PROTEINS: Structure, Function, and Genetics 17:297–309 (1993)

A Calculation Strategy for the Structure Determination of Symmetric Dimers by ^{1}H NMR

Michael Nilges
European Molecular Biology Laboratory, D-6900 Heidelberg, Germany

ABSTRACT **The structure determination of symmetric dimers by NMR is impeded by the ambiguity of inter- and intramonomer NOE crosspeaks. In this paper, a calculation strategy is presented that allows the calculation of dimer structures without resolving the ambiguity by additional experiments (like asymmetric labeling). The strategy employs a molecular dynamics-based simulated annealing approach to minimize a target function. The experimental part of the target function contains distance restraints that correctly describe the ambiguity of the NOE peaks, and a novel term that restrains the symmetry of the dimer without requiring the knowledge of the symmetry axis. The use of the method is illustrated by three examples, using experimentally obtained data and model data derived from a known structure. For the purpose of testing the method, it is assumed that *every* NOE crosspeak is ambiguous in all three cases. It is shown that the method is useful both in situations where the structure of a homologous protein is known and in ab initio structure determination. The method can be extended to higher order symmetric multimers.**

Key words: nuclear magnetic resonance, nuclear overhauser effect, dimer, solution structure, symmetry

INTRODUCTION

Increasing use has been made of ^{1}H-NMR to study the solution conformation of dimeric proteins.[1–7] The structure determination of symmetric homodimers by NMR is impeded by the fact that it is intrinsically impossible to distinguish between inter- and intramonomer NOEs. The only way to resolve this ambiguity completely is asymmetric labeling.[1,8,9]

Since the intra- and intermonomer distances between a given pair of protons are in general very different, a measured NOE can be assigned as intra- or intermonomer once the three-dimensional structure is known. For many NOEs, this distinction is possible even if the structure is only approximately known, for example, if it is the structure of a homologous protein. Thus, in an approach that has been likened to the molecular replacement technique in X-ray crystallography, Breg et al.[3] assigned the NOE spectrum of the Arc repressor by use of the known X-ray crystal structure coordinates of the homologous Metj repressor.[10] Several cycles of refinement and assignment were necessary until all the NOE crosspeaks could be accounted for. In fortunate cases it can be possible to sort out the different contributions from the spectra alone.[2,7]

Obviously, a method that would automate the process of distinguishing inter- and intramonomeric NOEs would have several advantages. Going through several cycles of assignments and structure calculation is rather tedious. Furthermore, there is unavoidable bias toward the assignments made in the first iteration. Ideally, the method should be completely automatic, so that all bias toward an initial model is avoided.

The strategy described in this report is based on minimization of a target function by simulated annealing. The problem of assignment of inter- and intramonomer NOEs is circumvented by adding to the NMR part of the target function the appropriate expression for an NOE-derived distance in a symmetric dimer. If one looks at the problem from a naive point of view, examining all possible combinations of inter- and intramonomer crosspeaks is a task of prohibitive combinatorial complexity, since for N crosspeaks, there are 2^N possibilities. Obviously, the choices are correlated through the three-dimensional structure, and much fewer possibilities will have to be searched. However, the task is much more complex than calculating the three-dimensional structure based on unambiguously assigned NOEs.

The strategy is tested with three data sets, two of which are experimental, one derived from a crystal structure. Calculations are performed both starting from approximate models, and without prior knowledge of the fold of the protein. The success and the present limitations of the approach are discussed.

Abbreviations: NOE, nuclear Overhauser effect; NMR, nuclear magnetic resonance; RMS, root mean square.

Received April 26, 1993; accepted May 26, 1993.
Address reprint requests to Dr. Michael Nilges, European Molecular Biology Laboratory, Meyerhofstrasse 1, D-6900 Heidelberg, Germany.

MATERIALS AND METHODS

NOE Target Function

The structure determination protocol described below makes use of the fact that strictly speaking *every* NOE crosspeak in a symmetric dimer has contributions from intramonomer and intermonomer effects:

$$NOE_{ij} = 2NOE_{ij}^{\text{intra}} + 2NOE_{ij}^{\text{inter}}. \quad (1)$$

Usually, one contribution is much bigger than the other, such that the NOE can be assigned as intra- or intermonomer. However, Eq. (1) is formally correct for every peak in the spectrum.

In the isolated spin pair approximation, the two contributions to the crosspeak depend on the corresponding intra- and intermonomer distances R_{ij}^{intra} and R_{ij}^{inter}:

$$\frac{d}{d\tau_m} NOE_{ij}^{\text{intra}} = cR_{ij}^{\text{intra}^{-6}} \quad (2)$$

$$\frac{d}{d\tau_m} NOE_{ij}^{\text{inter}} = cR_{ij}^{\text{inter}^{-6}} \quad (3)$$

where c is a proportionality constant. Thus, assuming that the order parameters and internal correlation times for the inter- and intramonomer vectors are identical, one obtains for the symmetric dimer

$$\frac{d}{d\tau_m} NOE_{ij} = 2c(R_{ij}^{\text{intra}^{-6}} = R_{ij}^{\text{inter}^{-6}}). \quad (4)$$

The total NOE in the dimer is therefore related to an "effective distance" that is calculated from the intra- and intermonomer distances

$$\frac{d}{d\tau_m} NOE_{ij} = 2c\overline{R}_{ij}^{-6} \quad (5)$$

with

$$\overline{R}_{ij} = (R_{ij}^{\text{intra}^{-6}} + R_{ij}^{\text{inter}^{-6}})^{-\frac{1}{6}}. \quad (6)$$

Using this expression, the restraints derived from the NOE spectrum of a symmetric dimer can be easily incorporated into a target function

$$E = E_{\text{chem}} + w_{\text{NOE}}E_{\text{NOE}} \quad (7)$$

where E_{chem} describes chemical knowledge (such as bond lengths, bond angles, planarity, chirality, and nonbonded interactions), E_{NOE} represents the NOE-derived data, and w_{NOE} is the weight on the experimental part of the target function. In the present paper, we use a harmonic "flat-bottom" potential (or square-well with harmonic walls) with a linear behavior for large deviations[11] for E_{NOE}

$$E_{\text{NOE}} = \sum_k \begin{cases} (L_k - \overline{R}_k)^2 & \text{if } \overline{R}_k < L_k \\ 0 & \text{if } L_k \le \overline{R}_k \le U_k \\ (\overline{R}_k - U_k)^2 & \text{if } U_k < \overline{R}_k < S \\ A(\overline{R}_k - S)^{-1} + B(\overline{R}_k - S) + C & \text{if } \overline{R}_k > S \end{cases} \quad (8)$$

where L_k and U_k are the lower and upper bounds on the effective distance derived from the slope of the NOE buildup, $\overline{R_k}$ is the effective distance measured in the current model for a particular restraint, and k runs over all the distance restraints. S is the value where the potential switches between the harmonic and asymptotic parts, B is the slope of the potential at large deviations, and the coefficients A and C are determined such that the potential is continuous and differentiable at S.

Switching between the intra- and intermonomer contributions to an NOE intensity, and thus the corresponding "assignment," is achieved by the potential form of the NOE pseudo-energy, which is heavily weighted toward the smaller distance. Equation (6) also describes situations correctly where both intra- *and* intermonomer contributions are present, i.e., for proton pairs close to the dyad axis of the dimer.

Obviously, the above expression [Eq. (6)] is closely related to the $<r^{-6}>$ average that has been used to treat, for example, not stereospecifically assigned methylene groups.[12] One difference between the two expressions is that in Eq. (6) no average is taken over the r^{-6} terms, but they are simply summed up. The effect is that the resulting effective distance is always smaller than the smallest of the contributing distances, while the $<r^{-6}>$ average distance lies between the smallest and the largest involved distance. For example, for two distances of 3 and 10 Å, the $<r^{-6}>$ average distance is 3.37 Å, while Eq. (6) yields 2.99 Å. For an ambiguous crosspeak, the latter is obviously the desired result. The more important difference, however, lies in the usage described in this paper. It will be shown that it is possible to use ambiguous crosspeaks on a large scale during the structure determination. The $<r^{-6}>$ average has so far been used only in situations where the assignment is of comparatively small consequences for the global fold of the protein, since the distances to the two protons of a methylene group, say, are fairly similar. By contrast, it is very important for the polypeptide fold of a symmetric dimer if a particular NOE is intra- or intermonomer, since the corresponding distances are in general very different.

Symmetry Target Function

In addition to the NOE-derived restraints on interatomic distances, the symmetry properties of the molecule are important experimental data that are

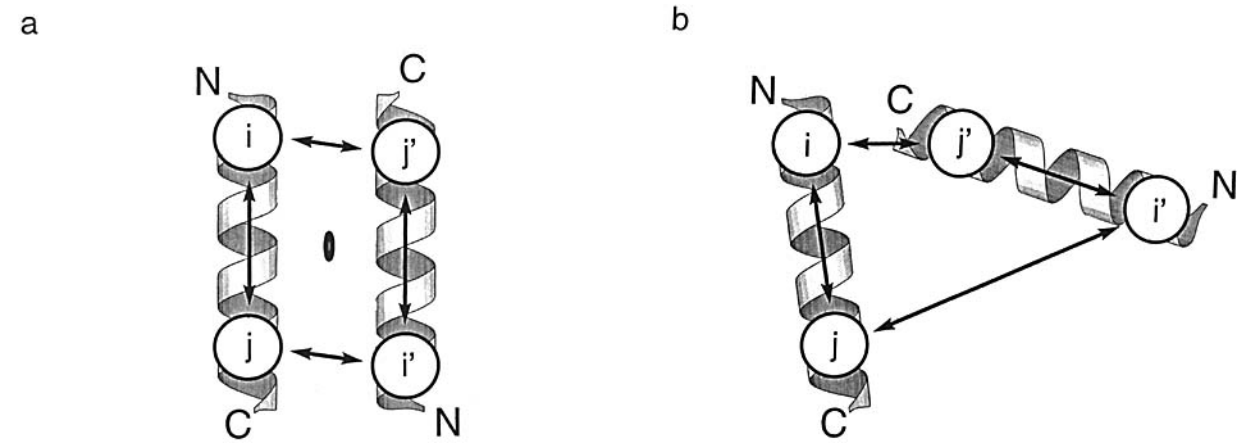

Fig. 1. (**a**) Symmetric dimer. The 2-fold symmetry axis of the dimer is indicated. The distances from *i* to *j'* and from *j* to *i'* are of equal size. (**b**) Configuration of identical monomers that does not possess a symmetry axis. The distances from *i* to *j'* and from *j* to *i'* are different.

crucial in the refinement protocols presented below. The fact that only one set of crosspeaks is observed in the NOE spectrum indicates that the molecule possesses a symmetry axis. Once the structure is known, the position of the symmetry axis can be determined. However, the calculation strategy presented in this paper does not assume the knowledge of the structure. In order to circumvent the problem of placing a symmetry axis in an unknown structure, a new symmetry restraining potential had to be devised.

In X-ray crystallography refinement, two methods of maintaining the symmetry in a multimeric molecule are in general use. The first, called "NCS restraints" in X-PLOR,[13] is a simple pseudo-energy term related to the RMS difference between the monomers. The second monomer is best fit[14] to the first, an average structure is calculated, and the monomers are then restrained to be similar to this average.[15] In other words, this pseudo-energy minimizes the atomic RMS difference between the monomers *without* taking the relative orientation of the monomers into account. In particular, the pseudo-energy term does *not* assume or impose a symmetry axis on the molecule. In fact, a conformation that perfectly satisfies this symmetry restraint alone will in general not possess a symmetry axis (Fig. 1).

The second option (called "NCS strict" in X-PLOR) assumes perfect symmetry. Only one monomer is simulated or refined, and the interaction with symmetry-related monomers is calculated by applying the appropriate symmetry operators. The position of the symmetry axis has to be defined.

For the present purpose, both these symmetry terms proved unsuitable. The "strict" symmetry option would require the knowledge of the symmetry axis in an unknown structure. Simply to impose it arbitrarily seemed inadvisable since it was expected that this would lead to convergence problems. On the other hand, initial calculations with the "NCS restraints" option alone showed that the calculated structures did not possess a symmetry axis, for the reasons outlined above.

The approach taken to ensure a symmetric molecule makes use of symmetry restraints on distances rather than coordinates, thus circumventing the problem of the placement of the symmetry axis. The difference between a particular distance and the symmetry-related distance is simply restrained to be 0. Since using $N(N-1)/2$ intramonomer and $N(N-1)/2$ intermonomer such distance pairs (N is the number of atoms in a monomer) is impractical, we have chosen to use a combination of the "NCS restraints" described above and the distance symmetry restraints. The former serve to minimize the atomic RMS difference between the two monomers, while the latter serves to arrange the monomers in a symmetric way. A small subset of all possible intermonomer distance pairs is sufficient. If the monomers were rigid bodies, restraints on 3 distance pairs would be enough. Different subsets were tried (each proton–proton distance for which there is a NOE, subsets of the C_α–C_α distances), with similar results. In the calculations reported below, only a small fraction of the possible distances was used, namely those between the C_α atom of residue i in monomer A to that of the C_α atom of residue $N_{\text{residue}} - i + 1$ in monomer B, where N_{residue} is the number of amino acid residues in a monomer.

Structure Determination Protocol

All calculations were performed with the X-ray and NMR refinement program X-PLOR version 3.0,[13] which had to be modified only slightly to incorporate the new symmetry term and the calculation of effective distances. The modified routines are part of the X-PLOR release 3.1, and the structure

TABLE I. Refinement Protocol for Symmetric Dimers

Stage	Conformational search	Cooling	Minimization
Temperature	2000 K	2000 K → 100 K	—
Masses*	100 a.m.u.	100 a.m.u.	—
Energy constants			
K_{bonds} [kcal/(mol Å^2)]	1,000	1,000	1,000
K_{angles} [kcal/(mol rad^2)]	125	125 → 500	500
K_{planar} [kcal/(mol rad^2)]	50	50 → 500	500
K_{repel} [kcal/(mol Å^4)]	0.003/0.1†	0.003 → 4	4
K_{NOE} [kcal/(mol Å^2)]	2	2 → 25	25
Asymptote	2	2	2
K_{NCS} [kcal/(mol Å^2)]	0.01	0.01 → 10	10
K_{symm} [kcal/(mol Å^2)]	1	1	1
Simulation time	≈ 50 psec	≈ 50 psec	250 steps

*Note that an increase in the mass has *exactly* the same effect as a reduction of the overall weight on the potential energy.
†For K_{repel} = 0.003 kcal/(mol Å^4) all nonbonded contacts were calculated in the conformational search phase (protocol 1), for K_{repel} = 0.1 kcal/(mol Å^4) the nonbonded interactions were computed only between C_α atoms (protocol 2).

determination protocols can be obtained on request from the author.

The general idea behind the protocols is similar to previously described protocols.[11,16] Local minima are avoided by a combination of a large reduction in the van der Waals interaction, allowing atoms to pass through each other,[17,18] the linear slope of the NOE restraint potential for large violations, and high temperature dynamics.

The special difficulties in the dimer calculations are a consequence of deep local minima in the NOE potential itself. These minima correspond to wrong "assignments" of intra- and intermonomer contributions to an NOE. The potential is strongly weighted toward the shorter distance. For example, the assignment of a particular NOE as intermonomer corresponds to the inter monomer distance being much shorter than the corresponding intramonomer distance. If this assignment is wrong, the intermonomer distance has to be increased and the intramonomer distance decreased. To achieve this, the minimization has to go "uphill" for some time, which is facilitated by the linear slope in the asymptotic region of the potential.

The first step of the calculation is the generation of a starting structure. If the fold of a homologous protein is known, the sequence of the new protein has to be mapped approximately onto the known structure. The side chains are built in arbitrary conformations, that is, without caring about van der Waals overlaps.

If the structure determination is to be attempted ab initio, ϕ and ψ angles are chosen randomly for a monomer, and the structure is subsequently simply duplicated: The two monomers are identical, and overlap completely in the starting conformation. This particular starting conformation is obviously symmetric, but the position and direction of the symmetry axis are not imposed. Also, this starting conformation does not bias a particular distance toward being intra- or intermonomer, as both distances are identical. If the intramonomer distance in the starting structure were larger, for example, the energy term would be biased toward the intermonomer distance.

The actual protocols are described in Table I. Phase I is a conformational search phase, where the temperature and the weights on the different terms in the target function are kept constant. Generally, most of the convergence toward the final structure is achieved during this phase. Two different methods of reducing the nonbonded interaction to allow the polypeptide chains to penetrate each other were used. In both, a "soft," repulsive van der Waals energy term replaces the Lennard–Jones energy term. In the first, the weight on the van der Waals interaction is substantially reduced to 0.003. In the second, the weight is set to a higher value (0.1), but nonbonded interactions are calculated only between C_α atoms whose radii are increased to 2.25 Å. This reduces the CPU time requirements of the conformational search phase. A similar method, which uses a two-atom representation for each residue to obtain approximate folds of proteins from NMR data, has been reported.[19]

Initial calculations had shown that a high weight on the "NCS restraints" term impeded convergence. To achieve convergence, the structures have to be allowed to deviate from symmetry during the conformational search phase. Some conformational energy constants (angles and planarity) are also softened during this phase, which has a similar effect as heating the system further. The NOE pseudo-potential changes from harmonic to the asymptotic form for deviations larger than 1.0 Å, with the slope of the asymptote set to 2.0.

In phase II, the system is cooled to 100 K, and all energy constants are slowly brought to their final

values. This phase is identical for both protocols, the van der Waals repulsion is always calculated for all atoms. Phase III consists of a short conjugate gradient minimization.

RESULTS

Calculations were carried out for three dimeric proteins. The first two test cases illustrate the situation which is similar to molecular replacement in X-ray crystallography, i.e., the approximate structure is known from a homologous protein or the same protein in a different environment. In the third example, the method is tested for the structure determination of a dimeric protein without assuming any prior knowledge of the structure.

Interleukin 8

This first example was chosen because of subtle but, in their total effect, striking differences between the X-ray and NMR structure.[2,20,21] The structure of the dimer consists of a six-stranded β-sheet, and two parallel α-helices that lie on top of the sheet. Part of the dimerization interface is made up by the central two stands of the sheet. The X-ray and NMR structures of the dimer differ most conspicuously by the distance between the two helices, which is about two Å larger in the NMR structure.[20,21]

The NMR structures and data were made available by G.M. Clore. The restrained minimized average structure (PDB[22] accession code 1IL8) is used as the reference NMR structure in the following. The data had to be modified for the calculations. The most important modification was the removal of all references to a specific monomer. For technical reasons, all pseudo-atoms were replaced by the nearest carbon atom (for methyl groups), or the restraints were entered in a form that avoids the use of pseudo-atoms.[23] This was necessary since in the present implementation it is not possible to use the distance averaging described in Eq. (6) and pseudo-atoms at the same time. (In X-PLOR, pseudo-atoms are not part of the standard residue library, but the average position of the equivalent or unassignable atoms is calculated at every minimization or dynamics step.)

The NOE distance data were complemented by restraints for those hydrogen bonds which had been used to determine the NMR structure of interleukin 8, and dihedral angle restraints derived from intraresidue and sequential NOEs and coupling constants. The use of the hydrogen bond restraints is justified, since all hydrogen bond restraints in the monomer–monomer interface were satisfied in the X-ray structure (data not shown).

Starting structure for all 10 calculations was the X-ray structure. Residues Ala-2 and Lys-3, which are missing in the X-ray structure but visible in the NMR spectra, were built in extended conformation, and hydrogens were added with the HBUILD command[24] in X-PLOR. Since the calculations started from the X-ray structure, the "assignments" of intra- and intermonomer NOEs were completely biased toward the X-ray structure. In the course of the refinement, 9 out of the 10 calculations resulted in structures without any NOE distance violations larger than 0.5 Å, and no dihedral angle violations larger than 5°.

All the structures converged toward the NMR structure, that is, they show a larger separation between the two helices than the X-ray structure.[20] Figure 2a and b shows an overlay of the converged structures with the NMR structure and the X-ray structure, respectively. The helix–helix separation in the converged structures is practically identical to that of the NMR structure, and different from that seen in the X-ray structure. Only small differences between the NMR structure and the nine converged structures can be found, which are located mostly in the surface loops (see Fig. 2) and not connected to the dimerization interface. These are due to the different treatment of the restraints involving methylene, propyl, and methyl groups, and differences in the nonbonded parameters used in versions 2.1 and 3.0 of X-PLOR.

The side chain conformation of the NMR structure and the average structure calculated from the nine converged structures is shown in Figure 3. For clarity, only one monomer and only side chains that are within 5 Å of the other monomer are shown. The positions of the side chains in the two structures are virtually identical.

RMS differences between the X-ray and the present converged structures are reported in Figure 4a and b in black lines. For comparison, the RMS differences between the NMR structure and the X-ray structure are shown as gray lines. The two lines agree very well. The largest RMS differences between X-ray and NMR structures involve exactly the same regions in the protein, namely, the region around residue 31 when fitted to a single monomer (Fig. 4a), and the C-terminal helix when fitted to the dimer (Fig. 4b).

A more stringent test is the comparison of the structures to the data with the original assignments.[2] In the 9 converged structures, no violations larger than 0.5 Å of the restraints using the original assignments are present. Using all restraint in the form of Eq. 1, the average RMS difference from experimental restraints is 0.027 Å. Using the original assignments for the analysis, the RMS difference is only slightly increased to 0.032 Å.

Metj Repressor

The choice of the second test case was motivated by the determination of the solution structure of the Arc repressor.[3] In this structure, and in the similar structure of the Metj repressor,[10] the two monomers are so intricately folded that it does not make sense

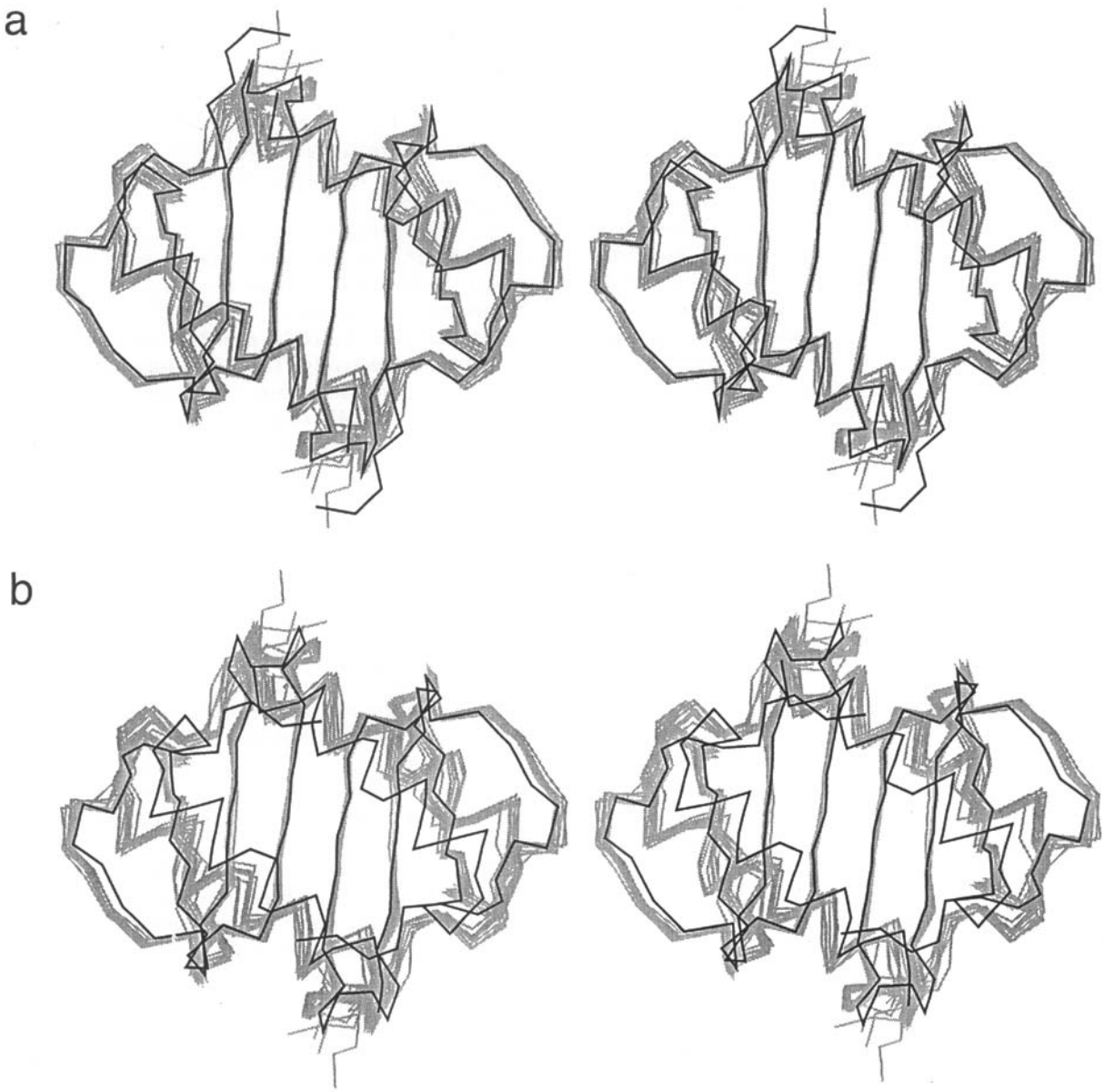

Fig. 2. Overlay of C_α traces the nine converged structures (gray lines) with the (**a**) NMR structure (black line) and (**b**) X-ray structure (black line).

to consider the monomers as independently folded "domains" of the dimeric protein. For the Arc repressor it has been shown by circular dichroism measurements that in fact dimerization, secondary and tertiary structure formation occur at the same time.[25,26] About half of the long-range NOEs in the Arc repressor were identified as inter monomer rather than intra monomer.[3] Part of the dimerization interface is made up by the N-terminal β-sheet (see Fig. 5a).

Derivation of a model data set

From the X-ray structure coordinates of the DNA binding domain of Metj that is structurally homologous to Arc (residues 15 to 72[3]), a distance set was derived that could reasonably be measured with an unlabeled sample. This included distances between α, amide, β, and distances to side chain protons that can generally be assigned (e.g., aromatic ring protons, isoleucines). All methylene, propyl, and methyl groups and equivalent protons on aromatic residues were treated as if they were unresolved. In particular, no stereospecific assignments or assignments of equivalent aromatic protons were assumed. For each pair of protons or unresolved groups of protons, the weighted sum described in Eq. (6) was evaluated. If the resulting effective distance was below 4.0 Å, it was classified into weak (< 4.0 Å), medium

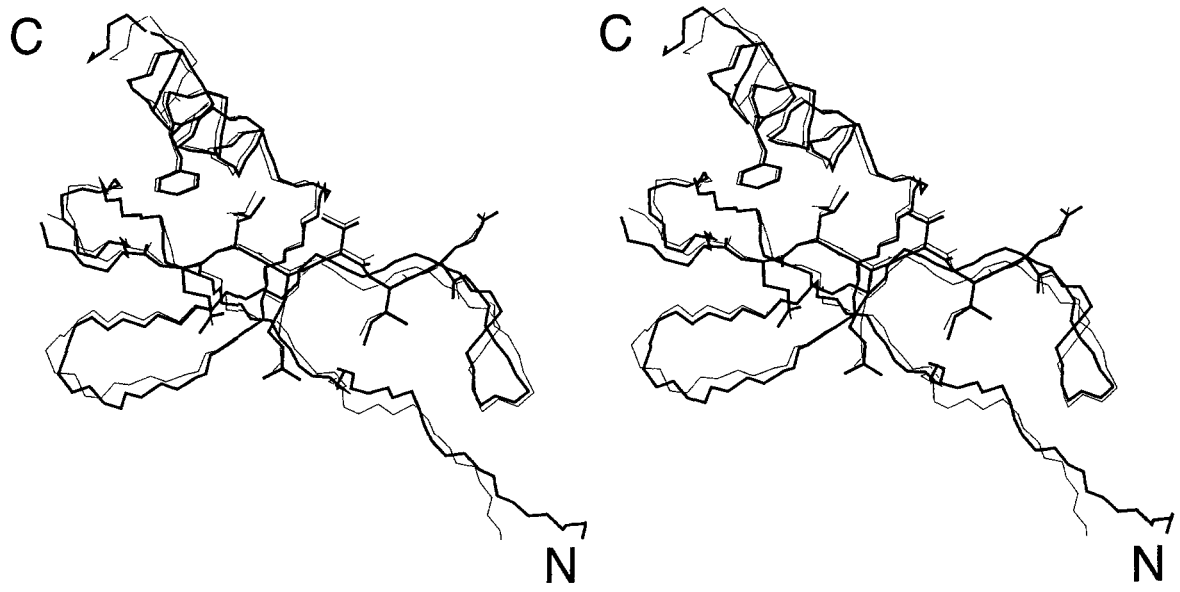

Fig. 3. Overlay of the NMR structure (thin line) with the average structure calculated from the ensemble of nine converged structures (thick line). The C_α trace of only one monomer and only side chains within 5 Å of the other monomer are shown.

(< 3.2 Å), or strong (< 2.5 Å), and the restraints were set to 1.8–4.5 , 1.8–3.5, or 1.8–2.7 Å, respectively. No distinction between inter- or intramonomer contribution was made. No intraresidue restraints were used. A total of 682 ambiguous restraints were generated in this way. A similar number of interresidue restraints were used in the NMR study of the Arc repressor.[3]

Initial structures

To mimic a real "molecular replacement" situation, where the homologous structure might differ systematically from the true structure of the molecule being refined, a series of starting structures was derived by distorting the X-ray structure. The first kind of distortion was brought about by a rotation of the secondary structure elements (the four helices in the dimer, and the antiparallel β-sheet between the N-terminal regions of the two monomers) by multiples of $\pi/10$. To this end, all side chain atoms were removed, and the backbone atoms alone were rotated. The resulting backbone structure was minimized by a short molecular dynamics run and conjugate gradient minimization, to remove all strain in the conformation. During minimization and dynamics, the rotated secondary structure elements were held in their new place by harmonic positional restraints,[27] while the pieces connecting them were allowed to move freely, and all nonbonded interactions were switched off. The side chains were then rebuilt in extended conformation regardless of packing considerations. One reason for choosing these particular distortions of the X-ray

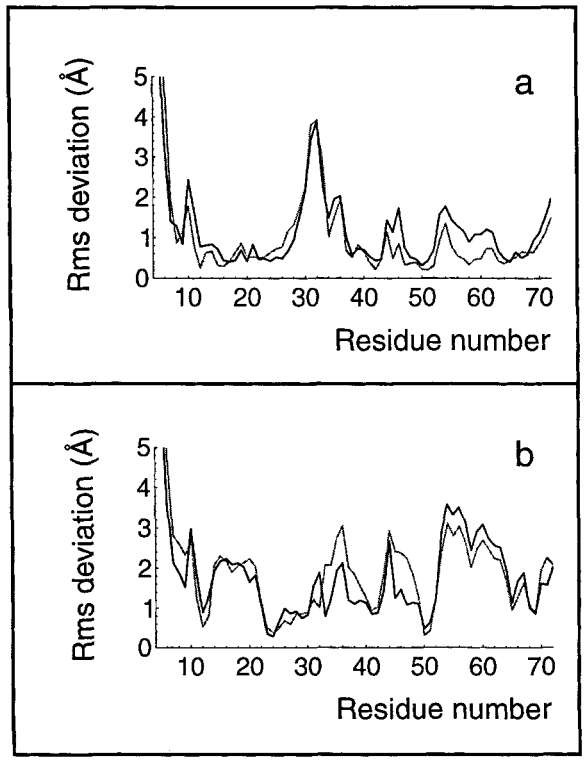

Fig. 4. Average atomic RMS differences between the nine converged structures and the X-ray structure (black line), and the NMR 1IL8 and the X-ray structure (gray line), for backbone atoms (N, C_α, C). (**a**) Best fit to monomer. (**b**) Best fit to dimer.

structure was the rotation of helix B of the Arc repressor with respect to its orientation in the crystal structure of the Metj repressor.[3]

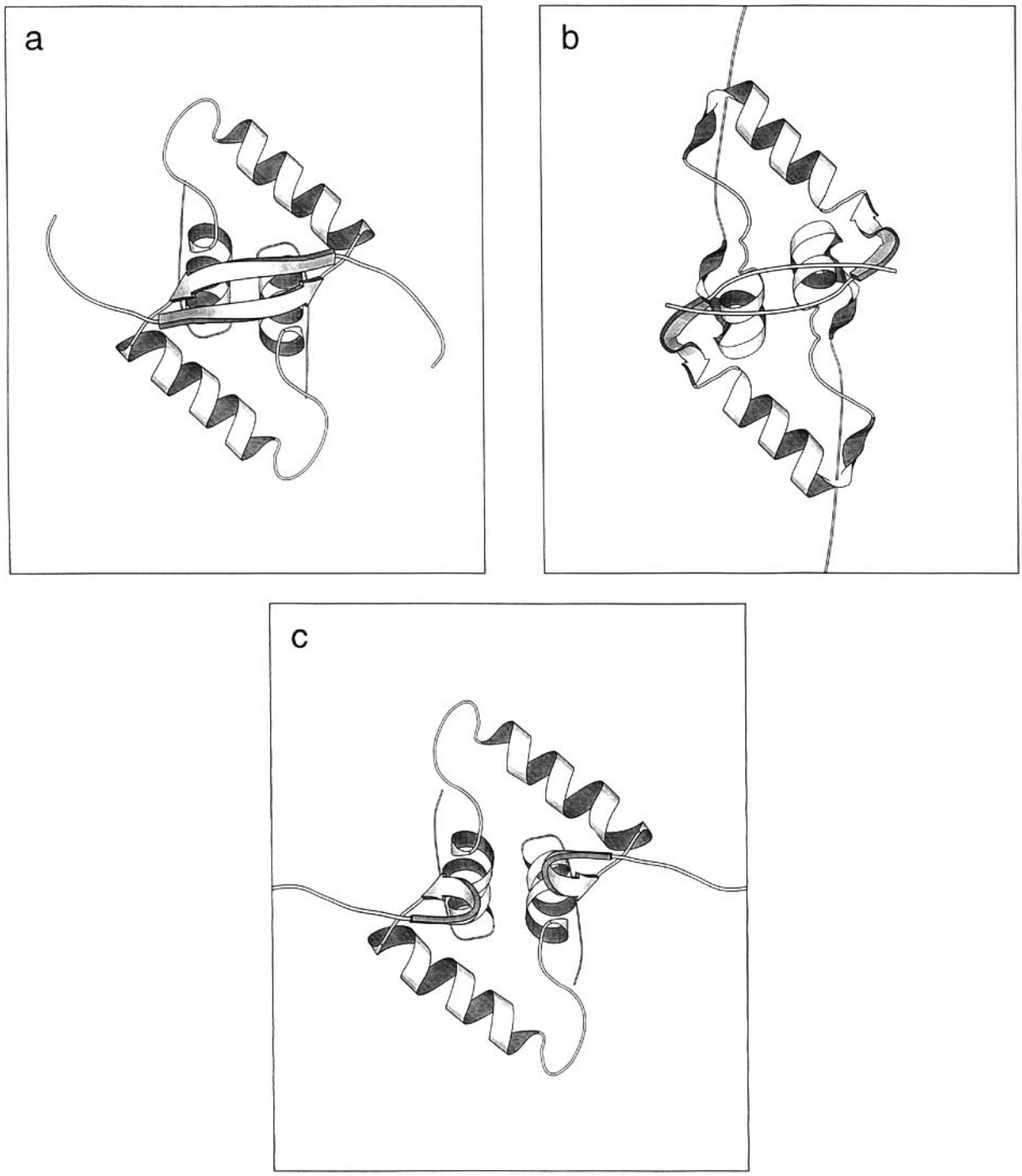

Fig. 5. MOLSCRIPT[32] plots of Metj DNA binding domain structures. (**a**) The X-ray structure.[10] (**b**) Initial structure for the calculation, shifted by six residues with respect to the X-ray structure. (**c**) Tight turns instead of the antiparallel β-sheet between the N-terminal regions of the protein.

The second kind of distortion was achieved by shifting the X-ray coordinates by one or several residues, such that the initial backbone coordinates of residues 20, for example, are those of residue 21. The missing pieces of backbone at the C-terminal end were built in extended conformation, and side chains were rebuilt in extended conformation as above. This particular distortion of the initial structure seemed of interest because of a systematic shift in the secondary structure assignment in the X-ray structure of the product of gene 5 of phage M13,[28] and the secondary structure apparent in the NMR

TABLE II. Results for the Metj Dimerization Domain, Rotation*

	Protocol 1	
Distortion	RMS_{NOE}, N_{viol}	$RMS_{X\text{-}ray}$
36°	0.01, 0	1.17
72°	0.01, 0	2.24
108°	0.01, 0	1.19
144°	0.08, 3	7.23
180°	0.04, 1	7.22
216°	0.01, 0	1.07
252°	0.01, 0	1.19
288°	0.01, 0	1.14
324°	0.01, 0	1.01
360°	0.01, 0	1.45

*See Table III footnotes for explanation of terms.

spectrum.[29] Figure 5b shows the structure with a shift of six residues with the secondary structure assignment of the undistorted X-ray structure.

Calculations and results

Tables II and III give a summary of the calculations performed, and the results. Most of the calculations starting from the structures generated by rotation of secondary structure elements converged to the correct structure. Only the two calculations with the largest rotations (144° and 180° for the four helices and the β-sheet) did not converge.

Convergence from the "shifted" starting structures is more difficult to achieve. This is partly due to the fact that this type of distortion is more severe than a simple rotation—a shift by two residues corresponds already to a rotation of the helices around 200° and a shift of the helix along its long axis by 3 Å. The radius of convergence of the present method is around 6 residues. Protocol 2, where the nonbonded contacts in the high temperature stage are only calculated between C_α atoms (see Materials and Methods), seems to perform slightly better than protocol 1. Figure 5b shows that a shift larger than six residues would completely disrupt the N-terminal β-sheet. This β-sheet poses a special difficulty, as figure 5c illustrates. Instead of the anti-parallel β-sheet, two independent tight turns have formed (i.e., inter monomer NOEs are falsely taken as intra monomer). The chance that a particular NOE is misinterpreted is high, since the energetic difference between the two conformations is small, and the tight turn conformation is more likely to be found as the search method works essentially in torsion angle space. However, this misfolded structure can be rejected on the basis of the worse overall fit to data (Table III).

Troponin C Ca^{2+} Binding Domain

In the third test case, the method is used to calculate a dimeric structure ab initio, without assuming any knowledge of the approximate fold of the protein. The peptide comprising the fourth Ca^{2+}-binding domain of troponin C forms homodimers in solution.[7] The solution structure[7] is shown in Figure 6a. The structure consists of a dimer of helix–loop–helix motifs and is similar to the arrangement of the Ca^{2+}-binding domains III and IV in the X-ray structure.[30]

The starting structures in this case were generated by assigning the backbone torsion angles randomly and overlaying the two monomers as described in Materials and Methods. A typical starting structure is shown in Figure 6b. For clarity, the secondary structure assignments of the final structure

TABLE III. Results for the Metj Dimerization Domain, Shift

	Protocol 1†		Protocol 2†	
Distortion*	RMS_{NOE},‡ N_{viol}§	$RMS_{X\text{-}ray}$**	RMS_{NOE},‡, N_{viol}§	$RMS_{X\text{-}ray}$**
1	0.01, 0	1.16	0.02, 0	1.32
2	0.01, 0	1.09	0.01, 0	1.18
3	0.01, 0	1.03	0.09, 5	2.55
4	0.10, 4	7.23	0.01, 0	1.00
5	0.17, 16	10.25	0.09, 6	12.10
6	0.19, 16	10.20	0.01, 0	1.81
7	0.17, 10	9.47	0.07, 2	10.10
8	0.12, 9	9.82	0.08, 6	12.27
9	0.04, 1	8.77	0.03, 1	8.56
10	0.15, 6	17.10	0.11, 6	14.16

*Number of residues by which the structure is shifted.
†See Table I.
‡RMS difference of effective interproton distances in the final structure [Eq. (6)] from the restraints.
§Number of violations of the distance restraints larger than 0.5 Å.
**Atomic RMS difference from the X-ray structure for backbone atoms (N, C_α, C) for residues 19 to 72.

TABLE IV. Troponin C Ca^{2+}-Binding Domain IV Dimer C_α RMS Differences

	NMR	S1	S2	S3	S4	S5	S6	S7	S8	X-ray*
NMR	—	1.2	1.9	2.2	2.6	2.1	2.1	1.8	3.7	2.2
S1	1.2	—	2.3	1.7	3.1	2.5	2.6	2.3	4.2	2.5
S2	2.3	2.7	—	3.6	1.8	0.9	1.4	1.2	2.5	1.2
S3	2.5	2.0	3.9	—	4.3	3.8	3.9	3.5	5.5	3.8
S4	2.8	3.4	2.0	4.7	—	1.3	1.1	1.6	1.4	1.8
S5	2.4	2.9	1.0	4.1	1.7	—	1.0	0.9	2.2	1.2
S6	2.2	2.8	1.9	4.2	1.2	1.6	—	1.4	1.9	1.3
S7	2.3	2.8	1.3	3.9	1.8	1.1	2.0	—	2.6	1.5
S8	4.1	4.6	2.6	5.8	1.7	2.3	2.4	2.6	—	2.3
X-ray	2.6	3.0	1.8	4.1	2.0	2.0	1.9	1.8	2.8	—

*X-Ray crystal structure coordinates[30] of residues Glu-95 to Arg-123 and Glu-131 to Glu-159, corresponding to the Ca^{2+}-binding domains III and IV. Values above the diagonal are RMS differences for monomer B and Ca^{2+} binding domain IV of the X-ray structure; values below the diagonal are for the dimer and the binding domains III and IV of the X-ray structure. All values are in Å.

were used to generate the ribbon drawings, and one of the monomers is slightly displaced.

The experimental data were made available by L. Kay. As in the case of interleukin 8 described above, all NOE restraints were entered as effective distances of the form described in Eqs. (1) and (6), and all references in the NOE restraint table to a specific monomer were removed, with the exception of the distances to the bound Ca^{2+} ions. The NOE restraints were complemented by torsion angle restraints.[7]

Apart from the generation of the initial structures, the calculations proceeded essentially as in the other two test cases. Since some of the structures that had converged to the correct topology after one pass of the protocol showed violations of the experimental data, the cooling phase of the protocol was repeated twice on all structures. Note that the initial random conformation is parallel (Fig. 6b), while the final conformation of the dimer is antiparallel (Fig. 6a). One monomer has to invert during the calculation.

Of 50 final structures 11 converged to essentially the fold reported originally,[7] 8 out of these show no violations of the experimental NOE restraints larger than 0.5 Å. The RMS differences from the experimental bound averaged to 0.03 Å for the NOE distance restraints and 0.08° for the dihedral angle restraints. Table IV gives a summary over the C_α RMS differences between the eight converged structures, the NMR structure[7] and Ca^{2+}-binding domains III and IV of the X-ray structure.[30] The RMS differences between structures are sometimes considerably larger than the 1.04 Å RMS difference from the average structure found in the NMR study.[7] This increased conformational flexibility is a consequence of the ambiguous distance restraints used in the present calculations. Only one of the eight converged structures converged to exactly the same conformation (structure S1 in Table IV). Figure 7 shows a comparison of this structure with the average NMR structure.[7] The C_α trace and the side chains of residues Leu-134 and Met-154 are depicted.

The larger RMS differences for the other six structures are a consequence of different "assignments" for some restraints. The most important difference involves the NOE between Leu-134 and Met-154 that was assigned as intermonomer.[7] In the six structures that show a larger difference from the original NMR structure this NOE is assigned as intramonomer in the course of the calculation. This assignment is not obvious, considering the large distance between the two involved residues in the polypeptide chain. It should be emphasized, however, that this assignment is equally plausible as the original assignment, since the structures satisfy the data equally well. In this context it is noteworthy that five out of the six structures (S2, S4, S5, S6, S7 in Table IV) are closer to the conformation found in the X-ray structure[30] than the structures with the NOE between Leu-134 and Met-154 assigned as intermonomer. Structure S2 is shown in Figure 7b. The consequence of the different assignment is a reorientation of the involved side chains and a closer packing of the two helices in the monomer. Structure S8 shows an even closer intramonomer packing of the helices than the X-ray structure, while structure S3 is somewhat more extended than the NMR structure. On the basis of the data *alone*, it seems that all these solutions are equally possible. Other criteria may lead to the rejection of some of the structures, like the prediction of NOEs from a structure that are not observed experimentally.

Most of the incorrectly folded final structures were conformations with separated monomers. Practically all NOEs were interpreted as intramonomer, and consequently the two helices in the monomer pack against each other instead of the equivalent helix of the other monomer (see Fig. 6c). The resulting structures "almost" satisfy the data, with RMS deviations from the distance restraints only about

two to three times as high as the converged dimeric structures (0.06–0.08 Å), and very few violations of distance restraints larger than 0.5 Å. The separated monomers correspond to a deep local minimum of the target function (see Fig. 6c). It is important to note that there is, however, a clear distinction between the NOE energies of the converged dimeric structures, and those of the lowest energy monomeric structures, so that it would have been possible to reject the monomeric structures.

CONCLUSIONS

A method is presented that circumvents the problem of assignment of inter- and intramonomer NOEs in symmetric dimers by using the appropriate expression for an NOE-derived effective distance in the dimer. It is shown that, combined with the symmetry properties of the molecule, the information content in the spectrum can be sufficient to determine the polypeptide fold of the monomers, and the relative position of the two monomers in the dimer. The strategy performs well if an approximate model is known, and also for ab initio structure determinations in cases where the fold of the dimer is not too complicated. The method finds alternate possibilities of assignments.

For the purpose of these test calculations, all NOEs were treated as ambiguous, while in practice it will be possible to assign certain NOE peaks on the basis of the spectra and the sequential assignments, like sequential peaks and those defining secondary structure. It is obvious that this would increase the radius of convergence of the method.

Special difficulties for the method are connected to residues close to the dyad axis of the dimer. Intermonomer peaks between equivalent residues are indistinguishable from intraresidue peaks; intermonomer peaks between equivalent protons on different monomers are on the diagonal and thus unmeasurable. Thus, dimers with the dyad axis running along the interface, like parallel leucine zippers,[4,5] are more difficult to investigate. The difficulties encountered with the N-terminus of the Metj domain indicate that a minimization strategy essentially operating in torsion angle space is not necessarily optimal. A new protocol is under development which random starting structures in cartesian coordinate space.[31]

A calculational method can in general not provide the same amount of information as an additional

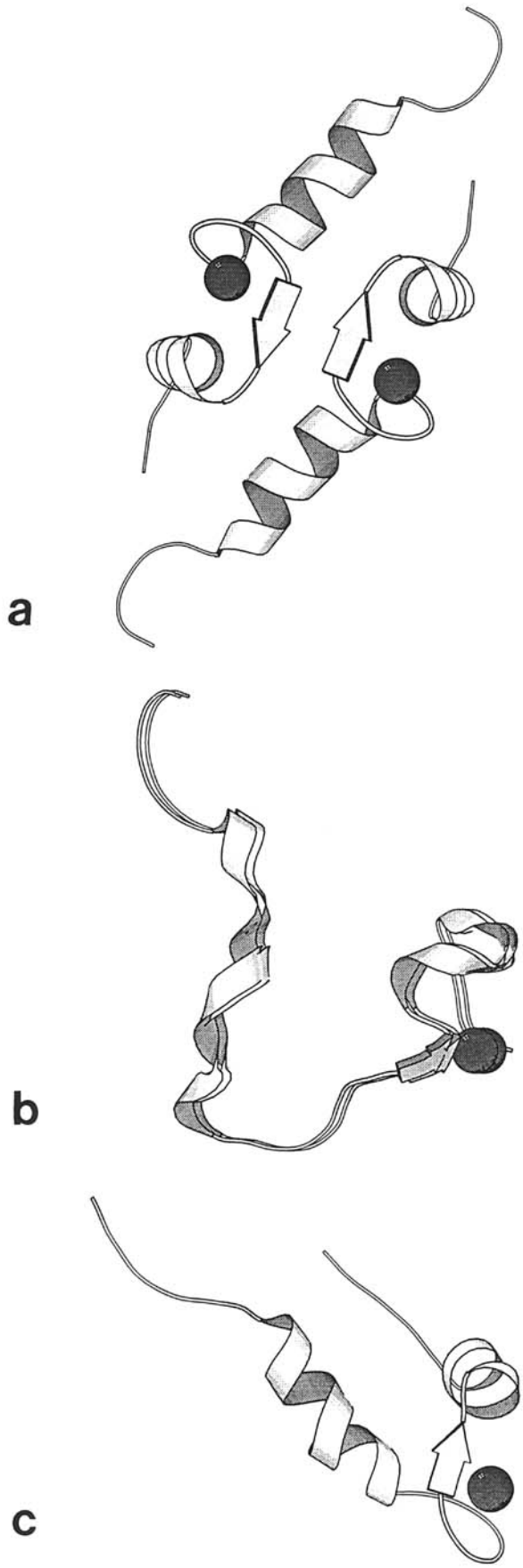

Fig. 6. MOLSCRIPT[32] plots of troponin C Ca^{2+}-binding domain IV structures. (**a**) The NMR structure of the troponin C Ca^{2+}-binding domain IV dimer.[7] (**b**) A typical random initial structure for the tropinin C Ca^{2+}-binding domain. The two monomers coincide completely. For clarity, one of the monomers is displaced slightly in the figure. (**c**) The monomers have separated during the calculation. This conformation does not satisfy the data as well as the dimeric structures, but is a deep local minimum.

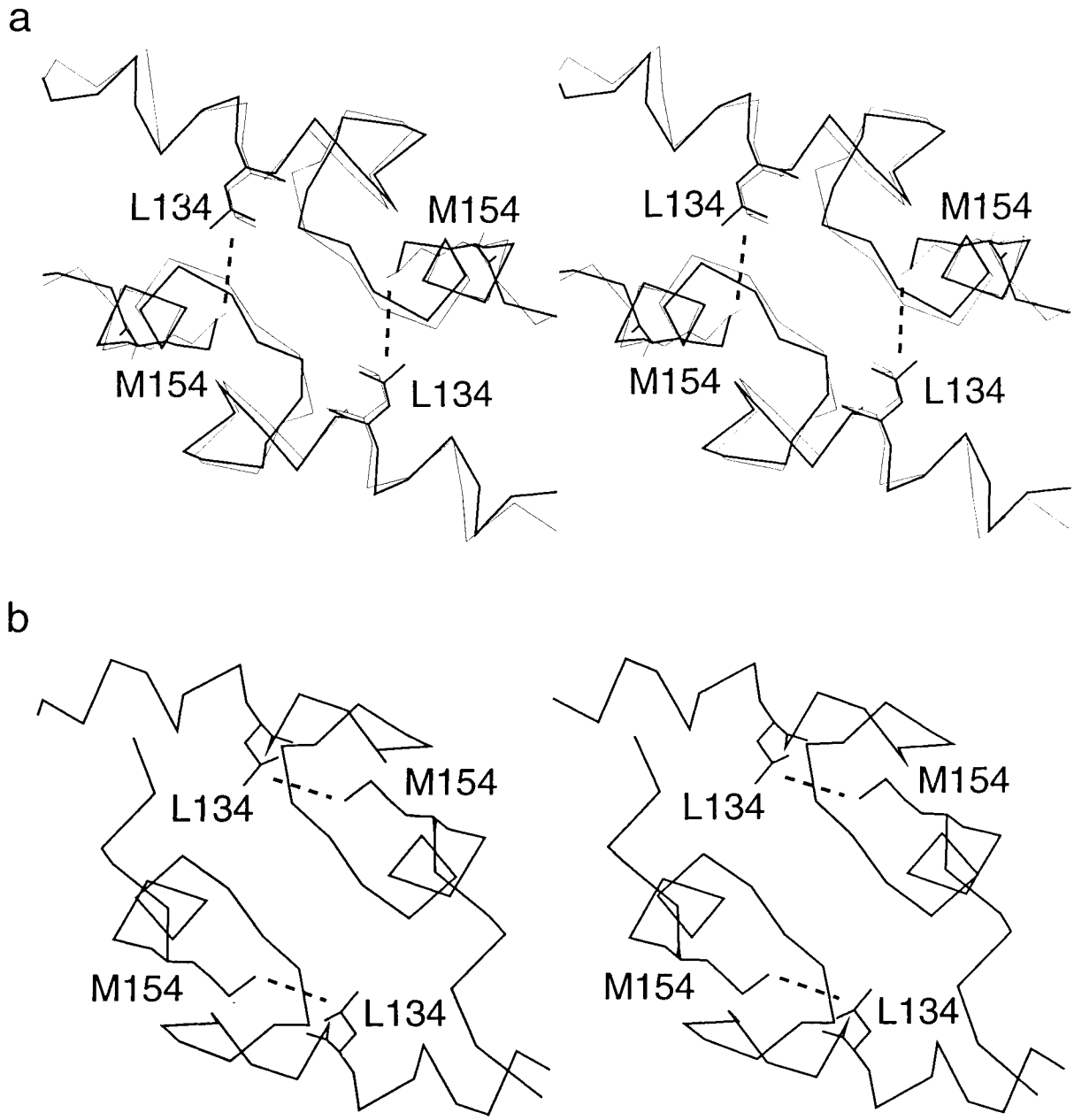

Fig. 7. Troponin C Ca^{2+}-binding domain IV dimerization interface. Only the C_α traces of the side chains of residues Leu-134 and Met-154 in the interface are shown. (**a**) The average NMR structure[7] is shown in thin lines, structures that converged to the same fold in thick lines. The NOE between Leu-134 and Met-154 is *inter*monomer. (**b**) The conformation as observed in five structures of the present calculation. The NOE between Leu-134 and Met-154 is *intra*monomer.

experiment, like an asymmetric labeling experiment.[1,8,9] Calculations can be helpful to decide if an asymmetric labeling experiment is really necessary to define the structure, and *which* residues have to be labeled if one uses an amino acid specific labeling method. However, quantitation of crosspeaks is difficult in asymmetric labeling experiments due to the lower signal-to-noise ratio in difference methods. In particular, it is difficult to exclude an intramonomer contribution to a crosspeak by means of these experiments. A conservative approach would be the combined use of *qualitative* results from an asymmetric labeling experiment and *quantitative* distance measurements from a normal NOE experiment, used in the form of Eq. (6) with the calculational method presented in this paper.

Generalizations of the method described in this paper are possible. In principle, the calculational method can be easily extended to higher order symmetric multimers, where the interpretation of asym-

metric labeling experiments will be increasingly difficult. The analysis described is also applicable to more general cases of ambiguous NOEs. Work in this direction is in progress.

ACKNOWLEDGMENTS

Initial work on this project was carried out while the author was a Howard Hughes postdoctoral fellow with A.T. Brünger at Yale University, who he thanks for continuing support and very helpful discussions. The author thanks L. Kay for the NMR data and coordinates of the troponin C Ca^{2+}-binding domain, S.E.V. Phillips for the X-ray coordinates of the Metj repressor, A. Wlodawer for the X-ray coordinates of interleukin 8, G.M. Clore for the NMR data and NMR coordinates of interleukin 8, A. Krukowsky for performing initial calculations on interleukin 8, and P. Folkers, R. Folmer, and A. Pastore for helpful discussions and a critical reading of the manuscript.

REFERENCES

1. Arrowsmith, C.H., Pachter, R., Altman, R.B., Iyer, S.B., Jardetzky, O. Sequence-specific ^{1}H NMR assignments and secondary structure in solution of *Escherichia coli trp* repressor. Biochemistry 29:6332–6341, 1990.
2. Clore, G.M., Appella, E., Yamada, M, Matsushima, H., Gronenborn, A.M. Three-dimensional structure of interleukin-8 in solution. Biochemistry 29:1689–1701, 1990.
3. Breg, J.N., van Opheusden, J.H.J., Burgering, M.J.M., Boelens, R., Kaptein, R. Structure of Arc repressor in solution: evidence for a family of β-sheet DNA binding proteins. Nature (London) 346:586–589, 1990.
4. Saudek, V., Pastore, A., Castiglione Morelli, M.A., Frank, R., Gausepohl, H., Gibson, T., Weih, F., Roesch, P. Solution structure of the DNA-binding domain of the yeast transcriptional activator protein GCN4. Protein Eng. 4:3–10, 1990.
5. Saudek, V., Pastore, A., Castiglione Morelli, M.A., Frank, R., Gausepohl, H., Gibson, T. The solution structure of a leucine zipper motif peptide. Protein Eng. 4:519–529, 1991.
6. Shaw, G.S., Hodges, R.S., Sykes, B.D. Calcium-induced peptide association to form an intact protein domain: ^{1}H NMR structural evidence. Science 249:280–283, 1990.
7. Kay, L.E., Forman-Kay, J.D., McCubbin, W.D., Kay, C.M. Solution structure of a polypeptide dimer comprising the fourth Ca(2+)-binding site of troponin C by nuclear magnetic resonance spectroscopy. Biochemistry 30:4323–4333, 1991.
8. Weiss, M.A. Distinguishing symmetry-related intramolecular and intermolecular nuclear Overhauser effects in a protein by asymmetric isotopic labelling. J. Magn. Reson. 86:626–632, 1990.
9. Folkers, P.J.M., Folmer, R.H.A., Konings, R.N.H., Hilbers, C.W. Overcoming the ambiguity problem encountered in the analysis of nuclear Overhauser magnetic resonance spectra of symmetric dimer proteins. J. Am. Chem. Soc., 115:3798–3799, 1993.
10. Rafferty, J.B., Somers, W.S., Saint-Girons, I., Phillips, S.E.V. Three-dimensional crystal structures of *Escherichia coli met* repressor with and without corepressor. Nature (London) 341:705–710, 1989.
11. Nilges, M., Gronenborn, A.M., Brünger, A.T., Clore, G.M. Determination of three-dimensional structures of proteins by simulated annealing with interproton distance restraints: Application to crambin, potato carboxypeptidase inhibitor and barley serine proteinase inhibitor 2. Protein Eng. 2:27–38, 1988.
12. Brünger, A.T., Clore, G.M., Gronenborn, A.M., Karplus, M. Three-dimensional structure of proteins determined by molecular dynamics with interproton distance restraints: Application to crambin. Proc. Natl. Acad. Sci. U.S.A. 83: 3801–3805, 1986.
13. Brünger, A.T. "X-PLOR. A System for X-ray Crystallography and NMR." New Haven: Yale University Press, 1992.
14. Kabsch, W. A solution for the best rotation to relate two sets of vectors. Acta Crystallogr. A32:922–923, 1976.
15. Konnert, J.H., Hendrickson, W.A. A restrained-parameter thermal-factor refinement procedure. Acta Crystallogr. A36:344–349, 1980.
16. Nilges, M., Kuszewski, J., Brünger, A.T. Sampling properties of simulated annealing and distance geometry. In "Computational Aspects of the Study of Biological Macromolecules by Nuclear Magnetic Resonance Spectroscopy." Hoch, J.C., Poulsen, F.M., Redfield, C., eds. New York: Plenum Press, 1991: 451–455.
17. Levitt, M. Protein folding by restrained energy minimization and molecular dynamics. J. Mol. Biol. 170:723–764, 1983.
18. Nilges, M., Clore, G.M., Gronenborn, A.M. Determination of three-dimensional structures of proteins from interproton distance data by hybrid distance geometry-dynamical simulated annealing calculations. FEBS Lett. 229:317–324, 1988.
19. Hoch, J.C., Stern, A.S. A method for determining overall protein fold from NMR distance restraints. J. Biomol. NMR 2:535–543, 1992.
20. Baldwin, E.T., Weber, I.T., St. Charles, R., Xuan, J.-C., Appella, E., Yamada, M., Matsushima, K., Edwards, B.F.P., Clore, G.M., Gronenborn, A.M., Wlodawer, A. Crystal structure of interleukin 8: Symbiosis of NMR and crystallography. Proc. Natl. Acad. Sci. U.S.A. 88:502–506, 1991.
21. Clore, G.M., Gronenborn, A.M. Comparison of the solution nuclear magnetic resonance and crystal structures of interleukin-8. Possible implications for the mechanism of receptor binding. J. Mol. Biol. 217:611–620, 1991.
22. Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Jr., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T., Tasumi, M. The Protein Data Bank: A computer-based archival file for macomolecular structures. J. Mol. Biol. 112:535–542, 1977.
23. Pardi, A., Hare, D.R., Selsted, M.E., Morrison, R.D., Bassolino, D.A., Bach, A.C. Solution structures of the rabbit neutophil defensin NP-5. J. Mol. Biol. 201:625–636, 1988.
24. Brünger, A.T., Karplus, M. Polar hydrogen positions in proteins: Empirical energy placement and neutron diffraction comparison. Proteins 4:148–156, 1988.
25. Bowie, J.U., Sauer, R.T. Equilibrium dissociation and unfolding of the Arc repressor dimer. Biochemistry 28:7139–7143, 1989.
26. Knight, K.L., Bowie, J.U., Vershon, A.K., Kelley, R.D., Sauer, R.T. The Arc and Mnt repressors. J. Biol. Chem. 364:3639–3642, 1989.
27. Bruccoleri, R.E., Karplus, M. Spatially constrained minimization of macromolecules. J. Comput. Chem. 7:165–175, 1986.
28. Brayer, G.D., McPherson, A. Refined structure of the gene 5 dna binding protein from bacteriophage fd. J. Mol. Biol. 169:565–596, 1983.
29. Folkers, P.J.M., van Duynhoven, J.P.M., Jonker, A.J., Harmsen, B.J.M., Konings, R.N.H., Hilbers, C.W. Sequence-specific ^{1}H-NMR assignment and secondary structure of the Tyr41 → His mutant of the single-stranded DNA binding protein, gene V protein, encoded by the filamentous bacteriophage M13. Eur. J. Biochem. 202:349–360, 1991.
30. Sundaralingam, M., Bergstrom, R., Strasberg, G., Rao, S.T., Roychowdhury, P., Greaser, M., Wang, B.C. Molecular structure of Troponin C from chicken skeletal muscle at 3-Angstrom resolution. Science 227:945–948, 1985.
31. Nilges, M., Clore, G.M., Gronenborn, A.M. Determination of three-dimensional structures of proteins from interproton distance data by dynamical simulated annealing from a random array of atoms. FEBS Lett. 239:129–136, 1988.
32. Kraulis, P. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. J. Appl. Crystallogr. 24:46–950, 1991.