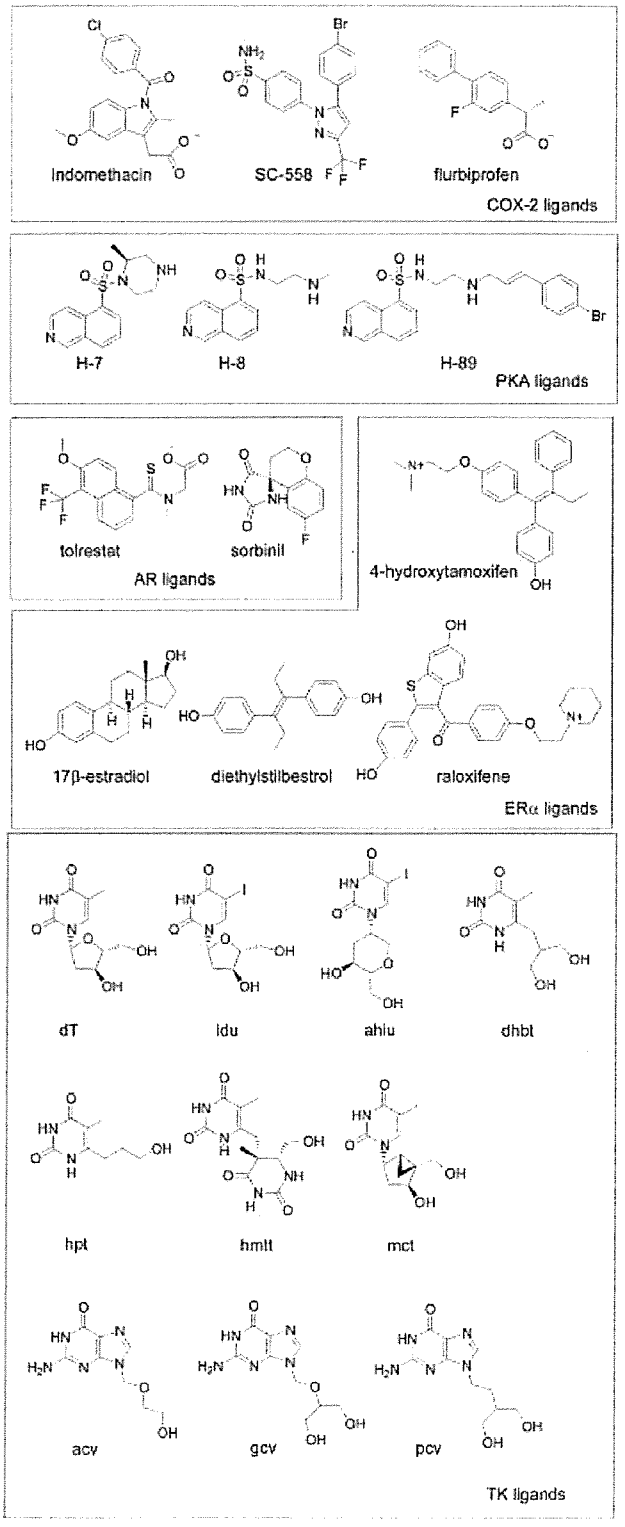



Fig. 2. The chemical structures of ligands used in the foreign-docking test.

determined by superposing the Cα atoms of the protein structure together with the foreign ligand on those of the target protein structure. It should be noted that intermolecular atomic clashes were seen in reference structures thus obtained in some cases of foreign-docking (i.e., AR02; ER03 and ER04; hmtt, mct, acv, gcv, and pcv for the TK system). Because the RMS deviation is not necessarily a sensitive barometer of the resemblance of the binding modes and might occasionally be inconsistent with researchers' intuitive evaluation, we also observed each docking model carefully on the 3D graphics to check how well the protein–ligand contacts in the corresponding crystal structure were reproduced, and to examine the merits and limitations of our approach.

## RESULTS

### Test Cases

All the test cases selected for docking validation of our approach are listed in Table I.

First, the test cases used in the evaluation of FlexX were adopted for the redocking test.[11] The test cases of FlexX include both rigid and flexible ligands, as well as several metalloproteins containing zinc or magnesium. Such protein–ligand systems are suitable for our purpose. The case of ribonuclease A and uridine vanadate was excluded, because the parameters for the vanadium atom have not been determined in the KMF parameter set.

Next, for the foreign-docking test, we chose five protein systems (22 protein–ligand cases), from the viewpoint of diversity of protein family, variety of shapes and physicochemical properties of the ligand-binding cavities, and potential value as a drug target. Although most of the foreign-docking test cases here were not used in the comparative evaluation studies of other well-known docking methods, TK has been used to validate the docking accuracies of many representative tools.[13,49,51] We used the same TK structure as was employed in the preceding validation reports (1kim, monomer A),[52] to allow direct comparison.

Among the five protein systems used in the foreign-docking test, cyclooxygenase-2 (COX-2) and cAMP-dependent protein kinase (PKA) show only small conformational differences of side chains among the cocrystal structures of test ligands. On the other hand, AR and estrogen receptor α (ERα) show substantial induced-fit motion of several side chains around the ligand-binding cavity in the cocrystal structures.[53,54] For ERα, especially large conformational differences, including movements of backbone atoms, can be observed between agonist- and antagonist-bound forms: the agonist-bound form has a small, closed binding site, while the antagonist-bound form has a wider, open binding site. Moreover, comparison of the two antagonist-bound structures (antagonists in this test: 4-hydroxytamoxifen and raloxifene) shows side-chain conformational changes of several residues that constitute the binding cavity, although the two agonist-bound structures are very similar. For TK as well, conformational changes of several side chains around the ligand-binding cavity are observed in the crystal structures of the complexes with test ligands,

**TABLE II. Closest RMS Deviations of the Two ADAM Docking Calculations Using Normal and vdW-Offset Grids**

| Code | Closest RMSD (Å) | |
|---|---|---|
| | Using normal grid | Using vdW-offset grid |
| Redocking test | | |
| 5tim | 1.19 | 1.25 |
| 1ldm | 0.52 | 0.67 |
| 2phh | 0.58 | 0.96 |
| 3ptb | 0.60 | 0.84 |
| 1ulb | 0.35 | 0.60 |
| 3tpi | 0.81 | 1.28 |
| 4tsl | 0.90 | 0.84 |
| 4dfr | 1.68 | 1.10 |
| 1stp | 0.75 | 0.71 |
| 1dwd | 1.83 | 1.77 |
| 1dwc | 1.50 | 1.29 |
| 1rnt | 0.87 | 0.93 |
| 1tmn | 1.80 | 1.39 |
| 4tln | 1.54 | 1.82 |
| 3cpa | 1.09 | 0.71 |
| 2ctc | 0.41 | 0.46 |
| 4phv | 4.47 | 1.02 |
| 121p | 1.72 | 1.28 |
| Foreign-docking test | | |
| COX01 | 1.34 | 1.17 |
| COX02 | 0.80 | 1.11 |
| COX03 | 0.76 | 1.18 |
| PKA01 | 0.78 | 0.46 |
| PKA02 | 1.05 | 0.88 |
| PKA03 | 1.26 | 0.97 |
| AR01 | 0.57 | 0.72 |
| AR02 | 0.63 | 0.67 |
| ER01 | 2.57 | 1.32 |
| ER02 | 6.21 | 0.64 |
| ER03 | 1.20 | 1.21 |
| ER04 | 5.10 | 1.98 |
| TK01 | 0.58 | 0.70 |
| TK02 | 0.56 | 0.86 |
| TK03 | 1.19 | 0.69 |
| TK04 | 0.74 | 0.83 |
| TK05 | 0.49 | 0.83 |
| TK06 | 1.19 | 0.95 |
| TK07 | 0.78 | 0.77 |
| TK08 | 1.88 | 1.34 |
| TK09 | 2.04 | 1.39 |
| TK10 | 1.11 | 1.56 |

The closest RMS deviations among the top 20 docking models are listed.

although positional changes of backbone atoms are not seen. As noted above, AR, ERα and TK should provide more challenging cases than COX-2 and PKA in this foreign-docking test.

### Effects of vdW-Offset Grid

We have shown that the vdW-offset grid contributed to successful virtual screening,[35] but we have not yet examined the effects of the vdW-offset grid on docking accuracy. Table II shows the closest RMS deviations (RMSD) among the top 20 docking models, comparing the two ADAM docking calculations using normal and vdW-offset grids. In the redocking test, the correct docking modes (i.e., RMSD < 2.0 Å) were predicted for almost all the test cases, when using either grid. The only incorrect case was the redocking calculation of HIV-1 protease/inhibitor (4phv) by using the normal grid (RMSD 4.47 Å). In the crystal structure of 4phv, the target protein has a closed binding cavity, and the large inhibitor fits there tightly. Owing to this tightness of fit, interatomic clashes would be likely to occur between protein and ligand in the docking process using the normal grid, and consequently, the correct mode would be eliminated by the energy criterion. The change of internal parameters of ADAM in the current version might make such cases more problematic than formerly, because the optimization steps for resolving the clashes are truncated as premises for the postdocking BLUTO optimization. When the vdW-offset grid was substituted for the normal one, the intermolecular clashes were reduced and the correct mode was successfully obtained in the case of 4phv (RMSD 1.02 Å).

In the foreign-docking test, comparable results were obtained from the two docking calculations using the normal and vdW-offset grids, for the cases of COX-2 and PKA, where the side-chain conformational changes are small for the test ligands, and even for the cases of AR and TK with larger induced-fit movements. However, for the case of ERα, with a drastic induced-fit, remarkable differences of the closest RMS deviations were seen between the two calculations. In particular, the effects of the vdW-offset grid were well demonstrated in the cases of 17β-estradiol (ER02) and raloxifene (ER04). Although the agonist estradiol is a ligand of small size, only incorrect modes where the ligand orientation was inverted from that of the cocrystal mode could be obtained by using the normal grid (RMSD 6.21 Å). The target ERα structure in antagonist-bound form (3ert) has an open ligand-binding cavity, but the region to which the agonist binds is somewhat narrower than that in the agonist-bound structure. This may account for the incorrect results in the case of estradiol. Also, in the case of raloxifene, the head moiety including benzothiophene and the phenol ring could not be properly fitted to the agonist-bound site of 3ert when the site was represented by the normal grid (RMSD 5.10 Å). By adopting the vdW-offset grid, the docking accuracy was improved dramatically for both estradiol (RMSD 0.64 Å) and raloxifene (RMSD 1.98 Å). These results illustrate the advantages of the vdW-offset grid in the treatment of induced-fit motion of the target protein.

As for the speed of ADAM docking calculations using the two grids, the CPU times were 1–72 s for the normal grid and 1–115 s for the vdW-offset grid, on our Linux machine (Intel Xeon CPU 3.06 GHz). The time was somewhat longer using the vdW-offset grid compared to the normal grid, because the protein cavity to be searched becomes wider when the vdW-offset strategy is adopted. However, the ADAM calculation was completed in a practical time in both cases. Moreover, in the application to virtual screening, the average computational time per molecule is greatly reduced, because most compounds in databases cannot bind stably to the target protein and the docking trials for

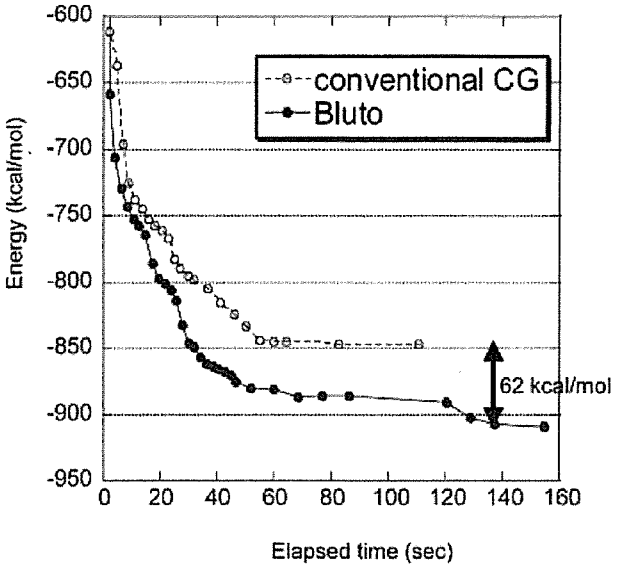



Fig. 3. Energy convergences in BLUTO and conventional CG optimization. In both optimizations, the same AMBER94 force field function and parameters were used.

such compounds are stopped at an early stage of calculation.

### Test of BLUTO Convergence

Prior to examination of the effects of combining the optimization process including protein and ligand atoms with the vdW-offset grid approach, we tested the performance of our new structure-optimization tool, BLUTO. Energy minimization of the PDB structure of bovine pancreatic trypsin inhibitor (BPTI; 1bpi)[55] was performed by using BLUTO and by using the conventional conjugated gradient (CG) method in the AMBER5.0 program.[56] In this test, all water molecules were excluded, and all of the 892 protein atoms contained in 58 residues were made movable in the optimization process. The dielectric constant was set at 1.0 r and nonbond cutoff at 10.0 Å, and the criterion of convergence was RMS <0.01 in terms of energy gradient. The calculation was executed on an SGI workstation (CPU 400 MHz R12000). Figure 3 compares the energy convergences in BLUTO and conventional CG optimization, both of which use the same AMBER94 force field function and parameters. The two methods were able to rapidly stabilize the energy value of BPTI from the initial state (207.4 kcal/mol), but the final minimized energy was 62 kcal/mol more stable in the BLUTO optimization (−909 kcal/mol) than in the conventional CG optimization (−847 kcal/mol). Moreover, BLUTO was superior in terms of the speed of energy reduction. To reach the minimized energy value with the conventional CG (−847 kcal/mol) took about 55 s using AMBER5.0, whereas BLUTO required only 30 s. In other words, the BLUTO method was able to stabilize the protein structure with only 55% of the elapsed time of conventional CG. Proteins (and protein–ligand complexes) cases other than BPTI showed a similar tendency (data not shown).

**TABLE III. RMS Deviations (in Å) for ADAM Docking Models and BLUTO-Optimized Models**

| | ADAM docking models using vdW-offset grid | | BLUTO-optimized models | |
|---|---|---|---|---|
| Code | Closest RMSD | RMSD of top-ranked model | Closest RMSD | RMSD of top-ranked model |
| Redocking test | | | | |
| 5tim | 1.25 | 1.62 | 1.39 | 1.41 |
| 1ldm | 0.67 | 1.60 | 0.34 | 0.43 |
| 2phh | 0.96 | 1.27 | 0.57 | 0.57 |
| 3ptb | 0.84 | 0.88 | 0.26 | 0.27 |
| 1ulb | 0.60 | 0.74 | 0.55 | 0.61 |
| 3tpi | 1.28 | 1.63 | 0.32 | 0.32 |
| 4tsl | 0.84 | 0.86 | 0.46 | 0.61 |
| 4dfr | 1.10 | 1.28 | 1.05 | 1.05 |
| 1stp | 0.71 | 1.13 | 0.42 | 0.42 |
| 1dwd | 1.77 | 3.92 | 1.27 | 1.53 |
| 1dwc | 1.29 | 1.29 | 0.81 | 1.99 |
| 1rnt | 0.93 | 1.01 | 0.41 | 0.92 |
| 1tmn | 1.39 | 2.05 | 0.87 | 8.99 |
| 4tln | 1.82 | 2.20 | 1.92 | 2.07 |
| 3cpa | 0.71 | 1.77 | 0.97 | 1.08 |
| 2ctc | 0.46 | 1.13 | 0.41 | 0.47 |
| 4phv | 1.02 | 1.02 | 0.63 | 0.63 |
| 121p | 1.28 | 1.31 | 1.08 | 1.08 |
| Foreign-docking test | | | | |
| COX01 | 1.17 | 1.30 | 0.66 | 0.91 |
| COX02 | 1.11 | 1.30 | 0.71 | 0.71 |
| COX03 | 1.18 | 2.02 | 0.86 | 1.98 |
| PKA01 | 0.46 | 0.76 | 0.65 | 0.76 |
| PKA02 | 0.88 | 0.98 | 0.68 | 0.78 |
| PKA03 | 0.97 | 1.49 | 1.03 | 1.21 |
| AR01 | 0.72 | 1.38 | 0.41 | 0.43 |
| AR02 | 0.67 | 0.67 | 0.44 | 0.44 |
| ER01 | 1.32 | 1.32 | 0.59 | 1.29 |
| ER02 | 0.64 | 1.28 | 0.32 | 0.63 |
| ER03 | 1.21 | 1.61 | 1.09 | 2.66 |
| ER04 | 1.98 | 4.97 | 1.06 | 1.06 |
| TK01 | 0.70 | 0.76 | 0.53 | 0.75 |
| TK02 | 0.86 | 0.86 | 0.44 | 0.63 |
| TK03 | 0.69 | 0.69 | 0.49 | 0.52 |
| TK04 | 0.83 | 4.22 | 0.76 | 0.92 |
| TK05 | 0.83 | 0.83 | 0.57 | 0.97 |
| TK06 | 0.95 | 1.14 | 0.95 | 1.42 |
| TK07 | 0.77 | 0.80 | 0.67 | 0.88 |
| TK08 | 1.34 | 2.63 | 1.35 | 1.89 |
| TK09 | 1.39 | 2.51 | 1.28 | 1.83 |
| TK10 | 1.56 | 6.31 | 1.35 | 1.70 |

The closest RMS deviations were selected from among the top 20 docking models.

### Effects of BLUTO Optimization

Next, we examined the effects of BLUTO optimization on the docking accuracy for various protein–ligand sets, especially those in which induced-fit is anticipated. Table III compares the RMS deviations of docking models that were subjected to the BLUTO optimization with those of ADAM docking models without BLUTO optimization. The closest RMS deviations of ADAM docking models were

satisfactory in all test cases, owing to the use of the vdW-offset grid. However, when the vdW-offset grid is used, the proper calculation of intermolecular energies cannot necessarily be expected, as mentioned in Materials and Methods. So, it is not surprising if the correct docking model cannot be selected as the top-scoring one in several cases. In this docking validation, typical examples are the cases of ER04, TK04, and TK10 in the foreign-docking test. Figure 4 shows the docking modes in the top-ranked models of ADAM output and in the top models reranked after BLUTO optimization, for ER04 and TK10. In these cases, small interatomic clashes between protein and ligand were observed in the top-ranked models of ADAM docking. A clash between the hydroxyl group on the benzothiophene ring in raloxifene and Gly521 Cα (2.7 Å) was seen in the top-ranked ADAM model for ER04 [Fig. 4(a)], while for TK10, the two oxygens in the ligand side chain clashed with Ala168 Cα (3.0 Å) and Tyr132 Cε (2.4 Å), respectively [Fig. 4 (c)]. Also, clashes between one of the terminal hydroxyl groups in the ligand dhbt and Ile100 Cγ2 (2.7 Å) were observed for TK04. Despite these intermolecular clashes, the energy values were low, because energy calculation was carried out using grids where the vdW energy curve was shifted. By performing the BLUTO optimization in such cases involving protein motion, the intermolecular clashes in the ADAM docking models are removed, and in consequence, proper energy evaluation becomes possible. In the above cases, the top-ranked models obtained after BLUTO optimization and reranking were in agreement with the cocrystal modes, as shown in Table III and Figure 4(b) and (d). The effect of the BLUTO optimization is well illustrated in the improvement of the docking modes of top-ranked models in these cases. Furthermore, BLUTO optimization also greatly improves the docking accuracy in cases where no marked interatomic clashes are observed in the ADAM top model (i.e., TK08), and even for the redocking test cases (i.e., 1dwd). The improved fitting between protein and ligand owing to the postdocking optimization should contribute to the superior energy evaluation and the reliable discrimination of correct solutions.

Conversely, in the cases of 1tmn and ER03, the top-ranked models after BLUTO optimization showed different binding modes from the crystal structures (RMSD: 8.99 Å for 1tmn, 2.66 Å for ER03), although the correct solutions were ranked as top in the ADAM output (RMSD: 2.05 Å for 1tmn, 1.61 Å for ER03). For the ER03 case, the rather large RMSD of the top-ranked model after BLUTO arises because the double bond at the center of diethylstilbestrol was reversed from that in the cocrystal mode. The binding positions of one of the aromatic rings and two ethyl groups were almost correct, but the other aromatic ring has shifted slightly from the cocrystal mode, being influenced by the inversion of the central double bond. In the case of 1tmn, the model where the whole ligand structure was completely reversed from the correct solution was chosen as the top-ranked solution by reranking. We consider that this incorrect result is not due to a negative effect of BLUTO optimization, but rather reflects a problem with the score used for reranking. The details are discussed later.

To assess the effects of BLUTO optimization further, we examined the RMSD changes produced by the optimization (Table IV). An RMSD change in Table IV means the difference between RMSD of the top-ranked solution after BLUTO and that of the corresponding solution before BLUTO. In 58% of all the test cases, the docking accuracy was improved by the BLUTO optimization (i.e., RMS deviations were reduced by more than 0.3 Å), and there was no case where the accuracy was degraded. This result clearly shows the effectiveness of BLUTO optimization. No marked difference in improvement of accuracy was observed in the foreign-docking test (59%) versus the redocking test (56%). In the current version of ADAM, the use of looser convergence criteria to achieve faster calculation means that the docking accuracy can be improved by the postdocking optimization even in the redocking cases, as well as in the foreign-docking cases.

As for the speed of BLUTO optimization, the CPU time was 2–18 s per docking model on our Linux machine (Intel Xeon CPU 3.06 GHz). This running time is fairly short for simple optimization of a protein–ligand complex structure. In the application to virtual screening, the BLUTO optimization is performed for only a small number of docking models for hit compounds that meet the user's criteria, so the adoption of BLUTO step would not cause an explosive increase of overall running time.

### Comparison with Other Docking Methods

Figure 5 and Table V compare the docking accuracy of our method with that of the other docking tools, for which validation reports have been published. In the redocking test (Fig. 5), both ADAM and ADAM + BLUTO showed comparable accuracy to Glide,[13] and superior accuracy to FlexX[11] (as for GOLD,[12] only a part of these redocking cases has been reported; http://www.ccdc.cam.ac.uk). Also, it is clear from Figure 5 that BLUTO optimization increased the number of "accurate" solutions (RMSD <1.2 Å), compared to the ADAM output. However, as noted in Introduction, a truly useful method must demonstrate power in foreign-docking cases, which are more relevant to practical molecular design. Table V compares the docking accuracy of our combinatorial approach and other well-known docking tools for TK in the foreign-docking test. Induced-fit motion is observed in the TK system, and correct docking modes were obtained for only a part of test ligands by the other docking tools, that is, DOCK, FlexX, GOLD,[49] Surflex,[51] and Glide.[13] On the other hand, ADAM + BLUTO gave correct solutions for all TK ligands. In particular, for the purine ligands (acv, gcv, pcv), which have different core structures from the cocrystallized ligand in 1kim, the other methods failed to generate the correct docking modes, while our method provided them as top-ranking solutions. To our knowledge, our approach offers better docking accuracy for the TK cases than any other currently available docking method.

When comparing the docking validation results of various methods, we should take account of differences in the

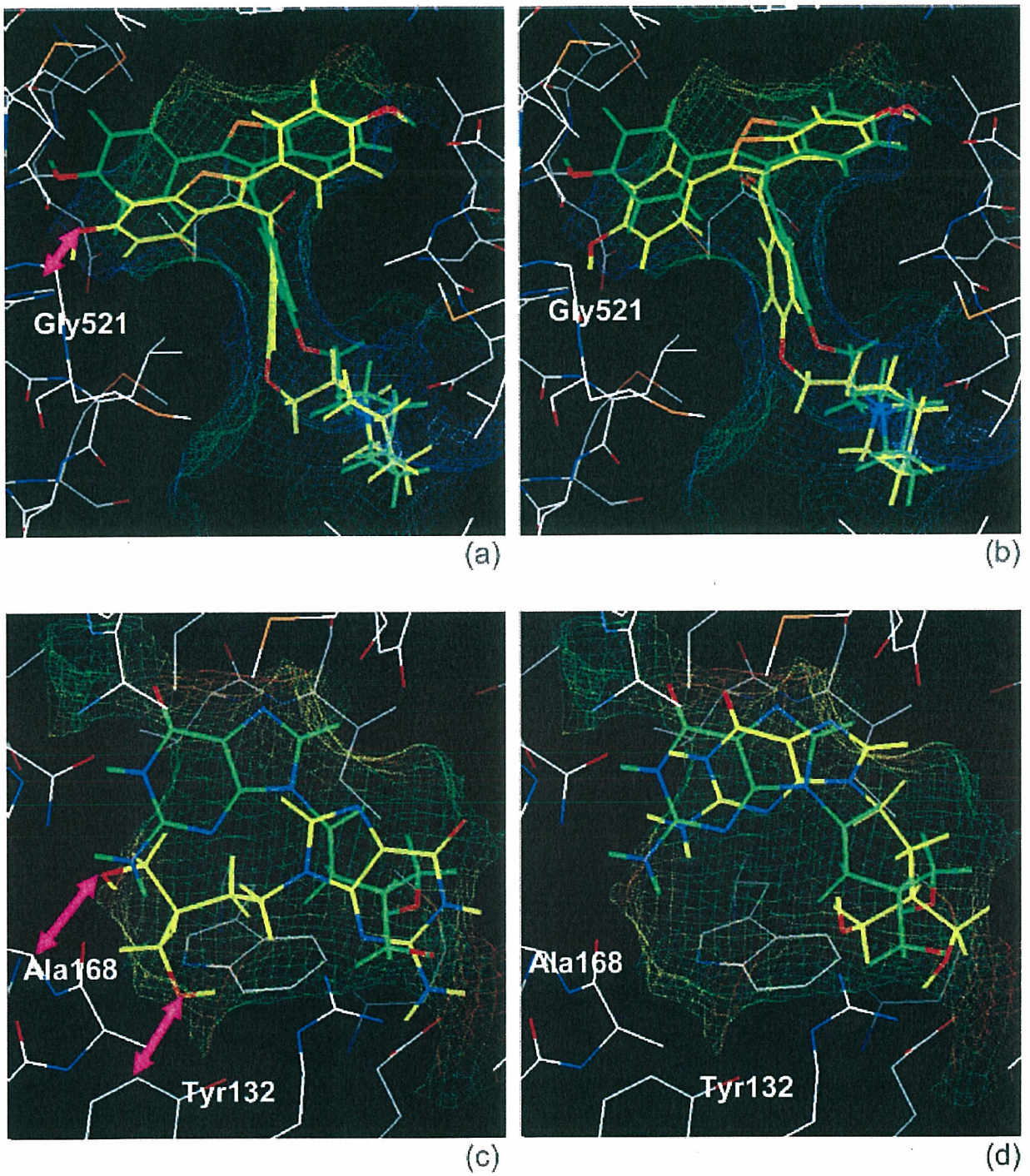



Fig. 4. Docking modes in the top-ranked models of ADAM output and in the top models reranked after BLUTO optimization, for foreign-docking of ER04 and TK10. Each docking model is compared with the reference ligand structure, which was obtained by superposition of the cocrystal structure on the target protein structure (3ert for ER04; 1kim for TK10). (**a**) Top model of ADAM docking (carbons and hydrogens are yellow) and reference structure (green) for ER04 case; (**b**) top model reranked after BLUTO optimization (yellow) and reference structure (green) for ER04 case; (**c**) top model of ADAM docking (yellow) and reference structure (green) for TK10 case; (**d**) top model reranked after BLUTO optimization (yellow) and reference structure (green) for TK10 case. The shape of the ligand-binding region is displayed as a birdcage model. The double-headed arrow in pink represents the interatomic clash observed in the docking model. Hydrogens in the protein are omitted for clarity. Note that intermolecular atomic clashes also exist between the superposed reference structure of the ligand and the target protein structure, in the cases of both ER04 and TK10 (arrows are not shown).

**TABLE IV. Rmsd Changes Produced by the BLUTO Optimization**

| | Accuracy was improved (RMSD change $< -0.3$ Å) | Not changed (RMSD change: $-0.3 \sim +0.3$ Å) | Accuracy was degraded (RMSD change $> +0.3$ Å) |
|---|---|---|---|
| Redocking test | 10 | 8 | 0 |
| Foreign-docking test | 13 | 9 | 0 |

RMSD change was calculated as follows: [RMSD of top solution after BLUTO] − [RMSD of the corresponding solution before BLUTO].

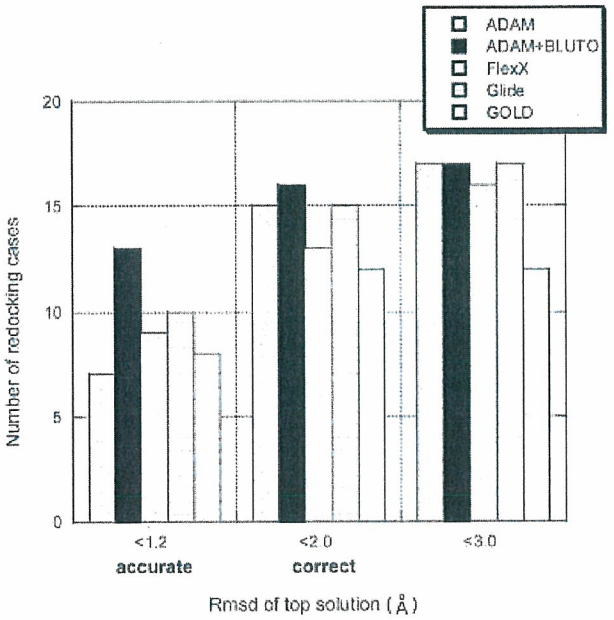



Fig. 5. Comparison of accuracy of several docking methods in the redocking test. The RMS deviations of top-ranked solutions in redocking test cases are compared, for ADAM, ADAM + BLUTO, FlexX, Glide, and GOLD. Data for FlexX, Glide, and GOLD are taken from refs. 11, 13, and 57, respectively. For GOLD, only 12 among the total of 18 redocking cases are available.

calculation conditions. One of the considerations is how the ligand input structure is prepared. In the redocking and foreign-docking tests in this article, the ligand input structures were computationally generated from the 2D chemical drawing, by using our Key3D program. On the other hand, in the redocking test of FlexX, GOLD and Glide (Fig. 5) and in the foreign-docking of TK by Glide (Table V), the ligand structures were originally taken from the PDB data (i.e., the answers of docking tests), and then those coordinates were varied by means of energy minimization and/or bond rotation (in the TK foreign-docking test of DOCK, FlexX, and GOLD, Bissantz et al. generated ligand input structures with Corina, and as for Surflex, the preparation of ligand structures is not clearly described in the article). Needless to say, such ligand preparation utilizing the answers of docking test is undesirable even if

**TABLE V. Accuracy in Foreign-Docking of TK Ligands**

| | RMSD (Å) of top solution | | | | | |
|---|---|---|---|---|---|---|
| Code (ligand) | ADAM + BLUTO | DOCK | FlexX | GOLD | Surflex | Glide |
| TK01 (dT) | 0.75 | 0.82 | 0.78 | 0.72 | 0.74 | 0.45 |
| TK02 (idu) | 0.63 | 9.33 | 1.03 | 0.77 | 1.05 | 0.35 |
| TK03 (ahiu) | 0.52 | 1.16 | 0.88 | 0.63 | 0.87 | 0.54 |
| TK04 (dhbt) | 0.92 | 2.02 | 3.65 | 0.93 | 0.96 | 0.68 |
| TK05 (hpt) | 0.97 | 1.02 | 4.18 | 0.49 | 1.90 | 1.58 |
| TK06 (hmtt) | 1.42 | 9.62 | 13.30 | 2.33 | 1.78 | 2.83 |
| TK07 (mct) | 0.88 | 7.56 | 1.11 | 1.19 | 0.87 | 0.79 |
| TK08 (acv) | 1.89 | 3.08 | 2.71 | 2.74 | 3.51 | 4.22 |
| TK09 (gcv) | 1.83 | 3.01 | 6.07 | 3.11 | 3.54 | 3.19 |
| TK10 (pcv) | 1.70 | 4.10 | 5.96 | 3.01 | 3.84 | 3.53 |

Data for DOCK, FlexX, and GOLD are taken from Bissantz et al.[49] Data for Surflex and Glide are taken from Jain[51] and Friesner et al.,[13] respectively.

the coordinates are modified, because no information on the binding geometry of a ligand would be available in advance in actual drug design. Some docking methods generate the ligand conformations without depending on the torsion angles in the input coordinates, but even for such methods, the bond lengths, angles and out-of-plane angles might still affect the docking results. In particular, when the ligand structure has nitrogen atoms with complicated orbital hybridization, the prediction of proper geometry (specifically, out-of-plane angles) around the nitrogens is not easy. Accordingly, it is expected that the use of computationally generated structures for ligand input will make the docking test more difficult. Considering the above issues, the high accuracy of our results under unfavorable conditions is impressive.

## DISCUSSION

We have developed an approach in which the vdW-offset grid and structure optimization involving the movements of protein atoms are effectively combined, to improve docking accuracy, especially in the systems with protein induced fit. The results of 18 redocking tests and 22 foreign-docking tests support the usefulness of our combinatorial approach, as shown in Figure 6. In a total of five cases (one redocking case and four foreign-docking cases) where the cocrystal modes could not be reproduced by the native ADAM process, the correct docking modes were obtained by adopting our approach. In particular, in the foreign-docking test on the ERα system, where a large induced-fit is experimentally observed, remarkable improvements in docking accuracy have been achieved by using our new approach, as described in Results. Because the handling of induced-fit motion is an inevitable issue in drug design, our achievement should be very valuable in practice.

In recent years, the redocking test has frequently been carried out for various docking tools. However, accuracy in the redocking test cannot show how helpful the docking tool would be for actual molecular design, especially for virtual screening, where a number of small molecules with

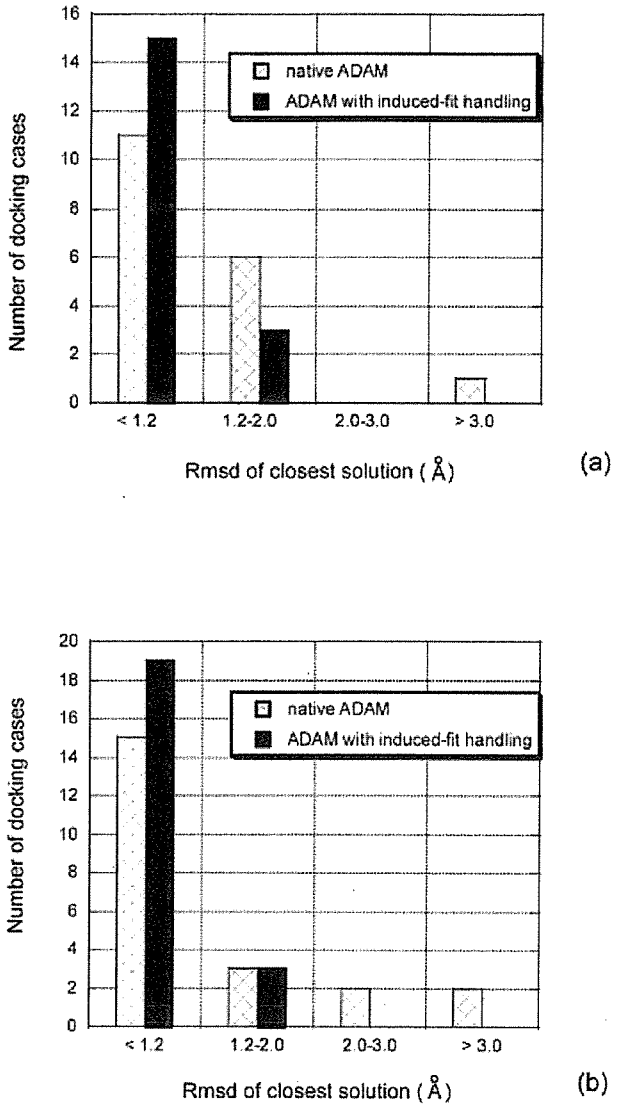



Fig. 6. Improvements of docking accuracy by our approach to handle induced fit. The RMS deviations of the closest docking models by native ADAM and by ADAM with induced-fit handling are compared for (a) redocking test, and (b) foreign-docking test.

various sizes and shapes are docked to a target protein structure. Rather, the redocking test should be regarded only as a basic functional check of the docking method. If a docking validation more closely resembling actual drug design, that is, the foreign-docking test, is adopted, the particular features and problems of each docking tool will become more apparent. Another method of docking validation is cross-docking, where the docking calculations are performed for all combinations of available protein conformers and test ligands. However, in molecular design, including virtual screening, the key point is how well a variety of ligands can be docked to a single favorable conformer of the target protein. So, we consider that the foreign-docking test using a representative conformer of each target protein is appropriate for evaluation of the usefulness of a docking tool. As for the ligand input structure, the 3D structure converted computationally from the 2D structural formula should be used for the docking assessment, instead of a structure based on the answer itself, derived from the PDB data. We believe our present docking validation directly reflects usefulness in actual drug design, because it has been performed focusing on the foreign-docking test, with the use of computationally generated structures as ligand input.

In the redocking test, our docking approach provided results superior or equal to those of other docking tools, despite having used more demanding calculation conditions. Moreover, in the case of TK with induced-fit motion, our approach gave the best results that have ever been reported. The high accuracy of our docking method is considered be due to the exhaustive and effective searching of the docking solution space by the ADAM algorithm, as well as the success in handling local induced-fit movements. The ADAM algorithm rarely fails to find the correct answer, because ADAM does not sample only discrete points in the docking solution space. Flexible docking methods other than ADAM generally perform conformational analysis of the ligand molecule in advance of the docking calculation or in the first step of docking, and a restricted number of the ligand conformations is selected to be used in the subsequent docking process. Methods of this type, for example, FlexX[11] and Glide,[13] have the common risk of getting only incorrect answers if the bound conformation is not included in the initial ligand conformation list. Surflex[51] has also a similar limitation, and in addition, the docking results seem to show dependency on the input ligand conformation. As an alternative approach, stochastic methods, which make use of the Monte Carlo method or the Genetic Algorithm, have been reported; representatives are GOLD[12] and AUTODOCK.[10] These methods also have a problem, in that the stochastic algorithm cannot fully cover the possibilities of ligand position and conformation in one docking run. Even if many runs are executed, the correct answer may not be obtained.

Our approach to handle induced-fit motion was successful in the foreign-docking test as well as the redocking test, but it is not surprising that our method also has some limitations. In our approach, drastic induced-fit motion including large movements of backbone atoms still cannot be handled. For instance, an ERα antagonist of large size cannot be docked properly to the agonist-bound protein structure, where the ligand-binding cavity is small and closed. This limitation in cases of major induced-fit is shared by many other approaches to protein flexibility, that is, methods handling only the side-chain motion and those using soft potential. A practical and simple strategy to overcome this difficulty would be to select the protein structure where the ligand-binding cavity is as wide as possible. In the foreign-docking of ERα ligands, our procedure yielded satisfactory results when we adopted the antagonist-bound form with a wider binding cavity as target protein structure. If a suitable crystal structure is not available, a protein conformation with a wider cavity

may be prepared by using the techniques of structure modeling and/or molecular dynamics on the basis of structural information about the target protein itself and/or homologous proteins, although this task may not be easy.

For further improvement in the docking accuracy, development of a reliable docking score is indispensable. In this validation study, the very simple, conventional score based on the force field energy was used. This score does not consider several important factors, such as desolvation costs and intramolecular energy changes of target protein, and this omission seems to have caused incorrect ranking of docking models in some cases. For example, the incorrect model was ranked first in the case of the 1tmn redocking test, even though a docking model that reproduced well the cocrystal mode was also generated. To tackle this difficult but important problem, we have been developing a more efficient docking score for properly discriminating the correct model. With this new score, the correct docking model was identified as the top-ranking result even for 1tmn in a preliminary test (data not shown). Further optimization of our new scoring system should lead to more accurate docking models that will give superior results in drug design.

## CONCLUSION

We have developed a simple and effective approach for handling the local induced-fit motion of the target protein in the flexible docking. To examine the effectiveness of our approach, validation tests were performed under severe conditions of ligand input structure. In the foreign-docking test cases where substantial induced-fit movements are observed in the cocrystal structures, the docking accuracy was improved remarkably by adopting our approach to treat local protein motions, and excellent results were obtained. It is noteworthy that our approach produced superior results to the other available methods in the foreign-docking test, which is directly relevant to usefulness in actual drug design. The limitation of our approach is that large conformational change of the target protein cannot be treated. In such cases, we recommend use of the protein structure in which the ligand-binding cavity is as wide as possible, selected from available crystal structures or computationally sampled conformations. We expect that the predictive power of our approach will be further improved by the incorporation of a new docking score, which is currently under development.

## REFERENCES


1. Najmanovich R, Kuttner J, Sobolev V, Edelman M. Side-chain flexibility in proteins upon ligand binding. Proteins 2000;39:261–268.
2. Ma B, Shatsky M, Wolfson HJ, Nussinov R. Multiple diverse ligands binding at a single protein site: a matter of pre-existing populations. Protein Sci 2002;11:184–197.
3. Gutteridge A, Thornton J. Conformational changes observed in enzyme crystal structures upon substrate binding. J Mol Biol 2005;346:21–28.
4. DesJarlais RL, Sheridan RP, Dixon JS, Kuntz ID, Venkataraghavan R. Docking flexible ligands to macromolecular receptors by molecular shape. J Med Chem 1986;29:2149–2153.
5. Leach AR, Kuntz ID. Conformational analysis of flexible ligands in macromolecular receptor sites. J Comp Chem 1992;13:730–748.
6. Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE. A geometric approach to macromolecule–ligand interactions. J Mol Biol 1982;161:269–288.
7. Yamada M, Itai A. Development of an efficient automated docking method. Chem Pharm Bull 1993;41:1200–1202.
8. Yamada M, Itai A Application and evaluation of the automated docking method. Chem Pharm Bull 1993;41:1203–1205.
9. Mizutani MY, Tomioka N, Itai A. Rational automatic search method for stable docking models of protein and ligand. J Mol Biol 1994;243:310–326.
10. Morris GM, Goodsell DS, Huey R, Olson AJ. Distributed automated docking of flexible ligands to proteins: parallel applications of AutoDock 2.4. J Comput Aided Mol Des 1996;10:293–304.
11. Rarey M, Kramer B, Lengauer T, Klebe G. A fast flexible docking method using an incremental construction algorithm. J Mol Biol 1996;261:470–489.
12. Jones G, Willett P, Glen RC, Leach AR, Taylor R. Development and validation of a genetic algorithm for flexible docking. J Mol Biol 1997;267:727–748.
13. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelley M, Perry JK, Shaw DE, Francis P, Shenkin PS. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J Med Chem 2004;47:1739–1749.
14. Kuntz ID. Structure-based strategies for drug design and discovery. Science 1992;257:1078–1082.
15. Shoichet BK, McGovern SL, Wei B, Irwin JJ. Lead discovery using molecular docking. Curr Opin Chem Biol 2002;6:439–446.
16. Schneider G., Böhm H-J. Virtual screening and fast automated docking methods. Drug Discov Today 2002;7:64–70.
17. Schnecke V, Swanson CA, Getzoff ED, Tainer JA, Kuhn LA. Screening a peptidyl database for potential ligands to proteins with side-chain flexibility. Proteins 1998;33:74–87.
18. Schnecke V, Kuhn LA. Virtual screening with salvation and ligand-induced complementarity. Perspect Drug Discov Design 2000;20:171–190.
19. Totrov M, Abagyan R. Flexible protein–ligand docking by global energy optimization in internal coordinates. Proteins 1997;Suppl. 1:215–220.
20. Leach AR. Ligand docking to proteins with discrete side-chain flexibility. J Mol Biol 1994;235:345–356.
21. Källblad P, Dean PM. Efficient conformational sampling of local side-chain flexibility. J Mol Biol 2003;326:1651–1665.
22. Frimurer TM, Peters GH, Iversen LF, Andersen HS, Møller NP, Olsen OH. Ligand-induced conformational changes: improved predictions of ligand binding conformations and affinities. Biophys J 2003;84:2273–2281.
23. Knegtel RM, Kuntz ID, Oshiro CM. Molecular docking to ensembles of protein structures. J Mol Biol 1997;266:424–440.
24. Claußen H, Buning C, Rarey M, Lengauer T. FlexE: efficient molecular docking considering protein structure variations. J Mol Biol 2001;308:377–395.
25. Österberg F, Morris GM, Sanner MF, Olson AJ, Goodsell DS. Automated docking to multiple target structures: incorporation of protein mobility and structural water heterogeneity in AutoDock. Proteins 2002;46:34–40.
26. Wei BQ, Weaver LH, Ferrari AM, Matthews BW, Shoichet BK. Testing a flexible-receptor docking algorithm in a model binding site. J Mol Biol 2004;337:1161–1182.
27. Broughton HB. A method for including protein flexibility in protein–ligand docking: improving tools for database mining and virtual screening. J Mol Graph Model 2000;18:247–257, 302–304.
28. Apostolakis J, Pluckthun A, Caflisch A. Docking small ligands in flexible binding sites. J Comp Chem 1998;19:21–37.
29. Sandak B, Wolfson HJ, Nussinov R. Flexible docking allowing induced fit in proteins: insights from an open to closed conformational isomers. Proteins 1998;32:159–174.
30. Zacharias M, Sklenar H. Harmonic modes as variables to approximately account for receptor flexibility in ligand-receptor docking simulations: application to DNA minor groove ligand complex. J Comp Chem 1999;20:287–300.
31. Vieth M, Hirst JD, Kolinski A, Brooks III CL. Assessing energy functions for flexible docking. J Comp Chem 1998;19:1612–1622.
32. Wu G, Robertson DH, Brooks CL III, Vieth M. Detailed analysis of grid-based molecular docking: a case study of CDOCKER—a CHARMm-based MD docking algorithm. J Comp Chem 2003:24: 1549–1562.

33. Ferrari AM, Wei BQ, Costantino L, Shoichet BK. Soft docking and multiple receptor conformations in virtual screening. J Med Chem 2004;47:5076–5084.
34. Iwata Y, Arisawa M, Hamada R, Kita Y, Mizutani MY, Tomioka N, Itai A, Miyamoto S. Discovery of novel aldose reductase inhibitors using a protein structure-based approach: 3D-database search followed by design and synthesis. J Med Chem 2001;44: 1718–1728.
35. Mizutani MY, Itai A. Efficient method for high-throughput virtual screening based on flexible docking: Discovery of novel acetylcholinesterase inhibitors. J Med Chem 2004;47:4818–4828.
36. Goto J, Kataoka R, Hirayama N. Ph4Dock: pharmacophore-based protein–ligand docking. J Med Chem 2004;47:6804–6811.
37. Makino S, Kuntz ID. Automated flexible ligand docking method and its application for database search. J Comp Chem 1997;18: 1812–1825.
38. Kellenberger E, Rodrigo J, Muller P, Rognan D. Comparative evaluation of eight docking tools for docking and virtual screening accuracy. Proteins 2004;57:225–242.
39. Kontoyianni M, McClellan LM, Sokol GS. Evaluation of docking performance: comparative data on docking algorithms. J Med Chem 2004;47:558–565.
40. Perola E, Walters WP, Charifson PS. A detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance. Proteins 2004;56:235–249.
41. Nissink JWM, Murray C, Hartshorn M, Verdonk ML, Cole JC, Taylor R. A new test set for validating predictions of protein-ligand interaction. Proteins 2002;49:457–471.
42. Erickson JA, Jalaie M, Robertson DH, Lewis RA, Vieth M. Lessons in molecular recognition: the effects of ligand and protein flexibility on molecular docking accuracy. J Med Chem. 2004;47:45–55.
43. Tomioka N, Itai A, Iitaka Y. A method for fast energy estimation and visualization of protein-ligand interaction. J Comput-Aided Mol Design 1987;1:197–210.
44. Tomioka N, Itai A. GREEN: a program package for docking studies in rational drug design. J Comput-Aided Mol Design 1994;8:347–366.
45. Liu DC, Nocedal J. On the limited memory BFGS method for large scale optimization. Math Program 1989;45:503–528.
46. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM Jr, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. J Am Chem Soc 1995;117: 5179–5197.
47. Berman HM, Battistuz T, Bhat TN, Bluhm WF, Bourne PE, Burkhardt K, Feng Z, Gilliland GL, Iype L, Jain S, Fagan P, Marvin J, Padilla D, Ravichandran V, Schneider B, Thanki N, Weissig H, Westbrook JD, Zardecki C. The Protein Data Bank. Acta Crystallogr D Biol Crystallogr 2002;58:899–907.
48. ISIS-Draw, Version 2.2.1. San Leandro: MDL Information Systems Inc.; 1999.
49. Bissantz C, Folkers G, Rognan D. Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. J Med Chem 2000;43:4759–4767.
50. MOPAC93. Japan: Fujitsu; 1993.
51. Jain AN. Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine. J Med Chem 2003;46: 499–511.
52. Champness JN, Bennett MS, Wien F, Visse R, Summers WC, Herdewijn P, de Clerq E, Ostrowski T, Jarvest RL, Sanderson MR. Exploring the active site of herpes simplex virus type-1 thymidine kinase by X-ray crystallography of complexes with aciclovir and other ligands. Proteins 1998;32:350–361.
53. Urzhumtsev A, Tête-Favier F, Mitschler A, Barbanton J, Barth P, Urzhumtseva L, Biellmann JF, Podjarny A, Moras D. A "specificity" pocket inferred from the crystal structures of the complexes of aldose reductase with the pharmaceutically important inhibitors tolrestat and sorbinil. Structure 1997;5:601–612.
54. Shiau AK, Barstad D, Loria PM, Cheng L, Kushner PJ, Agard DA, Greene GL. The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen. Cell 1998;95:927–937.
55. Parkin S, Rupp B, Hope H. Structure of bovine pancreatic trypsin inhibitor at 125 K definition of carboxyl-terminal residues Gly57 and Ala58. Acta Crystallogr D Biol Crystallogr 1996;52:18–29.
56. Case DA, Pearlman DA, Caldwell JW, Cheatham TE III, Ross WS, Simmerling CL, Darden TA, Merz KM, Stanton RV, Cheng AL, Vincent JJ, Crowley M, Ferguson DM, Radmer RJ, Seibel GL, Singh UC, Weiner PK, Kollman PA. AMBER, Version 5; San Francisco: University of California; 1997.

# Starting Point to Molecular Design: Efficient Automated 3D Model Builder Key3D

Miho Yamada MIZUTANI, Kensuke NAKAMURA,[1] Tazuko ICHINOSE, and Akiko ITAI*

*Institute of Medicinal Molecular Design, Inc.; 5–24–5 Hongo, Bunkyo-ku, Tokyo 113–0033, Japan.*
Received August 4, 2006; accepted October 2, 2006; published online October 6, 2006

**Obtaining three-dimensional (3D) structures from structural formulae is a crucial process in molecular design. We have developed a new 3D model builder, Key3D, in which the simplified distance geometry technique and structure optimization based on the MMFF force field are combined. In an evaluation study using 598 crystal structures, the high performance and accuracy of Key3D were demonstrated. In the "flexible-fitting" test, which is focused on practical usefulness in the molecular design process, 88% of the Key3D structures acceptably reproduced the reference crystal structures (root-mean-square deviation <0.6 Å) upon rotation of acyclic bonds. These results indicate that Key3D will be very effective in providing starting points for practical molecular design.**

**Key words** three-dimensional (3D) structure; drug design; Key3D; distance geometry; MMFF force field; MO calculation

Three-dimensional (3D) molecular structures are required in many fields of molecular design, such as docking simulation, virtual screening, molecular superposition and evaluation of quantitative structure–activity relationships, and methods to generate 3D structures of small molecules from two-dimensional (2D) structures play a key role in providing starting points.

A number of automated 3D model builders have been developed.[2] CORINA,[3] CONCORD,[4,5] and CONVERTER[6] are the most widely used methods. Two representative approaches to create 3D structures are a rule-based algorithm and the use of the distance geometry method. The rule-based approach is very rapid in general, and high accuracy can be expected with molecules for which all the components are incorporated in the internal rules and library. However, this approach is inevitably restricted by the number and quality of internal rules and the library size. Methods based on a distance geometry algorithm can handle a wide range of structures without the limitations of the rule-based approach, but require longer computational time.

Although various 3D model builders have been developed and subsequently improved in various respects, these methods may still not necessarily be adequate for providing starting points for actual molecular design. For instance, it has been reported that the accuracies of docking models which are generated starting from the CORINA structures are lower than those obtained using the structures optimized with the Merck molecular force field (MMFF94),[7—11] even though CORINA is one of the most sophisticated methods currently available.[12] We set out to develop a new 3D model builder with particular emphasis on the generation of high-quality 3D structures that would be suitable for molecular design. We named our method Key3D.

In developing Key3D, we assigned high priority to both accuracy of the converted 3D structure and the conversion rate. We chose an approach based on distance geometry technique, because it does not depend on a fragment library from which data may be missing. Regarding accuracy, we placed stronger emphasis on the accuracy of bond lengths, bond angles and ring conformations than on the conformations of acyclic moieties. The acyclic portion of ligand molecule has large conformational flexibility, and the conformation in the free state is often very different from that in the bound form to the target protein. So, the bound conformation of the ligand should be predicted in the molecular design stage, *e.g.*, by docking or molecular superposition, taking into account the intermolecular interaction with the target protein. Currently, many programs for molecular design can deal with conformational flexibility of an acyclic moiety of a ligand.[12—17] Besides structural accuracy and conversion rate, computational speed is also an important factor. However, there is a trade-off between structural accuracy and calculation speed, and in Key3D, structural accuracy is given priority over speed. This is because the generation of 3D structures needs to be performed only once when constructing a 3D database, and the 3D structures stored in the database can be used as many times as needed. Nevertheless, we tried to speed up the calculation as much as possible by significant simplification of the distance geometry procedure. Also, to enhance the quality of the created 3D structures, structure optimization is performed with MMFF94, which offers high accuracy and reliability in handling small organic molecules. In addition, molecular design based on 3D structures requires the proper handling of stereochemistry (*i.e.*, generation of structures with correct stereochemistry and enumeration of possible stereoisomers) and control of the protonation state, depending on the calculation model. Key3D was designed to meet these requirements.

In this paper, we describe in detail the Key3D algorithm. Then, we evaluate the performance of Key3D. For assessing the validity of the Key3D structures as inputs to molecular design tools, we have developed a "flexible-fitting test" method where the accuracy in terms of only bond lengths, bond angles and ring conformation can be estimated by allowing free rotation of acyclic bonds. Use of the Key3D structures as inputs for ADAM flexible docking, and validation of the docking accuracy have already been reported.[18]

## Methods

**Flowchart of Key3D** Figure 1 shows a flowchart of the Key3D method. Letters in parentheses in Fig. 1 correspond to the following sections.

(a) The structural formula prepared by using chemical structure drawing tools such as ISIS/DRAW (MDL Information Systems, Inc.) or ChemDraw

* To whom correspondence should be addressed. e-mail: itai@immd.co.jp

© 2006 Pharmaceutical Society of Japan

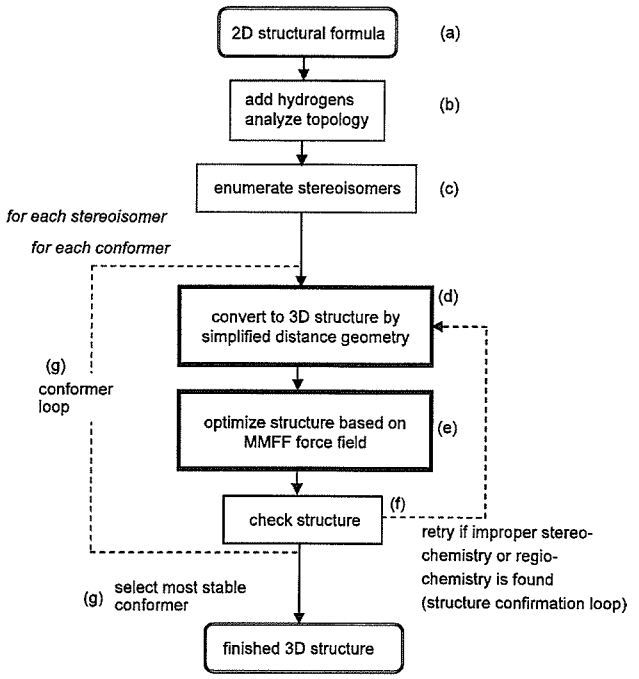



Fig. 1. Flowchart of Our 3D Model Builder, Key3D
Letters in parentheses correspond to the descriptions in the text.

(CambridgeSoft.Com) should be stored in the appropriate file format, *i.e.*, MDL format or MOL2 format of Tripos. Key3D reads this 2D topology data. Only atomic species and a connection table for the non-hydrogen atoms are needed as the 2D information. The elements that Key3D can deal with are H, C, N, O, Si, P, S, F, Cl, Br and I. Other elements are rarely found in general drug-like and bioactive molecules.

(b) The ring systems are identified, and the functional groups for which neutral and ionized forms are possible, *e.g.*, carboxylate, phosphate and amidino groups, are extracted. The protonation states for these groups can be designated by the user. According to the bond order and the designated protonation states, hydrogen atoms are added to the input structure.

(c) Analysis of stereochemistry is performed, and the stereoisomers to be generated are enumerated. When the stereochemistry is clearly defined in the input data (*e.g.*, as the bond records), Key3D creates the 3D structure of the corresponding isomer. If the molecule contains asymmetric centers with ambiguous stereochemistry, all the possible stereoisomers can be generated. However, when a large database has to be handled, it is impractical in terms of disk space to generate all possible $2^n$ isomers for $n$ asymmetric centers of ambiguous configuration. So, the number of chiral centers to be considered in creating stereoisomers is limited to a maximum of six in the current version, and the user also can control the number of isomers to be generated.

(d) For each stereoisomer, a 3D structure is constructed by means of the simplified distance geometry method, as described later.

(e) The constructed 3D structure is subjected to structure optimization based on MMFF94, to improve its quality. The original parameter set of MMFF94 accommodates a wide variety of small organic molecules, but not all. We have added some parameters to the original set so that most molecules in commercial compound libraries, such as Available Chemicals Directory (MDL Information Systems, Inc.) and MAYBRIDGE database, can be handled properly.

(f) Key3D examines whether the stereochemistry and regiochemistry of the generated 3D structure are consistent with the expected ones. If any inconsistency is found, Key3D retries the construction of 3D structure by returning to the distance geometry step (this cycle is called the "structure confirmation loop").

(g) The cycles of generation of 3D structure by the distance geometry method and structure optimization (d—f) are iterated several times, and as the result, a given number of conformers is produced for each stereoisomer ("conformer loop"). In general, satisfactory results can be obtained through three to five iterations. The conformer with the lowest MMFF energy is selected for each stereoisomer and output. The user can choose to output all of the created conformers.

**Simplified Distance Geometry in Key3D** We describe here the simplified distance geometry technique that has been used in generating 3D structures in Key3D. Good reviews of the standard distance geometry method have been published,[19,20] so we will focus on our modifications.

First, a distance bounds matrix that contains the maximum and minimum possible distances (upper and lower bounds) between each atom pair in the molecule is prepared as in the standard distance geometry method. The simplified method in Key3D skips the succeeding steps of smoothing, metrization and embedding, among which the metrization step requires a particularly large computational time. Instead of the crude Cartesian coordinates generated through these standard procedures, Key3D initially places atoms at random positions in a large 3D cube, 90 Å on a side. Then, minimization of an error function, which corresponds to the final step of the standard distance geometry, is performed. The error function is as follows:

$$E=\sum_{i=0}^{N-1}\sum_{j=i+1}^{N}\left(\max\left[0,\left(\frac{d_{ij}^2}{u_{ij}^2}-1\right)^2\right]+\max\left[0,\left(\frac{2l_{ij}^2}{l_{ij}^2+d_{ij}^2}-1\right)^2\right]\right)+c\sum_{k=0}^{N_{\text{chiral}}}(V_k-V_k^*)^2 \quad (1)$$

where $N$ is the number of atoms, $d_{ij}$ is the distance between atoms $i$ and $j$, $u_{ij}$ and $l_{ij}$ are the upper and lower bounds, $N_{\text{chiral}}$ is the number of chiral constraints, $V_k$ is the chiral volume for a given chiral constraint $k$, and $V_k^*$ is the desired value on that chiral volume. This error function is commonly used in the standard distance geometry calculation. The first term is for distance errors, and the second term corresponds to chiral errors. The coefficient $c$ for the chiral term has been introduced so that Key3D calculation reaches more desirable structures in terms of both accuracy of bond lengths and angles and correctness of stereochemistry and planarity. The value of 0.25 has been given to the coefficient $c$ for asymmetric centers and 1.0 for planar groups.

The stereochemistry is enforced to be the desired one by means of the chiral term of the error function, and the regiochemistry can also be controlled to some degree by the distance bounds of 1,4 atom pairs (*i.e.*, atoms separated by three bonds). Nevertheless, the minimization of the error function starting from random coordinates occasionally results in a 3D structure with incorrect stereochemistry or regiochemistry. To deal with this problem, the "structure confirmation loop" has been implemented. Key3D checks the structure after the structure optimization step, and if it has undesired stereochemistry or regiochemistry, Key3D goes back to the generation of random coordinates to try to construct another structure.

**Evaluation of Key3D. Test Set** A dataset of X-ray structures that Sadowski *et al.*[21] selected from the Cambridge Structural Database (CSD)[22] was used to evaluate the performance of Key3D. This test set does not focus on the drug-like structures that we place emphasis on, but it contains a wide variety of small molecules with high quality structures. The binding coordinates of ligand molecules are available from the Protein Data Bank (PDB),[23] but they were not used in our evaluation study. This is because the ligand coordinates in the PDB data have lower resolution than those in the CSD, and moreover, Key3D does not aim at the reproduction of bound conformation as stated in the introduction.

The test set contains a total of 639 structures, of which compounds including elements other than H, C, N, O, Si, P, S, F, Cl, Br and I are not targets of Key3D. We therefore used 598 molecules, excluding such compounds, in this validation study. The test set includes 213 molecules with one or more asymmetric centers (up to 17 asymmetric centers).

Sadowski *et al.* have carried out detailed comparative evaluation using this dataset for various 3D model builders including CORINA and CONCORD, and the results were reported in their paper.[21] However, their comparative study was carried out more than 10 years ago, so direct comparison of the performance between Key3D and the other methods studied in the paper is difficult.

**Evaluation Tests** Information only on atomic species, connectivity and stereochemistry was input to Key3D. Three cycles of conformer loop were designated in the Key3D calculation. The dielectric constant used in the MMFF structure optimization was set to 78.0.

We used the same test procedures as performed by Sadowski *et al.*,[21] with several modifications. The conversion rate, the number of program crashes, the number of asymmetric centers of incorrect stereochemistry, and the number of double bonds of incorrect regiochemistry were checked. To check CPU time, the calculation time on our Linux machine (Intel Xeon CPU 3.06 GHz) was measured. Close contact ratio (CCR),[21] which is the ratio be-

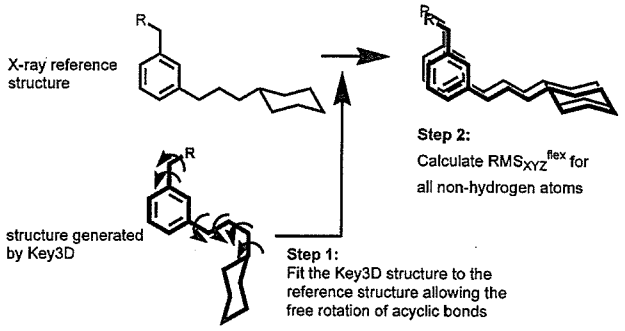



Fig. 2. Procedure of the Flexible-Fitting Test

tween the actual interatomic distance and a minimal acceptable distance, was measured as in the original procedure. The root-mean-square deviation (RMSD) for non-hydrogen atoms of the Key3D structure from the X-ray reference coordinates was calculated, as a measure of the accuracy of the created structures. We calculated RMSD for atoms included in ring systems ($RMS_{XYZ}^{rings}$), by means of the procedures described in the literature.

We also checked the occurrence rate of convergence error in single-point SCF calculation of the Key3D structure by using the MOPAC93 program,[24] although this test was not included in the report by Sadowski *et al.* The MO calculation occasionally fails to converge even for a 3D structure that seems to have been converted without problems, due to insufficient accuracy or energetic instability of the structure. We used the occurrence rate of convergence error in the MOPAC calculation as a measure of the suitability of the generated structures for general computational chemistry. The MNDO method was used for single-point SCF calculation.

**Flexible-Fitting Test** In the evaluation by Sadowski *et al.*, the RMSD were calculated in terms of the entire structure ($RMS_{XYZ}$), the ring moieties ($RMS_{XYZ}^{rings}$), and the torsion angles along the acyclic bonds ($RMS_{TA}^{chains}$). However, many molecular design methods in current use (*e.g.*, automated flexible docking, molecular superposition) are able to take into account the rotation of acyclic bonds, so there is a decreasing need for input structures in which the conformation of the acyclic moiety coincides with that in the crystal structure. Therefore, $RMS_{XYZ}$ and $RMS_{TA}^{chains}$ are considered to be less important, and were not included in this study. The factors that are more important for current molecular design tools are accurate geometries (bond lengths, bond angles, ring conformations), and accurate orientations of substituents in the ring systems (*i.e.*, axial or equatorial position) of the input structures. So, we have developed a procedure for calculating RMSD of all non-hydrogen atoms after fitting the generated structure to the reference crystal structure while allowing the rotation of acyclic bonds ($RMS_{XYZ}^{flex}$). This test procedure is named the flexible-fitting test. Figure 2 summarizes the method. In this procedure, the RMSD is minimized with the torsion angles along acyclic bonds of the generated structure as variables, by using the Simplex method.[25] If $RMS_{XYZ}^{flex}$ is close to 0 Å, the geometry of the rigid portion, *i.e.*, overall bond lengths and angles, ring conformations and the orientations of ring substituents, closely coincides with that of the reference structure.

## Results and Discussion

**Basic Performance** Table 1 summarizes the basic performance of Key3D. Key3D was able to convert all of the 598 molecules used in this study (*i.e.*, conversion rate= 100%). This test set corresponds to 93.6% of the 639 test structures in the original set. Key3D experienced no program crash during the calculations, and its robustness has been demonstrated. As for the CPU time, Key3D required an average of 0.49 s per molecule on our Linux machine (Intel Xeon CPU 3.06 GHz). Key3D is slower than the rule-based methods, but it is fast for a method based on distance geometry, which is generally time-consuming. The calculation time of Key3D is acceptable even for handling large compound libraries. No erroneous stereochemistry or regiochemistry was found in the Key3D structures.

Table 1. Performance of Key3D

| Evaluation item | |
|---|---|
| Conversion rate (%) | 94 (100)[a] |
| Program crashes | 0 |
| CPU time(s) per molecule | 0.49 |
| Erroneous stereochemistry | 0 |
| CCR>0.8 (%) | 100 |
| Convergence error in MOPAC 1SCF calculation | 0 |

*a*) Conversion rate of Key3D was 100% for the structures containing only H, C, N, O, Si, P, S and halogens, whereas it was 94% for all the original test set.

Table 2. Accuracy of the Ring Structures Created by Key3D

| $RMS_{XYZ}^{rings}$ (Å) | No. of structures (Percentage of the total) |
|---|---|
| <0.3 | 532 (89%) |
| 0.3—0.6 | 39 (7%) |
| 0.6—0.9 | 18 (3%) |
| 0.9—1.2 | 7 (1%) |
| 1.2—1.5 | 2 |
| >1.5 | 0 |
| Total | 598 |

Sadowski *et al.* established a criterion for acceptable structures without intramolecular clashes, based on analysis of the CCR values of the original X-ray structures; the criterion is CCR>0.8. All of the 3D structures generated by Key3D satisfied this criterion. It is a major advantage of methods based on distance geometry that few close contacts are included in the created structures, because severe clashes between atoms can be resolved during the step of minimizing the error function. Moreover, Key3D contains the further refinement step of structure optimization based on the MMFF force field so as to relieve structural distortion and intramolecular clashes that cannot be resolved only by the simplified distance geometry step.

In the single-point SCF calculation by using MOPAC93, all of the Key3D structures were successfully calculated without convergence failure. Apart from this validation study, we have rarely encountered MOPAC convergence failure for regular organic compound structures converted by Key3D. These results support the suitability of the Key3D structures for computational chemistry.

**RMSD from Reference Crystal Structure** Table 2 shows the distribution of the RMSD of the ring atoms ($RMS_{XYZ}^{rings}$) at intervals of 0.3 Å. This value reflects the accuracies of bond lengths, bond angles and conformations of the ring systems. In 89% of the structures, Key3D accurately reproduced the cyclic parts of crystal structures (RMSD<0.3 Å), and demonstrated high accuracy.

The RMSD of ring systems in the Key3D structures were rather large in some cases. There are several possible reasons for this. First, the number of iterations for conformer loops might not be sufficient. Since only three iterations were executed in this test calculation, the conformer resembling the crystal structure may not be produced, especially in the case of highly flexible ring systems. Key3D is able to perform conformational searches in which more conformers are generated and can output all conformers. For seven molecules with $RMS_{XYZ}^{rings}$ of more than 1.0 Å (Fig. 3), we carried out

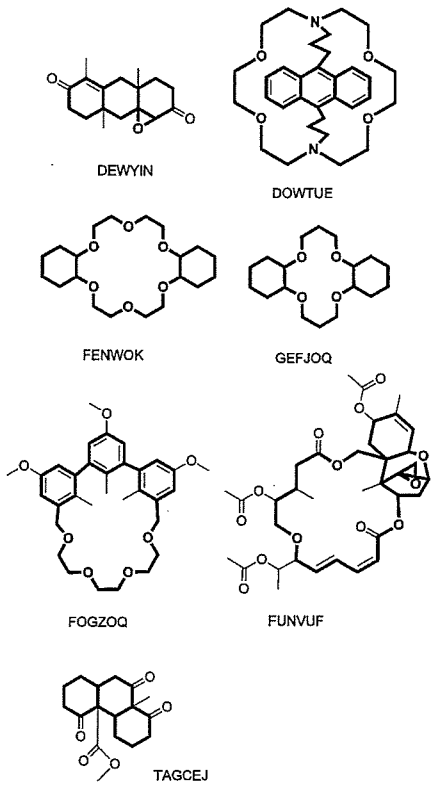



Fig. 3. Chemical Structures and CSD Refcodes of the Molecules with $RMS_{XYZ}^{ring}$ of More Than 1.0 Å

The ring moiety in bold line was used in estimating $RMS_{XYZ}^{ring}$ of each molecule. Stereochemical information is not shown for the sake of clarity.

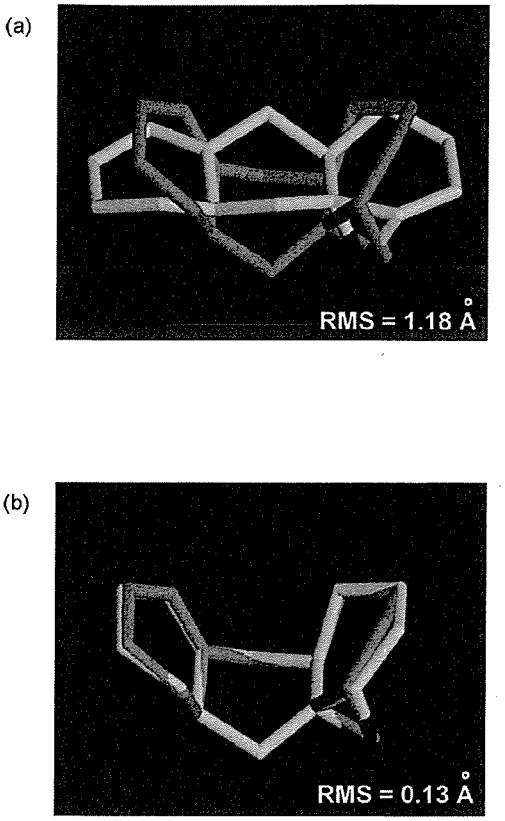



Fig. 4. Key3D Structures of DEWYIN, Obtained with Different Calculation Conditions; (a) Three Iterations of the Conformer Loop (Default Condition), (b) Conformational Search with 30 Conformer Loops

Each Key3D structure is compared with the reference crystal structure. The Key3D structure is shown in white, while the crystal structure is shown in gray. Only atoms included in the ring system are displayed for the sake of clarity.

Table 3. Effects of Conformational Search by Key3D on Accuracy of Ring Structures[a)]

| CSD refcode | The 3D structure obtained by the default conversion condition | The lowest energy structure obtained by conformational search | The smallest-RMSD structure obtained by conformational search |
|---|---|---|---|
| DEWYIN | 1.18 (53.0) | 0.13 (49.7) | 0.13 (49.7) |
| DOWTUE | 1.42 (134.2) | 0.83 (123.4) | 0.52 (125.8) |
| FENWOK | 1.21 (56.9) | 1.21 (56.9) | 0.69 (68.7) |
| FOGZOQ | 1.16 (153.4) | 0.97 (152.9) | 0.50 (155.8) |
| FUNVUF | 1.11 (151.0) | 0.52 (131.6) | 0.52 (131.6) |
| GEFJOQ | 1.14 (55.2) | 1.36 (46.7) | 0.54 (61.8) |
| TAGCEJ | 1.02 (65.8) | 0.05 (59.5) | 0.05 (59.5) |

*a*) The RMSD of the ring atoms is shown in Å unit. The value in parenthesis shows the MMFF force field energy (kcal/mol).

the Key3D calculation setting the number of the conformer loop to 30. As is shown in Table 3, for all the molecules, a ring conformation closer to the crystal reference structure was found among the 30 generated conformers (the smallest RMSD: 0.05—0.69 Å). The effects of conformational search with increased iterations, taking the case of DEWYIN as an example, is shown in Fig. 4. The smallest-RMSD structure obtained by conformational search had a very similar ring conformation to the crystal structure (RMSD 0.13 Å; Fig. 4b), while the ring conformation generated through only three iterations of the conformer loop was significantly different from the reference one (RMSD 1.18 Å; Fig. 4a). The second possible reason for large RMSD of ring systems is that accurate estimation of the relative stability of generated conformers is still difficult. In the results of conformational search with 30 iteration loops, the ring conformations of the lowest-energy structures were totally different from those of the crystal reference structures in the cases of FENWOK and GEFJOQ (Table 3). The MMFF94 that we have adopted for energy calculation in Key3D is one of the most accurate and reliable force fields for small organic molecules. However, it should be noted that the effect of crystal packing influences the stability of the crystal structure. Since most of the MMFF parameters are based on high-level *ab initio* MO calculations in the gas phase, it may be difficult for the force field to reproduce the effects of crystal packing. We consider that the reproduction of this effect is not necessarily important for the purpose of molecular design.

**Flexible-Fitting Test** Figure 5 shows the results of the flexible-fitting test of Key3D structures. This test evaluates how accurately the generated structures can reproduce the corresponding crystal structures when free rotations of acyclic bonds are allowed, as in flexible docking and flexible molecular superposition. As shown in Fig. 5, 76% of the total Key3D structures coincided with the reference crystal structures with $RMS_{XYZ}^{flex}$ of less than 0.3 Å. This result demonstrates high accuracy of the Key3D structures and the suitability of Key3D for use in current molecular design procedures. In our experience, an initial ligand structure with

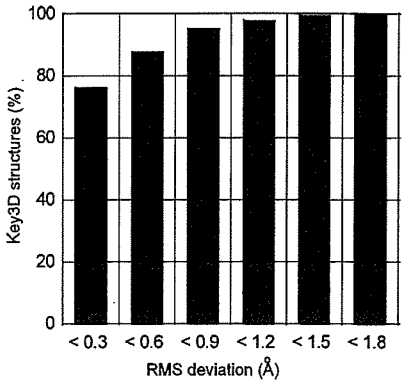



Fig. 5. Accuracy of the Key3D Structures Evaluated by the Flexible-Fitting Test

$RMS_{XYZ}^{flex}$ of less than 0.6 Å can afford an accurate docking model (in general, RMSD<1.2 Å). In this regard, 88% of the Key3D structures were of satisfactory quality. Moreover, further improvement of the accuracy is possible by increasing the number of cycles of the conformer loop. In this study, we used three cycles of the conformer loop. When we increased this to five cycles, the number of accurate Key3D structures ($RMS_{XYZ}^{flex}$<0.3 Å) increased by 2—3%, although more CPU time was required in proportion to the number of cycles of the conformer loop (data not shown).

In cases where rather large RMSD were observed in the flexible-fitting test, the difference in ring conformation was the major factor. Large RMSD were also observed when the orientation (axial or equatorial) of ring substituent(s) was inconsistent between reference and generated structures, no matter how accurate the other factors (bond length, bond angles, ring conformation) were. More elaborate conformational search by Key3D would lead to the acquisition of more appropriate structures in almost every case, as shown by the results for $RMS_{XYZ}^{rings}$.

**Features of the Key3D Algorithm** Here, we summarize the features of Key3D algorithm, based on the results of the evaluation tests described above. Key3D adopts a distance geometry method, which is more flexible than the rule-based approach that is inevitably restricted by the number and quality of internal rules and the library size. To improve the calculation speed of the distance geometry module, we have developed a highly simplified procedure centering on minimization of the error function. In consequence, both the accuracy of the generated structures and the calculation time are satisfactory for actual use in molecular design. The minimization step of the error function also contributes to the very low rate of conversion failure (in this evaluation study, there was no failure of conversion). Moreover, Key3D has been designed to provide good reproducibility of calculations, although a stochastic factor (*i.e.*, use of random numbers) is included in the algorithm. Another method of modified distance geometry depending on the refinement of error function has been developed by Spellmyer *et al.*[26] However, to our knowledge, their approach has not yet been validated as a practical 3D model builder. Key3D has the advantage of including all necessary functions for a practically useful and high-quality 3D builder, such as control of stereochemistry and structure optimization using a high-grade force field.

Key3D is also able to perform random conformational search. Since the conformational search by Key3D does not depend on a conformation library, which may restrict the variety of created conformers, Key3D can produce various stable conformations for any molecule. We showed above an example of the generation of a number of ring conformations. For the purpose of general 3D model building, the accuracy of Key3D structures can be improved by designating additional cycles of conformer loops. Thus, the user can easily adjust the balance between quality and speed in Key3D calculation.

Energy minimization of the intermediate structures constructed by the simplified distance geometry, on the basis of the MMFF force field, greatly contributes to the high quality of the Key3D final structures. In most 3D model builders, energy minimization is performed only roughly by using a simple force field, or no minimization is executed, because it requires significant computational time. However, we consider that elaborate structure optimization using an accurate force field is worth the cost of greater calculation time. Key3D has the advantage of directly generating reliable structures with no need for post-processing, and thus is suitable for practical use in actual drug design. For instance, when the Key3D structures were used for input structures in an evaluation test of flexible docking, high docking accuracy was achieved.[18] Furthermore, in our drug discovery projects based on virtual screening, far better results have been obtained by using Key3D than by using CONVERTER, which we had previously adopted.[27]

There remains some difficulty concerning the energy estimation. In our evaluation, the ring conformation in the lowest-energy structure generated by Key3D was not always consistent with that in the corresponding crystal structure. Nevertheless, ring conformers with small RMSD were found within a moderate energy range. Considering that more than one conformation may exist within the thermally accessible energy range, and that the most stable conformation may be different in the crystal, in solution, or in a protein-bound state, it is likely to be advantageous to provide several low-energy conformers as model structure candidates, for successful molecular design.

**Acknowledgments** We are grateful to Professor Johann Gasteiger and Dr. Christof Schwab (Universität Erlangen-Nürnberg) for kindly providing the coordinates of 639 CSD structures used in their evaluation scheme.

**References and Notes**

1) Present address: *Quantum Bioinformatics Team, Center for Promotion of Computational Science and Engineering, Japan Atomic Energy Agency; 8–1 Umemidai, Kizu-cho, Soraku-gun, Kyoto 619–0215, Japan.*
2) Sadowski J., Gasteiger J., *Chem. Rev.*, **93**, 2567—2581 (1993).
3) Gasteiger J., Rudolph C., Sadowski J., *Tetrahedron Comp. Methodol.*, **3**, 537—547 (1990).
4) Pearlman R. S., *Chem. Des. Autom. News*, **2**, 1—6 (1987).
5) Rusinko A., Sheridan R. P., Nilakantan R., Haraki K. S., Bauman N., Venkataraghavan R., *J. Chem. Inf. Comput. Sci.*, **29**, 251—255 (1989).
6) Crippen G. M., Havel T. F., *Acta Cryst.*, **A34**, 282—284 (1978).
7) Halgren T. A., *J. Comp. Chem.*, **17**, 490—519 (1996).
8) Halgren T. A., *J. Comp. Chem.*, **17**, 520—552 (1996).
9) Halgren T. A., *J. Comp. Chem.*, **17**, 553—586 (1996).
10) Halgren T. A., Nachbar R. B., *J. Comp. Chem.*, **17**, 587—615 (1996).
11) Halgren T. A., *J. Comp. Chem.*, **17**, 616—641 (1996).
12) Friesner R. A., Banks J. L., Murphy R. B., Halgren T. A., Klicic J. J., Mainz D. T., Repasky M. P., Knoll E. H., Shelley M., Perry J. K., Shaw

D. E., Francis P., Shenkin P. S., *J. Med. Chem.*, **47**, 1739—1749 (2004).
13) DesJarlais R. L., Sheridan R. P., Dixon J. S., Kuntz I. D., Venkataraghavan R., *J. Med. Chem.*, **29**, 2149—2153 (1986).
14) Leach A. R., Kuntz I. D., *J. Comp. Chem.*, **13**, 730—748 (1992).
15) Mizutani M. Y., Tomioka N., Itai A., *J. Mol. Biol.*, **243**, 310—326 (1994).
16) Jones G., Willett P., Glen R. C., Leach A. R., Taylor R., *J. Mol. Biol.*, **267**, 727—748 (1997).
17) Kato Y., Inoue A., Yamada M., Tomioka N., Itai A., *J. Comput.-Aided Mol. Des.*, **6**, 475—486 (1992).
18) Mizutani M. Y., Takamatsu Y., Ichinose T., Nakamura K., Itai A., *PROTEINS*, **63**, 878—891 (2006).
19) Crippen G. M., Havel T. F., "Distance Geometry and Molecular Conformation," ed. by Bawden D., Research Studies Press (Wiley), New York, 1988.
20) Blaney J. M., Dixon J. S., "Reviews in Computational Chemistry," Vol. 5, Chap. 6, ed. by Lipkowitz K. B., Boyd D. B., VCH. Publishers, New York, 1994, pp. 299—335.
21) Sadowski J., Gasteiger J., Klebe G., *J. Chem. Inf. Comput. Sci.*, **34**, 1000—1008 (1994).
22) Allen F. H., Davies J. E., Galloy J. J., Johnson O., Kennard O., Macrae C. F., Mitchell E. M., Mitchell G. F., Smith J. M., Watson D. G., *J. Chem. Inf. Comput. Sci.*, **31**, 187—204 (1991).
23) Berman H. M., Battistuz T., Bhat T. N., Bluhm W. F., Bourne P. E., Burkhardt K., Feng Z., Gilliland G. L., Iype L., Jain S., Fagan P., Marvin J., Padilla D., Ravichandran V., Schneider B., Thanki N., Weissig H., Westbrook J. D., Zardecki C., *Acta Cryst.*, **D58**, 899—907 (2002).
24) MOPAC93, Fujitsu Limited, Tokyo, 1993.
25) Nelder J. A., Mead R., *Computer J.*, **7**, 308—313 (1965).
26) Spellmeyer D. C., Wong A. K., Bower M. J., Blaney J. M., *J. Mol. Graph. Model.*, **15**, 18—36 (1997).
27) Mizutani M. Y., Itai A., *J. Med. Chem.*, **47**, 4818—4828 (2004).

細胞バンク

第4回

# JCRB細胞バンク事業の概要

小原有弘* 水澤 博*

Point

①厚生労働省によって設立されたJCRB細胞バンクは生命科学研究の研究支援として高品質な細胞を研究者に提供している.
②研究に使用している細胞のマイコプラズマ汚染の現状は深刻で,研究への悪影響を知らずに研究利用している研究者が多い.
③STR-PCR法による細胞個体識別はクロスカルチャーコンタミネーションの発見に非常に有用である.
④クロスカルチャーコンタミネーションの問題は生命科学研究の根底を揺るがしかねない大きな問題である.
⑤ウイルス汚染検査の確立,ES細胞・体性幹細胞の供給,細胞特性解析などJCRB細胞バンクの今後の役割は大きい.

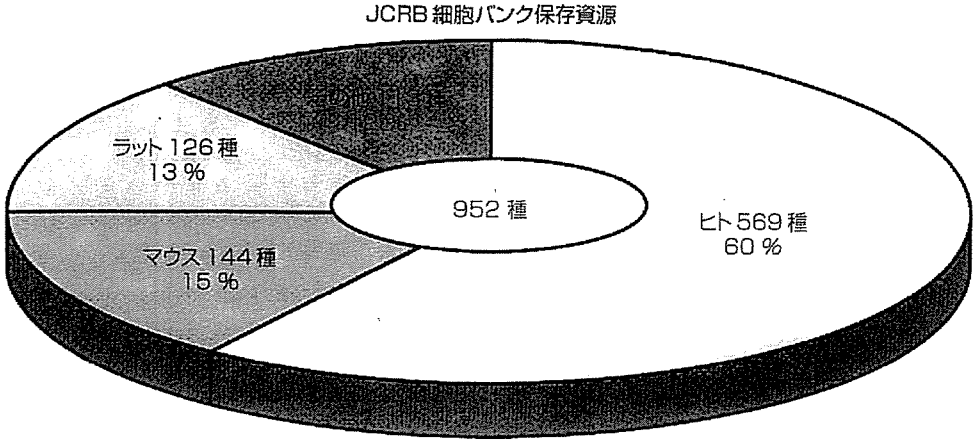
Key Words /JCRB細胞バンク,マイコプラズマ汚染,細胞個体識別,STR-PCR法,クロスカルチャーコンタミネーション

## ● はじめに

1985年,わが国最初の細胞バンクとして厚生労働省(当時厚生省)によってJCRB細胞バンク(http://cellbank.nibio.go.jp/)が誕生した.当時すでに,ヒトに由来する培養細胞は生命科学研究に欠かせない研究材料となっており,対がん10ヵ年総合戦略(1984年～1993年)における,研究の基盤整備の一環として,日本がん研究資源バンク(Japanese Cancer Research Resources Bank:JCRB)の整備が進められた.そのなかで『培養細胞研究資源』に関するものとして,国立医薬品食品衛生研究所(当時国立衛生試験所,変異遺伝部)に設立されたのがJCRB細胞バンクである.

こうしてがん研究の支援をおもな目的として細胞バンクがスタートし,生命科学研究全体の発展とともに細胞バンク事業も発展してきた.現在は(財)ヒューマンサイエンス振興財団(HS財団)との協力体制をとり,JCRB細胞バンクが細胞の収集,品質管理,長期安定保存を担当し,HS財団が細胞分譲に関する業務を担い,迅速な分譲体制を維持・確立している.また,2005年4月1日,国立医薬品食品衛生研究所にあったJCRB細胞バンクは大阪府北部の新しい街「彩都(茨木市)」の一角に建設された独立行政法人医薬基盤研究所に移転して新たなスタートを切ることになった.

*KOHARA Arihiro, MIZUSAWA Hiroshi/(独)医薬基盤研究所・生物資源研究部




分譲実績(2002年1月～2005年11月)

| 年 | 海外 | 国内企業 | 研究機関 | 合計 |
|---|---|---|---|---|
| 2002 | − | − | − | 2406 |
| 2003 | 222 | 682 | 2095 | 2999 |
| 2004 | 294 | 727 | 2152 | 3173 |
| 2005 | 190 | 684 | 1863 | 2737 |

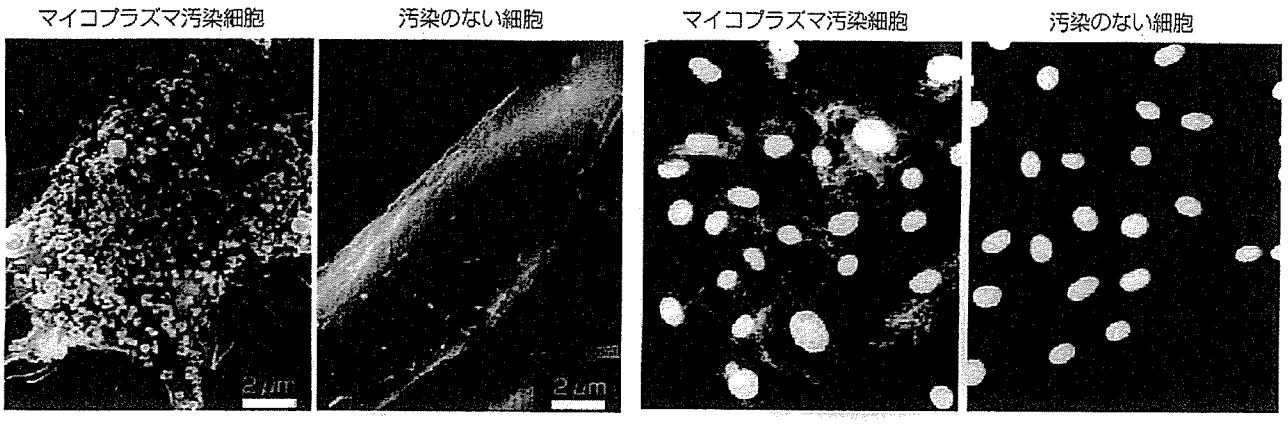
**図❶ JCRB細胞バンクの資源保有数と分譲実績**(筆者作成)
JCRB細胞バンクが保有する細胞資源数は952種となっており，そのうち約60％がヒト由来の細胞である．分譲は年間3000アンプルを越え徐々に増加している．

## 1 細胞バンクの概要

JCRB細胞バンクの事業内容は，①研究に有用な培養細胞研究資源の収集，②収集した培養細胞研究資源の品質管理，③十分な数の細胞アンプルの長期安定的保存，④培養細胞に関連する新しい研究資源の開発および資源化に関する研究開発，⑤培養細胞研究資源の品質管理手法の開発研究，⑥培養細胞研究資源に関連する情報の収集と提供，⑦分譲を担当するHS財団への新規細胞株の分譲用細胞の提供，となっている．収集・登録した細胞は毎年40株程度ずつ増加しており，現在では952種となっている．また，年間に分譲するアンプル数も3000を越えて年々増加する傾向にある(図❶)．品質管理に関しては汚染検査，細胞個体識別，細胞性状確認を実施しており，後述するマイコプラズマ非汚染で，しかも細胞個体識別を確認した培養細胞を提供していることは，JCRB細胞バンクが研究者に提供する研究資源の品質の高さを示すものであり，研究者より好評を得ている．

## 2 マイコプラズマ汚染は古くて新しい課題である

培養細胞を用いて研究している研究者のなかでもマイコプラズマ汚染による研究への悪影響を知らない研究者は多いようである(図❷)．マイコプラズマは自己増殖能をもつ細菌の1/10ぐらいの大きさの微生物であり，培養細胞と共存して増殖するが，汚染しても培地が濁ったりしないので混入に気づきにくい．もし，自分が研究に使っている細胞がマイコプラズマ汚染されていたらどうであろうか？その意味は十分考える必要がある．マイコプラズマ汚染のためせっかくおこなった研究に再現性がなく，信頼されない研究となってしまうかもしれない．われわれ細胞バンクには培養の専門家から多くの細胞が寄託されるが，その約20％にマイコプラズマ汚染が見つかるのが現状である．専門家とよばれる研究者が使用していた細胞にマイコプラズマ汚染率が高いことを考え




**図❷ マイコプラズマによる汚染**
マイコプラズマ汚染された細胞の走査型電子顕微鏡像(ヒューマンサイエンス研究資源バンク：吉田東歩博士撮影)とヘキスト 33258 で染色される細胞の蛍光染色像．マイコプラズマにはいろいろな種類が存在するが，培養細胞では細胞表面に付着して細胞と共生する．走査型電子顕微鏡像では細胞表面の粒子状のものがマイコプラズマである．ヘキストによる核酸染色は培養細胞の核だけではなく，マイコプラズマの核酸も染色されるのでマイコプラズマ汚染が検出できる．

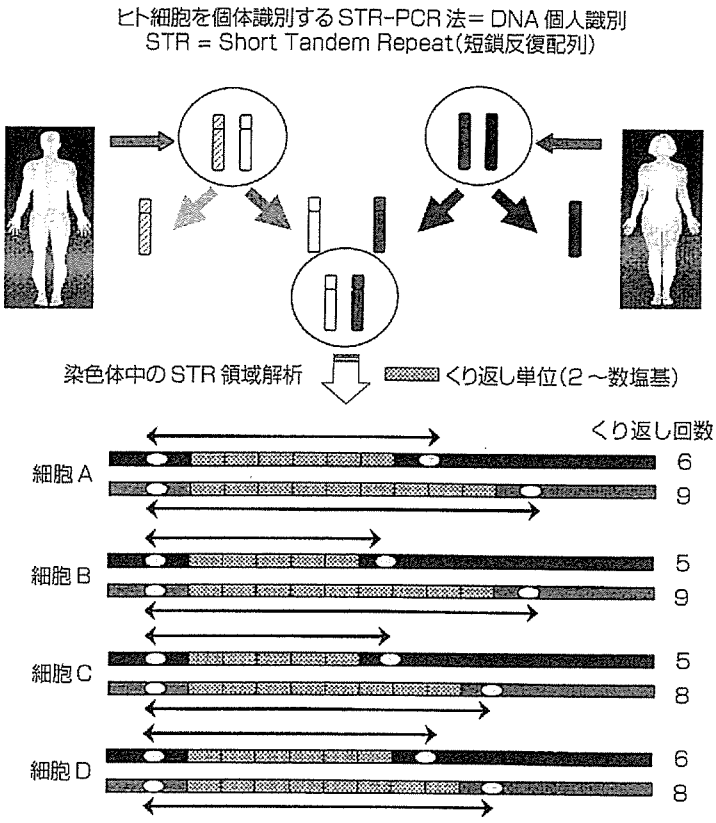
ると，広く普及した培養細胞の汚染率は相当高いと容易に予測できる．一度汚染した細胞は通常の培養方法で用いる抗生物質ではなかなか除去することは難しく増殖を抑える程度となってしまう．また，凍結保存しても細胞と同様にマイコプラズマも生き残ってしまう．培養細胞を研究に利用されている場合には，是非一度自分の細胞を調べてみることを勧めたい．

では，何が原因で汚染が広がってゆくのであろうか？その答えには三つあげられる．第一に培養実験室の環境からの汚染である．培地などをこぼしてしまったりしたとき，すぐにしっかりと拭き取らないと何らかの培養操作の過程でほかの細胞へ混入することがある．第二に培地の使い回しによる汚染があげられる．同じ培地を用いるのだから細胞が違っても大きなビンから同じ培地を使用するとか，ほかの研究者と培地を共有してしまうなどよく聞く話ではないだろうか？これが汚染の拡大に繋がっているのである．第三に人の唾液を通じての汚染が考えられる．マイコプラズマは人の上部気道や尿生殖器に常在する微生物である．培養する際のおしゃべりが汚染の原因の一つとも考えられている．これらのことから考えると汚染を拡大しないようにするには，培養環境を清潔に保つ，培地は「1 培地－1 細胞」，おしゃべりは厳禁(マスク着用)が原則となる．

## ❸ STR-PCR 法はヒト由来細胞株を個体識別する

1951 年に HeLa 細胞が樹立されたことが刺激となり，多くの研究者が続々と新しいヒト由来の培養細胞の樹立に成功した[1]．そこに問題を提起したのが Gartler[2]であり，アイソザイムの分析から HeLa 以降樹立されたヒト細胞の多くが HeLa 細胞ではないかと疑ったのだった．1900 年代の後半には英国の Jeffreys[3]らが DNA フィンガープリント法を報告し，ヒトを DNA レベルで識別することが可能だと紹介した．その後，この方法は犯罪捜査への応用という視点から急速に研究が進み，現在では Short Tandem Repeat-Polymerase Chain Reaction (STR-PCR)法として迅速かつ精密な分析法として定着しつつある．STR-PCR 法とはゲノム中に存在する 2～数個の塩基からなるくり返し配列[(CAG)n，(GC)n など]のくり返し回数に個人差があることから，その出現回数を分析することによって個体識別する方法である(図❸)．

JCRB 細胞バンクではこの方法を 1999 年末から培養細胞に取り入れて，収集したヒト由来の培養細胞に関する調査を実施してデータを蓄積し，『ヒト培養細胞識別データベース』を構築した．このデータベースを利用しておこなった多種類のヒト細胞の比較で明らかに




**図❸ STR-PCR法の原理**(筆者作成)
細胞Aから細胞Dの4種類の細胞に存在するSTR領域の短鎖反復配列の構造を模式的に示した．高等動物の染色体は母方と父方に由来するものがペアになっており，STR(短鎖反復配列)領域も対になって2つ存在する．STRは2～数塩基程度のきわめて短い塩基配列が反復している構造であり，この領域の外側に適当なPCRプライマーを設定してSTRを挟む100～300塩基程度の長さのDNA鎖をPCR法によって増幅してその反応産物をジェネティックアナライザーによって解析する．

なったことは，ヒト培養細胞には意外と多くのクロスカルチャーコンタミネーション(細胞の入れ替わり)が発生していたということである．これまでJCRB細胞バンクで収集したヒトに由来する培養細胞数は560種(全体の60％)であるが，そのうち32種(約6％弱)にクロスカルチャーコンタミネーションが見つかったのである．この結果は詳細な実験手法を含めてJCRB細胞バンクのホームページで公開しているので是非参考にしていただきたい(http://cellbank.nibio.go.jp/中，『JCRB Cell Bank』のCell IDの欄に公開している)．

最近，JCRB細胞バンクで判明した一例を挙げてみたい．唾液腺がん由来の細胞としてわが国で樹立された細胞が歯学領域の研究者の間で広く研究利用されており，ヨーロッパの細胞バンクにも登録されていた．JCRB細胞バンクにこの細胞のSTR-PCR解析の依頼があり，調べた結果HeLa細胞と同一のヒトから樹立されたもの，つまりはHeLa細胞の入れ替わりと判定された．その後広く普及したこの細胞をヨーロッパの細胞バンクや日本国内の研究者より数種集めて再解析をおこなったが，結果はHeLa細胞由来であることを示した．細胞の入れ替わりがどの時点で起こったのか明らかではないが，現在その細胞を利用して研究している研究者は研究成果をどのように発表してよいのか苦悩している．このようにクロスカルチャーコンタミネーションは生命科学研究の根底を揺るがす深刻な問題となりかねないものなのである．

## ❹ JCRB細胞バンクの今後の役割は大きい

以上紹介したようにJCRB細胞バンクは，細胞の収集と品質管理を重視した運営をおこなってきたが，近年，ヒトゲノムの全塩基配列が決定されて個人単位の遺伝子解析が容易に実施できるようになった．その結果個人の遺伝情報が垂れ流しになるのではないかという心配から保護をする必要性が話題になり，それに関連して生命倫理という新たな課題が浮上している．もちろんこれまでも倫理問題はなかったわけではないが国内ではあまり関心をもたれなかった．そこで，細胞バンクで重視している研究材料がヒトに由来する細胞であることを考える

と，研究倫理を研究課題として細胞バンクでも独自に調査研究する必要があるのではないだろうかと考えはじめたところである．

また，培養細胞の品質管理については，①マイコプラズマ汚染の有無，②細菌・真菌汚染の有無，③アイソザイム分析による由来動物種の確認，④STR-PCR法によるヒト細胞の個体識別を実施する体制を確立してきた一方，まだ導入していない品質管理の課題としてはウイルスによる汚染検査の問題がある．再生医療や細胞治療に向けて非常に多くの研究開発が進んでいるが，細胞を用いた医療の実現にはウイルス汚染検査が必須である．JCRB細胞バンクでもウイルス汚染検査体制の確立が急務であるとして研究開発に着手した．

収集する細胞に関しても非常に多様かつ有用な細胞が次々と樹立されている．とくに近年ではES細胞や体性幹細胞とよばれる多分化能をもった増殖性の細胞の研究利用が進んでいる．これはヒトの正常な(がんではない)状態を*in vitro*で模倣できる非常に有用な研究材料であり，これらの研究材料を供給することにJCRB細胞バンクも積極的に取り組んでいる．

研究者が研究に利用する細胞を選択するには細胞特性の情報が非常に重要であり，これまでにもJCRB細胞バンクではそれらの特性を解析する研究を実施している．今後も染色体の詳細な解析，遺伝子発現情報の付加，細胞増殖過程の動画記録などの情報提供に取り組み，研究者の支援に努めていきたいと考えている．

《細胞寄託，品質管理に関する問い合わせ》
JCRB細胞バンク
独立行政法人医薬基盤研究所，生物資源研究部，
細胞資源研究室
〒567-0085 大阪府茨木市彩都あさぎ7-6-8
電話：072-641-9819，FAX：072-641-9851
e-mail：cell@nibio.go.jp
ホームページ：http://cellbank.nibio.go.jp/

## 文献

1) Gey G.O *et al*：Tissue culture studies of the proliferative capacity of cervical carcinoma and normal epithelium. *Cancer Res* **12**：264-265, 1951
2) Gartler SM *et al*：Apparent Hela cell contamination of human heteroploid cell lines. *Nature* **217**：750-751, 1968
3) Jeffreys AJ *et al*：Individual-specific 'fingerprints' of human DNA. *Nature* **316**：76-79, 1985