Supplementary Information for

"Modulation of Terahertz Absorption by single mutation of rhodopsin mimics"

Yunyu Wang; Yongnan Hu; Jiajia Meng; Xubiao Peng*; Qing Zhao*.

Center for Quantum Technology Research, Key Laboratory of Advanced Optoelectronic Quantum Architecture and Measurements (MOE), School of Physics, Beijing Institute of Technology, Beijing, 100081, China.

Corresponding Author:

Xubiao Peng: xubiaopeng@bit.edu.cn;

Qing Zhao: qzhaoyuping@bit.edu.cn.

This PDF file includes:

Supporting text;

Figures S1 to S8.

Supporting Information Text

Refractive index and absorption coefficient calculations

According to the method proposed by Duvillaret et al. in 1996 for determining the optical parameters of terahertz waves in samples, the main peak $E_{sam}(t)$ and the first echo peak $E_{echo1}(t)$ in the time-domain spectra can be well separated for thick samples. Therefore, the required optical parameters can be derived using only the information from the main peak.

In the actual experiment, the time-domain spectrum $E_{ref}(t)$ of terahertz wave E_0 after passing through the reference object is obtained as the reference signal. Under the same environment and temperature conditions, the time domain spectrum $E_{sam}(t)$ after passing through the sample is measured as the sample signal. These signals are then Fourier transformed to obtain their frequency spectra $E_{ref}(\omega)$ and $E_{sam}(\omega)$. Since most materials that absorb terahertz waves exhibit a small extinction coefficient ($k_s \ll 1$) and follow the Beer-Lambert Law, the following relation can be derived through simplification:

$$\frac{E_{sam}(\omega)}{E_{ref}(\omega)} = \frac{4n_s(\omega)}{[n_r(\omega)+n_s(\omega)]^2} \cdot \exp\left[-i\left(n_s(\omega) - i\frac{\alpha_s(\omega)c}{2\omega} - n_r(\omega)\right)\frac{\omega d}{c}\right]$$

$$= \frac{4n_s(\omega)}{[n_r(\omega)+n_s(\omega)]^2} \cdot exp[-\frac{\alpha_s(\omega)d}{2}] \cdot \exp\left[-i\left(n_s(\omega) - n_r(\omega)\right)\frac{\omega d}{c}\right]$$

$$= A(\omega) \cdot exp[i\varphi(\omega)]$$

where $A(\omega)$ represents the ratio of the amplitude of the signal in the frequency domain between the protein sample and the reference PTFE, and $\varphi(\omega)$ represents the phase difference of the signal in the frequency domain between the protein sample and the reference PTFE.

Therefore, it can be obtained from the above formula:

$$A(\omega) = \frac{4n_s(\omega)}{[n_r(\omega) + n_s(\omega)]^2} \cdot exp[-\frac{\alpha_s(\omega)d}{2}]$$

$$\varphi(\omega) = -\left(n_s(\omega) - n_r(\omega)\right)\frac{\omega d}{c}$$

Finally, the refractive index $n_s(\omega)$ and the absorption coefficient $\alpha_s(\omega)$ in the terahertz can be expressed as, respectively:

$$n_s(\omega) = n_r(\omega) - \frac{c}{\omega d}\varphi(\omega)$$

$$\alpha_s(\omega) = -\frac{2}{d}ln\frac{A(\omega)[n_r(\omega) + n_s(\omega)]^2}{4n_s(\omega)}$$

where $n_s(\omega)$ is the refractive index of the sample. $\alpha_s(\omega)$ is the absorption coefficient of the sample in the terahertz band. $n_r(\omega)$ is the refractive index of the reference. For PTFE, we take constant 1.433 here. c is the propagation velocity of the THz wave in a nitrogen environment, ω is the angular frequency, and d is the sample thickness.

Figure S1

a.

Protein Sequence

```
          10        20        30        40        50
R1  PNFSGNWKII  RSENFEELLK  VLGVNVMLRK  IAVAAASKYA  VEIKQEGDTF
          10        20        30        40        50
R2  PNFSGNWKII  RSENFEELLK  VLGVNVMLRK  IWVAAASKYA  VEIKQEGDTF

          60        70        80        90       100
R1  YIKVSTTVYT  TEINFKVGEE  FEEQTVDGRP  CKSLVKWESE  NKMVCEQKLL
          60        70        80        90       100
R2  YIKVSTTVYT  TEINFKVGEE  FEEQTVDGRP  CKSLVKWESE  NKMVCEQKLL

          110       120       130       137
R1  KGEGPKTSWT  KELTNDGELI  LTMTADDVVC  TQVFVRE
          110       120       130       137
R2  KGEGPKTSWT  KELTNDGELI  LTMTADDVVC  TQVFVRE
```

b.

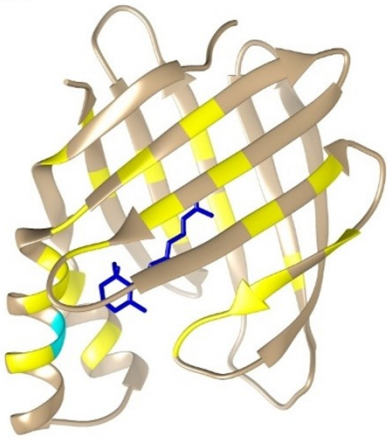


Figure S1: a. Aligned sequence of R1 and R2. The mutation A32W are highlight in yellow. The residues that are within 5Å from retinal with binding are colored in red font. **b. The 3D structure of the rhodopsin mimics.** The structure in dark blue is the retinal, the bright yellow part is the residues less than 5Å away from the retinal after binding with the protein, and the mutation A32W is colored in cyan. Notably, the A32W is also within 5Å from the retinal.

Figure S2

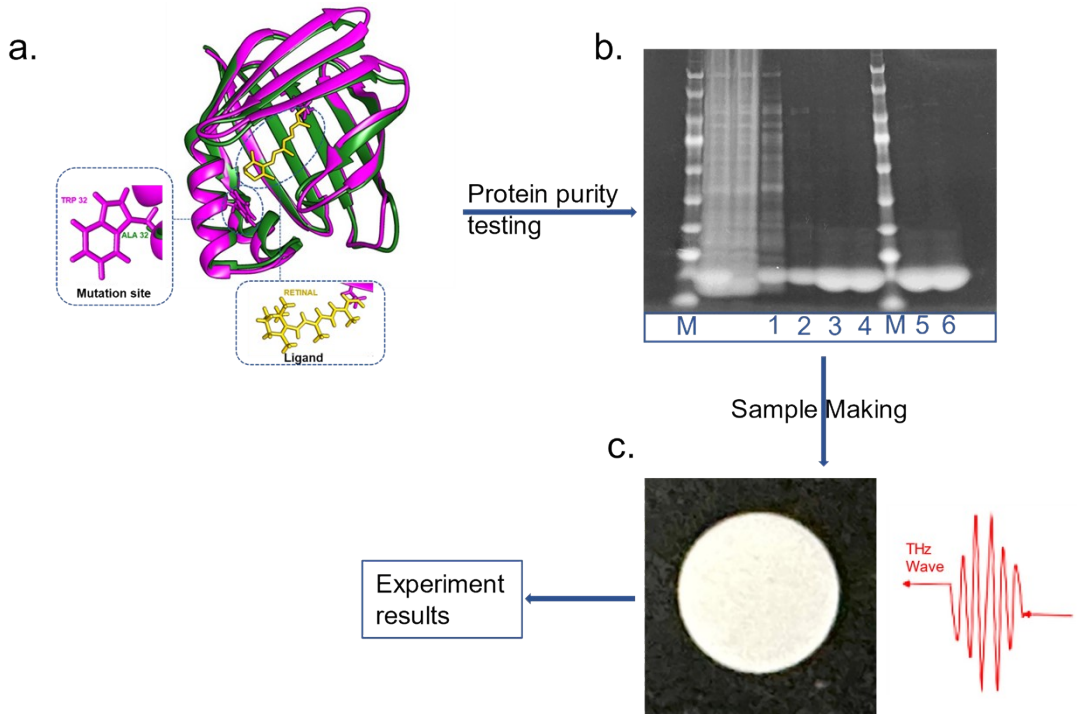


Figure S2: Schematic diagram of the experiment. a. Schematic diagram of rhodopsin mimic structure. The green part is the structure of mutant R1, and the purple part is the structure of mutant R2. R1 and R2 mutants differ only in one A32W residue. Among them, the residue includes alanine in R1 and tryptophan in R2. Both mutants have a packet, with the all-trans-retinal chromophore (SI Fig.1 dark blue structure) bound to the pocket. The yellow part is the ligand for two proteins: all-trans retinal. **b.** Schematic diagram of sample purity detected by electrophoresis, where bands 3 and 5 are purified R1 protein bands and 4 and 6 are purified R2 protein bands, both of which meet the experimental requirements. **c.** Schematic diagram of the terahertz absorption measurement on the solid tablet sample.

Figure S3

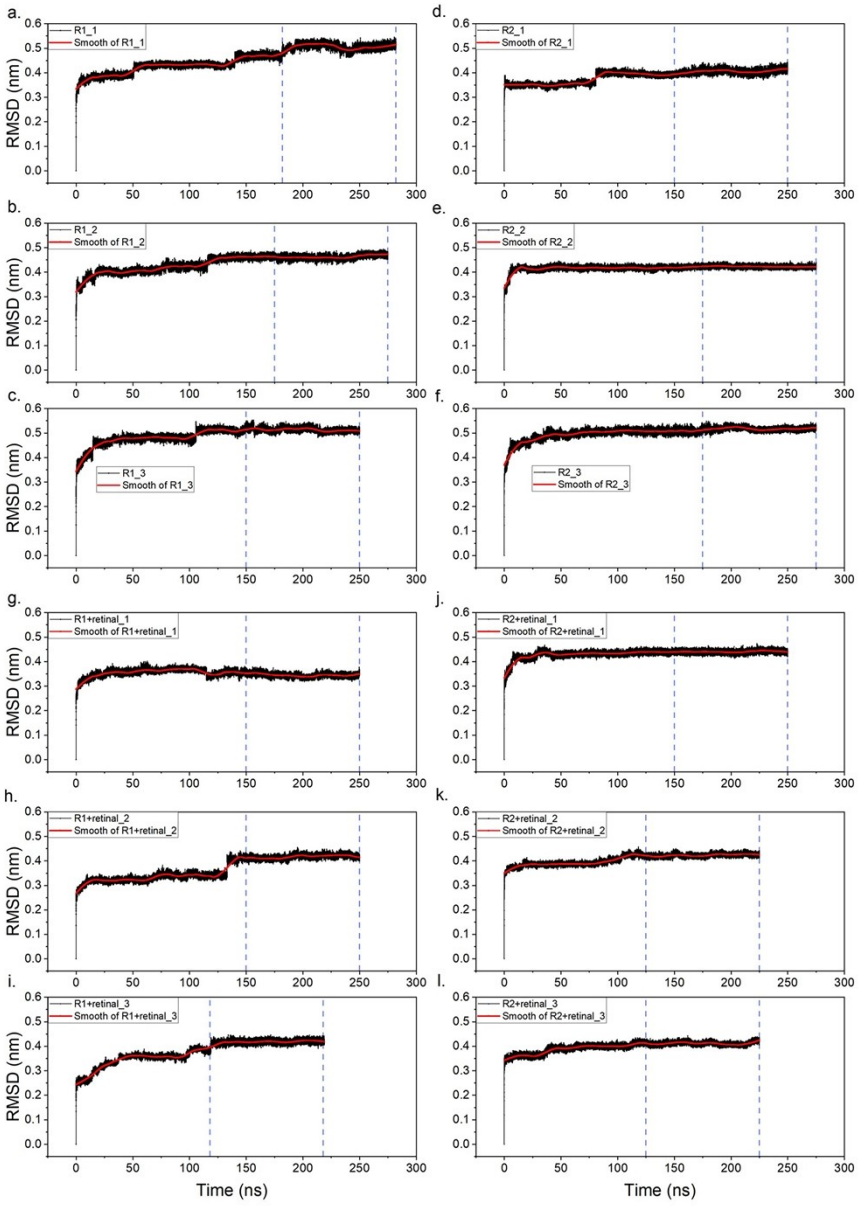


Figure S3 Schematic diagram of Root Mean Square Deviation (RMSD) of rhodopsin mimics. Figure a-c: Schematic diagram of RMSD of protein R1. **Figure d-f:** Schematic diagram of RMSD of protein R2. **Figure g-i:** Schematic diagram of RMSD of protein R1 after binding with retinal. **Figure j-l:** Schematic diagram of RMSD of protein R2 after binding with retinal. In each diagram, the equilibrium region marked by the blue dashed line represents the trajectory file section used for calculating the dipole moment autocorrelation function.

Figure S4

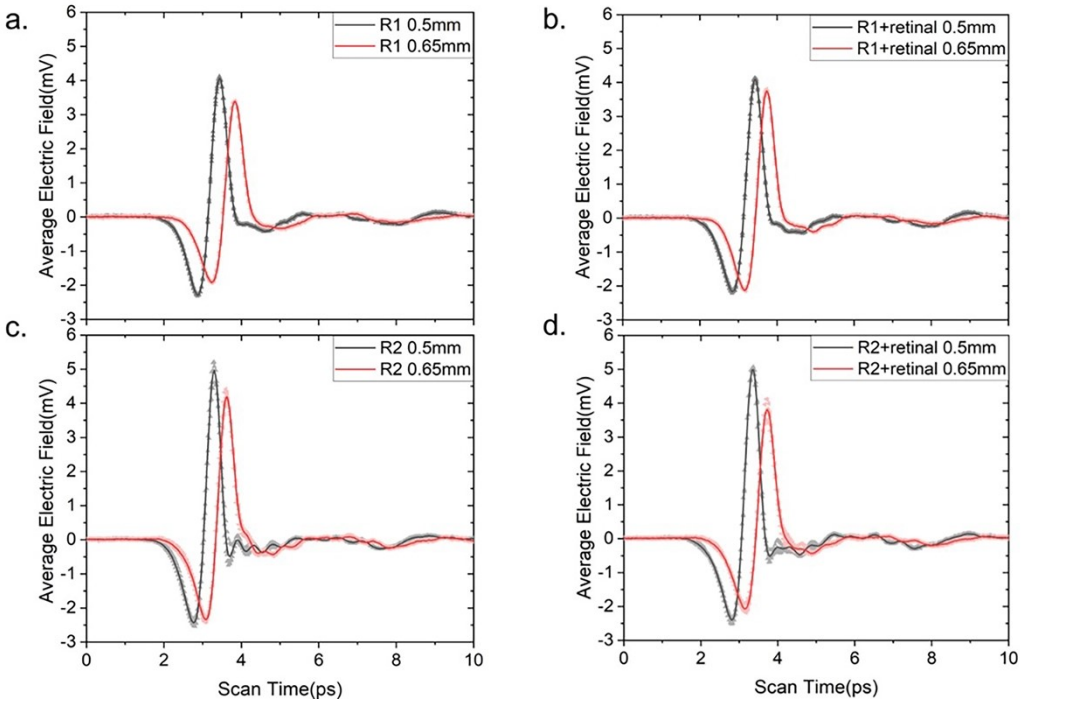


Figure S4: Terahertz time-domain spectra of rhodopsin mimics with thickness of 0.5mm and 0.65mm. a. Terahertz time domain spectra of R1. **b.** Terahertz time-domain spectra of R1 with retinal binding. **c.** Terahertz time domain spectra of R2. **d.** Terahertz time domain spectra of R2 with retinal binding.

Figure S5

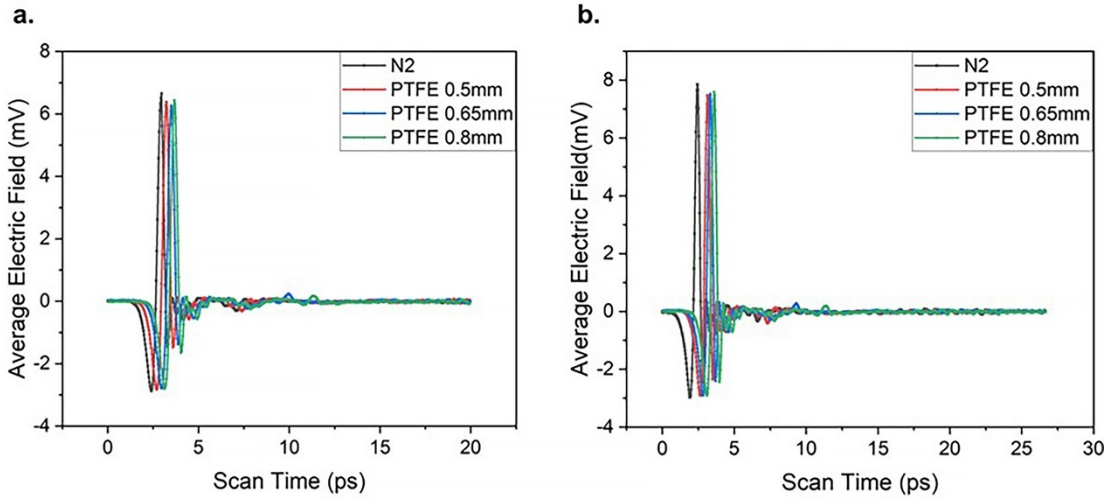


Figure S5: Terahertz time-domain spectra of the N_2 and reference PTFE of different thicknesses. a. Terahertz time-domain spectra of the reference material PTFE used in the experiment on R1. **b.** Terahertz time-domain spectra of the reference material PTFE used in the experiment on R2.

Figure S6

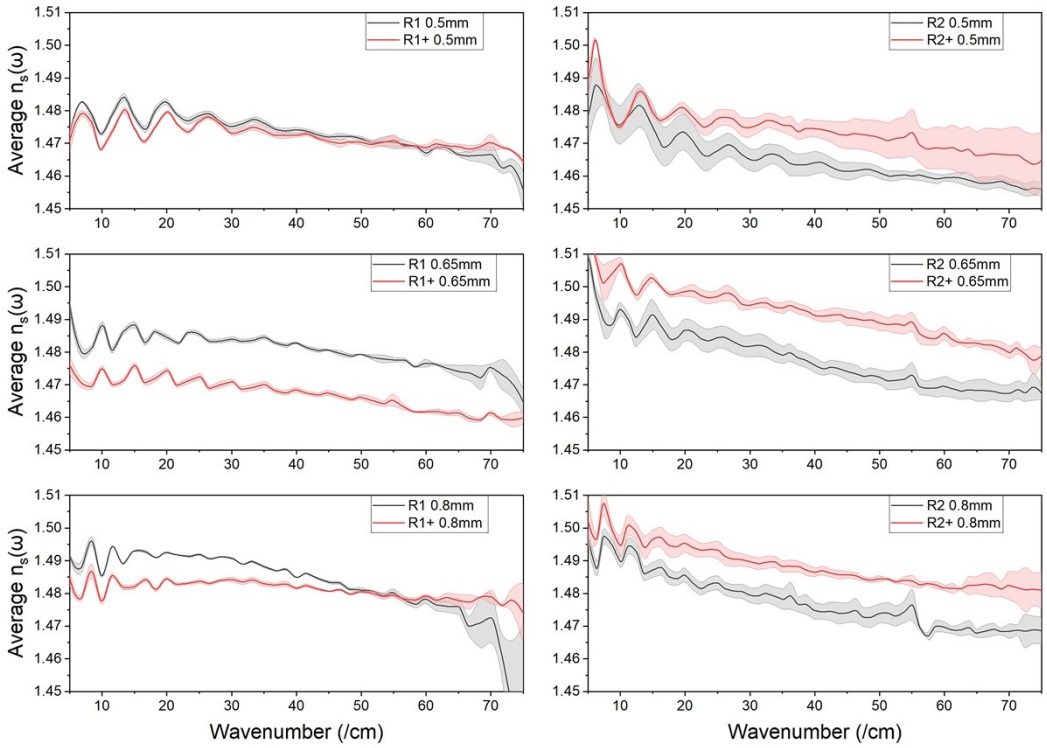
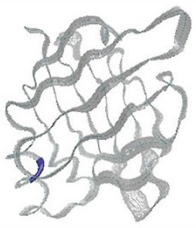
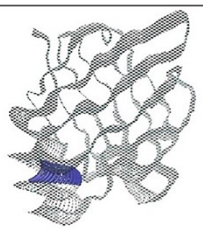
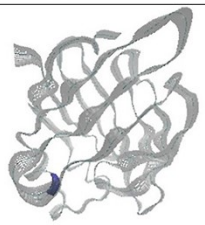
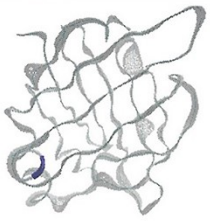
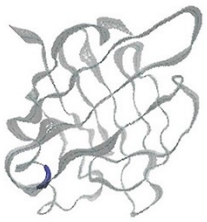
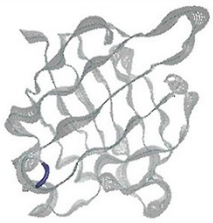
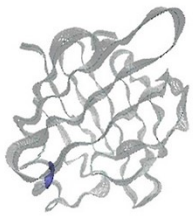
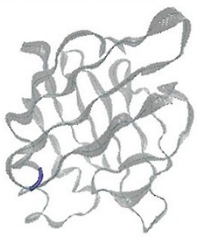
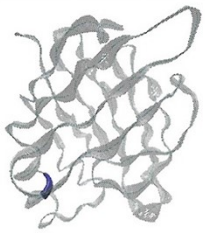


Figure S6: Relative refractive indices of R1 and R2 samples with different thicknesses before and after binding with retinal, with the shaded area representing the standard deviation from three independent measurements.

Figure S7

a:

PROTEIN R1			
Absorption Frequency without Binding	Utmost Vibration area (Directions included)	Absorption Frequency with Binding	Utmost Vibration area (Directions included)
8.8/cm		7.9/cm	
13.7/cm		13.2/cm	
35.4/cm		35.0/cm	
56.8/cm		55.6/cm	
65.2/cm		67.8/cm	

b:

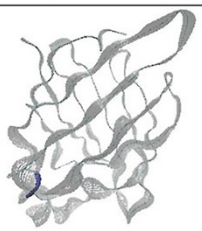
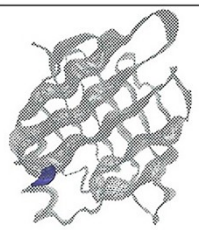
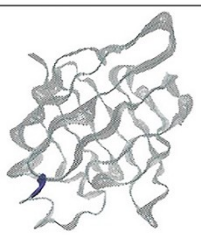
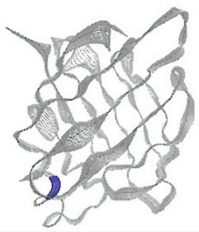
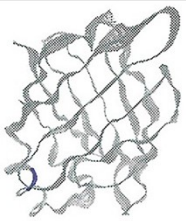
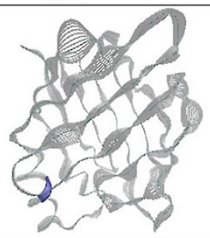
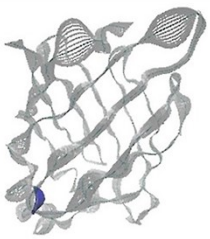
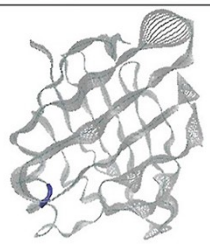
PROTEIN R2			
Absorption Frequency without Binding	Utmost Vibration area (Directions included)	Absorption Frequency with Binding	Utmost Vibration area (Directions included)
7.5/cm		7.6/cm	
11.9/cm		12.5/cm	
40.6/cm		40.3/cm	
58.9/cm		59.4/cm	
68.8/cm		66.9/cm	

Figure S7: a: Vibration direction of R1 protein at different eigenfrequencies without and with retinal binding. b: Vibration direction of R2 protein at different eigenfrequencies without and with retinal binding. (The gray part represents the protein vibration trajectory, while the dark blue part represents the mutation sites of two proteins).

Figure S8

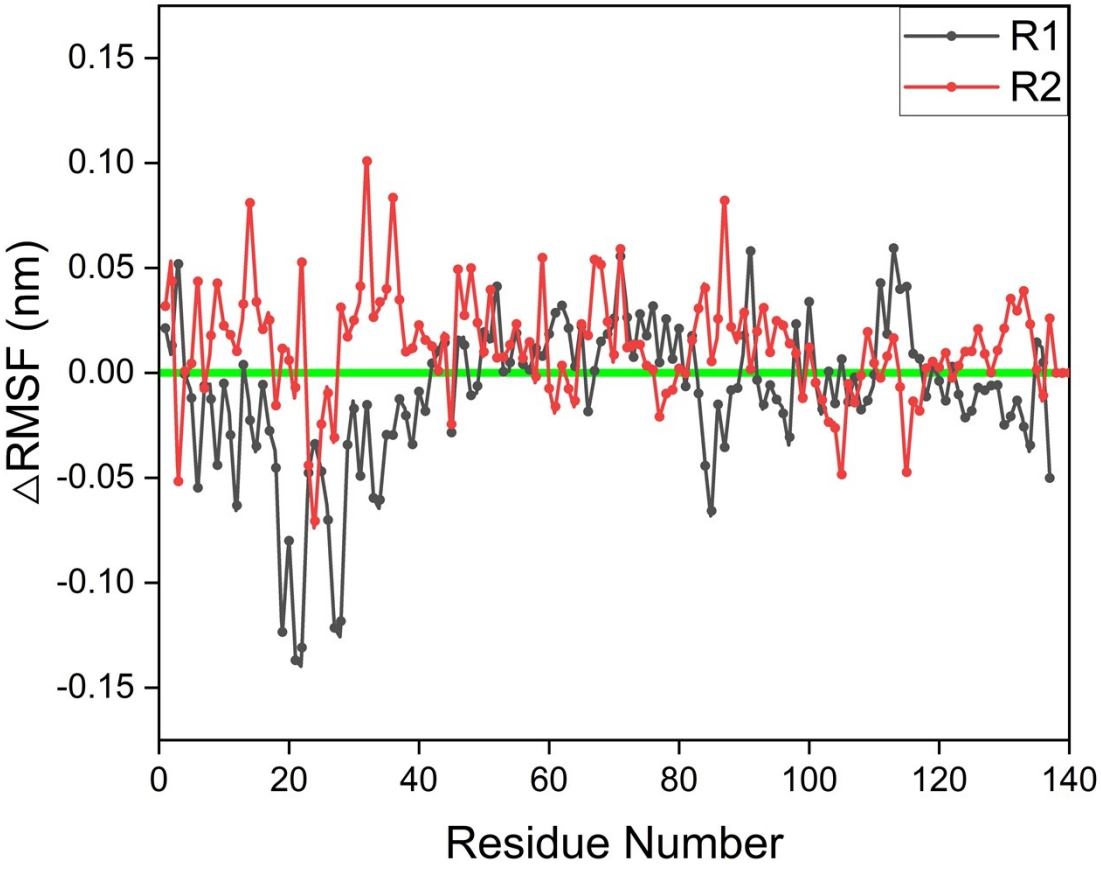


Figure S8: Comparison of RMSF differences (ΔRMSF=RMSF $_{\text{with retinal binding}}$- RMSF $_{\text{without retinal binding}}$) between R1 and R2. The x-axis represents the residue number, while the y-axis shows ΔRMSF (in nm). The black data line corresponds to the results for R1, and the red data line represents the results for R2. RMSF values were derived from three independent MD simulations, averaged over the last 100 ns of the equilibrated trajectory.