Journal of Biomolecular NMR, 1 (1991) 3-11 ESCOM

J-Bio NMR 009

Multi-conformational peptide dynamics derived from NMR data: A new search algorithm and its application to antamanide

R. Brüschweiler, M. Blackledge and R.R. Ernst

Laboratorium für Physikalische Chemie, Eidgenössische Technische Hochschule, CH-8092 Zürich, Switzerland

Received 24 January 1991 Accepted 4 February 1991

Keywords: Molecular dynamics; Biomolecular structure; Conformational search; Multiple conformations; Antamanide; NMR; T_{1e} relaxation

SUMMARY

A search algorithm, called MEDUSA, is presented which allows the determination of multiple conformations of biomolecules in solution with exchange rate constants typically between 10^3 and 10^7 s⁻¹ on the basis of experimental high-resolution NMR data. Multiples of structures are generated which are consistent as ensembles with NMR cross-relaxation rates (NOESY, ROESY), scalar J-coupling constants, and $T_{1\rho}$ measurements. The algorithm is applied to the cyclic decapeptide antamanide dissolved in chloroform. The characteristic radio-frequency field dependence of the $T_{1\rho}$ relaxation rates found for the NH protons of Val¹ and Phe⁶ can be explained by a dynamical exchange between two structures.

INTRODUCTION

Most proteins and peptides are flexible, with a dynamic equilibrium between various local energy minima, sometimes called conformational substates (Austin et al., 1975). It is known that conformational flexibility is of importance for biological function and for the folding process that proceeds through numerous conformations spanning a relatively large phase space volume (Kim and Baldwin, 1990; Wright et al., 1988). Obtaining information on transient conformations is therefore of great interest. Numerous experimental approaches have been proposed, such as using the temperature dependence of B factors from X-ray diffraction (Petsko and Ringe, 1984; Ringe and Petsko, 1985), quasi-elastic neutron scattering (Cusack, 1989), fluorescence depolarization (Beecham and Brand, 1984), ultrasonic absorption measurements (Slutsky et al., 1989) and NMR studies (Lipari and Szabo, 1982; Kopple et al., 1988; Nusser et al., 1988; Clore et al., 1990). All

0925-2738/\$ 5.00 © 1991 ESCOM Science Publishers B.V.

these techniques have some virtues and many limitations. X-ray diffraction, although informative about local disorder, does not provide rate constants nor insight into correlated motional processes, and the relevance of the results is restricted by the single-crystal environment. Quasi-elastic neutron scattering and ultrasonic absorption allow the measurement of correlation functions of the motional processes but provide no local information.

NMR has some of the most attractive features. It uses local sensors that are sensitive to motion over a wide time scale (Abragam, 1961; Wüthrich, 1986; Ernst et al., 1987). Slow dynamics can be followed in real time after an initial perturbation of the chemical equilibrium, possibly combined with a pulse labelling technique (Roder and Wüthrich, 1986; Roder et al., 1988; Udgaonkar and Baldwin, 1988). Conformational processes with exchange lifetimes τ_e in the range $10^{-4} \leq \tau_e \leq$ 10^{-1} s can be investigated by line shape studies. Faster intramolecular processes may be monitored through relaxation measurements. The effects on relaxation are dependent on the ratio τ_e/τ_c with the overall molecular tumbling correlation time τ_c . For $\tau_e/\tau_c >> 1$, population-weighted conformationally averaged relaxation and cross-relaxation rates are observed. For $0.1 \leq \tau_e/\tau_e \leq 10$, intramolecular dynamics becomes directly relaxation-active, while for $\tau_e/\tau_c << 1$ conformational averaging of the molecular structural parameters occurs that depends on the details of the motional process (Brüschweiler et al., to be published). In addition, it is possible to sense, by rotating frame relaxation measurements (Jones, 1966), processes that modulate the chemical shift in the range $10^{-6} \leq \tau_e \leq 10^{-3}$ s (Deverell et al., 1970).

In the standard procedure of NMR structure determination, NOESY or ROESY (Jeener et al., 1979; Kumar et al., 1981; Bothner-By et al., 1984) cross-peak intensities are translated into distances r_{kl}^{NOE} between nuclei k and / (Wagner and Wüthrich, 1982). In a mono-conformational system, the measured r_{kl}^{NOE} can be used in a distance geometry (DG) (Havel and Wüthrich, 1984; Braun and Go, 1985; Crippen and Havel, 1988) or restrained molecular dynamics (rMD) algorithm (Clore et al., 1985; Kaptein et al., 1985) to determine the three-dimensional biomolecular structure. When it turns out to be impossible to find a structure consistent with all measured NOE/ROEs and the holonomic structural data (bond lengths and bond angles), it is likely that several dynamically interchanging conformations are involved. On the other hand, the finding of a structure which fulfills all constraints does not prove the absence of multiple conformations. An approach to apply rMD also for dynamic structures has recently been proposed by Torda et al. (1990). It uses a time-variable NOE-restrained potential that depends on the past of the trajectory.

This contribution describes a novel algorithm for processing NMR information that allows the extraction of multiple conformations for exchange processes slow compared to the molecular tumbling correlation time τ_c , i.e. $\tau_c/\tau_c >> 1$, where population-averaged cross-relaxation rates are observed.

SEARCH PROTOCOL

The search protocol is called MEDUSA standing for Multiconformational Evaluation of Distance information Using a Stochastically constrained minimization Algorithm.

It interprets experimental cross-relaxation data in the following way:

a) The detection of an NOE/ROE yielding a distance r_{kl}^{NOE} implies that there is at least one signifi-

cantly populated conformation, in which the proton pair (k, l) has a distance $r_{kl} < r_{kl}^{NOE}$. A distance constraint defined in this sense will be abbreviated by DC.

b) The absence of an NOE/ROE means that there is no significantly populated conformation with $r_{kl} < r_{thres}$. The threshold distance r_{thres} depends on the maximum distance r_{max} of a rigid structure for which an NOE/ROE could be detected and on the significance level σ of the smallest considered population $(0 < \sigma < 1)$: $r_{thres} = r_{max} \cdot \sigma^{1/6}$. For $r_{max} = 3.5$ Å and $\sigma = 0.15$, for example, one obtains $r_{thres} = 2.6$ Å. In the following, a constraint derived from the absence of an NOE/ROE between two spins will be called an Anti Distance Constraint (ADC). ADC's, that exceed the DC's in number by far, are useful despite the fact that the restriction on the conformational space provided by a DC is stronger. The further a conformational search progresses towards the final structure, the higher the significance of certain ADC's can become. ADC's can easily be incorporated into a multi-conformational search. The use of ADC's in a mono-conformational search was shown to improve the accuracy of the NMR structure of a lac repressor headpiece (De Vlieg et al., 1986).

The philosophy of MEDUSA relies on the fact that the individual conformations may violate some of the DC's, which are fulfilled only by the entire dynamic set of substates. If, for example, two structures I and II are complementary with respect to two DC's belonging to the proton pairs (i,j) and (k,l) in the sense, that they fulfill the inequalities

$$r_{ii}(I) \leq r_{ii}^{NOE}$$
 and $r_{ii}(II) > r_{ii}^{NOE}$

and at the same time

$$\mathbf{r}_{kl}(\mathbf{I}) \geq \mathbf{r}_{kl}^{\text{NOE}}$$
 and $\mathbf{r}_{kl}(\mathbf{II}) < \mathbf{r}_{kl}^{\text{NOE}}$.

it is likely that a suitably weighted dynamic equilibrium between the two structures (I) and (II) can be found which will satisfy the two DC's. This is a consequence of the monotonous and rapidly decaying l/r^6 dependence. MEDUSA is directed to create sets of physically reasonable structures fulfilling inequalities of this type.

The algorithm consists of two parts. At first, a large number of static structures is generated, each of which fulfills a subset of the DC's, all ADC's, and having an energy below a set threshold. In a second step, the structures are combined in pairs, triples or larger clusters to fulfill all constraints including all DC's, ADC's, J-couplings, and $T_{1\rho}$ data. The structures are created as follows:

- 1. An initial structure S is chosen from a set of possibilities, such as X-ray structure, structure from MD simulations, and randomly chosen structures.
- 2. A randomly permuted list of the experimentally measured DC's is established and the first DC is selected from this list.
- 3. An energy threshold E_{thres} is chosen randomly within a preset interval $[E_{min}, E_{max}]$ (This avoids the difficult choice of a fixed threshold that may either be too loose or too restrictive).
- The selected DC (e.g. acting between spins I_k and I_l), together with all previous DC's, is added to the molecular force field of CHARMM (Brooks et al., 1983) in the form of a semiparabolic energy term,



а

b

I



Fig. 1. Structural pair of the antamanide exchange system with the minimum least squares error Q (Eq. 2) and fulfilling the $T_{1\rho}$ restrictions imposed by the experiments. (a) Schematic representation of the hydrogen bonding networks by arrows. The amino acid residues are labeled by their one-letter codes. (b) 3D structural stereoview representation including hydrogen bonds indicated by dotted lines. Pro⁸ is in the top left corner and Pro³ in the bottom right corner with a clockwise increase of the residue numbers indicated in (a). The optimal populations of the two structures I and II are 47% and 53%, respectively.

7

$$E_{kl} = \begin{cases} K(r_{kl} - r_{kl}^{NOE})^2, & \text{if } r_{kl} \ge r_{kl}^{NOE} \\ 0, & \text{if } r_{kl} < r_{kl}^{NOE} \end{cases}$$
(1)

An energy minimization, e.g. with 500 steps of ABNR (Adapted Base Newton-Raphson algorithm; Brooks et al., 1983), yields a new structure S'. Its energy E' is determined in the absence of distance constraints.

- (a) If $E' < E_{thres}$, S is set equal to S'.
- (b) If $E' > E_{thres}$, the tested (and violated) DC is removed from the list and from the potential.
- 5. The next DC is selected from the list of step 2 and step 4 is repeated. When the list of DC's is exhausted, the structure S is energy-minimized again, this time for a potential that includes all accepted DC's and all ADC's, the latter also in the form of semiparabolic energy terms. The coordinates of the resulting structure are stored, together with the list of accepted DC's, and the procedure is continued at step 2 or, less frequently, at step 1.

The random permutation of the DC's in step 2 is crucial for an adequate covering of the conformational space. In the presence of multiple conformations, the final structures depend strongly on the order of the successively applied DC's.

When a sufficient number of structures (usually at least several hundreds) has been generated in this manner, all possible combinations each containing N_s structures (in practice, N_s=2, 3, or rarely more) are formed. Each combination is tested as a potential exchange system with yet unknown probabilities p_x for the N_s structures. The probabilities are optimized to minimize, in the dynamic average, the deviations from the experimental DC's and J-coupling constraints (leading to angular constraints or AC's). For this purpose, the weighted mean square error

$$Q = A \sum_{(kl)} \left\{ \sum_{x}^{N_{x}} p_{x} J_{kl}^{x} - J_{kl}^{exp} \right\}^{2} + B \sum_{(kl)} \left\{ \sum_{x}^{N_{x}} p_{x} (r_{kl}^{x})^{-6} - (r_{kl}^{NOE})^{-6} \right\}^{2}$$
(2)

where $A = q/[\sum_{(kl)} (J_{kl}^{exp})^2]$

and $B = (1-q) / [\sum_{(kl)} (r_{kl}^{NOE})^{-12}]$

is minimized by zeroing the first partial derivatives with respect to the probabilities p_{α} . The number q, with $0 \le q \le 1$, allows one to shift the emphasis between distance constraints and angular constraints (J couplings). The J-coupling constants J_{kl}^{α} are computed from Karplus-type equations (Bystrov, 1976). Different weighting of the various terms in Eq. 2 is conceivable, e.g. to enhance the importance of long-range interactions.

In a last step, structural sets with a low mean square error Q are checked for their compatibility with the available $T_{1\rho}$ data in terms of feasible exchange mechanisms.

APPLICATION TO ANTAMANIDE

The algorithm described above was applied to the cyclic decapeptide antamanide $(-Val^1 - Pro^2 - Pro^3 - Ala^4 - Phe^5 - Phe^6 - Pro^7 - Pro^8 - Phe^9 - Phe^{10} -)$. Early ultrasonic absorption measurements in dioxane have indicated an unidentified conformational exchange



Fig. 2. Number of class structures for antamanide plotted as a function of the threshold value Δ_{thres} of the dihedral angle rms deviation that defines the angular space of each class.

process with $\tau_c \simeq 10^{-6}$ s (Burgermeister et al., 1974). A more recent NMR study of antamanide, combined with restrained MD simulations (Kessler et al., 1988, 1989), suggested the presence of two or four different structures in chloroform, one being essentially identical to the X-ray structure (Karle et al., 1979). The rate of this process has however not been determined. We attempted to reconcile these two studies based on $T_{1\rho}$ -measurements in chloroform (Ernst et al., 1990; Blackledge et al., to be published). They revealed a dynamic process that affects primarily the NH protons of Val¹ and Phe⁶ with an exchange time constant $\tau_e \simeq 27 \, \mu$ s at 320 K.



Fig. 3. Histogram of the pairs of class structures for $\Delta_{thres} = 3^{\circ}$ for the antamanide exchange system as a function of the mean square error Q that is a measure for the quality of fitting the experimental NOE and J-coupling measurements. The pair of structures suggested by Kessler et al. (1989) has a Q value indicated by the arrow a. The single structure with the best fit is associated with the Q value indicated by arrow b. The number of pair structures which conform, in addition, also to the hydrogen bonding network dynamics suggested by the T_{1p} measurements are shown in the insert, again for $\Delta_{tresh} = 3^{\circ}$. The 19 pairs with minimum error Q are all close to the structural pair visualized in Fig. 1.

To determine the substate structures, 23 backbone-backbone DC's were derived from crosspeak intensities in NOESY and ROESY spectra (Blackledge et al., to be published). In addition, 108 relevant ADC's were found. HNC_aH vicinal J-coupling constants were extracted from homonuclear COSY spectra while heteronuclear $J({}^{13}C'NC_{a}H)$ and $J(HNC_{a}{}^{13}C')$ couplings were derived from ¹H-detected heteronuclear COSY, ¹³C-detected COLOC, and ¹³C-detected J-resolved 2D experiments.

As starting structures for the structure generation algorithm, we chose the X-ray structure (Karle et al., 1979), the structure B proposed by Kessler et al. (1989), the structure produced by a 400 ps MD simulation given by Brüschweiler et al. (to be published), and some randomly selected structures. The energy thresholds were chosen in the interval 560 to 800 kJ/mol (whereby the lowest energy ever reached with the CHARMm force field parameters 'parallh19' was 556 kJ/mol). The constant K in Eq. 1 was set to $13 \cdot 10^6$ kJ/Å². The proposed algorithm produced 1176 structures. Pair and triple combinations of these structures have been optimized to minimize Q of Eq. 2. It was found that pair combinations gave adequate fits.

As an example, Fig. 1 shows the conformational pair with the smallest mean square error Q among the $\binom{1176}{2}$ combinations conforming well to the $T_{1\rho}$ measurements as just two hydrogen bonds involving the NH protons of Val¹ and Phe⁶ are broken, a feature that is not fulfilled by the structures proposed by Kessler et al. (1989). This is discussed in more detail by Blackledge et al. (to be published).

To survey the variety of structures generated by MEDUSA, the structures are sorted into classes based on the similarity of their backbone dihedral angles φ_i and ψ_i . Two structures S_{α} and S_{β} are said to belong to the same (backbone) class when their rms deviation $\Delta_{\alpha\beta}$ is smaller than a given threshold Δ_{tresh} . The rms deviation $\Delta_{\alpha\beta}$ is defined by

$$\Delta_{\alpha\beta}^{2} = \frac{1}{2n-1} \sum_{i=1}^{n} \left[(\varphi_{i}^{\alpha} - \varphi_{i}^{\beta})^{2} + (\psi_{i}^{\alpha} - \psi_{i}^{\beta})^{2} \right]$$
(3)

where n is the number of residues of the (cyclic) peptide. The conformation S with the lowest energy within each class is called the class structure. The sorting procedure starts with the first structure of a randomly sorted list and selects all those structures that belong to the same class. The next class is started with the first alien structure of the list, and the procedure is continued until all structures are classified. Although the obtained classification depends on the order of structures in the list, it provides insight into the diversity of the structures.

Figure 2 presents the number of classes as a function of the rms angular threshold Δ_{thres} . It demonstrates that the set of structures is relatively wide spread. When Δ_{thres} is set equal to 10°, 77 classes of structures result. For $\Delta_{\text{thres}} = 20^{\circ}$, one obtains 28 classes, and for $\Delta_{\text{thres}} = 40^{\circ}$, the number of classes is reduced to 8.

To reduce the computational effort in combining all 1176 structures in pairs, triples, or larger clusters, it is also possible to combine only class structures for a given threshold Δ_{thres} . Figure 3 presents a histogram of pairwise combined class structures for $\Delta_{thres} = 3^{\circ}$ plotted against the mean square error Q of Eq. 2. It shows all possible pairs and, in the upper right corner, the pairs that also conform to the T_{1p} measurements which require the hydrogen bonds involving the NH protons of Val¹ and Phe⁶ to be broken. For the pair of structures suggested by Kessler et al. (1989) one finds Q = 0.15, while the best single-structure fit leads to Q = 0.16 compared to Q = 0.12 for

the best pair structure. The figure shows that there is a considerable number of possible structures that are only slightly disfavored by the error measure Q. It can be concluded that the structures shown in Fig. 1 are possible but not the only candidates to model reality. It can also not be excluded that other hydrogen-bond dynamics mechanisms, involving further, 'accidentally' $T_{1\rho}$ -inactive NH protons, are taking place.

CONCLUSIONS

The algorithm MEDUSA proposed in this paper allows a systematic and comprehensive interpretation of the experimental constraints in terms of the dynamic structure of multi-conformational peptides and proteins. It produces a large family of feasible exchange systems rather than emphasizing a single average structure. It can also provide good insight into the character of a possibly ill-determined experimental data set.

A special feature of MEDUSA is its relative insensitivity to the selected initial structures. By the step-wise introduction of randomly ordered distance constraints it is possible to scan even with a small number of starting structures a large part of the conformational space spanned by the NMR distance information together with the holonomic restrictions, provided that a sufficiently strong weight is given to the distance constraints in determining the potential. Nevertheless, it is advisable to spread the initial structures as well as possible to avoid accidental confining of the search to a conformational subdomain.

With the proposed procedure it is more difficult to adequately cover the conformational domain in the case of missing DC's that lead to undetermined substructures. By constrained energy minimization it is no longer possible to access the entire set of physically feasible conformations. In this situation, it is possible to combine the energy minimization using randomized DC's with molecular dynamics simulation. It is, for example, conceivable to perform after the introduction of all compatible constraints and energy minimization a restrained molecular dynamics simulation to explore the remaining motional degrees of freedom of the obtained structure. This allows then to account for flexibility of the individual substates that are finally combined to describe the multi-conformational equilibrium.

ACKNOWLEDGEMENTS

We thank Prof. H. Kessler for providing a sample of antamanide and Prof. C. Griesinger for experimental advice in the initial stage of this project. Helpful comments by Prof. W. van Gunsteren are acknowledged. The authors are grateful to Prof. M. Karplus for providing a copy of the program CHARMM with the parameter set 'parallh19'. This work was supported in part by the Swiss National Science Foundation. The generous allowance to use an AMX 500 spectrometer at Bruker in Karlsruhe is gratefully acknowledged. The manuscript has been processed by Mrs. I. Müller.

REFERENCES

Beechem, J.M. and Brand, L. (1985) Annu. Rev. Biochem., 54, 43-71.

- Blackledge, M., Brüschweiler, R., Griesinger, C. and Ernst, R.R. (1990) In Proc. 10th Int. Biophysics Congress, Vancouver, Abstract Volume, p.109.
- Blackledge, M., Brüschweiler, R., Griesinger, C. and Ernst, R.R., to be published.
- Bothner-By, A.A., Stephens, R.L., Lee, J., Warren, C.D. and Jeanloz, R.W. (1984) J. Am. Chem. Soc., 106, 811-813.
- Braun, W. and Go, N. (1985) J. Mol. Biol., 186, 611-626.
- Brooks, B.R., Bruccoleri, R.E., Olafsen, B.D., States, D.J., Swaminathan, S. and Karplus, M. (1983) J. Comput. Chem., 4, 187-217.
- Brüschweiler, R., Roux, B., Blackledge, M., Griesinger, C., Karplus, M. and Ernst, R.R., to be published.
- Burgermeister, W., Wieland, T. and Winkler, R. (1974) Eur. J. Biochem., 44, 311-316.
- Bystrov, V.F. (1976) Prog. NMR Spectrosc., 10, 41-81.
- Clore, G.M., Gronenborn, A.M., Brünger, A.T. and Karplus, M. (1985) J. Mol. Biol., 186, 435-455.
- Clore, G.M., Driscoll, P.C., Wingfield, P.T. and Gronenborn, A.M. (1990) Biochemistry, 29, 7387-7401.
- Crippen, G.M. and Havel, T.F. (1988) Distance Geometry and Molecular Conformation, Wiley, New York.
- Cusack, S. (1989) Chem. Scripta, 29A, 103-107.
- Deverell, C., Morgan, R.E. and Strange, J.H. (1970) Mol. Phys., 18, 553-559.
- de Vlieg, J., Boelens, R., Scheek, R.M., Kaptein, R. and van Gunsteren, W.F. (1986) Israel J. Chem., 27, 181-188.
- Ernst, R.R., Blackledge, M., Boentges, S., Briand, J., Brüschweiler, R., Ernst, M., Griesinger, C., Mádi, Z.L., Schulte-Herbrüggen, T. and Sørensen, O.W. (1991) In *Proteins: Structure, Dynamics and Design* (Eds. Renugopalakrishnan, V., Carey, P.R., Smith, I.C.P., Huang, S.-G. and Storer, A.C.) ESCOM, Leiden, pp. 11-28.
- Ernst, R.R., Bodenhausen, G. and Wokaun, A. (1987) Principles of NMR in One and Two Dimensions, Clarendon Press, Oxford.
- Havel, T.F. and Wüthrich, K. (1984) Bull. Math. Biol., 46, 673-698.
- Jeener, J., Meier, B.H., Bachmann, P. and Ernst, R.R. (1979) J. Chem. Phys., 71, 4546-4553.
- Jones, G.P. (1966) Phys. Rev., 148, 332-335.
- Kaptein, R., Zuiderweg, E.R.P., Scheek, R.M., Boelens, R. and van Gunsteren, W.F. (1985) J. Mol. Biol., 182, 179-182.
- Karle, I.L., Wieland, T., Schermer, D. and Ottenheym, H.C.J. (1979) Proc. Natl. Acad. Sci. USA, 76, 1532-1536.
- Kessler, H., Griesinger, C., Lautz, J., Müller, A., van Gunsteren, W.F. and Berendsen, H.C.J. (1988) J. Am. Chem. Soc., 110, 3394-3396.
- Kessler, H., Bats, J.W., Lautz, J. and Müller, A. (1989) Liebigs Ann., 913-928.
- Kim, P.S. and Baldwin, R.L. (1990) Annu. Rev. Biochem., 59, 631-660.
- Kopple, K.D., Wang, Y.-S., Cheng, A.G. and Bhandary, K.K. (1988) J. Am. Chem. Soc., 110, 4168-4176.
- Kumar, A., Wagner, G., Ernst, R.R. and Wüthrich, K. (1981) J. Am. Chem. Soc., 103, 3654-3658.
- Lipari, G. and Szabo, A. (1982) J. Am. Chem. Soc., 104, 4546-4559, 4559-4570.
- Nusser, W., Kimmich, R. and Winter, F. (1988) J. Phys. Chem., 92, 6808-6814.
- Petsko, G.A. and Ringe, D. (1984) Annu. Rev. Biophys. Bioeng., 13, 331-371.
- Ringe, D. and Petsko, G.A. (1985) Progr. Biophys. Molec. Biol., 45, 197-235.
- Roder, H. and Wüthrich, K. (1986) Proteins, 1, 34-42.
- Roder, H., Elöve, G.A. and Englander, S.W. (1988) Nature, 335, 700-704.
- Slutsky, L.J., Allin, C.D., Madsen, L. and Brady, J. (1989) Ultrason. Symp. Proc., 2, 965-968.
- Torda, A.E., Scheek, R.M. and van Gunsteren, W.F. (1990) J. Mol. Biol., 214, 223-235.
- Udgaonkar, J.B. and Baldwin, R.L. (1988) Nature, 335, 694-699.
- Wagner, G. and Wüthrich, K. (1982) J. Mol. Biol., 160, 343-361.
- Wright, P.E., Dyson, H.J. and Lerner, R.L. (1988) Biochemistry, 27, 7167-7175.
- Wüthrich, K. (1986) NMR of Proteins and Nucleic Acids, Wiley, New York.