Making Optimal Use of Empirical Energy Functions: Force-Field Parameterization in Crystal Space

Elmar Krieger,[1]* Tom Darden,[2] Sander B. Nabuurs,[1] Alexei Finkelstein,[3] and Gert Vriend[1]
[1]*Center for Molecular and Biomolecular Informatics, University of Nijmegen, Nijmegen, The Netherlands*
[2]*Laboratory of Structural Biology, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina*
[3]*Institute of Protein Research, Russian Academy of Sciences, Pushchino, Moscow Region, Russia*

ABSTRACT **Today's energy functions are not able yet to distinguish reliably between correct and almost correct protein models. Improving these near-native models is currently a major bottle-neck in homology modeling or experimental structure determination at low resolution. Increasingly accurate energy functions are required to complete the "last mile of the protein folding problem," for example during a molecular dynamics simulation. We present a new approach to reach this goal. For 50 high resolution X-ray structures, the complete unit cell was reconstructed, including disordered water molecules, counter ions, and hydrogen atoms. Simulations were then run at the pH at which the crystal was solved, while force-field parameters were iteratively adjusted so that the damage done to the structures was minimal. Starting with initial parameters from the AMBER force field, the optimization procedure converged at a new force field called YAMBER (Yet Another Model Building and Energy Refinement force field), which is shown to do significantly less damage to X-ray structures, often move homology models in the right direction, and occasionally make them look like experimental structures. Application of YAMBER during the CASP5 structure prediction experiment yielded a model for target 176 that was ranked first among 150 submissions. Due to its compatibility with the well-established AMBER format, YAMBER can be used by almost any molecular dynamics program. The parameters are freely available from www.yasara.org/yamber. Proteins 2004;57:678–683.**

INTRODUCTION

Thanks to the exponential growth in processing power, the atomistic simulation of proteins has become feasible on personal computers, allowing scientists to routinely analyze internal motions[1,2] or the effects of point mutations on protein stability.[3] Because accurate quantum chemical calculations take orders of magnitude too long, such simulations are based on empirical energy functions, like the AMBER,[4] CHARMm,[5] or GROMOS[6] molecular dynamics force fields.

While many questions can be answered by today's force fields, the biannual CASP meetings regularly show that one major goal has not been reached yet: the successful refinement of homology models.[7] With structural genomics producing a rapidly growing number of templates for homology modeling,[8–10] bridging the accuracy gap between homology models and high resolution X-ray structures becomes increasingly important. However, in close proximity to the native structure (0.5–3Å Cα RMSD), energy functions lose their discriminative power due to the often very small structural and energetic differences involved.[11] As a conclusion, highest force-field accuracy is required when refining homology models.[12]

Fitting force-field parameters is a very tedious task, usually involving quantum chemical calculations on small molecules.[13,14] And even perfectly accurate parameters cannot guarantee success, because the mathematical form of the energy function is an approximation by itself. Recently, we demonstrated a property we call "force-field equivalence:" the force-field parameters that performed best at improving homology models were virtually the same as those that did minimum damage to real structures during an energy minimization.[15] While it takes nanosecond simulations to check if a homology model improves during a molecular dynamics run, only picoseconds are required to minimize a real experimental structure. "Force-field equivalence" thus allows a gain factor of 1000 in computing time when judging the suitability of a force field for model refinement. This in turn permits to increase force-field accuracy with a rather uncompromising approach: a "self-parameterizing force field," the parameters of which are iteratively optimized to minimize the damage done to a training set of high resolution X-ray structures during a simulation.[15] This method leads to a consistent set of parameters that optimally fit the given force-field equation.

It is obvious that the protein structures in the training set should be as close to reality as possible, otherwise the force field might learn to reproduce features that simply do not exist. Here we describe a novel way of achieving that: the large-scale reconstruction of crystallographic unit cells, including disordered water molecules and counter ions, as well as hydrogen atoms. The latter turned out to be

*Correspondence to: Elmar Krieger, CMBI, Toernooiveld 1,6525ED Nijmegen, The Netherlands. E-mail: elmar.krieger@yasara.org

Received 8 December 2003; Accepted 7 May 2004

Published online 12 August 2004 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20251

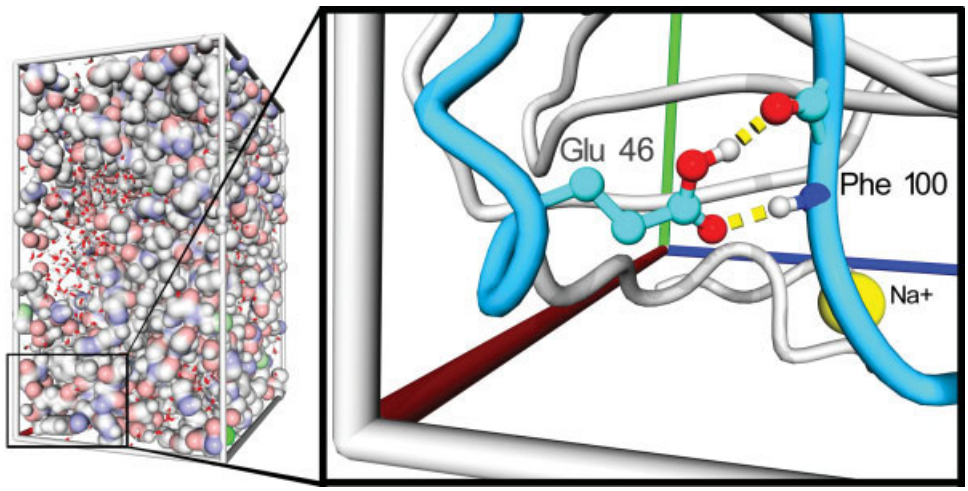

Fig. 1. Reconstructed unit cell of PDB entry 1FUS, containing four chains of ribonuclease F1. The magnified area on the right shows residue Glu 46 contacting the backbone of Phe 100. The predicted pKa for Glu 46 is 4.1, while the protein was crystallized at pH 3.5, making a protonation very likely. A predicted sodium counter-ion is also shown. Image created with YASARA.

especially important, as can be seen from Figure 1. This example shows the unit cell of a ribonuclease F1 crystal (PDB[16] ID 1FUS[17]), with the side-chain carboxyl group of residue Glu 46 contacting the backbone of residue Phe 100. This close contact of one carboxyl- and one backbone oxygen is very unfavorable, unless there is a proton trapped in between. If the force field was optimized without this proton present, it would memorize a completely unrealistic interaction pattern. In addition to a hydrogen-bonding network optimizer,[18] a new method for pKa prediction in protein crystals was required to assign the protonation states of ionizable groups (Krieger et al., forthcoming).

Optimizing a force field in crystal space makes sure that all the forces giving rise to the experimentally observed structure are also present during the simulation and can be considered when fitting the parameters. As the physics acting in a crystal are the same as in solution, the resulting force field can be used in both environments.

MATERIALS AND METHODS

Reconstruction of Crystallographic Unit Cells

Using the PDBFINDER database,[19] all X-ray structures without uncommon ligands and with unit cells smaller than 260,000 Å^3 and were selected. From these, a non-redundant set (30% sequence identity cutoff) was extracted and sorted by a combined resolution/R-factor quality indicator.[20] The top 50 structures were chosen and divided into optimization (even numbers) and validation sets (uneven numbers). For every structure, WHAT IF[21] was run to add symmetry related chains based on the spacegroup information in the PDB file, and to optimize the hydrogen-bonding network.[18] Then YASARA was used to predict the pKa values of the ionizable groups in the crystal, assign their protonation states based on the pH at which the crystal was solved (retrieved from literature if not specified in the PDB file) and fill the cell with water. Water molecules in the original PDB files were kept if they were closer than 5 Å to the protein. Using an iterative procedure, the AMBER99 electrostatic potential[13] was evaluated at all water molecules,[22] and the one with the lowest or highest potential was turned into a sodium or chloride counter ion, respectively, until the cell was neutral. Then a startup simulation was run for 5 ps using the protocol described below, with all heavy protein atoms fixed, so that the solvent molecules could smoothly cover the protein surface. Finally a short steepest descent minimization of all atoms was done to remove severe bumps in the protein. The 25 structures in the optimization set were 1et1, 1bqk, 1k1b, 1fus, 1ijv, 1ptf, 1g2b, 1bd8, 1h75, 1ifc, 1ajj, 1aho, 1kf3, 1jo8, 1d4t, 1hyp, 1bkr, 1a62, 1faz, 1aac, 1hcv, 2ovo, 1exr, 2erl, and 1kth; the validation set consisted of 1psr, 2pth, 1ihr, 1g9o, 1fux, 1qtw, 1eqt, 1jek, 2a0b, 1gk7, 1c7k, 1d0d, 1g2r, 2ygs, 1fl0, 1hka, 1g2q, 1im5, 1gvp, 1hg7, 1i2t, 1f94, 2igd, 1cuo, and 1gdu.

Simulations of Crystals and Models

All simulations were run with YASARA (www.yasara.org), using a multiple time step of 1 fs for intramolecular and 2 fs for intermolecular forces. A 7.9 Å cutoff was taken for Lennard–Jones forces and the direct space portion of the electrostatic forces, which were calculated using the Particle Mesh Ewald method[23] with a grid spacing <1 Å, 4th order B-splines and a tolerance of 10^{-4} for the direct space sum. Simulated annealing minimizations started at 298 K, velocities were scaled down with 0.9 every ten steps for a total time of 5 ps. While it is tempting to let the unit cells relax during the energy minimizations and use the deviations as an additional force-field quality indicator, we decided against it due to mainly two reasons: first, the required pressure calculations have been shown to be negatively influenced by the truncation of long-range Van der Waals interactions,[23] potentially leading to optimization artifacts in our case. Second, there is no unambigu-

TABLE 1. The 37 Optimization Parameters of the YAMBER Force Field†

Parameters	Description
1	Common scaling factor for bond stretching force constants
2	Common scaling factor for angle bending force constants
3	Scaling factor for Lennard-Jones forces between 1–4 bonded atoms
4	Scaling factor for electrostatic forces between 1–4 bonded atoms
5	Common scaling factor for all torsion energies excluding peptide bond
6	Energy barrier of the peptide bond
7	Improper dihedral barrier for carbonyl and carboxyl groups
8	Improper dihedral barrier for all other planar groups
9–24	VdW radii of the following AMBER atom types: H, HC, HI, HP, HA, H4, H5, O, OH, C, CA, CT, N, S, C0, Zn
25–28	VdW contact energies of hydrogen, carbon, nitrogen, and oxygen
29	Charge shift c for N (+c) and H (−c) in peptide bonds
30	Charge shift c for N (+c) and H (−c) in aromatic NH groups
31	Charge shift c for N (+c) and H (−c/2) in NH2 groups
32	Charge shift c for N (+c) and H (−c/3) in NH3 groups
33	Charge shift c for O (+c), H (−2c/3) and C (−c/3) in hydroxyl groups
34	Charge shift c for C (+c) and O (−c) in carbonyl groups
35	Charge shift c for C (+c) and O (−c/2) in carboxyl groups
36	Charge shift c for C (+c) and H (−c/3) in methyl groups
37	Charge shift c for C (+c) and H (−c) in aromatic rings

†Charge shift parameters are simply added to the AMBER charges, their distribution ensures that the overall net charge does not change.

ous way of combining the unit cell deviation with the atomic RMSD to arrive at a single optimization progress indicator.

Molecular dynamics simulations of crystals[24] were carried out at the temperature chosen during structure determination, the unit cells were again kept fixed (NVT ensemble). Simulations of homology models were set up in the same way as the crystal simulations (with respect to the placement of water molecules, counter ions and optimization of the hydrogen-bonding network to determine e.g., histidine protonation states), with cells 13 Å larger than the protein along each axis, and then run at 298 K and constant pressure (NPT ensemble) to account for volume changes due to the much larger fluctuations of homology models in solution when compared to X-ray structures in a crystal environment. The temperature was adjusted using a Berendsen thermostat[25] based on the time-averaged temperature, i.e., to minimize the impact of temperature control, velocities were rescaled only about every ~100 simulation steps, whenever the average of the last ~100 measured temperatures converged. To allow a direct comparison, the 25 homology models were the same as in our previous study.[15]

Calculation of RMSDs and B-Factors

To reduce the amount of noise in the data, various precautions were taken. In Figures 2 and 3, all atoms with alternate locations in the original PDB file were excluded from RMSD calculations. In Figure 4, the models were scanned for flexible N- and C-terminal tails (Cα atoms that have less than x other Cα atoms within 7 Å, where x = 3 for the first and last residue, x = 4 for the second and second-last, and x = 5 for all other residues). Those tails were excluded, as well as three models in the upper half of Figure 4 that could not be expected to contribute useful data: 1RB9 and 451C because they contained an iron-sulfur cluster and a hem group for which no suitable force-field parameters were available, and 1BX7, because we found previously that there was no "correct" structure due to the protein's high flexibility.[15]

B-factors were calculated from the last nanosecond of the simulation as described previously,[26] using only backbone atoms and ignoring the side-chains to avoid artificially high results caused by rotamer flips. Of the 25 crystals in the validation set, the 20 structures that did not have backbone atoms with alternate locations were considered.

RESULTS

Force-Field Optimization

Initial force-field parameters were borrowed from the AMBER99 force field, which has been shown earlier to be very accurate.[27] Because the total number of AMBER force-field parameters is much larger than what can possibly be optimized, a subset of 37 parameters was chosen (Table I). The majority of these (20) describe Van der Waals radii and contact energies, which are usually among the most difficult to parameterize. Nine parameters capture shifts in the charge distribution, and the remaining eight relate to bonds, angles, and torsions.

The parameters were optimized using a Monte Carlo search algorithm.[15] The quality of each parameter set was evaluated by a simulated-annealing minimization of 25 protein crystals. A lower RMSD from the initial structures meant higher quality. Due to the huge computational requirements of this procedure, the Models@Home distributed computing system was used[28] (freely available from www.yasara.org/models).

The force-field optimization progress is shown in Figure 2. From an initial value of 0.467 Å measured after minimi-

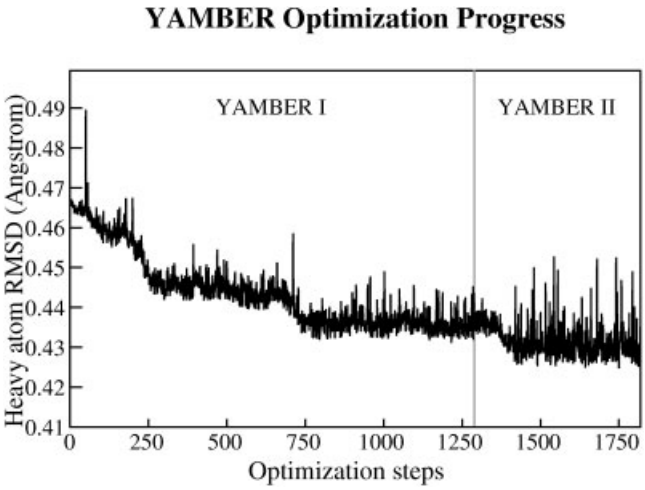

Fig. 2. Average heavy atom RMSD of 25 protein crystals after 5-ps simulated annealing runs, as a function of force-field optimization step.

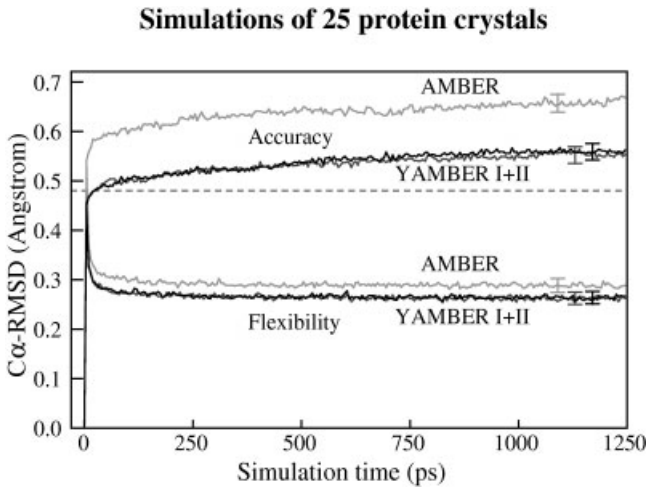

Fig. 3. Molecular dynamics simulations of 25 protein crystals with three different force fields. Shown is the average Cα RMSD from the starting structure (upper part, indicating accuracy) and from the previous simulation snapshot 5 ps before (lower part, indicating flexibility). YAMBER is shown in dark gray, YAMBER 2 in black. The horizontal line at 0.48 Å shows the limit of experimental uncertainty. Error bars were derived from the last 25% of the simulation by calculating the standard deviation for every protein and averaging over all proteins.

zation with the AMBER99 force field, the RMSD dropped quickly during the first 250 parameter optimization steps, until it hit an extended local minimum, from where it escaped around step 700. Then there was virtually no progress until step 1300, therefore we assumed convergence at 0.431 Å. This force field is called "YAMBER 1," and except for different parameters, it is still the same as AMBER. We then decided to switch bond length and angle parameters to the Engh&Huber dataset,[29] which required the introduction of 16 new atom types. This was mainly done to provide a slightly different starting point in search space and to avoid small systematic deviations from the WHAT_CHECK[30] standard values. Indeed, we found another noticeable improvement around step 1400, and finally stopped the procedure at step 1800, with an RMSD of 0.425 Å. This resulting force field is referred to as "YAMBER 2."

Force-Field Evaluation

During the force-field optimization procedure, the improvement at every step is very small. Consequently, simulated annealing runs had to be used, because the randomness of a true molecular dynamics simulation at constant temperature completely masks the progress signal. To investigate how the reduction in simulated annealing RMSD affects the behavior of the two YAMBER force fields under "real-life" conditions, we ran molecular dynamics simulations of another 25 protein crystals, sharing less than 30% sequence identity with those in the optimization set. The simulation temperatures were the same as during experimental structure determination. As can be seen from the top part of Figure 3, the Cα RMSDs from the starting structure after 1.25 ns are significantly lower for the YAMBER 1 (0.552 ± 0.017 Å) and YAMBER 2 (0.559 ± 0.017 Å) force fields, than for AMBER (0.657 ± 0.018 Å), which provided the initial parameters.

The horizontal line at 0.48 Å marks the border of experimental uncertainty (i.e., the average RMSD observed if the same structure is solved at high resolution by different research groups and refinement programs[15]). As soon as the RMSD crosses this line, one can say that a structure got worse during the simulation. The optimization procedure therefore allowed to bridge ~60% of the gap between the maximum allowed RMSD and the initially measured RMSD.

One could argue that the RMSD of single snapshots is not the ideal indicator of force-field accuracy, because the X-ray diffraction pattern provides only a time-averaged view of the protein. Consequently, one should compare it with the time-averaged structure in the simulation. We therefore superimposed the snapshots covering the last 25% of the simulation on the X-ray structure and averaged the atom coordinates. The resulting RMSDs are 0.611 for AMBER, and 0.506/0.517 Å for YAMBER1/2, which is only a small improvement considering the fact that the answer was implicitly given by providing the X-ray structure as a superposition target. The individual simulation snapshots are not equally spread around the true structure, but cluster at a different spot in conformational space, indicating that the source of the difference is indeed the force-field accuracy.

Any parameter optimization procedure carries the inherent danger of producing artifacts. With the simulated annealing approach used here, the YAMBER force fields were trained to have stable energy minima as close to true structures as possible. An alternative view is that they were optimized not to move proteins away from the starting structure, so one might ask if they move proteins at all. This question was answered by measuring the RMSD between two consecutive simulation snapshots (saved in 5-ps intervals), which is a good indicator of protein flexibility. The bottom part of Figure 3 shows that all three force fields lie close together at 0.288 ± 0.014 (AMBER), 0.262 ± 0.013 (YAMBER 1), and 0.263 ± 0.013

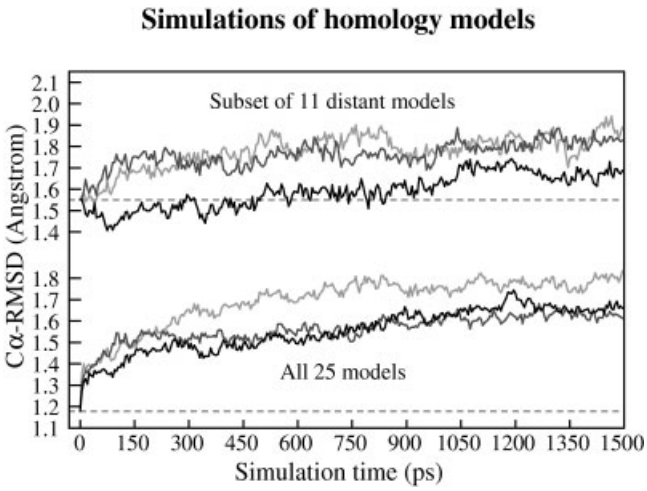

Fig. 4. Molecular dynamics simulations of 25 homology models in solution, using three different force fields. AMBER is shown in light gray, YAMBER in dark grey and YAMBER 2 in black. The average Cα-RMSD from the target is shown on the vertical axis, the dotted lines indicates the initial models before any refinement. The upper half shows a subset of 11 distant models, where template resolution divided by percentage sequence identity to the target is >0.04.

Å (YAMBER 2). The overall flexibility of proteins simulated with YAMBER is thus slightly smaller, but less than two standard deviations away from AMBER, while the accuracy increased by six standard deviations. To investigate this finding in more detail, we calculated the B-factors from the simulations and compared them to the experimental values. As expected, the in silico B-factors are lower, because 1-ns simulations are usually not enough to sample conformational space exhaustively (http://amber.scripps.edu/tutorial/integrase/loop13.htm), despite some encouraging findings.[31] The average B-factor is 17 for the experimental structures, 15 for AMBER, and 11 for YAMBER1/2. While proteins simulated with YAMBER are thus indeed "stiffer" on the 1-ns time-scale, they nevertheless show a better fit to the experimental data: the RMSD between calculated and experimental B-factors is 25 for AMBER, 21 for YAMBER 1, and 20 for YAMBER 2.

Refinement of Homology Models

To evaluate the performance of YAMBER in model refinement, we ran simulations for a set of 25 randomly chosen homology models, some of which are very similar to the target. We previously concluded that it does not make sense to indiscriminately refine all homology models. Only the more distant models, where template (X-ray) resolution divided by percentage sequence identity with the target is larger than 0.04, are good candidates for successful refinement.[15] Eleven (11) of the 25 models match this criterion and can be simulated reliably. The difference between these two sets is shown in Figure 4. As expected, one still cannot blindly run simulations for any model built. For the complete set, the Cα-RMSD from the targets increases during the unrestrained simulation: averaged over the entire simulation time (1.5 ns), it is 0.51 (AMBER) and 0.38/0.37 Å (YAMBER1/2) higher than the initial RMSD (bottom half of Fig. 4). While a temporary increase in RMSD may be required by individual models to overcome energy barriers, the fact that it happens immediately in the beginning and at a rather low starting RMSD of 1.35 Å indicates that the reason for the jump is the same as in Figure 3—a lack of force-field accuracy. The subset of distant models is shown in the upper half of Figure 4. Because it is smaller, the average RMSD is influenced more strongly by the random noise inherent to molecular dynamics simulations. Nevertheless, YAMBER 2 still performs best and actually can move models in the right direction, towards the target coordinates. After about 750 ps, the YAMBER 2 curve crosses the initial line again. However, this does not mean that models generally get worse after a certain simulation time. Closer inspection shows that after 750 ps, those models that can be improved reach a stable state, while the hopeless cases continue to go in the wrong direction and eventually pull the average across the line (e.g., after 750 ps, five of the eleven models still have a lower RMSD [by 0.190 Å on average], and keep this improvement till the end of the simulation [0.193 Å]).

The bad models can fortunately be identified using structure validation tools. During CASP5, we used WHAT_CHECK[30] to pick out the pearls, and at least for target 176, our model was ranked first among 150 submissions (http://predictioncenter.llnl.gov/casp5/pubresult5/CASP_BROWSER/DATA.html/ 3d_T0176_all.html).

DISCUSSION AND CONCLUSION

When looking at the final force-field parameters, one obvious question is: which parameters changed and why? In a best case scenario, all parameter changes would seem random, indicating the absence of systematic errors in the initial force field as well as in the optimization procedure. And essentially this was the case: Van der Waals radii shifted up in six and down in eight cases, charges increased in five and decreased in three cases. Two systematic changes were noted however: the scaling factors for the Van der Waals and electrostatic forces between 1–4 bonded atoms shifted both down considerably (from 0.5 and 0.83 to 0.27 and 0.65, respectively), and the Van der Waals contact energies increased in all cases, almost doubling for hydrogen and carbon atoms. In the AMBER force field, hydrogen and carbon have a very small contact energy (0.015 and 0.09–0.11 kcal/mol) when compared to nitrogen (0.17) and oxygen (0.21). In YAMBER 2, carbon and nitrogen came out equal (0.19 kcal/mol). This increase in Van der Waals attraction may explain the slightly lower flexibility of proteins simulated with the YAMBER force fields (Fig. 3). An additional reason may be the backbone hydrogens bonds, which got stronger during the parameter optimization (in contrast to other hydrogen bonds involving side-chain hydroxyl and amino groups that got weaker).

The fact that YAMBER 2 performed best not only in its trained area, the minimization of protein crystals (Fig. 2), but also in the quite different application of homology model refinement, reaffirms our previous finding that protein force-field parameters do not strongly depend on a particular field of research. Therefore it seems likely that

the YAMBER force fields will be more generally applicable.

In our previous work, the top improvement we found for any model was 0.25 Å. We concluded that the problem was not the force-field accuracy, but just the fact that the models "got stuck" too early during the simulated annealing minimization. Here, we used molecular dynamics simulations to search conformational space, which are much less likely to get trapped in local minima, thereby raising our hopes for a significantly better result. An analysis of all 75 model trajectories yielded quite a surprise: the highest improvement was found for a protein G model, simulated with YAMBER 2. The Cα-RMSD dropped from an initial value of 1.74 Å, which is typical for a close homology model, all the way down to 0.7 Å, which corresponds to a medium-resolution X-ray structure. So at least in this case, we could observe the metamorphosis of a model to an experimental-like structure.

ACKNOWLEDGMENTS

We would like to thank Herman Berendsen and Alan Mark for their helpful hints, and D.N. Ivankov, M.Yu. Lobanov, I. Litvinov, N.S. Bogatyreva, O.V. Galzitskaya, M.S. Kondratova, S.O. Garbuzinskii, M.A. Roytberg, who were team members at CASP-5. We also thank all colleagues at the CMBI for using the Models@Home screensaver, which provided the computational resources. This work was supported in part by the European community (5^{th} Framework program NMRQUAL Contract Number QLG2-CT-2000-01313), by NWO grant 047.009.021, by the program "Physical and Chemical Biology" of the Russian Academy of Sciences and by an International Research Scholar's Award to A.V.F. from the Howard Hughes Medical Institute.

REFERENCES

1. Karplus M, McCammon JA. Molecular dynamics simulations of biomolecules. Nat Struct Biol 2002;9:646–652.
2. de Groot BL, van Aalten DM, Scheek RM, Amadei A, Vriend G, Berendsen HJ. Prediction of protein conformational freedom from distance constraints. Proteins 1997;29:240–251.
3. el-Bastawissy E, Knaggs MH, Gilbert IH. Molecular dynamics simulations of wild-type and point mutation human prion protein at normal and elevated temperature. J Mol Graph Model 2001;20: 145–154.
4. 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.
5. MacKerell Jr., A.D., Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 1998;102:3586–3616.
6. Van Gunsteren WF, Billeter SR, Eising AA, Hünenberger PH, Krüger P, Mark AE, Scott WRP, Tironi IG. Biomolecular simulation: The GROMOS96 , manual and user guide. vdf Hochschulverlag, ETH Zürich; 1996.
7. Tramontano A, Leplae R, Morea V. Analysis and assessment of comparative modeling predictions in CASP4. Proteins 2001; Suppl.5:22–38.
8. Sanchez R, Sali A. ModBase: a database of comparative protein structure models. Bioinformatics 1999;15:1060–1061.
9. Peitsch MC, Schwede T, Guex N. Automated protein modelling—the proteome in 3D. Pharmacogenomics 2000;1:257–266.
10. Teichmann SA, Murzin AG, Chothia C. Determination of protein function, evolution and interactions by structural genomics. Curr Opin Struct Biol 2001;11:354–363.
11. Huang ES, Subbiah S, Tsai J, Levitt M. Using a hydrophobic contact potential to evaluate native and near-native folds generated by molecular dynamics simulations. J Mol Biol 1996;257:716–725.
12. Lee MR, Tsai J, Baker D, Kollman PA. Molecular dynamics in the endgame of protein structure prediction. JMolBiol 2001;313:417–430.
13. Wang J, Cieplak P, Kollman PA. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? J Comp Chem 2000;21:1049–1074.
14. Ren P, Ponder JW. Consistent treatment of inter- and intramolecular polarization in molecular mechanics calculations. J Comp Chem 2002;23:1497–1506.
15. Krieger E, Koraimann G, Vriend G. Increasing the precision of comparative models with YASARA NOVA—a self-parameterizing force field. Proteins 2002;47:393–402.
16. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids Res 2000;28:235–242.
17. Vassylyev DG, Katayanagi K, Ishikawa K, Tsujimoto-Hirano M, Danno M, Pahler A, Matsumoto O, Matsushima M, Yoshida H, Morikawa K. Crystal structures of ribonuclease F1 of Fusarium moniliforme in its free form and in complex with 2′GMP. J Mol Biol 1993;230:979–996.
18. Hooft RWW, Sander C, Vriend G. Positioning hydrogen atoms by optimizing hydrogen-bond networks in protein structures. Proteins 1996;26:363–376.
19. Hooft RWW, Sander C, Scharf M, Vriend G. The PDBFINDER database: a summary of PDB, DSSP and HSSP information with added value. Comput Appl Biosci 1996;12:525–529.
20. Hooft RWW, Sander C, Vriend G. Verification of protein structures: side-chain planarity. J Appl Cryst 1996;29:714–716.
21. Vriend G. WHAT IF—A molecular modeling and drug design program. J Mol Graph 1990;8:52–56.
22. Walser R, Hunenberger PH, Van Gunsteren WF. Comparison of different schemes to treat long-range electrostatic interactions in molecular dynamics simulations of a protein crystal. Proteins 2001;43:509–519.
23. Essman U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald method. J Chem Phys 1995;103: 8577–8593.
24. Van Gunsteren WF, Karplus M. Protein dynamics in solution and in a crystalline environment: a molecular dynamics study. Biochemistry 1982;21:2259–2274.
25. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. J Chem Phys 1984;81:3684–3690.
26. Eastman P, Pellegrini M, Doniach S. Protein flexibility in solution and in crystals. J Chem Phys 1999;110:10141–10152.
27. Liu H, Elstner M, Kaxiras E, Frauenheim T, Hermans J, Yang W. Quantum mechanics simulation of protein dynamics on long timescale. Proteins 2001;44:484–489.
28. Krieger E, Vriend G. Models@Home: distributed computing in bioinformatics using a screensaver based approach. Bioinformatics 2002;18:315–318.
29. Engh RA, Huber R. Accurate bond and angle parameters for X-ray protein structure refinement. Acta Crystallogr A 1991;47:392–400.
30. Hooft RWW, Vriend G, Sander C, Abola EE. Errors in protein structures. Nature 1996;381:272–272.
31. Clarage JB, Romo T, Andrews BK, Pettitt BM, Phillips GNJ. A sampling problem in molecular dynamics simulations of macromolecules. Proc Natl Acad Sci USA 1995;92:3288.