Journal of Biomolecular NMR **29:** 325–338, 2004.

Model building of a protein-protein complexed structure using saturation transfer and residual dipolar coupling without paired intermolecular NOE

Tomoki Matsuda[a,**], Takahisa Ikegami[a], Nobuyuki Nakajima[a], Toshio Yamazaki[b] & Haruki Nakamura[a,*]

[a]*Institute for Protein Research, Osaka University, 3-2 Yamadaoka, Suita, Osaka 565-0871, Japan, and* [b]*RIKEN Genome Science Center, 1-7-22 Suehiro, Tsurumi, Yokohama, Kanagawa 230-0045, Japan*

Received 16 December 2003; Accepted 15 March 2004

Key words: protein-protein complex, residual dipolar coupling, rigid body docking, saturation transfer, simulated annealing

Abstract

For understanding the precise mechanisms of molecular recognition of proteins, three-dimensional structural analyses of the protein-protein complexes are essential. For this purpose, a new method to reveal complex structures was developed with the assistance of saturation transfer (SAT) and residual dipolar coupling (RDC) by heteronuclear NMR experiments, without any paired intermolecular NOE information. The SAT and RDC experiments provide the information of the interfacial residues and the relative orientations of the two protein molecules, respectively. Docking simulation was then made to reconstruct a complex conformation, which satisfies the SAT and RDC data. The method was applied to the CAD-ICAD complex structure, which was previously determined by the NOE-distance geometry method. The quality of the current model was evaluated.

Abbreviations: ASA – accessible surface area; HSQC – heteronuclear single quantum correlation; MD – molecular dynamics; NOE – nuclear Overhauser effect; TROSY – Transverse-Relaxation Optimized Spectroscopy

Introduction

Biological functions are much more mediated by their complex forms rather than monomer forms. Therefore, more protein structures in the states of complexes have to be analyzed to discuss wider ranges of biological systems at molecular levels.

However, structure determinations of complexes are still challenging. In particular by NMR, the structural information to be obtained is complicated and the procedures are therefore time-consuming. For example, to determine the relative position between molecules or domains, intermolecular or inter-domains NOEs (nuclear Overhauser effects) have generally been used. The number of such NOE cross-peaks is normally so few, the peaks are weak in their sensitivities, and NOE spectra are so complicated due to signal overlaps. Furthermore, the procedure often requires the assignments of most ^{1}H atoms in side chains in advance. Even if the structure of each monomer had already been determined in its free form, it would be cumbersome to assign the chemical shifts of proton nuclei in most side chains as well as in the main chain for subsequent analyses of intermolecular NOEs.

Prediction of a complex structure only from each monomer structure produces many candidates, which are hard to narrow down, because rigorous principles for protein-protein interactions have not yet been established. To overcome this problem, a combination of a structure-based prediction and simple experimental data for protein-protein binding interfaces is suitable.

Here, we propose a method by which the relative positions between two proteins (domains) in a

*To whom correspondence should be addressed. E-mail: harukin@protein.osaka-u.ac.jp

**Present address: Riken Genome Science Center, 1-7-22 Suehiro, Tsurumi, Yokohama, Kanagawa 230-0045, Japan.

complex form are identified by results from saturation transfer (SAT) experiments and residual dipolar coupling (RDC) constants of the amide groups. The SAT method has been developed by Takahashi et al. (2000), using a cross saturation phenomenon, which can detect residues on binding surfaces more precisely than conventional chemical shift perturbation (Foster et al., 1998) or H-D exchange experiments (Paterson et al., 1990). Furthermore, data obtained by this method directly relate to the distances between intermolecular atom pairs, contrary to the conventional methods. However, relative orientations between the components cannot be accurately determined, because the SAT method does not provide detailed atom pair information at the interaction site. On the contrary, RDC constants provide relative orientations accurately without any information on relative translational positioning. Therefore, both data are complementary to each other, and their combination is expected to determine the precise and accurate relative positions between the components.

We applied this method to the complex between the CAD domain of mouse caspase-activated deoxyribonuclease (CAD-CD: residues 1–87) and the CAD domain of its inhibitor (ICAD-CD: residues 1–100), which had been previously determined using the conventional NOE-distance geometry method by our group (Otomo et al., 2000).

Materials and methods

Sample preparation

Mouse CAD-CD and mouse ICAD-CD were prepared for the NMR experiments. Hereafter, CAD-CD and ICAD-CD are written as CAD and ICAD, respectively, in this paper for simplicity. To detect the region of a part of the complex molecule responsible for the interaction with another by cross saturation measurements, a pair of a ^{2}H, ^{15}N-labeled molecule and a non-labeled one were prepared. The former labeled sample was obtained by culturing the transformed bacteria in an M9 minimum medium with 99.9% D_2O containing $^{2}H_6$-glucose and $^{15}NH_4Cl$ as the sole carbon and nitrogen sources. To detect RDC constants for the amide ^{1}H and ^{15}N pairs in the mainchains, both proteins labeled with ^{15}N were prepared.

Each protein was overproduced in *E. coli* BL21(DE3) harboring the previously constructed plasmid (Otomo et al., 2000). The cells were cultured in an M9 minimum-medium, and the gene expression was induced by 1 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG). Both proteins were purified separately as described earlier (Otomo et al., 2000). Each sample dissolved in 20 mM potassium phosphate buffer (pH 6.4) containing 1 mM ethylene diamine-N,N,N′,N′-tetraacetic acid (EDTA), 20 mM dithiothreitol (DTT), and 10% D_2O was mixed with each other, so that the concentrations of CAD and ICAD were 0.20 and 0.22 mM, respectively. For the measurements of RDC in an anisotropic phase a lyotropic liquid crystalline phase was formed by addition of hexaethyleneglycol monododecylether, $C_{12}E_5$, ($CH_3(CH_2)_{11}(OCH_2CH_2)_5OH$, EC No. 221-281-8, M.W. 450.7, purchased from SigmaP-8550) and 1-hexanol as described in (Rueckert et al., 2000). The weight percent for the ratio of $C_{12}E_5$ to water and the molar ratio of $C_{12}E_5$ to 1-hexanol were 5% and 0.96, respectively.

Cross-saturation experiment

A cross-saturation experiment was employed with a Bruker DRX-800 NMR spectrometer, operated at the ^{1}H resonance frequency of 800.33 MHz, at 308 K basically in the same way as described in Takahashi et al. (2000). The ^{1}H-^{15}N heteronuclear single quantum correlation (HSQC) part was constructed with a normal INEPT type, not with a TROSY type as depicted in the paper (Takahashi et al., 2000), containing WATERGATE just before the direct detection period. Saturation was applied with a WURST-20 adiabatic pulse with a frequency sweep range of 2000 Hz, a duration of 10 ms, a maximum amplitude of 178.412 Hz, and an excitation center at the −2561 Hz off-resonance position by phase modulation. The simulation confirmed that the pulse excited only the aliphatic resonance region and had no effect on the water resonance. The saturation time lasted 2.0 s by 200 times of sequential repeats of this pulse. In the reference sequence, which was combined with the saturation sequence in an interleaved manner, the excitation center for the adiabatic pulse was shifted to −30 000 Hz off-resonance, so that it should have no effect on the protein resonance region but should maintain the other conditions such as sample heating by the pulses. A series of spectra was processed in the same way with the program NMRPipe (Delaglio, 1995) and analyzed with software Pipp (Garret et al., 1991).

Simulation of saturation transfer (SAT) phenomena

The reduction ratios of the peak intensities in the ^{1}H-^{15}N HSQC spectra with and without cross saturation were simulated using the proton coordinates of the CAD-ICAD complex. Although only aliphatic protons in the saturation-donor molecule were directly affected by the selective radio frequency field applied, all protons in the donor molecule were regarded as being well-saturated through spin diffusion during the saturation time. Relaxation matrix for the amide protons of the saturation-acceptor molecule was formulated from the distances between the protons in the atom coordinates of the complex structure (Dobson et al., 1982). The Bloch equation utilizing this matrix was formularized and solved to calculate the reduction ratio of the peak intensities for each amide proton in the saturation-acceptor molecule (Matsuda et al., 2004). The simulated ratios were applied to the calculation of the pseudo potential energy E_{sat}, defined by,

$$E_{\mathrm{sat}} = \sum \left(\eta^{\mathrm{sim}} - \eta^{\mathrm{exp}}\right)^2 /2, \tag{1}$$

where η^{sim} and η^{exp} were the simulated and experimentally measured peak reduction ratios, respectively. Only the reduction ratios that were less than 0.85 in the experiments were used for the calculation.

Measurement of scalar and dipolar coupling constants

A series of $^{1}J_{\mathrm{HN}}$-modulated two-dimensional (2D) ^{1}H-^{15}N HSQC-J spectra were measured according to the procedure described in (Tjandra et al., 1996) at 303 K with DRX-800 spectrometer. The anisotropic and isotropic phase samples were used to obtain the sum of the scalar J and dipolar D coupling constants and only the J coupling constant, respectively, for each amide group. The delay for the coupling evolution, T, which was inserted just before the ^{15}N chemical shift evolution period in the HSQC-J pulse sequence containing a gradient echo for the coherence selection, was 0.005, 6.0, 9.0, 12.0, 15.0, 18.0, 21.0, 24.0, 27.0, 30.0, and 33.0 ms. The procedure of the measurements was programmed in an interleaved manner to decrease otherwise possible systematic errors among the measurements containing the various delays. The delay T was sequentially changed prior to the increment of the ^{15}N evolution period, t_1. The sum of the scalar and dipolar coupling constants, $J + D$, for each amide group was determined by fitting the measured modulation of resonance intensities to the theoretical model in almost the same way as the method developed by Tjandra et al. (1996).

Molecular orientations from RDC information

The residual dipolar couplings are calculated from the three-dimensional (3D) structures of the protein molecules by usually using the method from Bax et al. (2001).

Each molecule was aligned independently in a direction such that the back-calculated dipolar coupling values from all N-H vectors were consistent with the experimental data by rotating the molecule with the Euler angles (3 parameters) and changing the magnitudes of the alignment tensor (2 parameters). Practically, a Powell minimization method (Press et al., 1992) was used to minimize the following target function:

$$E_{\mathrm{dip}} = \sum\nolimits_{\mathrm{i}} \left\{ \mathrm{D}^{\mathrm{exp}}_{\mathrm{NHi}} - \mathrm{D}^{\mathrm{calc}}_{\mathrm{NHi}} \right\}^2 /\sigma^2_{\mathrm{NHi}}, \tag{2}$$

where $\mathrm{D}^{\mathrm{exp}}_{\mathrm{NHi}}$ and $\mathrm{D}^{\mathrm{calc}}_{\mathrm{NH}}$ are the observed and calculated dipolar coupling values, respectively, and $^{\sigma}$NHi is the experimental uncertainties.

Molecular dynamics simulation

In order to obtain reasonable structures of protein complexes, we created a molecular dynamics (MD) simulation, modifying an MD program, PRESTO ver.3 (Morikami et al., 1992; Nakajima et al., 2000), to include RDC and SAT restraints. The intramolecular NOEs and the J-coupling constants provided information on the distances between the intramolecular atom pairs and the dihedral angles, and their restraints, E_{NOE} and E_J, which maintain the global conformation of each molecule, were added to the AMBER96 force field, E_{AMBER}. The total potential energy modified for MD, E_{MD}, was expressed as,

$$\begin{aligned} E_{MD} \;=\; & E_{AMBER} + W_{dip} E_{dip} + W_{sat} E_{sat} \\ & + W_{NOE} E_{NOE} + W_J E_J . \end{aligned} \tag{3}$$

Here, the weight factors for the corresponding restraints were set as $W_{dip} = 0.1$, $W_{sat} = 50$, $W_{NOE} = 10$, and $W_J = 500$. All simulations were carried out *in vacuo* with the SHAKE option applied to all the bonds involving hydrogen atoms with a unit time step of 0.5 fs. Following the constant temperature (300 K) condition for 2.5 ps, the system was then gradually cooled down to 10 K for 7.5 ps. After the MD simulation, the energy minimization with a 1000-step

conjugate gradient was applied. We performed five independent simulations, to each of which different random velocities were applied during the initial stage.

Results

Cross saturation experiments for the CAD-ICAD complex

Two kinds of molecular complexes (i) between the ^{2}H and ^{15}N doubly-labeled CAD and unlabeled ICAD, and (ii) between the ^{2}H and ^{15}N doubly-labeled ICAD and unlabeled CAD were prepared, and dissolved in 90% $^{1}H_2O$ / 10% $^{2}H_2O$. As shown in Figures 1a–d, the comparison between the ^{1}H-^{15}N HSQC spectra for the complexes with and without irradiating the aliphatic protons showed distinct reductions in the cross peak intensities for the cases with irradiation.

In the above case (i), the intensities of CAD were significantly reduced for the residues Val 11, Leu 13, Lys 21, Phe 22, Gly 23, Val 24, and Ala 25 by irradiating ICAD, as shown in Figures 1a and b. In contrast, in the case (ii), the peak intensities of ICAD for the residues Ile69, Asp74, Tyr75, Phe76, and Leu77 were greatly reduced by irradiating CAD, as shown in Figures 1c and d. The reduced intensity ratios for the individual residues of CAD and ICAD are plotted by open circles in Figures 2a and b, for the cases (i) and (ii), respectively.

Determination of dipolar coupling constants

Overall, 56 and 54 RDC constants for CAD, and 66 and 61 for ICAD were determined for the isotropic and anisotropic phases, respectively. The other RDC signals were discarded because the intensities of the corresponding peaks were too low. The average constants and uncertainties were 93.84 and 0.17 Hz for the isotropic phase and 93.61 and 0.50 Hz for the anisotropic phase. The larger uncertainties for the latter case were due to lower signal-to-noise ratios of the spectra caused by more inhomogeneous shimming and dipolar interactions with various other nuclei located in a vicinity of the corresponding nuclei in space. Dipolar coupling constants (D) were determined as the subtraction of the coupling constant extracted from the data for the isotropic phase (J) from that for the anisotropic phase ($J + D$) for each residue. As a result, 54 and 61 D values were obtained for CAD and ICAD, respectively, ranging from -31.12 to 21.38 Hz with an average uncertainty of 0.54 Hz.

Rigid body docking using RDC and SAT data

The observed RDC data did not fit well with the data that were back-calculated from the structure determined previously by the NOE-distance geometry method (PDB code, 1F2R). Therefore the structure of the CAD-ICAD complex was refined by adding the restraints derived from the current RDC data. Hereafter, refinement the structure was called NOE&RDC structure.

Each monomer structure of CAD and ICAD (residues 5–85 of CAD and 15–98 of ICAD) without the flexible terminals was extracted from the refined structure NOE&RDC, and they were used as the initial monomer structures for the following rigid body docking calculations. The procedure of the rigid body docking is composed of two steps, (i) a molecular alignment step by rotations and (ii) a docking step by translations.

In the first step, each molecule was aligned independently along its alignment tensor axes, as described in the Methods section. This procedure provided a set of N-H vectors, which satisfied the observed RDC values as shown in Figure 3.

The axial (A_a) and rhombic (A_r) components of the alignment tensor of CAD were 10.275×10^{-4} and 1.200×10^{-4}, respectively. Those of ICAD were 11.505×10^{-4}, 1.297×10^{-4}, respectively. The quality factor (Q) was calculated following to the next equation.

$$Q = \left(\sum_i \left(D_{NHi}^{exp} - D_{NHi}^{calc} \right)^2 \Big/ \sum_i \left(D_{NHi}^{exp} - D_{av} \right)^2 \right)^{1/2}, \quad (4)$$

where D_{NHi}^{exp} and D_{NHi}^{calc} stand for the observed and calculated dipolar coupling values, respectively, and D_{av} is the average of the observed values. Dobrodumov and Gronenborn (2003) describes that the Q factors are generally distributed between 0.15 and 0.5 for correct models. In our case, the Q factors of the CAD and ICAD were 0.13, small enough to show a good quality of the coordinates. A good correlation was observed between the measured and back-calculated dipolar coupling constants with the fitting values (χ^2) of CAD (3.4623×10^2), and ICAD (3.5656×10^2), respectively. They agreed well with the values for the NOE&RDC structure. Four different arrangements of CAD with respect to that of ICAD were possible by 180° rotation about the alignment axes as shown in Figure 4a: orientation A to D.

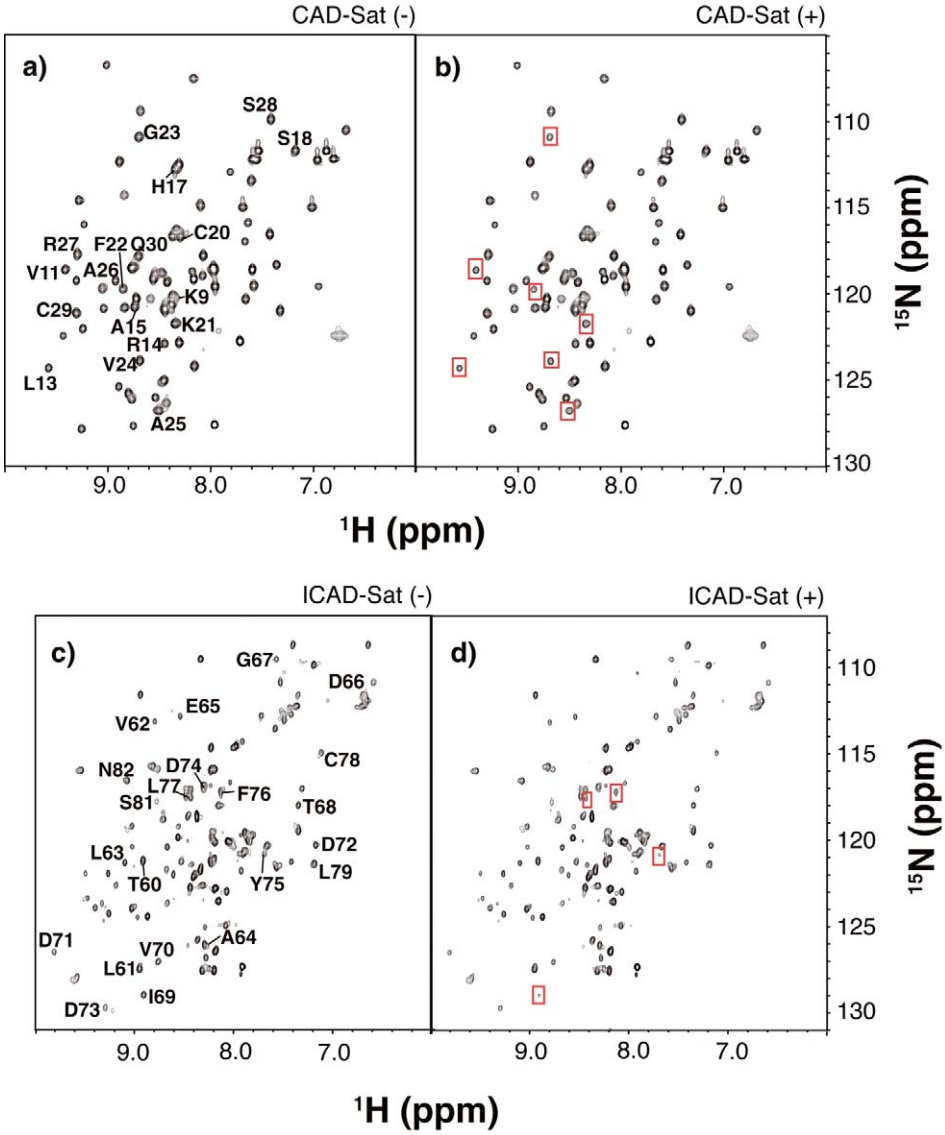

Figure 1. ^{1}H-^{15}N HSQC spectra observed for the ^{2}H, ^{15}N-labeled CAD when the non-labeled ICAD was (a) not irradiated, and (b) irradiated. ^{1}H-^{15}N HSQC spectra observed for the ^{2}H, ^{15}N-labeled ICAD when the non-labeled CAD was (c) not irradiated, and (d) irradiated. Residue numbers are labeled for the peaks on the binding surface on the left spectra. The peaks are enclosed by red rectangular frames on (b) and (d), for which the reduction ratios of the intensities with irradiation to those without irradiation were less than 0.7. The samples were dissolved in 90% H_2O/10% $^{2}H_2O$ containing 20 mM potassium phosphate buffer at pH 6.4. The concentrations of CAD and ICAD were 0.20 and 0.22 mM, respectively. The spectra were recorded at 308 K on a Bruker DRX spectrometer operating at the ^{1}H frequency of 800 MHz.

In the second step, the results of the SAT experiments shown in Figures 2a and b identified the molecular interface between the CAD and ICAD. The center of the binding surface was defined as the center of the gravity for the amide protons in the saturation acceptor molecule, for which the ratios of the peak reduction by saturation were less than 0.8. The saturation-donor molecule was placed at a position where its center of the gravity was on the extension of the vector directed from the center of the gravity of the acceptor molecule to the center of the binding surface (see Figure 4b). The distance between the centers of gravity for the donor and acceptor molecules was set to 30.0 Å (The distance between the intermolecular atom pair closest to each other was about 5 Å), and the four

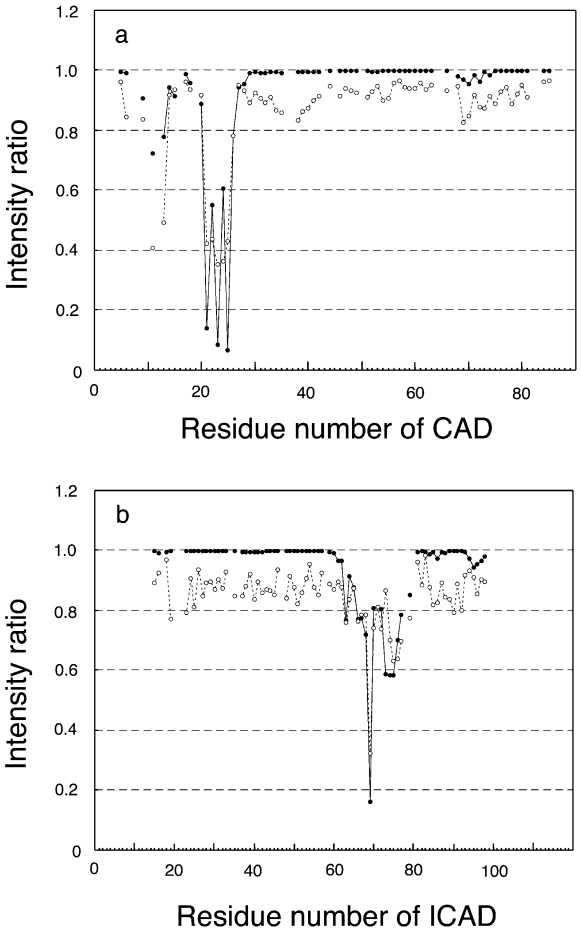

Figure 2. Plots of the intensity ratios of the cross-peaks originating from backbone amide groups in the saturation transfer experiments. The measured data (open circles with broken lines), and the simulated data from the relaxation matrix methods (filled circles with solid lines) based on the CAD-ICAD complex structure with the best-scored arrangement after the rigid body minimization are shown. (a) CAD and (b) ICAD were regarded as the SAT acceptors for the simulations, respectively.

relative arrangements were considered as mentioned above.

In order to expand the search space, sub-arrangements for the acceptor molecule were created from each of the four arrangements by translating the molecule on square lattice points (Figure 4b). The lattice was placed so that its center should be identical to the center of gravity of the acceptor molecule and that a vector normal to the lattice plane should be parallel to the vector directed from the center of gravity for the acceptor molecule to the center of the binding surface. Five lattice points were set at intervals of 5 Å for one side of the square, and 25 sub-arrangements were

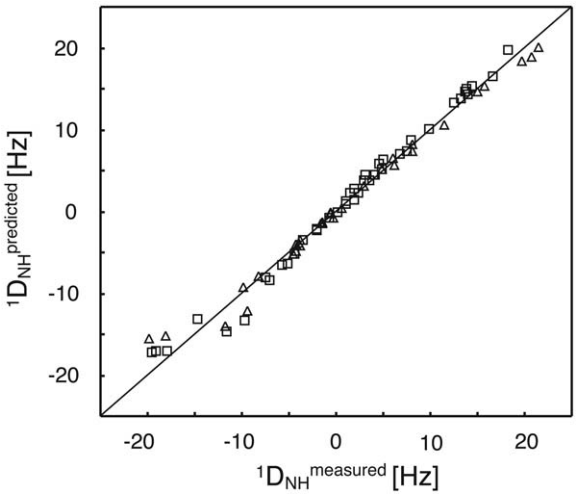

Figure 3. Comparison of the experimentally measured dipolar couplings for the CAD-ICAD complex with the back-calculated values, displayed by the program MODULE (Dosset et al., 2001) on the basis of the coordinates resulted from the molecular alignment at the first step of the rigid body minimization. Open triangles and squares represent the values for CAD and ICAD, respectively.

created from one of the four original arrangements. A total of 100 sub-arrangements was used as the initial coordinates for the subsequent docking simulation.

The saturation-donor molecule was then fixed on the coordinate for the docking, and the acceptor molecule was moved to minimize a target function, E_{RB}, by translation with three degrees of freedom in the Cartesian coordinate. The conventional Powell minimization method was applied.

Here, E_{RB} is composed of E_{sat} in Equation 1 and the intermolecular van der Waals energy, E_{vdw}:

$$E_{RB} = E_{vdw} + W_{sat}^{RB} E_{sat} \big/ N_{sat} \ , \tag{5}$$

where W_{SAT}^{RB} is the weight of the E_{sat} contribution, and was set at 10^3. N_{sat} was the number of the amide protons used for each E_{sat} calculation. The force field of AMBER parm96 (Cornell et al., 1995) was used for calculations of E_{vdw}. We carried out three types of simulations: One of them used only the SAT data for CAD ($N_{sat} = 13$), another used those for ICAD ($N_{sat} = 31$), and the remaining used both of them ($N_{sat} = 44$). In the last case, the initial locations of CAD and ICAD for the rigid docking procedure were determined so that CAD and ICAD were regarded as the saturation-acceptor and donor, respectively.

When the SAT information for both the CAD and ICAD molecules was applied, the orientation-C among the initial relative arrangements provided distinctively small E_{sat} values as the result of the

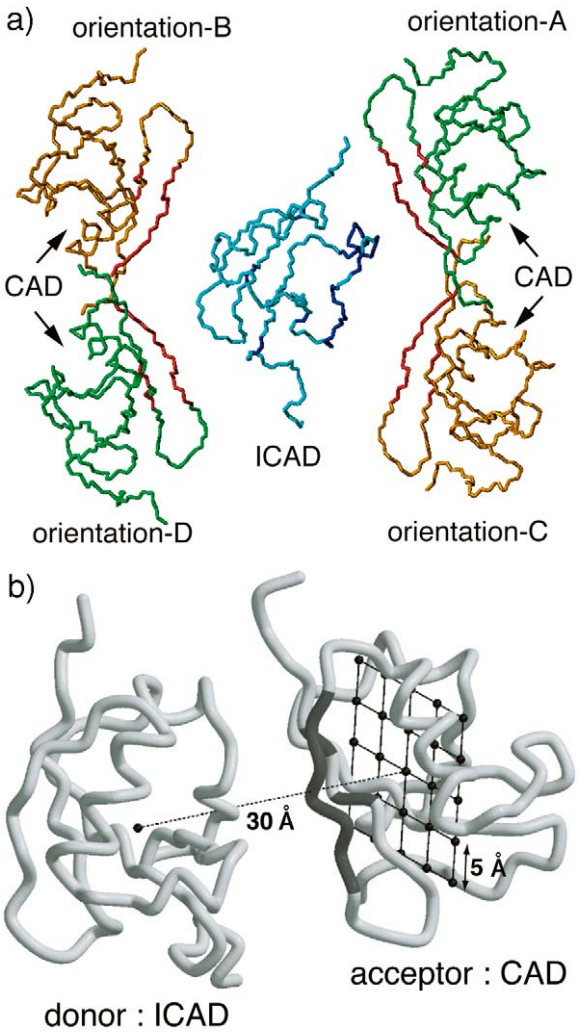

Figure 4. a) Four kinds of relative orientations generated from the results of rotational rigid body minimization using the RDC data. ICAD (colored cyan) is fixed and the relative orientations of CAD (orientations-A to D: A and D are colored green; B and C are colored orange) are represented. The residues responsible for the binding predicted from the SAT experiments, for which the peak reduction ratios were less than 0.8, are colored red and dark blue for CAD and ICAD, respectively. The orientation-C has almost the same CAD-ICAD relative orientation as that of the NOE&RDC structure. The picture was made with a program MOLMOL (Koradi et al., 1996). b) An initial arrangement for the docking simulation, when CAD is regarded as a SAT acceptor. Residues for which the SAT experimental peak reduction ratios of amide protons were less than 0.8 are colored black. The dashed line links the centers of gravity for both molecules with the distance between them of 30 Å. Interval of the squared lattice is 5 Å. Sub-arrangements were generated by translating the displayed CAD, such that its center of gravity corresponded to each lattice point. When the SAT data for both molecules were used, CAD was regarded as a saturation-acceptor and ICAD a saturation-donor. The picture was made with a program Molscript (Kraulis, 1991).

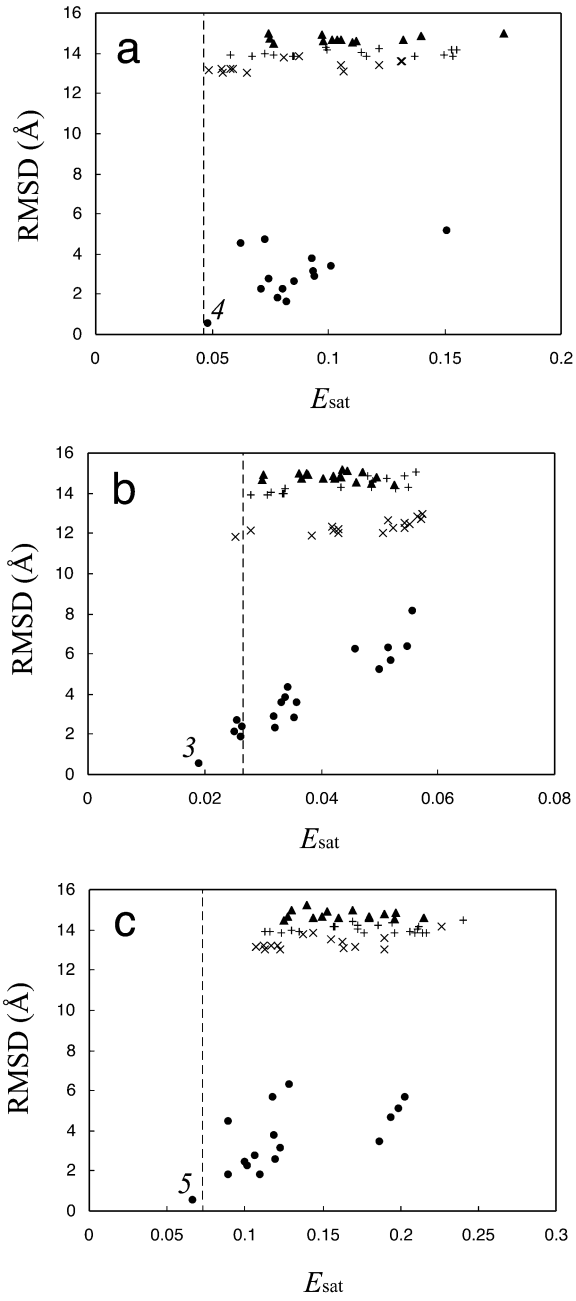

Figure 5. Plots showing the correlation between the target function E_{sat} and RMSD from the NOE&RDC structure. Results of the translational rigid body minimization, using the SAT data of (a) only CAD, (b) only ICAD and (c) both of them, are shown. Each of the three types of simulation was started from the four relative orientations shown in Figure 4a (▲, A; +, B; ●, C; and ×, D), each of which included 25 sub-arrangements shown in Figure 2b. Broken lines indicate the E_{sat} values for the NOE&RDC structure for each case: (a) 0.046, (b) 0.027 and (c) 0.073. The number of the overlapped points is indicated for the cluster converged to almost the same position with the smallest E_{sat} and RMSD.

simulation, as shown in Figure 5c, which implied that, the orientation-C arrangement should be correct among the four possible ones. In fact, as shown in Figures 5a–c, about half of the initial arrangements in the orientation-C always converged to structures similar to the reference structure NOE&RDC, with the root mean square deviations (RMSD) from the reference structure of less than 3.0 Å for the backbone heavy atoms.

The resultant structure that yielded the smallest E_{sat} energy in the docking simulation using the SAT data for both molecules had an RMSD of 0.56 Å from the reference structure (NOE&RDC), as shown in Figure 7a. In Figures 2a and b, the filled circles stand for the peak reduction ratios for the individual residues simulated for the SAT experiments based on this structure, and they are consistent with the experimental values.

Structure refinement by molecular dynamics simulations

To cope with possible small structural changes that may arise during the complex formation, molecular dynamic simulations including all atoms with the restraints based on the RDC and SAT experimental data were performed with a simulated annealing method. The structures provided by the rigid body docking using the SAT data for both CAD and ICAD molecules, for which the E_{sat} energies were within the lowest five, were applied to the simulation, and five structures were independently constructed from each of the five starting structures. Among the total of 25 resultant structures, we selected 10 structures that had the lowest total energies including E_{sat} and E_{dip}. The structures and RMSD statistics are indicated in Figure 7b and Table 1.

For these 10 structures, the average number of the violations in a total of 2383 intramolecular distance restraints applied was 1.2 ± 0.4, and that in a total of 87 dihedral angle restraints applied was 0.9 ± 0.3 per molecule. The RMSD values of these violations from the intramolecular NOE and dihedral angle restraints applied, calculated with XPLOR, were 0.03 ± 0.02 Å and $0.99 \pm 0.10°$, respectively. These results indicate that the global structure of each protein was almost maintained in the MD simulation.

The average number of the violations in 73 intermolecular NOEs was 3.1, and the RMSD from the restraints was 0.20 ± 0.08 Å. Thus, although we used no intermolecular NOE information, most of the intermolecular restraints were satisfied in the final complex models.

For this final molecular dynamics simulation step the full relaxation method was used, in which most of the computation time was spent. The MD simulation from one initial structure was completed within about 12 h when using one CPU of SGI origin 3400. Since the candidates of the complex structures were already selected by rigid body minimization, we carried out only a few simulations. Thus, the time consumed for the full relaxation matrix method was not critical in the current procedure.

Table 1. Structural statistics for the ensemble of 10 structures

Violations		
NOE restraints (Number > 0.5 Å)		
Intramolecular NOE restraints (±sd)	1.2	± 0.4
Intermolecular NOE restraints (±sd)	3.1	± 1.5
Dihedral angles (±sd) (Number < 5°)	0.90	± 0.32
RMSD (±sd) from experimental restraints		
Intramolecular NOE restraints (Å)	0.0306	± 0.0016
Intermolecular NOE restraints (Å)	0.200	± 0.082
Dihedral angles (°)	0.993	± 0.100
Residual dipolar couplings (Hz)	2.48	± 0.09
Q-factor[a]	0.24	± 0.01
Contacting surface area		
ΔASA[b] (Å^2)	376	± 13
PROCHECK Ramachandran plot statistics[c]		
Most favored regions (%)	72.8	
Additionally allowed regions (%)	22.7	
Generously allowed regions (%)	3.2	
Disallowed regions (%)	1.3	
RMSD (±sd) from the NOE&RDC-MD structure		
Backbone heavy atoms (Å)		
All	1.22	± 0.18
Rigid regions[d]	0.666	± 0.103

[a]The quality (Q) factor calculated according to Cornilescu et al. (1998), comparing the agreement between experimental dipolar coupling and those predicted from the structure.
[b]Difference between the water-accessible surface areas for the residues on the binding surface where the peak reduction ratios of the SAT experiment are less than 0.7.
[c]The PROCHECK program (Lawskowski et al., 1993) was used to evaluate the quality of the structures.
[d]The regions having the secondly structures commonly among the best 10 structures and the NOE&RDC-MD structure. Residues 9-14, 21-26, 29-39, 49-51, 57-58, 63-65 and 72-74 in CAD, and 19-24, 32-36, 39-49, 60-63, 75-78 and 84-88 in ICAD.

Docking simulation using the individually determined monomer structure of the CAD

In order to assess validity of the current docking system against the initial structures in which side-chain conformations were different from those in the complex structure, we performed docking simulation using the previously determined monomer structure of the CAD molecule (Uegaki et al., 2000) instead of the CAD in the complexed form. Docking procedure was similar to that described above. For the second step of rigid body minimization, SAT data for both CAD and ICAD were used.

After the rigid body minimization, the top eight arrangements of those yielding the best 10 E_{sat} values had correct alignments and four of them showed RMSD of 2.0 Å from the NOE&RDC structure (Figure 8a). The top three (cluster I) and the next top four (cluster II) were converged independently. However, the 9th and 10th arrangements had incorrect alignments and the difference between E_{sat} of the 8th and 9th structures was relatively small (0.091: 8th; 0.093: 9th) (Figure 8a). Although the evaluation using E_{sat} alone was not enough to discriminate correct structures from wrong ones due to conformational differences between the side-chains in the free and docked states, another index of ΔASA could give a clear distinction (Figure 8b).

Following the rigid body minimization, we performed MD simulation by starting with the arrangements with the best six E_{sat} values, half of them belonging to cluster I and another half to cluster II. Five different initial velocities were supplied to each starting condition and the resulting arrangements included in the best five total energies were independently analyzed for each cluster. For cluster II, the average number of the violations observed for the intermolcular NOE restraints (6.4 ± 3.3) was less than 10% of the total number applied (73) and RMSD from the NOE&RDC-MD structure was similar to the results of the simulation using each part of the complexed structure (Table 2 and Figure 9b). Therefore these structures are promising candidates of the complex.

Table 2. Structural statistics for the ensemble of 5 resultant structures of the MD simulation following the rigid body minimization using the independently determined monomer structure of CAD

Cluster I		
Violations		
NOE restraints (Number > 0.5 Å)		
Intramolecular NOE restraints (±sd)	0.0	± 0.0
Intermolecular NOE restraints (±sd)	31.8	± 2.0
Dihedral angles (±sd) (Number > 5°)	2.20	± 0.44
RMSD (±sd) from experimental restraints		
Intramolecular NOE restraints (Å)	0.0239	± 0.0014
Intermolecular NOE restraints (Å)	1.72	± 0.25
Dihedral angles (°)	2.35	± 0.70
Residual dipolar coupling (Hz)	2.44	± 0.02
Q factor	0.24	± 0.01
Contacting surface area		
ΔASA (Å^2)	317	± 20
PROCHECK Ramachandran plot statistics		
Most favored regions (%)	73.5	
Additionally allowed regions (%)	22.7	
Generously allowed regions (%)	2.2	
Disallowed regions (%)	1.5	
RMSD (±sd) from the NOE&RDC-MD structure		
Backbone heavy atoms (Å)		
All	2.43	± 0.25
Rigid region	1.96	± 0.23
Cluster II		
Violations		
NOE restraints (Number > 0.5 Å)		
Intramolecular NOE restraints (±sd)	1.0	± 0.7
Intermolecular NOE restraints (±sd)	6.4	± 3.3
Dihedral angles (±sd) (Number > 5°)	2.40	± 0.54
RMSD (±sd) from experimental restraints		
Intramolecular NOE restraints (Å)	0.0267	± 0.0033
Intermolecular NOE restraints (Å)	0.37	± 0.15
Dihedral angles (°)	2.28	± 0.65
Residual dipolar coupling (Hz)	2.49	± 0.15
Q factor	0.25	± 0.02
Contacting surface area		
ΔASA (Å^2)	369	± 15
PROCHECK Ramachandran plot statistics		
Most favored regions (%)	71.9	
Additionally allowed regions (%)	24.6	
Generously allowed regions (%)	2.5	
Disallowed regions (%)	1.0	
RMSD (±sd) from the NOE&RDC-MD structure		
Backbone heavy atoms (Å)		
All	1.47	± 0.15
Rigid region	0.798	± 0.134

Discussion

Comparison with the docking simulation using chemical shift perturbation

Recently, a combination of chemical shift perturbation and RDC data for molecular docking simulation

has also been developed by several groups (Clore and Schwieters, 2003; Dobrodumov and Gronenborn 2003; Dominguez et al., 2003; McCoy and Wyss, 2002). Although the sophisticated method can predict extremely accurate complex structures, it includes a visual inspection process to determine which residues are within the binding surface. This process is inevitable, because the predicted interaction residues often contain false positives in chemical shift perturbation. One of the reasons is that, the magnitudes of chemical shift perturbations may be largely affected during complex formations by small orientational changes in aromatic side chains, which can lead to large chemical shift perturbations by the ring current effects. In fact, such unexpected perturbations were also observed in our previous chemical shift perturbation study for the CAD surface of the CAD-ICAD complex as shown in Figure 10c (Uegaki et al., 2000; Otomo et al., 2000). The corresponding intermolecular NOEs were never observed, and those residues were not close to ICAD (M1, K9, R14, G36, V38, R39, F40 and N68) (see Figures 10a and b). Moreover, the forces to attract the molecular pairs in the above method are based on the artificial target functions, which have no theoretical background.

In our study, we have performed protein-protein docking simulations using RDC and SAT data. The SAT data successfully predicted the binding surface, and the peak reductions directly reflected the corresponding proton-proton distances (Figure 10d). The SAT data were introduced in the form of a pseudo-potential energy, which was proportional to the square of the difference between the experimental data and those back-calculated on the complex structure in each simulation step, by solving the Bloch equation describing the state of the saturation transfer among spins. Because this potential energy strictly depends on the distances between two associated components, it serves as a theoretical binding force by numerically setting up a threshold that selects significant SAT data.

Evaluation of the complexed model structures

The predictions of complex structures through the simulation using SAT data from either component were found to be difficult. When the SAT data for only CAD or ICAD were used, the boundary of the E_{sat} value that should separate nearly correct structures from wrong ones became ambiguous (Figures 5a, b). E_{sat} values will decrease as long as the residues receiving cross-saturation on the saturation acceptor molecule face the donor molecule, and they do not depend so much on which residues on the donor face the acceptor. Thus, E_{sat} values do not always become the smallest at the correct structures. On the other hand, the usage of the SAT data for both molecules at the same time makes both interaction regions face each other, and clearly separates the correct arrangement from the others by the indication of E_{sat} (Figure 5c). Accordingly, reliable predictions of complex structures require the SAT data for both molecules.

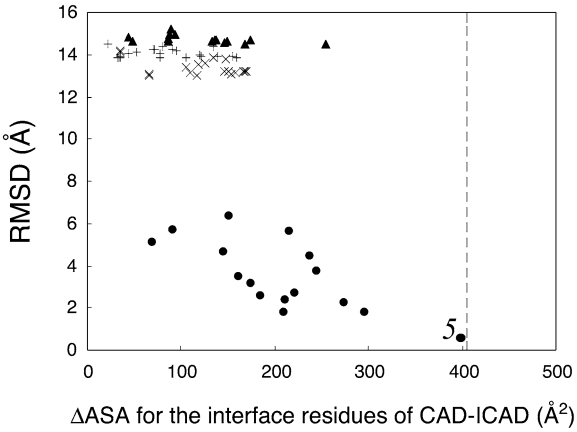

Figure 6. Plots showing the correlation between ΔASA (the difference of the water-accessible surface areas for the residues on the binding surface for which the peak reduction ratios were less than 0.7) and the RMSD from the NOE&RDC structure. Results of the translational rigid body minimization using the SAT data for both molecules are represented. Notations of the plotted points are the same as in Figure 5. The broken line indicates the ΔASA calculated from the NOE&RDC structure: 405 Å^2. The number of the overlapped points is indicated for the cluster converged to almost the same position where the ΔASA is the largest and the RMSD is the smallest.

When the SAT data are available for both molecules, an index other than E_{sat} can also be used to exclude the cases where the interaction regions do not apparently come face to face. One of the most effective indices is ΔASA, which represents the difference between the water-accessible surface areas in the free and complex states within the regions in which the amide peaks were quite saturated by SAT. Larger ΔASA values indicate larger interaction areas; thus the complex structures in such cases can be expected to be closer to the correct ones. In contrast, if the saturated regions do not contribute to the interaction at all, the summation of ΔASA for both molecules becomes 0. This was actually observed in the ΔASA of ICAD, which was calculated for the amide peaks saturated to <70% intensities, in all the final structures where the interaction regions did not face each

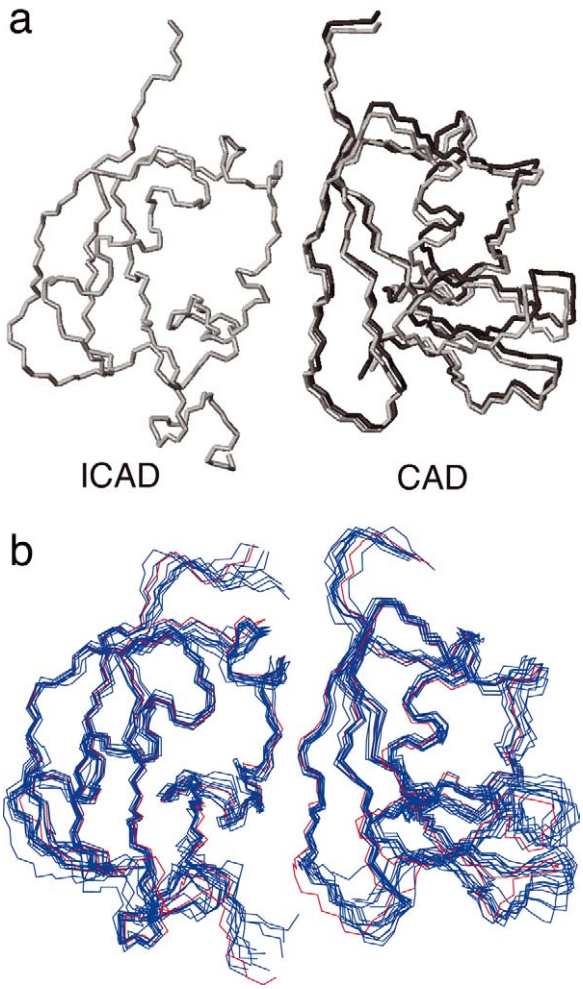

Figure 7. a) The complex structure with the best-scored arrangement after the rigid body docking simulation using the SAT data for both molecules (dark gray lines). Only the backbone structure is shown. The ICAD molecule is superimposed on that of the NOE&RDC structure (light gray lines). The RMSD of the backbone heavy atoms of CAD is 0.56 Å. The figure was made with the program MOLMOL (Koradi et al., 1996). b) The final 10 structures (blue lines) superimposed on the refined reference structure (NOE&RDC-MD: red lines).

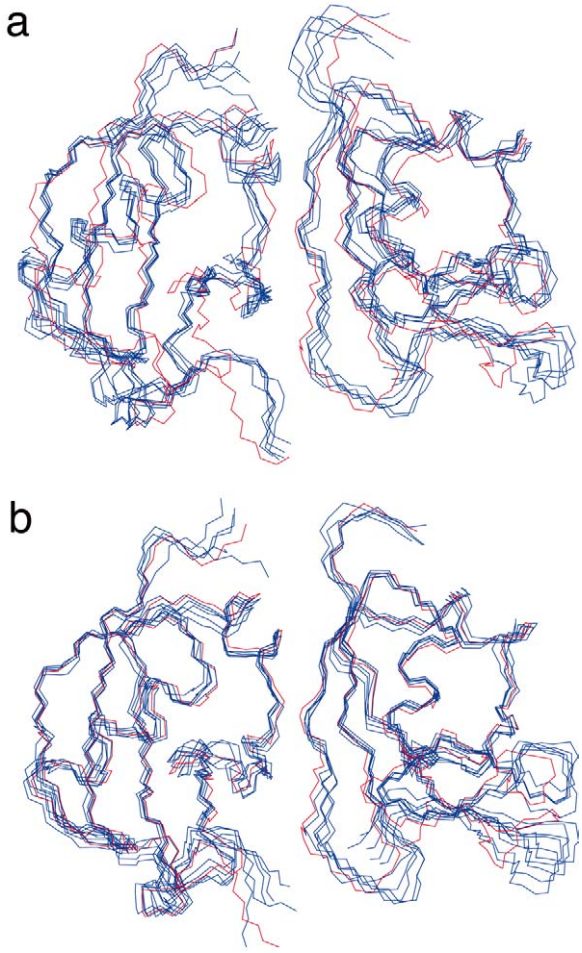

Figure 9. The final 10 structures (blue lines) started from the arrangements in (a) cluster I and (b) cluster II are superimposed on the refined reference structure (NOE&RDC-MD: red lines). The figures were made with the program MOLMOL (Koradi et al., 1996).

other (orientation-B and -D). Although orientation-A consisted of pairs of the molecules whose saturated regions face each other in a wrong way, a combinational use of large E_{sat} and relatively small ΔASA values successfully excluded orientation-A from the correct orientation-C (Figure 6). And, the summation of ΔASA for the two molecules and the RMSD from the correct structure were so well correlated even among the final structures of orientation-C.

Moreover, when the CAD monomer structure was used for simulation, the ΔASA value for cluster II (≈343 Å^2) was larger than that for cluster I (≈276 Å^2), and the former (cluster II, RMSD ≈ 2.0 Å) was more similar to the NOE&RDC structure than the latter (cluster I, RMSD ≈ 3.2 Å), showing that ΔASA can be a reliable index. Therefore, ΔASA is expected to be one of the best indices for selecting the correct structures.

The cases where the binding surface on one of the molecules in the complex does not face the other can be screened out at the stage of the initial arrangement generation. We defined a vector extending from the center of gravity to the center of the binding surface shown by the SAT experiment for a saturation acceptor molecule and placed the other (saturation donor) molecule so that its center of gravity should be placed on an extension of the vector. If SAT data for both

molecules are available, such a vector can be defined for each of them. If the binding surfaces on both molecules are facing each other, the inner product of the normalized vectors for these molecules should be smaller than 0 and near −1.0. The inner products calculated for the initial arrangements of this study were −0.73, 0.89, −0.86 and 0.70 for the orientations-A to -D, respectively. The values of the orientations-B and -D are larger than 0, and the ΔASA values calculated after the docking simulation were also very small. Thus, the initial arrangements having the inner product larger than 0 can be discarded.

Comparison between the complexed structures by the current method and those using intermolecular NOEs

To compare the final 10 structures calculated using SAT and RDC data with those determined using intermolecular NOEs and RDC, we also performed the same MD procedure starting with the NOE&RDC structure to generate five different structures, all of which had no violations in the intermolecular NOE restraints. We selected the best structure with the lowest energy, and call it NOE&RDC-MD. The RMSD between the final 10 structures and the NOE&RDC-MD structure was 1.22 ± 0.18 Å and 0.67 ± 0.10 Å, for the backbone heavy atoms of the full regions and for the regions having the secondary structures, respectively (Table 1). The final 10 structures were derived from the simulation in which the separated components of the complex were used. Although we used no paired intermolecular NOE information for the former, its global fold was very similar to that of NOE&RDC-MD and all the secondary structures were located almost at the same positions (Figure 7b). Since at first we docked two monomer structures, each of which was determined as a part of the complex structure, we applied the current simulation to the system using previously determined CAD monomer structure. After the rigid body simulation, the resultant seven structures, which had the lowest E_{sat} energy, were converged to the two clusters. These structures were applied to the MD simulation, and the resultant structures of cluster II were also similar to the NOE&RDC-MD structure (RMSD from the NOE&RDC-MD structure was 1.47 ± 0.15 Å and 0.798 ± 0.134 Å, in relation to the backbone heavy atoms of the full regions and of the regions having the secondary structures, respectively (Table 2)). Interface of the complex is similar in the NOE&RDC-MD and the model. For the residues which had intermolecular

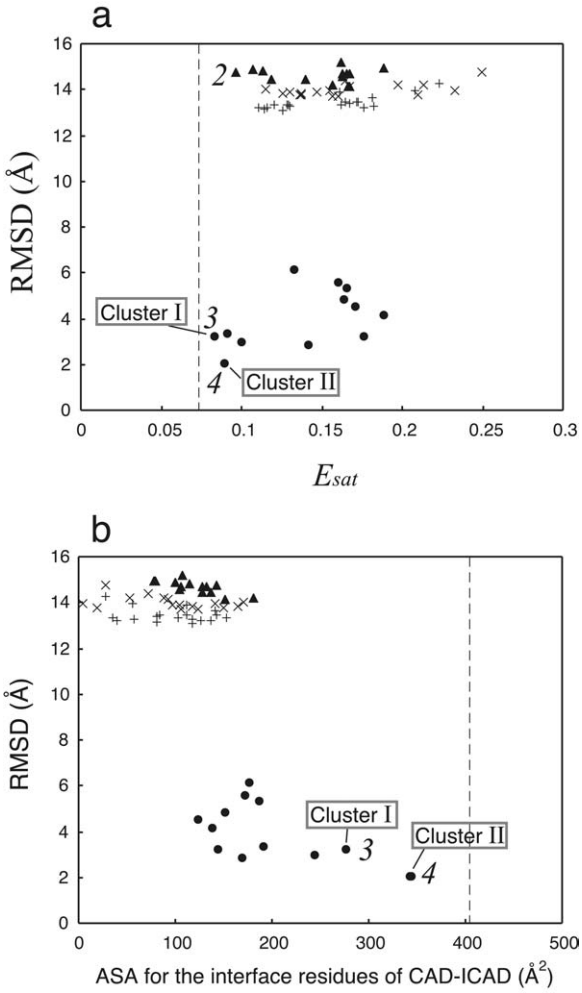

Figure 8. Plots showing the correlation between (a) the target function E_{sat} and RMSD from the NOE&RDC structure, and (b) ΔASA and RMSD. Results of the translational rigid body minimization in the second step using the monomer structure of CAD and the structure of ICAD extracted from the complex structure are shown. The simulations utilized the SAT data from both CAD and ICAD. Simulations were started from the four kinds of relative orientations, which were almost the same as shown in Figure 4a (▲, A; +, B; •, C; and ×, D). Broken lines indicate the (a) E_{sat} values or (b) ΔASA from the NOE&RDC structure. The number of the overlapped points is indicated for the clusters for which E_{sat} was less than 0.1 and ΔASA was larger than 200 Å^2.

atom pairs within 3.0 Å most of salt bridges, hydrophobic interactions and the hydrogen bonds (Figure 10a) were the same. The resultant structures of cluster I had many violations of the intermolecular NOE restraints, and thus these are not appropriate. However, RMSD for cluster I from the reference structure decreased (from 3.2 Å to 2.43 ± 0.25 Å) and ΔASA increased significantly (from 276 Å^2 to

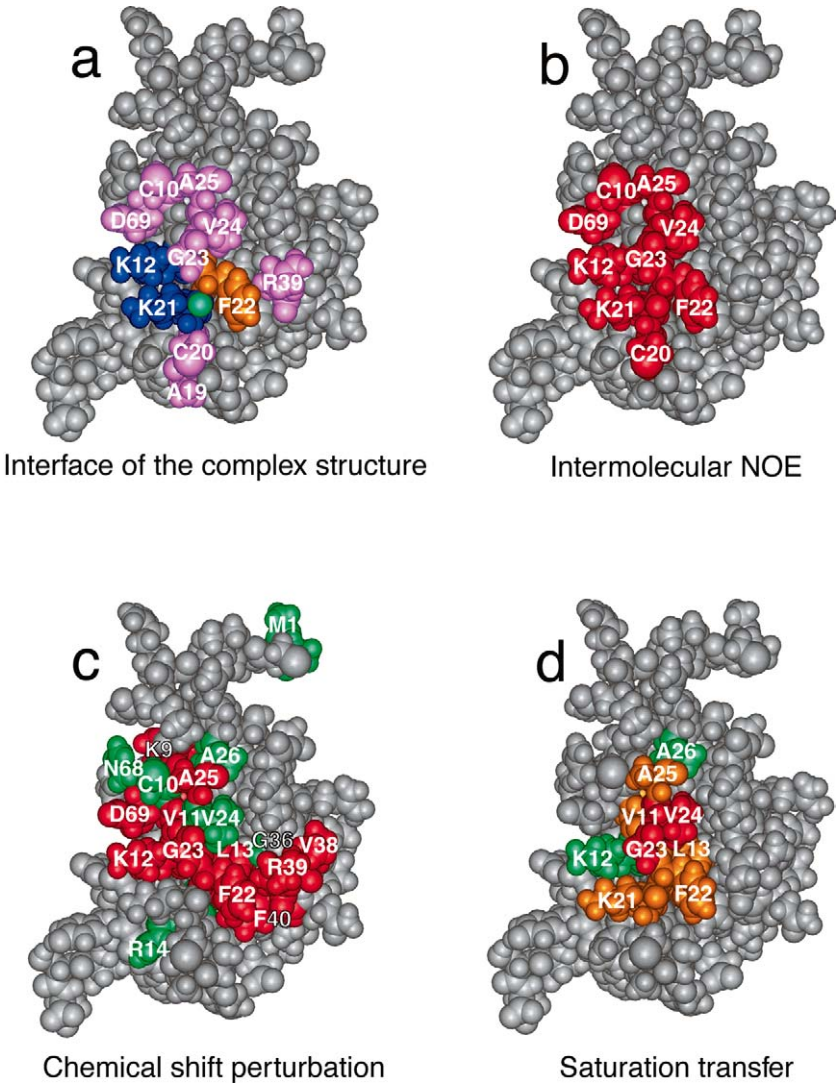

Figure 10. Comparison of the binding surfaces on CAD in the CAD-ICAD complex, of (a) previously determined complex structure, detected by (b) the intermolecular NOE assignments, (c) the chemical shift perturbation and (d) the saturation transfer. (a) The residues, which have intermolecular atom pairs within 3.0 Å, are colored purple, and especially for the residues, for which salt-bridges and hydrophobic interaction, was reported is colored blue and orange, respectively (Otomo et al., 2000). Furthermore, oxygen of carbonyl group in K21, which makes hydrogen bond with amide group of I69 of ICAD, is colored cyan (Otomo, et al., 2000). (b) Red colored residues are assigned by the intermolecular NOEs. (c) Residues showing absolute values of the corresponding chemical shift differences ($[(\Delta H_N)^2 + (\Delta N(\times 0.15)^2]^{1/2}$) of more than 0.3 ppm are colored red, and those of 0.2–0.3 ppm are colored green. (d) Residues with the intensity ratios of <0.4, 0.4–0.5 and 0.5–0.8 are colored red, orange and green, respectively.

317 ± 20 Å^2) during the MD simulation (Table 2 and Figure 9a). These results strongly suggest that the last MD simulation is valuable to bring the model closer to the correct structure. In the future, we should apply this system on other protein complexes and confirm that these convenient methods are not limited to the CAD-ICAD system.

Our method may also be helpful for the difficult cases of previous method such as when only a few intermolecular NOE restraints are available. We can also use this method to create the initial intermolecular positions previous to the other simulation or to confirm the validity of the calculated complex structures, with the information contained in the intramolecular relationship of distance from each amide proton to the protons of the other molecule.

Inquiries about the availability and distribution of the programs concerning the current protein-protein

docking system should be addressed to the corresponding author.

Conclusions

We have developed a method that simulates protein-protein docking on the basis of residual dipolar coupling (RDC) constants and saturation transfer (SAT) data. The former information basically restricts the relative orientations between the components in a complex, and the latter reveals the interaction interface. This method was applied to a molecular complex, CAD-ICAD. The RDC data were obtained for the amide groups of both molecules by aligning the complex in a liquid crystalline phase in a static magnetic field. The SAT data were also acquired for the individual molecules by mixing non-labeled either molecule and the ^{2}H, ^{15}N -labeled other, and applying radio-frequency pulses designed for the saturation. We performed rigid body docking to create the complex structure using the RDC and SAT data. After the rigid body docking, molecular dynamics simulations were performed with the restraint forces to satisfy the RDC and SAT experimental data. The results agreed well with the reference structure, which was separately obtained with the intermolecular NOE and RDC information.

Acknowledgements

The authors thank Drs Hideo Takahashi in AIST-JBIRC and Ichio Shimada in AIST-JBIRC and the University of Tokyo for their helpful advice on the saturation transfer experiments. We also wish to thank Dr Shigekazu Nagata (Graduate School of Frontier Biosciences, Osaka Univ.) for the gift of the CAD-ICAD samples.

References

Bax, A., Kontaxis, G. and Tjandra, N. (2001) In *Methods in Enzymology*, Vol. 339, James, T.L., Doetsch, V. and Schmitz, U. (Eds.), Academic Press, San Diego, pp. 127–174.

Clore, G.M. and Schwieters, C.D. (2003) *J. Am. Chem. Soc.*, **125**, 2902–2912.

Cordier, F., Dingley, A.J. and Grzesiek, S. (1999) *J. Biomol. NMR*, **13**, 175–180.

Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, Jr. K.M., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W. and Kollman, P.A. (1995) *J. Am. Chem. Soc.*, **117**, 5179–5197.

Cornilescu, G., Marquardt, J.L., Ottiger, M. and Bax, A. (1998) *J. Am. Chem. Soc.*, **120**, 6836–6837.

Delaglio, F., Grzesiek, S., Vuister, G., Zhu, G., Pfeifer, J. and Bax, A. (1995) *J. Biomol. NMR*, **6**, 277–293.

Dobrodumov A. and Gronenborn A. (2003) *PROTEINS*, **53**, 18–32.

Dobson, C.M., Olejniczak, E.T., Poulsen, F.M. and Ratcliffe, R.G. (1982) *J. Magn. Reson.*, **48**, 97–111.

Dominguez C., Boelens R. and Bonvin A.M.J.J. (2003) *J. Am. Chem. Soc.*, **125**, 1731–1737.

Dosset, P., Hus, J.C., Marion, D. and Blackledge, M. (2001) *J. Biomol. NMR*, **20**, 223–231.

Foster, M.P., Wuttke, D.S., Clemens, K.R., Jahnke, W., Radhakrishnan, I., Tennant, L., Reymond, M., Chung, J. and Wright, P.E. (1998) *J. Biomol. NMR*, **12**, 51–71.

Garret, D.S., Powers, R., Gronenborn, A.M. and Clore, G.M. (1991) *J. Magn. Reson.*, **95**, 214–220.

Güntert, P., Mumenthaler, C. and Wüthrich, K. (1997) *J. Mol. Biol.*, **273**, 283–298.

Koradi, R., Billeter, M. and Wüthrich, K. (1996) *J. Mol. Graph.*, **14**, 51–55.

Kraulis, P.J. (1991) *J. Appl. Crystallogr.*, **24**, 946–950.

Lerche, M.H., Meissner, A., Poulsen, F.M. and Sørensen, O.W. (1999) *J. Magn. Reson.*, **140**, 259–263.

Matsuda, T., Nakajima, N., Toshio, Y. and Nakamura H. (2004) *J. Mol. Recog.*, **17**, 41–50.

McCoy M.A. and Wyss D.F. (2002) *J. Am. Chem. Soc.*, **124**, 2104–2105.

Morikami, K., Nakai, T., Kidera, A., Saito, M. and Nakamura. H. (1992) *Comput. Chem.*, **16**, 243–248.

Nakajima, N., Higo, J., Kidera, A. and Nakamura, H. (2000) *J. Mol. Biol.*, **296**, 197–216.

Nakanishi, T., Miyazawa, M., Sakakura, M., Terasawa, H., Takahashi, H. and Shimada, I. (2002) *J. Mol. Biol.*, **318**, 245–249.

Nishida, N., Sumikawa, H., Sakakura, M., Shimba, N., Takahashi, H., Terasawa, H., Suzuki, E.I. and Shimada, I. (2003) *Nat. Struct. Biol.*, **101**, 53–58.

Ojennus, D.D., Mitton-Fry, R.M. and Wuttke, D.S. (1999) *J. Biomol. NMR*, **14**, 175–179.

Omichinski, J.G., Pedone, P.V., Felsenfeld, G., Gronenborn, A.M. and Colre, G.M. (1997) *Nat. Struct. Biol.*, **4**, 122–132.

Otomo, T., Sakahira, H., Uegaki, K., Nagata, S. and Yamazaki, T. (2000) *Nat. Struct. Biol.*, **7**, 658–662.

Paterson, Y., Englander, S.W. and Roder, H. (1990) *Science*, **249**, 755–759.

Press, W.H., Teukolsky, S.A., Vetterling, W.T. and Flannery, B.P. (1992) In *Numerical Recipies in C*, 2nd edn., Cambridge University Press, Cambridge, U.K., pp. 412–420.

Riek, R., Fiaux, J., Bertelsen, E.B., Horwich, A.L. and Wüthrich, K. (2002) *J. Am. Chem. Soc.*, **124**, 12144–12153.

Rueckert, M. and Otting, G. (2000) *J. Am. Chem. Soc.*, **122**, 7793–7797.

Schwieters, C.D., Kuszewski, J., Tjandra, N. and Clore, G.M. (2003) *J. Magn. Reson.*, **160**, 66–74.

Takahashi, H., Nakanishi, T., Kami, K., Arata, Y. and Shimada, I. (2000) *Nat. Struct. Biol.*, **7**, 220–223.

Tjandra, N., Grzesiek, S. and Bax, A. (1996) *J. Am. Chem. Soc.*, **118**, 6264–6272.

Uegaki, K., Otomo, T., Sakahira, H., Shimizu, M., Yumoto, N., Kyogoku, Y., Nagata, S. and Yamazaki, T. (2000) *J. Mol. Biol.*, **297**, 1121–1128.

Zweckstetter, M. and Bax, M. (2001) *J. Biolmol. NMR*, **20**, 365–377.