Crystallographic R Factor Refinement by Molecular Dynamics

Axel T. Brunger, John Kuriyan, Martin Karplus

Science, New Series, Volume 235, Issue 4787 (Jan. 23, 1987), 458-460.

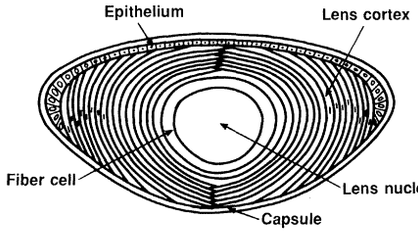

Fig. 3. Schematic cross section of a mammalian lens.

quences from −75 to −26 are highly conserved among the different members of the γ-crystallin gene family (*11, 13*), while those further upstream show little homology (*6*), it is possible that the latter sequences are involved in the differential expression of the γ-crystallin genes during lens development.

Overbeek and co-workers (*14*) have previously shown that a DNA fragment containing the mouse αA-crystallin promoter coupled to the bacterial gene for chloramphenicol acetyltransferase (CAT) directed tissue-specific expression of CAT as detected in extracts prepared from both lens epithelium and fiber cells. However, this approach did not provide the detailed in situ resolution afforded by the β-gal assay described here. While β-gal constructs have been used successfully in P-element transformation of *Drosophila* (*15*), our study demonstrates that β-gal can be used to localize gene activity in situ in transgenic mice.

Slit lamp examination of the lenses of mice expressing the γ2-crystallin–β-gal hybrid gene revealed no evidence of cataract formation, suggesting that high level expression of the bacterial enzyme is not deleterious to lens function. We cannot exclude the possibility that high levels of β-gal activity are injurious in other developmental contexts, a possibility that is being investigated with various promoter–β-gal constructs in transgenic mice. The use of β-gal as a reporter gene in transgenic mice also assumes that post-transcriptional processes affecting its expression would be the same in all cell types. At the present time this is only an assumption. The lower limits of enzyme activity detectable in tissue sections remain to be determined. However, β-gal–based hybrid genes hold considerable potential for studying the developmental regulation of gene expression in situ in transgenic mice.

REFERENCES AND NOTES

1. M. L. Breitman *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 7762 (1984); J. W. McAvoy, *Differentiation* **17**, 137 (1980); D. S. McDevitt and S. K. Brahma, in *Cell Biology of the Eye*, D. S. McDevitt, Ed. (Academic Press, New York, 1982), pp. 143–191.
2. J. Piatigorsky, *Differentiation* **19**, 134 (1981).
3. I. Kabasawa, Y. Tsunematsu, G. W. Barber, K. Kinoshita, *Exp. Eye Res.* **24**, 437 (1977).
4. J. Papaconstantinou, *Science* **156**, 338 (1967); C. Slinsby and L. R. Miller, *Exp. Eye Res.* **37**, 517 (1983); J. S. Zigler *et al.*, *Invest. Ophthalmol. Vis. Sci.* **26**, 525 (1985).
5. S. Lok *et al.*, *Mol. Cell. Biol.* **5**, 2221 (1985).
6. S. O. Meakin, L.-C. Tsui, M. L. Breitman, in preparation.
7. C. V. Hall, P. E. Jacob, G. M. Ringold, F. Lee, *J. Mol. Appl. Genet.* **2**, 101 (1983).
8. B. D. Lake, *Histochem. J.* **6**, 211 (1974).
9. D. R. Goring *et al.*, unpublished data.
10. J. W. McAvoy, *J. Embryol. Exp. Morph.* **44**, 149 (1978); ———, *ibid.* **45**, 271 (1978).
11. M. Murer-Orlando, R. C. Patterson, S. Lok, L.-C. Tsui, M. L. Breitman, *Dev. Biol.*, in press.
12. S. Lok, M. L. Breitman, L.-C. Tsui, unpublished data.
13. S. Lok *et al.*, *Nucleic Acids Res.* **12**, 4517 (1984); S. O. Meakin, M. L. Breitman, L.-C. Tsui, *Mol. Cell. Biol.* **5**, 1408 (1985).
14. P. A. Overbeek, A. B. Chepelinsky, J. S. Khillan, J. Piatigorsky, H. Westphal, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 7815 (1985).
15. J. T. Lis, J. A. Simon, C. A. Sutton, *Cell* **35**, 403 (1983).
16. The γ2-crystallin sequences were excised from pγ2CAT1281 (*5*) and inserted into the Hind III site in the β-gal vector pCH126. The plasmid pCH126 is identical to pCH110 as described (*7*) except that the SV40 promoter (Pvu II–Hind III fragment) was removed and the Hind III site regenerated (F. Lee, personal communication).
17. Three-week-old animals were screened for the insertions of the γ2-crystallin–β-gal hybrid gene by Southern blot analysis [E. M. Southern, *J. Mol. Biol.* **98**, 503 (1975)] of tail DNA digested with Pst I, a non-cutter of the test construct. The probe was a 3.7-kb Hind III–Bam HI fragment spanning the *lacZ* fusion gene (see Fig. 1A) and labeled with the random-priming procedure [A. P. Feinberg and B. Vogelstein, *Anal. Biochem.* **132**, 6 (1983)]. The single integrations detected in animals B1, B3, B4, and B5 and the two present in B2 were subsequently characterized by digestion of tail DNA with Sac I and Hpa I; enzymes that produced one and two cuts in the hybrid gene, respectively (see Fig. 1A). Digested DNA's (5 μg) were separated by electrophoresis through 0.7% agarose gels, transferred to Zetabind filters (AMF Cuno), and prehybridized and hybridized as described [G. M. Wahl, M. Stern, G. R. Stark, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 3683 (1979)]. Filters were washed at room temperature once in 2× standard saline citrate (SSC), 0.1% SDS and once in 0.1× SSC, 0.1% SDS; followed by four washes at 65°C in 0.1× SSC, 0.1% SDS. Dried filters were exposed to Kodak XAR-5 film for 1 to 3 days at −70°C between two Dupont Lightning Plus intensifying screens.
18. Whole eyes were frozen in liquid nitrogen and mounted in OCT compound (Fisher) for cryosectioning. Sections (7 μm) were cut, air-dried, and incubated overnight in a buffer containing X-gal [2 m*M* 5-bromo-4-chloro-3-indolyl-β-D-galactoside (Sigma), 3 m*M* potassium ferricyanide, and 3 m*M* potassium ferrocyanide in phosphate-buffered saline, *p*H 7.0]. Sections were then counterstained with hematoxylin and eosin and mounted in Canada balsam. Tissue sections prepared similarly from brain, heart, liver, kidney, and spleen showed no detectable bacterial enzyme activity, although endogenous β-gal activity was detectable under a different assay condition with McIlvaine buffer (120 m*M* NaCl, 3 m*M* potassium ferricyanide, 3 m*M* potassium ferrocyanide, and 2 m*M* X-gal, *p*H 4.2) (D. R. Goring *et al.*, unpublished data).
19. We wish to thank F. Lee of DNAX, Inc., Palo Alto, CA, for providing the plasmid pCH126; M. Musarella for slit lamp examination of transgenic mice; R. Worton, L. Siminovitch, and A. Bernstein for encouragement; and R. Gravel for helpful discussions. This research was supported by grants from the Medical Research Council (MRC) of Canada (L.-C.T., M.L.B., and J.R.), National Cancer Institute (NCI) of Canada (J.R.), and the Natural Sciences and Engineering Research Council of Canada (J.R.). D.G. is a recipient of a Student Fellowship from MRC. J.R. is a NCI Research Associate. M.L.B. and L.-C.T. are Research Scholars of MRC and the Canadian Cystic Fibrosis Foundation, respectively.

16 September 1986; accepted 21 November 1986

Crystallographic *R* Factor Refinement by Molecular Dynamics

AXEL T. BRÜNGER, JOHN KURIYAN, MARTIN KARPLUS

Molecular dynamics was used to refine macromolecular structures by incorporating the difference between the observed crystallographic structure factor amplitude and that calculated from an assumed atomic model into the total energy of the system. The method has a radius of convergence that is larger than that of conventional restrained least-squares refinement. Test cases showed that the need for manual corrections during refinement of macromolecular crystal structures is reduced. In crambin, the dynamics calculation moved residues that were misplaced by more than 3 angstroms into the correct positions without human intervention.

CRYSTALLOGRAPHIC STRUCTURE DETERMINATIONS by x-ray or neutron diffraction generally proceed in two stages. First, the phases of the measured reflections are estimated and a low- to medium-resolution model of the protein is constructed and second, more precise information about the structure is obtained by refining the parameters of the molecular model against the crystallographic data (*1*). The refinement is performed by minimizing the crystallographic *R* factor, which is defined as the difference between the observed [$|F_{obs}(h\ k\ l)|$] and calculated [$|F_{calc}(h\ k\ l)|$] structure factor amplitudes,

$$R = \frac{\sum_{h\,k\,l} ||F_{obs}(h\ k\ l)| - |F_{calc}(h\ k\ l)||}{\sum_{h\,k\,l} |F_{obs}(h\ k\ l)|} \quad (1)$$

where $h\ k\ l$ are the reciprocal lattice points of the crystal.

Conventional refinement involves a series of steps, each of which consists of a few cycles of least-squares refinement with stereochemical and internal packing constraints or restraints (*2–5*) that are followed

A. T. Brünger and M. Karplus, Department of Chemistry, Harvard University, Cambridge, MA 02138.
J. Kuriyan, Department of Chemistry, Harvard University, Cambridge, MA 02138, and Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139.

by rebuilding the model structure with interactive computer graphics (*6*). Finally, solvent molecules are included and alternative conformations for some atoms in the protein may be introduced. Conventional refinement is time-consuming, because the limited radius of convergence of least-squares algorithms necessitates the periodic examination of electron density maps which are computed with various combinations of F_{obs} and F_{calc} as amplitudes and with phases calculated from the model structure. The radius of convergence is given theoretically by $d/4$, where d is the highest resolution reflection included in the least-squares refinement (*5*). Although reduction of the resolution increases the radius of convergence, it also decreases the information available from the experimental data. Restrained least-squares refinement in general does not correct residues that are displaced by more than 1 Å. Also, the least-squares refinement process is easily trapped in a local minimum so that human intervention is required. We refer to the approach that combines restrained least-squares refinement with model-building as "manual refinement."

Simulated annealing (*7*), which makes use of molecular dynamics to explore the conformational space of the molecule, can help overcome the local-minimum problem. This has been shown in the application of molecular dynamics to structure refinement with nuclear magnetic resonance (NMR) data (*8*). However, in contrast to the NMR application (*8*), the initial model for crystallographic refinement cannot be arbitrary. It has to be relatively close to the correct geometry to provide an adequate approximation to the phases of the structure factors.

Molecular dynamics simulations were performed by solving Newton's equations of motion (*9*) with forces on the atoms derived from an empirical potential energy that described stereochemical and nonbonding interactions (*10*). Molecular dynamics was incorporated into the crystallographic refinement by adding the effective potential energy

$$E_{sf} = S\sum_{h\,k\,l} [|F_{obs}(h\,k\,l)| - |F_{calc}(h\,k\,l)|]^2 \quad (2)$$

to the empirical potential energy used in the CHARMM program (*11*). The effective potential energy E_{sf} describes the differences between the observed structure factor amplitudes and those calculated from the atomic model; it is identical to the function used in standard least-squares refinement methods (*5*). The scale factor S was chosen such that the gradient of E_{sf} was comparable in magnitude to the gradient of the empirical potential energy of a molecular dynamics simulation in which S was set to zero. To reduce the computational time required to compute E_{sf} and its derivatives, we used a harmonic approximation for the dependence of E_{sf} on the atomic coordinates (*5*); once any atom moved by more than 0.4 Å, E_{sf} and the derivatives for all atomic coordinates were recomputed. We refer to refinement by molecular dynamics with the effective potential energy E_{sf} as "MD-refinement." Although variable temperature factors and solvent molecules can be included in a straightforward manner, the present applications used a constant temperature factor and did not include solvent.

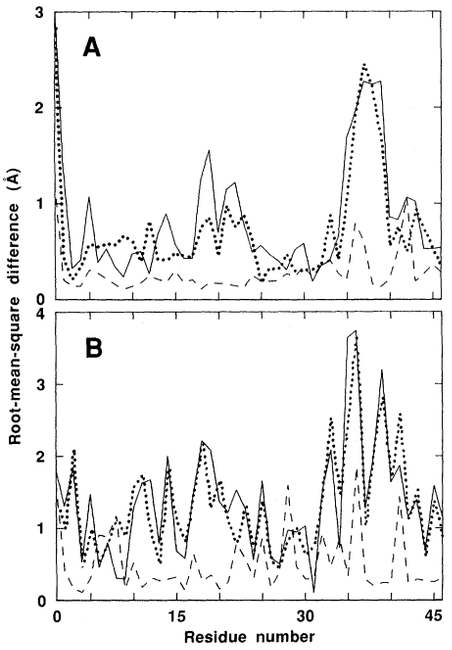

Fig. 1. Atomic rms difference from the manually refined x-ray structure of crambin for the initial NMR-derived structure (——), the restrained least-squares-only refined structure (. . . .), and the MD-refined structure (– – –); the rms differences are plotted as a function of residue number for (**A**) backbone (C, N, Cα) and (**B**) side chain atoms.

As a first test the method was applied to crambin, a small protein of 46 amino acids, for which high-resolution x-ray diffraction data and a refined structure (determined by resolved anomalous phasing and conventional least-squares refinement with model-building) are available (*12*). The initial structure for the MD-refinement was obtained from an NMR structure determination that used simulated data (*8*); the orientation and position of the NMR-derived crambin molecule in the unit cell was determined by molecular replacement (*13*). The root-mean-square (rms) differences for residue positions of this initial structure and the final manually refined structure (*12*) are as large as 3.5 Å, with particularly large differences for residues 34 to 40 (Fig. 1); the R factor of the initial structure is 0.56 at 2 Å resolution (Table 1). MD-refinement at 3000 K started with 4 Å resolution data for 2.5 psec and was extended to 3 Å resolution for 2.5 psec and finally to 2 Å resolution for 5 psec. This procedure was followed by several cycles of minimization that reduced the atomic rms deviations to 0.34 and 0.56 Å for the backbone and side chain atoms, respectively (Table 1). The MD-refined structure has only a few significant differences from the manually refined structure (Fig. 1). During the MD-refinement, some atoms in residues 35 to 40 moved by more than 3 Å (Fig. 1B). The refinement of the crambin structure has been achieved by starting from the initial NMR-structure and continuing without human intervention.

Table 1. Refinement of crambin and of α-amylase inhibitor. The R factor for crambin was computed with constant temperature factors (B = 10 Å^2 for side chain atoms, B = 6 Å^2 for backbone atoms) at 2 Å resolution; the R factor for α-amylase inhibitor was computed with constant temperature factors (B = 20 Å^2 for side chain atoms, B = 15 Å^2 for backbone atoms) at 2.06 Å resolution.

Structure	R factor	Δ_{bond}* (Å)	Δ_{angle}* (degrees)	rms difference (Å)†	
				Backbone	Side chain
Crambin					
Initial	0.56	0.017	4.13	1.06	1.40
Least-squares refined‡	0.381	0.086	11.06	0.92	1.30
MD-refined	0.294	0.029	4.48	0.34	0.56
Manually refined§	0.258	0.014	2.53	0.08	0.12
α-Amylase inhibitor					
Initial	0.405	0.024	3.36	0.40	0.97
Least-squares refined‡	0.305	0.017	3.49	0.22	0.85
MD-refined	0.271	0.015	3.03	0.23	0.89
Manually refined§	0.272	0.014	2.88	0.12	0.19

*Root-mean-square deviations of bond lengths and bond angles from ideality. †Atomic rms differences from the manually refined structures (*12, 15*). ‡For crambin, 6, 13, and 14 cycles of PROLSQ (*2*) at 4, 3, and 2 Å resolution without model-building were used; the large values for Δ_{bond} and Δ_{angle} in the table are caused by loose stereochemical restraints that were applied to obtain a lower R factor. For the α-amylase inhibitor, 50 cycles of conjugant gradient minimization (*11*) that included E_{sf} (Eq. 2) at 2.06 Å resolution without model-building were used. §For comparison, the manually refined structures (*12, 15*) were subjected to refinement without solvent and with constant temperature factors.

The R factor (0.294) of the MD-refined structure is somewhat higher than the R factor (0.258) of the manually refined structure without solvent and with constant temperature factors; minor model-building is required to correct this difference. The refinement required approximately 1 hour of central processing unit (CPU) time on a CRAY 1; structure factor calculations accounted for about half this time.

As a control, the initial NMR-derived structure was refined without rebuilding by a restrained least-squares method (*2*) that started at 4 Å resolution and then increased the resolution to 3 Å, and finally to 2 Å. The R factor dropped to 0.381, but the very bad stereochemistry and large deviation from the manually refined structure (Table 1) indicated that this structure did not converge to the correct result; residues 34 to 40 did not move (Fig. 1) and substantial model-building would be required to correct the structure. Thus, restrained least-squares refinement in the absence of model-building did not produce the large conformational changes that occurred in MD-refinement.

In another application, MD-refinement was used with human α-lactalbumin, a protein composed of 123 amino acids, for which an initial structure had been obtained by molecular replacement using baboon α-lactalbumin, phasing with the molecular replacement model and preliminary model-building (*14*); conventional refinement had not been done. The initial and the MD-refined structure have R factors of 0.537 and 0.276 at 2.5 Å resolution, respectively; the deviations of bond lengths from ideality of the initial and the MD-refined structure are 0.042 and 0.034 Å, respectively. The rms difference between the initial and MD-refined structure is 1.06 Å and 2.05 Å for backbone and side chain atoms, respectively; eight atoms moved by more than 5 Å during the MD-refinement.

The applicability of the method to a case for which a good initial model is available was examined by using MD-refinement with the α-amylase inhibitor Hoe-467A, a protein composed of 74 amino acids, for which high-resolution x-ray diffraction data and a refined structure exist (*15*); the results are given in Table 1. The initial structure was built from an isomorphous-replacement electron density map and has a relatively low R factor (0.405 at 2.06 Å resolution). MD-refinement was performed at 2.06 Å resolution for 2.5 psec followed by minimization. MD-refinement produced a better R factor than least-squares refinement without model-building (0.271 compared to 0.305, see Table 1) and the same value as manual model-building without solvent and with constant temperature factors.

REFERENCES AND NOTES

1. H. W. Wyckoff, C. H. W. Hirs, S. N. Timasheff, Eds., *Diffraction Methods for Biological Macromolecules, Part B* of *Methods Enzymol.* **115** (1985).
2. W. A. Hendrickson, in (*1*), p. 252; J. H. Konnert and W. A. Hendrickson, *Acta Crystallogr. Sect. A* **36**, 614 (1980).
3. D. S. Moss and A. J. Morffew, *Comput. Chem.* **6**, 1 (1982).
4. J. L. Sussman, S. R. Holbrook, G. M. Church, S. H. Kim, *Acta Crystallogr. Sect. A* **33**, 800 (1977).
5. A. Jack and M. Levitt, *ibid.* **34**, 931 (1978).
6. T. A. Jones, in *Computational Crystallography*, D. Sayre, Ed. (Clarendon, Oxford, 1982), p. 303.
7. S. Kirkpatrick, C. D. Gelatt, Jr., M. P. Vecchi, *Science* **220**, 671 (1983).
8. A. T. Brünger, G. M. Clore, A. M. Gronenborn, M. Karplus, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 3801 (1986).
9. L. Verlet, *Phys. Rev.* **159**, 98 (1967).
10. M. Karplus and J. A. McCammon, *Annu. Rev. Biochem.* **52**, 263 (1983).
11. B. R. Brooks *et al.*, *J. Comput. Chem.* **4**, 187 (1983).
12. W. A. Hendrickson and M. M. Teeter, *Nature (London)* **290**, 107 (1981).
13. A. T. Brünger *et al.*, *Science*, in press.
14. R. E. Fenna, personal communication.
15. J. W. Pflugrath *et al.*, *J. Mol. Biol.* **189**, 383 (1986).

16. We thank M. M. Teeter (*12*) and J. W. Pflugrath (*15*) for providing data, and G. M. Clore, R. E. Fenna, R. Huber, G. A. Petsko, and J. W. Pflugrath for useful discussions. Supported in part by a grant from the National Science Foundation. Computer time on the CRAY 1 and CRAY 2 at the Minnesota Supercomputer Center was supplied by a grant from the Office of Advanced Scientific Computing of the National Science Foundation.

2 September 1986; accepted 26 November 1986

Thunderstorms: An Important Mechanism in the Transport of Air Pollutants

R. R. DICKERSON, G. J. HUFFMAN, W. T. LUKE, L. J. NUNNERMACKER, K. E. PICKERING, A. C. D. LESLIE, C. G. LINDSEY, W. G. N. SLINN, T. J. KELLY, P. H. DAUM, A. C. DELANY, J. P. GREENBERG, P. R. ZIMMERMAN, J. F. BOATMAN, J. D. RAY, D. H. STEDMAN

Acid deposition and photochemical smog are urban air pollution problems, and they remain localized as long as the sulfur, nitrogen, and hydrocarbon pollutants are confined to the lower troposphere (below about 1-kilometer altitude) where they are short-lived. If, however, the contaminants are rapidly transported to the upper troposphere, then their atmospheric residence times grow and their range of influence expands dramatically. Although this vertical transport ameliorates some of the effects of acid rain by diluting atmospheric acids, it exacerbates global tropospheric ozone production by redistributing the necessary nitrogen catalysts. Results of recent computer simulations suggest that thunderstorms are one means of rapid vertical transport. To test this hypothesis, several research aircraft near a midwestern thunderstorm measured carbon monoxide, hydrocarbons, ozone, and reactive nitrogen compounds. Their concentrations were much greater in the outflow region of the storm, up to 11 kilometers in altitude, than in surrounding air. Trace gas measurements can thus be used to track the motion of air in and around a cloud. Thunderstorms may transform local air pollution problems into regional or global atmospheric chemistry problems.

OZONE (O_3) CONTROLS MUCH OF the chemistry of the global atmosphere. When the sun shines on air polluted with automobile exhaust, photochemical reactions of nitrogen dioxide (NO_2), nonmethane hydrocarbons, and carbon monoxide (CO) produce "Los Angeles–type smog" containing high concentrations of O_3 and other oxidants. When these oxidizing agents then react with NO_2 to form nitric acid (HNO_3) and with SO_2 to form sulfuric acid (H_2SO_4), acid rain can result. On a local scale, photochemical smog and acid rain (more accurately called acid deposition) are serious environmental problems; how far these phenomena extend on the regional or global scale is a major unanswered question in atmospheric sciences (*1–3*).

Pollutants travel farther at higher altitudes. In the lowest kilometer of the atmosphere, the planetary boundary layer (PBL), friction with the earth's surface reduces wind speeds. A temperature inversion often isolates the air at the top of the PBL from the rest of the troposphere; the troposphere,

R. R. Dickerson, G. J. Huffman, W. T. Luke, L. J. Nunnermacker, K. E. Pickering, Department of Meteorology, University of Maryland, College Park, MD 20742.
A. C. D. Leslie, C. G. Lindsey, W. G. N. Slinn, Pacific Northwest Laboratory, Richland, WA 99352.
T. J. Kelly and P. H. Daum, Brookhaven National Laboratory, Upton, NY 11973.
A. C. Delany, J. P. Greenberg, P. R. Zimmerman, National Center for Atmospheric Research, Boulder, CO 80307.
J. F. Boatman, National Oceanic and Atmospheric Administration, Boulder, CO 80303.
J. D. Ray and D. H. Stedman, Department of Chemistry, University of Denver, Denver, CO 80210.