Supplementary Text S1. Docking calculation scheme on MEGADOCK

In MEGADOCK [S1,S2], one considers a protein and a docking-target as a receptor (Rec) and a ligand (Lig), respectively. All molecules are allocated on a 3D voxel space with grid-point spacing of 1.2 Å. Then, the score is assigned to each voxel (l,m,n). To evaluate the score several contributions are evaluated in advance. The shape complementarity scores G_{Rec} and G_{Lig} of Rec and Lig are represented as

$$G_{\mathrm{Rec}}(l,m,n) = \begin{cases} \text{\# of Rec atoms within } \left(3.6\text{Å} + r_{vdW}\ of \text{ Rec atoms in the voxel } (l,m,n)\right) & \text{(open space)} \\ -45 & \text{(inside of Rec)} \end{cases} \quad (1),$$

and

$$G_{\mathrm{Lig}}(l,m,n) = \begin{cases} 0 & \text{(solvent accessible surface layer of Lig)} \\ 1 & \text{(solvent excluding surface layer of Lig)} \\ 1 & \text{(core of Lig)} \\ 0 & \text{(open space)} \end{cases} \quad (2),$$

where G_{Rec} and G_{Lig} are assigned to the receptor and ligand voxels, respectively. r_{vdW} is the van der Waals radius of an atom. The real Pairwise Shape Complementarity (rPSC) score is represented as follows;

$$\mathrm{rPSC}(\alpha,\beta,\gamma) = \sum_{l=1}^{N}\sum_{m=1}^{N}\sum_{n=1}^{N} G_{\mathrm{Rec}}(l,m,n)G_{\mathrm{Lig}}(l+\alpha,m+\beta,n+\gamma) \quad (3),$$

where (α,β,γ) is a vector of the ligand translation. Furthermore, MEGADOCK takes into account the electronic interactions of each amino acid residue as a physicochemical score ELEC. The electric field ϕ_a is assigned to each voxel $a \in \Omega$ as follows;

$$\phi_a = \sum_{b\in\Omega} \frac{q_b}{\varepsilon(r_{ab})r_{ab}} \quad (4),$$

$$\varepsilon(r) = \begin{cases} 4 & (r \le 6\text{Å}) \\ 38r - 224 & (6\text{Å} < r < 8\text{Å}) \\ 80 & (8\text{Å} \le r) \end{cases} \quad (5),$$

where q_b is the charge of a voxel, $b \in \Omega$, r_{ab} is the Euclidean distance between a and b, and $\varepsilon(r)$ is a distance-dependent dielectric function. The electrostatic terms E_{Rec} and E_{Lig} are decided by charge of each voxel, $q(l,m,n)$, and the atomic charge of residues is decided by CHARMM19 [S3].

$$E_{\mathrm{Rec}}(l,m,n)=\begin{cases}\phi_{(l,m,n)} & \text{(entire voxel excluding core)}\\ 0 & \text{(core of Rec)}\end{cases}\quad (6),$$

$$E_{\mathrm{Lig}}(l,m,n)=q(l,m,n)\quad (7),$$

$$\mathrm{ELEC}(\alpha,\beta,\gamma)=\sum_{l=1}^{N}\sum_{m=1}^{N}\sum_{n=1}^{N}E_{\mathrm{Rec}}(l,m,n)E_{\mathrm{Lig}}(l+\alpha,m+\beta,n+\gamma)\quad (8),$$

The docking score *S* is represented by above terms as

$$R(l,m,n)=G_{\mathrm{Rec}}(l,m,n)+iE_{\mathrm{Rec}}(l,m,n)\quad (9),$$

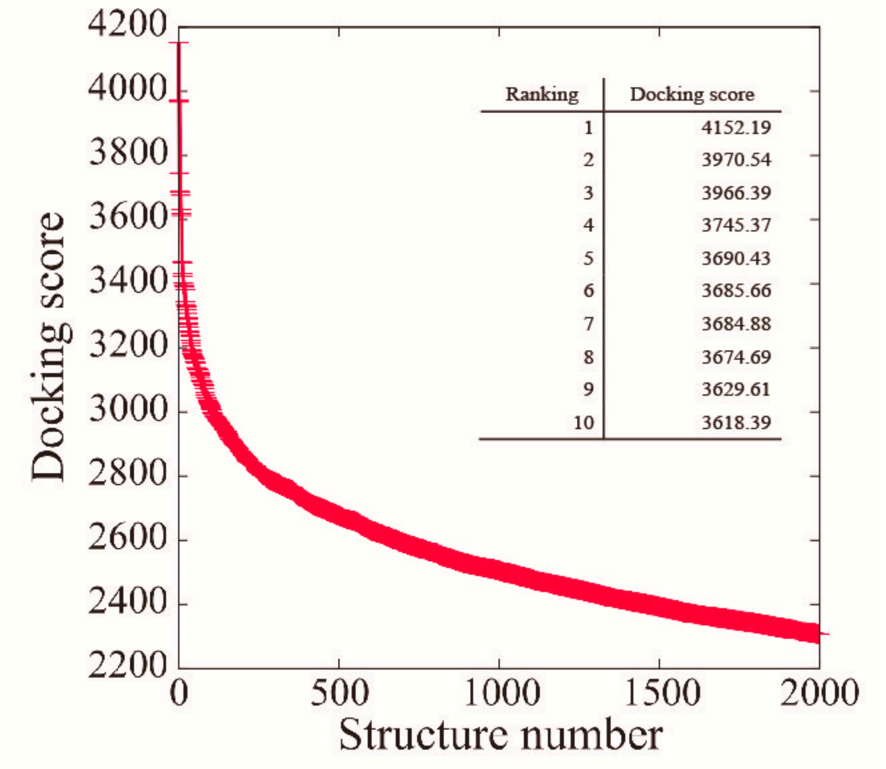
$$L(l,m,n)=G_{\mathrm{Lig}}(l,m,n)+i\omega E_{\mathrm{Lig}}(l,m,n)\quad (10),$$

$$\begin{aligned}S(\alpha,\beta,\gamma)&=\Re\left[\sum_{l=1}^{N}\sum_{m=1}^{N}\sum_{n=1}^{N}R(l,m,n)L(l+\alpha,m+\beta,n+\gamma)\right]\\&=\mathrm{rPSC}(\alpha,\beta,\gamma)-\omega\mathrm{ELEC}(\alpha,\beta,\gamma)\qquad (11),\end{aligned}$$

where $\Re[z]$ is the real part of z, *i* is an imaginary unit $\sqrt{-1}$, and ω is a weight parameter. Since, the summation over real space grid is time consumaing, the docking score *S* is evaluated by using the fast Fourier transform (FFT) algorithm. By denoting the discrete Fourier transforms (DFTs), and inverse discrete Fourier transforms (IFTs), respectively, *S* can be rewritten as

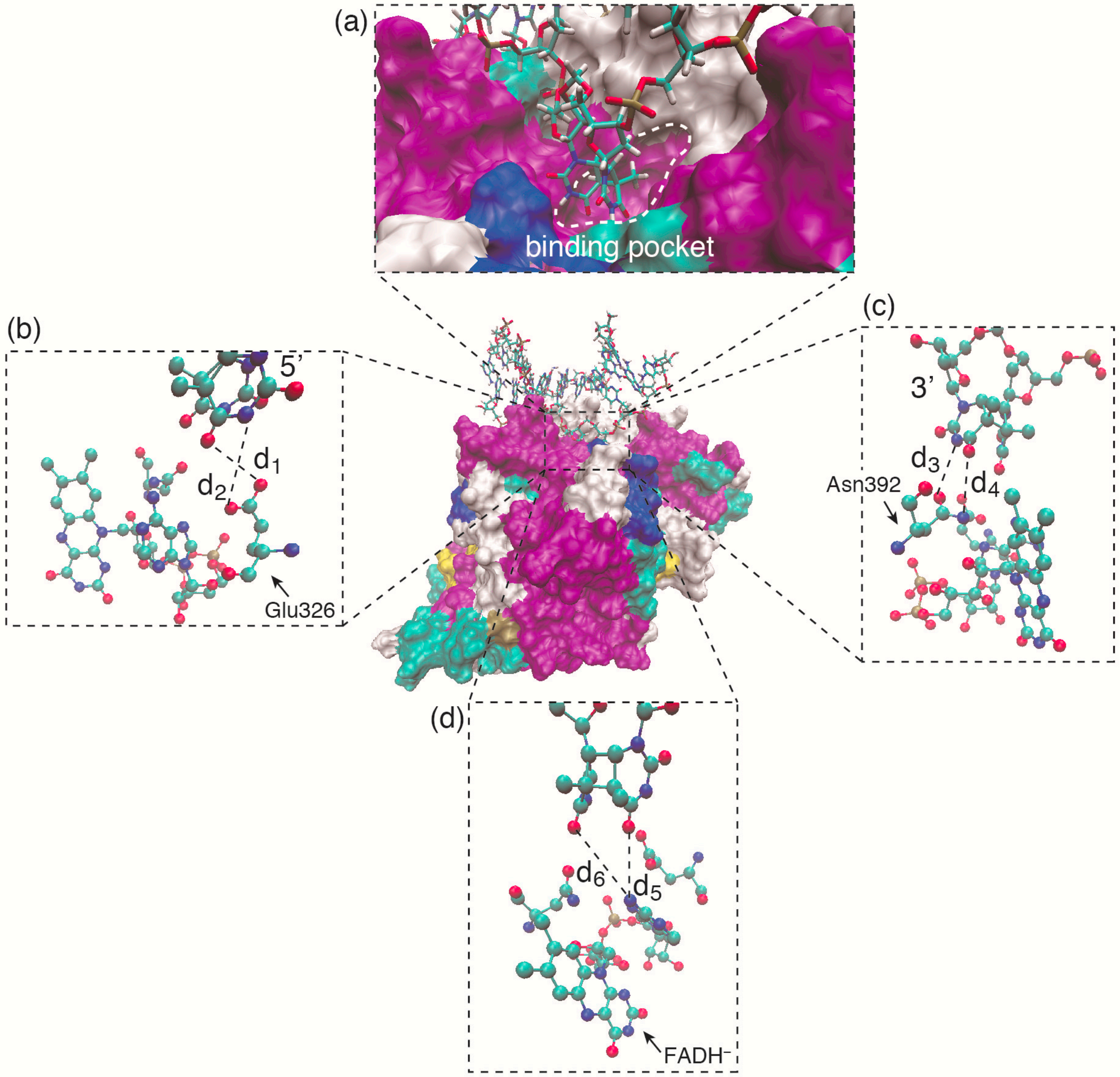
$$S(\alpha,\beta,\gamma)=\Re\left[\mathrm{IFT}\left[\mathrm{DFT}[R(l,m,n)]^{*}\,\mathrm{DFT}[L(l,m,n)]\right]\right]\quad (12).$$

where z* is the complex conjugate of z.

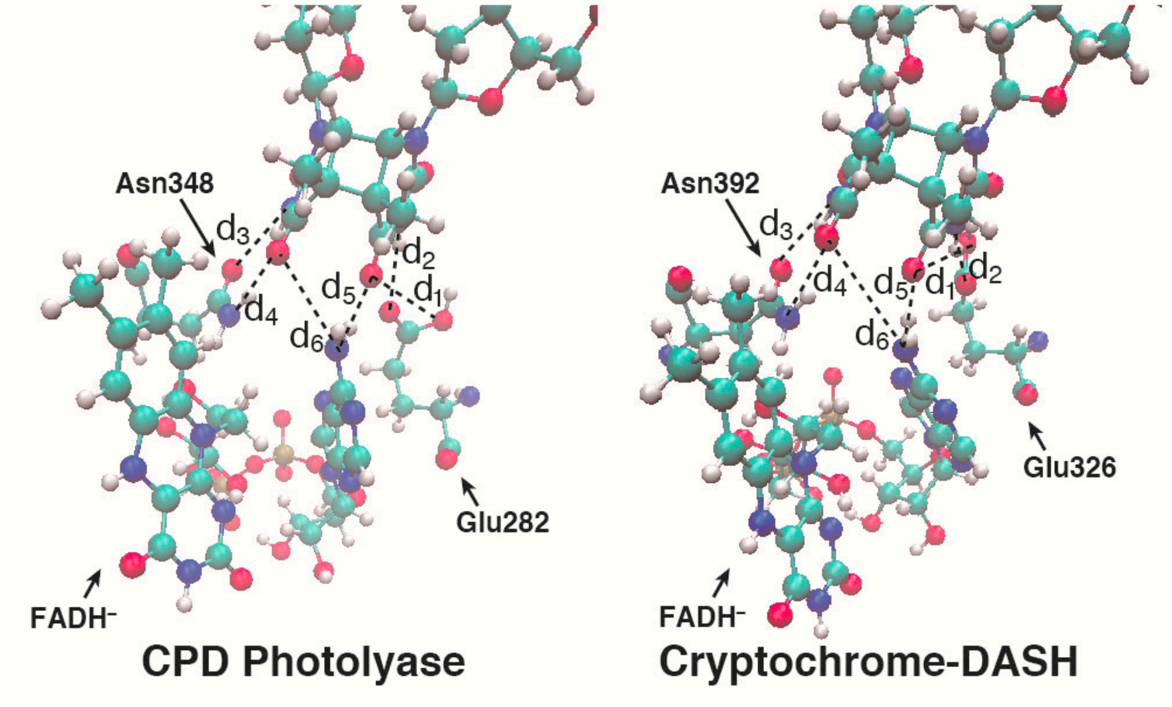


Ranking	Docking score
1	4152.19
2	3970.54
3	3966.39
4	3745.37
5	3690.43
6	3685.66
7	3684.88
8	3674.69
9	3629.61
10	3618.39

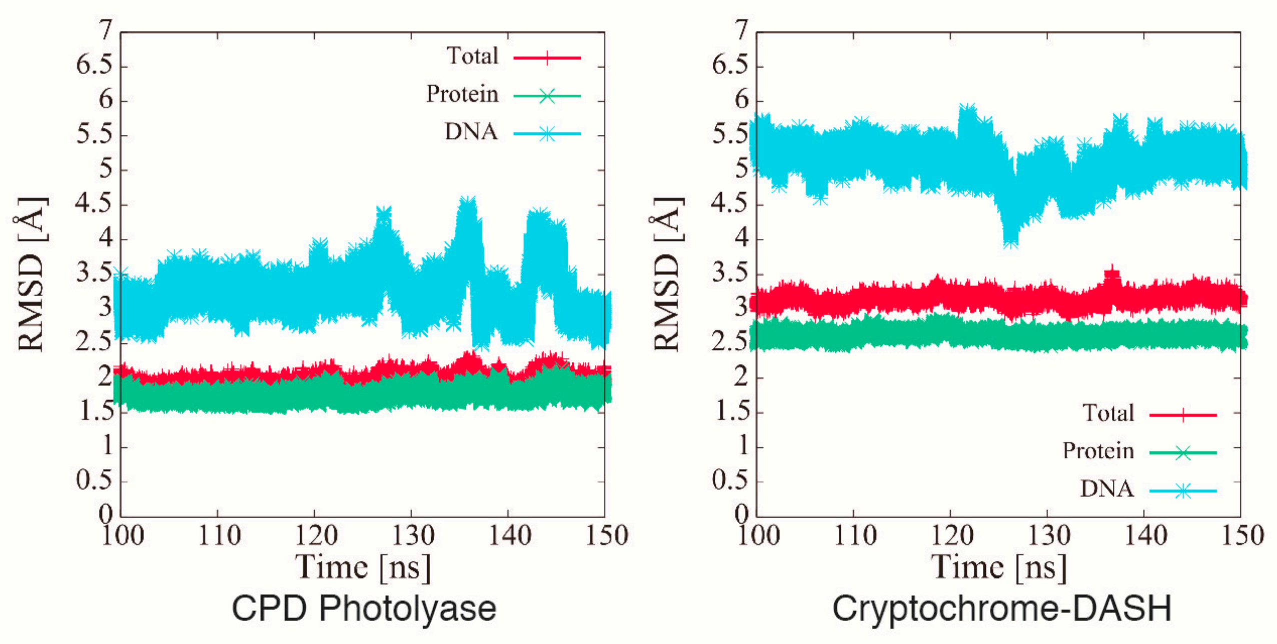
Supplementary Figure S1. Docking score for receptor (CRY-DASH (PDB ID: 2VTB)) and ligand (UV-damaged duplex DNA (PDB ID: 1TEZ)). The top 10 docking scores are also displayed in the panel.



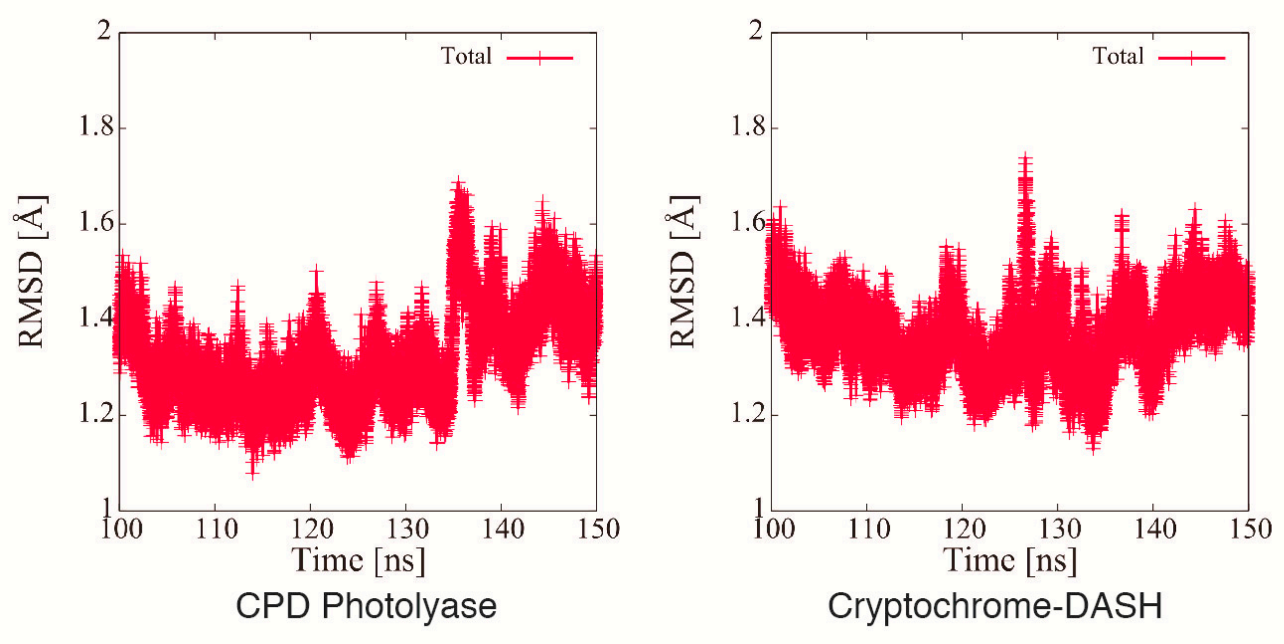
Supplementary Figure S2. Model structure of CRY-DASH complex with best docking score. (a) CPD nearby the binding pocket, (b) 5'side of CPD and Glu326, (c) the 3'side of CPD and Asn392, and (d) the CPD and $FADH^-$ in the binding pocket. d_1, d_2, d_3, d_4, d_5, and d_6 are the distances between N_2 of CPD and O_1 of Glu326, between O_1 of CPD and O_1 of Clu326, between N_1 of CPD and O of Asn392, between O_2 of CPD and N of Asn392, between O_1 of CPD and N of $FADH^-$, and between O_2 of CPD and N of $FADH^-$, respectively. And d_1, d_2, d_3, d_4, d_5, and d_6 are 4.35, 4.93, 4.32, 4.16, 4.07, and 5.06 Å, respectively, while the corresponding distances for CPD-PHR are 2.88, 3.53, 3.00, 3.36, 3.12, and 3.16 Å, respectively.



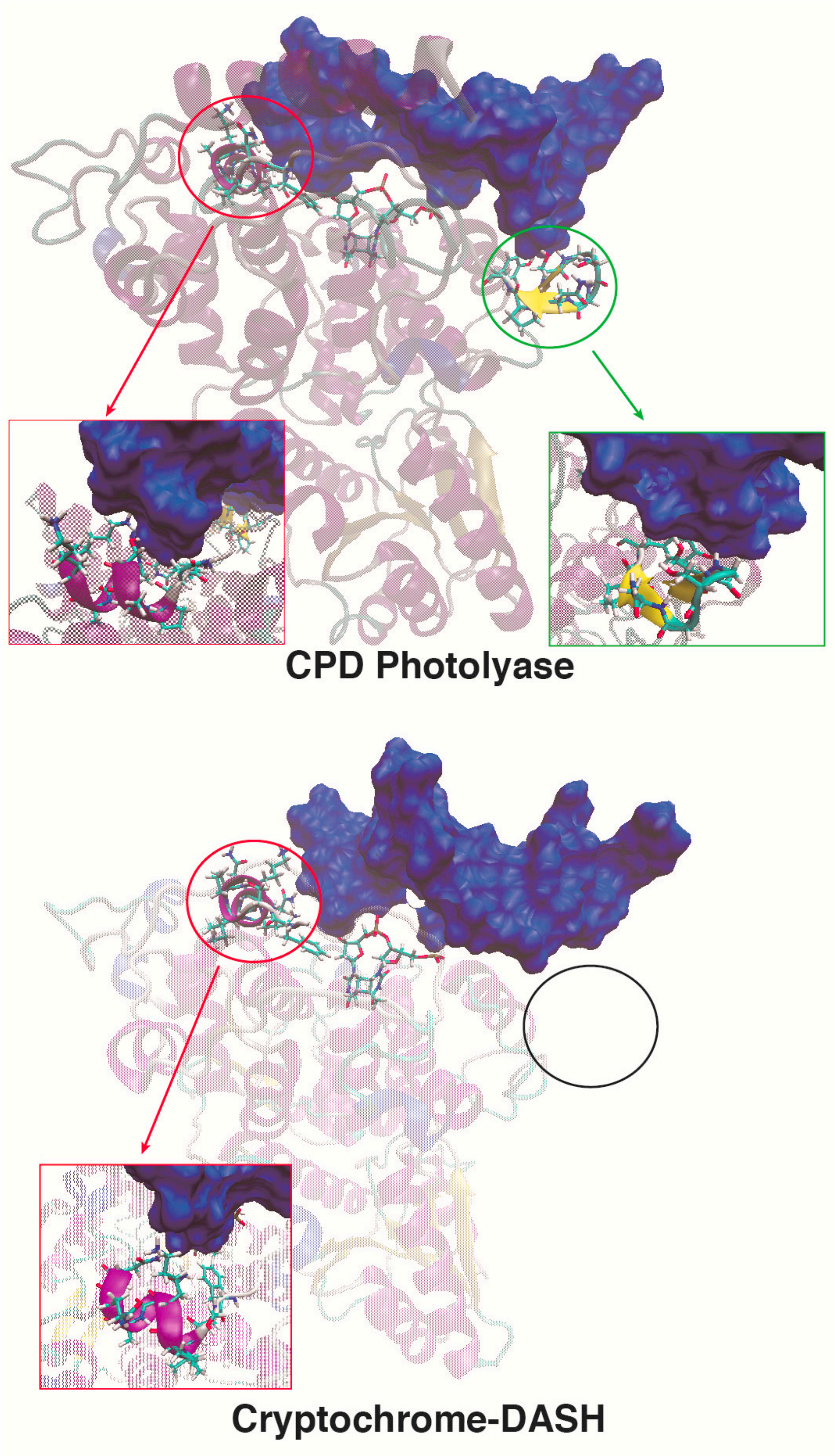
Supplementary Figure S3. Structures after energy minimization of CPD-PHR and CRY-DASH. Here distances defined in Supplementary Figure S2**,** d_1, d_2, d_3, d_4, d_5, and d_6**,** are 4.98, 3.37, 2.96, 2.76, 3.02, and 3.53 Å in CPD-PHR (left), and 5.24, 2.80, 2.93, 2.85, 2.86, and 3.91 Å for CRY-DASH (right), respectively.



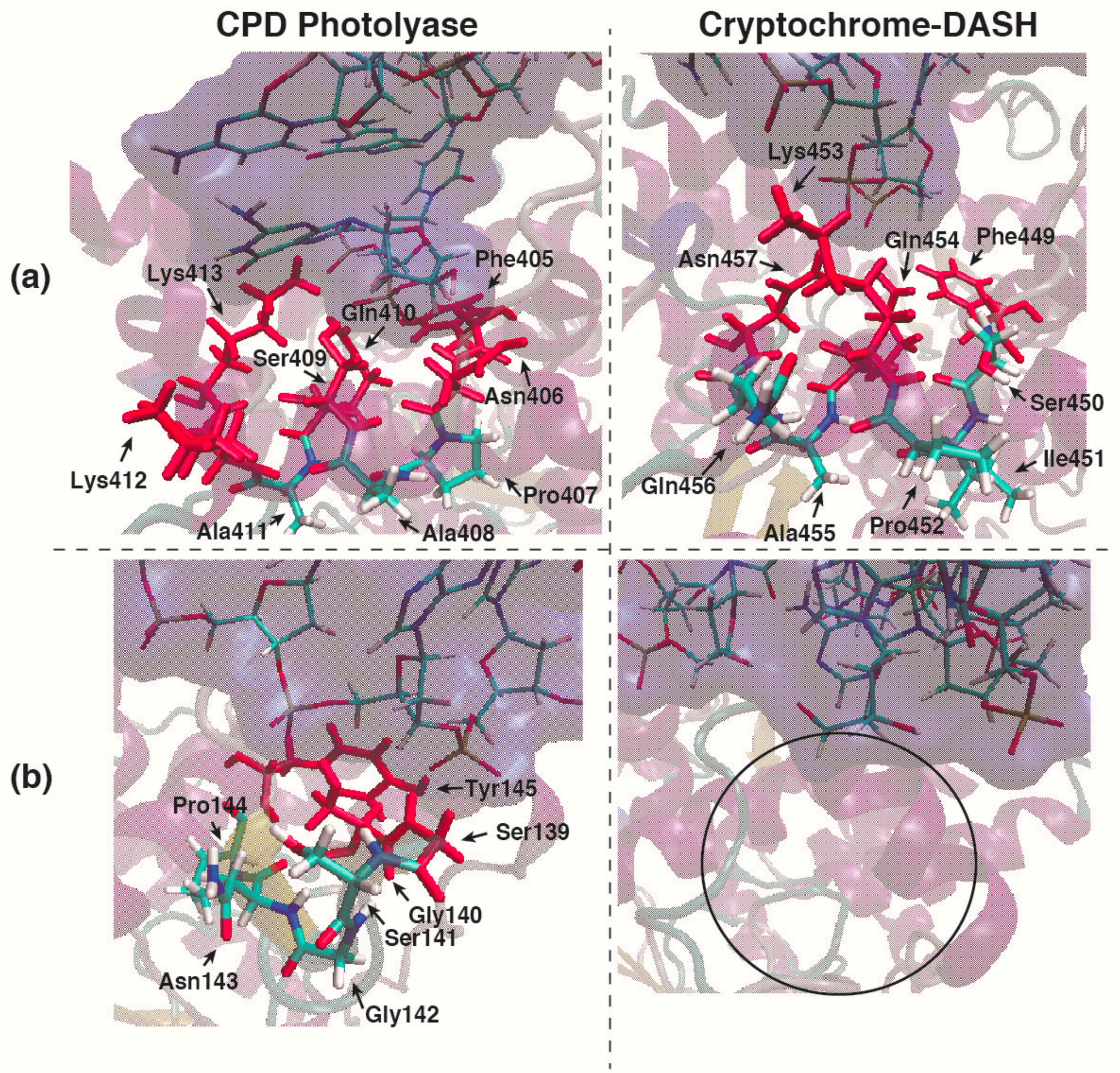
Supplementary Figure S4. Root mean square deviation (RMSD) measured from a initial structure for each system. Here, RMSD is calculated for last 50 ns of total 150 ns MD simulation. For CPD-PHR (left), averaged total RMSD, protein part and DNA part of RMSDs are 1.98±0.09 Å, 1.75±0.07 Å, and 3.25±0.32 Å, respectively. For CRY-DASH (right), those are 3.15±0.08 Å, 2.63±0.05 Å, and 5.12±0.24 Å, respectively.



Supplementary Figure S5. Root mean square deviation (RMSD) measured from an average structure for each system. Average total RMSDs of CPD-PHR (left) and CRY-DASH (right) are 1.32±0.01 Å and 1.37±0.08 Å, respectively.



Supplementary Figure S6. Aspect of position that interact with DNA in protein surface. Red, green, and black circles represent an α helix commonly found in the both proteins, a loop part in CPD-PHR, and a missing loop part in CRY-DASH compared with CPD-PHR. Blue surface represents duplex DNA.



Supplementary Figure S7. Amino acid residues contributing to the DNA binding located at ptotein surface. (a), the amino acid residues of red circle part in Figure S6 for CPD-PHR are Phe405, Asn406, Pro407, Ala408, Ser409, Gln410, Ala411, Lys412, and Lys413, while they for CRY-DASH are Phe449, Ser450, Ile451, Pro452, Lys453, Gln454, Ala455, Gln456, and Asn457. (b) Those of green circle part in Figure S6 for CPD-PHR are Ser139, Gly140, Ser141, Gly142, Asn143, Pro144, and Tyr145, while CRY-DASH does not have the loop moiety (black circle). The red color licorices mean the amino acid residues with the partial binding free energy lower than -1.0 kcal mol^{-1}.

Supplementary References

[S1] Ohue, M., Matsuzaki, Y., Uchikoga, N., Ishida, T. & Akiyama, Y. MEGADOCK: An all-to-all protein-protein interaction prediction system using tertiary structure data. *Protein Pept. Lett.* **21**, 766–778 (2014).

[S2] Ohue. M., Shimoda, T., Suzuki, S., Matsuzaki, Y., Ishida, T. & Akiyama, Y. MEGADOCK4.0: an ultra-high-performance protein-protein docking software for heterogeneous supercomputers. *Bioinformatics* **30**, 3281–3283 (2014).

[S3] Brooks, B. R. Bruccoleri, R. E. Olafson, B. D. States, D. J. Swaminathan, S. & Karplus, M. CHARMM: a program for macro-molecular energy, minimization, and dynamics calculations. *J. Comput. Chem*. **4**, 187–217 (1983).