Normal Modes as Refinement Parameters for the F-Actin Model

Monique M. Tirion,* Daniel ben-Avraham,* Michael Lorenz,‡ and Kenneth C. Holmes‡
*Department of Physics, Clarkson University, Potsdam, New York 13699-5820 USA; ‡Max-Planck Institute for Medical Research, Jahnstrasse 29, 6900 Heidelberg, Germany

ABSTRACT The slow normal modes of G-actin were used as structural parameters to refine the F-actin model against 8-Å resolution x-ray fiber diffraction data. The slowest frequency normal modes of G-actin pertain to collective rearrangements of domains, motions that are characterized by correlation lengths on the order of the resolution of the fiber diffraction data. Using a small number of normal mode degrees of freedom (≤ 12) improved the fit to the data significantly. The refined model of F-actin shows that the nucleotide binding cleft has narrowed and that the DNase I binding loop has twisted to a lower radius, consistent with other refinement techniques and electron microscopy data. The methodology of a normal mode refinement is described, and the results, as applied to actin, are detailed.

INTRODUCTION

Many biological polymers, such as collagen, Rec A, tubulin, DNA, and F-actin do not form crystalline arrays suitable for structural analysis by x-ray crystallography. However, some biological polymers can be induced to form axially oriented, one-dimensional arrays appropriate for fiber diffraction. The resulting diffraction patterns consist of characteristic layer lines, the spacing of which is determined by the helical symmetry of the diffracting helical polymers. The polymers in a sample align axially, not azimuthally, and hence the diffraction patterns consist of rotationally averaged, overlapping diffraction maxima. The structural interpretation of these diffraction patterns is limited by the resolution of the data, the spacing of the layer lines, and the extent of the alignment of the sample. Typically, x-ray fiber diffraction data permit the determination of anywhere from tens to hundreds of structural parameters (Makowski, 1991). Clearly this is not sufficient to refine the Cartesian coordinates of every atom in the polymers, as in x-ray crystallography. The challenge, therefore, in refining the structure of biopolymers lies in defining a set of structural refinement parameters suitable for fiber diffraction data.

In some instances, the crystalline structure of the monomers comprising a helical polymer is known, as for example in the cases of tobacco mosaic virus, Rec A, and F-actin. In these instances it becomes possible to construct an initial model of the polymer helix, as was done in the case of F-actin (Holmes et al., 1990). Further refinement of this initial model, reflecting structural modifications of the monomer as it is incorporated into the filament, requires a choice of structural parameters. In the case of F-actin, several have been tested. Initially, the monomer was divided into four separate regions, corresponding to the four subdomains of G-actin, and each region was refined independently as rigid bodies (for a total of $4 \times 6 - 2 = 22$ degrees of freedom). This approach lowered the R factor, the fit of the computed diffraction pattern to the observed diffraction data. However, refinement of domains as independent rigid bodies could not maintain proper interdomain contacts, or stereochemistry, a necessary prerequisite for an atomic model.

Lorenz has refined the F-actin model by a method termed directed mutation (Lorenz et al., 1993). In this instance many more degrees of freedom are permitted and a very good fit of the computed diffraction pattern to the data is obtained (R factor; 0.07). This approach is based on rigid body refinements of arbitrarily defined domains. Proper stereochemistry is maintained by repeated intervention of energy minimization during refinement. A total of 375 domains may be defined (corresponding to the number of residues in actin), each with six rigid body degrees of freedom, allowing the algorithm to sample a total of 2250 structural parameters. However, an examination of the structure obtained with the directed mutation algorithm shows that bond lengths and bond angles changed very little during refinement; hence in practice as few as 750 main chain dihedral angles were sampled. Although repeated minimization runs with different starting seeds converged to very similar final structures (Lorenz et al., 1994), this technique provides no formal assurance against overfitting the fiber diffraction data with too many structural parameters.

As an alternative to rigid body refinements, we sought a refinement algorithm that uses a minimum number of adjustable parameters while maintaining a proper stereochemistry as a natural consequence of the choice of generalized coordinates. As the actin filament can be assembled from G-actin monomers without the hydrolysis of ATP or presence of enzymes, we expect structural modifications within G-actin that are attainable without overcoming large energy barriers, structural modifications that can be achieved by thermal activation. The softest degrees of freedom in a protein that permit local as well as collective rearrangements of atoms are those that pertain to dihedral modifications. The energy required to adjust a dihedral angle is typically two orders of magnitude smaller than adjusting a bond length or

Received for publication 18 May 1994 and in final form 14 October 1994.
Address reprint requests to Dr. Monique M. Tirion, Department of Physics, Clarkson University, Box 5820, Potsdam, NY 13699-5820. Tel.: 315-268-2375; Fax: 315-268-6670; E-mail: tirion@craft.camp.clarkson.edu.

0006-3495/95/01/05/08 $2.00

angle. Hence we sought a refinement algorithm that uses only dihedral angles as a subset of the total degrees of freedom accessible to a protein (a subset of 1368 ψ, ϕ, and ξ torsional degrees of freedom out of >10,600 Cartesian degrees of freedom in G-actin).

Direct refinement with all dihedral degrees of freedom to optimize the fit to fiber diffraction data has two drawbacks, however. First, such an approach permits structural adjustments beyond the resolution of the data (one could overfit the data); and second, it ignores nonbonded, stereochemical constraints. It was for these reasons that we chose to pursue a novel approach to fiber refinement, the use of normal modes of the monomer as refinement parameters for the filament structure.

The vibrational motions of a protein are most easily described by its normal modes, a set of vibrational frequencies and their associated atomic motion for small oscillations about an equilibrium, or preferred, minimum energy configuration. The modes are normal in the sense that they are mutually orthogonal; the motion described by one mode cannot be achieved by any other superposition of modes, each of which describes a unique frequency. The overall motion of a protein for small oscillations about an equilibrium configuration is described by a superposition of all computed modes. The slowest frequency modes describe collective motions of domains, such as scissor-type opening/closing and twisting of domains, whereas the faster modes pertain to small, rapid oscillations of sidechains and small groups of atoms about their minimum energy configurations (Levitt et al., 1985; Gō et al., 1983; and Brooks Karplus, 1983). The correlation length of the motion, in other words, decreases as the frequency increases. This fact provides a natural means to fit the fiber data with a sensible number of structural parameters. The motion described by modes beyond a certain frequency will pertain to structural modifications smaller than the resolution of the data and will be ignored by a normal mode refinement algorithm. As there is always a risk with other methods to overfit the fiber data, this is one obvious advantage to using normal modes as refinement parameters.

Finally, because normal modes describe thermally activated motions, i.e., motions attainable without encountering large energy barriers due to stearic clashes, we reasoned they might describe the type of structural modifications that the G-actin monomer utilizes as it is incorporated into the filament. For these reasons we used the slowest modes as structural parameters in a least-squares minimization algorithm to refine the F-actin structure. In the next sections we describe the details of the method and the resulting refined model of the actin filament.

MATERIALS AND METHODS

Preparation of actin

G-actin was prepared as described in Mannherz et al. (1977). G-actin was polymerized into F-actin by adding 100 mM KCl at pH 7.4. Phalloidin was added to G-actin in slight excess. A solution of 2–6 mg/ml F-actin was concentrated in an Amicon cell and drawn into x-ray capillaries of 0.5 mm diameter as described in Popp et al. (1987).

Recording and processing of data

X-ray diffraction data were taken with an Elliott rotating anode GX 18 operating at 35 kV and 50 mA with double mirror focusing optics and a cylindrical camera. X-ray diagrams were recorded on a four-film stack (CEA REFLEX 25) to cover the wide range of data (Holmes et al., 1990). The films were scanned with an Optronics drum scanner at a spatial resolution of 50 μm. The data were then corrected for polarization and mapped into reciprocal spaced coordinates with the program PROFIDA (Lorenz and Holmes, 1993). Pixel resolution in reciprocal space is 0.0025 nm^{-1}/pixel. Data sets are stored in two-dimensional arrays of size 256 × 256 in 2-byte values.

Normal mode refinement algorithm

The normal modes of the ternary system, G-actin:ADP:Ca^{2+}, were obtained as described previously (Tirion and ben-Avraham, 1993). The analysis utilized all main chain ϕ (except Pro) and ψ and all side chain ξ torsional angles, as well as all dihedrals of the ADP nucleotide, for a total of 1384 degrees of freedom. The energy function, L79, developed by Levitt (1983), uses standard van der Waals terms as well as directional hydrogen bonds. The analysis yields 1384 eigenfrequencies and their associated eigenmodes, extending from periods of 17 to 0.01 ps.

We used a nonlinear least-squares algorithm to determine the magnitude of the contribution of each structural parameter so as to minimize ξ^2 (see, for example, Press et al., 1990)

$$\xi^2 = \sum_{x,y} \frac{1}{W^2(x,y)} |I_{calc}(x,y) - I_{obs}(x,y)|^2. \quad (1)$$

Here the $I_{calc}(x,y)$ and $I_{obs}(x,y)$ refer to the calculated and observed diffraction intensities at pixel location (x,y), respectively. W is a weighting factor equal to $I_{obs}(x,y) + \alpha$, where α is chosen to suppress low signal noise fluctuations. (For a derivation of the Fourier transform structure factors of a helix, see Klug et al., 1958. For a derivation of the azimuthal averaging and convolutions required to obtain the computed diffraction intensities, $I_{calc}(x,y)$ from the structure factors, see Holmes and Barrington-Leigh, 1974). Eq. 1 implies that the observed and computed intensities, I_{calc} and I_{obs}, have been suitably scaled so that the total observed and computed intensities match.

The least-squares minimization algorithm requires the evaluation of the second derivative matrix, or Hessian, of the ξ^2 error function with respect to the structural parameters. This was done numerically:

$$B_{kl} = \sum_{x,y} \frac{1}{W^2(x,y)} \left[\frac{I^k_{calc}(x,y) - I_{calc}(x,y)}{\xi_k}\right] \times \left[\frac{I^l_{calc}(x,y) - I_{calc}(x,y)}{\xi_l}\right] \quad (2)$$

(It can be shown that the second derivative terms may be ignored in the sum, reducing the Hessian to a product of first derivatives; see, for example, Press et al., 1990.) I^i_{calc} refers to the computed intensity of the current model after a trial shift by a small increment, ξ_i, of the structural parameter, q_i. The q_i consist of n normal modes, as well as three rigid body rotational and one radial degrees of freedom for the monomer, for a total of $N = n + 4$ generalized coordinates. The magnitudes of ξ_i were tested and set to 0.05° for dihedral updates (arbitrarily about the largest main chain component of the nth eigenvector), to 0.5 Å for translations, and to 0.5° for rotations of the rigid body degrees of freedom.

The set of shifts, Δq_i, that minimize ξ^2 is given by the solutions of the set of equations:

$$\sum_i B_{ki} \Delta q_i = A_k, \quad (3)$$

where A_k is the partial derivative of ξ^2 with respect to q_k: $A_k = -\frac{1}{2}\xi^2/\partial q_k$, or

$$A_k = \sum_{x,y} \frac{1}{W^2(x,y)} [I_{obs}(x,y) - I_{calc}(x,y)] \times \left[\frac{I^k_{calc}(x,y) - I_{calc}(x,y)}{\xi_k}\right] \quad (4)$$

This computation is repeated iteratively until the shifts, Δq_i, converge and produce no change in the error function, ξ^2.

Stereochemical constraints

The solution of Eq. 3 determines the optimal contributions, Δq_i, required to minimize the difference between I_{calc} and I_{obs}. There is, in other words, no restriction placed on the magnitude, Δq_i, of the contribution of each degree of freedom to the refined model. An important test for the success of a normal mode refinement will be to confirm that only small updates in the normal mode degrees of freedom are effected; violations will appear as unfavorable nonbonded interactions. We therefore analyzed the refined coordinates of each run for stereochemical clashes with XPLOR (Brünger, 1990). Any unfavorable nonbonded interactions were relieved by a Powell energy minimization (see, for example, Press et al., 1990). The resulting root mean square, rms, deviations of the refined coordinates from the subsequent energy-minimized coordinates, σ^{E}_{rms}, were computed by the algorithm of Kabsch (1976).

Fiber diffraction residual

In accordance with previous work (Holmes et al., 1990; Lorenz et al., 1993) we define a fiber diffraction residual that is closely related to the ξ^2 error function and also resembles the familiar crystallographic residual

$$R_{cryst} = \frac{\sum_{x,y} ||F_{obs}(x, y)| - |F_{calc}(x, y)||}{\sum_{x,y} |F_{obs}(x, y)|}.$$

The sum extends over all data points, (x, y), and F_{obs} and F_{calc} refer to the observed and calculated diffraction structure factors, respectively. In fiber diffraction, however, $|F|$ is generally not equal to $I^{1/2}$ due to the azimuthal averaging of the data. Hence we define the fiber diffraction residual, R_F, as

$$R_F \equiv \frac{\sum_{x,y} |I_{obs}(x, y) - I_{calc}(x, y)|^2}{\sum_{x,y} I_{obs}(x, y)^2}. \quad (5)$$

Both definitions share the same minimum; at $F_{calc} = F_{obs}$ both residuals are zero. However, the magnitude of the residuals as well as the steepness of the functions about the minimum will vary slightly; the same atomic configuration will not assure that $R_{cryst} = R_F$. Therefore, only the efficiency of the minimization is affected by the different choices of definition of residual.

RESULTS

R_F factor

First we tested how increasing the number of normal mode degrees of freedom, n, reduced R_F. The residual decreased smoothly as n increased (Table 1). The initial R_F factor, obtained with four rigid body radial and rotational degrees of freedom, was 19.3%, like the original Holmes model (Holmes et al., 1990). (We included the phalloidin coordinates by using a separate rigid body refinement for this small heptapeptide. Phalloidin refined to the same location as published by Lorenz et al. (1993).) Including the slowest vibrational mode in the refinement ($n = 1$), for a total of five adjustable structural parameters, decreased the R_F factor to 17.0%. The 12 slowest modes (16 degrees of freedom) lowered the R_F factor to 13.0%.

Increasing the number of modes in the refinement beyond 12 did not further reduce the R_F factor. This is not surprising given the limited resolution of the fiber diffraction data. The motion described by the nth eigenvector (with eigenfrequency ω_n) is characterized by a typical coherence length that decreases as the mode number increases. Apparently, the coherence length of the motion associated with eigenfrequencies faster than ω_{12} is smaller than the resolution of the data. It is interesting to note that 72% of the total thermal motion of the atoms at 300 K can be accounted for by the 12 slowest modes (Tirion and ben-Avraham, 1993).

Stereochemistry

We checked the stereochemistries of the refined structures (for values of n from 1 to 12) for unfavorable nonbonded interactions. The number of intramonomer clashes were minor, in each case fewer than 40 of a total of ~120,000 nonbonded interactions, which were readily relieved by energy minimization. Also the intermonomer interactions between helically interacting monomers were analyzed for unfavorable interactions (by using the noncrystallographic symmetry option in XPLOR (Brünger, 1990)). In all cases, any minor clashes that resulted from the refinement were relieved by energy minimization. The rms deviations, σ, of the energy-minimized models from the refined but unminimized coordinates are reported in Table 1 and are seen to be quite small. The R_F factors of the energy-minimized coordinates were also recomputed and were not noticeably affected by the energy minimization (Table 1).

The lack of serious interatomic clashes as a result of the least-squares minimization is significant. By using normal modes as degrees of freedom, the refinement proceeds, for small updates, along stereochemically allowable paths. However, larger updates, by exceeding the harmonic limit, necessarily result in enormously unfavorable overlaps that result in anomalous diffraction maxima disallowed by the data. The faster modes produce structural changes beyond the resolution of the data and therefore cannot be used to overcome

TABLE 1 R_F factors and rms deviations of refined models

	n								
	0	1	2	3	4	6	8	10	12
R_F	0.193	0.170	0.160	0.158	0.156	0.149	0.144	0.144	0.136
σ (Å)		0.42	0.43	0.41	0.41	0.41	0.50	0.56	0.44
R_E		0.165	0.166	0.164	0.161	0.147	0.141	0.130	0.130
σ_{SD} (Å)		1.24	1.29	1.29	1.28	1.25	1.35	1.93	1.49
σ_{LD} (Å)		1.49	1.22	1.19	1.23	1.26	1.30	1.13	1.11

n, number of normal mode degrees of freedom included in the refinement; R_F, fiber diffraction residuals as defined in Eq. 2.5; R_E, residuals obtained after the refined coordinates are energy minimized. σ records the rms deviation between the refined and subsequent energy-minimized models. σ_{SD} and σ_{LD} give the rms deviations of the refined, energy-minimized coordinates from the crystal coordinates by superposing only the small domains (SD) or large domains (LD), respectively.

nonphysical geometries. Hence this technique is self-regulating in the sense that the data determine to which modes it is sensitive, within the resolution limit.

Next, we determined whether the models obtained with the various values of n refined to similar three-dimensional configurations. We therefore computed the rms deviations of the main chain atoms of the models from each other. As seen in Table 2, the refined models agree well. The rms deviations between the models never exceed 2.4 Å, and the maximal deviation of any one main chain atom never exceeds 2.9 Å. Indeed, the locations of maximal differences between the models are confined to three places: the COOH-terminal arm (residues 364–372), the DNase I binding loop (residues 42–45), and residues 230–247 in subdomain 4. These regions have the largest computed and observed B factors and are apparently not uniquely oriented within the resolution of the data by the refinements.

Contribution of the slow modes

To understand the structural modifications required to improve the fit of the crystal G-actin coordinates to the F-actin data, we superposed the large domain (LD) of each refined model onto the equivalent region of the crystal coordinates. Similarly, we superposed the small domain (SD) of each model onto the equivalent residues in the crystal structure. The resulting rms deviations of the corresponding superpositions are presented in Table 1, under σ_{rms}(LD) and σ_{rms}(SD). As can be seen from these values, each domain superposes onto the crystal coordinates very well; in no instance is the net rms deviation larger than 1.5 Å. This means that the intradomain structure of G-actin has changed very little upon incorporation into the filament, whereas the interdomain and, to some extent, the intersubdomain orientations have shifted relative to each other.

The structure of one actin monomer resulting from refinement with 12 normal modes plus 4 rigid body degrees of freedom is shown in Fig. 1 and the resulting computed and observed diffraction pattern in Fig. 2. The top panel of Fig. 1 shows a C_α tracing of the refined and crystal coordinates of the actin monomer such that the large domains are optimally superposed (σ_{rms}(LD) = 1.11 Å, see Table 1). The C_α atoms in the large domain superpose very well in the two structures, except for residues 262–274 and 230–250. Residues 262–274 form a loop between subdomains 3 and 4, interrupting an otherwise long α-helix straddling these two subdomains. This loop, implicated in stabilizing the F-actin structure (Holmes et al., 1990; Chen et al., 1993), has shifted laterally away from the nucleotide binding cleft (Ile C_α 267 has moved 7.5 Å). It is interesting to note that with only the slowest modes available, the position of this loop can still adjust independently of the rest of the large domain.

Also residues 230–250 in subdomain 4 have shifted up, in the orientation of Fig. 1, in a manner so as to close the nucleotide binding cleft. Meanwhile, the small domain has shifted to further close the nucleotide binding cleft. The small domain has also twisted, propeller-style, so that subdomain 2 has swung a bit forward.

Similarly, the bottom panel of Fig. 1 shows a C_α tracing of the refined and crystal coordinates of the actin monomer so that the small domains are optimally superposed (σ_{rms}(SD) = 1.49 Å, see Table 1). All of subdomain 1 and most of subdomain 2, except for residues 36–52, superpose very well. Residues 36–52 includes the DNase I binding loop, which has narrowed, twisting sideways and up. Indeed, this loop now seems poised to form two β-strands, with Gly-48 situated at a turn, as if to make a β-bend. The COOH-terminal arm does not superpose as well as the rest of the small domain, having shifted further back than the rest of subdomain 1. This indicates also that the COOH-terminal region is quite flexible and not as tightly constrained as the rest of the small domain. It is clear from Fig. 1*B* how the domains have shifted so as to close the ADP binding cleft. Residues 223–233 in subdomain 4 shifted up to a higher axial location and in to a lower radial location, also resulting in the narrowed nucleotide binding cleft.

Filament structure

The filament model obtained with 12 normal modes degrees of freedom, plus phalloidin, is shown in Fig. 3. The number of intermonomer interactions have increased relative to the original model of Holmes et al. (1990), suggesting that the structure is in a more stable configuration. (Note that no

TABLE 2 The rms and maximal deviations, in Ångstroms, between models refined by using different numbers of normal mode degrees of freedom

	n							
m	1	2	3	4	6	8	10	12
1		1.06	1.10	1.09	1.29	1.56	2.20	2.34
2	0.90		0.31	0.42	0.65	1.14	2.15	2.13
3	0.96	0.21		0.25	0.69	1.12	2.08	2.06
4	0.99	0.32	0.16		0.65	1.08	2.08	2.07
6	2.00	1.18	1.18	1.08		0.78	1.87	1.84
8	1.89	1.32	1.32	1.24	0.37		1.69	1.55
10	1.50	1.23	1.22	1.20	1.49	1.36		1.01
12	2.77	2.64	2.86	2.88	2.57	2.45	1.91	

The upper right triangle records the rms deviations between the refined, energy-minimized models obtained by using n and m normal mode degrees of freedom. The rms deviations are computed for all main chain atoms. The lower left triangle records the corresponding maximum deviations contributing to these rms deviations.

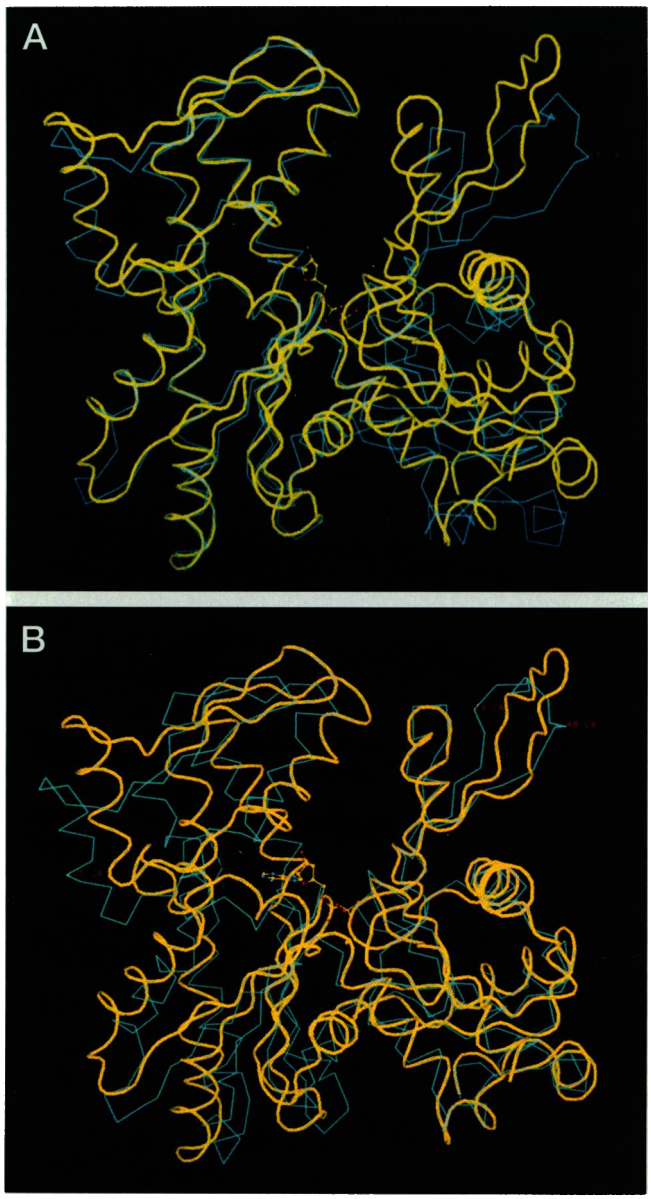

FIGURE 1 Overlays of the refined C_α coordinates (yellow) and the crystal, G-actin coordinates (Kabsch et al., 1990). In the top panel the main chain atoms of the large domains of the two actin structures have been optimally superposed, whereas in the bottom panel only the small domain main chain atoms have been superposed. The closing of the nucleotide binding cleft as a result of the refinement is clearly evident. The monomers are situated so that the large domain is on the left and the large subdomains (1 and 3) are at the bottom, in conformity with the convention established by Kabsch and co-workers (1990).

loops have been rebuilt to carry out this analysis, not the 262–274 nor the 40–50 DNase I binding loops. Rebuilding either loop manually, as was done for the 262–274 "plug" by Holmes and co-workers (1990), would certainly stabilize the diagonal one-start helix contacts further.) Indeed, a normal mode analysis of the F-actin filament confirms that the current model leads to a significantly more rigid structure, as compared with the Holmes or Lorenz models (ben-Avraham and Tirion, in preparation).

The most striking features of the refined actin filament, besides the closing of the nucleotide-binding cleft, include the realignment of the DNase I binding loop. This loop has narrowed, twisting sideways and up so as to bring residues on the outside to a lower radius. In this model, the loop refined to a location directly under the cleft between subdomains 1 and 3, near residues 146–148 leading from the small to the large domain, and residues 168–170 in the loop connecting the last 2 strands of the β-sheet in subdomain 4. Furthermore, subdomain 3 and residues 223–250 in subdomain 4 moved back, radially, filling in the axis cavity somewhat. Residues 243–245 are situated so as to fit into the gap between 323–325 and 287 in subdomain 3 above it.

The model is consistent with stereochemical data. Elzinga and Phelan (1984) cross-linked Lys-191 with Cys-374 and were able to show the side chains of these two residues were no more than 12–14 Å apart. In this model the C_α atoms of these residues are 20.4 Å apart, consistent with this finding. Similarly Hegyi et al. (1992) cross-linked Gln-41 and Lys-113 of two adjacent monomers. They concluded that the distance between the two C_α carbons is not greater than 22.3 Å. This model places them 23.8 Å apart. Finally, the model was superposed on the electron microscopy (EM) data of Milligan et al. (1990) and showed good agreement. This is especially true for loop 38–52 in subdomain 2, which in the original model is at a higher radius than the EM data allow. The normal mode refinement has brought this loop to a lower radius, so that it no longer extends out of the EM data.

Comparison with Lorenz model

The Lorenz model of F-actin (Lorenz et al., 1993) achieved a much lower R_F factor, as the normal mode refinement makes a more restricted, coarse-grained search of conformational space than does the directed mutation algorithm. Excluding the DNase I binding loop and the hydrophobic plug between subdomains 3 and 4, which were rebuilt manually by Lorenz et al. (1993), the rms fit of the main chain atoms of the two models is 3.4 Å, with a maximal deviation of 8.2 Å at residue 350 in the COOH-terminal arm. However, many features of the two models are preserved, including a closing of the nucleotide binding cleft. The most striking difference between the models is in the orientation of the DNase I binding loop. The normal mode refinement has brought this loop to a higher axial and lower radial location, whereas the directed mutation algorithm brought this loop to a lower axial and lower radial location. The location of this loop seems to be critical to the dynamical spectrum of F-actin and may indeed fluctuate between various configurations, depending on the pH, ion concentration, and other factors (Orlova and Egelman, 1993).

DISCUSSION

Normal modes as refinement parameters

Normal modes provide a small, sensible set of basis vectors with which to refine atomic models of biopolymers against fiber diffraction data. As the fiber diffraction data are of limited resolution, it is important to choose a set of refinement parameters that will not overfit the data. The motion

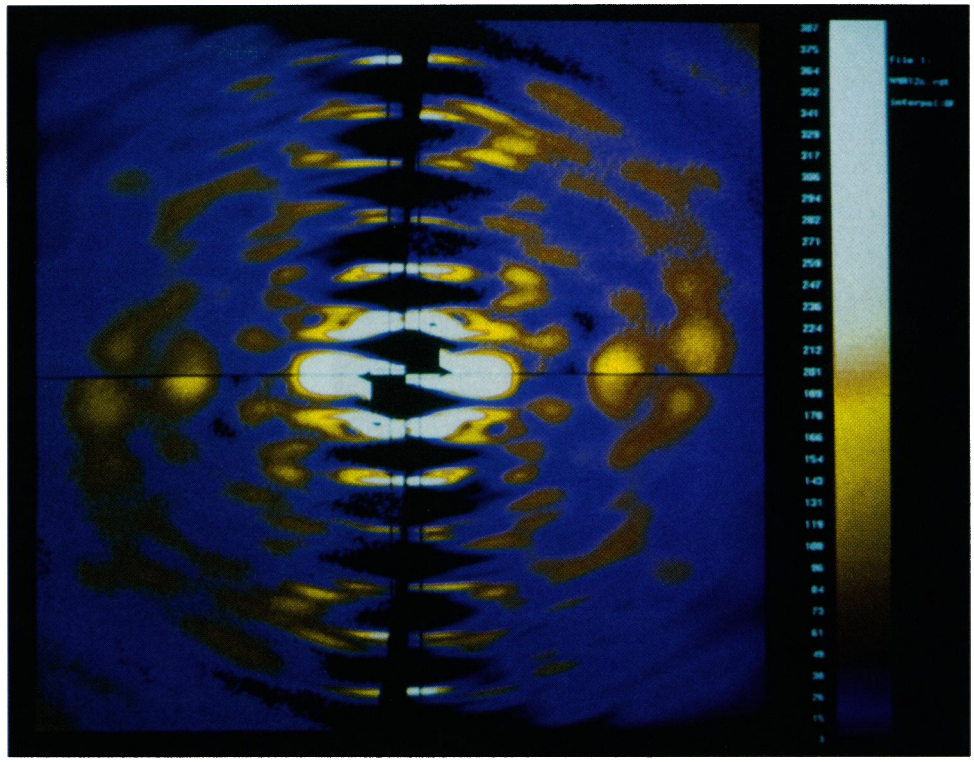

FIGURE 2 The x-ray fiber diffraction data shown in the upper right and lower left quadrants, compared with the computed diffraction pattern in the upper left and lower right quadrants for the actin model obtained with 12 normal mode degrees of freedom.

described by each normal mode has a characteristic correlation length associated with it; the slower frequency modes pertain to motions that involve collective rearrangements of large domains and regions, whereas the faster modes pertain to small oscillations of side chains and groups of atoms about their equilibrium configurations. Beyond a certain frequency, in other words, the normal modes describe conformational changes that cannot be detected by limited resolution data. Indeed, our data indicate that using more than the 12 slowest normal modes of the actin monomer does not improve the R_F factor. Using normal modes as refinement parameters, therefore, assures that the data is not overfit with too many structural parameters.

Furthermore, given the limitation of the resolution, it is prudent to rely on constraints imposed by proper stereochemistry to restrict the total number of configurations the monomer can access in the search algorithm. Configurational searches that ignore the stereochemical constraints imposed by excluded volume, proper patterns of hydrogen bonding in β-sheets and α-helices, and van der Waals interactions, for example, are not efficient in this sense. The computed normal modes identify those motions that are readily accessible to the monomer, motions that can be attained without the need to overcome large energy barriers due to steric clashes and breakage of favorable nonbonded interactions. A normal mode refinement, therefore, attempts to model the structural transitions of the monomer as it is incorporated into the filament, by using these flexibilities as degrees of freedom.

It should be borne in mind, however, that the modes are not pathways. Their use as refinement parameters are limited to small deformations. We did not achieve as low an R_F factor as Lorenz et al. (1993), and clearly a number of features of the diffraction data are not reproduced by this analysis. This suggests that not all structural adjustments in the G-actin to F-actin transition are modeled by the slow normal modes of the monomer. One option may be to recompute the normal modes of the refined coordinates. However, an NMA requires an initial energy minimization that, in the case of F-actin, is not practical due to the large size of the system. Furthermore, refining a monomer outside the filament merely brings the coordinates back to the crystal coordinates. Additional adjustments, such as manual rebuilds of particular loops and regions, as done by Lorenz and co-workers (1993), are necessary to further improve the fit of the model to the data.

Modeling the G- to F-actin transition

The normal mode refinement of the F-actin model confirms certain structural features of the G- to F-actin transition and highlights aspects that remain to be resolved. It should be stressed at this point that the Lorenz refinement established, and this independent approach confirms, that a solution to the x-ray fiber diffraction data of F-actin exists, a solution that is structurally similar to the crystal G-actin configuration. No dramatic interdomain shifts or major rebuilds of large regions need to be invoked to successfully reproduce the x-ray data. We find a small intra domain shift between the large and small domains and a few mobile surface loops reoriented by the refinements.

The closure of the nucleotide binding cleft is seen in every refinement undertaken to date: in initial rigid body refine-

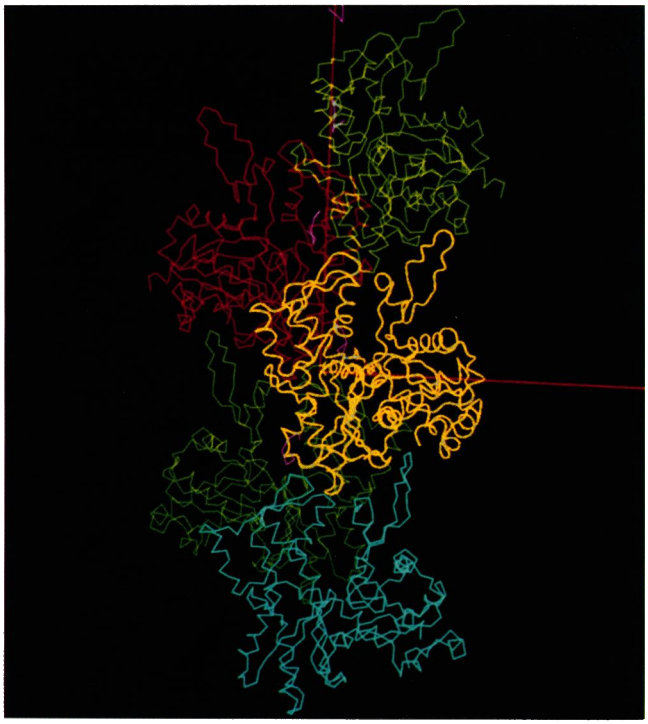

FIGURE 3 F-actin model resulting from the refinement with 12 normal mode degrees of freedom. Shown are five helically related monomers, plus phalloidin, in pink. The fiber axis is vertical. Subdomain 2 extends towards the protomer above it, interacting with the cleft between subdomains 1 and 3.

ments of the four subdomains separately (Holmes et al., 1993), in the directed mutation algorithm (Lorenz et al., 1993), as well as in this independent approach to fiber refinement. Each refinement confirms that subdomain 2 comes to a lower radius and subdomain 4 undergoes shifts whereby the nucleotide binding cleft between these domains narrows. Experimental work by Miki and Kouyama (1994) with energy transfer measurements of fluorescently labeled probes, also confirms this aspect of the F-actin structure. Presumably the closure of the cleft pertains to the F-actin:ADP monomer configuration, as the refinements were based on data obtained from long-lived muscle filaments. This closure will certainly affect the nature of the protein to nucleotide and cation interaction; perhaps this feature of the actin assembly process triggers the hydrolysis of the bound ATP.

The normal mode refinement reinforces another observation made about the G- to F-actin transition: besides the closing of the nucleotide binding cleft no dramatic structural changes are required to obtain a very good fit to the x-ray data. Indeed, the refinements indicate that subdomains 1 and 3, except for the COOH terminus, are consistently resolved by the data, and undergo no obvious structural modifications within the filament. Only a few surface regions and loops appear to undergo shifts, including the 38–52 DNase I binding loop in subdomain 2, the COOH-terminal arm (359–375), and the 262–274 hydrophobic plug between subdomains 3 and 4.

The DNase I binding loop in subdomain 2 is not uniquely oriented by the refinements. In the Lorenz model this loop is situated on the surface of subdomain 2, where it is able to make close longitudinal contacts with subdomain 3 of the neighboring (i + 2) monomer. In the current refinement, however, this loop has twisted up, so that it is positioned to interact not just with subdomain 3 but also with subdomain 1 of unit i + 2 in the long-pitch helix. Both orientations are plausible; there are no obvious stereochemical reasons why either orientation is preferred. Cross-linking experiments between Lys-113 and Gln-41 indicate these residues should be no further apart than 22.3 A (Hegyi et al., 1992), consistent with the current model. Furthermore, EM reconstructions of F-actin bound with the *Limulus* acrosomal process protein, scruin, show a bridge of intrastrand density between subdomains 2 and 1 (Owen and DeRosier, 1993), a finding consistent with this refinement. However, recent experiments also indicate that subdomain 2 may assume different configurations in response to a number of experimental conditions such as the identity of the bound nucleotide and cation (Orlova and Egelman, 1992, 1993) and suggest that this region may be quite flexible (Orlova and Egelman, 1993). It remains to be resolved what role this loop plays in filament integrity and flexibility, in nucleotide exchange, and in other possible functions.

Also the COOH terminus is shifted by the refinements, consistent with findings that Cys-374 undergoes large structural adjustments upon incorporation into the filament (Orlova and Egelman, 1992) and with separate findings that this residue is quite mobile in the submillisecond time regime (Thomas et al., 1979; dos Remedios et al., 1987). Our refinements suggest this arm shifts away from the cleft between subdomains 1 and 3, where it might interact with monomer i + 2 in the long-pitch helix, toward the 223–230 helix in subdomain 4 of monomer i + 1 of the other strand, strengthening diagonal contacts between monomers.

The hydrophobic 262–274 plug between subdomains 3 and 4 seems quite flexible according to the normal mode analysis (Tirion and ben-Avraham, 1993) and the normal mode refinement. However, a rebuild of this loop into a β-bend, postulated to extend into a hydrophobic cavity in the neighboring long-pitch strand, cannot be achieved by a normal mode refinement. Upon examination of the filament structure it becomes apparent that these three regions from three neighboring monomers, the 38–52 loop, the COOH terminus, and the hydrophobic plug, are all in close proximity. Residues 262–274 are able to interact with residues 166–173 and 285–288 in subdomain 3 of monomer i + 1 in the neighboring long-pitch helix, as well as with subdomain 2 (residues 63–66 and 38–40) of monomer i − 1 of the same long-pitch strand. Similarly, the COOH-terminal arm seems poised to interact with, and therefore potentially influence structurally, subdomain 4 (residues 223–230) of monomer i − 1 in the neighboring long-pitch helix. These interactions suggest that the bonding between neighboring actin monomers is a cooperative process, one where the final configuration of the various loops and regions within one monomer

is not achieved until all the pieces (or all neighboring monomers) are in place. This observation is supported by the report of Owen and DeRosier (1993), who suggested, based on their work with EM reconstructions, that the COOH terminus, subdomain 2, and helix 223–230 from three neighboring monomers may interact cooperatively to produce a stable filament.

Given the possibility that the binding process between actin monomers is a collective, cooperative process, it will be difficult to model the G-actin to F-actin transition by examining one monomer in isolation. The filament will need to be refined as a whole, integral unit. Neither the directed mutation algorithm nor the normal mode refinement algorithm adequately monitors the collective, stereochemical repercussions of intermonomer interactions when refining the (isolated) monomers. One important aspect for feature research, therefore, will be to develop practical methods to refine three to five monomers simultaneously.

We thank W. Kabsch for useful suggestions and comments.
This material is based upon work supported by the National Science Foundation under grant MCB-9316109 (to M. M. Tirion and D. ben-Avraham). Coordinates of the normal mode refined model of actin are available upon request by E-mail from tirion@craft.camp.clarkson.edu.

REFERENCES

Brooks, B., and B. Karplus. 1983. Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor. *Biophysics.* 80:6571–6575.

Brünger, A. T. 1990. X-PLOR (Version 2.1) Manual. Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT.

Chen, X., R. K. Cook, and P. A. Rubinstein. 1993. Yeast actin with a mutation in the "hydrophobic plug" between subdomains 3 and 4 (L266D) displays a cold-sensitive polymerization defect. *J. Cell Biol.* 123: 1185–1195.

dos Remedios, C., M. Miki, and J. A. Barden. 1987. *J. Muscle Res. Cell. Motil.* 8:97–117.

Elzinga, M., and J. J. Phelan. 1984. F-actin is intermolecularly crosslinked by N,N′-*p*-phenylene-dimaleimide through lysine-191 and cysteine-372. *Proc. Natl. Acad. Sci. USA.* 81:6599–6602.

Gō, N., T. Noguti, and T. Nishikawa. 1983. Dynamics of a small globular protein in terms of low frequency vibrational modes. *Proc. Natl. Acad. Sci. USA.* 80:3696–3700.

Hegyi, G., H. Michel, J. Shabanowitz, D. F. Hunt, N. Chatterjee, G. Healy-Louie, and M. Elzinga. 1992. Gln-41 is intermolecularly cross-linked to Lys-113 in F-actin by N-(4-azidobenzoyl)-putrescine. *Protein Sci.* 1:132–144.

Holmes, K. C., and J. Barrington Leigh. 1974. The effect of disorientation on the intensity distribution of non-crystalline fibres. I. Theory *Acta Cryst. A* 30:635–638.

Holmes, K. C., D. Popp, W. Gebhard, and W. Kabsch. 1990. Atomic model of the actin filament. *Nature.* 347:44–49.

Holmes, K. C., M. M. Tirion, D. Popp, M. Lorenz, W. Kabsch, and R. A. Milligan. 1993. A comparison of the atomic model of F-actin with cryo-electron micrographs of actin and decorated actin. *In* Mechanism of Myofilament Sliding in Muscle Contraction. H. Sugi and G. H. Pollack, editors. Plenum Publishing, New York. 15–24.

Kabsch, W. 1976. A solution for the best rotation to relate two sets of vectors. *Acta Cryst.* A32:922–923.

Kabsch, W., H. G. Mannherz, D. Suck, E. Pai, and K. C. Holmes. 1990. Atomic structure of the actin:DNase I complex. *Nature.* 347:37–44.

Klug, A., F. H. C. Crick, and H. W. Wyckoff. 1958. Diffraction by helical structures. *Acta Cryst.* 11:199–213.

Levitt, M. 1983. Molecular dynamics of native protein. I. Computer simulation of trajectories. *J. Mol. Biol.* 168:595–620.

Levitt, M., C. Sander, and P. S. Stern. 1985. Protein normal-mode dynamics: trypsin inhibitor, crambin, ribonuclease, and lysozyme. *J. Mol. Biol.* 181: 423–447.

Lorenz, M., and K. C. Holmes. 1993. Computer processing and analysis of x-ray fiber diffraction data. *J. Appl. Crystallogr.* 26:82–91.

Lorenz M., D. Popp, and K. C. Holmes. 1993. Refinement of the F-actin model against x-ray fiber diffraction data by the use of a directed mutation algorithm. *J. Mol. Biol.* 234:826–836.

Makowski, L. 1991. An estimate of the number of structural parameters measurable from a fiber diffraction pattern. *Acta. Cryst.* A47:562.

Mannherz, H. G., W. Kabsch, and R. Leberman. 1977. Crystals of skeletal muscle actin:pancreatic DNase I complex. *FEBS Lett.* 73:141–143.

Miki, M., and T. Kouyama. 1994. Domain motion in actin observed by fluorescence resonance energy transfer. *Biochemistry* 33: 10171–10177.

Milligan, R. A., M. Whittaker, and D. Safer. 1990. Molecular structure of F-actin and location of surface binding proteins. *Nature.* 348: 217–221.

Orlova, A., and E. H. Egelman. 1992. Structural basis for the destabilization of F-actin by phosphate release following ATP hydrolysis. *J. Mol. Biol.* 227:1043–1053.

Orlova, A., and E.H. Egelman. 1993. A conformational change in the actin subunit can change the flexibility of the actin filament. *J. Mol. Biol.* 232:334–341. Owen, C., and D. DeRosier. 1993. A 13-Å map of the actin-scruin filament from the limulus acrosomal process. *J. Cell Biol.* 123:337–344.

Owen, C., and D. DeRosier. 1993. A 13-Å map of the actin-scruin filament from the limulus acrosomal process. *J. Cell. Biol.* 123:337–344.

Popp, D., V. V. Lednev, and W. Jahn. 1987. Methods of preparing well-oriented sols of F-actin containing filaments suitable for x-ray diffraction. *J. Mol. Biol.* 181:423–447.

Press, W. H., B. P. Flannery, W. T. Teukolsky, and W. T. Vetterling. 1990. Numerical Recipes: The Art of Scientific Computing. Cambridge University Press, Cambridge.

Thomas, D. D., J. C. Seidel, and J. Gergely. 1979. Rotational dynamics of spin-labeled F-actin in the sub-millisecond time range. *J. Mol. Biol.* 132: 257–273.

Tirion, M. M., and D. ben-Avraham. 1993. Normal Mode Analysis of G-Actin. *J. Mol. Biol.* 230:186–195.