calculated the binding stability of each. The binding mode in the crystal was identified as having the lowest binding energy. We conclude that our method can be used to distinguish the most plausible binding modes for interacting molecules.

## 2. Materials and methods

### 2.1. Thermodynamic cycle for free energy calculation

In calculating the binding free energy from a molecular simulation, the following thermodynamic cycle should be considered:

$$\begin{array}{ccc} \textit{Lysozyme+triNAG in vacuo} & \xrightarrow{\Delta G_v} & \textit{Complex in vacuo} \\ \downarrow \Delta\mu_{Lyz}+\Delta\mu_{NAG} & & \downarrow \Delta\mu_{Complex} \\ \textit{Lysozyme+triNAG in solution} & \xrightarrow{\Delta G_s} & \textit{Complex in solution} \end{array} \quad (1)$$

where $\Delta G_v$ and $\Delta G_s$ represent the binding free energy of lysozyme and triNAG *in vacuo* and in solution, respectively, and $\Delta\mu_{Lyz}$, $\Delta\mu_{NAG}$, and $\Delta\mu_{Complex}$ represent the solvation free energies (transfer energy from vacuum to solution) of lysozyme, triNAG, and their complex, respectively. The binding free energy of interest in this case, $\Delta G = \Delta G_s$, is given by

$$\Delta G = \Delta G_v + \Delta\mu_{Complex} - (\Delta\mu_{Lyz} + \Delta\mu_{NAG}), \quad (2)$$

$\Delta G_v$ depends only on the conformational energy and entropy of the solutes, as shown in Eq. (3):

$$\Delta G_v = E_{Complex} - (E_{Lyz} + E_{NAG}) - T\{S_{Complex} - (S_{Lyz} + S_{NAG})\}, \quad (3)$$

where $E_{Lyz}$, $E_{NAG}$, $E_{Complex}$, $S_{Lyz}$, $S_{NAG}$ and $S_{Complex}$ represent the conformational (potential) energies of lysozyme, triNAG, and their complex and the entropies of lysozyme, triNAG, and their complex *in vacuo*, respectively. Therefore, from Eqs. (2) and (3), the binding free energy in solution is obtained as,

$$\begin{aligned} \Delta G &= (E_{Complex} - TS_{Complex} + \Delta\mu_{Complex}) - (E_{Lyz} - TS_{Lyz} + \Delta\mu_{Lyz}) \\ &\quad - (E_{NAG} - TS_{NAG} + \Delta\mu_{NAG}) \\ &= \Delta E - T\Delta S + \Delta\Delta\mu. \end{aligned} \quad (4)$$

where,

$$\Delta E = E_{Complex} - (E_{Lyz} + E_{NAG}), \quad (5)$$

$$\Delta\Delta\mu = \Delta\mu_{Complex} - (\Delta\mu_{Lyz} + \Delta\mu_{NAG}), \quad (6)$$

$$\Delta S = S_{Complex} - (S_{Lyz} + S_{NAG}). \quad (7)$$

The conformational energy is directly calculated in MD simulation. Solvation free energy is calculated using the energy representation method (see next subsection). In each case, solute entropy can be decomposed into the contributions from translational, rotational and internal motion of the solute [18],

$$S = S^{trans} + S^{rot} + S^{int}, \quad (8)$$

which are defined by,

$$S^{trans} = k_B \ln\left\{\frac{(2\pi M k_B T)^{3/2}}{h^3} e^{5/2} V\right\}, \quad (9)$$

$$S^{rot} = k_B \ln\left\{\frac{(2\pi k_B T)^{3/2}(I_X I_Y I_Z)^{1/2}}{h^3} e^{3/2} 8\pi^2\right\}, \quad (10)$$

$$S^{int} = \sum_i k_B\left\{\frac{1}{\alpha_i(\exp\alpha_i^{-1} - 1)} - \ln(1 - \exp(-\alpha_i^{-1}))\right\}, \quad (11)$$

where $k_B$, $h$, $T$, $M$ and $V$ are Boltzmann constant, Planck constant, absolute temperature, mass of solute and volume (1 l/mol), respectively. $I_X$, $I_Y$ and $I_Z$ represent solute principal moments of inertia. $S^{int}$ is entropy contribution from internal motion of solute, which is similar to so-called configurational entropy obtained if probability distribution of internal motion of solute is approximated to be deduced from the covariance matrix of coordinates [19]. The configurational entropy contains configurational integral of potential term only, but $S^{int}$ also contains kinetic term. Eq. (11) is the formula obtained for a system undergoing quantum normal mode vibrations for each normal mode with angular frequency $\omega_i$, $\alpha_i = 2\pi k_B T/h\omega_i$. In this Letter, $\omega_i$ is obtained as the effective frequency of principal mode [20] determined by the principal component analysis of solute MD simulation *in vacuo* using all the solute atoms in the mass-weighted form. It should be noted that Eq. (11) can be understood as more generalized form of classical form in principle and that the classical form is limited to be valid only when $\alpha_i > 1$. If this condition is not satisfied, the classical form of $S^{int}$ can be negative, which is inacceptable from physical point of view as entropy should not be negative; the classical entropy of a harmonic oscillator diverges towards negative infinity as frequency increases. The present 'semi-quantum' treatment in which classically-determined angular frequency is used to the quantum form of entropy is considered to be reasonable [21–24]. The empirical form of the $S^{int}$ proposed by Schlitter [21,22] were also often used, which provides the value close to Eq. (11).

To obtain the value of $\Delta G$, we need to calculate all the energy terms in Eq. (4). However, to distinguish the most plausible binding modes from the model complex structures, we only need to compare the first term of Eq. (4), $(E_{Complex} - TS_{Complex} + \Delta\mu_{Complex})$, because the other terms depend on the monomer structures, which are common in all the cases.

### 2.2. All-atom computation of solvation free energy using the energy representation method

In the present Letter, a set of protein-ligand complex candidate structures generated by docking programs were subjected to all-atom MD simulation and analysis of the solvation free-energy. Calculating the solvation free energy with an all-atom force field using standard and numerically exact methods such as free-energy perturbation and thermodynamic integration is highly prohibitive for proteins, however [1]. We therefore employed an approximate method of free-energy computation, namely, the method of energy representation (ER) [6–8]. In the ER method, a set of energy distribution functions are obtained only from simulations of the solution system of interest and the pure solvent system, and these functions are then substituted into an approximate functional for the solvation free energy calculation. Since the method does not require simulations for intermediate states of solute insertion, the computational time required to determine the solvation free energy is reduced considerably compared to exact methods that require sampling a large number of intermediate states. The performance of the ER method was assessed in a previous paper, and the error in the functional was observed to be not larger than the error due to the use of a force field [9]. Other studies have shown that all-atom, free-energy analysis is feasible for protein molecules composed of up to a few hundred residues [25–27]. In the ER method the solvation free energy $\Delta\mu$ can be determined from the following equation:

$$\Delta\mu = \int d\varepsilon\, \varepsilon\rho^e(\varepsilon) - G[\rho^e, \rho_0^e, \chi_0^e], \quad (12)$$

where $\rho^e(\varepsilon)$ represents the average distribution (histogram) of the pair energy $\varepsilon$ between the solute (lysozyme-triNAG in the present Letter) and the solvent (water). $\rho_0^e(\varepsilon)$ and $\chi_0^e(\varepsilon, \varepsilon')$ represent the

density of the states of solute-solvent pair interactions and the solvent-solvent pair correlation, respectively. The first term in Eq. (12) denotes the average sum of the solute-solvent interaction energy in the solution system of interest, and the second term approximates the effect of solvent reorganization. $G$ is the functional of $\rho^e(\varepsilon)$, $\rho_0^e(\varepsilon)$ and $\chi_0^e(\varepsilon,\varepsilon')$. The explicit form of this free-energy functional is described elsewhere [7,28]. Note that ε denotes the pair energy between the solute and solvent, not the total sum of the solute-solvent energy. The solution system containing the solute of interest is simulated to determine $\rho^e(\varepsilon)$. An additional simulation of pure solvent must be conducted in order to determine $\rho_0^e(\varepsilon)$ and $\chi_0^e(\varepsilon,\varepsilon')$, which represent the density of states of solute-solvent pair interaction and the solvent-solvent pair correlation, respectively. In this case, the solute molecule is inserted into the pure solvent system as a test particle; the test particle is inserted without disturbing the solvent configuration. Overlap with the solvent is allowed in the test-particle insertion, and the overlapping configurations contribute to the solvation free energy as the excluded-volume term.

### 2.3. Simulation procedure

We examined four binding modes in the present Letter: the crystal structure of lysozyme-triNAG complex (PDB ID: 1HEW) [17] and three structures generated by distinct docking programs, AutoDock3 [14], GOLD [15] and Molegro Virtual Docker (MVD) [16]. The main purpose of the complex model structure generation was not to predict complex structure but to generate complex model structures different from the crystal binding mode. Docking calculations were conducted for the lysozyme monomer structure taken from the complex structure (1HEW), which should be regarded as a relatively easy setup for docking calculation because monomer structure should be employed for real prediction but the docked structure was used in this Letter. Compared to monomer lysozyme structure (PDB ID: 193L) [29], root-mean-square deviation of 1HEW backbone atoms is 0.37 Å, which is relatively small. The lysozyme–triNAG complex structure corresponding to the top-scored binding mode predicted by each docking program was selected from one set of docking simulation and examined in the following simulation.

Simulations were performed using the NAMD 2.9 package [30]. MD simulations were performed for lysozyme–triNAG complex structure *in vacuo* and in solution and pure solvent. In all solution cases and pure water case, the systems initially contained 24,635 TIP3P (CHARMM) model water in a rectangular box (8-nm × 8-nm × 12-nm) which is 2.5 times as large as the lysozyme–triNAG complex along all box axes. Before solvation, a 50000-step energy minimization of the solutes was performed using the conjugate gradient method. The simulated systems were modeled using the CHARM22 [31] with CMAP [32] (lysozyme) and CHARMM36 [33,34] (triNAG) force field. The equation of motion is integrated every 2 fs with velocity Verlet integrator with H-bond SHAKE/RATTLE. The long-range Coulomb energy was evaluated using the particle mesh Ewald method with 12-Å real-space cut off. Real-space contribution only was considered in MD simulations *in vacuo*. The pure solvent and solution systems were simulated under constant temperature (300 K) and pressure (1 atm) using Langevin thermostat and barostat. MD simulations *in vacuo* were simulated under constant temperature using Langevin thermostat.

Stability of the four complex structures in solution was first examined by relatively long MD simulation for 20 ns without positional restraints/constraints. As shown in Results, the crystal binding mode is very stable but the other complex forms drift significantly. To select the lowest energy complex, we conducted distinct MD simulations. Positional restraints with force constant 0.1 kcal/mol Å$^2$ were applied to all the heavy atoms of complex to avoid significant conformational drift from the initial structures. A 1-ns simulation in NPT was carried out for production after a 1-ns equilibration run. Atomic coordinates were stored at 100-fs intervals. The energy distribution function $\rho^e$ was calculated from the coordinates of the solution systems (see Eq. (12)). To calculate the energy distribution functions $\rho_0^e$ and $\chi_0^e$, MD simulation of a pure water solvent was also conducted. Following a 1-ns equilibration simulation, 100-ps sampling simulation was performed while coordinates were saved every 100 fs. Solute configurations used for solute insertion were taken from the MD simulations *in vacuo*. For this purpose a 1-ns MD simulation were performed after a 1-ns equilibration. Positional restraints with force constant 0.1 kcal/mol Å$^2$ were applied to all the heavy atoms of complex. To generate snapshots similar to structures in solution, the reference used in positional restraints was the same with that used in solution. Atomic coordinates were stored at 10-fs intervals. A solute insertion was conducted 1000 times for each solvent configuration. Insertions of a solute were performed using random translations. No random rotations for insertions were conducted in this Letter. The total number of solute insertions was performed up to 1 000 000 times.

The solvation free energy was calculated from the distribution functions as described above using the Energy Representation Module (ERMOD, http://sourceforge.net/projects/ermod/). Conformational energy and entropy of the solutes were calculated from the 1-ns MD trajectories *in vacuo* (see Eqs (8)-(11)). After elimination of overall translational and rotational contributions with best-fit to the average structure, all-atom principal component analysis was performed to determine the effective frequency of the principal modes from the 1-ns trajectory. High frequency modes corresponding to the number of the hydrogen and translational/rotational degrees of freedom were not used for entropy calculation with Eq. (11).

## 3. Results and discussions

### 3.1. Structures of complex model structures

Figure 1 shows structures of the four binding modes for the lysozyme–triNAG complex. As described above, in the crystal form triNAG binds to the A–B–C sites of lysozyme, but triNAG binds to the B–C–D sites in the GOLD and MVD docking models and to the C–D–E sites in the AutoDock model. These four binding modes were evaluated by AutoDock3, GOLD, and MVD with individual docking scores (Table 1). Note that AutoDock3 and MVD evaluate the complex structures with energy, which means lower value is better, whereas GOLD returns fitness score which is better for higher value. The ranks of the crystal structure out of 4 structures were the third in AutoDock3 and GOLD, and the forth in MVD. Therefore, the crystal structure was not judged to be good binding mode by the docking programs. It should be also noted that the best binding mode found by docking program could depend on parameter setting. The MVD binding mode was evaluated as the best also with AutoDock score, suggesting that this binding mode happened to be not examined in AutoDock docking because all the possible configurations are not necessarily generated. As far as in our experiences, docking prediction can often give reasonable results comparable to crystal structures. The present results suggest that prediction of lysozyme--triNAG complex is a relatively difficult problem for docking prediction. Since all the sites (A–F) should work as the binding sites for longer glycan, accurate energy calculation might be needed.

Stability of the four lysozyme–triNAG complex structures was examined by the 20-ns MD simulations. Time evolutions of root-mean-square-deviation (RMSD) of the complex structures from the initial structures are shown in Figure 2a. To calculate RMSD,

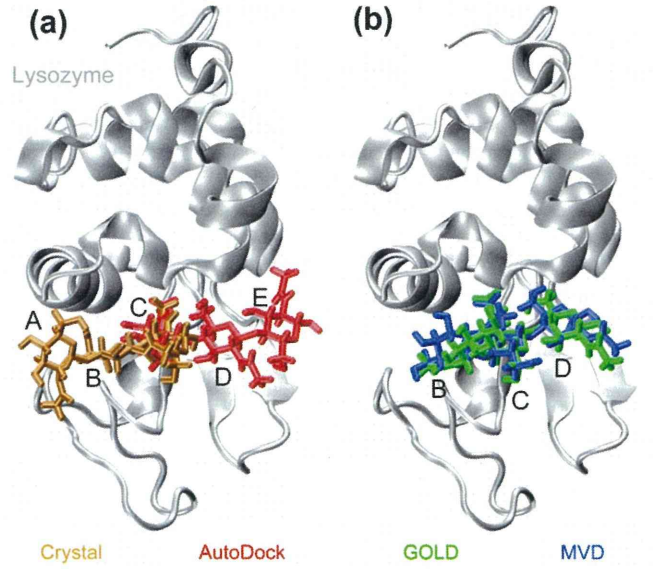


**Figure 1.** Model complex structures of hen egg-white lysozyme and triNAG. Structure of lysozyme (gray) and (a) TriNAG in the crystal (brown) and the binding modes predicted by AutoDock 3.0 (red), (b) GOLD (green) and Molegro Virtual Docker (MVD: blue). The symbols A–E represent the binding pockets for glucosamine. Structures were illustrated by VMD [35].

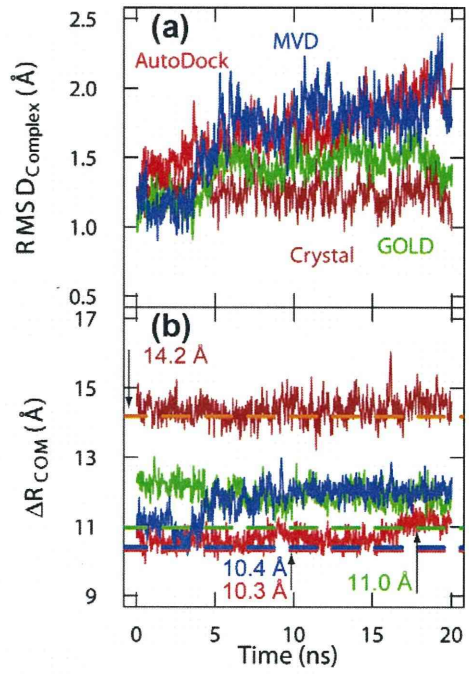


**Figure 2.** Time evolution of (a) root-mean-square-deviation (RMSD) of the complex structures from the initial structures and (b) center-of-mass distance between lysozyme and triNAG ($\Delta R_{COM}$). Broken lines represent $\Delta R_{COM}$ values for the initial structures (14.2, 10.3, 11.0, and 10.4 Å for Crystal, AutoDock, GOLD, and MVD, respectively).

**Table 1**
Docking scores for four lysozyme–triNAG complex models obtained by AutoDock, GOLD and MVD.

| Model | AutoDock[a] | GOLD[b] | MVD[c] |
|---|---|---|---|
| Crystal | −7.1 | 33.4 | −101.1 |
| AutoDock | −11.9 | 30.6 | −127.9 |
| GOLD | −5.1 | 54.2 | −114.4 |
| MVD | −13.1 | 48.9 | −182.0 |

[a] Docked energy evaluated by AutoDock3.05.
[b] Fitness of gold score calculated by GOLD5.1.
[c] Energy evaluated by MVD2.3.

the backbone heavy atoms of the protein and the atoms that form triNAG rings were superposed onto the initial structure in best-fit position (translation and orientation were adjusted to give the minimum RMSD value) for each snapshot. The RMSD value in MD simulations starting from the crystal structure (~1.2 Å) is small and very stable. In the other cases, the RMSD values were larger. The RMSD values for the AutoDock and MVD models continuously drifted during the MD simulation, suggesting that these two binding modes were not stable. Figure 2b shows the time evolution of center-of-mass distance between lysozyme and triNAG. TriNAG in the crystal form stayed in the distance around the initial structure and was stable during the 20-ns MD. In the other cases, triNAG positions significantly shifted from the initial structures. This suggests that binding free calculation with Eq. (4) for the crystal binding mode is expected to be performed reasonably but difficult for the other cases. Therefore, we conducted the calculation of $(E_{Complex} - TS_{Complex} + \Delta\mu_{Complex})$ with the weak positional restraints (see Section 2) to select the lowest energy binding mode in the next subsection.

### 3.2. Binding mode evaluation with $(E_{Complex} - TS_{Complex} + \Delta\mu_{Complex})$

Table 2 shows the results of energy analyses. The crystal binding mode was identified as the lowest energy structure in $(E_{Complex} - TS_{Complex} + \Delta\mu_{Complex})$. The solvation free energy for the crystal binding mode is the lowest, consistent to the fact that the first NAG relatively protrudes into the solvent in the crystal (binding to the A–B–C sites) as seen in Figure 1a. Solvent accessible surface area of the complex interface of the crystal, AutoDock, GOLD and MVD binding modes in the initial structures were 857.8, 950.3, 979.3 and 1025.9 Å$^2$, respectively. Although the value for $E$ determined in the AutoDock binding mode was the lowest among the four modes examined, $(E_{Complex} - TS_{Complex} + \Delta\mu_{Complex})$ value was higher than the crystal binding mode. Consistent with the experimental results, the crystal binding mode was found to have the lowest binding energy. The three docking programs failed to predict the correct binding mode. These results suggest that the docking programs did not adequately account for solvation effects.

## 4. Conclusions

We propose a method for calculating the binding free energy of protein-ligand complexes that combines all-atom molecular dynamics simulation of solutes and pure water with the ER method. Four distinct lysozyme–triNAG complex binding modes were investigated. The crystal binding mode, in which the TriNAG molecule protrudes slightly into solvent upon binding, was identified as having the lowest binding energy. Our results indicate the difficulty in selecting correct binding mode without accurate energy calculation. We conclude that the proposed method for calculating binding energy values is useful for distinguishing the most plausible binding mode complex-forming molecules. The proposed method requires few ns MD simulations of pure water, solution, and solute *in vacuo*. To evaluate multiple binding modes, only few ns MD simulations of solution and solute *in vacuo* are performed for each binding mode; the former is much faster than the FEP-type calculation as mentioned in Introduction and the latter computationally costs only few percent of the former. Only one MD simulation of pure water is required because the result is used

**Table 2**
Energy difference among lysozyme–triNAG complex models in kcal/mol.

| Model | $E_{Complex}$[a] | $\Delta\mu_{Complex}$[a] | $TS^{trans}$ | $TS^{rot}$[a] | $TS^{int}$[b] | $TS_{Complex}$ | $E_{Complex} + \Delta\mu_{Complex} - TS_{Complex}$ |
|---|---|---|---|---|---|---|---|
| Crystal | −2499.9 ± 2.7 | −698.3 ± 3.8 | 14.5 | 16.4 ± 0.0 | 1544.9 | 1575.8 | −4774.0 ± 4.7 |
| AutoDock | −2537.9 ± 3.2 | −663.3 ± 14.3 | 14.5 | 16.4 ± 0.0 | 1533.3 | 1564.1 | −4765.3 ± 14.6 |
| GOLD | −2510.4 ± 2.8 | −660.2 ± 9.7 | 14.5 | 16.4 ± 0.0 | 1542.6 | 1573.5 | −4744.1 ± 10.1 |
| MVD | −2512.5 ± 2.5 | −669.8 ± 5.6 | 14.5 | 16.4 ± 0.0 | 1532.2 | 1563.1 | −4745.3 ± 6.1 |

[a] The error is expressed at 95% confidence from ten 100-ps simulations.
[b] Calculated from 1-ns simulation.

for the solute insertion and it can be shared among all the complex models. Therefore, this computational cost is roughly proportional to the number of binding modes examined. Overall, the proposed method is computationally not demanding compared to the FEP-type calculation. It should be also noted that this calculation can be applied to large molecules, for example, proteins as shown in this Letter, as well as small compounds.

## Acknowledgements

This research was supported by the Strategic Programs for Innovative Research (SPIRE) of the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan, the Computational Materials Science Initiative (CMSI), Japan, Japan Society for the Promotion of Science (JSPS) Grants-in-Aid for Science Research (B), and MEXT Grants-in-Aid for Science Research in Innovative Areas to A.K. and N.M. The computation in this Letter was conducted using the supercomputers at the Research Center for Computational Science, Okazaki Research Facilities, National Institute of Natural Science and our PC clusters, as well as the facilities of the Supercomputer Center, Institute for Solid State Physics, University of Tokyo.

## References

[1] P. Kollman, Chem. Rev. 93 (1993) 2395.
[2] M.R. Shirts, D.L. Mobley, J.D. Chodera, Annu. Rep. Comput. Chem. 3 (2007) 41.
[3] J.D. Chodera, D.L. Mobley, M.R. Shirts, R.W. Dixon, K. Branson, V.S. Pande, Curr. Opin. Struct. Biol. 21 (2011) 150.
[4] R.W. Zwanzig, J. Chem. Phys. 22 (1954) 1420.
[5] J. Åqvist, C. Medina, J.E. Samuelsson, Protein Eng. 7 (1994) 385.
[6] N. Matubayasi, M. Nakahara, J. Chem. Phys. 113 (2000) 6070.
[7] N. Matubayasi, M. Nakahara, J. Chem. Phys. 117 (2002) 3605 (J. Chem. Phys. 118 (2003) 2446 (erratum)).
[8] N. Matubayasi, M. Nakahara, J. Chem. Phys. 119 (2003) 9686.
[9] Y. Karino, M.V. Fedorov, N. Matubayasi, Chem. Phys. Lett. 496 (2010) 351.
[10] R. Wolfenden, L. Andersson, P.M. Cullis, C.C.B. Southgate, Biochemistry (Mosc.) 20 (1981) 849.
[11] M.R. Shirts, V.S. Pande, J. Chem. Phys. 122 (2005).
[12] K. Maenaka, M. Matsushima, H. Song, F. Sunada, K. Watanabe, I. Kumagai, J. Mol. Biol. 247 (1995) 281.
[13] N. Kamiya, Y. Yonezawa, H. Nakamura, J. Higo, Proteins-Struct., Funct., Bioinf. 70 (2008) 41.
[14] G.M. Morris, D.S. Goodsell, R.S. Halliday, R. Huey, W.E. Hart, R.K. Belew, A.J. Olson, J. Comput. Chem. 19 (1998) 1639.
[15] G. Jones, P. Willett, R.C. Glen, A.R. Leach, R. Taylor, J. Mol. Biol. 267 (1997) 727.
[16] R. Thomsen, M.H. Christensen, J. Med. Chem. 49 (2006) 3315.
[17] J.C. Cheetham, P.J. Artymiuk, D.C. Phillips, J. Mol. Biol. 224 (1992) 613.
[18] D.A. McQuarrie, Statistical Mechanics, University Science Books, Sausalito, 2000.
[19] M. Karplus, J.N. Kushick, Macromolecules 14 (1981) 325.
[20] A. Kitao, F. Hirata, N. Go, Chem. Phys. 158 (1991) 447.
[21] J. Schlitter, Chem. Phys. Lett. 215 (1993) 617.
[22] I. Andricioaei, M. Karplus, J. Chem. Phys. 115 (2001) 6289.
[23] R. Baron, W.F. van Gunsteren, P.H. Hünenberger, Trends Phys. Chem. 11 (2006) 87.
[24] J. Numata, M. Wan, E.-W. Knapp, Genome Inform. 18 (2007) 191.
[25] H. Saito, N. Matubayasi, K. Nishikawa, H. Nagao, Chem. Phys. Lett. 497 (2010) 218.
[26] Y. Karino, N. Matubayasi, J. Chem. Phys. 134 (2011).
[27] K. Takemura, H. Guo, S. Sakuraba, N. Matubayasi, A. Kitao, J. Chem. Phys. 137 (2012) 215105.
[28] N. Matubayasi, W. Shinoda, M. Nakahara, J. Chem. Phys. 128 (2008) 195107.
[29] M.C. Vaney, S. Maignan, M. RiesKautt, A. Ducruix, Acta Crystallogr. D: Biol. Crystallogr. 52 (1996) 505.
[30] J.C. Phillips et al., J. Comput. Chem. 26 (2005) 1781.
[31] A.D. MacKerell et al., J. Phys. Chem. B 102 (1998) 3586.
[32] A.D. Mackerell Jr., M. Feig, C.L. Brooks 3rd, J. Comput. Chem. 25 (2004) 1400.
[33] O. Guvench et al., J. Chem. Theory Comput. 7 (2011) 3162.
[34] T. Wiseman, S. Williston, J.F. Brandts, L.N. Lin, Anal. Biochem. 179 (1989) 131.
[35] W. Humphrey, A. Dalke, K. Schulten, J. Mol. Graph. Model. 14 (1996) 33.

*Published online 24 November 2012*

*Nucleic Acids Research, 2013, Vol. 41, No. 2* ***1355–1362***
*doi:10.1093/nar/gks1132*

# Anti-prion activity of an RNA aptamer and its structural basis

**Tsukasa Mashima[1,2], Fumiko Nishikawa[3], Yuji O. Kamatari[4], Hiromichi Fujiwara[1,2], Masayuki Saimura[1], Takashi Nagata[1,2], Tsutomu Kodaki[1,2], Satoshi Nishikawa[3], Kazuo Kuwata[5,6] and Masato Katahira[1,2,6,*]**

[1]Institute of Advanced Energy, [2]Graduate School of Energy Science, Kyoto University, Gokasho, Uji, Kyoto 611-0011, [3]National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Ibaraki 305-8566, [4]Life Science Research Center, Gifu University, 1-1 Yanagido, Gifu 501-1194, [5]United Graduate School of Drug Discovery and Medical Information Sciences, Gifu University, 1-1 Yanagido, Gifu 501-1193 and [6]CREST, Japan Science and Technology Agency, 4-1-8 Honcho, Kawaguchi, Saitama 332-0012, Japan

Received August 9, 2012; Revised October 20, 2012; Accepted October 22, 2012

## ABSTRACT

**Prion proteins (PrPs) cause prion diseases, such as bovine spongiform encephalopathy. The conversion of a normal cellular form ($PrP^C$) of PrP into an abnormal form ($PrP^{Sc}$) is thought to be associated with the pathogenesis. An RNA aptamer that tightly binds to and stabilizes $PrP^C$ is expected to block this conversion and to thereby prevent prion diseases. Here, we show that an RNA aptamer comprising only 12 residues, r(GGAGGAGGAGGA) (R12), reduces the $PrP^{Sc}$ level in mouse neuronal cells persistently infected with the transmissible spongiform encephalopathy agent. Nuclear magnetic resonance analysis revealed that R12, folded into a unique quadruplex structure, forms a dimer and that each monomer simultaneously binds to two portions of the N-terminal half of $PrP^C$, resulting in tight binding. Electrostatic and stacking interactions contribute to the affinity of each portion. Our results demonstrate the therapeutic potential of an RNA aptamer as to prion diseases.**

## INTRODUCTION

Prion protein (PrP) is almost ubiquitously expressed and is highly conserved in mammals. PrP exhibits two alternative forms; a normal cellular form ($PrP^C$), which is a soluble α-helix-rich isoform, and an abnormal form ($PrP^{Sc}$), which is an insoluble β-sheet-rich isoform and resistant to cleavage by proteinase K (1,2). The details of the structure of $PrP^{Sc}$ and the mechanism underlying the conversion of $PrP^C$ to $PrP^{Sc}$ remain unknown. Prion diseases, also known as transmissible spongiform encephalopathies (TSEs), are invariably fatal neurodegenerative disorders of mammals characterized by the accumulation of $PrP^{Sc}$ in the central nervous system (1–7). These disorders include Creutzfeldt–Jakob disease in humans, bovine spongiform encephalopathy in cattle and scrapie in sheep.

Considerable efforts have been made to develop a compound that inhibits the accumulation of $PrP^{Sc}$ in prion-infected cells (8–17). An RNA aptamer that tightly binds to $PrP^C$ is expected to stabilize $PrP^C$ and, thus, to block the conversion to $PrP^{Sc}$. Therefore, such an RNA aptamer may prevent prion diseases. However, attempts in this context have been limited, and a structural basis of the binding of an RNA aptamer to $PrP^C$, which would facilitate such an application, has not been available. We identified RNA aptamers that tightly bind to bovine $PrP^C$ ($bPrP^C$) (18,19). We revealed that the RNA aptamer specifically binds to $PrP^C$ among various kinds of proteins present in bovine brain homogenate (18). Surprisingly, among the RNA aptamers, a short RNA comprising only 12 residues, r(GGAGGAGGAGGA) (R12), binds to bPrP strongly, the dissociation constant being $8.5 \times 10^{-9}$ M (18). We already reported the unique quadruplex structure of R12 (20). Here, we have revealed that R12 actually exhibits anti-prion activity on the basis of the assay with mouse neuronal cells. Nuclear magnetic resonance (NMR) analysis elucidates the structural basis of tight binding of R12 with $PrP^C$ that causes anti-prion activity.

## MATERIALS AND METHODS

### Sample preparation

R12 [r(GGAGGAGGAGGA)], D12 [d(GGAGGAGGAGGA)] and U12 [$r(U)_{12}$], synthesized, purified by

*To whom correspondence should be addressed. Tel: +81 774 38 3517; Fax: +81 774 38 3524; Email: katahira@iae.kyoto-u.ac.jp

high-performance liquid chromatography and desalted, were purchased from Sigma-Aldrich and Nippon Seihun, respectively. P1 (residues 25–35 of bPrP: $NH_2$-SKKRPKPGGGWN-COOH), P16 (residues 108–119: $NH_2$-GQWNKPSKPKTN-COOH) and mutant P16 peptides (K5A, K8A, K10A and W3A, respectively), synthesized and purified by high-performance liquid chromatography, were purchased from Sigma-Aldrich or Hipep Laboratories. Wild-type bPrP-N (residues 25–131) and mutant bPrP-N (residues 25–131, 5WS) were expressed with an *Escherichia coli* system and purified using a reported protocol (21). Full-length bPrP (residues 25–241) was purchased from Alicon.

### Evaluation of anti-prion activity

The anti-prion activity of either R12 or D12 was examined using mouse neuronal cells (GT1–7) persistently infected with the human TSE agent (Fukuoka-1 strain), designated as GT + FK, as described previously (14,16). The cells were grown and maintained at 37°C under 5% of $CO_2$ in Dulbecco's modified Eagle's medium (Invitrogen) supplemented with 10% of fetal bovine serum (Equitech-bio), 50 U/ml of penicillin G sodium and 50 μg/ml of streptomycin sulphate (Invitrogen). Approximately $1.5 \times 10^5$ cells were plated in each well of a six-well plate, and treatment with either R12 or D12 was started 15 h later. R12 (0.25 mM) and D12 (0.35 mM) were dissolved in a buffer solution containing 10 mM of K-phosphate (pH 6.2) and 100 mM of KCl, respectively. Either 80 μl of the R12 solution or 57 μl of the D12 solution was added to 2 ml of the medium, with the final concentration of both R12 and D12 being 10 μM. As a control, either 80 or 57 μl of the buffer solution was added to the medium. For the conditions with an RNase inhibitor, 800 U of RNasin (Promega) was added to the medium 5 min before the addition of the R12 solution. As a reference nucleic acid, U12 was added to the medium to the final concentration of 10 μM after the addition of the RNase inhibitor. After 72 h of treatment, cells were lysed in 150 μl of 1× Triton X-100-deoxycholate lysis buffer [150 mM of NaCl, 0.5% of Triton X-100, 0.5% of sodium deoxycholate, 50 mM of Tris–HCl (pH 7.5)], and the supernatant was collected. Samples were digested with 20 μg/ml of proteinase K for 30 min at 37°C. Western blotting for $PrP^{Sc}$ was performed as described previously (22). As the primary antibody, PrP M-20 antibody (Santa Cruz Biotechnology) was used to detect $PrP^{Sc}$. The signals were visualized with Super-Signal (Pierce Biotechnology) and scanned using a LAS-1000 UV mini analyser (Fuji Film). The density of $PrP^{Sc}$ in each solution was measured and compared with that in the control treated with just the buffer solution. Each assay with either R12 or D12 was repeated eight times.

### Microchip electrophoresis assay

A Hitachi SV1210 microchip CE system (Hitachi Electronics Engineering) was used for microchip electrophoresis (ME) analysis, as described previously (19,20). R12 was incubated in the binding buffer comprising 20 mM of Tris–HCl (pH 7.5) and 10 mM of KCl under the re-folding conditions: 90°C for 2 min, 72°C for 5 min, 55°C for 5 min and 37°C for 2 min and then kept at 25°C. After the addition of 10 nM of Cy5-2′-deoxyuridine 5′-triphosphate (dUTP) (Amersham Biosciences), the internal standard, R12 (100 nM), was incubated with a peptide (0 and 100 μM) in the binding buffer of 10 μl for 15 min and was then analysed by ME (20). The observed peak intensity of R12 was divided by the peak intensity of the internal standard, which gave the relative peak intensity. The relative peak intensity value of R12 was normalized by dividing by the relative peak intensity of R12 in the absence of peptide, which gave the normalized peak intensity. The relative binding of R12 was defined as 1 (the normalized peak intensity) and was calculated for the wild-type and mutant P16 peptide (Figure 3H).

### Filter binding assay

The 5′-end γ-$^{32}$P-labelled R12 (<10 nM) was mixed with varying concentrations of either a peptide or protein to give a total volume of 25 μl in a solution comprising 20 mM of Tris–HCl (pH 7.5) and 10 mM of KCl. After 20 min incubation, each mixture was passed through a nitrocellulose filter and washed with 500 μl of the solution. The amount of bound RNA was measured with BAS 2500 (Fuji Film), and binding activities were calculated as the percentage of input RNA in the peptide- or protein-RNA complex retained on the filter. We determined the dissociation constant ($K_d$) using GraphPad PRISM and non-linear regression curve fitting, and a one site-binding hyperbola equation {RNA binding (%) = $B_{max} \times$ (peptide or protein)/[$K_d$ + (peptide or protein)], where $B_{max}$ is the maximum bound at the saturating peptide or protein concentration} (18).

### Titration and structure determination

In titration experiments, either P1, P16 or bPrP-N(25–131, 5WS) was added step-by-step to R12 in a solution comprising 10 mM of K-phosphate (pH 6.2) and 10 mM of KCl at 30°C. Structural determination of the R12:P16 complex was carried out under the same conditions. NMR spectra, nuclear overhauser effect spectroscopy (NOESY), total correlation spectroscopy (TOCSY), double-quantum filtered correlation spectroscopy (DQF-COSY), $^1$H-$^{13}$C heteronuclear single quantum coherence (HSQC) and heteronuclear multiple bond coherence with jump and return solvent suppression (JRHMBC) were recorded with Bruker DRX600 and DRX800 spectrometers equipped with a cryoprobe. Distance and dihedral constraints were obtained as described previously (20) (Supplementary Table S1). Structural calculations were carried out using these constraints and a simulated annealing protocol supplied with XPLOR-NIH v. 2.26 (23,24), as described previously (20,25). The R12:P16 complex comprises two R12 monomers and two P16 monomers, as explained in the main text. Ten final structures were selected from 100 calculations on the basis of the criterion of the smallest residual energy term. The statistics are shown in Supplementary Table S1. The final ensemble of ten structures of the R12:P16 complex has 13.1 and 86.9% in the most favoured and additional allowed regions, respectively, for the P16 peptide. Molecular images were

generated with MOLMOL (http://www.mol.biol.ethz.ch/groups/wuthrich_group/software) and University of California, San Francisco (UCSF) Chimera (http://www.cgl.ucsf.edu/chimera/).

## RESULTS

### Anti-prion activity of R12

The anti-prion activity of R12 has been examined using mouse neuronal cells (GT1-7) persistently infected with the human TSE agent (Fukuoka-1 strain), designated as GT+FK. R12 was found to inhibit the $PrP^{Sc}$ formation in GT+FK at 10 μM (second and third lanes in Figure 1). R12 reduced the $PrP^{Sc}$ level to 65.0% of that in the control (Table 1, $P < 0.001$). When an RNase inhibitor, RNasin, was added to the medium, R12 reduced the $PrP^{Sc}$ level even more, that is, to 49.4% (eighth and ninth lanes in Figure 1 and Table 1, $P < 0.001$). The observed difference between in absence and presence of an RNase inhibitor (Table 1, $P < 0.05$) is supposed to be because of the repressed degradation of R12 in the medium caused by the RNase inhibitor. No reduction of the $PrP^{Sc}$ level was observed for U12 even in the presence of the RNase inhibitor (Table 1). Therefore, the reduction is specific to R12. The DNA version of R12, d(GGAGGAGGAGGA) (D12), also tightly binds to $bPrP^{C}$, the dissociation constant being $6.4 \times 10^{-8}$ M (18). D12 reduced the $PrP^{Sc}$ level to 75.5% (fifth and sixth lanes in Figure 1 and Table 1, $P < 0.01$), although its inhibitory effect is less than that of R12. Thus, we have demonstrated that R12, and also D12 to a lesser extent, exhibits anti-prion activity towards cells persistently infected with the TSE agent.

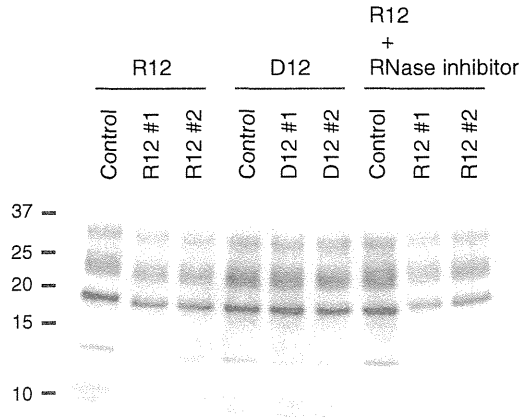


**Figure 1.** Anti-prion activity of r(GGAGGAGGAGGA) (R12) and d(GGAGGAGGAGGA) (D12). Western blotting of $PrP^{Sc}$ in GT+FK cells after treatment with either 10 μM R12 or D12. Two independent experiments, #1 and #2, are shown. The control was treated with just the buffer solution. The treatment with R12 was also performed in the presence of an RNase inhibitor, RNasin. Molecular mass markers are shown at the left.

### Structure of R12 in complex with the binding peptide of bPrP

Then, we tried to elucidate the structural basis of the tight binding of R12 to $bPrP^{C}$. The structure of $bPrP^{C}$ is available (21); the C-terminal half of $bPrP^{C}$ is composed of three α-helices and two short β-strands, whereas the N-terminal half is basically flexible and disordered. Two portions of bPrP were identified as binding sites for R12 by epitope mapping using ME (20). They are residues 25–35 (SKKRPKPGGGWN, designated as P1) and 108–119 (GQWNKPSKPKTN, designated as P16) of bPrP. Both binding sites commonly include three Lys residues and one Trp residue. The dissociation constants for the R12:P1 and R12:P16 complexes are $1.7 \times 10^{-5}$ M and $1.0 \times 10^{-5}$ M, respectively, being similar to each other. Chemical shift perturbation of each imino proton of R12 was traced during the course of titration with P16 (Figure 2A). The perturbation reached a plateau when the P16/R12 molar ratio was 3.0. Thus, the perturbation at each molar ratio was divided by the perturbation at the molar ratio of 3.0 and defined as a normalized perturbation for each imino proton. Then, the normalized perturbation was plotted against the P16/R12 molar ratio (Figure 2B). The initial slope of the plot almost coincides with that expected for the formation of the R12:P16 = 1:1 complex. The deviation of the experimental points from the initial slope for larger P16/R12 molar ratios is rationally understood, when the dissociation constant of $1.0 \times 10^{-5}$ M and the concentrations of P16 and R12 applied for each point are taken into account, which also supports the formation of the 1:1 complex. As described later in the text, R12 forms a dimer in the complex. Therefore, the complex should comprise two R12 monomers and two P16 monomers.

**Table 1.** Anti-prion activities of R12 and D12*

| Treatment | Relative $PrP^{Sc}$ level (%) |
|---|---|
| Control | 100 |
| R12 | 65.0 ± 9.3 |
| R12 + RNase inhibitor | 49.4 ± 17.4 |
| U12 + RNase inhibitor | 98.8 ± 5.8 |
| D12 | 75.5 ± 17.0 |

*Each assay was repeated eight times.
**$P < 0.05$; ***$P < 0.01$; ****$P < 0.001$.

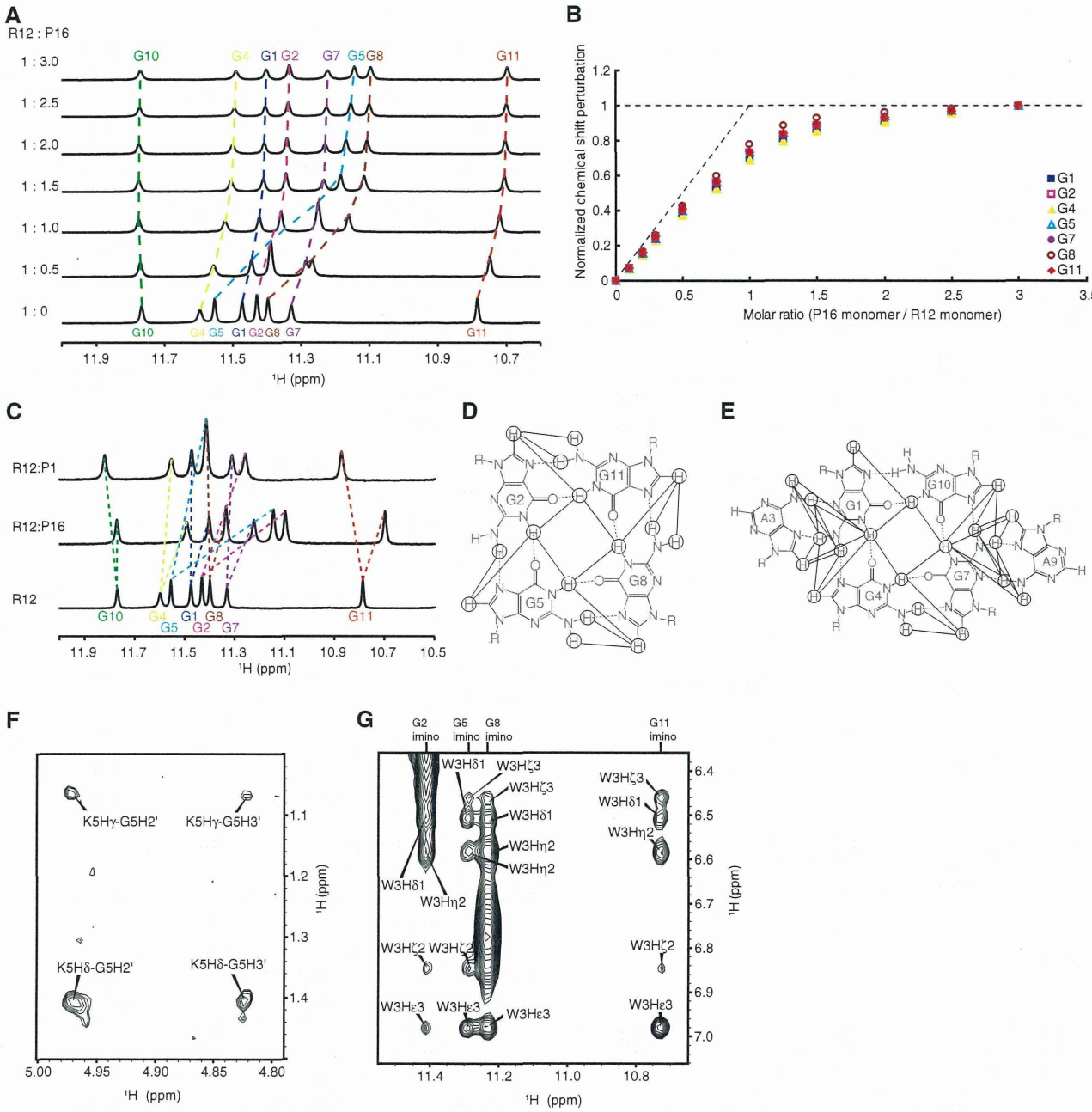


**Figure 2.** NMR analysis of the R12:P16 complex. (**A**) The imino proton spectra of R12 in the course of the titration with P16, the assignments being indicated by residue numbers. The numbering of R12 is G1G2A3G4G5A6G7G8A9G10G11A12. The molar ratio is indicated at the left. (**B**) Normalized chemical shift perturbation for each imino proton of R12 in the course of the titration with P16. (**C**) Imino proton spectra of R12, the R12:P1 complex and the R12:P16 complex with the assignments. (**D** and **E**) The G:G:G:G tetrad (D) and G(:A):G:G(:A):G hexad (E) planes. The observed NOESY cross peaks and hydrogen bonds are indicated by solid and dotted lines, respectively. (**F** and **G**) Intermolecular NOESY cross peaks observed for the K5 (F) and W3 (G) residues of P16, respectively. The numbering of P16 is G1Q2W3N4K5P6S7K8P9K10T11N12.

The imino proton spectrum of the R12:P1 complex is similar to that of the R12:P16 complex (Figure 2C). In fact, the chemical shift perturbation of R12 on binding of P1 is similar to that on binding of P16, except for that for a G11 residue (Supplementary Figure S1). This suggests that the interaction of R12 with P1 is similar to that with P16, although some difference in interaction and/or structure may be present around G11. The similar dissociation constants obtained for the R12:P1 and R12:P16 complexes are consistent with this idea. Thus, to elucidate the interaction of R12 with bPrP, we have determined the structure of R12 in a complex with P16 as a representative.

The resonance assignments of R12 in the R12:P16 complex were made (Supplementary Figure S2A and B) in the same way as reported for free R12 (20). Then, the G:G:G:G tetrad and G(:A):G:G(:A):G hexad structures were identified on the basis of the observation of characteristic NOESY cross peaks (Figure 2D and E), as for free R12 (20). The distance between H2 and H1′ of an

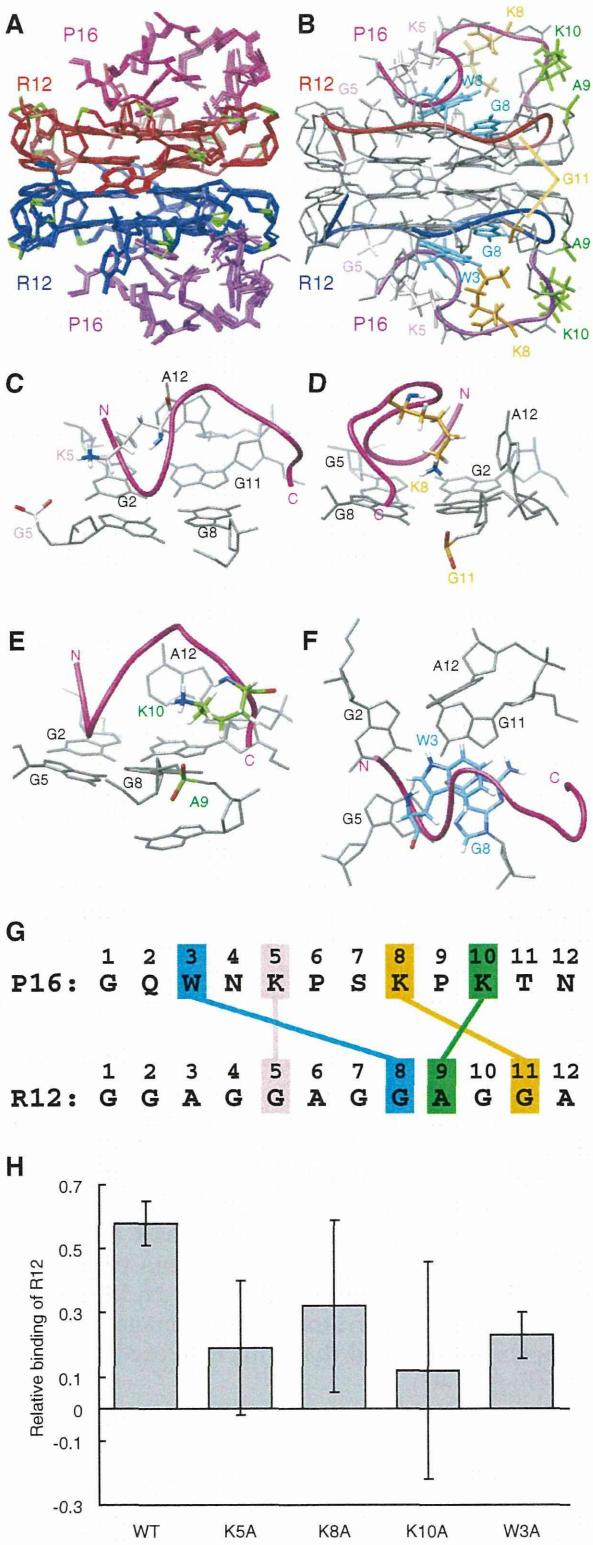


**Figure 3.** The structure of the R12:P16 complex. (**A**) NMR structure ensemble of the R12:P16 complex. The two R12 monomers are coloured red and blue, respectively, phosphate groups being coloured green. The two P16 monomers are coloured magenta and purple, respectively. (**B**) The representative structure with the lowest energy. The chains connecting the C3′ and C4′ atoms of each R12 monomer are indicated by red and blue tubes, respectively. The chains connecting the $C^{\alpha}$ atoms of each P16 monomer are indicated by magenta and purple tubes, respectively. The pairs K5 and G5, K8 and G11, K10 and A9 and W3 and G8 are coloured pink, orange, green and blue, respectively.

adenosine residue is larger than 4.5 Å for any conformation (26). Therefore, observation of strong NOESY cross peaks between H2 and H1′ for an A3 residue and also for an A9 residue (Supplementary Figure S2C) is an indication that R12 in the complex forms a dimer, as free R12 does (20). The dimer formation of R12 in the complex was supported by many other NOESY cross peaks as well. Resonance assignments of P16 in the R12:P16 complex were also made (Supplementary Figure S2D) using a standard protocol (26). Then, many intermolecular NOESY cross peaks were identified on the basis of the resonance assignments (Figure 2F and G).

The structure of the R12:P16 complex comprising two R12 monomers and two P16 monomers was determined on the basis of distance and dihedral angle constraints (Figure 3). The structure of the complex reveals the origin of the high affinity. Three Lys residues of P16 are involved in the electrostatic interactions with the phosphate groups of R12; K5 of P16 with G5 of R12, K8 with G11 and K10 with A9, respectively (Figure 3C–E and G). The average of the closest distance for 10 final structures between an $H^{\zeta}$ atom of the Lys residue and an oxygen atom of a phosphate group is 3.71 Å, 3.67 Å and 4.65 Å, respectively. In addition, a stacking interaction was found between the rings of W3 of P16 and G8 of R12 (Figure 3F and G). These interactions were biochemically confirmed by ME using mutant P16 peptides. A decrease in affinity to R12 was observed for four mutant P16 peptides in which either K5, K8, K10 or W3 was replaced by an alanine residue, respectively (Figure 3H).

## Overall architecture responsible for tight trapping of bPrP by R12

As described earlier in the text, R12 is suggested to interact with P1 in a similar way as it does with P16, the interaction involving three lysine residues and one tryptophan residue, which both P1 and P16 commonly possess. Therefore, it is likely that in the R12:bPrP complex, one monomer of the R12 dimer interacts with P1 and the other monomer with P16 (Figure 4A). To confirm this, a titration experiment on the N-terminal region of bPrP, bPrP-N(25–131), which contains the two binding sites, P1 (residues 25–35) and P16 (residues 108–119), was performed. bPrP-N(25–131) tightly binds to R12. The dissociation constant for the R12:bPrP-N(25-131) complex was $2.1 \times 10^{-8}$ M, which is roughly comparable with that for the R12: full-length bPrP(25–241) complex, $8.5 \times 10^{-9}$ M. Unfortunately, precipitation was observed in the course of the titration with bPrP-N(25–131), probably because of the low solubility of the complex, which hinders detailed analysis. To improve its solubility,

**Figure 3.** Continued
(**C–F**) Close-up views of intermolecular interactions between K5 of P16 and G5 of R12 (C), K8 and G11 (D), K10 and A9 (E) and W3 and G8 (F), respectively. (**G**) Summary of interactions. The electrostatic interactions between K5 of P16 and G5 of R12, K8 and G11 and K10 and A9 and the stacking interaction between W3 and G8 are indicated. (**H**) The relative binding is shown for wild-type P16 and four mutant P16, in which either K5, K8, K10 or W3 was replaced by an alanine residue, respectively.

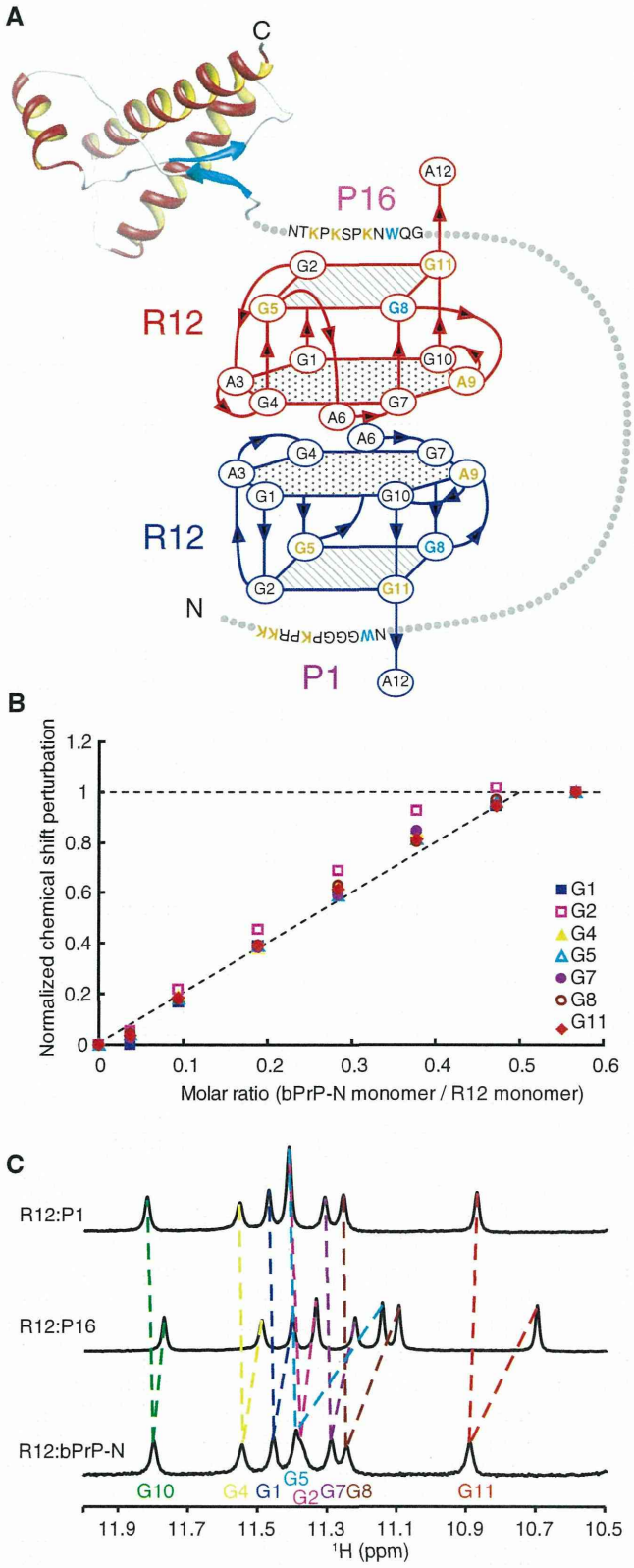


**Figure 4.** Overall architecture of tight trapping of bPrP by R12. (**A**) Overall architecture of the complex between bPrP and R12. The tetrad and hexad planes of R12 are indicated by a square and a hexagon, respectively. The lysine and guanosine residues involved in the electrostatic interaction are coloured orange, and the tryptophan and guanosine residues involved in the stacking interaction are coloured blue, respectively. The structure of the C-terminal region of bPrP is drawn on the basis of the coordinates under accession number 1DX0 in the Protein Data Bank (21). (**B**) Normalized chemical shift perturbation for each imino proton of R12 in the course of the titration with bPrP-N(25–131, 5WS). (**C**) Imino proton spectra of the R12:P1 complex, the R12:P16 complex and the R12:bPrP-N(25–131, 5WS) complex with the assignments.

five tryptophan residues of the octapeptide repeat region of bPrP-N(25–131), Trp 60, Trp 68, Trp76, Trp 84 and Trp 92 were replaced by serine residues, resulting in bPrP-N(25–131, 5WS). It should be noted that the tryptophan residues that are involved in P1 and P16 and interact with R12 are Trp 34 and Trp 110, respectively, and that they are not replaced. In fact, it was found that bPrP-N(25–131, 5WS) also tightly binds to R12, the dissociation constant being $5.5 \times 10^{-8}$ M, which is almost comparable with that for wild-type bPrP-N(25–131). Precipitation was not observed during the titration with bPrP-N(25–131, 5WS). The chemical shift perturbation of each imino proton of R12 in the course of the titration with bPrP-N(25–131, 5WS) was traced. The perturbation reached a plateau when the bPrP-N/R12 molar ratio was 0.56. Thus, the perturbation at each molar ratio was divided by the perturbation at the molar ratio of 0.56 and defined as a normalized perturbation for each imino proton. The normalized perturbation was plotted against the molar ratio (Figure 4B). The initial slope of the plot almost coincides with that expected for the formation of the R12:bPrP-N(25–131, 5WS) = 1:0.5 complex. Because of the dimer formation of R12, this indicates that the complex comprises two R12 monomers and one bPrP-N(25–131, 5WS) molecule. This is consistent with the idea aforementioned that one monomer of the R12 dimer interacts with P1 of bPrP and that the other monomer with P16 of the same bPrP molecule (Figure 4A). The deduced simultaneous binding of the R12 dimer with two binding sites of a single bPrP molecule must also contribute to the high affinity because the double binding could ideally increase the affinity by as much as the square of the binding constant if each interaction is perfectly achieved, respectively (27).

Only one resonance is observed for each imino proton of R12 in complex with bPrP-N(25–131, 5WS) (Figure 4C), although two imino protons of each monomer of R12 could potentially be non-equivalent in the complex. This observation suggests that each R12 monomer switches binding site, P1 and P16 and vice versa, in fast exchange regime on an NMR time scale. It is consistent with this idea that each imino proton resonance of R12 in the complex with bPrP-N(25–131, 5WS) appears between the corresponding resonances of R12 in the complex with P1 and that with P16, except for a G11 residue (Figure 4C).

## DISCUSSION

We have demonstrated that R12 reduces the $PrP^{Sc}$ level of mouse neuronal cells persistently infected with the human TSE agent. It is supposed that R12, which tightly binds to