**<Supplementary Information>**

**Exploring ligand binding pathways on proteins using hypersound–accelerated molecular dynamics**

**Mitsugu Araki1,\*, Shigeyuki Matsumoto2, Gert-Jan Bekker3, Yuta Isaka4, Yukari Sagae1, Narutoshi Kamiya5, and Yasushi Okuno1,2,\***

1 Graduate School of Medicine, Kyoto University, 53 Shogoin-Kawaharacho, Sakyo-ku, Kyoto 606-8507, Japan

2 Medical Sciences Innovation Hub Program, RIKEN Cluster for Science, Technology and Innovation Hub, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama City, Kanagawa 230-0045, Japan

3 Institute for Protein Research, Osaka University, 3-2 Yamadaoka, Suita, Osaka 565-0871, Japan

4 Research and Development Group for *In Silico* Drug Discovery, Center for Cluster Development and Coordination (CCD), Foundation for Biomedical Research and Innovation at Kobe (FBRI), 6-3-5, Minatojima-minamimachi, Chuo-ku, Kobe, Hyogo 650-0047, Japan

5 Graduate School of Simulation Studies, University of Hyogo, 7-1-28 Minatojima-Minamimachi, Chuo-ku, Kobe, Hyogo 650-0047, Japan

**\*Corresponding authors:**

Yasushi Okuno

Graduate School of Medicine, Kyoto University, 53 Shogoin-Kawaharacho, Sakyo-ku, Kyoto 606-8507, Japan. Phone: +81-75-751-4881, Email: [okuno.yasushi.4c@kyoto-u.ac.jp](mailto:okuno.yasushi.4c@kyoto-u.ac.jp)

Mitsugu Araki

Graduate School of Medicine, Kyoto University, 53 Shogoin-Kawaharacho, Sakyo-ku, Kyoto 606-8507, Japan. Phone: +81-75-751-4881, Email: [araki.mitsugu.6w@kyoto-u.ac.jp](mailto:araki.mitsugu.6w@kyoto-u.ac.jp)

**Contents**

**1 Materials and Methods ······································································· 4**

**1.1 Model systems and force fields** **·························································· 4**

**1.2 Modeling of shock waves ···································································5**

**1.3 MD simulations ············································································· 5**

**1.4 Analysis of MD simulations of liquid water ·········································· 6**

**1.5 Analysis of ligand binding within different CDK2 pockets ························ 8**

**1.6 Advanced analysis of CS3 and CS242 binding to the ATP pocket of CDK2 ·· 8**

**1.7 Estimation of kinetic parameters for the CDK2-ligand binding process ······· 9**

**1.8 Effects of increasing the solvent or solvent/ligand temperature on the probability of observing the ligand binding event ······································ 11**

**1.9 Identification of specific ligand binding sites on the CDK2 surface ··············13**

**2 Supplementary Figures ··································································· 14**

**3 Supplementary Tables ···································································· 33**

**4 Supplementary Movies ··································································· 38**

**5 References ··················································································· 39**

**1. Materials and Methods**

**1.1 Model systems and force fields**

We modeled the binding of CDK2 to two ATP-competitive inhibitors, CS3 and CS242, and two allosteric inhibitors, 2AN and 9YZ. The initial structural data of human CDK2 were obtained from the Protein Data Bank (PDB) and the Community Structure-Activity Resource (CSAR) (<http://www.csardock.org>) databases 1. Based on cocrystal structures (PDB IDs: 4EK5 (CS3), 4FKQ (CS242), 3PXF (2AN), and 5OO0 (9YZ)), disordered loops and flexible side chains were modeled and refined using the structure preparation module in the MOE program 2, and the dominant protonation state at pH 7.0 was assigned to titratable residues. Considering that a high concentration of ligands enhances the probability of capturing the protein-ligand binding 3, 50 ligands were randomly placed around the protein and away from the binding site (> 17 Å) by translating the ligand in the bound crystal structure.

The ligands were protonated to give net charges of 0 (CS3, CS242, and 9YZ) or -1 (2AN), reflecting the dominant protonation states at neutral pH. GAMESS was used to optimize the structure of each ligand and calculate its electrostatic potential at the HF/6-31G\* level 4, after which the atomic partial charges were obtained *via* the restrained electrostatic potential (RESP) approach 5. The other potential parameters of the ligands were obtained by the general AMBER force field (GAFF) 6 using the antechamber module of AMBER Tools 12. The AMBER ff99SB-ILDN force field 7 was used for the protein and ions, while water was modeled with the TIP3P potential 8. Approximately 18,000 water molecules were placed around the protein model in an 8.4×8.4×8.4 nm3 cubic box. In addition, approximately 60 sodium and chloride ions (corresponding to 150 mM NaCl) were introduced into the simulation box to neutralize all systems, except for the CDK2-2AN complex, for which the NaCl concentration was decreased to 10 mM because of the high concentration of the charged ligand. Based on the volume of the simulation box (592.7 nm3) and the number of ligands (50), the ligand concentration was calculated to be 138 mM, which is much higher than the typical concentrations used in biochemical assays; however, the enhanced ligand diffusion by hypersound irradiation indicates that aggregation of ligand molecules is successfully prevented (Table 1). For the liquid water system, a total of 20,068 water molecules were included in an 8.5×8.5×8.5 nm3 cubic box.

**1.2 Modeling of shock waves**

The isotropic hypersound irradiation of the solute was modeled by generating six different shock waves sequentially propagating from each face of the cubic simulation box (X0, Y0, Z0, X1, Y1, and Z1) toward its center (Fig. 1A, top). Six shock waves were sequentially irradiated in the +X, +Y, +Z, -X, -Y, and -Z directions, and a delay (corresponding to the *T*int interval) was applied between each series of shock waves to prevent temperature increase. Each shock wave consisted of five cycles of 16*N* time steps: 80 velocity pulses (= 16×5 cycles) were applied every *N* MD step (Fig. 1A, bottom). In each pulse, hypersound-induced velocities are defined as:

*vi*= *vmax*×cos(2π×*m* / 16*N*) (for *i* = X0, Y0, Z0, m = 0, *N*, 2*N*…)

*vi*= -*vmax*×cos(2π×*m* / 16*N*) (for *i* = X1, Y1, Z1, m = 0, *N*, 2*N*…) ,

where *vmax* is the maximum velocity assigned to the pulse and *m* is the time step number, were added to the thermal velocities of the water molecules located within 1 nm of each surface to model locally originated shock waves. The parameters for shock waves applied to the liquid water and solvated protein-ligand systems are reported in Supplementary Section 1.3 and Supplementary Table 2, respectively. A modified version of the GROMACS 4.6.5 program 9 was used to model the shock waves.

**1.3 MD simulations**

MD simulations with periodic boundary conditions were carried out using the GROMACS 4.6.5 program on the K computer, Cybermedia Center at Osaka University, and Global Scientific Information and Computing Center at Tokyo Institute of Technology (Japan). Electrostatic interactions were calculated using the particle mesh Ewald (PME) method 10 with a cutoff radius of 10 Å, unless stated otherwise; van der Waals interactions were cut off at 10 Å. The P-LINCS algorithm was employed to constrain all bond lengths at their equilibrium value of 11. After energy minimization, each system was equilibrated as described in the following subsections. A time step of 2 fs was used in all MD runs.

1. Liquid water

The system was equilibrated for 1 ns in a constant number of molecules, volume, and temperature (NVT) ensemble. Production runs were also conducted in the NVT ensemble. Electrostatic interactions were cut off at 11 Å. Production runs of 5 ns were performed with and without hypersound irradiation. The *N*, *vmax*, and *Tint* parameters in the hypersound-perturbed MD simulations were set to 50, 400 m s-1, and 2,400N, respectively. The cooling effect of Nose-Hoover 12 13, stochastic velocity rescaling 14, and Berendsen 15 thermostats on the hypersound-perturbed MD simulation was examined, showing that all the thermostats with a time constant of 0.1 or 0.3 ps rapidly decreased the excess energy and the elevation of the total kinetic energy returned to the baseline level before the next shock wave pulse (Supplementary Fig. 5). The mass density, pressure, and kinetic energy of the system were analyzed using MD trajectories obtained with a Nose-Hoover thermostat with a time constant of 0.3 ps and calculated using the coordinates and velocities saved every 2 fs.

1. Protein-ligand systems

Each system was equilibrated for 100 ps under NVT conditions, followed by an MD run of 100 ps in a constant number of molecules, pressure, and temperature (NPT) ensemble, with positional restraints applied on protein heavy atoms. Production runs were then conducted under NPT conditions without positional restraints. The temperature was maintained at 298 K by stochastic velocity rescaling 14, and a Parrinello-Rahman barostat was used to maintain the pressure at 1 bar 16. The temperature and pressure time constants were set to 0.1 and 2 ps, respectively. A total of 283, 369, 100, and 100 independent production runs of 100 ns (with different atomic velocities) were performed for the CDK2-CS3, CDK2-CS242, CDK2-2AN, and CDK2-9YZ systems, respectively. In addition, 1,137 (CS3), 362 (CS242), 100 (2AN), and 100 (9YZ) production runs were performed under hypersound irradiation using the parameters summarized in Supplementary Table 2.

**1.4 Analysis of MD simulations of liquid water**

The mass density, pressure, and kinetic energy in the hypersound-perturbed MD simulations of liquid water were estimated by focusing on wave propagation along the X direction, as described below.

The mass density was estimated at 82 different X-points, based on the number of molecules located within ±0.2 nm of each point. The kinetic energy (*kx*) was calculated as , where *M* is the mass of a water molecule and *<vx>* is the X component of the velocity, averaged over all water molecules located within ±0.2 nm from the corresponding X-point. Under hypersound irradiation, *kx* was estimated to be 0.4–0.5 kcal/mol at the center of the simulation box (X = 4 nm, Fig. 1D). The instantaneous temperature in this region was estimated to be 400–500 K, based on the *kx* value of bulk water at 300 K (~ 0.3 kcal/mol, corresponding to *RT*/2).

The pressure of water in the +X direction of the cubic simulation box was estimated from the X components of the velocities of the water molecules that crossed the YZ plane at a given X during the observation time *Δt*, according to the modified van der Waals equation for liquid systems:

[1]

where *m* is the mass of a water molecule, *S* is the area of the YZ plane, *vxi* is the X component of the velocity of the *i*-th water molecule, *a* is the intermolecular attractive force constant (determined as described below), *Na* is Avogadro’s number, and *Vm* is the molar volume, which was calculated to be 0.0183 L mol-1 based on the volume of the simulation box (8.53 nm3), and the number of water molecules contained in it (20,068). We initially performed a conventional MD simulation of 50 ps, and the first term of equation [1], , was calculated to be 1.298×108 Pa based on the water molecules that crossed the YZ plane at X = 2 nm (corresponding to the mid-point between the origin and the center of the simulation box) during a *Δt* interval of 50 ps. Using the saturated vapor pressure of water at 298 K (*P* = 0.032×105 Pa), the *a* parameter was estimated to be 0.423 (atm L2 mol-2). The pressure under hypersound irradiation was then determined from the hypersound-perturbed MD trajectory, using the estimated *a* value and the sum of the *vx* values of the water molecules that crossed the YZ plane at each selected X point during a *Δt* interval of 0.4 ps.

**1.5 Analysis of ligand binding within different CDK2 pockets**

For each ligand, we analyzed the MD trajectories of the system containing the CDK2 protein and 50 ligand molecules. Ligand binding within individual CDK2 sites (ATP pocket, allosteric site 1, and allosteric site 2) was considered to occur if at least two distances between an atom belonging to the protein pocket (see below) and any ligand heavy atom were below 5 Å. The following atoms of the protein pocket were used in the distance calculation: Val18 (beta carbon, Cβ) and Leu134 (gamma carbon, Cγ) for the ATP pocket, Tyr15 (zeta carbon, Cζ) and Leu55 (gamma carbon, Cγ) for allosteric site 1, and Cys177 (gamma carbon, Cγ) and Trp227 (indole nitrogen, Nε) for allosteric site 2. Advanced analysis of CS3 and CS242 binding to the ATP pocket is described in the Supplementary section 1.6.

**1.6 Advanced analysis of CS3 and CS242 binding to the ATP pocket of CDK2**

For the ATP-competitive inhibitors (CS3 and CS242), whose experimental binding structures and *kon* values are available, the occurrence of a binding event to the ATP pocket was assessed using stricter criteria, as follows. First, we identified trajectories that satisfied two conditions: (1) a distance between Val18 Cβ and any ligand heavy atom ≦ 5 Å and (2) the RMSD of the ligand from the crystallographic pose below 9 Å. Next, entry into the ATP pocket was confirmed by visual inspection of these trajectories, using the VMD software 17. Finally, we identified 67 (CS3) and 14 (CS242) MD trajectories that captured binding events.

In approximately half of these MD trajectories, the bound state was unstable, and the ligand separated from the ATP pocket within 1–40 ns. However, in the remaining trajectories, the ligand remained stably bound to the protein until the end of the simulation; these trajectories were thus extended to 200 ns to further examine the behavior of the bound ligands.

Principal component and conformational clustering analyses of the ligand binding poses observed in representative 27 (CS3) and 14 (CS242) MD trajectories were performed as follows: after removing the overall translation and rotation of the protein, the covariance matrix was calculated using the Cartesian coordinates of the ligand and diagonalized to obtain the PC eigenvectors. The first three principal components (PC1, PC2, and PC3) accounted for 40%, 33%, and 23% of the variance for CS3, respectively, while PC1, PC2, and PC3 accounted for 42%, 29%, and 20% of the variance for CS242, respectively. Conformational clustering of the binding poses into an optimal number of clusters was then performed on the first three PCs (PC1–PC3) using the X-means clustering method 18. The bound states of CS3 and CS242 on the ATP pocket were grouped into 10 and 7 conformational clusters, respectively, one of which corresponded to the crystallographic pose 1, indicating that some of these binding conformations are commonly observed in the 27 (CS3) and 14 (CS242) trajectories.

**1.7 Estimation of kinetic parameters for the CDK2-ligand binding process**

The association rate constant under hypersound irradiation (*kon*), activation energy (*E*), diffusion constant of the solute (*D*), steric factor (*P*), frequency factor (*A*), and effective temperature under hypersound irradiation (*T*) were estimated as follows, using the experimental *kon* values measured without any perturbation and the trajectories obtained from conventional and hypersound-perturbed MD simulations.

The kinetics of the binding between protein (P) and ligand (L) were analyzed according to the following reaction scheme:

where PL is the protein-ligand complex.

The second-order reaction rate is defined as:

[2]

where [P], [L], and [PL] are the concentrations of the protein, ligand, and protein–ligand complex, respectively. The initial binding rate is proportional to the initial concentrations of P and L ([P]0 and [L]0, respectively). If [P] ≈ [P]0 and [L] ≈ [L]0, the following relation can be derived by solving equation [2]:

[3]

The [P]0 and [L]0 values in the present simulations of the CDK2-ligand binding were 2.8 and 138 mM, respectively. Based on the experimentally determined *kon* values of CS3 and CS242 [3.35 × 105 and 3.21 × 104 M-1 s-1, respectively 1], the fractions of the CDK2-ligand complexes after 100 ns were expected to be 0.46% (CS3) and 0.044% (CS242). The probabilities of observing the stable ligand binding event in the 100-ns conventional MD simulations of CS3 and CS242 were 0.4% (= 1/283) and 0.3% (= 1/369) (Supplementary Table 2), respectively. Under hypersound irradiation with *N* = 50 steps, *vmax* = 400 m/s, and *Tint* = 2,400 N, which are the parameters predominantly used in the simulations (Supplementary Table 2), and using [PL]/[P]0 ratios of 9/177 (CS3) or 6/227 (CS242), corresponding to the proportions of MD trajectories that exhibited stable ligand binding (Supplementary Table 2), the *kon* values were estimated to be 3.68 × 106 (CS3) and 1.92 × 106 M-1 s-1 (CS242).

The *kon* constant can also be described using the Arrhenius equation:

[4]

where *R* is the gas constant. To estimate *E*, the potential energy and free energy differences between the unbound state and the highest-energy transition state were averaged over the 9 (CS3) and 6 (CS242) trajectories. The *E* values estimated from potential energy trajectories were 3.9 ± 1.8 (CS3) and 6.7 ± 2.4 (CS242) kcal mol-1 (p=0.02, one-sided Student *t*-test), while those estimated from free energy trajectories produced from the free energy landscapes (Supplementary Fig. 7) were ‒0.71 ± 0.23 (CS3) and ‒0.42 ± 0.18 (CS242) kcal mol-1 (p=0.01, one-sided Student *t*-test). According to a kinetic model involving a “doorway state” located between the unbound and bound states, the frequency factor can be approximated by the diffusion-controlled rate constant 19:

[5]

where *NA*, *DP* (*DL*), *R\**, and *P* are Avogadro’s number, the diffusion constant of the protein (ligand), the critical protein-ligand distance, and the steric factor, respectively. In this study, *R\** was set to 1 nm, and we assumed *DP* << *DL*. To calculate *DL*, the mean-square displacement of the 50 ligands during an MD simulation of the solvated CDK2-ligand system was averaged over 10 independent simulations. The diffusion constants (*DL\_conv*) of CS3 and CS242 estimated from the conventional MD simulations were 0.17 ± 0.05 × 10-5 and 0.19 ± 0.07 × 10-5 cm2 s-1, respectively, while those estimated from the MD runs under hypersound irradiation with *N* = 50 steps, *v*max = 400 m/s, and *Tint* = 2,400 N (*DL\_hyper*) were 0.61 ± 0.16 × 10-5 (CS3) and 0.30 ± 0.10 × 10-5 cm2 s-1 (CS242). Using the *DL\_conv* and experimental *kon* values along with the *E* parameter estimated from the potential energy difference in equations [4] and [5], the steric factors of CS3 and CS242 were calculated as 10-0.76 and 100.20, respectively. According to equation [5], the frequency factors (*A*) without hypersound irradiation were calculated to be 108.35±0.13 M-1 s-1 (CS3) and 109.36±0.17 M-1 s-1 (CS242), while those obtained under hypersound irradiation were 108.91±0.12 M-1 s-1 (CS3) and 109.56±0.15 M-1 s-1 (CS242). Finally, the effective temperatures under hypersound irradiation calculated from equation [4] were 362 K for CS3 and 445 K for CS242.

**1.8 Effects of increasing the solvent or solvent/ligand temperature on the probability of observing the ligand binding event**

To assess how enhancing the thermal motions of the solvent or ligand molecules affects the probability of observing the ligand binding event, we performed a conventional MD protocol in which the water or ligand diffusion coefficients were adjusted to the values observed in the hypersound-perturbed MD simulation.

The diffusion coefficient of the water molecules in the solvated CDK2-ligand system was estimated to be 4.7 ± 0.1 × 10-5 cm2/s (conventional MDs at 298 K) or 5.5 ± 0.1 × 10-5 cm2/s (hypersound-perturbed MDs with *N* = 50 steps, *v*max = 400 m/s, and *Tint* = 2,400 N). The increased water diffusion coefficient was obtained using a conventional MD protocol in which the temperature of the solvent was increased to 309 K while that of the protein and ligands was maintained at 298 K. This protocol may be the closest to the hypersound-perturbed MD simulation method, since additional velocities were only applied to solvent molecules. However, the diffusion coefficients of the ligands remained 0.18 ± 0.07 × 10-5 cm2/s (CS3) and 0.18 ± 0.05 × 10-5 cm2/s (CS242), which are almost equivalent to those estimated from the conventional MD simulations (*i.e.,* 0.17 ± 0.05 × 10-5 cm2/s for CS3 and 0.19 ± 0.07 × 10-5 cm2/s for CS242 (Table 1)). In addition, the probability of observing the ligand-binding event in this type of conventional MD simulation was estimated to be only 2% (2/100, corresponding to 2 out of 100 MD runs resulting in binding) for CS3 and 3% (3/100) for CS242, which are significantly lower than that in the hypersound-perturbed MD simulations (12.4% for CS3 and 4.8% for CS242 (Supplementary Table 2)).

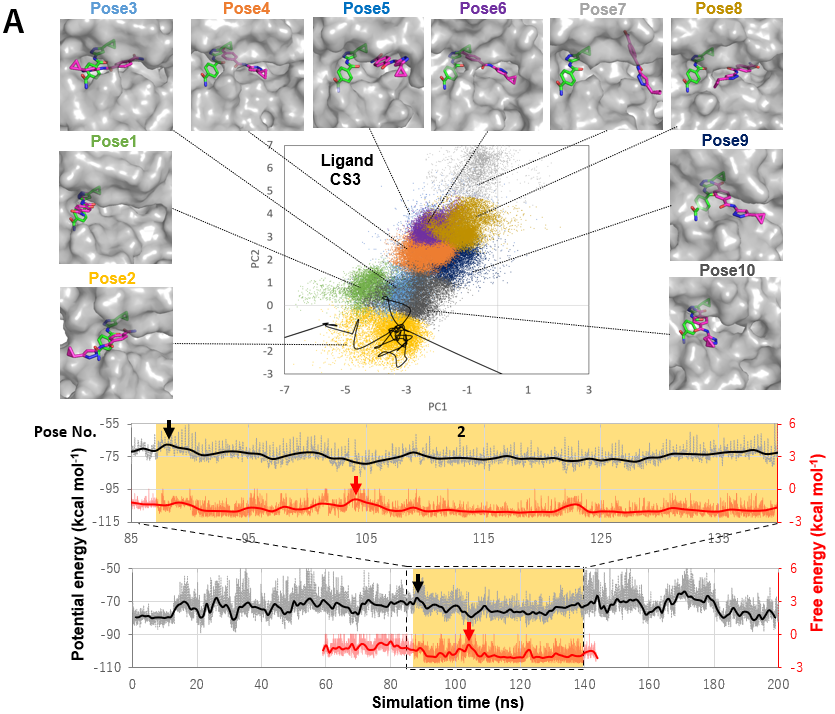
The diffusion coefficients of CS3 and CS242 estimated from the MD runs under hypersound irradiation with the same parameters described above were 0.61 ± 0.16 × 10–5 cm2/s (CS3) and 0.30 ± 0.10 × 10–5 cm2/s (CS242) (Table 1). The diffusion coefficients close to these values were obtained using a conventional MD protocol in which the temperature of the ligands and solvent was increased to 375 K (CS3) or 355 K (CS242), while that of the protein was maintained at 298K. The probability of observing the ligand-binding event in this type of conventional MD simulation was estimated to be 14% (14/100) for CS3 and 4% (4/100) for CS242, which is almost equivalent to that in the hypersound-perturbed MD simulations. However, this type of simulation (different temperatures of protein and ligand/solvent) is unrealistic, and also induces a partial collapse of rigidly structured regions in CDK2 (Supplementary Fig. 6), presumably because of the excessively increased diffusion coefficient of water molecules (11.6 ± 0.1 × 10–5 cm2/s at 375 K or 9.5 ± 0.1 × 10–5 cm2/s at 355 K). On the other hand, when the ligands and solvent were coupled separately to temperature baths at 375 K/355 K and 309 K, respectively, the diffusion coefficients of the ligands remained 0.21 ± 0.09 × 10-5 cm2/s (CS3) and 0.22 ± 0.05 × 10-5 cm2/s (CS242), suggesting that the water and ligand diffusion coefficients measured in the hypersound-perturbed MD simulation cannot be reproduced by the use of a separate temperature bath for the ligands and solvent. These results demonstrate the distinct effects of hypersound irradiation and temperature increase.

**1.9 Identification of specific ligand binding sites on the CDK2 surface**

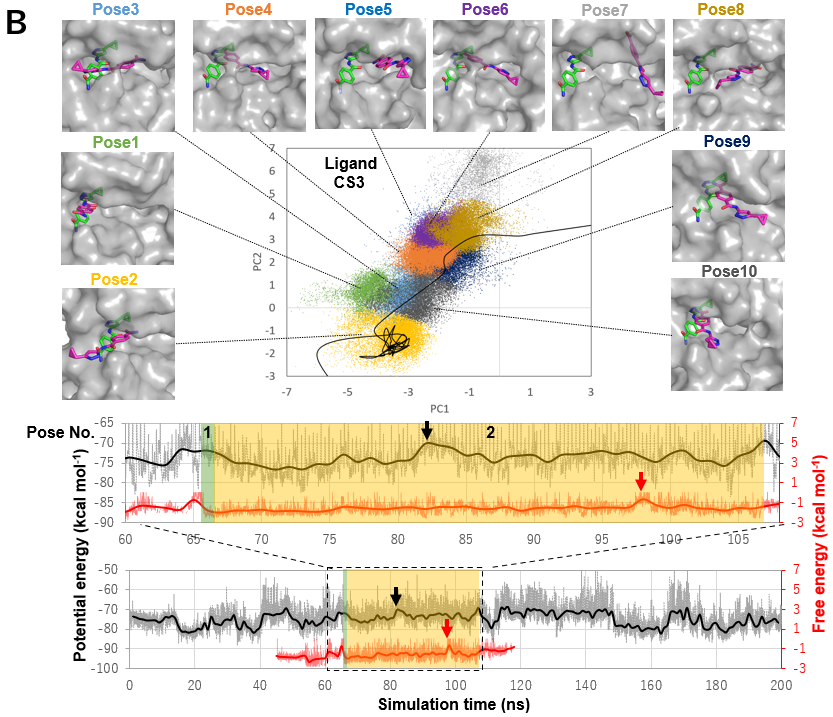
The specific binding sites of each of the ATP-competitive inhibitors (CS3 and CS242) and allosteric inhibitors (2AN and 9YZ) on the CDK2 surface were determined as follows. First, the root-mean-square fluctuation (RMSF) of the ligand was calculated every 10 ns of the individual 100-ns hypersound-perturbed MD trajectories obtained with *N* = 50 steps, *Tint* = 2,400 N, and *v*max = 400 m/s (CS3, CS242, and 9YZ) or *v*max = 300 m/s (2AN). If the value was below 3 Å, a stable CDK2-ligand complex was considered to be formed during the 10-ns period, and residues that interact with the ligand (<5 Å) were extracted from the mean coordinates of the protein and ligand. Next, the frequency of ligand interactions at each CDK2 residue (*fint*) was calculated across all stable complex structures and normalized by the number of MD trajectories. Finally, after excluding residues that frequently interacted with all ligands (*fint* of more than 0.1) as nonspecific binding sites, residues with higher *fint* values were identified as specific binding sites.

**2. Supplementary Figures**

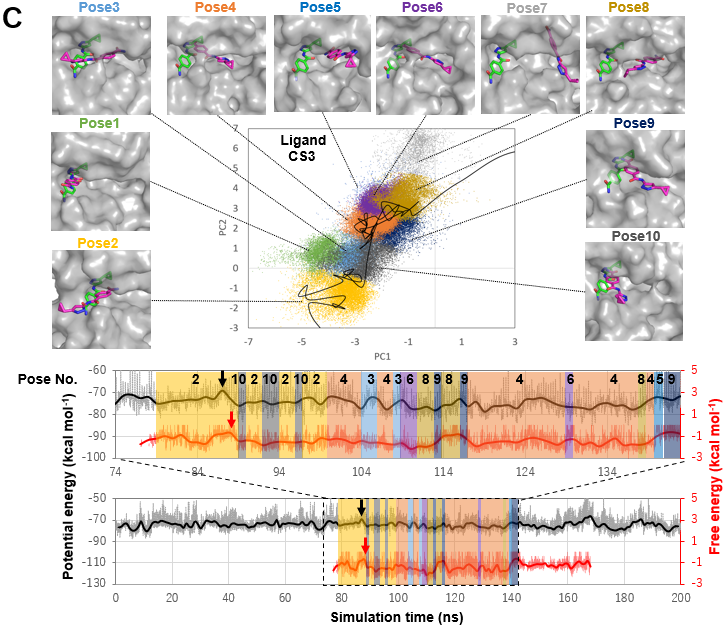
**Supplementary Figure 1**

****

**Supplementary Figure 1 (Continued)**

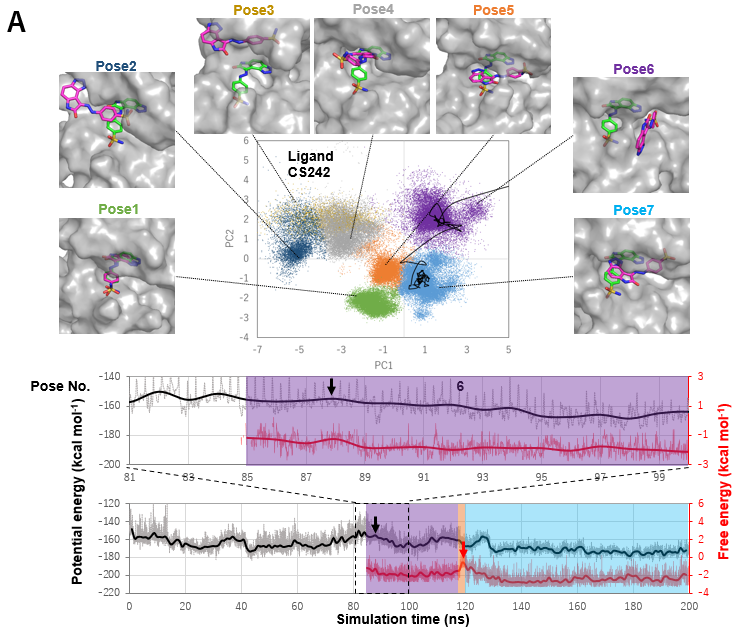


**Supplementary Figure 1 (Continued)**

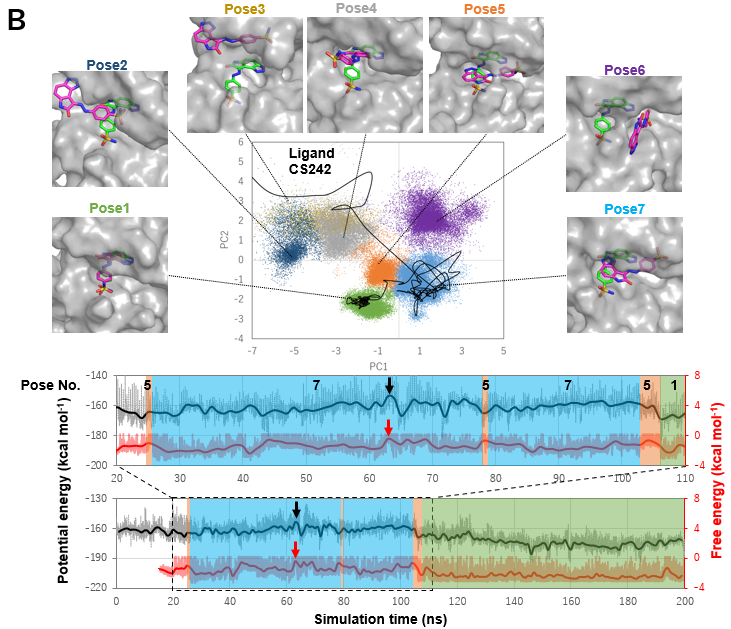


**Supplementary Figure 1.** Three representative binding pathways of the CS3 ligand to the ATP-binding pocket of CDK2. In pathway (A), the transition state occurs upon entry into the CDK2 pocket; in (B), the transition state is reached during conformational rearrangement in the pocket interior, whereas both ligand binding and unbinding are observed in pathway (C). (top) Projections of binding conformations observed in the whole set of MD trajectories (colored dots) and of a representative binding pathway (black line) onto the first and second principal components (PC1 and PC2) calculated from PCA (see the Supplementary section 1.6). Ten representative binding poses (magenta sticks) on CDK2 (gray surfaces) are shown along with the crystallographic pose (green sticks), the closest conformation to which was designated as Pose 1. (bottom) Potential energy (black) and free energy (red) trajectories corresponding to the pathway shown in the PCA map. The potential energy was calculated as the sum of the intraligand and intermolecular (protein-ligand and ligand-solvent) contributions. The free energy trajectory was produced from the free energy landscape with respect to (PC1, PC2, and PC3) (Supplementary Fig. 7). The highest-energy transition state is indicated by a black (potential energy) or red (free energy) arrow. An enlarged view of these trajectories close to the highest-energy transition state is also shown in the panel above the whole trajectories. Time intervals in which ligand binding was observed are highlighted in the same color as that used for the binding conformation in the top panel. In 5 out of the 9 binding pathways observed in the hypersound-perturbed MD simulations with *N* = 50 steps, *vmax* = 400 m/s, and *Tint* = 2,400 N, the binding pose assigned to the highest-energy transition state is the same between potential energy and free energy trajectories.

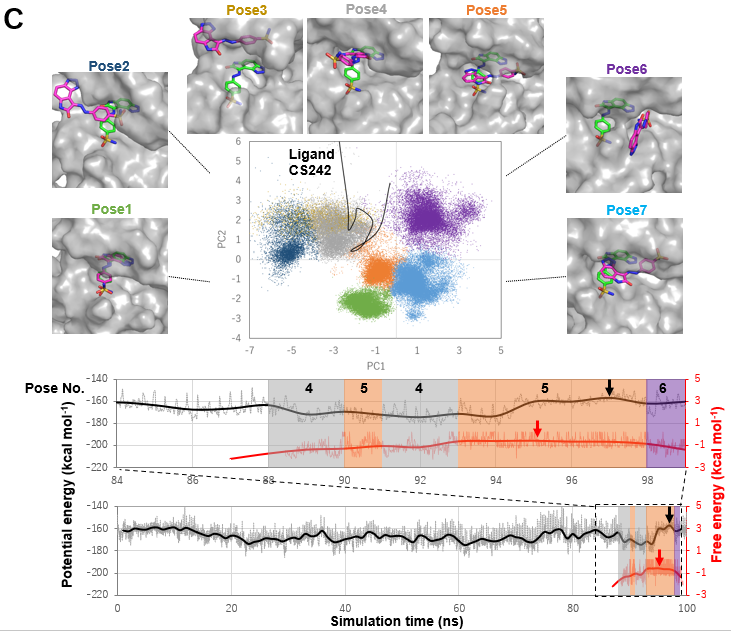
**Supplementary Figure 2**

****

**Supplementary Figure 2 (Continued)**

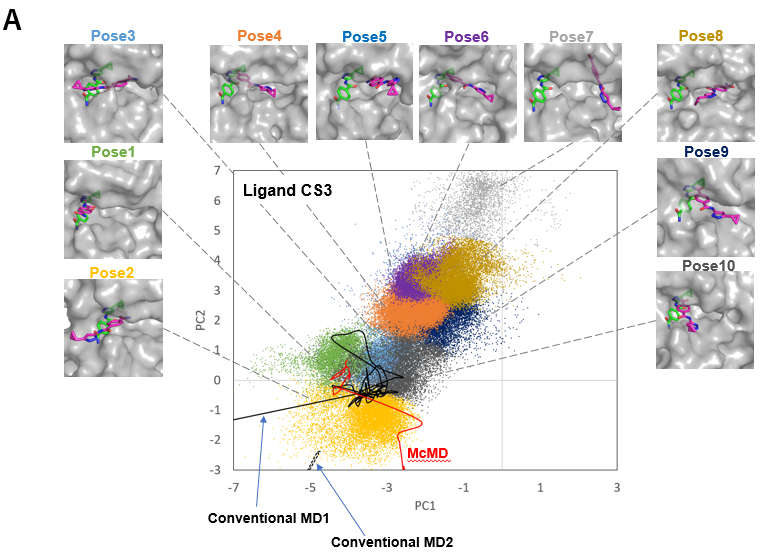
****

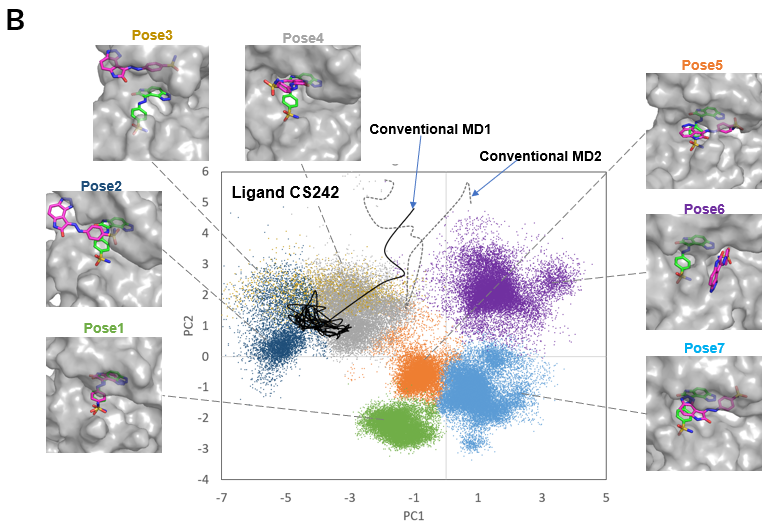
**Supplementary Figure 2 (Continued)**



**Supplementary Figure 2.** Three representative binding pathways of the CS242 ligand to the ATP-binding pocket of CDK2. In pathway (A), the transition state occurs upon entry into the CDK2 pocket; in (B), the transition state is reached during conformational rearrangement in the pocket interior, whereas both ligand binding and unbinding are observed in pathway (C). (top) Projections of binding conformations observed in the whole set of MD trajectories (colored dots) and of a representative binding pathway (black line) onto the first and second principal components (PC1 and PC2) calculated from PCA (see the Supplementary section 1.6). Seven representative binding poses (magenta sticks) on CDK2 (gray surfaces) are shown along with the crystallographic pose (green sticks), the closest conformation to which was designated as Pose 1. (bottom) Potential energy (black) and free energy (red) trajectories corresponding to the pathway shown in the PCA map. The potential energy was calculated as the sum of the intraligand and intermolecular (protein-ligand and ligand-solvent) contributions. The free energy trajectory was produced from the free energy landscape with respect to (PC1, PC2, and PC3) (Supplementary Fig. 7). The highest-energy transition state is indicated by a black (potential energy) or red (free energy) arrow. An enlarged view of these trajectories close to the highest-energy transition state is also shown in the panel above the whole trajectories. Time intervals in which ligand binding was observed are highlighted in the same color as that used for the binding conformation in the top panel. In 4 out of the 6 binding pathways observed in the hypersound-perturbed MD simulations with *N* = 50 steps, *vmax* = 400 m/s, and *Tint* = 2,400 N, the binding pose assigned to the highest-energy transition state is the same between potential energy and free energy trajectories.

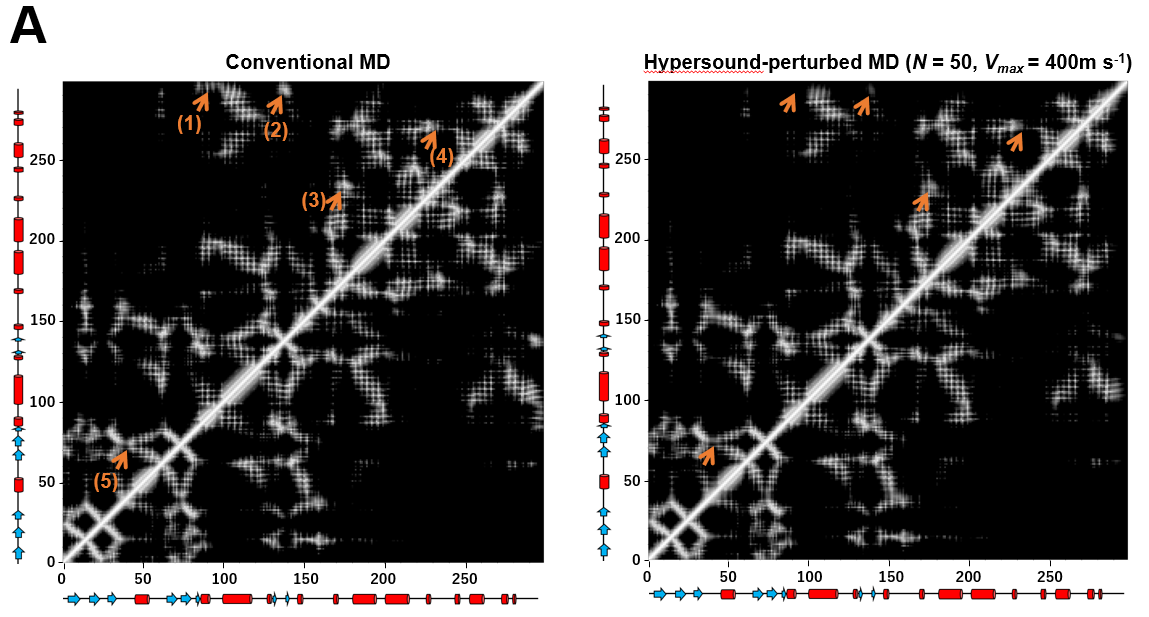
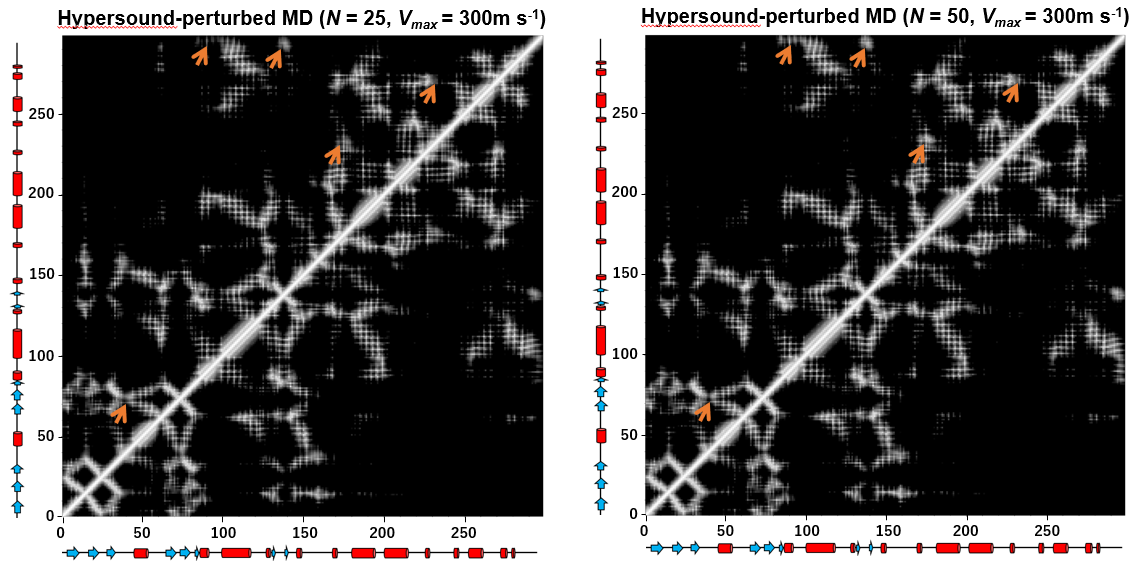
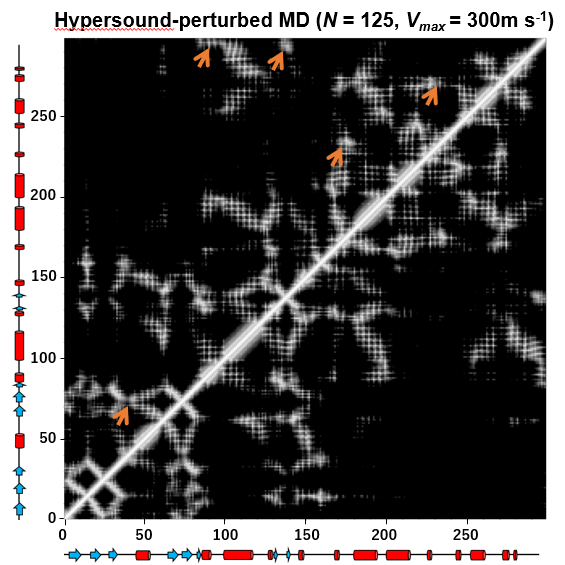
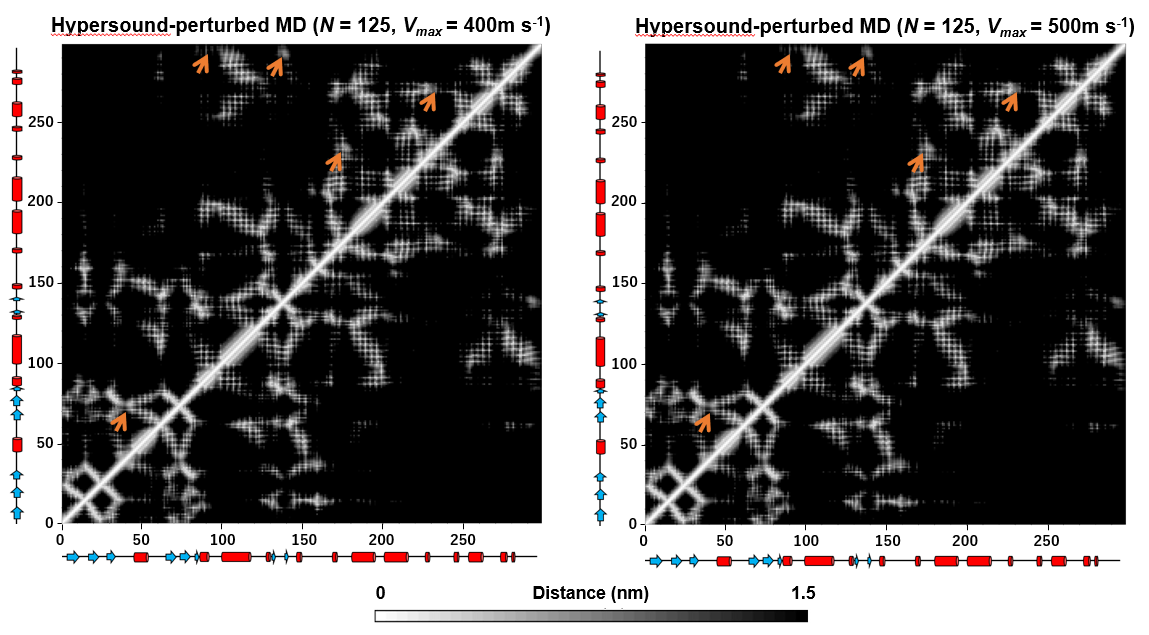
**Supplementary Figure 3**



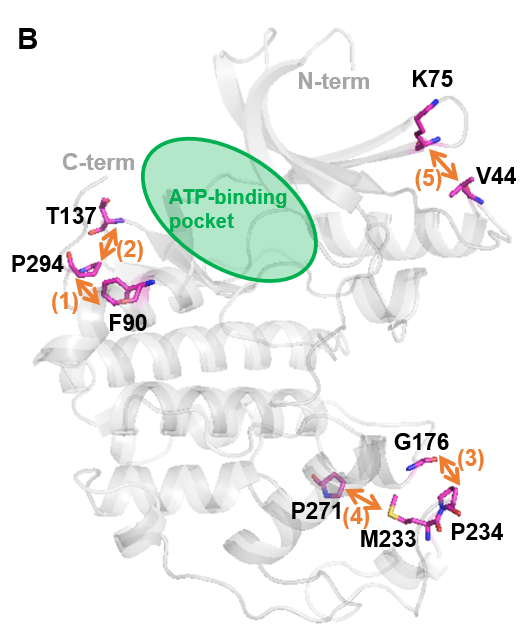


**Supplementary Figure 3.** Binding pathways of the CS3 (A) and CS242 (B) ligands to the ATP-binding pocket of CDK2 captured by conventional and multicanonical MD (McMD, another type of advanced MD simulations used to efficiently explore protein conformational space 20. Projections of binding conformations observed in the whole set of simulations (colored dots) and of binding pathways (lines) projected onto the first and second principal components (PC1 and PC2) calculated from PCA (see the Supplementary section 1.6). Two binding pathways (conventional MD1 and MD2) observed in conventional MD simulations are indicated by black solid and dotted lines, while a CS3 binding pathway predicted by McMD 21 is indicated by a red line. Ten (CS3) or seven (CS242) representative binding poses (magenta) are shown along with the crystallographic pose (green), the closest conformation to which was designated as Pose 1.

**Supplementary Figure 4**

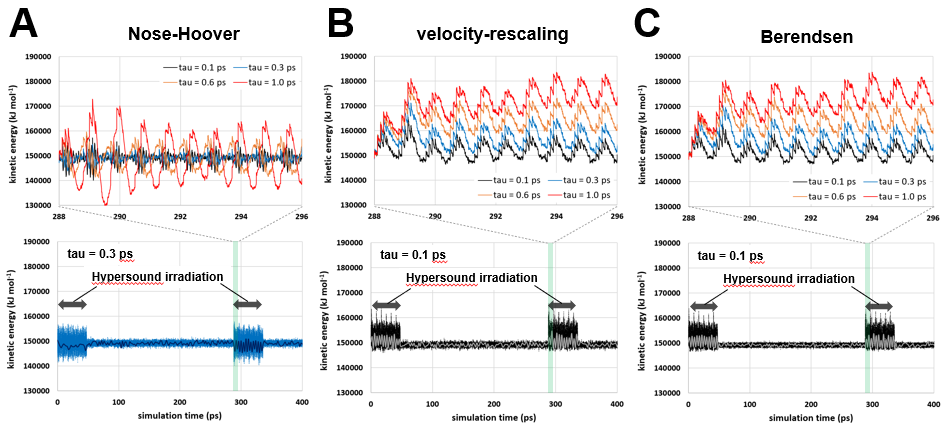
****

**Supplementary Figure 4 (Continued)**

****

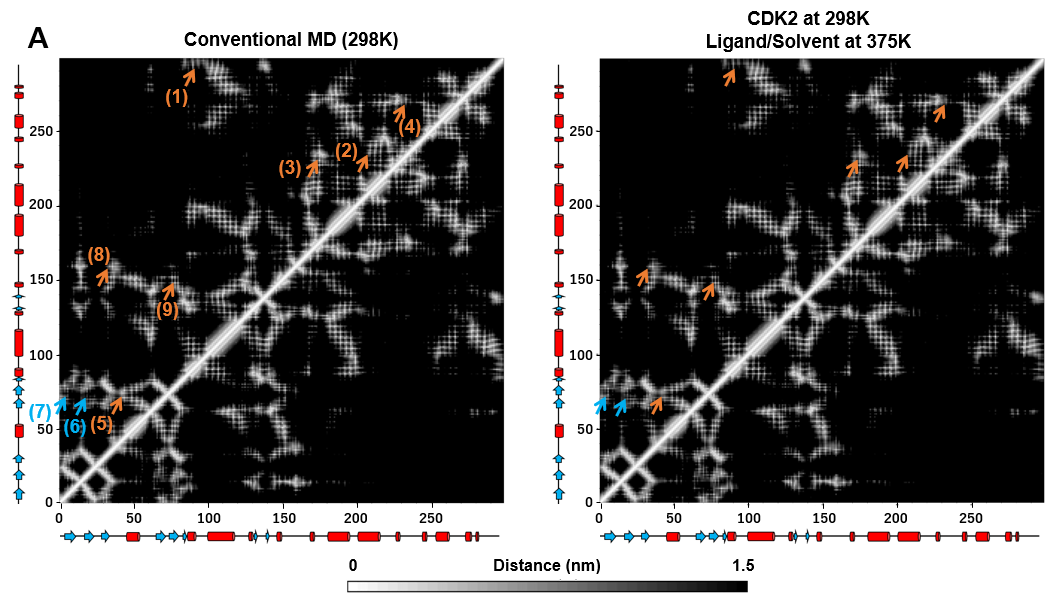
**Supplementary Figure 4.** Effect of hypersound shock waves on the native structure of the CDK2 kinase. (A) Native contact maps of CDK2 in the presence of CS3, obtained from conventional and hypersound-perturbed MD simulations. The shortest distances between residue pairs were determined using trajectories of 50–100 ns extracted from 10 independent MD simulations of 100 ns without or with hypersound irradiation, where the *N* and *vmax* values used are indicated at the top of each figure; *Tint* of 2,400N was used across experiments. The inter-residue contacts whose intensities were attenuated by an increase in hypersound frequency (proportional to 1/*N*) or amplitude (*vmax*) are indicated by orange arrows. (B) Structural location of CDK2 residues affected by hypersound irradiation. Inter-residue contacts corresponding to the numbered arrows in (A) are shown in the native structure of the CDK2 kinase (PDBID: 4EK5), suggesting that hypersound shock waves only perturbed interactions involving the C-terminus or flexible loop regions.

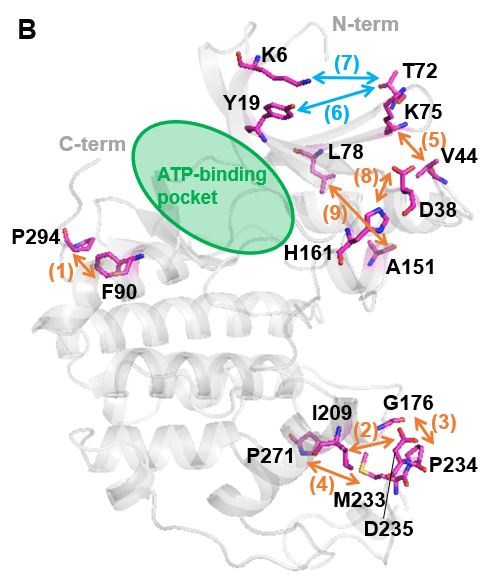
**Supplementary Figure 5**



**Supplementary Figure 5.** Relaxation of the hypersound-induced excess energy using (A) Nose-Hoover, (B) stochastic velocity rescaling, and (C) Berendsen thermostats. Bottom: The total kinetic energy of the liquid water system is plotted every 2 fs (thin lines), and smoothed by a window average of 2 ps (thick lines). A time constant for temperature coupling (t) is indicated in each panel. The total kinetic energy averaged across the intervals in which hypersound shock waves were generated (indicated by arrows) corresponds to 296 K, 301 K, and 301 K for the plots for Nose-Hoover, stochastic velocity rescaling, and Berendsen thermostats, respectively. Top: An enlarged view of a region ranging from 288 to 296 ps, in which the first shock wave (in the +X direction) in the second series of shock waves was generated, is also shown with the kinetic energy trajectories with different t values.

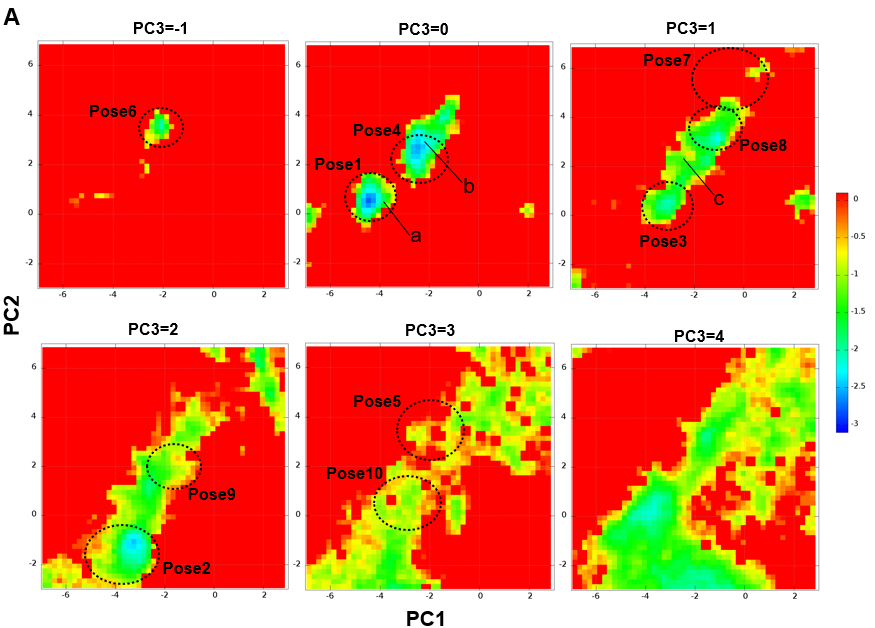
**Supplementary Figure 6**

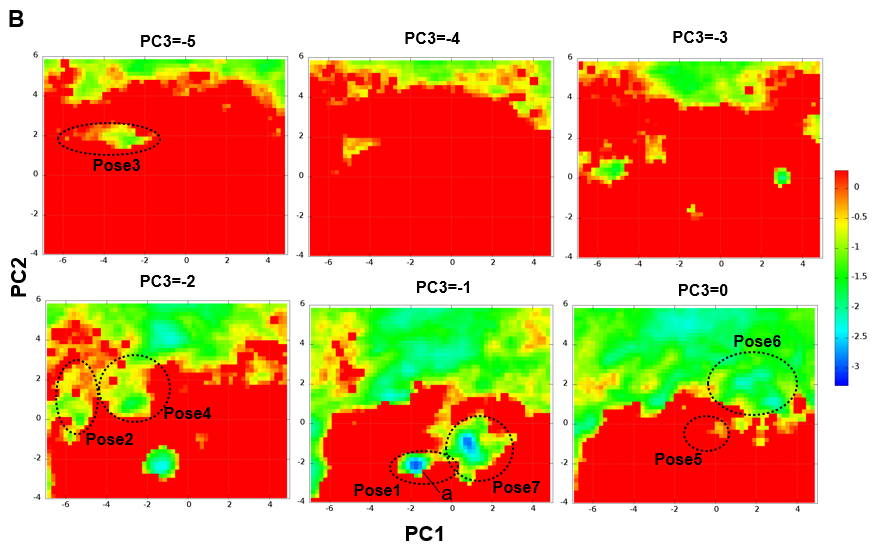




**Supplementary Figure 6.** Effects of the high ligand/solvent temperature on the native structure of the CDK2 kinase. (A) Native contact maps of CDK2 in the presence of CS3, obtained from conventional MDs at 298 K and high temperature MDs in which the temperature of the ligands and solvent was increased to 375 K while that of the protein was maintained at 298 K. The shortest distances between residue pairs were determined using trajectories of 50–100 ns extracted from 10 independent 100-ns MD simulations. Inter-residue contacts whose intensities were attenuated with an increase in the ligand/solvent temperature are indicated by orange arrows. Non-native contacts generated in the high temperature MD simulation are indicated by cyan arrows. (B) Structural location of CDK2 residues affected by the high temperature MD simulation. Inter-residue contacts corresponding to the numbered arrows in (A) are shown in the native structure of the CDK2 kinase (PDBID: 4EK5), suggesting that the high ligand/solvent temperature not only perturbed interactions involving the C-terminus or flexible loop regions but also induced a partial collapse of rigidly structured regions in the N-lobe of CDK2.

**Supplementary Figure 7**





**Supplementary Figure 7.** The free energy landscapes of (A) CS3 and (B) CS242 binding to the ATP binding pocket of CDK2, which are described as a function of the first three principal components (PC1, PC2, and PC3) calculated from PCA (the Supplementary section 1.6), and displayed as 2D slices at different PC3 values. Each landscape was calculated from the normalized probability distribution, according to G(PC1, PC2, PC3) = ‒RT ln (P(PC1, PC2, PC3) / Punbound), where R is the gas constant, T is the absolute temperature, and P is the probability density estimated from the hypersound-perturbed MD simulations with *N* = 50 steps, *vmax* = 400 m/s, and *Tint* = 2,400 N. Punbound is the probability density of the unbound state, and is set to an average density across regions of the PC space in which the ligand is away from the CDK2 surface (>5Å). The 10 (CS3) and 7 (CS242) representative binding poses, which correspond to those shown in Fig. 2A and 2B and Supplementary Fig. 1 and 2 are marked by dotted circles. The locations a-c on the landscape correspond to the crystallographic pose (a) and metastable binding poses estimated by McMD (b and c), which correspond to local minima b and c on Fig. S10A of 21, respectively. The closest conformational cluster to the crystallographic pose (Pose 1) as well as its immediately-preceding state (Pose 4 for CS3 and Pose 7 for CS242) along observed binding pathways (Fig. 2) are successfully assigned to the most stable binding poses. Also, metastable binding poses estimated from the McMD are captured also by the hypersound-perturbed MD simulation, suggesting its high sampling efficiency for ligand binding conformations. The free energy trajectory corresponding to each of 9 (CS3) and 6 (CS242) binding pathways observed in these simulations was produced from the free energy landscapes (Fig. 2A and Supplementary Fig. 1 (CS3)) (Fig. 2B and Supplementary Fig. 2 (CS242)).

**3. Supplementary Tables**

**Supplementary Table 1.** Calculated thermodynamic properties of liquid water

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Simulation conditions | Temperature (K) | Diffusion constant (×10-5 cm2 s-1) | Kinetic energy (kcal mol-1) | Potential energy  (kcal mol-1) | Total energy  (kcal mol-1) |
| Conventional MD at 298 K | 298.0 ± 1.2 | 5.79 ± 0.02 | 1.78 ± 0.01 | -9.61 ± 0.01 | -7.83 ± 0.01 |
| Hypersound-perturbed MD  at 298 K a | 295.5 ± 0.1 (297.5 ± 2.2) | 6.30 ± 0.10  (5.97 ± 0.06) | 1.76 ± 0.00  (1.78 ± 0.01) | -9.77 ± 0.01  (-9.64 ± 0.07) | -8.00 ± 0.01  (-7.86 ± 0.08) |
| Conventional MD at 305 K | 305.0 ± 1.3 | 6.40 ± 0.07 | 1.82 ± 0.01 | -9.53 ± 0.01 | -7.71 ± 0.01 |
| Conventional MD at 313 K | 313.0 ± 0.1 | 7.00 ± 0.02 | 1.87 ± 0.01 | -9.44 ± 0.01 | -7.57 ± 0.01 |
| Conventional MD at 328 K | 328.0 ± 1.3 | 8.19 ± 0.02 | 1.96 ± 0.01 | -9.28 ± 0.01 | -7.33 ± 0.01 |

a The thermodynamic parameters were calculated using the 0–0.048, 0.288–0.336, 0.576–0.624, 0.864–0.912, 1.152–1.200, 1.440–1.488, 1.728–1.776, 2.016–2.064, 2.304–2.352, 2.592–2.640, 2.880–2.928, 3.168–3.216, 3.456–3.504, 3.744–3.792, 4.032–4.080, 4.320–4.368, 4.608–4.656, and 4.896–4.944 ns trajectories (corresponding to the intervals in which hypersound shock waves were generated), extracted from a 5-ns MD simulation. Averages calculated over the whole 5-ns trajectory, which includes time intervals between shock wave generation (*T*int of 240 ps in Fig. 1A), are indicated in parentheses.

**Supplementary Table 2.** Summary of simulations of ligand binding to the ATP pocket of CDK2

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ligand | method | number of 100-ns simulations | *N* a | *vmax*  (m s-1) a | *Tint* a | number (percentage) of binding events | number of stable binding events b |
| CS3 | Conventional MD | 283 | - | - | - | 2 (0.7%) | 1 |
| Hypersound-perturbed MD | 180 | 25 | 300 | 2,400N | 22 | 7 |
| 200 | 50 | 300 | 2,400N | 6 | 1 |
| 177 | 50 | 400 | 2,400N | 22 (12.4%) | 9 |
| 180 | 125 | 300 | 2,400N | 0 | 0 |
| 200 | 125 | 400 | 2,400N | 3 | 0 |
| 20 | 125 | 500 | 1,440N | 3 | 2 |
| 180 | 125 | 500 | 2,400N | 11 | 5 |
| total: 1,137 |  |  |  | 67 | 24 |
| CS242 | Conventional MD | 369 | - | - | - | 2 (0.5%) | 1 |
| Hypersound-perturbed MD | 64 | 50 | 300 | 2,400N | 2 | 1 |
| 227 | 50 | 400 | 2,400N | 11 (4.8%) | 6 |
| 8 | 125 | 500 | 480N | 0 | 0 |
| 59 | 125 | 500 | 1,440N | 1 | 0 |
| 4 | 125 | 700 | 2,400N | 0 | 0 |
| total: 362 |  |  |  | 14 | 7 |
| 2AN | Conventional MD | 100 | - | - | - | 2 (2.0%) | 0 |
| Hypersound-perturbed MD | 100 | 50 | 300 | 2,400N | 8 (8.0%) | 1 |
| 9YZ | Conventional MD | 100 | - | - | - | 7 (7.0%) | 2 |
| Hypersound-perturbed MD | 100 | 50 | 400 | 2,400N | 21 (21.0%) | 1 |

a The *N*, *vmax*, and *Tint* parameters are defined in Fig. 1A. *N* values of 25, 50, and 125 correspond to hypersound frequencies of 1250, 625, and 250 GHz, respectively.

b Number of MD trajectories in which the formed protein-ligand complex remained stable until the end of the simulation (100 ns).

**Supplementary Table 3.** Hypersound parameter dependence of probabilities of observing CS3 binding to the ATP pocket of CDK2a

|  |  |  |  |
| --- | --- | --- | --- |
|  | *vmax* = 300 (m s-1) | *vmax* = 400 (m s-1) | *vmax* = 500 (m s-1) |
| *N =* 25 | 12.2% (22/180) b | N/A c | N/A c |
| *N =* 50 | 3.0% (6/200) | 12.4% (22/177) | N/A c |
| *N =* 125 | 0.0% (0/180) b | 1.5% (3/200) | 6.1% (11/180) |

a The probabilities are extracted from Supplementary Table 2 to show their *N* or *vmax* dependences, where *Tint* is fixed at 2,400 N. The number of binding events out of the total number of 100-ns MD runs are indicated in parentheses.

b Computation speeds of hypersound-perturbed MDs with (*N* = 25 steps and *v*max = 300 m/s) and (*N* = 125 steps and *v*max = 300 m/s) were 40.6 ± 2.6 and 42.4 ± 2.3. ns/day, respectively, while that of conventional MDs was 45.7 ± 0.9. These MD simulations were performed using seven OpenMP threads on a high-performance computing infrastructure equipped with Intel(R) Xeon(R) CPU E5-2680 v4 and NVIDIA Tesla P100 GPGPUs. The computation speed was estimated from ten independent simulations.

c N/A indicates that the probabilities could not be estimated because hypersound-perturbed MD simulations crashed because of the high frequency and/or amplitude of the shock waves.

**Supplementary Table 4.** Probabilities of observing ligand binding within different CDK2 pockets in conventional and hypersound-perturbed MD simulationsa

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Ligand and simulation type | ATP-binding site | Allosteric site 1 | Allosteric site 2 | Number of 100-ns simulations |
| CS3 |  |  |  |  |
| Conventional MD | 0.7% (2) | 0.0% (0) | 39.2% (111) | 283 |
| Hypersound-perturbed MD b | 12.4% (22) | 2.3% (4) | 87.6% (155) | 177 |
| CS242 |  |  |  |  |
| Conventional MD | 0.5% (2) | 0.0% (0) | 32.2% (119) | 369 |
| Hypersound-perturbed MD b | 4.8% (11) | 0.9% (2) | 57.7% (131) | 227 |
| 2AN |  |  |  |  |
| Conventional MD | 2.0% (2) | 36.0% (36) | 99.0% (99) | 100 |
| Hypersound-perturbed MD c | 8.0% (8) | 39.0% (39) | 100.0% (100) | 100 |
| 9YZ |  |  |  |  |
| Conventional MD | 7.0% (7) | 0.0% (0) | 95.0% (95) | 100 |
| Hypersound-perturbed MD b | 21.0% (21) | 4.0% (4) | 100.0% (100) | 100 |

a The number of binding events is noted in parentheses. The maximum number of binding events per trajectory was set to one, even if multiple events were observed.

b Probabilities were estimated from hypersound-perturbed MDs with *N* = 50 steps, *Tint* = 2,400 N, and *v*max = 400 m/s.

c Probabilities were estimated from hypersound-perturbed MDs with *N* = 50 steps, *Tint* = 2,400 N, and *v*max = 300 m/s.

**4. Supplementary Movies**

**Supplementary Movie 1.** Hypersound-perturbed MD simulation (50 ps) of liquid water. Hypersound shock waves were sequentially generated from each of the X0, Y0, Z0, X1, Y1, and Z1 surfaces every 8 ps. Water molecules are displayed as red sticks.

**Supplementary Movie 2.** MD simulation of CDK2-CS3 binding (120 of 200 ns) under hypersound irradiation. CDK2 and CS3 are displayed as surface and stick models, respectively. The corresponding potential energy [combining intraligand and intermolecular (protein-ligand + ligand-solvent) components] trajectory is also shown. The last 80 ns of the trajectory were not included in the movie because no significant conformational changes occurred in the ligand.

**Supplementary Movie 3.** MD simulation of CDK2-CS242 binding (120 of 200 ns) under hypersound irradiation. CDK2 and CS242 are represented as surface and stick models, respectively. The corresponding potential energy trajectory [combining intraligand and intermolecular (protein-ligand + ligand-solvent) components] is also shown. The last 80 ns of the trajectory were omitted because no significant conformational changes occurred in the ligand.

**5. References**

1. Dunbar J. B., Jr.*, et al.* CSAR benchmark exercise of 2010: selection of the protein-ligand complexes. *J. Chem. Inf. Model.* 2011, **51**(9)**:** 2036-2046.

2. Molecular Operating Environment (MOE). 2016.08 ed. Chemical Computing Group Inc., 1010 Sherbrooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7; 2016.

3. Takemura K., Sato C.& Kitao A. ColDock: Concentrated Ligand Docking with All-Atom Molecular Dynamics Simulation. *J. Phys. Chem. B* 2018, **122**(29)**:** 7191-7200.

4. Schmidt M. W.*, et al.* General atomic and molecular electronic structure system. *J. Comput. Chem.* 1993, **14**(11)**:** 1347-1363.

5. Bayly C. I., Cieplak P., Cornell W.& Kollman P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* 1993, **97**(40)**:** 10269-10280.

6. Wang J., Wolf R. M., Caldwell J. W., Kollman P. A.& Case D. A. Development and testing of a general amber force field. *J. Comput. Chem.* 2004, **25**(9)**:** 1157-1174.

7. Lindorff-Larsen K.*, et al.* Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* 2010, **78**(8)**:** 1950-1958.

8. Jorgensen W. L., Chandrasekhar J., Madura J. D., Impey R. W.& Klein M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 1983, **79**(2)**:** 926-935.

9. Hess B., Kutzner C., van der Spoel D.& Lindahl E. GROMACS 4:  Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* 2008, **4**(3)**:** 435-447.

10. Darden T., York D.& Pedersen L. Particle mesh Ewald: An N⋅log(N) method for Ewald sums in large systems. *J. Chem. Phys.* 1993, **98**(12)**:** 10089-10092.

11. Hess B., Bekker H., Berendsen H. J. C.& Fraaije J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* 1997, **18**(12)**:** 1463-1472.

12. Nose S. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* 1984, **52:** 255-268.

13. Hoover W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* 1985, **31:** 1695-1697.

14. Bussi G., Donadio D.& Parrinello M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* 2007, **126**(1)**:** 014101.

15. Berendsen H. J. C., Postma J. P. M., Vangunsteren W. F., Dinola A.& Haak J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 1984, **81**(8)**:** 3684–3690.

16. Parrinello M.& Rahman A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* 1981, **52**(12)**:** 7182-7190.

17. Humphrey W., Dalke A.& Schulten K. VMD: visual molecular dynamics. *J. Mol. Graph.* 1996, **14**(1)**:** 33-38, 27-38.

18. Ishioka T. Extended K-means with an efficient estimation of the number of clusters. *Intelligent Data Engineering and Automated Learning — IDEAL 2000*, 2000, pp 17-22.

19. Mondal J., Friesner R. A.& Berne B. J. Role of Desolvation in Thermodynamics and Kinetics of Ligand Binding to a Kinase. *J. Chem. Theory. Comput.* 2014, **10**(12)**:** 5696-5705.

20. Nakajima N., Nakamura H.& Kidera A. Multicanonical ensemble generated by molecular dynamics simulation for enhanced conforma conformational sampling of peptides. *J. Phys. Chem. B* 1997, **101**(5)**:** 817-824.

21. Bekker G. J.*, et al.* Accurate Prediction of Complex Structure and Affinity for a Flexible Protein Receptor and Its Inhibitor. *J. Chem. Theory. Comput.* 2017, **13**(6)**:** 2389-2399.