Volume 236, number 1, 71–76 FEB 06163 August 1988

Protein dynamics and distance determination by NOE measurements

David M. LeMaster[+], Lewis E. Kay*, Axel T. Brünger[+°] and J.H. Prestegard*

*Departments of [+]Molecular Biophysics and Biochemistry and *Chemistry, Yale University* and *[°]The Howard Hughes Medical Institute, New Haven, CT 06511, USA*

Received 8 June 1988

Present analysis procedures for NMR structure determination of macromolecules presuppose fixed internuclear distances. Improvement of the precision of the requisite NOE information has stimulated the use of more quantitative distance constraints thus necessitating examination as to whether the assumption of a rigid model systematically biases the distance estimates. Analysis using the simple $\langle r^{-6} \rangle$ dependence of NOE buildup rates seriously underestimates the correct distance for spatially proximal proton pairs having fluctuations comparable to those observed in X-ray temperature factor analysis. However, by calculating the proper generalized order parameter it is shown that for nuclei undergoing rapid isotropic uncorrelated fluctuations the effective distance is identical to the distance between the mean positions of the nuclei. Similar analysis of molecular dynamics simulation data from bovine pancreatic trypsin inhibitor indicates that the distance obtained from the generalized order parameter predicts the distance between the mean positions to within a few percent regardless of the degree of correlation of the pairwise motion for virtually all main chain and dynamically constrained side chain protons.

Protein molecular dynamics; NOE; Distance determination

1. INTRODUCTION

Distance constraints derived from NOE measurements have proven sufficient for the determination of the solution structures of several small proteins (e.g. [1–5]). In most cases NOE intensities obtained from NOESY experiments have been used only semi-quantitatively by classification according to weak, medium and strong crosspeaks. Using these rather coarse constraints it has been possible to determine the overall course of the mainchain with rms deviations from known X-ray structures of roughly 1.5–2.5 Å. Sidechain conformations as well as local mainchain conformations are generally less well determined as might be expected from a qualitatively analogous moderate resolution X-ray structural analysis [6]. With more accurate distance constraints it should become possible to improve the effective resolution of NMR structure determinations.

Problems such as baseline distortions (e.g. T1 noise), spin diffusion and sensitivity limitations have been the major reasons for the previous use of the semi-quantitative NOE analysis. However, recently techniques have been developed to suppress T1 noise [7–9] and spin diffusion effects [10] as well as to allow for more reliable internal calibration of crosspeak intensities [11–13]. These advances have stimulated attempts to obtain precise distance information. All such calculations have assumed that the internuclear distances are fixed and that these vectors have similar autocorrelation behavior. Although rigorous analysis of relaxation requires calculation of the full spectral density function, this is not presently feasible. Fur-

Correspondence address: D.M. LeMaster, Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06511, USA

Abbreviations: NOE, nuclear Overhauser enhancement; BPTI, bovine pancreatic trypsin inhibitor; rms, root mean square

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/88/$3.50

thermore such a calculation requires a structural model and hence could only be of use in the final stages of verification in the structural determination.

Of primary concern in quantitative NOE analysis is whether the assumption of rigid internuclear distances introduces systematic errors in the distance measurements. From the point of view of structural analysis the related question is whether the distance between the mean nuclear positions can be determined from NMR data since only this measure of internuclear distance will give rise in general to an internally consistent set of distance constraints. X-ray temperature factor analysis and molecular dynamics simulations monitor the rapid vibrational and librational motions which give rise to a dispersion of the nuclear positions of the order of 0.5 Å at room temperature [14]. Since NOE buildup rates are generally analyzed in terms of a simple r^{-6} dependence, one might anticipate that the apparent distance for two such fluctuating nuclei would be biased heavily toward their distance of closest approach. For example, two protons undergoing independent Gaussian fluctuations with rms deviations of 0.5 Å whose mean positions are 3.0 Å apart would have an apparent separation of 2.2 Å using a r^{-6} dependence.

A proper analysis requires that angular as well as radial fluctuations be considered. Angular dispersion tends to counteract the radial fluctuations by resulting in an apparent increase in internuclear distance. Nevertheless it has been suggested that significant systematic underestimates of the internuclear distances still remain [15]. More recent analyses [16,17], in part based on molecular dynamics simulation using a unified atom model for protonated carbons, suggest this systematic bias may be less than earlier estimated.

In order to examine the question of systematic bias in the distance estimates more carefully, we have carried out a more extensive full atom molecular dynamics simulation of bovine pancreatic trypsin inhibitor and analyzed the results in terms of a simple random fluctuation model.

2. METHODS

The computational strategy for the calculation of dynamics trajectories of proteins, including the effects of water, using stochastic boundary molecular dynamics has been discussed in detail in a previous publication [18]. We discuss here only those points which are specific to the present calculation.

The X-ray structure of BPTI refined to an R factor of 16.2% at 1.5 Å [19,20] was used as a starting point for the simulations. The stochastic boundary dynamics method partitions the system into three discrete domains: the reaction region, consisting of the region of interest; the buffer region, consisting of the region immediately surrounding the reaction region and the excluded region, consisting of the remainder of the system. In the present study a reaction region with a 16 Å radius centered at the C_α of tyrosine 35 was chosen. All mainchain atoms within 16 Å of the reference point were included in the reaction region while all sidechain atoms of any particular residue were labeled as reaction region atoms if any of the sidechain atoms of that residue were within the 16 Å radius. All atoms between 16 Å and 18 Å from the reference point were labeled as buffer region atoms. Atoms outside a 18 Å radius from the reference point were excluded from the calculations. The system was hydrated and equilibrated using a procedure described previously [18]. After hydration and equilibration the resulting system contained 762 protein atoms and 604 water molecules. A dielectric constant of unity was employed throughout the system and nonbonded interactions up to 11 Å were included.

The simulation was carried out using the OPLS force field [21] with several modifications. Nonpolar hydrogens were included in the force field with charges obtained from AMBER [22]. The charges of the directly bonded carbons were modified to preserve neutrality. The ϵ and δ values of the Lennard-Jones potential for nonpolar hydrogens and protonated carbons were taken from AMBER. All other parameters were from the OPLS force field.

For the NOE calculations only proton pairs which lay within 14 Å of the reference point were considered in order to minimize boundary effects. Although the reaction region is predominantly filled with residues of the core β-sheet domain, it seems unlikely that the general results discussed herein will be strongly dependent on secondary structural conformation. The NOE analysis presented involves averages taken at 0.1 ps intervals over 215 ps of simulation. Analysis of the first and second 100 ps segments yielded equivalent results. The length of the simulation was chosen for other purposes and presumably a period significantly less than 100 ps [17] should prove sufficient for analyses similar to that described here.

The random fluctuation calculations were performed by choosing points with x, y and z coordinates sampled from a Gaussian distribution of the appropriate standard deviation. Points from two such distributions were then displaced by the mean separation and the generalized order parameter calculation was carried out. Twenty to fifty thousand pairs of points were used for each calculation.

3. RESULTS AND DISCUSSION

The initial rate of NOE buildup between an isolated proton pair is proportional to their cross-relaxation rate σ_{ij} given by the formula

$$\sigma_{ij} = \frac{6\pi}{5}\gamma^4\hbar^2[2J_{ij}(2\omega) - (1/3)J_{ij}(0)]$$

where γ is the gyromagnetic ratio, ω is the ^1H Larmor frequency and $J_{ij}(\ )$ are components of a spectral density function.

$$J^n_{ij}(\omega) = \int_0^\infty \left\langle \frac{Y^2_n(\theta(t),\phi(t))\, Y^2_n(\theta(0),\phi(0))}{r^3_{ij}(t) r^3_{ij}(0)} \right\rangle e^{i\omega t} dt \qquad (1)$$

where Y^2_n are the second order spherical harmonics [23].

In this analysis we wish to focus our attention exclusively on the effects of those fluctuations which define the thermal ellipsoids. It is these rapid small amplitude fluctuations which dominate the thermodynamic behavior in the native state and have been studied extensively in comparisons between molecular dynamics simulations and X-ray temperature factor analysis. In order for differential effects of these motions to be manifested directly in comparisons of observed NOE buildup rates, the lower frequency internal motions and the overall molecular reorientation must be similar for the various NOE crosspeaks considered.

If one assumes that the dissipation of correlated motions within these thermal ellipsoids is rapid compared to the larger scale internal motions and the molecular tumbling, this component of the internal motion correlation function can be factored out as

$$\frac{4\pi}{5}\sum_{-2}^{2}\left[\left\langle \frac{Y^2_n(\theta(0),\phi(0))}{r^3_{ij}} \right\rangle\right]^2 \equiv S^2 \qquad (2)$$

where S is defined as the generalized order parameter [17,24]. This formula has been shown to be valid for many cross-relaxation interactions in proteins for an extended plateau region which exists after the first few picoseconds in which thermal correlations dissipate and before the onset of larger scale motions such as methyl rotations with correlation times of 20–200 ps [17]. The ^1H-^1H relaxation effects of the correlated motion within the thermal ellipsoids which dissipates within the first few picoseconds have been shown to be negligible in globular proteins [17,25]. In the plateau region where the equilibrium distribution within the thermal ellipsoid is effectively sampled, the ensemble average indicated in the generalized order parameter equation can be replaced by the corresponding time average obtained from molecular dynamics calculations. The inverse of the sixth root of the left side of eqn 2 has the dimension of distance and serves to define the effective distance measured by cross-relaxation experiments.

Effective distances were calculated from the 215 ps BPTI molecular dynamics simulation for various pairwise proton interactions. Since the assumption of similar correlation functions for lower frequency motions is essential for this analysis, protons on methyl groups as well as sidechain methylene protons past the β-positions on the generally mobile arginine, lysine, glutamine and glutamic acid residues have been excluded. In order to examine the effects of correlated motion two subsets of the proton pairs were considered: those pairs having one or two variable dihedral angles between the interacting protons and those pairs arising from positions separated by more than two residues in the sequence (i.e. 'long range'). Within these two subsets only those proton pairs were used for which neither the initial equilibrated distance nor the mean distance was greater than 4.5 Å and for which the difference between the initial and the mean distances were less than 0.5 Å.

In fig.1A is given the histogram of occurrences of the effective distance derived from the order parameter calculation divided by the distance between the mean positions for 206 proton pairs separated by one or two dihedral angles and in fig.1B is shown the corresponding histogram for 122 long range proton pairs. It is striking how accurately the generalized order parameter calculation serves to estimate the internuclear distance. For the long range proton pairs on average the effective distance is within 0.4% of the distance between the mean nuclear positions and the standard deviation σ of the distribution is only 2.3%. Even in the case of the proton pairs separated by only one or two flexible dihedral angles for which correlated motion might be expected to be more significant the mean effective distance is within 2.8% of the distance between the mean positions.

Insight into the precise cancellation of the effects of radial and angular fluctuations on the generalized order parameter calculation can be gained by reference to the earlier mentioned case of two nuclei separated by 3.0 Å both of which undergo 0.5 Å isotropic Gaussian fluctuations. In contrast to the greater than 25% discrepancy in the

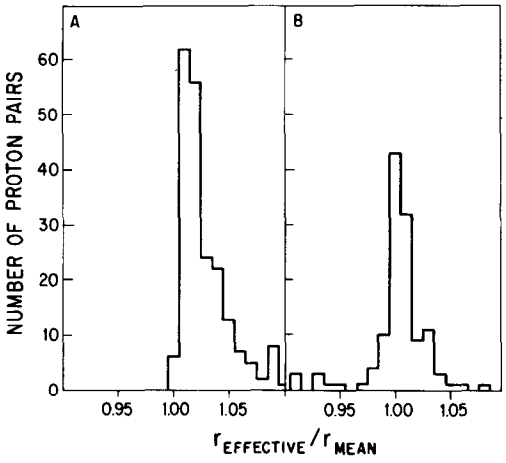

Fig.1. Effective distance measurement derived from BPTI simulation of high frequency cross-relaxation. The generalized order parameter was calculated for 216 proton pairs separated by one or two variable dihedral angles (panel A) and 122 proton pairs arising from residues at least three residues apart in sequence (panel B). Normalized to the distance between the mean nuclear positions, the average ratio was 1.028 for panel A and 1.004 for panel B.

distance estimate using a $\langle r^{-6} \rangle$ dependence, a generalized order parameter calculation yields a deviation of less than 0.2%. This effect can be understood theoretically by evaluation of the generalized order parameter for a spherical shell of density with respect to an exterior point. We first note that the $e^{ni\phi}$ factor present in all spherical harmonic terms except Y_0^2 go to zero upon integration from 0 to 2π. Hence we need only evaluate the integral

$$\int_{\theta=0}^{\pi}\int_{\phi=0}^{2\pi} \tfrac{1}{2}\frac{(3\cos^2\theta - 1)}{r^3}A^2\sin\theta'\,d\theta'\,d\phi \Big/ \int_{\theta=0}^{\pi}\int_{\phi=0}^{2\pi} A^2\sin\theta'\,d\theta'\,d\phi$$

Using the law of cosines to substitute for $\cos^2\theta$ and $\sin\theta'\,d\theta'$ one obtains

$$\int_{D-A}^{D+A} \tfrac{1}{4}\frac{(3(\frac{r^2+D^2-A^2}{2rD}) - 1)}{r^3}\cdot\frac{r}{AD}\cdot dr = \frac{1}{D^3}$$

where the variables are defined as in fig.2. Since this relationship must hold for all points of a second spherically symmetric distribution, by symmetry the converse must apply. Hence for any pair

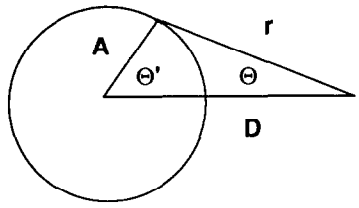

Fig.2. Calculation of the generalized order parameter for a spherical shell with respect to an exterior point.

of nuclei undergoing isotropic uncorrelated motion in non-overlapping distributions, the NOE distance estimate based on these motions is identically equal to the separation between their mean positions.

The close correspondence between the results obtained from the molecular dynamics simulation and those derived from the random fluctuation model does not imply that the atomic motions are in fact isotropic and uncorrelated. Considerable discussion has focused on the anisotropy of atomic motions in molecular dynamics simulations and the resulting inappropriateness of the isotropic temperature factors standardly used in X-ray analysis [26]. The possibility that the order parameter calculations are not particularly sensitive to the asymmetry of the nuclear dispersion can be eliminated by reference to fig.3. The dependence of the dipolar interaction on the asymmetry of the motion was studied by applying the independent fluctuation model to nuclei having prolate and oblate ellipsoids of dispersion with the unique axis oriented along the internuclear vector. In fig.3 are plotted contours of the ratio of effective distance to the distance between the mean positions as a function of the axial ratio of the dispersion ellipsoids and of the distance between the means normalized to the dispersion along the internuclear vector. Significant discrepancies between the effective distance and the distance between the mean positions can arise under physically plausible conditions. If we consider the case of a 0.5 Å dispersion along the interatomic vector and a 1.0 Å dispersion along the perpendicular axes at 3.0 Å separation, the effective distance is 23% greater than the distance between the mean positions. Particularly striking is how gradually the effect of the asymmetry decreases with distance. At 4.0 Å and 5.0 Å the corresponding discrepancies are 15% and 10%, respectively.

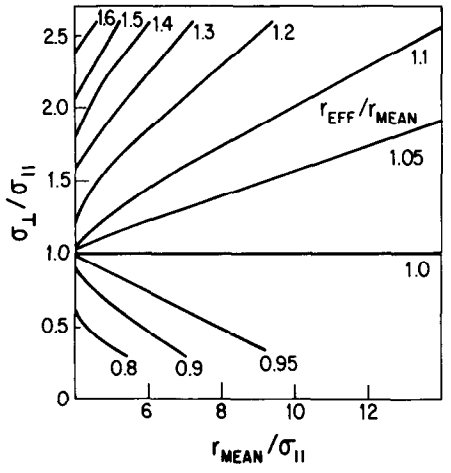

Fig.3. Effective NOE distance as a function of asymmetry of the thermal ellipsoids of the interacting proton pairs. Using the random fluctuation approximation the generalized order parameter was calculated for two protons having oblate ($\sigma_\perp/\sigma_\parallel > 1$) or prolate ($\sigma_\perp/\sigma_\parallel < 1$) thermal ellipsoids aligned along the internuclear vector. Constant value contours of the ratio of the resultant effective distance to the distance between the mean nuclear positions are plotted as a function of the ellipsoid asymmetry along the ordinate and the normalized separation distance along the abscissa.

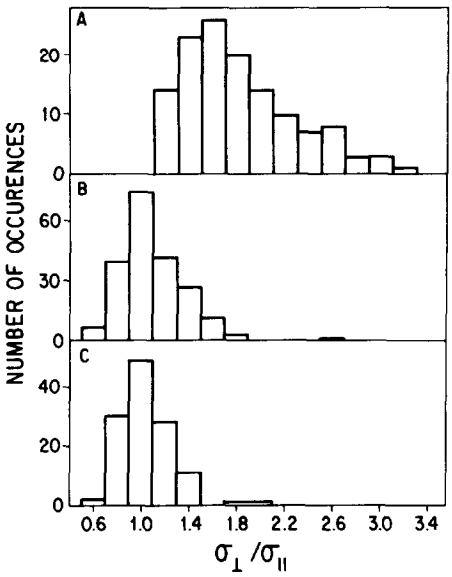

Fig.4. Correlated asymmetry of the thermal ellipsoids of proton pairs in BPTI simulation. The asymmetry of the individual thermal ellipsoids for those nuclei used in the proton pair calculations is presented in panel A with the minor axis of the ellipsoid serving as the reference axis. In panel B and C are shown comparable calculations for proton pairs separated by one or two dihedral bonds and long range proton pairs, respectively. In this case the calculation for both ellipsoids are made relative to the internuclear vector and the geometric mean of their asymmetry factors is plotted.

For comparison the axial ratio between the rms dispersions perpendicular and parallel to the minor axis of the individual thermal ellipsoids obtained from the BPTI simulation for the atoms included in the proton pairs considered in fig.1 is plotted in fig.4A. Note that there is a high degree of asymmetry in the distribution. The apparent contradiction with the previous result of effective isotropy in the generalized order parameter calculation on the same data can be resolved by noting that the NMR experiment does not sense the asymmetry per se but rather the difference between the radial and angular components with respect to the internuclear vector which in general does not lie along a principal axis of the thermal ellipsoid. If instead we consider the components of the thermal ellipsoids parallel and perpendicular to the internuclear vector as in fig.4B and C we see that the evidence of the anisotropy of the individual nuclear dispersions is largely obscured. In particular for the long range proton pairs considered in fig.4C, using the thermal ellipsoids obtained from the simulation, the mean and standard deviations of the $\sigma_\perp/\sigma_\parallel$ values are calculated to be 1.035 and 0.19, respectively, if there is no correlation between the orientation of the thermal ellipsoids and the internuclear vector. This compares quite well with a mean and standard deviation of 1.04 and 0.21 for the $\sigma_\perp/\sigma_\parallel$ obtained from the orientations which were actually observed in the BPTI simulations. In the cases where correlation of the orientation of the thermal ellipsoids with the internuclear vector gives rise to large asymmetries, the corresponding generalized order parameter calculations indicate the anticipated bias in apparent distance.

A useful measure of the degree of correlated motion can be gained by comparison of the standard deviation of the internuclear distance obtained from the molecular dynamics simulation and the corresponding value obtained by assuming random fluctuations within the thermal ellipsoids of each pair of protons. For the proton pairs separated by only one or two flexible dihedral angles the average values are 0.16 Å and 0.66 Å, respectively. Hence, correlated motions contribute the major share of the thermal ellipsoids for these proton pairs. The

apparent isotropy implied by the generalized order parameter calculation suggests that the reduced dispersion of the internuclear distances obtained in the simulation should be accompanied by a reduced angular dispersion as is observed. Calculation of the corresponding values for the long range proton pairs (0.40 Å and 0.61 Å, respectively) indicate that, as expected, random fluctuation provides a much more realistic model in this case.

In summary it appears that most proton pairwise interactions in proteins not involving methyl groups or flexible sidechains on the external surface can be effectively modeled by the use of single Gaussian thermal ellipsoids for which the correlation of rapid thermal fluctuations dissipates much more rapidly than the onset of dynamically significant larger scale motions. When combined with the additional observation that there is relatively little correlation of the orientation of the thermal ellipsoids with the internuclear vectors, it is concluded that the generalized order parameter calculation gives a highly accurate measure of the distance between the mean nuclear positions as predicted from the simple random fluctuation model. Hence for these dynamically constrained nuclei the analysis of globular proteins based on a rigid model are well justified within the present precision of NOE measurements. On the other hand these results also indicate that for such systems proton-proton dipolar interactions are unlikely to provide a useful probe of small amplitude fluctuations in the picosecond timescale.

It should be noted that the case of an effective isotropy of motion may not apply equally to all macromolecular systems. For example, in studies of DNA duplex structures the interproton vectors directed approximately along the helix axis provide the main source of useful distance constraints. These may conceivably have significant differences between the radial and angular components of fluctuation. In such a case sequence variations in writhing behavior would be incorrectly interpreted as local bending using the standard rigid model analysis. Whether such asymmetry in the molecular motion is practically significant will be addressed in future molecular dynamics simulations similar to those described here.

Acknowledgements: This investigation was supported by NIH Project Grants GM-38779 and GM-33225 and Program Grant GM-22778. D.L. would like to thank F.M. Richards for his support during this work.

REFERENCES

[1] Braun, W., Wider, G., Lee, K.H. and Wüthrich, K. (1983) J. Mol. Biol. 169, 921–948.
[2] Kline, A.D., Braun, W. and Wüthrich, K. (1986) J. Mol. Biol. 189, 377–382.
[3] Clore, G.M., Sukumaran, D.K., Nilges, M., Zarbock, J. and Gronenborn, A.M. (1987) EMBO J. 6, 529–537.
[4] Moore, J.M., Case, D.A., Chazin, W.J., Gippert, G.P., Havel, T.F., Powls, R. and Wright, P.E. (1988) Science 240, 314–317.
[5] Holak, T.A., Kearsley, S.K., Kim, Y. and Prestegard, J.H. (1988) Biochemistry, in press.
[6] Wagner, G., Braun, W., Havel, T.F., Schaumann, T., Go, N. and Wüthrich, K. (1987) J. Mol. Biol. 196, 611–639.
[7] Bodenhausen, G. and Ernst, R.R. (1982) Mol. Phys. 47, 319–328.
[8] Otting, G., Widmer, H., Wagner, G. and Wüthrich, K. (1986) J. Magn. Res. 66, 187–193.
[9] Barsukov, I.L. and Arseniev, A.S. (1978) J. Magn. Res. 73, 148–149.
[10] LeMaster, D.M. and Richards, F.M. (1988) Biochemistry 27, 142–150.
[11] Denk, W., Baumann, R. and Wagner, G. (1986) J. Magn. Res. 67, 386–390.
[12] Macura, S., Farmer, B.T. and Brown, L.R. (1986) J. Magn. Res. 70, 493–499.
[13] Holak, T.A., Scarsdale, J.N. and Prestegard, J.H. (1987) J. Magn. Res. 74, 546–549.
[14] Ringe, D. and Petsko, G.A. (1985) Prog. Biophys. Mol. Biol. 45, 197–235.
[15] Bothner-By, A.A. and Spevacek, J. (1982) Pure Appl. Chem. 54, 569–574.
[16] Keepers, J.W. and James, T.L. (1984) J. Magn. Res. 57, 404–426.
[17] Olejniczak, E.T., Dobson, C.M., Karplus, M. and Levy, R.M. (1984) J. Am. Chem. Soc. 106, 1923–1930.
[18] Brünger, A.T., Huber, R. and Karplus, M. (1987) Biochemistry 26, 5153–5162.
[19] Deisenhofer, J. and Steigemann, W. (1975) Acta Cryst. B31, 238–250.
[20] Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M. (1977) J. Mol. Biol. 112, 535–542.
[21] Jorgensen, W.L. and Tirado-Rives, J. (1988) J. Am. Chem. Soc. 110, 1657–1666.
[22] Weiner, P.K. and Kollman, P.A. (1981) J. Comput. Chem. 2, 287–303.
[23] Brink, D.M. and Satchler, G.R. (1968) Angular Momentum, Clarendon Press, Oxford.
[24] Lipari, G. and Szabo, A. (1982) J. Am. Chem. Soc. 104, 4546–4559.
[25] Swaminatham, S., Ichiye, T., Van Gunsteren, W. and Karplus, M. (1982) Biochemistry 21, 5230–5241.
[26] Kuriyan, J., Petsko, G.A., Levy, R.M. and Karplus, M. (1986) J. Mol. Biol. 190, 227–254.