J. Mol. Biol. (1991) 217, 531–540

Structure Determination of the *Antp(C39 → S)* Homeodomain from Nuclear Magnetic Resonance Data in Solution using a Novel Strategy for the Structure Calculation with the Programs DIANA, CALIBA, HABAS and GLOMSA

Peter Güntert[1], Yan Qiu Qian[1], Gottfried Otting[1], Martin Müller[2], Walter Gehring[2] and Kurt Wüthrich[1]

[1]*Institut für Molekularbiologie und Biophysik*
Eidgenösische Technische Hochschule-Hönggerberg
CH-8093 Zürich, Switzerland

[2]*Biozentrum der Universität Basel, Abteilung Zellbiologie*
Klingelbergstrasse 70, CH-4056 Basel, Switzerland

(Received 22 June 1990; accepted 2 October 1990)

The structure of a mutant *Antennapedia* homeodomain, *Antp(C39 → S)*, from *Drosophila melanogaster* was determined using a set of new programs introduced in the accompanying paper. An input dataset of 957 distance constraints and 171 dihedral angle constraints was collected using two-dimensional n.m.r. experiments with the ^{15}N-labeled protein. The resulting high quality structure for *Antp(C39 → S)*, with an average root-mean-square deviation of 0·53 Å between the backbone atoms of residues 7 to 59 in 20 energy-refined distance geometry structures and the mean structure, is nearly identical to the previously reported structure of the wild-type *Antp* homeodomain. The only significant difference is in the connection between helices III and IV, which was found to be less kinked than was indicated by the structure determination for *Antp*. The main emphasis of the presentation in this paper is on a detailed account of the practical use of a novel strategy for the computation of nuclear magnetic resonance structures of proteins with the combined use of the programs DIANA, CALIBA, HABAS and GLOMSA.

1. Introduction

The protein structure determination of a mutant *Antp*† homeodomain described in this paper relates to a fundamental problem of developmental biology, i.e. the conversion of the genetic information contained in the egg cell of eukaryotes into the complex structure of the embryo. In the fruit fly *Drosophila melangogaster*, three classes of genes have been described that specify its body plan (Gehring, 1987). Genes from all three classes contain a highly conserved coding fragment of about 180 bp, which is commonly referred to as the *homeobox* (McGinnis *et al.*, 1984*a*,*b*; Scott & Weiner, 1984). Highly homologous homeoboxes have been found in various other species, including organisms as different as nematodes and humans, and a similarly high degree of homology is encountered among the *homeodomains* encoded by the homeoboxes (Scott *et al.*, 1989). Homeodomains are responsible for the DNA-binding of the entire protein encoded by the homeobox-containing genes, for example, the *antennapedia* (*Antp*) protein. A recombinant *Antp* homeodomain polypeptide of 68 amino acid residues has been shown to retain the ability of specific binding to the DNA (Müller *et al.*, 1988). In a project aimed at elucidating the mechanism of homeodomain–DNA interactions, the three-dimensional structure of this recombinant *Antp* homeodomain was solved by nuclear magnetic resonance (n.m.r.) spectroscopy (Otting *et al.*, 1988; Qian *et al.*, 1989; Billeter *et al.*, 1990*a*).

For structural studies of homeodomain–DNA

† Abbreviations used: *Antp*, *Antennapedia*; *Antp(C39 → S)* homeodomain, *Antp* homeodomain with the residue Cys39 replaced by Ser; bp, base-pair(s); n.m.r., nuclear magnetic resonance; NOE, nuclear Overhauser enhancement; NOESY, 2-dimensional nuclear Overhauser enhancement spectroscopy; COSY, 2-dimensional correlated spectroscopy; c.p.u., central processing unit; r.m.s.d., root-mean-square deviation; p.p.m., parts per million.

0022-2836/91/030531-10 $03.00/0

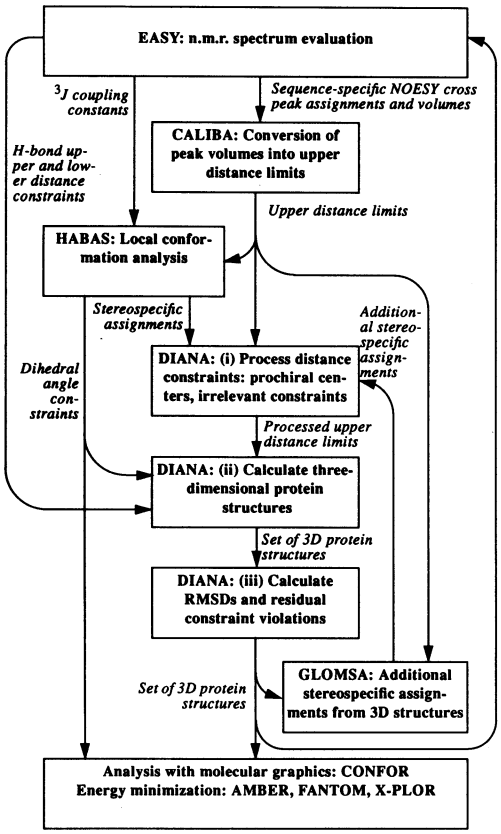

Figure 1. Scheme outlining the computation of a 3-dimensional protein structure from n.m.r. data using the program DIANA together with supporting programs for n.m.r. spectral analysis and structure refinement. See the text for details.

complexes by n.m.r. in solution, the wild-type *Antp* homeodomain would have a drawback because it contains a single Cys, so that the experiments require the addition of a reducing agent. The mutant *Antp* homeodomain with Cys39 replaced by Ser, *Antp(C39 → S)*, was therefore prepared. The structure determination of the *Antp(C39 → S)* homeodomain and the comparison of the three-dimensional structure of this mutant protein with that of the wild-type *Antp* homeodomain, which are presented here, were therefore an essential platform for the structure determination of an *Antp(C39 → S)*–DNA complex (Otting *et al.*, 1990).

In addition to presenting the *Antp(C39 → S)* structure, the major purpose of this paper is to give a detailed description of the practice of a novel strategy for protein structure calculations from n.m.r. data, which was introduced in the accompanying paper (Güntert *et al.*, 1990). This new procedure makes use of a new implementation of the variable target function algorithm (Braun & Gō, 1985) in the program DIANA. Efficient structure calculations are afforded by a user-friendly package of four programs, where DIANA is used in combination with CALIBA (Güntert *et al.*, 1990), HABAS (Güntert *et al.*, 1989) and GLOMSA (Güntert *et al.*, 1990).

2. Experimental Procedures

(a) *Preparation of the* Antp(C39 → S) *homeodomain*

The cysteinyl residue in position 39 of the *Antp* homeodomain peptide expressed from the expression vector pAop2 (Müller *et al.*, 1988) was changed to Ser using site-directed mutagenesis (Amersham oligonucleotide-directed *in vitro* mutagenesis system version 2). An appropriate fragment of pAAop2 was subcloned into Bluescript M13+ (STRATAGENE). Single-stranded DNA was prepared according to the manufacturer's protocol. The mutagenesis oligonucleotide was purified over an oligonucleotide purification cartridge (Applied Bio-systems). Its sequence was (mismatch is underlined): 5′-CACGCCCTG-TCCCTCACGGAG-3′. The correct structure of the mutagenized vector was confirmed by sequencing. A 170 bp *Kpn*I–*Hpa*I fragment was cloned back into pAAop2 to obtain the new expression vector pAop2CS.

The homeodomain peptide expressed from pAop2CS was purified as described (Müller *et al.*, 1988). In addition, a ^{15}N-labeled *Antp(C39 → S)* homeodomain was isolated from bacteria grown in 5 mM-$(^{15}NH_4)_2SO_4$, 22 mM-KH_2PO_4, 39 mM-Na_2H-PO_4, 8·5 mM-NaCl, 0·5% (w/v) glucose, 0·1 mM-$CaCl_2$, 1 mM-$MgSO_4$, 0·5 μM-$FeCl_3$, 5 μg thiamin/ml. The *Antp(C39 → S)* homeodomain no longer dimerizes *via* an S–S bridge, as has been observed for the wild-type homeodomain (Müller & Gehring, unpublished results). The *in vitro* DNA-binding properties of the *Antp* homeodomain are not altered by the exchange of Cys39 with Ser (Affolter *et al.*, 1990). The stability of the folded conformation of the 2 proteins with respect to thermal denaturation was found to be identical at pH 6·9, with a denaturation temperature of 60 °C.

(b) *n.m.r. measurements*

n.m.r. spectra of the *Antp(C39 → S)* homeodomain were recorded, assigned and evaluated in the same way as in the previous structure determination of the wild-type *Antp* homeodomain (Otting *et al.*, 1988; Billeter *et al.*, 1990*a*), except that NOESY cross-peak volumes were evaluated by a more sophisticated method using reference lineshapes that allow for correct quantification of overlapping peaks (Denk *et al.*, 1986; Eccles *et al.*, 1989), and $^3J_{HN\alpha}$ scalar coupling constants were measured with a *J*-modulated (^{15}N, ^{1}H)-COSY experiment with the ^{15}N-labeled protein (Neri *et al.*, 1990). A list of the proton chemical shifts in *Antp(C39 → S)* is given in Table 1; apart from the immediate vicinity of Ser39, they are very similar to those of the wild-type *Antp* homeodomain (Billeter *et al.*, 1990*a*). A total of 1201 NOESY cross-peaks could unambiguously be assigned and their volumes be determined, and 63 $^3J_{HN\alpha}$ and 88 $^3J_{\alpha\beta}$ scalar coupling constants were measured. In addition, 25 hydrogen bonds were identified from a combined analysis of NOE constraints and data on amide proton exchange (Wüthrich *et al.*, 1984; Wüthrich, 1986).

3. Computation of the Three-dimensional Structure from n.m.r. Data

We follow the procedure outlined in Figure 1 when using the new programs CALIBA, DIANA and GLOMSA (Güntert *et al.*, 1990, accompanying paper) in conjunction with other software for protein structure determination available in our laboratory. This outline starts with the analysis of the

Table 1

^1H chemical shifts, δ (p.p.m.), and stereospecific assignments for the Antp(39 → S) *homeodomain at pH 4·3 and 20 °C*

Residue	Chemical shift δ (p.p.m.)†			
	NH	αH	βH	Others
Met0		4·10	2·10, 2·10	γCH_2 2·55, –; εCH_3 2·13
Arg1	8·74	4·33	1·75, 1·75	γCH_2 1·58, –; δCH_2 3·14, –; εNH 7·18
Lys2	8·55	4·24	1·77, 1·70	γCH_2 1·41, 1·35; δCH_2 1·57, –; εCH_2 2·94, –; ζNH_3^+ 7·45
Arg3	8·49	4·29	1·81, 1·74	γCH_2 1·62, –; δCH_2 3·14, –; εNH 7·18
Gly4	8·45	3·98, 3·86		
Arg5	8·28	4·28	1·82, 1·69	γCH_2 1·66, 1·52; δCH_2 3·12, –; εNH 7·15
Gln6	8·55	4·31	*2·02, 1·95*	γCH_2 2·29, 2·21; εNH_2 *7·48, 6·81*
Thr7	8·15	4·17	3·97	γCH_3 0·95
Tyr8	8·14	4·89	*3·03, 2·67*	δH 6·98, –; εH 6·68, –
Thr9	9·14	4·40	4·80	γCH_3 1·31
Arg10	9·06	4·05	*1·92, 1·86*	γCH_2 1·72, 1·63; δCH_2 3·21, –; εNH 7·28
Tyr11	8·41	4·12	*3·20, 2·72*	δH 7·02, –; εH 6·70, –
Gln12	7·73	3·49	*2·57, 1·54*	γCH_2 *2·69, 2·37*; εNH_2 *7·67, 6·80*
Thr13	8·45	3·34	4·24	γCH_3 1·26
Leu14	8·16	3·89	*1·78, 1·46*	γH 1·68; δCH_3 *0·83, 0·82*
Glu15	7·73	3·82	*1·78, 1·68*	
Leu16	8·17	3·47	*0·42, −1·21*	γH 1·04; δCH_3 *0·37, −0·68*
Glu17	8·18	4·19	*2·11, 1·97*	γCH_2 2·50, 2·31
Lys18	7·76	3·94	1·86, 1·86	γCH_2 1·52, 1·31; δCH_2 1·63, –; εCH_2 2·85, –
Glu19	7·93	4·18	2·30, 2·30	γCH_2 2·47, –
Phe20	8·90	4·39	*3·07, 2·99*	δH 6·86, –; εH 6·96, –; ζH 6·54
His21	7·80	4·05	3·22, 3·22	δ^2H 6·94; ε^1H 8·62
Phe22	7·57	4·38	3·19, 3·14	δ, ε, ζH 7·31
Asn23	8·29	4·50	2·71, –	δNH_2 *7·92, 7·16*
Arg24	8·15	3·58	*1·19, 0·89*	γCH_2 0·85, −0·07; δCH_2 1·82, –; εNH 6·26; ηNH_2 6·52, –
Tyr25	7·64	4·50	*2·85, 2·59*	δH 7·04, –; εH 6·71, –
Leu26	8·35	4·51	*1·44, 1·15*	γH 0·61; δCH_3 *0·38, 0·14*
Thr27	7·51	4·34	4·59	γCH_3 1·28
Arg28	8·83	3·76	*1·84, 1·77*	γCH_2 1·61, 1·57; δCH_2 3·14, –; εNH 7·66
Arg29	8·29	3·86	1·81, 1·62	γCH_2 1·49, –; δCH_2 3·12, –; εNH 7·23
Arg30	7·64	4·05	1·82, 1·71	γCH_2 1·62, 1·51; δCH_2 *3·44, 3·02*; εNH 7·73; ηNH_2 *6·99, 6·92*
Arg31	8·43	3·64	*1·98, 1·77*	γCH_2 *1·54, 1·39*; δCH_2 3·45, 3·25; εNH 8·04; ηNH_2 *7·07, 6·69*
Ile32	7·92	3·45	1·79	γCH_2 1·67, 0·97; γCH_3 0·76; δCH_3 0·76
Glu33	7·71	4·00	*2·25, 2·02*	γCH_2 *2·52, 2·27*
Ile34	8·62	3·77	1·70	γCH_2 1·66, 1·02; γCH_3 0·80; δCH_3 0·67
Ala35	8·29	3·69	1·40	
His36	7·98	4·44	3·32, 3·32	δ^2H 7·36; ε^1H 8·56
Ala37	8·25	4·11	1·51	
Leu38	8·26	4·52	*1·64, 1·52*	γH 1·89; δCH_3 *0·68, 0·62*
Ser39	7·80	4·06	4·12, 4·10	
Leu40	8·13	4·89	1·54, 1·27	γH 1·52; δCH_3 *0·85, 0·62*
Thr41	8·76	4·54	4·72	γCH_3 1·27
Glu42	9·08	3·63	1·99, 1·99	γCH_3 2·29, –
Arg43	8·30	3·96	1·89, 1·75	γCH_2 1·62, –; δCH_2 3·18, –; εNH 7·30
Gln44	7·95	4·00	*2·58, 2·00*	γCH_2 2·53, 2·41; εNH_2 *7·40, 6·84*
Ile45	7·97	3·85	2·11	γCH_2 1·53, 1·45; γCH_3 0·92; δCH_3 0·67
Lys46	8·51	3·96	*2·08, 1·88*	γCH_2 1·62, 1·38; δCH_2 1·68, –; εCH_2 2·78, –; ζNH_3^+ 7·48
Ile47	8·52	3·79	1·90	γCH_2 1·69, 1·27; γCH_3 0·96; δCH_3 0·82
Trp48	8·12	3·88	*3·42, 3·24*	δ^1H 6·05; ε^3H 6·74; ε^1NH 9·92; ζ^2H 7·13; ζ^3H 5·63; η^2H 6·36
Phe49	9·01	3·78	*3·49, 3·21*	δH 7·68, –; εH 7·45, –; ζH 7·14
Gln50	8·22	3·97	2·24, 2·13	γCH_2 *2·50, 2·43*; εNH_2 *7·54, 6·89*
Asn51	8·28	4·31	*2·59, 2·40*	δNH_2 *7.40, 6.89*
Arg52	8·31	3·50	*0·65, −0·38*	γCH_2 −0·22, −0·64; δCH_2 2·37, 1·95; εNH 9·52; ηNH_2 7·07, –
Arg53	8·30	4·18	2·20, 1·87	γCH_2 2·24, 1·33; δCH_2 2·82, 2·53; εNH 7·60; ηNH_2 *7·50, 6·94*
Met54	7·51	4·25	*2·19, 2·14*	γCH_2 *2·73, 2·61*; εCH_3 2·13
Lys55	7·48	3·96	1·68, 1·68	γCH_2 *1·34, 1·14*; δCH_2 1·44, –; εCH_2 2·91, –; ζNH_3^+ 7·46
Trp56	8·13	4·44	*3·64, 3·29*	δ^1H 7·36; ε^3H 7·56; ε^1NH 10·22; ζ^2H 7·38; ζ^3H 7·09; η^2H 7·16
Lys57	8·23	3·74	1·89, 1·89	γCH_2 1·41, –; δCH_2 1·68, 1·59; εCH_2 2·95, –; ζNH_3^+ 7·44
Lys58	7·61	4·01	1·83, 1·83	γCH_2 1·47, 1·35; δCH_2 1·60, –; εCH_2 2·90, –; ζNH_3^+ 7·52
Glu59	7·90	4·07	1·88, 1·88	γCH_2 *2·35, 2·18*
Asn60	7·80	4·47	2·32, 2·00	δNH_2 *6·95, 6·36*
Lys61	7·95	4·22	*1·79, 1·72*	γCH_2 1·35, –; δCH_2 1·59, –; εCH_2 2·89, –; ζNH_3^+ 7·55
Thr62	8·08	4·25	4·16	γCH_3 1·14
Lys63	8·28	4·24	1·81, 1·74	γCH_2 1·42, 1·40; δCH_2 1·55, –; εCH_2 2·93, –
Gly64	8·37	3·89, –		
Glu65	8·08	4·61	*2·04, 1·85*	γCH_2 2·31, –
Pro66		4·38	2·21, 1·91	γCH_2 1·99, 1·93; δCH_2 3·75, 3·66
Gly67	8·04	3·73, –		

† For methylene groups, 2 chemical shifts are given only when 2 resolved signals were observed, or when the presence of the 2 degenerate signals had been established unambiguously by the observation of a remote peak in a 2-quantum experiment in ^2H_2O. Otherwise, the 2nd methylene proton is represented by a dash. For diastereotopic pairs of methylene protons or isopropyl methyl groups, stereospecific individual assignments are printed in italics, and the first number is the chemical shift of the proton or methyl group with the lower branch number, e.g. the β^2 proton, or the δ^1 methyl group of leucine. The chemical shifts that differ from the corresponding shifts in the wild-type *Antp* homeodomain by 0·05 p.p.m. or more are underlined.

n.m.r. spectra using the interactive program EASY (Eccles *et al.*, 1989), which produces lists of the resonance assignments, NOESY cross-peak volumes, and values for the spin-spin coupling constants in a suitable format as input for CALIBA and HABAS. In the next steps, upper bounds for interatomic distances are derived from the NOESY cross-peak volumes by CALIBA, and dihedral angle constraints and stereospecific assignments are obtained by a combined analysis of local NOE distance constraints and ^3J scalar coupling constants using HABAS (Güntert *et al.*, 1989). The program DIANA then performs three different functions. (1) Distance constraints with pairs of diastereotopic substituents that were not individually assigned are processed so as to allow for both possible stereospecific assignments, and irrelevant distance constraints exceeding the maximal distances allowed by the covalent structure are eliminated. (2) The structure calculation yields a set of three-dimensional protein structures, of which a subset of "good structures" is selected by the criterion of having small final target function values. These are used for further analysis and refinement. (3) r.m.s.d. values between pairs of good structures are calculated, and tables of the residual constraint violations are prepared. GLOMSA then compares the experimental upper distance constraints with the corresponding distances in the set of good structures from DIANA in order to obtain additional stereospecific assignments. With the thus updated, refined input data set, new structures are computed with DIANA. The three-dimensional structures calculated by DIANA may also be used for refinements of the n.m.r. spectral analysis, e.g. to assign additional NOESY cross-peaks, or to backcalculate NOESY spectra. In general, the determination of a high-quality protein structure requires several rounds of DIANA calculations with successively more refined input constraint data sets. The good structures selected from the final DIANA run are subjected to energy minimization, e.g. with the programs AMBER (Weiner & Kollman, 1981), FANTOM (Schaumann *et al.*, 1990), or X-PLOR (Brünger *et al.*, 1987). Eventually, the structures are visually analyzed with the molecular graphics program CONFOR (Billeter *et al.*, 1985*a*,*b*).

(a) *Conversion of NOE intensities into upper distance limits*

The program CALIBA was used to derive upper distance limits from NOESY cross-peak intensities. We applied four different "calibration curves", i.e. relationships between cross-peak volume and upper distance limit, for backbone, side-chain, backbone–methyl, and side-chain–methyl cross-peaks. Here, backbone peaks manifest NOEs between NH or $C^\alpha H$ and either another $C^\alpha H$, NH, or non-methyl $C^\beta H$; backbone–methyl peaks are between NH or $C^\alpha H$ and a methyl group of Ala; side-chain peaks correspond to all other NOES that do not involve methyl groups; side-chain–methyl peaks are the same except that one or two methyl groups are involved. For the final DIANA calculation with *Antp(C39→S)*, the calibration curves were selected based on plots of cross-peak volumes *versus* average proton–proton distances in a set of preliminary structures. Whereas for the backbone peaks, the usual $1/b^6$ dependence fitted well to the data in such a plot, this was not the case for the side-chain peaks. Based on a manual best fit, we chose a $1/b^4$ dependence for this class. For cross-peaks with methyl protons we used the uniform averaging model (Braun *et al.*, 1981):

$$V = \frac{C}{b^6}, \qquad \text{for backbone peaks (498 peaks);}$$

$$V = \frac{0{\cdot}11C}{b^4}, \qquad \text{for side-chain peaks (456 peaks);}$$

$$V = \frac{0{\cdot}12C}{b-d_0}\left(\frac{1}{d_0^5} - \frac{1}{b^5}\right),$$

for backbone–methyl peaks (10 peaks);

$$V = \frac{0{\cdot}29C}{b-d_0}\left(\frac{1}{d_0^5} - \frac{1}{b^5}\right),$$

for side-chain–methyl peaks (237 peaks).

C is a constant that depends on the (arbitrary) scaling of the NOESY spectra, and $d_0 = 1{\cdot}9$ Å (1 Å = 0·1 nm) denotes the minimal sterically allowed distance between two protons. The proton–proton upper distance bounds generated by CALIBA were in all cases confined to the range 2·4 to 4·8 Å.

(b) *Dihedral angle constraints and stereospecific assignments from HABAS*

In the local conformation analysis by HABAS (Güntert *et al.*, 1989), 151 experimental vicinal scalar coupling constants, 213 intraresidual and sequential NOE upper distance constraints, and 39 relational constraints derived from NOEs (Güntert *et al.*, 1989) were used. The relational constraints for pairs of diastereotopic ligands were obtained automatically by CALIBA; the program generated a relational constraint whenever the volume of a cross-peak with one out of a pair of diastereotopic protons was more than 1·5 times that of the corresponding peak with the other diastereotopic proton. In all, 122 direct constraints for ϕ and ψ dihedral angles, 49 direct constraints for χ^1 dihedral angles, and 14 stereospecific assignments for $C^\beta H_2$ groups were thus found.

(c) *Processing of upper distance limits by DIANA*

Of the 1201 upper distance limits, 363 were found to be irrelevant, i.e. they are either independent of the conformation, or there exists no conformation that would violate the constraint. Together with 50 upper distance limits used to account for the 25

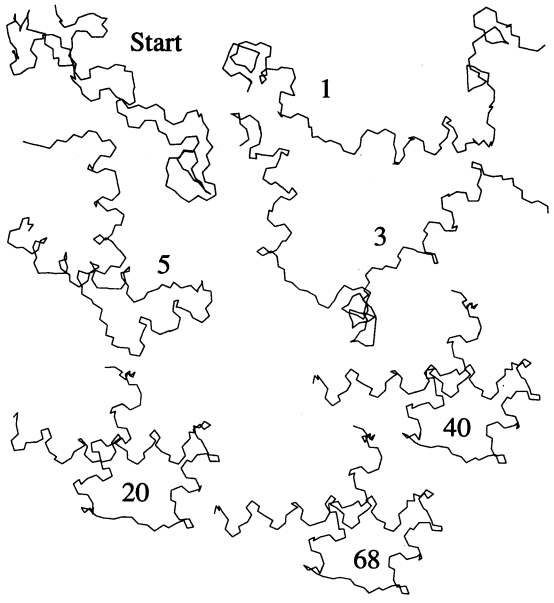

Figure 2. Intermediate structures during the minimization with the program DIANA of the *Antp(C39 → S)* homeodomain conformation with the smallest final target function value among the 250 conformations that were calculated. The backbone atoms of residues 7 to 59 of the random start conformation, and of the conformations at the end of the minimization levels $L = 1$, 3, 5, 20, 40 and 68 are shown.

experimentally identified hydrogen bonds (Williamson *et al.*, 1985), there were thus 888 relevant upper distance limits before further processing of NOE constraints involving prochiral centers without stereospecific assignments. Such processing of all NOEs with pairs of diastereotopic protons for which no stereospecific assignments were available in the input for the final DIANA calculation (see below) increased the number of upper distance limits slightly to 907.

(d) *Structure calculation with DIANA*

In this section we describe the final DIANA calculations. Preliminary structures used to refine the calibration curves in CALIBA and to obtain stereospecific assignments by GLOMSA were obtained in a similar way, except that smaller numbers of structures were calculated. The final input included 907 upper distance constraints, 50 lower distance constraints for H-bonds, 171 dihedral angle constraints, stereospecific assignments for 35 out of a total of 120 prochiral CH_2 groups and for all isopropyl groups, and individual proton assignments for ten out of a total of 22 NH_2 groups of Asn, Gln and Arg, which were determined from intramolecular NOEs with the nearest CH_2 group.

A total of 250 structure calculations were performed, using the same input data but different starting conformations, with random dihedral angles distributed uniformly between $-180°$ and $+180°$. For each conformation, the target function was minimized at the 24 levels (L in eqn (7) of Güntert *et al.*, 1990) 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60 and twice 68. Throughout the calculation, the weighting factors w_u and w_l for explicit upper and lower distance limits were set to unity, and the weighting factor w_a for dihedral angle constraints were set to 5 Å^2. For levels 0 to 60, the weighting factor w_v for van der Waals constraints was 0·2; for the first minimization step at level 68, it was increased to 0·6, and for the final minimization step at level 68 to 2·0. For minimization levels 1 to 5 and 68, the maximal number of target function evaluations was set to 500, for all others it was set to 300. Thus, the maximal number of target function evaluations for the calculation of one conformation was 8600. Calculations were done on a Cray X-MP computer and required 2·0 minutes of c.p.u. time per conformation, which added up to 8·4 hours of c.p.u. time for all 250 conformations.

For the structure with the smallest final target function value, the course of the DIANA calculation is illustrated in Figure 2, which shows the randomly generated start conformation, intermediate conformations at the end of the minimization levels $L = 1$, 3, 5, 20 and 40, and the final structure at the end of the minimization with the second level $L = 68$. At the minimization level $L = 3$, the helices are clearly formed, and the nearly final global fold is first apparent at the level $L = 20$.

(e) *Stereospecific assignments from GLOMSA*

Prior to the final run of DIANA, additional individual assignments for protons in CH_2 groups and isopropyl methyl groups were determined by matching the list of upper distance limits to a set of 17 preliminary structures using the program GLOMSA. There were 13 assignments for β-methylene groups (in addition to the 14 assignments found previously with HABAS), seven for γ-methylene groups, one for a δ-methylene group of arginine, and five for isopropyl groups of leucine.

Examples of the output from the program GLOMSA are afforded by Table 2. Stereospecific assignments are proposed by GLOMSA if $|\Delta d| > 0{\cdot}2$ Å, $|n_{\Delta d}| > 13$ and $|\Delta b| > 0{\cdot}2$ Å unless there is a relational constraint (see Table 2), in which case no limit is imposed on $|\Delta b|$ (for the notation used, see Güntert *et al.*, 1990). If these requirements are met, GLOMSA proposes stereospecific assignments based on comparing the signs of Δd and $n_{\Delta d}$ with that of Δb, or the relational constraint, respectively. If these three parameters have the same signs, then the assignment assumed in the input, I, is proposed, and otherwise the reversed assignment R. The β-methylene group of Arg10 is an example where unambiguous stereospecific assignments were obtained from GLOMSA, whereas this was not possible for the β-methylene group of Arg29. For the β-methylene group of Trp48, unambiguous stereospecific assignments could again be obtained. The last two examples in Table 2 show that stereospecific assignments can also be obtained for other

Table 2

GLOMSA results for selected prochiral centers of Antp(C39 → S) *homeodomain*

Prochiral center	NOEs to	Constraints (Å)† b_1		b_2	Δb (Å)	$n_{\Delta d}$‡	$\overline{\Delta d}$ (Å)	Assignment§
Arg10 $C^{\beta}H_2$	Arg10 HN	3·0		3·2	−0·2	−17	−0·8	*I*
	Arg10 H^{α}	3·0	<			−17	−0·3	*I*
	Tyr11 HN	3·6	>	3·3	0·3	17	1·1	*I*
Arg29 $C^{\beta}H_2$	Arg29 H^{ε}	4·8	>	4·8	0·0	−10	0·0	
	Arg29 HN	3·4	>	2·8	0·6	−12	−0·7	
	Arg29 H^{α}	2·7	<	3·0	−0·3	−11	−0·2	
Trp48 $C^{\beta}H_2$	Trp48 $H^{\varepsilon 3}$	3·4	<			17	1·5	*R*
	Trp48 $H^{\delta 1}$		>	3·2		−17	−1·1	*R*
Gln12 $C^{\gamma}H_2$	Gln12 HN	4·2		4·5	−0·3	17	0·2	*R*
	Gln12 H^{α}	3·1	>	2·5	0·6	17	1·1	*I*
Leu26 $C^{\delta 1,2}H_3$	Leu26 H^{α}	3·6	<			10	0·8	
	Thr27 HN	5·8	<	5·8	0·0	10	0·5	
	Arg31 HN	5·8	>	5·8	0·0	−14	−0·2	*R*
	Arg31 H^{α}	5·8	>	4·4	0·6	−13	−0·7	*R*
	Phe49 H^{δ}	7·8	>	7·8	0·0	−16	−1·8	*R*

† In the input, individual assignments are arbitrarily assumed for the 2 diastereotopic ligands. b_1 denotes the upper bound for the distance between the 1st ligand of the prochiral center, e.g. $H^{\beta 2}$ in a β-methylene group, and the atom in the 2nd column of the Table; b_2 denotes the corresponding upper bound for the 2nd atom of the diastereotopic pair, e.g. $H^{\beta 3}$ in a β-methylene group. Relational constraints are given between the 2 columns labeled b_1 and b_2: > denotes a relational constraint that requires the distance b_1 to be bigger than b_2, < denotes a relational constraint with the opposite meaning. $\Delta b = b_1 - b_2$.

‡ $n_{\Delta d}$ denotes the bigger of the 2 numbers of conformations for which the difference between the distances d_1 and d_2, $d_1 - d_2$, has the same sign. d_1 and d_2 are the proton–proton distances in the protein structure that correspond to the upper bounds b_1 and b_2 in the input for the structure calculation. If there are more conformations with $d_1 - d_2 > 0$ than with $d_1 - d_2 < 0$, the sign of $n_{\Delta d}$ is positive, otherwise it is negative. $\overline{\Delta d}$ denotes the average over all structures of the distance difference $d_1 - d_2$. Mathematical definitions of the quantities $n_{\Delta d}$ and $\overline{\Delta d}$ are given in eqns (16) and (17) of the accompanying paper (Güntert *et al.*, 1990).

§ Stereospecific assignments proposed by GLOMSA: *I* denotes the stereospecific assignment assumed in the input, *R* denotes the reversed stereospecific assignment. The thresholds were chosen such that a definite assignment is proposed only if $|\Delta b| > 0{\cdot}2$ Å (no threshold for $|\Delta b|$ when there is a relational constraint), and if $|n_{\Delta d}| > 13$ and $|\overline{\Delta d}| > 0{\cdot}2$ Å. In this case, the assignment *I* is proposed if the sign of Δb (or the sign implied in the relational constraint) matches the signs of $n_{\Delta d}$ and $\overline{\Delta d}$, otherwise the opposite assignment *R* is proposed.

groups than $C^{\beta}H_2$, for example γ-methylene groups and leucine isopropyl groups. For the γ-methylene protons of Gln12, two contradicting stereospecific assignments were proposed by GLOMSA on the basis of two distance constraint pairs. In this case, we accepted the *I* assignment obtained from the NOEs to H^{α} because the two upper limits of 3·1 and 2·5 Å correspond to much bigger NOE intensity differences than the two limits of 4·2 and 4·5 Å that lead to the *R* assignment. In the case of the isopropyl group of Leu26, the individual constraint pairs do not yield completely unambiguous evidence, because $|n_{\Delta d}| < 17$ (a stereospecific assignment proposed by GLOMSA based on a single pair of distance constraints would be accepted only if $n_{\Delta d} = \pm 17$), but since we have consistent assignments from three distance constraint pairs, the stereospecific assignment is still considered to be unambiguous.

(f) *Restrained energy minimization with AMBER*

The 20 *Antp(C39 → S)* DIANA structures with smallest final target function values were subjected to restrained energy minimization using a modified version of the program AMBER (Weiner & Kollman, 1981), which includes pseudoenergy terms for distance constraints and dihedral angle constraints. The pseudoenergy was proportional to the sixth power of the distance constraint violations, and adjusted such that violations of 0·2 Å for distance constraints, and 5° for angle constraints corresponded to $k_BT/2$ at room temperature (Billeter *et al.*, 1990*b*).

An analysis of the 20 structures before and after restrained energy minimization is afforded by Table 3. The best 20 DIANA structures before energy minimization have low target function values, and fulfil the NOE distance constraints and dihedral angle constraints almost perfectly. The moderate AMBER energies of the structures before energy minimization indicate that bad steric overlaps were successfully removed by DIANA. For all 20 DIANA structures, AMBER was able to find a low-energy conformation in the near vicinity, with only moderate increase of the sum of constraint violations. The average of the r.m.s.d. (McLachlan, 1979) between the DIANA structures and the corresponding energy-minimized structures was 0·24(±0·07) Å for the backbone atoms N, C^{α} and C′

Table 3

Analysis of the 20 best Antp(C39 → S) *structures obtained from DIANA before and after restrained energy minimization with the program AMBER*

Quantity	Average value ± standard deviation (range)†			
	Before energy minimization		After energy minimization	
Target function (Å²)‡	1·67±0·32	(0·89 . . . 2·12)		
AMBER energy (kcal/mol)	40±65	(−89 . . . 181)	−829±61	(−976 . . . −735)
NOE constraint violations:				
Number >0·2 Å	2·7±1·3	(0 . . . 5)	1·9±1·1	(0 . . . 5)
Sum (Å)	5·7±0·8	(4·4 . . . 6·9)	11·1±1·0	(9·1 . . . 13·4)
Maximum (Å)	0·34±0·10	(0·20 . . . 0·57)	0·21±0·01	(0·19 . . . 0·24)
Angle constraint violations:				
Number >5°	0·8±0·8	(0 . . . 3)	3·6±1·5	(1 . . . 7)
Sum (deg.)	28·0±8·1	(16·2 . . . 47·2)	67·9±11·8	(45·5 . . . 93·5)
Maximum (deg.)	5·2±1·4	(2·4 . . . 7·2)	6·4±1·0	(5·3 . . . 9·4)
Pairwise r.m.s.d. (Å):				
Backbone 0–67	5·22±1·27	(2·18 . . . 8·55)	5·24±1·27	(2·17 . . . 8·54)
Backbone 7–59	0·77±0·18	(0·43 . . . 1·31)	0·76±0·18	(0·41 . . . 1·32)
Core§	0·81±0·15	(0·50 . . . 1·26)	0·80±0·16	(0·45 . . . 1·26)
Heavy atoms 7–59	1·77±0·22	(1·35 . . . 2·38)	1·77±0·22	(1·36 . . . 2·37)
r.m.s.d. to mean structure (Å):				
Backbone 0–67	3·69±1·06	(2·70 . . . 5·47)	3·68±1·09	(2·81 . . . 5·67)
Backbone 7–59	0·53±0·17	(0·37 . . . 0·78)	0·55±0·18	(0·39 . . . 0·83)
Core§	0·56±0·16	(0·43 . . . 0·77)	0·57±0·17	(0·43 . . . 0·82)
Heavy atoms 7–59	1·22±0·31	(1·01 . . . 1·65)	1·23±0·32	(1·03 . . . 1·68)

A total of 250 structures were calculated with DIANA, but only the 20 structures with the smallest final target function values were subjected to energy minimization.

† Of the individual values for the 20 structures.

‡ The weighting factors for experimental upper and lower distance limits were 1, for van der Waals lower distance limits 2, and for dihedral angle constraints 5 Å². For the structures after energy minimization, the target function is undefined because these structures do not have the ECEPP standard geometry.

§ Includes the backbone atoms N, C^{α}, C′ of residues 7 to 59, and the side-chain heavy atoms of the residues 13, 16, 26, 34, 35, 37, 38, 40, 45, 48, 49. See Billeter *et al.* (1990*a*) for details.

in the complete polypeptide chain 0 to 67. The energy-minimized structures exhibit smaller maximal violations of NOE distance constraints but larger sums of NOE constraint violations because the corresponding penalty term in the modified AMBER program is much steeper than in DIANA (Billeter *et al.*, 1990*b*).

The r.m.s.d. values for the whole protein are large (Table 3) because there are almost no conformational constraints for the chain ends of residues 0 to 6 and 60 to 67. In contrast, the r.m.s.d for residues 7 to 59 are small and thus representative of a high-quality n.m.r. structure. The r.m.s.d. for the protein core, i.e. the backbone of residues 7 to 59 and the side-chain heavy atoms of the residues that are not exposed to the solvent (Billeter *et al.*, 1990*a*), are virtually identical to the r.m.s.d. for the backbone of residues 7 to 59 alone. A stereo view of the three-dimensional structure of the *Antp(C39 → S)* homeodomain represented by a group of 20 energy-minimized conformers (Wüthrich, 1986) is displayed in Figure 3.

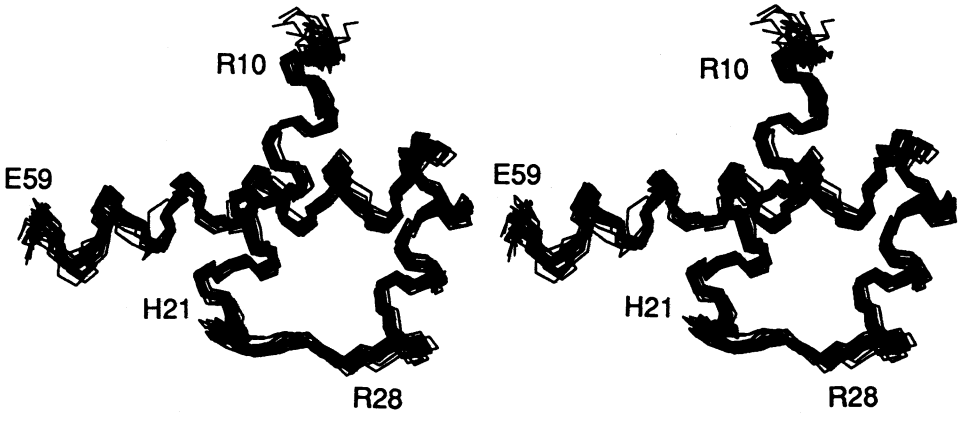

Figure 3. Stereo view of the superposition of the residues 7 to 59 of the 20 energy-minimized *Antp(C39 → S)* homeodomain structures of Table 3. The bonds connecting the backbone atoms N, C^{α} and C′ are shown.

Table 4

Comparison of the Antp(C39 → S) *structure with the wild-type* Antp *homeodomain structure*

Atom set	Average of the pairwise r.m.s.d. (Å)†	r.m.s.d. between the mean structures (Å)‡
Backbone 0–67	5·44 ± 1·35	2·64
Backbone 7–59	1·25 ± 0·26	1·01
Backbone 7–59 and core side-chains§	1·27 ± 0·27	1·00
Heavy atoms 7–59	2·35 ± 0·26	1·47
Backbone 7–52	0·84 ± 0·12	0·52

The solution conformation of *Antp(C39 → S)* is represented by the 20 energy-minimized DIANA structures of Table 3, and that of the wild-type *Antp* homeodomain by a similar set of 19 structures (Billeter *et al.*, 1990*a*).

† The average was taken of all 380 pairwise comparisons between the 20 energy-minimized *Antp(C39 → S)* structures and the 19 energy-minimized wild-type *Antp* structures.

‡ For both species, the mean structure was obtained by taking the arithmetic mean of the Cartesian atom co-ordinates after a fit for minimal r.m.s.d. of the backbone atoms N, C^{α} and C′ of the residues 7 to 59 of the individual structures to the best structure.

§ Includes the backbone atoms N, C^{α} and C′ of the residues 7 to 59 and the side-chain heavy atoms of the residues 13, 16, 26, 34, 35, 37, 38, 40, 45, 48, 49. See Billeter *et al.* (1990*a*) for details.

4. Comparison of *Antp(C39 → S)* with the wild-type *Antp* Homeodomain

Overall, the *Antp(C39 → S)* structure coincides closely with the previously determined structure of the wild-type *Antp* homeodomain (Qian *et al.*, 1989; Billeter *et al.*, 1990*a*). Global r.m.s.d. values between the structures of the two proteins are given in Table 4. They are only slightly bigger than the corresponding average r.m.s.d. for the 20 energy-minimized *Antp(C39 → S)* conformers (Table 3), or the 19 conformers computed for the wild-type *Antp* homeodomain (Table 5 of Billeter *et al.*, 1990*a*). The high r.m.s.d. value for the backbone atoms of residues 0 to 67 is due to the fact that the chain ends 0 to 6 and 60 to 67 are only poorly constrained by

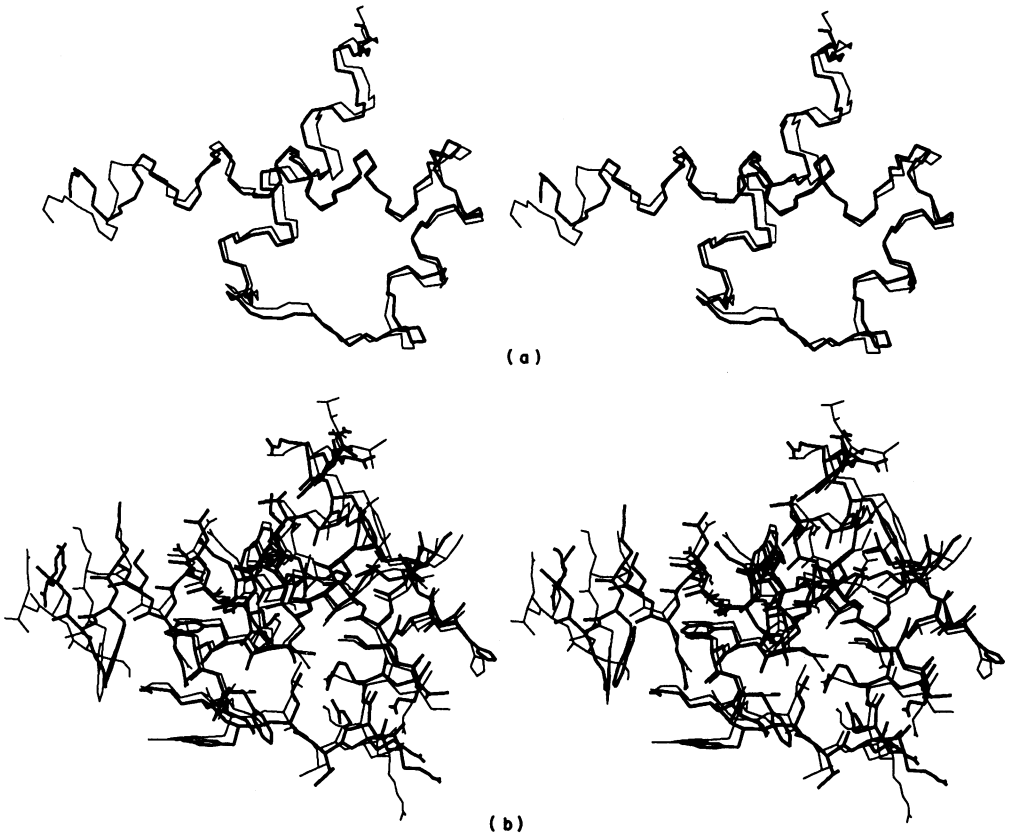

Figure 4. Stereo views of the superposition of the residues 7 to 59 of the mean structure of the 20 energy-minimized *Antp(C39 → S)* homeodomain structures (thick lines) with the mean of the 19 energy-minimized wild-type *Antp* homeodomain structures (thin lines). (a) The backbone atoms N, C^{α} and C′ are shown. (b) All heavy atoms are shown.

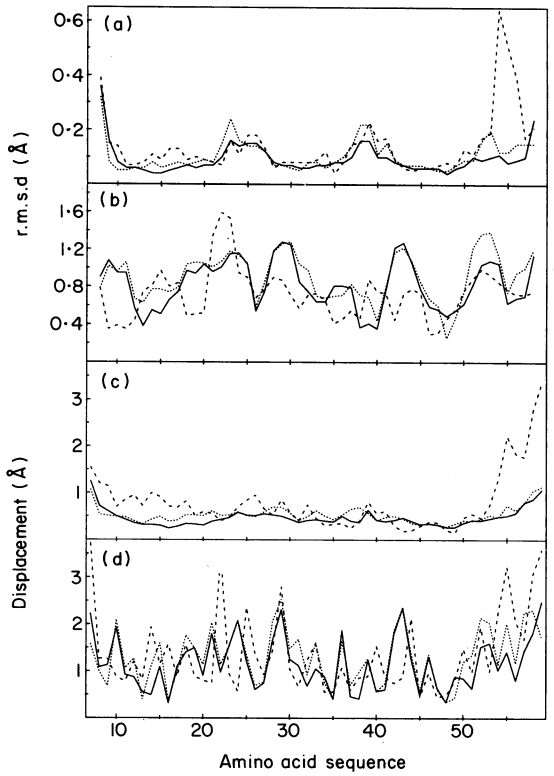

Figure 5. Plots *versus* the amino acid sequence of local r.m.s.d. and of average displacements for residues 7 to 59 of the *Antp(C39 → S)* and the wild-type *Antp* homeodomain structure. Local r.m.s.d. for segments of 3 residues were obtained after a local best fit of the corresponding segment, and the r.m.s.d. values are plotted at the position of the central residue. Displacements are r.m.s.d. calculated for the backbone atoms N, C^{α} and C′, or the side-chain heavy atoms, respectively, of 1 residue after a global best fit of the backbone atoms N, C^{α} and C′ of the residues 7 to 59. Continuous lines correspond to the average of the pairwise local r.m.s.d., or displacements, respectively, between the 20 energy-minimized *Antp(C39 → S)* structures and the mean structure obtained from these. Dotted lines correspond to the average of the pairwise local r.m.s.d., or displacements, respectively, between the 19 energy-minimized wild type *Antp* structures (Billeter *et al.*, 1990*a*) and the mean structure obtained from these. Broken lines indicate the local r.m.s.d., or displacements, respectively, between the mean of the 20 energy-minimized *Antp(C39 → S)* structures and the mean of the 19 energy-miinimized *Antp* structures. (a) Local backbone r.m.s.d.; (b) local heavy-atom r.m.s.d.; (c) backbone displacements; (d) side-chain heavy-atom displacements.

the n.m.r. data (Billeter *et al.*, 1990*a*). A visual impression of the close similarity of the structures of the two proteins is given by Figure 4. From Figure 4(a) it is apparent that the only notable difference between the backbone conformations of the two mean structures is found in residues 54 to 59 of helix IV, which is also clearly manifested in the small r.m.s.d. for residues 7 to 52 when compared to the backbone atoms of residues 7 to 59 (Table 4). Close similarity of the two structures is also supported by the near-identity of the ^{1}H chemical shifts (Table 1).

The local backbone r.m.s.d. (Figure 5(a)) show that the local backbone conformation was determined somewhat more precisely in the *Antp(C39 → S)* structure than in the structure of the wild-type *Antp* homeodomain, since for most residues the continuous line is below the dotted line. Between the two mean structures the only clear-cut difference is in the region of residues 54 to 56. The local heavy-atom r.m.s.d. values (Fig. 5(b)) exhibit a difference between the two mean structures in the region of Phe22. Since this residue is located on the surface, this difference is, however, not really significant. The backbone displacements (local r.m.s.d. obtained after a global best fit) show significant differences only for residues 53 to 59.

Figure 4(b) affords a demonstration that the close similarity between the two structures extends all the way to the molecular surface, with the aforementioned exception of Phe22. It also provides a clue for rationalizing the major difference between the two structures in the region of helix IV. The peptide group between Arg53 and Met54 is rotated by about 180° in *Antp* with respect to its position in *Antp(C39 → S)*, where it adopts a standard α-helical conformation. In *Antp*, the unusual conformation of this peptide group caused helix IV to open up and to be positioned further away from helix III. A slightly bigger kink between helices III and IV was then enforced by long-range NOEs between Phe20, Arg24 and Trp56. The 180° rotation of the peptide bond linking residues 53 and 54 was probably caused by short intraresidual upper distance limits of 3·4 and 4·0 Å, respectively, between HN and the two γ-methylene protons of Met54 (Billeter *et al.*, 1990*a*). In *Antp(C39 → S)*, the corresponding upper limits used were 4·4 and 4·5 Å, respectively.

The close similarity of the *Antp* and *Antp(C39 → S)* structures does not come as a surprise in view of the conservative nature of this substitution of a surface residue. From the viewpoint of the structure determination method, the near identity of the two structures is quite satisfying, considering that both structures were obtained completely independently and with somewhat different strategies for the structure calculations. Overall, a higher quality experimental input was collected for *Antp(C39 → S)* than for the wild-type *Antp* homeodomain, primarily because of the availability of a ^{15}N-labeled protein preparation, which enabled improved measurements of the spin–spin coupling constants (Neri *et al.*, 1990) and because of more extensive NOESY cross peak assignments. As a result, the presently described *Antp(C39 → S)* structure has a higher overall quality than the structure obtained for the wild-type *Antp* homeodomain. Therefore we tend to believe that the present structure also gives a more correct picture of the relative orientation between the helices III and IV, i.e. with a less pronounced kink than in the structure reported for the wild-type

Antp homeodomain (Qian *et. al.*, 1989; Billeter *et. al.*, 1990*a*). However, the fact that the junction between the helices III and IV is located outside of the globular core of the protein makes it difficult to unambiguously settle this question at this point.

We thank Dr M. Billeter for helpful discussions and Mr R. Marani for the careful processing of the manuscript. Financial support by the Schweizerischer Nationalfonds (projects 31.25174.88 and K03.613.87) and the use of the Cray X-MP/28 of the ETH Zürich are gratefully acknowledged.

References

Affolter, M., Percival-Smith, A., Müller, M., Leupin, W. & Gehring, W. J. (1990). *Proc. Nat. Acad. Sci., U.S.A.* **87**, 4093–4097.

Billeter, M., Engeli, M. & Wüthrich, K. (1985*a*). *J. Mol. Graph.* **3**, 79–83.

Billeter, M., Engeli, M. & Wüthrich, K. (1985*b*). *J. Mol. Graph.* **3**, 97–98.

Billeter, M., Qian, Y. Q., Otting, G., Müller, M., Gehring, W. J. & Wüthrich, K. (1990*a*). *J. Mol. Biol.* **214**, 183–197.

Billeter, M., Schaumann, T., Braun, W. & Wüthrich, K. (1990*b*). *Biopolymers,* **29**, 695–706.

Braun, W. & Gō, N. (1985). *J. Mol. Biol.* **186**, 611–626.

Braun, W., Bösch, C., Brown, L. R., Gō, N. & Wüthrich, K. (1981). *Biochim. Biophys. Acta,* **667**, 377–396.

Brünger, A. T., Kuriyan, J. & Karplus, M. (1987). *Science,* **235**, 458–460.

Denk, W., Baumann, R. & Wagner, G. (1986). *J. Magn. Reson.* **67**, 386–390.

Eccles, C., Billeter, M., Güntert, P. & Wüthrich, K. (1989). *Abstracts 10th Meeting of the International Society of Magnetic Resonance,* Morzine, France, July 16–21, 1989, p. S50.

Gehring, W. J. (1987). *Science,* **236**, 1245–1252.

Güntert, P., Braun, W., Billeter, M. & Wüthrich, K. (1989). *J. Amer. Chem. Soc.* **111**, 3997–4004.

Güntert, P., Braun, W. & Wüthrich, K. (1991). *J. Mol. Biol.* **217**, 517–530.

McGinnis, W., Levine, M. S., Hafen, E., Kuroiwa, A. & Gehring, W. J. (1984*a*). *Nature (London),* **308**, 428–433.

McGinnis, W., Hart, C. P., Gehring, W. J. & Ruddle, F. H. (1984*b*). *Cell,* **38**, 675–680.

McLachlan, A. D. (1979). *J. Mol. Biol.* **128**, 49–79.

Müller, M., Affolter, M., Leupin, W., Otting, G., Wüthrich, K. & Gehring, W. J. (1988). *EMBO J.* **7**, 4299–4304.

Neri, D., Otting, G. & Wüthrich, K. (1990). *J. Amer. Chem. Soc.* **112**, 3663–3665.

Otting, G., Qian, Y. Q., Müller, M., Gehring, W. J. & Wüthrich, K. (1988). *EMBO J.* **7**, 4305–4309.

Otting, G., Qian, Y. Q., Billeter, M., Müller, M., Affolter, M., Gehring, W. J. & Wüthrich, K. (1990). *EMBO J.* **9**, 3085–3092.

Qian, Y. Q., Billeter, M., Otting, G., Müller, M., Gehring, W. J. & Wüthrich, K. (1989). *Cell,* **59**, 573–580.

Schaumann, T., Braun, W. & Wüthrich, K. (1990). *Biopolymers,* **29**, 679–694.

Scott, M. P. & Weiner, A. J. (1984). *Proc. Nat. Acad. Sci., U.S.A.* **81**, 4115–4119.

Scott, M. P., Tamkun, J. W. & Hartzell, G. W., III (1989). *Biochim. Biophys. Acta,* **989**, 25–48.

Weiner, P. K. & Kollman, P. A. (1981). *J. Comp. Chem.* **2**, 287–303.

Williamson, M. P., Havel, T. F. & Wüthrich, K. (1985). *J. Mol. Biol.* **182**, 295–315.

Wüthrich, K. (1986). *NMR of Proteins and Nucleic Acids,* Wiley, New York.

Wüthrich, K., Billeter, M. & Braun, W. (1984). *J. Mol. Biol.* **180**, 715–740.

Edited by P. E. Wright