Several dilutions of Mbo I (New England Biolabs) were tested to obtain a range of DNA fragments in the 20- to 1000-kb size range; a concentration of 0.025 unit per microgram of DNA gave optimum results. A scaled-up digest was done on the large agarose block, which was then placed in a preparative 1% agarose PFGE gel for DNA sizing, using annealed λ phage DNA size markers (*24*). The gel band corresponding to the approximate size range of 80 to 130 kb was cut out, placed in a dialysis bag with electrophoresis buffer, and electroeluted in the PFGE box. The gel fragment was removed, and the DNA was extensively dialyzed against TE. Analysis of a portion by PFGE confirmed that the general range of DNA sizes was in the expected distribution, and that the DNA had not sheared. Four micrograms of the size-selected DNA was further diluted to a concentration of 0.22 μg/ml and 2 μg of a purified 219-bp Bam HI–ended supF gene (*14, 17*) was added. After equilibration in ligase buffer for 2 hours, T_4 DNA ligase (New England Biolabs) was added to a concentration of 1.6 U/μl and ligation was carried out for 12 hours at 4°C. A second aliquot of ligase was then added and another 12-hour ligation was carried out. After addition of 20 μg of carrier transfer RNA, the DNA was precipitated with ethanol, resuspended in 100 μl of TE, and digested to completion with Eco RI. After phenol extraction and ethanol precipitation, the genomic DNA was ligated into a 3:1 molar excess of Eco RI–cut λCh3AΔlac. This is an A*am* B*am* phage that will accept Eco RI inserts from 0 to 12 kb; this vector was generated by replacing the Eco RI–Bam HI fragment of λCh3A (*35*) with the Eco RI–Bgl II polylinker from the miniplasmid πVX (*36*). The ligated DNA was packaged in vitro (*37*) and plated on the sup$^-$ strain MC1061. A total of 2 million clones was obtained; a study of ten random clones indicated that all contained from one to three supF genes within the phage insert, in each instance flanked by genomic DNA segments.

17. R. J. Dunn *et al.*, *J. Biol. Chem.* **256**, 6109 (1981).
18. C. S. Cooper *et al.*, *Nature (London)* **311**, 29 (1984).
19. M. Dean *et al.*, *ibid.* **318**, 385 (1985).
20. M. Park *et al.*, *Cell* **45**, 895 (1986).
21. F. Collins, J. Cole, M. Iannuzzi, J. Richards, unpublished data.
22. D. C. Schwartz and C. R. Cantor, *Cell* **37**, 67 (1984).
23. G. F. Carle and M. V. Olson, *Nucleic Acids Res.* **12**, 5647 (1984).
24. C. L. Smith *et al.*, in *Genetic Engineering*, J. K. Setlow and A. Hollaender, Eds. (Plenum, New York, 1986), vol. 8, pp. 45–70.
25. C. L. Smith and C. R. Cantor, *Methods Enzymol.*, in press.
26. G. F. Carle, M. Frank, M. V. Olson, *Science* **232**, 65 (1986).
27. K. Gardiner, W. Laas, D. Patterson, *Somatic Cell Mol. Genet.* **12**, 185 (1986).
28. R. White *et al.*, *Cold Spring Harbor Symp. Quant. Biol.*, in press.
29. P. J. Scambler, H.-Y. Law, R. Williamson, C. S. Cooper, *Nucleic Acids Res.* **14**, 7159 (1986).
30. M. Park *et al.*, *Cold Spring Harbor Symp. Quant. Biol.*, in press.
31. A. Poustka *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 4129 (1984).
32. P. F. R. Little and S. H. Cross, *ibid.* **82**, 3159 (1985).
33. A. R. Wyman, L. B. Wolfe, D. Botstein, *ibid.*, p. 2880.
34. A. Poustka *et al.*, *Nature (London)* **325**, 353 (1987).
35. F. R. Blattner *et al.*, *Science* **196**, 161 (1977).
36. B. Seed, *Nucleic Acids Res.* **11**, 2427 (1983).
37. B. Hohn and K. Murray, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 3259 (1977).
38. A. P. Feinberg and B. Vogelstein, *Analyt. Biochem.* **137**, 266 (1984).

39. We thank S. Weissman and D. Ginsburg for valuable discussions; B. Handelin, T. Lugo, K. Fournier, and D. Housman for supplying the hybrid cell line HDm20s; F. Blattner for the λCh3AΔlac phage vector; and M. Roth for assistance with PFGE. This research was supported by NIH grant GM34960 and a Hartford Foundation fellowship grant to F.S.C.

15 December 1986; accepted 29 January 1987

Solution of a Protein Crystal Structure with a Model Obtained from NMR Interproton Distance Restraints

AXEL T. BRÜNGER, ROBERT L. CAMPBELL, G. MARIUS CLORE, ANGELA M. GRONENBORN, MARTIN KARPLUS, GREGORY A. PETSKO, MARTHA M. TEETER

Model calculations were performed to test the possibility of solving crystal structures of proteins by Patterson search techniques with three-dimensional structures obtained from nuclear magnetic resonance (NMR) interproton distance restraints. Structures for crambin obtained from simulated NMR data were used as the test system; the root-mean-square deviations of the NMR structures from the x-ray structure were 1.5 to 2.2 Å for backbone atoms and 2.0 to 2.8 Å for side-chain atoms. Patterson searches were made to determine the orientation and position of the NMR structures in the unit cell. The correct solution was obtained by comparing the rotation function results of several of the NMR structures and the average structure derived from them. Conventional refinement techniques reduced the *R* factor from 0.43 at 4 Å resolution to 0.27 at 2 Å resolution without inclusion of water molecules. The partially refined structure has root-mean-square backbone and side-chain atom deviations from the x-ray structure of 0.5 and 1.3 Å, respectively.

IN PROTEIN CRYSTALLOGRAPHY, THE initial determination of phases by multiple isomorphous replacement may be difficult because heavy-atom derivatives of the crystal are not available. If the structure of a similar or homologous molecule is known, the "molecular replacement method" (*1–3*) can be tried for the initial phasing. It involves the rotation and translation of the known structure in the unit cell of the target crystal to obtain the best match between the Patterson function calculated from the model and that from the observed diffraction data. The optimally oriented and translated model structure is used to begin conventional x-ray crystallographic refinement; this approach may or may not succeed (*1–3*). Criteria for determining when the molecular replacement method should work are available only for specific cases (*3*).

With the use of nuclear magnetic resonance (NMR) techniques (*4, 5*) a large number of approximate interproton distances for certain proteins (up to 80 residues) can be obtained (*6, 7*) and used to build three-dimensional structural models. Because of the limitations on the number and the range (only <5 Å) of these distances, the NMR data must be complemented with information about covalent structure (derived from bond lengths, bond angles, and dihedral angles) and internal packing requirements (derived from van der Waals repulsions). Restrained molecular dynamics, in which the energy function is augmented by effective potentials obtained from interproton distances (*8–10*), can add sufficient information to determine the solution structure of a protein. The method has been shown to yield converged structures with root-mean-square (rms) deviations from the x-ray structure of 1.5 to 3.0 Å (*9, 10*). Similar results have been obtained with distance geometry algorithms (*11, 12*). Recently the structure of the α-amylase inhibitor tendamistat has been solved independently by x-ray crystallography and NMR spectroscopy (*13, 14*); a comparison of the two structures indicates reasonable agreement.

We show in this report that structures obtained for a protein in solution with NMR data can be used to solve the crystal structure of the same protein by molecular replacement (*1–3*). As a test we use crambin (*15*), a protein of 46 residues, for which NMR structures with simulated data have been obtained (*9, 10*). We show that the individual NMR structures are not sufficiently accurate to solve the rotation and translation functions uniquely. Use of the refined x-ray structure itself as a model demonstrates that crambin is a difficult case for the molecular replacement method. We obtained the correct solution when we averaged over an ensemble of NMR structures or used the average structure derived from them. We demonstrate that the averaged

A. T. Brünger and M. Karplus, Department of Chemistry, Harvard University, Cambridge, MA 02138.
R. L. Campbell and G. A. Petsko, Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139.
G. M. Clore and A. M. Gronenborn, Max-Planck-Institut für Biochemie, D-8033 Martinsried bei München, Federal Republic of Germany.
M. M. Teeter, Department of Chemistry, Boston College, Chestnut Hill, MA 02167.

Table 1. Atomic rms differences between the NMR structures and the x-ray structure.

NMR structure	Atomic rms difference (Å)	
	Backbone	Side chain
S_1	1.4	2.2
S_2	1.5	2.2
S_3	2.0	2.6
S_4	1.8	2.7
S_5	1.5	2.2
$S_{\langle i \rangle}$	1.1	1.4

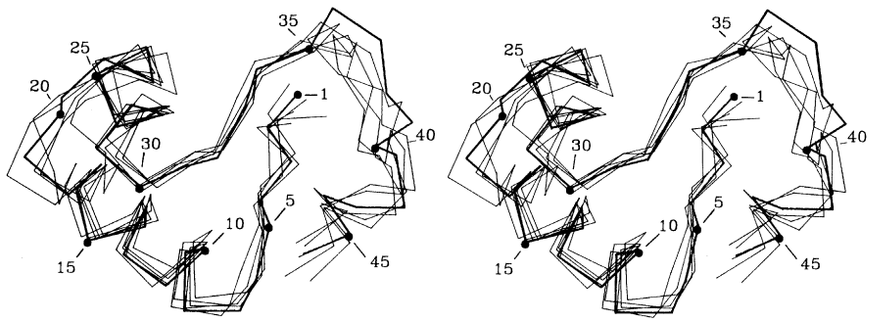

Fig. 1. Best fit of the superposition of the Cα atoms of NMR-restrained dynamics structures (thin lines) on the x-ray structure (thick lines).

NMR structure can be used in the molecular replacement method and converges to the refined crystal structure when standard model building and refinement techniques are employed.

Five NMR structures were taken from a previous study with simulated data (*9, 10*). The five structures S_1 to S_5 [RDII′, RDIB′, RDIC′, RDID′, RDIE′ in (*10*)] were obtained from a number of initial structures (that included a totally unfolded polypeptide chain) by restrained molecular dynamics that incorporated NMR interproton distance restraints and the three disulfide bridges present in crambin. The latter were necessary because the strong scattering of the sulfur atoms has a large influence on the Patterson function; in the convergence of NMR structures, the disulfide bridges do not play an important role. The five structures were obtained with different molecular dynamics protocols. To facilitate comparisons, structures S_2 to S_5 were superimposed on S_1 with the rigid body least-squares method (*16*); only Cα atoms were considered. The coordinates of the superimposed structures (S_1 to S_5) were averaged and the resulting structure was subjected to energy minimization to improve the stereochemistry and to relax bad nonbonded contacts (*10*); we call this structure $S_{\langle i \rangle}$. The 1.5 Å resolution crambin x-ray structure (*15*), which served as a control, was also superimposed on the NMR structure S_1; we call this structure $S_{x\text{-}ray}$. The ensemble of superimposed structures (S_1 to S_5 and $S_{x\text{-}ray}$) is shown in Fig. 1; the rms differences of S_1 to S_5 and $S_{\langle i \rangle}$ with respect to $S_{x\text{-}ray}$ are listed in Table 1. Each NMR structure S_i explores a somewhat different region in the neighborhood of the x-ray structure; each is closer to the x-ray structure than to any of the other structures (*9, 10*). The averaged structure is closer to the x-ray structure than any of the NMR structures (Table 1).

The rotation search is based on the comparison of Patterson functions

$$PAT(r) = 1/V \sum_{hkl} |F(hkl)|^2 \exp[-2\pi i(hx + ky + lz)] \quad (1)$$

from a model structure and a crystal structure, where $|F(hkl)|$ are the observed or calculated structure factor amplitudes at reciprocal lattice points *hkl*. The Patterson function for the model structure, PAT_1, and the observed reflection data, PAT_2, can be computed without phase information. The rotation function $RF(\Omega)$ that evaluates the agreement between PAT_1 and PAT_2 is defined by (*17*)

$$RF(\Omega) = \int PAT_1(\mathbf{r})\, PAT_2(\Omega\mathbf{r})\, dV \quad (2)$$

where Ω is an operator that rotates the coordinate system of PAT_2 with respect to PAT_1. Because $RF(\Omega)$ depends only on intramolecular distance vectors of PAT_1, it is possible to first determine the rotational orientation of the molecule by maximizing Eq. 2 as a function of Ω, and to then determine the relative translational position of the molecule with respect to the symmetry-related molecules in the unit cell.

The NMR structures were embedded in a cubic unit cell with *P*1 symmetry, 100 Å in size, so that the largest intramolecular distance vector (30 Å) is smaller than any intermolecular distance vector; the center of mass of the molecule was placed at the origin. As the crambin crystal symmetry is

Table 2. List of five largest peaks in rotation and translation search. The Eulerian angles $[(\theta_1, \theta_2, \theta_3) = RF]$ are listed for the rotation function where *RF* is the value of the rotation function scaled to the smallest (*RF* = 0.0) and largest (*RF* = 100.0) peak. The intermolecular distance vectors $[(x,z) = TF]$ are listed for the translation function, where *TF* is the value of the translation function scaled to the smallest (*TF* = 0.0) and largest (*TF* = 100.0) peak and *x*,*z* are in fractional units multiplied by 100. The mean and standard deviation σ are computed over all peaks. The correct peak is marked by an asterisk.

$RF(S_1)$	$RF(S_2)$	$RF(S_3)$	$RF(S_4)$
(129,70,268) = 100	(255,80,300) = 100	(128,80, 68) = 100	(320,70,184) = 100
(199,60,217) = 96	(74,80,299) = 89	(170,80, 65) = 99	(335,70,182) = 99
(23,60,212) = 95	(139,40, 52) = 89	(49,90, 7) = 95	(100,70,188) = 96
(208,60,207) = 92	(63,50,331) = 86	(228,80,288) = 94	(329,70,176) = 94
(19,20,322) = 91	(316,40, 55) = 86	(176,80, 71) = 94	(326,70,190) = 93
mean = 37, σ = 13	mean = 38, σ = 14	mean = 41, σ = 14	mean = 39, σ = 13
$RF(S_5)$	$RF(S_{\langle i \rangle})$	$\langle RF(S_i) \rangle$	$RF(S_{x\text{-}ray})$
(24,50,200) = 100	(306,80,201) = 100*	(105,90, 7) = 100	(306,80,201) = 100*
(301,90,203) = 99	(192,80,207) = 95	(35,50,188) = 100	(300,80,195) = 87
(115,70,186) = 98	(145,60,144) = 94	(243,70,141) = 99	(260,80,170) = 86
(300,80,195) = 97	(143,60,143) = 93	(306,80,201) = 98*	(270,80,285) = 83
(307,80,187) = 95	(97,40,254) = 93	(10,80,190) = 96	(260,50,298) = 83
mean = 42, σ = 14	mean = 44, σ = 14	mean = 46, σ = 14	mean = 36, σ = 14
$TF(S_1)$	$TF(S_2)$	$TF(S_3)$	$TF(S_4)$
(7, 79) = 100	(84, 13) = 100	(100, 79) = 100	(46, 57) = 100*
(10, 84) = 97	(86, 13) = 95	(0, 79) = 98	(94, 13) = 97
(7, 84) = 97	(5, 75) = 95	(2, 79) = 98	(98, 0) = 96
(10, 79) = 91	(84, 9) = 94	(0, 75) = 95	(98,100) = 95
(2, 31) = 91	(98, 26) = 93	(98, 75) = 94	(94, 18) = 93
mean = 52, σ = 18	mean = 49, σ = 18	mean = 43, σ = 16	mean = 50, σ = 19
$TF(S_5)$	$TF(S_{\langle i \rangle})$	$\langle TF(S_i) \rangle$	$TF(S_{x\text{-}ray})$
(46, 26) = 100	(46, 57) = 100*	(46, 57) = 100*	(46, 57) = 100*
(48, 26) = 99	(43, 57) = 96	(46, 62) = 99	(43, 62) = 99
(46, 31) = 93	(46, 62) = 92	(43, 62) = 94	(46, 62) = 98
(48, 22) = 91	(43, 62) = 89	(43, 57) = 92	(43, 57) = 97
(46, 57) = 89*	(48, 57) = 84	(48, 62) = 88	(41, 62) = 78
mean = 50, σ = 18	mean = 42, σ = 15	mean = 45, σ = 18	mean = 36, σ = 14

monoclinic $P2_1$, *b*-axis unique, the symmetry of the rotation function is $Pbn2_1$ (*18*). The rotation search can therefore be restricted to the asymmetric unit ($\theta_1 = 0 \ldots 360°$, $\theta_2 = 0 \ldots 90°$, $\theta_3 = 0 \ldots 360°$) in Eulerian angle space (θ_1, θ_2, θ_3). The symmetry-related molecule is then located at ($-\theta_1$, $\pi - \theta_2$, $\pi + \theta_3$). The search was carried out in [$\theta_+ = \theta_1 + \theta_3$, θ_2, $\theta_- = \theta_1 - \theta_3$] space, which ensures a homogenous angle density distribution (*19, 20*).

As a control and to establish the optimum parameters for the rotation function search we computed the rotation function of $S_{x\text{-ray}}$. The five largest peaks of the rotation function for $S_{x\text{-ray}}$ are listed in Table 2; a θ_2 section of the rotation function through the largest peak ($\theta_1 = 306°$, $\theta_2 = 80°$, $\theta_3 = 201°$) is shown in Fig. 2A. The largest peak corresponds to the correct orientation of the structure.

The parameters of the rotation search (number of reflections, resolution range, treatment of the Patterson origin, *B* factors of the model structure, and which atoms were included in the model structure) were varied systematically. Optimal results (shown in Table 2 and Fig. 2A) were obtained with the largest 30% of all measured crystal intensities from 10 to 5 Å resolution, the *B* factors of all atoms set to 8.0, and the original Patterson origin. Similar conditions have led to solutions of protein crystal structures by the molecular replacement method (*1–3*). Even when the x-ray structure was the model, the rotation function failed to produce the correct orientation unless all nonhydrogen atoms were included in the calculation of the model Patterson function. This contrasts with other successful applications of the method where backbone atoms and β-carbons sufficed (*1–3*). Use of only backbone atoms plus β-carbons had been attempted because these coordinates were predicted most accurately by the NMR-restrained dynamics (Table 1).

Even when optimum conditions were used in the $S_{x\text{-ray}}$ control calculation, the largest peak was only one standard deviation (σ) level above other erroneous peaks in the rotation function. This shows that crambin is a difficult system for the molecular replacement method, probably because of crystal packing and the high content of ordered solvent in the crystal lattice.

None of the NMR structures S_1 to S_5 produced the correct orientation ($\theta_1 = 306°$, $\theta_2 = 80°$, $\theta_3 = 201°$) in the rotation search (Table 2); indeed, the correct orientation (within 5° for all three angles) is not even among the five largest peaks. However, the regularized average structure $S_{\langle i \rangle}$ yielded the correct result (Table 2 and Fig. 2B). The correct peak of the rotation function for $S_{\langle i \rangle}$

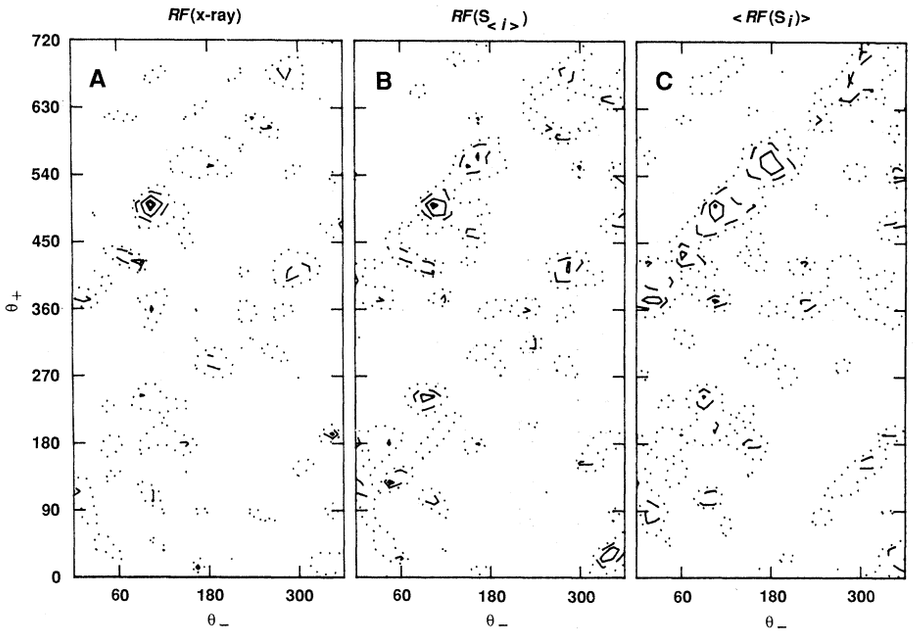

Fig. 2. Sections at $\theta_2 = 80°$ through the crambin rotation function; in (**A**) the x-ray structure and in (**B**) the averaged NMR structure $S_{\langle i \rangle}$ is used; in (**C**) the average of the rotation functions of the NMR structures S_i is plotted; contour levels are at σ, 2σ, 3σ, and 4σ.

is only 1/2 σ above the other peaks. To confirm the result with an alternative method, the rotation functions from the different NMR structures were averaged:

$$\langle RF(S_i) \rangle = 1/5 \sum_{i=1}^{5} RF(S_i) \qquad (3)$$

This procedure produced a peak at the correct orientation with different and equally large error peaks (Table 2); however, the error peaks are weighted differently. This is apparent in the θ_2 sections plotted for the two averaging possibilities in Fig. 2, B and C. The correct peak at $\theta_+ = 501°$, $\theta_- = 105°$ is somewhat broader in $\langle RF(S_i) \rangle$ than it is in $RF(S_{\langle i \rangle})$. The pattern in Fig. 2, B and C, is similar to the pattern observed in Fig. 2A for the x-ray structure, but the error peaks have smaller weights in case of the x-ray structure.

To increase the precision of the solution of the rotation function, a fine-grid search was carried out with $RF(S_{\langle i \rangle})$ around the correct peak ($\theta_1 = 306°$, $\theta_2 = 80°$, $\theta_3 = 201°$) in 1° steps for θ_2. This yielded a maximum at ($\theta_1 = 306.5°$, $\theta_2 = 78°$, $\theta_3 = 196.5°$) that corresponds to a 4° misorientation around the axis ($x = 0.75$ Å, $y = -0.29$ Å, $z = -0.59$ Å) in real space with regard to the correctly oriented x-ray structure. We took the rotation function search to be solved at this point; all struc-

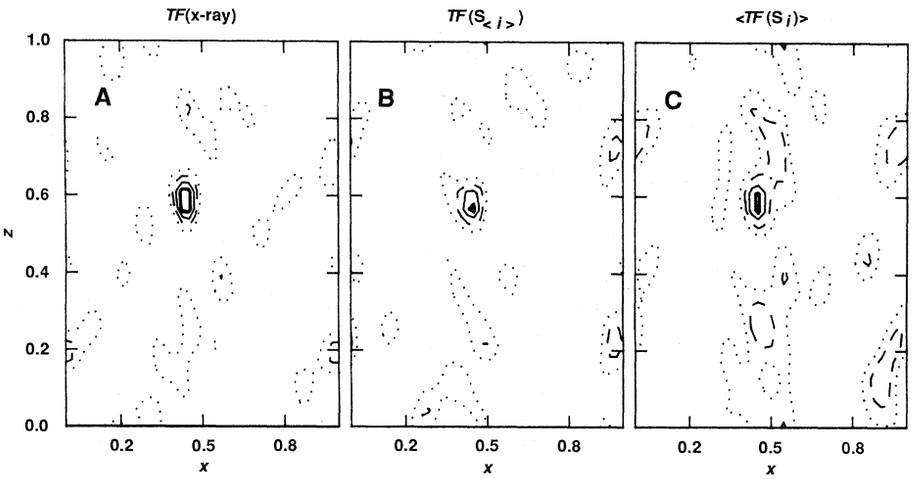

Fig. 3. Sections at $y = 1/2$ through the crambin translation function: in (**A**) the x-ray structure and in (**B**) the averaged NMR structure $S_{\langle i \rangle}$ is used; in (**C**) the average of the translation functions of the NMR structures S_i is plotted. All structures were oriented in the same manner as the optimal solution of the rotation function. Contour levels are at σ, 2σ, 3σ, and 4σ.

tures (S_i, $S_{\langle i \rangle}$, and $S_{x\text{-ray}}$) were rotated in this direction.

To determine the position of the molecule in the unit cell, we used the translation function of Crowther and Blow (*21*):

$$TF = \sum_{hkl} \{|F_{obs}(hkl)|^2 - [F_{calc}(hkl)^2 + F_{calc}(\bar{h}\bar{k}\bar{l})^2]\} \times F_{calc}(hkl)F^*_{calc}(\bar{h}\bar{k}\bar{l}) \exp[-2\pi i(hx + ky + lz)] \quad (4)$$

where $|F_{obs}(hkl)|$ are the observed structure factor amplitudes and $F_{calc}(hkl)$ and $F_{calc}(\bar{h}\bar{k}\bar{l})$ ($\bar{h} = -h$ and $\bar{l} = -l$) are the calculated structure factors of the model structure and its symmetry mate, and $F^*_{calc}(hkl)$ is the complex conjugate of $F_{calc}(hkl)$. The translation function TF describes the overlap of the intermolecular distance vectors between the two symmetry-related molecules in space group $P2_1$. This implies that the relative y-component is a constant at one-half the b axial length. Therefore, the search for a maximum of TF needed be done only in x,z space.

A control calculation with the x-ray structure was performed (Table 2 and Fig. 3A). We included all x-ray data between 10 and 4 Å resolution and computed the Fourier transform of the all-heavy-atom models to 4 Å resolution with the B factors set equal to 8.0; the grid size for the search was 1.0 Å.

The results of the translation search for the NMR structures (Table 2 and Fig. 3) closely resemble those for the rotation search; that is, the average NMR structure $TF(S_{\langle i \rangle})$, the average of the translation functions of the NMR structures $\langle TF(S_i) \rangle$ and the x-ray structure produced the same prominent and correct peak at ($x = 0.46$, $z = 0.57$) in fractional units; only S_4 and S_5 produce the correct peak within one σ below the maximum peak. Furthermore, the correct peak is broader; the weights for the error peaks are different for $\langle TF(S_i) \rangle$ and $TF(S_{\langle i \rangle})$.

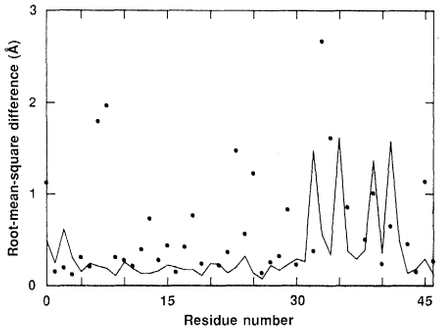

Fig. 4. Atomic rms difference from the x-ray structure for backbone (—) and side-chain (●) atoms for the structure after stage 7 of the refinement (see Table 3) as a function of residue number.

A fine-grid search of the translation function around the point $x = 0.46$, $y = 0.57$ with $TF(S_{\langle i \rangle})$ yielded an optimum at ($x = 0.448$, $z = 0.588$) in fractional units. All structures were then translated according to the $TF(S_{\langle i \rangle})$ solution, which corresponds to a displacement of ($x = -0.05$ Å, $z = 0.1$ Å) in real space with regard to the correctly positioned x-ray structure. This completed the determination of the position and orientation of the NMR structures.

The final stages of the solution of the protein crystal structure involved repeated cycles of refinement and model building to minimize the R factor

$$R = \sum_{hkl} ||F_{obs}| - |F_{calc}|| / \sum_{hkl} |F_{obs}| \quad (5)$$

The various stages of the refinement process are summarized in Table 3. The average NMR structure $S_{\langle i \rangle}$ was oriented and positioned according to the rotation and translation function results, which yielded an R factor of 0.45 at 4 Å resolution. In stage 1 only the orientation and position of the molecule were refined; the radius of convergence of this method is 5° in orientation and 1 Å in position (*22*). This rigid body refinement reduced the misorientation to about 1° and the displacement vector to ($x = 0.02$ Å, $z = -0.04$ Å). Stages 2 and 3 involved restrained least-squares refinement with data of successively higher resolution, which lowered the R factor to 0.38. There was no human intervention that could have biased the results.

Inspection of the ($2F_{obs} - F_{calc}$) map calculated after stage 3 showed that the region that involved residues 35 to 40 required the most rebuilding, because some atoms were not in electron density at 1σ and the electron density was somewhat fragmented. This is not surprising; comparison of the NMR structures with the x-ray structure in Fig. 1 reveals backbone deviations between 2 to 3 Å for residues 35 through 40. Most other regions of the molecule showed good density. Difference maps ($F_{obs} - F_{calc}$) were computed and the molecule was partially rebuilt. Three additional cycles of model building and refinement, during which the conformations of several side chains were modified, yielded an R factor of 0.27 at 2 Å resolution. Density now appeared that could be interpreted as ordered water.

The refinement was continued until further refinement at higher resolution would be straightforward. No attempt was made to include water molecules in the refinement, although the ordered solvent structure is an essential part of the crambin crystal (*23*). Note that the R factor of the published x-ray structure is 0.25 at 2 Å when all water molecules are omitted in the computation of the calculated structure factors.

The deviation of the refined $S_{\langle i \rangle}$ structure from the x-ray structure is 0.5 Å and 1.3 Å for backbone and side-chain atoms, respectively. The atomic rms differences as a function of residue number are shown in Fig. 4. The atoms that deviate by more than 2 Å are: Ile^7 Cγ2, Cδ; Val^8 Cγ2; Glu^{23} Oε2; Ile^{25} Cδ; Cys^{32} O; Ile^{33} Cγ1, Cδ; Ile^{34} Cδ; Ile^{35} Cβ, Cγ1, Cγ2, Cδ, O; Thr^{39} O; Pro^{41} O. For some of these (Ile^7 and Ile^{25}) conformational or compositional heterogeneity has been reported (*15*). Further, the temperature factors for residues 35 through 40 in the x-ray structure are larger than those of the rest of the molecule (*15*). We considered the refinement as completed for the purposes of the model study. It has been demonstrated that the average NMR structure can be refined to an R factor of 0.29 at 2 Å resolution without manual rebuilding by a new crystallographic refinement method that uses molecular dynamics (*24*).

Table 3. Stages of refinement and model building. The different stages can be described as follows. Stage 1 involved rotation and translation of the molecule as a rigid body. Stage 2 involved 20 cycles of the program ROTLSQ (obtained from W. Hendrickson; parameters were: $B = 8.0$; 10 to 4 Å resolution; *afsig* = 100; *bfsig* = 0). Stage 3 involved 6, 13, and 14 cycles of the program PROLSQ (*25*) (parameters were: $B = 8.0$; low resolution limit, 10 Å; *afsig* = 100; *bfsig* = 0) at resolutions of 4, 3, and 2 Å, respectively. Stage 4 involved rebuilding of the model with program FRODO (*26, 27*), followed by 10 cycles of program PROLSQ with $B = 8.0$ and 10 additional cycles with B factor refinement. Stages 5 through 7 involved model rebuilding with program FRODO followed by 10 cycles of B factor refinement with program PROLSQ. N represents the number of distances that deviate by more than 1σ. E_{int} is calculated from bond lengths, bond angles, dihedral angles, and improper torsional angle energies with the CHARMM program (*28*).

Stage	R factor	Resolution (Å)	rms difference (Å)		N	E_{int}
			Backbone	Side chain		
1	0.45	4	1.1	1.4	119	363
2	0.43	4	1.1	1.4	119	363
3	0.38	2	0.9	1.3	408	2244
4	0.32	2	0.8	1.4	351	1682
5	0.31	2	0.7	1.3	305	1573
6	0.30	2	0.5	1.2	251	1218
7	0.27	2	0.3	1.1	186	1058

REFERENCES AND NOTES

1. M. G. Rossman, Ed., *The Molecular Replacement Method* (International Science Review No. 13, Gordon & Breach, New York, 1972).
2. E. E. Lattman, *Methods Enzymol.* **115**, 55 (1985).
3. P. A. Machin, Ed., *Molecular Replacement*, Proceedings of the Daresbury Study Weekend (Science and Engineering Research Council, The Librarian, Daresbury Laboratory, Daresbury, United Kingdom), Daresbury, February 1985.
4. W. P. Aue, E. Bartholdi, R. R. Ernst, *J. Chem. Phys.* **64**, 2229 (1976).
5. J. Jeener, B. H. Meier, P. Bachman, R. R. Ernst, *ibid.* **71**, 4546 (1976).
6. A. D. Kline and K. Wüthrich, *J. Mol. Biol.* **183**, 503 (1985).
7. J. Zarbock, G. M. Clore, A. M. Gronenborn, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 7628 (1986).
8. R. Kaptein, E. R. P. Zuiderweg, R. M. Scheck, R. Boeleus, W. F. van Guusteren, *J. Mol. Biol.* **182**, 179 (1985).
9. A. T. Brünger, G. M. Clore, A. M. Gronenborn, M. Karplus, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 3801 (1986).
10. G. M. Clore, A. T. Brünger, A. M. Gronenborn, M. Karplus, *J. Mol. Biol.* **191**, 523 (1986).
11. T. F. Havel and K. Wüthrich, *ibid.* **182**, 281 (1985).
12. W. Braun and N. Go, *ibid.* **186**, 611 (1985).
13. A. D. Kline, W. Braun, K. Wüthrich, *ibid.* **189**, 377 (1986).
14. J. W. Pflugrath, G. Wiegand, R. Huber, *ibid.*, p. 383.
15. W. A. Hendrickson and M. M. Teeter, *Nature (London)* **290**, 107 (1981).
16. W. Kabsch, *Acta Crystallogr. Sect. A* **32**, 922 (1976).
17. M. G. Rossmann and D. M. Blow, *Acta Crystallogr.* **15**, 26 (1962).
18. S. N. Rao, J.-H. Jih, J. A. Hartsuck, *Acta Crystallogr. Sect. A* **36**, 878 (1980).
19. The angular grid parameter was set to 10° in the program RF described in E. E. Lattman, *Acta Crystallogr. Sect. B* **28**, 1065 (1972). Modifications are described in (*20*).
20. P. D. Martin, thesis, Wayne State University (1977); S. W. Mowbray, thesis, Massachusetts Institute of Technology (1983).
21. R. A. Crowther and D. M. Blow, *Acta Crystallogr.* **23**, 544 (1967). Program TF was obtained from E. E. Lattman; modifications are as described in (*20*).
22. R. Huber and M. Schneider, *J. Appl. Crystallogr.* **18**, 165 (1985).
23. M. M. Teeter, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 6014 (1984).
24. A. T. Brünger, J. Kuriyan, M. Karplus, *Science* **235**, 458 (1987).
25. W. A. Hendrickson, *Methods Enzymol.* **115**, 252 (1985).
26. T. A. Jones, in *Computational Crystallography*, D. Sayre, Ed. (Clarendon, Oxford, 1982), p. 303.
27. J. W. Pflugrath, M. A. Saper, F. A. Quiocho, in *Methods and Applications in Crystallographic Computing*, S. Hall and T. Ashida, Eds. (Clarendon, Oxford, 1984), p. 404.
28. B. R. Brooks *et al.*, *J. Comput. Chem.* **4**, 187 (1983).
29. We thank W. A. Hendrickson, R. Huber, J. Kuriyan, and D. Ringe for discussions. The work was partially supported by the National Science Foundation. Computations were done at the University of Minnesota Supercomputer Center with a grant from the Office of Scientific Computing of the NSF.

13 August 1986; accepted 29 December 1986

Isolation of an Olfactory cDNA: Similarity to Retinol-Binding Protein Suggests a Role in Olfaction

KYU-HO LEE, REBECCA G. WELLS, RANDALL R. REED

Molecular cloning techniques were used to isolate and characterize a protein possibly involved in the signal transducing system in olfactory tissue of the frog *Rana pipiens*. A complementary DNA library was constructed with messenger RNA obtained from frog olfactory neuroepithelium. A 700–base pair complementary DNA clone encoding a protein with a molecular weight of 20,300 was identified by differential hybridization analysis with polyadenylated RNA from olfactory epithelium and nonsensory respiratory epithelium. The messenger RNA corresponding to this clone was abundant in the cells of Bowman's glands in olfactory tissue but not in respiratory epithelium nor in several other tissues. The predicted sequence of this protein is homologous to members of a family of proteins that bind and transport small molecules in serum, suggesting that this protein may also bind and transport odorants in the mucus secreted by Bowman's glands.

THE PRIMARY EVENTS OF THE OLFACTORY response—odorant recognition and signal transduction—occur in the receptor neurons of the olfactory neuroepithelium. The apical processes of these bipolar neurons lie at the luminal surface of the tissue and are covered with long cilia (*1*). The mucosal layer contains two other cell types—supporting cells and neuroblast-like progenitor basal cells from which mature neurons continually differentiate. In the submucosa just beneath the basal lamina are the Bowman's glands, which secrete the mucus that bathes the olfactory cilia. This mucous layer is thought to play a major role in the solubilization and concentration of airborne odorants (*2*).

Transduction appears to be mediated by odorant-specific protein receptors in the ciliary membranes (*1*), and several investigators have identified specific odorant-binding activities in olfactory tissue extracts (*3–5*). However, purified molecular components of the transduction system have yet to be fully characterized and their activities correlated with actual signal transduction in olfactory neurons.

We have used molecular cloning techniques (*6, 7*) to isolate and characterize individual products specific for olfactory tissue. We constructed a complementary DNA (cDNA) library of approximately 4500 clones using polyadenylated [poly(A)$^+$] messenger RNA (mRNA) obtained from olfactory mucosa of the frog *Rana pipiens*. Approximately 10% of 2400 clones screened were olfactory-specific; that is, they showed strong hybridization to a probe derived from olfactory mRNA, and weak to negligible hybridization to an mRNA probe derived from respiratory epithelium. The six clones that showed the most intense olfactory-specific hybridization were studied further by restriction mapping and nucleotide sequencing. Analysis with various combinations of restriction enzymes indicated that five of the six clones had similar restriction maps.

Analysis of the nucleotide sequence of both strands of clone 5B5, the longest of the five similar clones, revealed a single long open reading frame putatively encoding a 182–amino acid protein with a predicted molecular size of 20.3 kD (Fig. 1). The sequence contained no consensus *N*-linked glycosylation signal sites but had a single 13–amino acid hydrophobic region at the amino terminus, similar to the signal sequence of many secreted proteins. Hybridization analysis of total RNA from various tissues with an M13 strand-specific probe containing the central Pst I–Hind III fragment from the 5B5 coding region indicated that the approximately 850-base message represented by this cloned cDNA was produced in large quantity in the frog olfactory neuroepithelium but not detectable in frog brain, liver, or respiratory tissue (Fig. 2A).

The predicted properties of the 20.3-kD secreted protein that was specific for olfactory tissue coincided with those of a soluble protein of approximately 20 kD observed in extracts of olfactory tissue (Fig. 2B). This protein, not present in extracts of respiratory tissue, was also seen when total RNA from olfactory epithelium from frog was analyzed in a rabbit reticulocyte lysate in vitro translation system (Fig. 2C).

To demonstrate that this protein corresponded to the product of the olfactory-specific mRNA we had examined, we attempted to specifically arrest the translation of this species in vitro (Fig. 2C). Addition to the translation reaction of SP6-transcribed RNA complementary to the message specifically blocked the translation of the 20-

Howard Hughes Medical Institute Laboratory of Genetics, Department of Molecular Biology and Genetics, Johns Hopkins University School of Medicine, Baltimore, MD 21205.