have made it possible to predict the binding energy of p53 to DNA from the configuration of molecular dynamics simulations of the complex structure. We performed validation studies using FTD, an anti-tumor drug that exerts its pharmacological effect by incorporation into DNA, and confirmed that the binding affinity of the p53-binding BAX sequence to the p53 tetramer was increased by FTD incorporation. This indicates that FTD tends to interact strongly with DNA-binding proteins when incorporated into DNA. On the other hand, the binding affinity of FTD-containing DNA was greatly reduced in p53-binding sequences extracted from FTD-resistant cells compared to that of normal DNA. In other words, it was suggested that randomly substituted thymidine for FTDs in resistant cells may have acquired resistance by entering positions that prevented binding to DNA-binding proteins. We believe that this procedure is a versatile procedure that can take into account the binding energies of not only the p53 protein, but also other DNA-binding proteins and other biomolecules, thus increasing the importance of computational science in the life sciences.

Introduction

The p53 protein has been reported to role of an important tumor suppressor. It acts primarily as a transcription factor, binding to specific DNA base sequences in response to stresses such as DNA damage and activating genes that promote apoptosis and permanent cell cycle arrest. The various functions of p53 such as DNA binding depend on the multimeric structure of p53 which forms a tetramer^{1,2}. As mutations in the DNA-binding domain of the p53 protein have been identified in many cancers, many studies related to their mutations were reported^{1,3-6}. Recently, it has become more performable to investigate the structural different of wild type and some p53 mutants in terms of theoretical views⁶. It has been widely known that these mutations result in the binding strength changes of the DNA-binding domain in response to the conformational changes in each mutant. In other words, understanding “how the p53 protein tetramer recognizes DNA sequences” and “how strongly it binds to them” is very important to understand the relationship between binding sequences and the strength of the functional responses. Therefore, many studies have also reported focused on how the p53 molecule achieves 'rapid' and 'accurate' target discovery in living cells⁷⁻¹⁰. However, recent study has suggested that the difference between the binding energy obtained for a binding motif and the binding energy obtained for other sequences could be very small¹⁰. In other word, it would be very difficult to experimentally observe such the small energy difference and predict the functional responses of DNA-binding proteins. Therefore, theoretical approaches might be useful. A number of methods have been developed and used to predict how low-molecular compounds would bind to a target protein, such as docking simulations. However, to our knowledge, there are no previous methods that could predict the binding mode of DNA-binding proteins and discuss their binding energies with sufficient accuracy. Thus, in this study, we propose a procedure for predicting the binding energies of arbitrary DNA sequence to p53

protein, taking into account the thermal molecular behavior of them in biological environment to understand p53 response with theoretical methods. Firstly, in our procedure, the p53 tetramer and target DNAs' complex structures are predicted based on conformational homology of complex structural data registered in Protein data bank (PDB). Subsequently, molecular dynamics (MD) simulations are performed using the initial structures of their biological molecules to sample the thermodynamic behaviors in biological environment. After that, the binding energies are estimated with their sampling conformations. Additionally, to validate the proposed procedure, we have performed the case calculations using antitumor drug, trifluorothymidine (FTD), which is known to have a highly effective antitumor potency by incorporating with DNA.

Results And Discussion

Skeleton brief of our procedure for predicting the binding poses and energies

To understanding the functional response of DNA-binding proteins, the procedure for our predicting the binding mode of a p53 protein with an arbitrary DNA sequences and estimating the binding energy in a molecular vibratory environment in vivo is developed. As shown in Fig. 1, our procedure is divided into two stages: (1) binding mode prediction and (2) energy calculation, which are described in detail below.

(1) Prediction the complex of p53 protein and each DNA sequence.

In the case of docking simulations of DNA for p53 protein, the number of conformations that must be performed is too enormous, in addition, it might be extrapolated that the value of docking score as an index could not be different based on the experimental facts reported by Y. Itoh¹⁰. Therefore, in order to predict how DNA-binding proteins recognize and bind to their target DNA, we imported the concept of not a docking simulation but a conformational homology (Fig. 1A). That is, using the previously reported crystal structures, we predict binding configurations by considering the positional relationships between the residue loci at the DNA-binding interface of the P53 protein and the binding motifs (5'- C(A/T)(T/A)G - 3'). In concrete terms, at first, from the PDB database, we extracted the structures where binding between the p53 protein and DNA sequence is observed, and superposed the conformations of the amino acids in P53 protein on the binding interface so that the root mean square deviation (RMSD) is smaller. Similarly, the binding motifs on the DNA structure parts were superposed as well. As results, we were able to shed light on the average binding between the p53 protein and the binding motif part. Based on the binding position of recognition motif in above average binding structure, the initial binding conformation is created by replacing it with the DNA of the target sequence. The binding structure predict here does not have to be a precise predicting structure. This is because the energy minimization will be performed with molecular mechanics (MM) for this predicted structure and the after thermodynamical behavior in biological environment will be sampled later with molecular dynamics simulations.

(2) Estimation the DNA binding energy to p53 protein.

As shown in Fig. 1B, in order to sample the molecular behavior of the created p53-DNA complex in biological environment (water phase and 310 K: $\sim 37^\circ\text{C}$), water molecules and counter ions (Na^+) are placed around the predicted structure to mimic the biological environment. The purpose of placing counter ions is to neutralize the total charge across the MD space to be considered. After then, energy minimization of this complex structure is performed. It makes sampling stably the molecular behavior possible. Using the stabilized structure as the initial structure, molecular dynamics simulations are performed for thermodynamic sampling, which is carried out after an elevated temperature process from 0 K to 310 K and equilibration process for a sufficient amount of time. Using this molecular sampling, we calculate the binding energy, which is estimate from the energy difference between the complex formation state and the dissociation state. This makes it possible to obtain the binding energy as a distribution in this calculation process.

Validation Study using antitumor drug incorporating to DNA.

(1) Antitumor drug and test sequences used for verifications

To validate our process, we selected the curious compound, FTD, which possesses a highly effective antitumor potency¹¹⁻¹³. The structural difference between FTD and normal thymidine is only a substituted group, where is methyl group or trifluoro methyl group (Fig. 2A). The combination of FTD and tipiracil hydrochloride has already been approved as a cancer treatment by various institutions including the FDA (US Food and Drug Administration)¹⁴. FTD has been reported to induce p53-dependent sustained arrest during G2 phase¹⁵. Therefore, there might be changes in the binding affinity of FTD-incorporated DNA to p53 protein. From our previous study, as shown in Fig. 2B, it is presumed that thymidine adjacent to adenine and guanine having a part of electron rich structure is likely to replace FTD in DNA replication¹⁶. This is because the fluorine atoms at the trifluoromethyl group cause an attractive interaction with an electron-rich molecular orbital via a halogen bonding.

In first validation case, normal type and FTD incorporating type of the BAX response elements' sequences binding to p53 were used (Fig. 2C-D). In these figures about p53-binding sequences, some yellow half-tone screening means the area of p53 recognition motif. At this validation case, the thymidine in BAX sequence surrounded by electron rich base was substituted to FTD as shown in Fig. 2D.

In second validation case, we used p53-binding sequences extracted from cancer cell lines that had become resistant to FTDs due to continuously exposed to it over a long period of time. As shown in Fig. 3A, for above sequence extraction, the resistant cells were disrupted and then captured using a consensus binding sequence for the p53 protein. The complementary strand was added based on the captured DNA sequence and is shown in Fig. 3B. Based on our previous theoretical predictions¹⁶, we validated the sequences with FTD substitutions at thymidine positions adjacent to electron-rich bases (Fig. 3C).

(2) Thermodynamical stabilization and binding energies

In accordance with the proposed method described above, the binding complex structures of p53 tetramer were predicted to each DNA, which is BAX sequence (Supplementary Conformational data 1), its sequence incorporating FTD (Supplementary Conformational data 2), p53 binding sequence from FTD resistant cell (Supplementary Conformational data 3), and its binding sequence incorporating FTD (Supplementary Conformational data 4). For these complexes, the conformational structures were sampled with molecular dynamics method in biological temperature ($\sim 37\text{ }^{\circ}\text{C} = 310\text{ K}$) after the equilibration for sufficient time. Figure 4A-B shows the RMSDs of the DNA positions calculated for these sampling structures after superposing the part of p53 protein structures, for BAX sequence and binding sequence from FTD resistant cell, respectively. The variation for a normal type of DNA and a FTD containing DNA are shown as blue line and orange line, respectively. In the BAX sequence, DNA incorporating FTD was found to have lower thermal oscillations than normal type DNA, while normal type DNA from FTD-resistant cells, in contrast, showed increased thermal stability compared to DNA incorporating DNA. Due to these differences in thermal stability, Fig. 5A-B shows the difference in the binding energy distributions fitted to a Gaussian type function, in which the blue line and orange line show the distribution for a normal type of DNA and a DNA incorporating FTD, respectively. The pre-fit distributions and each p-values are shown in Figure S1, S2 as box plots. In this test cases the binding energies were calculated at the MM level with Amber Force Field, which might not be quantitative, but it does qualitatively represent the property. As the DNA recognition probability of p53 has been reported to be very low¹⁰, the binding energy difference caused by sequence differences is also expected to be very small. In fact, the increase in binding energy caused by FTD incorporated DNA was also shown to be slight. However, this means that FTD-incorporated DNA binds to DNA-binding proteins more strongly and is easier to recognize than normal DNA. On the other hand, in p53-binding sequences extracted from FTD-resistant cells, we found that the FTD-containing DNA greatly reduced its binding affinity compared to normal DNA. In other words, it could be suggested that the randomly substituted thymidine to FTD, in resistant cells, might have acquired resistance to the FTDs by entering a position that prevents them from binding to DNA-binding proteins. Using this procedure, the binding energy could be calculated for any mutation-containing p53 in the same way. Not only the case of p53, it also allows us to analyze in terms of the structural and energetical views.

Conclusion

Via a conformational prediction of the arbitrary DNA complex to p53 using theoretical methods such as molecular mechanics and molecular dynamics, we could establish a procedure for predicting p53 binding energies. The validation studies using an antitumor drug, FTD, that exerts its pharmacological effects by incorporating into DNA showed an increased binding affinity for p53 tetramer. As estimated from previous studies, the increase in binding energy was not great even in the presence of increased p53 function. On the other hand, with the cell line having resistance for FTD, the decrease in binding energy was extremely large. The establishment of this procedure made it possible to predict the functional response of p53 that cannot be predicted by only experiments by considering the binding energy difference between any DNA sequence and a control sequence. We believe that this procedure is highly

versatile and applicable to consider the binding energies of not only p53 protein but also the other DNA binding protein and the other biomolecules and raises the importance of computational science in the life sciences.

Methods

Energy minimization and Molecular dynamics simulations.

The energy minimizations were performed using the AMBER 18 program package¹⁷ for each complex structure in ~ 33,000 water molecules. The AMBER 99 force field¹⁸, general AMBER force field (GAFF)¹⁹ and TIP3P²⁰ force field were used for the complex and water molecules, respectively. In this validation case, GAFF was used for FTD. After the energy minimization calculations, the MD simulations (NVT ensemble) were performed at 310 K (~ 37 °C) with periodic boundary conditions using the minimized structure as the initial structure to sample some conformations in biological environment for each structure. The time step was 0.2 fs and the total simulation time was 10 ns (50,000,000 steps).

Binding energies estimation.

We extracted 5,000 conformations from each structure sampled thermodynamically in a mimicked biological environment. The energies of the p53 protein alone, the DNA alone, and their complexes were calculated. And the binding energies were estimated by the energy difference between their complexes from the sum of the energies of the components according to the following equation.

$$\Delta E_{\text{bind}} = (E_{\text{p53}} - E_{\text{DNA}}) - E_{\text{Complex}}$$

Each energy estimation was performed with Gaussian 16 program package²¹. In this validation study, we have calculated the MM level by loading Amber 99 force field and GAFF into Gaussian 16. In order to be able to convert the program to energy calculation at the *ab initio* molecular orbital level, we have made a program that discharges in Gaussian input format.

Extraction of DNA from resistant cells

Colon cancer cell line (DLD1) and FTD resistant DLD1 were cultured in the manner described in the previous report²². DNA derived from these cell lines were extracted using QIAamp DNA Mini Kit (Quiagen Cat. No. 51304).

Declarations

Acknowledgments

This work received the following financial support: grants-in-aid of research by companies; Osaka Cancer Society and Takeda Science Foundation.

Author contributions

JK designed the analyses and performed all computational analyses. JK, HH, and TS performed identifying and gathering information on issues related to p53. MK, AA, TO, and MT performed some experiments. JK, HH, MK, AA, TO, MT and TS have discussed this research in detail. JK, HH, MK, and TS wrote the article.

Competing interests

There is no competing interesting.

References

1. Toledo, F., Wahl, G. M., Regulating the p53 pathway: in vitro hypotheses, in vivo veritas. *Nat Rev Cancer*. **6**, 909-923 (2006)
2. Itahana, Y, Ke, H., Zhang, Y., p53 Oligomerization Is Essential for Its C-terminal Lysine Acetylation. *Biol. Chem*. **284**, 5158-5164 (2009).
3. Green D. R. & Kroemer G., Cytoplasmic functions of the tumour suppressor p53. *Nature* **458**, 1127-1130 (2009)
4. Olivier, M., *et al.*, The IARC TP53 database: new online mutation analysis and recommendations to users. *Mut*. **19**, 607-614 (2002)
5. Baugh, E. H., *et al.*, Why are there hotspot mutations in the *TP53* gene in human cancers? *Cell Death Different*. **25**, 154-160 (2018)
6. Pradhan, M. R., *et al.*, Simulations of mutant p53 DNA binding domains reveal a novel druggable pocket. *Acid Res*. **47**, 1637-1652 (2019)
7. Kamagata, K., *et al.*, How p53 molecules solve the target DNA search problem: A Review. *J. Mol. Sci*. **21**, 1031 (2020)
8. El-Deiry, W. S., *et al.*, Definition of a consensus binding site for p53. *Genet*. **1**, 45-49 (1992)
9. Chen, Y., *et al.*, Structure of p53 binding to the BAX response element reveals DNA unwinding and compression to accommodate base-pair insertion. *Acid Res*. **41**, 8368-8376 (2013)
10. Itoh, Y., *et al.*, Activation of p53 facilitates the target search in DNA by enhancing the target recognition probability. *Mol. Biol*. **17**, 2916-2930 (2016)
11. Heidelberger, C., *et al.*, Syntheses of 5-trifluoromethyluracil and 5-trifluoromethyl- 2'-deoxyuridine. *Med. Chem*. **7**, 1-5 (1964)
12. Ryan, K. J., *et al.*, Chemical Synthesis of 2'-Deoxy-5-(trifluoromethyl)uridine and the α -Anomer. *Org. Chem*. **31**, 1181-1184 (1966)
13. Fujiwara, Y., *et al.*, Fluorinated Pyrimidines XXXVII. Effects of 5-Trifluoromethyl-2'-deoxyuridine on the Synthesis of Deoxyribonucleic Acid of Mammalian Cells in Culture. *Pharmacol*. **6**, 273-280 (1970)

14. Mayer, R. J., *et al.*, Randomized Trial of TAS-102 for Refractory Metastatic Colorectal Cancer. *Engl. J. Med.* **372**, 1909–1919 (2015).
15. Matsuoka, K., *et al.*, Trifluridine induces p53-Dependent Sustained G₂ Phase Arrest with Its Massive Misincorporation into DNA and Few DNA Strand Breaks. *Cancer Ther.* **14**, 1004-1013 (2015)
16. Koseki, J., *et al.*, Theoretical analyses and experimental validation of the effects caused by the fluorinated substituent modification of DNA. *Rep.* **10**, 1138 (2020)
17. Case, D. A., *et al.*, *AMBER 18* 18th, University of California, San Francisco (2018)
18. Wang, J., *et al.*, How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *Comp. Chem.* **21**, 1049–1074 (2000)
19. Wang, J., *et al.*, Development and testing of a general amber force field. *Comput. Chem.* **25**, 1157–1174 (2004).
20. Jorgensen, W. L., *et al.*, Comparison of simple potential functions for simulating liquid water. *Chem. Phys.* **79**, 926–935 (1983)
21. Frisch, M. J., *et al.*, *Gaussian 16*, revision B.01; Gaussian Inc.: Wallingford, CT. (2016)
22. Tsunekuni, K., *et al.*, MicroRNA profiles involved in trifluridine resistance. *Oncotarget* **8**, 53017-53027 (2017)

Figures

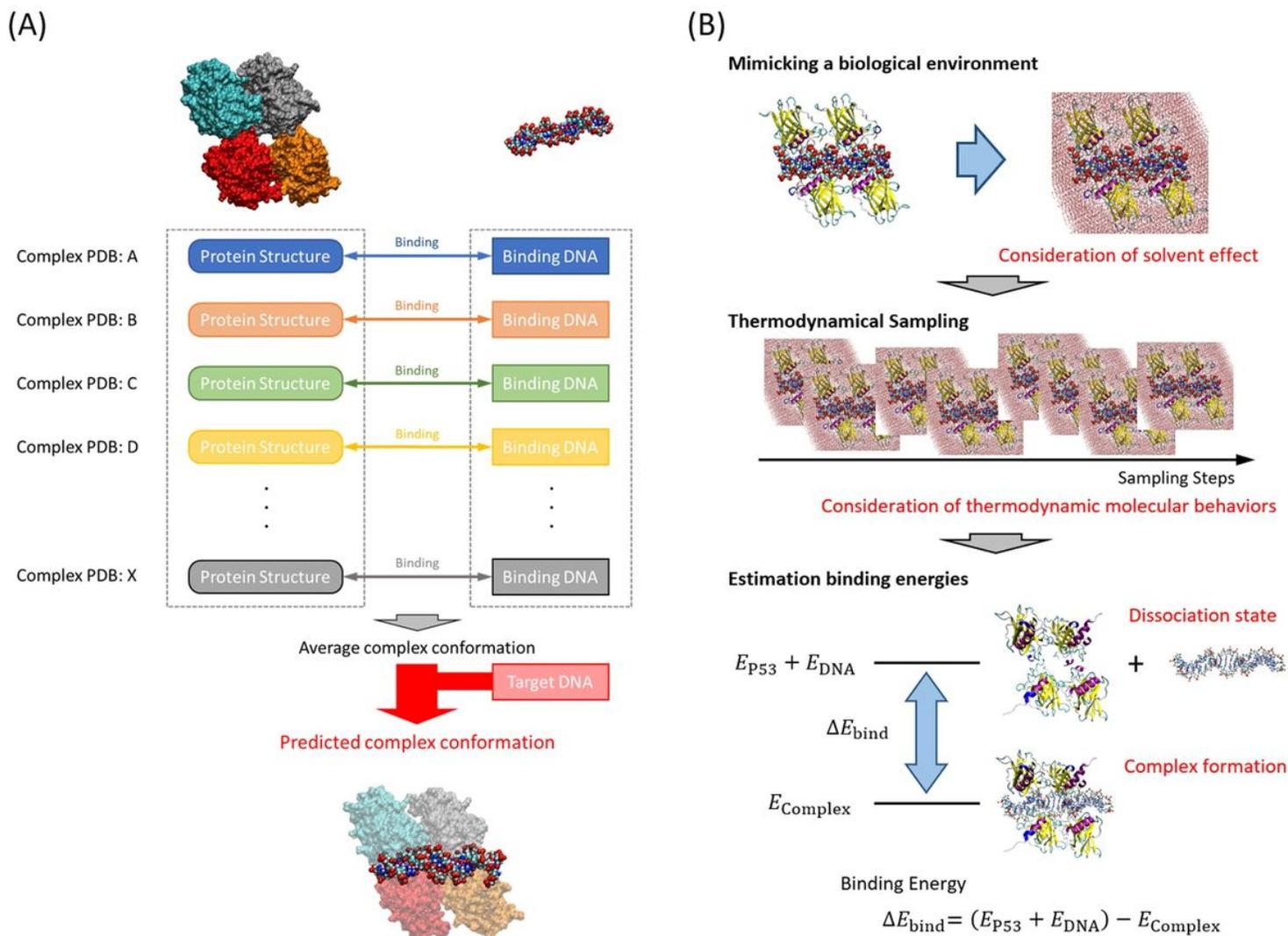


Figure 1

The schematic diagram of our proposed theoretical prediction system of stabilization complex conformation and the binding affinity. (A) After determining the average conformation of the reported PDB crystal structures to binding motif sites, the binding structures are predicted based on the sequence similarity of the target DNA. (B) Thermodynamic sampling will be performed in a mimicking biological environment for the predicted complex structure. And the after, the binding energies are estimated to account for the energy difference between the dissociated and complex states.

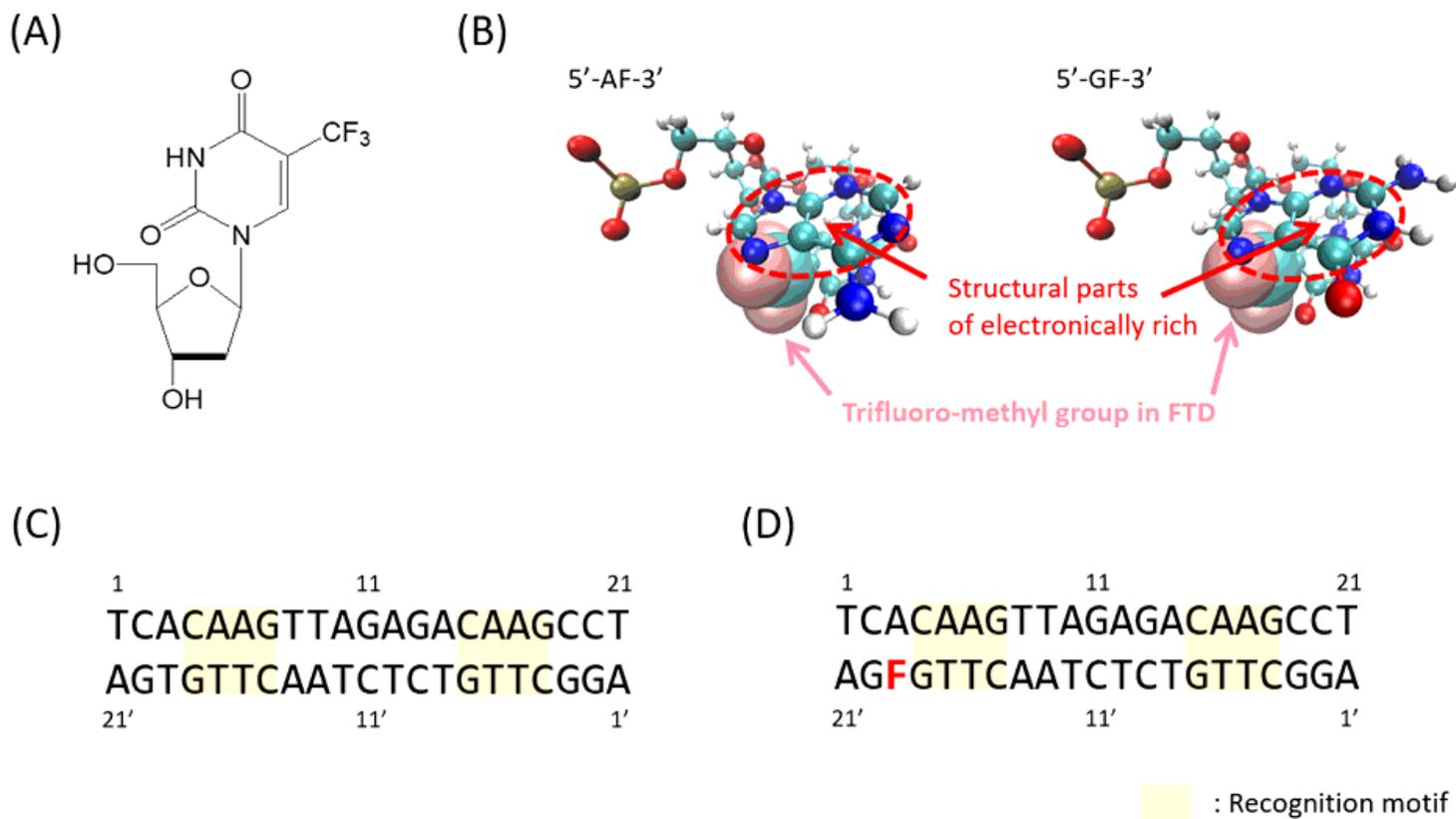


Figure 2

Structural features of FTD, and test sequences binding p53. (A) The structural formula of FTD and (B) examples of adjacent bases that stabilize FTD. The presence of an electron-rich structure in the adjacent bases stabilizes the coordination through an attractive interaction with the trifluoromethyl group of the FTD. And the BAX response elements' sequences of (C) normal type and (D) FTD incorporated.

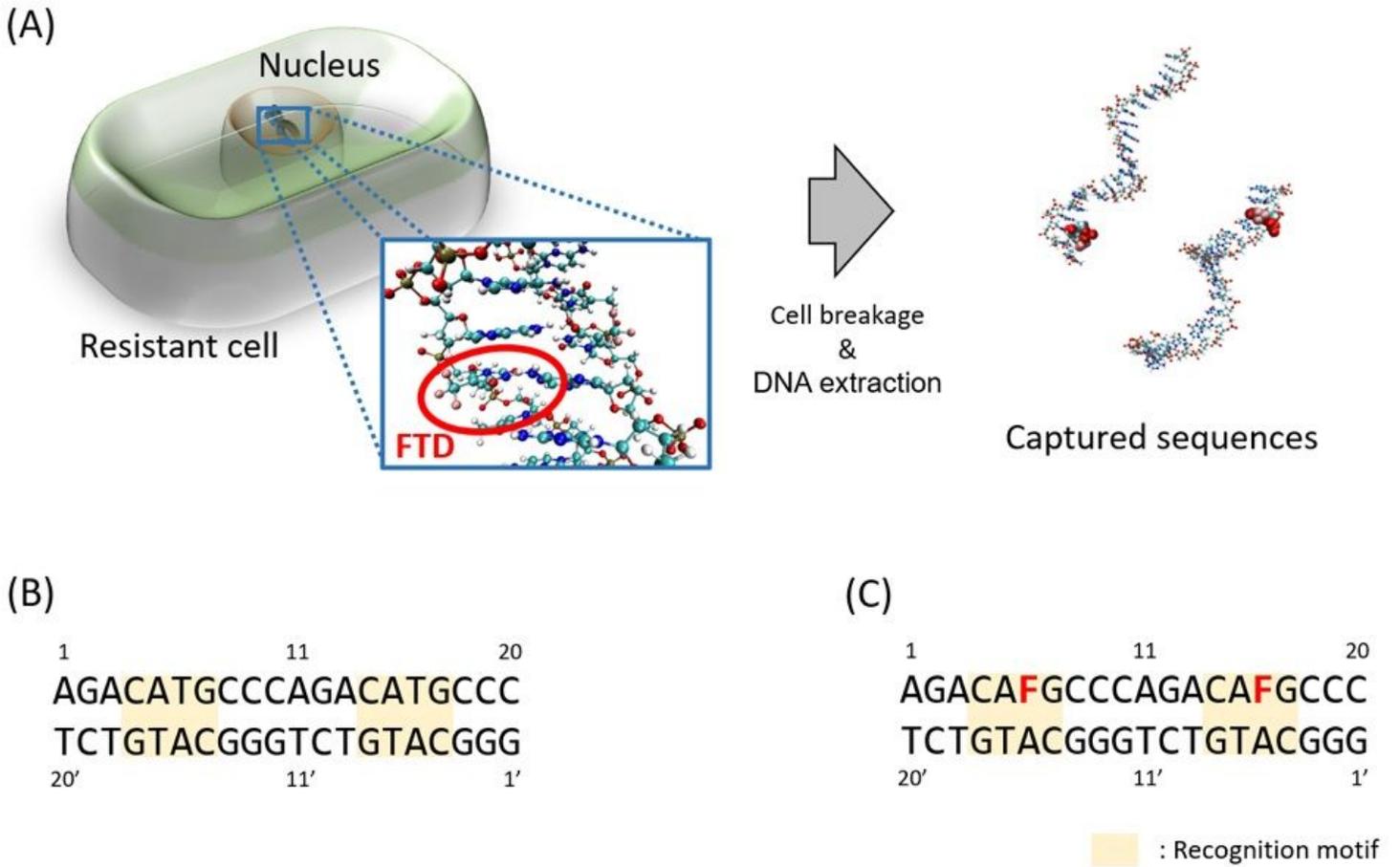


Figure 3

Validation using p53-binding DNA extracted from FTD-resistant cells. (A) The resistant strains were crushed and p53-binding sequences were extracted. The DNA sequences of the (B) normal and (C) FTD-containing types used for verification.

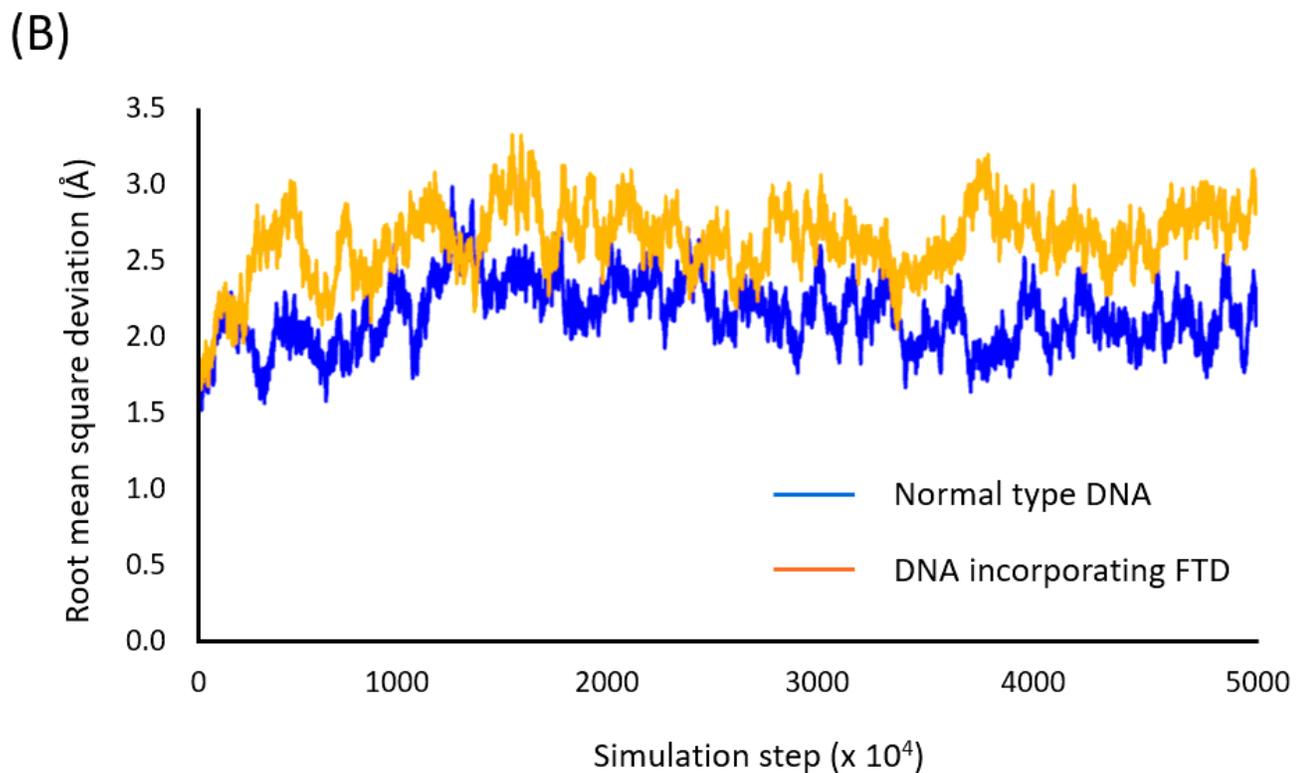
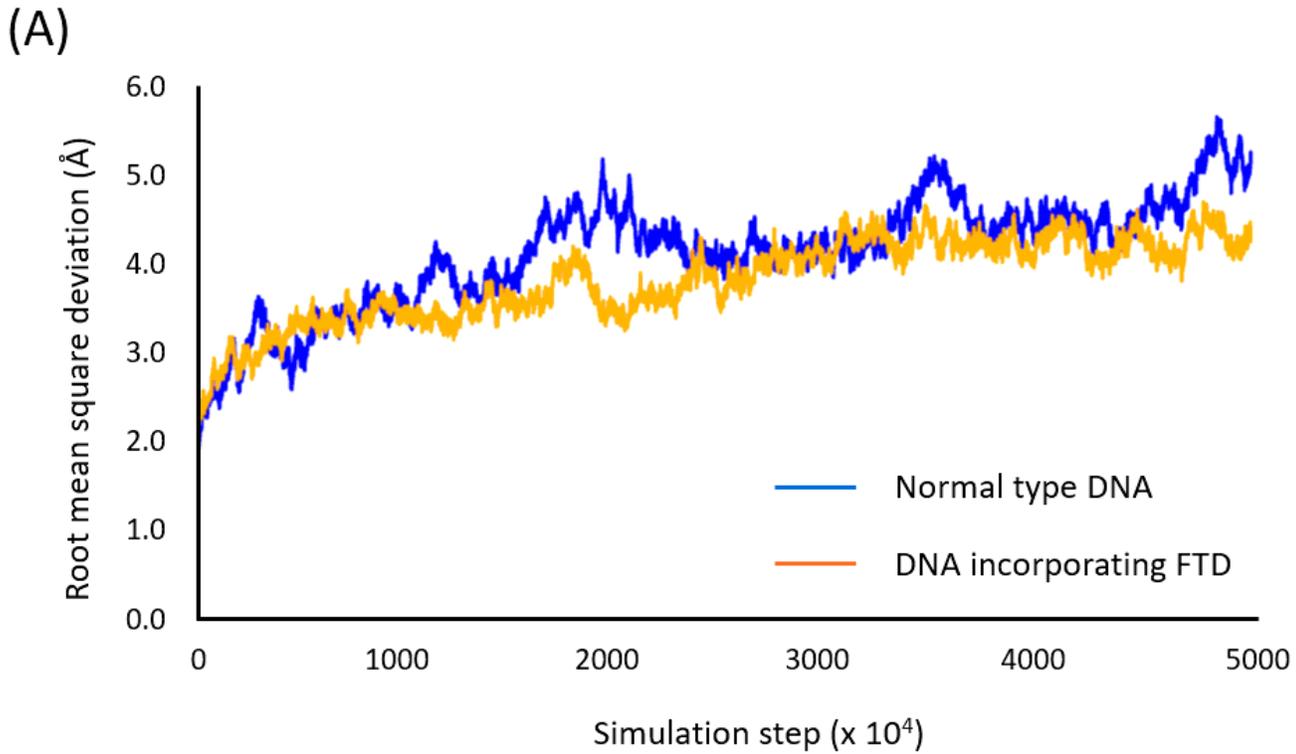


Figure 4

The root mean square deviation (RMSD) analyses for (A) the BAX response elements' sequences and (B) the p53-binding sequences extracted from FTD-resistant cells. The variation for a normal type of DNA and a FTD incorporated DNA are shown as blue line and orange line, respectively.

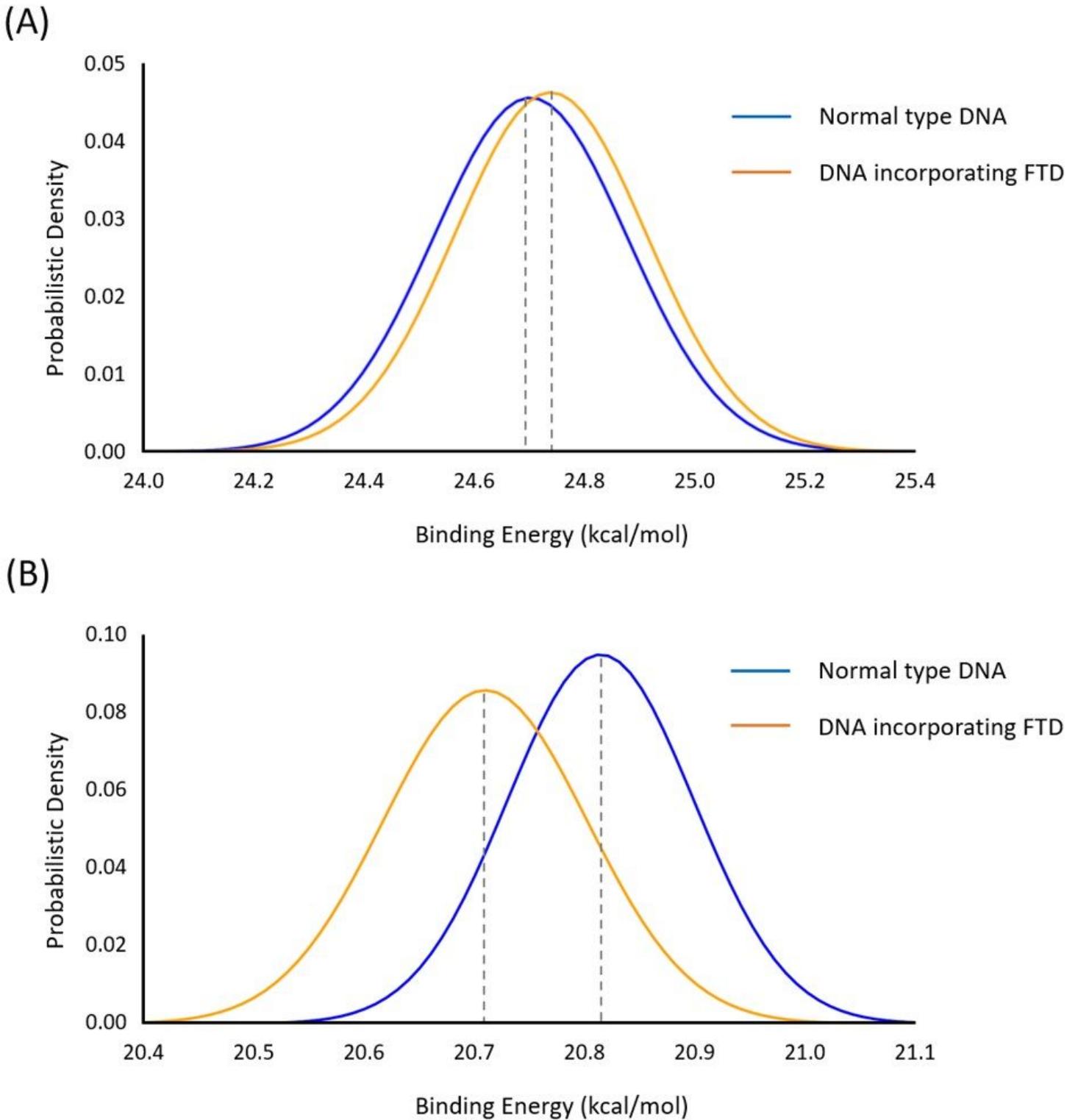


Figure 5

The distributions of P53 binding energy for (A) the BAX response elements' sequences and (B) the p53-binding sequences extracted from FTD-resistant cells. The blue line and orange line show the distribution for a normal type of DNA and a FTD containing DNA, respectively.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SCD1BAXnormaltypeDNA.pdb](#)
- [SCD2BAXDNAincorporatingFTD.pdb](#)
- [SCD3ResistantCellnormaltypeDNA.pdb](#)
- [SCD4ResistantCellDNAincorporatingFTD.pdb](#)
- [SuplInfo.docx](#)