doi:10.1016/j.jmb.2004.01.009 J. Mol. Biol. (2004) 337, 443–451

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

The Solution NMR Structure of *Antheraea polyphemus* PBP Provides New Insight into Pheromone Recognition by Pheromone-binding Proteins

Smita Mohanty[1]*, Sergey Zubkov[1] and Angela M. Gronenborn[2]

[1]*Department of Biochemistry and Cell Biology, State University of New York at Stony Brook, Stony Brook, NY 11794-5215, USA*

[2]*Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda MD 20892, USA*

**Corresponding author*

Pheromone-binding proteins (PBPs) located in the antennae of male moth species play an important role in olfaction. They are carrier proteins, believed to transport volatile hydrophobic pheromone molecules across the aqueous sensillar lymph to the membrane-bound G protein-coupled olfactory receptor proteins. The roles of PBPs in molecular recognition and the mechanisms of pheromone binding and release are poorly understood. Here, we report the NMR structure of a PBP from the giant silk moth *Antheraea polyphemus*. This is the first structure of a PBP with specific acetate-binding function *in vivo*. The protein consists of nine α-helices: α1a (residues 2–5), α1b (8–12), α1c (16–23), α2 (27–34), α3a (46–52), α3b (54–59), α4 (70–79), α5 (84–100) and α6 (107–125), held together by three disulfide bridges: 19–54, 50–108 and 97–117. A large hydrophobic cavity is located inside the protein, lined with side-chains from all nine helices. The acetate-binding site is located at the narrow end of the cavity formed by the helices α3b and α4. The pheromone can enter this cavity through an opening between the helix α1a, the C-terminal end of the helix α6, and the loop between α2 and α3a. We suggest that Trp37 may play an important role in the initial interaction with the ligand. Our analysis also shows that Asn53 plays the key role in recognition of acetate pheromones specifically, while Phe12, Phe36, Trp37, Phe76, and Phe118 are responsible for non-specific binding, and Leu8 and Ser9 may play a role in ligand chain length recognition.

Keywords: Antheraea polyphemus; PBP; pheromone-binding protein; olfaction; NMR structure

Introduction

Olfaction is an extremely important process for a large number of animals. Indeed, their survival depends on the recognition and discrimination of chemical signals from the environment. Some insects, particularly Lepidoptera male moths, have an exquisitely sensitive olfactory system that is capable of perceiving airborne pheromone molecules released by the females and responding to them over great distances.[1] They are capable of distinguishing between closely related pheromones of different species. The organs responsible for olfaction are the sensory hair (sensilla trichoidea) on the surface of the moth's antennae. Each sensillum contains several dendrite endings of the olfactory neurons surrounded by aqueous protein-rich lymph. Lepidopteran pheromones are volatile hydrophobic molecules, insoluble in aqueous media. Upon entering the sensillum, the pheromone binds to a small water-soluble carrier protein, pheromone-binding protein (PBP), which transports it to the dendrite membrane. This is the first step in a series of sequential events leading to the activation of a G protein-coupled receptor (GPCR) mediated signal transduction cascade in olfaction.[2]

Abbreviations used: PBP, pheromone-binding protein; GPCR, G protein-coupled receptor; OBP, odorant-binding protein; GOBP, general OBP; ApolPBP, *Antheraea polyphemus* PBP; HSQC, heteronuclear single quantum coherence; NOE, nuclear Overhauser effect; IPTG, isopropyl-β-D-thiogalactopyranoside; NOESY, NOE spectroscopy; rms, root-mean-square; rmsd, rms deviation.

E-mail address of the corresponding author: smita.mohanty@sunysb.edu

PBPs belong to a family of proteins called odorant-binding proteins (OBPs). They are highly soluble acidic proteins with molecular masses of 14–16 kDa. While PBPs are found in male moths only, general odorant-binding proteins (GOBPs), the other OBP subfamily, are found in both sexes. GOBPs are believed to respond to a range of volatile plant and food odorants.[3] The PBPs from different moth species share over 50% sequence identity and about 30% identity with GOBPs. All insect OBPs have six strictly conserved cysteine residues (Figure 1). The individual roles of each of these proteins in olfaction are not well understood. Since PBPs can bind pheromone molecules, it has been proposed that they can either transport the hydrophobic pheromone molecules through the aqueous sensillar lymph to the membrane-bound GPCR proteins,[2] or that they can remove the pheromones from the receptors,[4] or have a dual function. It has been also shown that PBP–pheromone complexes, rather than free pheromones, activate the olfactory receptors.[5] PBPs exhibit specificity in binding their respective pheromones,[6,7] although they are sometimes capable of binding other similar molecules.[8,9] It has been suggested that despite their *in vitro* promiscuity, *in vivo* each PBP might be tuned to interact with a specific pheromone in a unique way.[9–11]

The PBP of the giant silk moth *Antheraea polyphemus* (ApolPBP) was the first PBP that was identified by its pheromone binding capability.[12] This protein binds (6*E*,11*Z*)-hexadeca-6,11-dienyl-1-acetate (Figure 2), the most abundant pheromone for *A. polyphemus*. At present three different pheromone-binding proteins have been found in this species[13] (Figure 1). However, no three-dimensional structure has been determined either by X-ray crystallography or by NMR spectroscopy for any of these proteins. ApolPBP has over 50% sequence identity with PBPs from other moth species but differs in substrate specificity.[6,9,10]

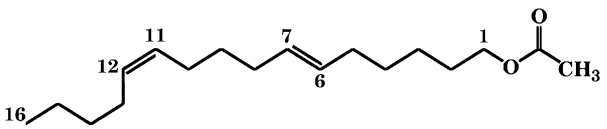

Figure 2. Chemical structure of the pheromone of *Antheraea polyphemus*, (6*E*,11*Z*)-hexadeca-6,11-dienyl-1-acetate.

In order to address the questions of pheromone recognition, specificity, and the mechanism of acetate binding by proteins, we have initiated a detailed structural study of the giant silk moth *A. polyphemus* PBP. We observe a conformational transition for this protein between pH 5.0 and 6.0.[14] It has been reported that ApolPBP binds pheromone only at pH above 6.0.[15] Here, we report the first NMR structure of the pheromone-binding conformation of ApolPBP determined at pH 6.3. This structure reveals the details of the putative acetate-binding site and provides sensible delineation of the roles of the residues involved in pheromone binding.

Results and Discussion

Despite over 50% sequence identity and six strictly conserved cysteine residues, PBPs from different moth species have different substrate specificity.[6] It is therefore of considerable interest to investigate the three-dimensional structures of a variety of PBPs in order to uncover structural and conformational features responsible for pheromone

```
                1        10        20        30        40        50        60        70
ApolPBP   SPEIMKNLSNNFGKAMDQCKDELSLPDSVVADLYNFWKDDYVMTDRLAGCAINCLATKLDVVDPDGNLHH
ApolPBP2  SPEIMKNLCMNYGKTMDQCKQELGLPDSVINDLYNFWKDDYVMTDRLAGCAINCLSTKLDIVDPDGNLHH
ApolPBP3  SQEIMKTMTLTFTKGLDACKKEMDLPDTVDVDFNNFWKEDYVVTNRDAGCAIVCLASKINLVDSMGILIH
BmPBP     SQEVMKNLSLNFGKALDECKKEMTLTDAINEDFYNFWKEGYEIKNRETGCAIMCLSTKLNMLDPEGNLHH
AperPBP1  SPEIIKNLSQNFCKAMDQCKQELNIPDSVIADLYNFWKDDYVMTDRLAGCAINCMATKLDVVDPDGNLHH
HvirPBP   SQDVMKNLSMNFAKPLEDCKKEMDLPDSVTTDFYNFWKEGYEFTNRHTGCASLCLSSKLELLDQEMKLHH
AvelPBP   SQDVIKGMTLNFRKGLDECKKEMNLPDSINADFYNFWKDDHVLSNRDTGCAIMCLSSKLELVS-DGKLHH
OnubPBP   SQDVMKQMTINFGKALDTCRKELDLPDSINADFYNFWKEGYELSNRQTGCAIMCLSSKLDLVDPEGKLHH
CrosPBP   SADVVKGMTLNFGKGLEECKKEMNLPDSINADFYNFWKDDHVLTNRDTGCAIMCLSSKLELVS-DGKLHH
LdisPBP1  SKEVMKQMTINFAKPMEACKQELNVPDAVMQDFFNFWKEGYQITNRE-GCVILCLAKKLELLDQDMNLHH

                80        90        100       110       120       130       140
ApolPBP   GNAKDFAMKHGADETMAQQLVDIIHGCEKSAPP-NDDKCMKTIDVAMCFKKEIHKLNWVPNMDLVIGEVLAEV 142
ApolPBP2  GNAKEFAMKHGADDGMAQQLVDIIHRCEKSTPP-NDDKCTKTMDIAMCFKKEIHKLNWVPNMDLVVCEVLAVV 142
ApolPBP3  GSAHEFAKQHGADDNMAKQLSDTLHTCETIIGT-GNDECTRALHVANCFKVEMHKLDWAPSMDLIIGELLAEI 142
BmPBP     GNAMEFAKKHGADETMAQQLIDIVHGCEKSTPA-NDDKCIWTLGVATCFKAEIHKLNWAPSMDVAVGEILAEV 142
AperPBP1  GNAKEFAMKHGADASMAQQLVDIIHGCEKSAPP-NDDKCMKTIDVAMCFKKEIHKLNWVPDMDVVLGEVLAEV 142
HvirPBP   GKAQEFAKKHGADDAMAKQLVDMIHGCSQSTPDATDDPCMKALNVAKCFKAKIHELNWAPSMELVVGEVLAEV 143
AvelPBP   GNTFDYAKQHGADETVAQQLVDLIHSCEKSLPD-LEDPCMKVLEWAKCFKTEIHKLNWAPSVEVLAAEMLAEV 141
OnubPBP   GNTHEFAKKHGADDSMAKQLVELIHKCEGSVAD-DPDACMKVLNIAKCFKAEIHKLNWAPSMDLIVAEVLAEV 142
CrosPBP   GNTLEYAKQHGADDTVAQQLVDLIHNCEKALPD-LEDPCMKVLEWAKCFKIEIHKLNWAPSMDVLAGEMLAEI 141
LdisPBP1  GKAMEFAMKHGADEAMAKQLLDIKHSCEKVITIVADDPCQTMLNLAMCFKAEIHKLDWAPTLDVAVGKLLADT 143
```

Figure 1. Primary sequences of the PBPs of the moths *Antheraea polyphemus* (GenBank accession numbers X17559, AJ277266, AJ277267), *Bombyx mori* (acc. no. X94987), *Antheraea pernyi* (acc. no. X96773), *Heliothis virescens* (acc. no. X96861) *Argyrotaenia velutinana* (acc. no. AF177641), *Ostrinia nubilalis* (acc. no. AF133643), *Choristoneura rosaceana* (acc. no. AF177652), *Lymantria dispar* (acc. no. AF007867). Conserved residues are shown on yellow background. Residues expected to be involved in ligand binding are indicated with white letters on red background.

binding. The solution structure of ApolPBP determined and analyzed here will aid in this endeavor.

NMR structure of ApolPBP

ApolPBP at pH 6.3 consists of nine α-helices with residues 2–5 (α1a), 8–12 (α1b), 16–23 (α1c), 27–34 (α2), 46–52 (α3a), 54–59 (α3b), 70–79 (α4), 84–100 (α5) and 107–125 (α6). The secondary structure elements are shown in Figure 3. The C^{β} chemical shifts of all six cysteine residues are equal to or larger than 35 ppm, confirming that they are involved in disulfide bridges. The bridge between Cys19 and Cys54 connects helices α2 and α3b, the Cys50-Cys108 disulfide bridge connects α3a and α5, and the helices α5 and α6 are linked *via* Cys97 and Cys117. The helices are packed in a roughly globular structure (residues 1–125) with an unstructured C terminus (residues 126–142) extending into the solvent and a large hydrophobic cavity enclosed inside. Stereo views of the superposition of 20 lowest energy structures and a ribbon diagram of the overall structure are displayed in Figure 4.

The ApolPBP primary sequence contains 17 aspartate, seven glutamate, 13 lysine, one arginine and five histidine residues. Out of these 43 polar amino acid residues, six are involved in the formation of salt-bridges (Asp106-Arg46, Glu22-Lys58 and Glu98-Lys110), two are partially solvent-accessible (His69 and Lys74) and the rest are fully solvent-accessible, which accounts for the protein's high solubility. Closely packed helices have a range of crossing angles: α4–α5 (35°), α5–α6 (62°), α3a–α6 (73°), α3a–α2 (55°), α1c–α3b (78°). The helix α1b is connected to the loop between α3a and α3b with a hydrogen bond, and α1a does not have close contacts with other helices and exhibits increased flexibility as evidenced by weak or missing nuclear Overhauser effect (NOE) resonances.

The structure of the acetate-binding ApolPBP is different from the structure of the alcohol-binding PBP from *Bombyx mori*[16] with an rmsd value of 4.6 Å for the backbone atoms. In the case of ApolPBP the third helix of the ideal hexahelical fold is broken up into two helices: α3a and α3b, with only one non-helical residue, Asn53, at the junction. This feature plays an important role in the ligand-binding site as discussed below. This residue is involved in an N-capping interaction accepting a hydrogen bond from the side-chain of Thr57 to its carbonyl. The helix α1c is tethered to α3b by the Cys19-Cys54 disulfide bridge and a Glu22-Lys58 salt-bridge. Glu22 and Lys58 are strictly conserved through lepidopteran PBPs (Figure 1). The side-chain of Ser9 from the helix α1b forms a hydrogen bond with the carbonyl of Ile52 from α3a. These interactions lead to the kink of 61° between α3a and α3b. This causes the arrangement of the helices around α3 to be different from the ideal helix packing, resulting in the split of the helix α1 into three helices α1a, α1b and α1c.

Several parts of the ApolPBP backbone exhibit increased mobility as apparent from weak or missing NOE resonances resulting in large local rmsd values in the ensemble of structures. The flexible regions comprise helices α1a and α1b (residues 1–11), the loop between α1c and α2 (residues 26–28), part of the loop between α2 and α3a (residues 34–39), part of the helix α3b (residues 54–56), beginning of the loop between α3b and α4 (residues 60–64), and the beginning of the disordered C terminus (residues 126–129).

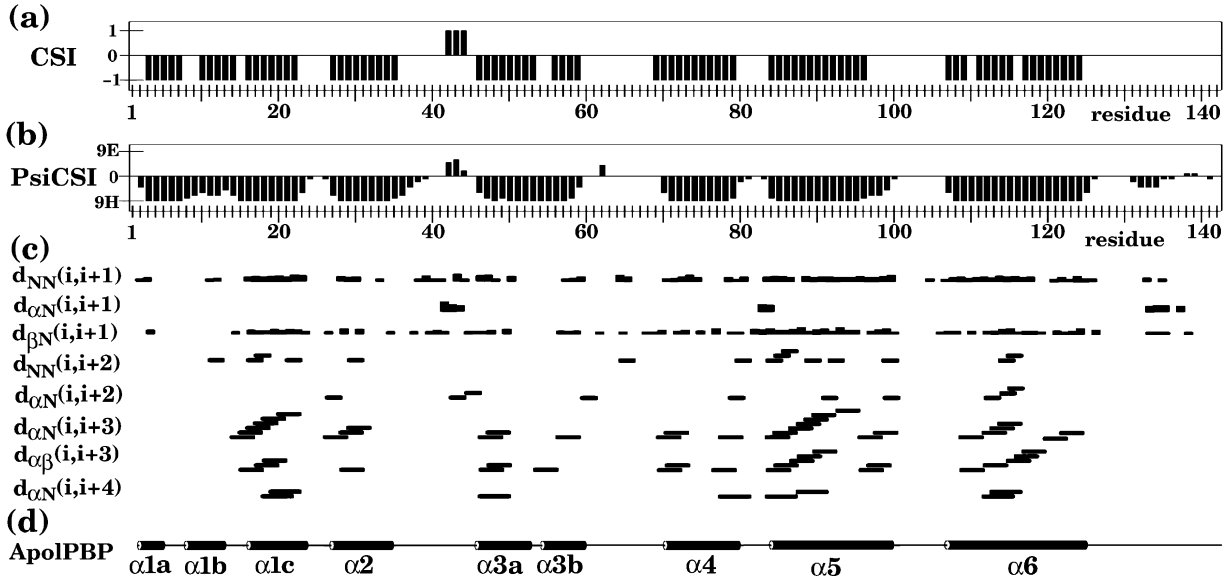

Figure 3. (a) Secondary structure of ApolPBP as predicted by CSI and (b) PsiCSI. Helical conformations are indicated by negative bars, extended conformations by positive. (c) Summary of sequential NOE contacts. (d) Secondary structure of ApolPBP as observed in the final set of energy-minimized conformations.

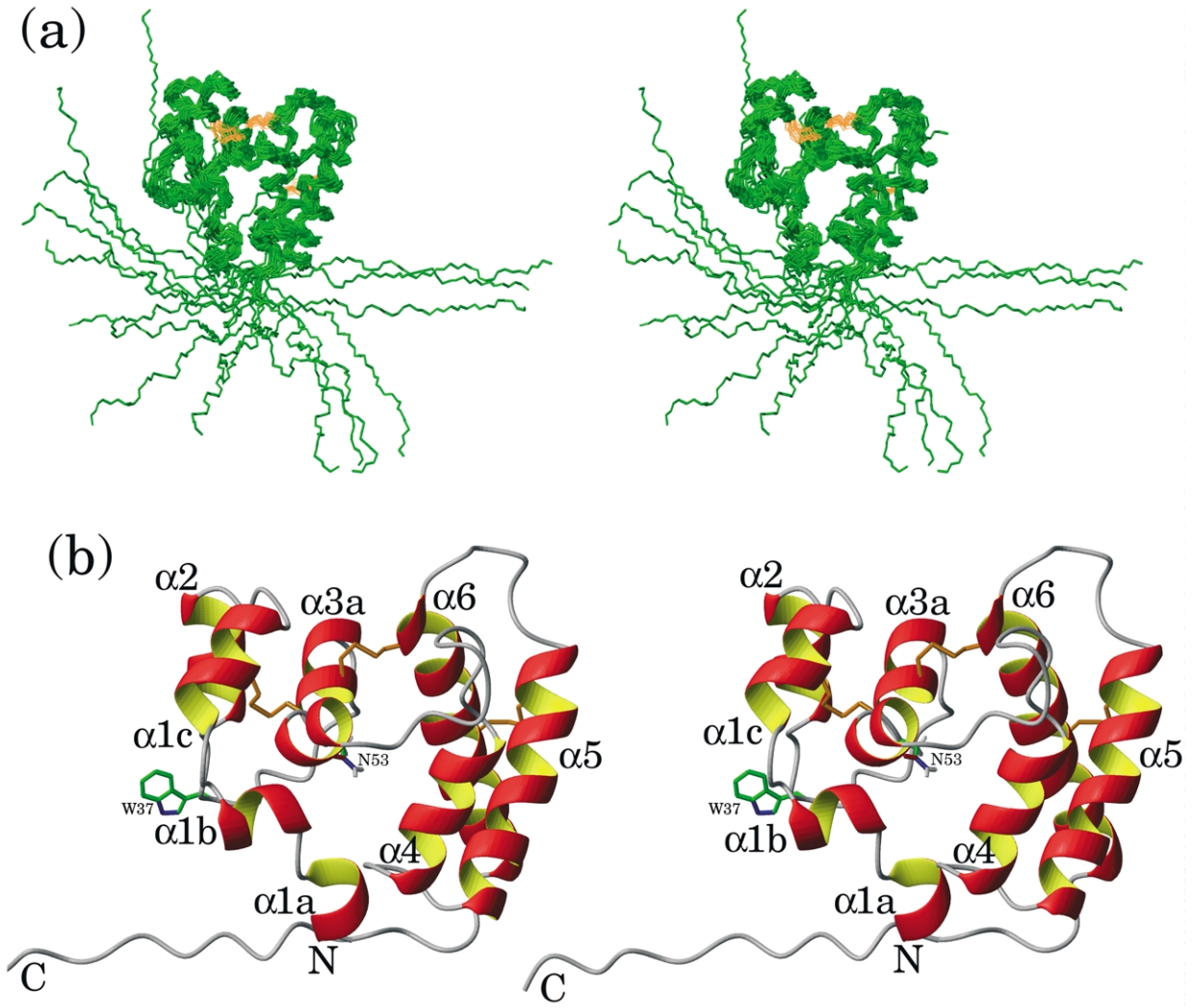

Figure 4. Stereo views of the three-dimensional structures of ApolPBP. (a) Superposition of the 20 energy-minimized and water-refined conformations of ApolPBP. Backbone of the residues 1–125 is shown in green and the disulfide bridges are shown in yellow. (b) Ribbon drawing of one of the ApolPBP structures. Helices, N and C termini, and the residues Asn53 and Trp37 are indicated.

Temperature and pH effect

In order to investigate the effect of temperature and pH on the conformation of the protein, 2D {^{15}N,^{1}H}-HSQC NMR spectra of ApolPBP were recorded over the temperature range 5–40 °C (data not shown) and pH range 4–7 (Figure 5). No significant differences were found in the spectra collected at different temperatures, indicating absence of major conformational changes or denaturation. The heteronuclear single quantum coherence (HSQC) spectra recorded at pH 4, 5, 6, and 7 are displayed in Figure 5. As can be noted, a significantly different pattern of resonances was observed at pH 4 and 5, compared to pH 6 and 7. Comparison of the resonance positions in the spectra shows that only 27% of resonances are in identical positions between pH 5 and 6, while 59% are identical between pH 4 and 5 and 70% between pH 6 and 7. This implies that different conformations are predominant at pH 4 and pH 7, and that a major conformational change occurs between pH 5 and 6. The HSQC spectra at pH 6 and 7 are more dispersed than the spectra at pH 4 and 5, indicating a more structured conformation at basic pH. This conformational change is not confined to a few residues or the residues in the binding cavity, rather an overall conformational change involving most residues. This conformational transition may explain the loss of pheromone binding ability observed for ApolPBP at acidic pH.[15]

Pheromone-binding cavity in ApolPBP

The structure of ApolPBP contains a large, 282(±40) Å^3 hydrophobic cavity (Figure 6(a)). This cavity is lined by the side-chains of 29 residues:

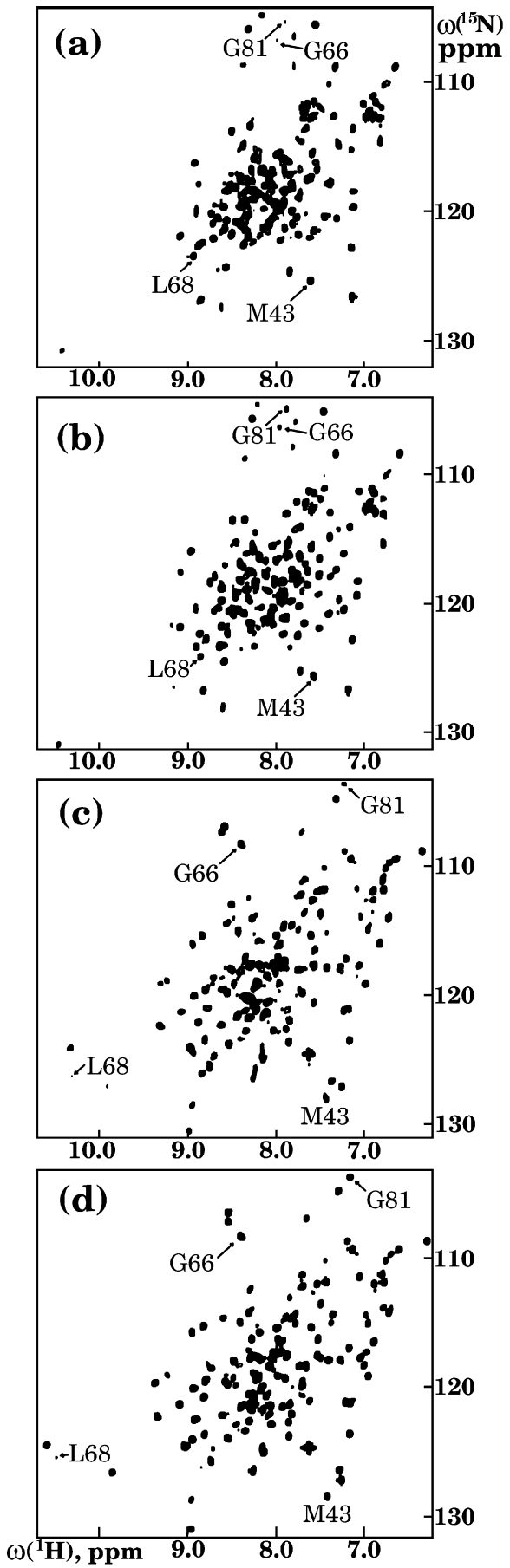

Figure 5. The 2D {^{1}H,^{15}N}-HSQC spectra of ApolPBP: (a) pH 4.0, (b) 5.0, (c) 6.0 and (d) 7.0. Labels indicate some of the residues different between acidic and basic conformations.

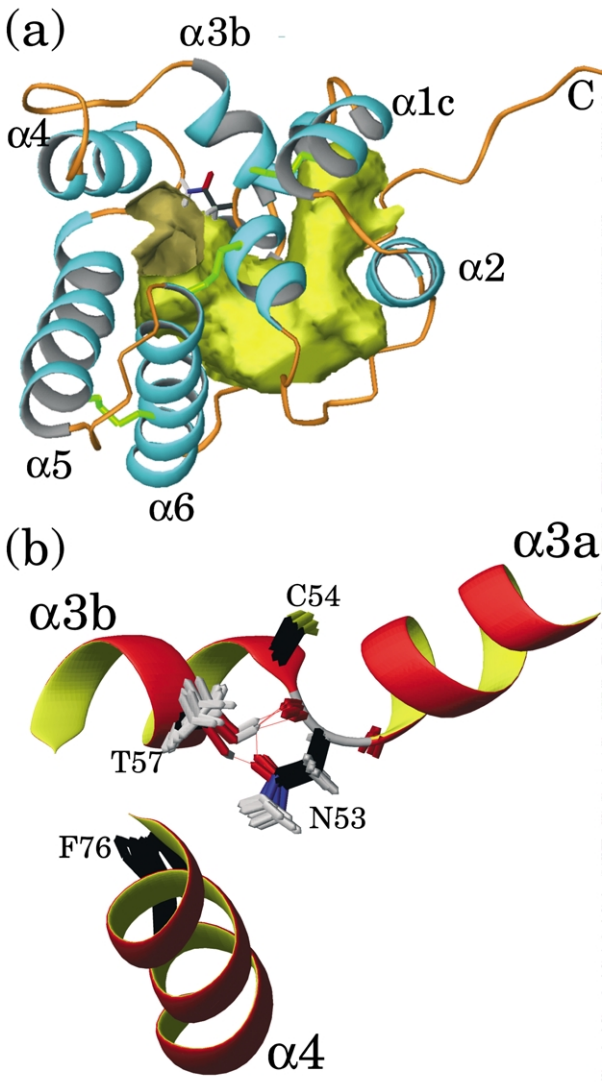

Figure 6. (a) Binding cavity of ApolPBP viewed from the side of the helix α3a. (b) Close-up view of the acetate-binding site of ApolPBP in an ensemble of five NMR structures.

Ile4 and Met5 of the helix α1a, Asn7, Leu8, Ser9 and Phe12 of α1b, Met16 of α1c, Leu33 of α2, Tyr34, Asn35, Phe36 and Met43 from the loop between α2 and α3a, Ala48 and Ile52 of α3a, Asn53, Ala56 and Thr57 of α3b, Val61 from the loop between α3b and α4, Ala73, Lys74, Phe76 and Ala77 of α4, Leu90 and Ile94 of α5, Thr111, Ile112, Ala115, Phe118, Ile122 of α6. Most of these residues are hydrophobic, as expected for a cavity designed to bind hydrophobic pheromones. The cavity is crescent-shaped, wider in the middle and narrow at the ends (Figure 6(a)). There is an

opening of approximately 15 Å^2 between helix α3b and the N-terminal ends of α4 and α6. Another opening of approximately 20 Å^2 is located between helix α1a, the C-terminal end of the helix α6 and the loop between α2 and α3a.

Structure of the acetate-binding site

The three-dimensional structure of ApolPBP is the first structure of a PBP with acetate binding function. This allows for the first time a comparison with the alcohol-binding PBP from *B. mori*, for which a recent *ab initio* study[17] identified nine residues involved in direct interaction with the bound pheromone. These are Leu8, Ser9, Phe12, Phe36, Trp37, Ser56, Met61, Phe76, and Phe118. Of these nine residues only two are not conserved in ApolPBP, namely Ser56 and Met61, both of which form the *B. mori* alcohol-binding site.[17,18] In ApolPBP positions 56 and 61 are occupied by hydrophobic residues alanine and valine, respectively. Interestingly, a comparison of the primary sequences of this polypeptide chain (residues 56–61) among the PBPs from different moth species (Figure 1) does not reveal any obvious amino acid residues suited to distinguish the PBPs that bind acetates from those that bind alcohols or other types of pheromones. However, the region encompassing residues 50–61 in the three-dimensional structure of ApolPBP (Figure 6(b)) clearly shows that the side-chain of Asn53 is located at the end of the hydrophobic cavity and directed towards the anticipated location of the bound pheromone. Asparagine side-chains were shown to form amide-carboxyl hydrogen bonds with acetate ligands in such proteins as esterase-like catalytic antibodies[19] and predicted to play this role in muscarinic acetylcholine receptors.[20,21] Considering this, and the fact that the Asn53 side-chain amide is the only hydrogen bond donor at the narrow end of the hydrophobic cavity of ApolPBP (Figure 6a), we propose that Asn53 is responsible for the interaction with the acetate group of the pheromone. It should also be noted that additional selectivity and stability of the ApolPBP acetate-binding site may be caused by CH/π interaction[22] between the acetate methyl protons and the phenyl ring of Phe76 suitably located at the wall of the cavity between helices α4 and α3b (Figure 6(b)). Since all the binding site residues except those located in the stretch 53–61 are conserved between PBPs with different functions, we further suggest that this region plays the crucial role in the PBP substrate specificity by recognizing the terminal polar group of the ligand.

Roles of the other amino acid residues involved in ligand binding

The five aromatic residues Phe12, Phe36, Trp37, Phe76, and Phe118 are strictly conserved throughout all known lepidopteran PBPs and GOBPs with Phe76 and Phe12 replaced by tyrosine in rare cases. These residues are therefore most clearly responsible for non-specific binding of a wide range of unsaturated aliphatic odorant molecules. In the solution structure of ApolPBP all these aromatic residues are found inside the cavity with the exception of Trp37, which is solvent-accessible and forms a part of the mouth of the second, larger opening. This, together with the fact that the second opening is formed by the polypeptide chains that exhibit considerable mobility (helix α1a, C-terminal end of helix α6 and the loop between α2 and α3a) allows us to present a tentative pheromone-binding scenario. The unsaturated aliphatic pheromone chain interacts with the aromatic side-chain of Trp37 before entering the internal cavity of the PBP through the opening between α1a, α6, and the α2–α3a loop, causing a conformational change in the loop, that takes Trp37 into the cavity. In fact, we observe the same pattern in the alcohol-binding *B. mori* PBP where Trp37 is found outside the cavity in the unliganded protein[16] but is located inside the cavity in the complex,[18] which suggests that our mechanistic proposal may be of general nature.

The two remaining residues expected to interact with the ligand are Leu8 and Ser9. These residues are not conserved among all PBPs. Leu8 is replaced by methionine in several species and Ser9 is sometimes replaced by threonine or cysteine (Figure 1). Comparison of the primary sequences of lepidopteran PBPs (Figure 1) shows that Leu8 and Ser9 are found in the PBPs from the moths *A. polyphemus*, *B. mori*, *Antheraea pernyi*, and *Heliothis virescens*. For these four species the most abundant pheromones have 16-carbon chains. The residues Met8 and Thr9 are found in the PBPs from the moths *Ostrinia nubilalis*, *Argyrotaenia velutinana* and *Choristoneura rosaceana*, and for these species the most abundant pheromones have 14-carbon chains. Given the fact that the side-chains of both methionine and threonine are larger than those of leucine and serine, these bulky side-chains may restrict the available space for the pheromone in the cavity, making it more favorable to have the short 14-carbon chain pheromones in combination with the longer side-chains and the long 16-carbon chain pheromones with the shorter side-chains. However, many moths use a mixture of pheromone compounds, often including both 16 and 14-carbon chains, and therefore other combinations of the residues are also found in positions 8 and 9. Although it is likely that these two residues are involved in pheromone chain length recognition, more experimental data are clearly needed to clarify their roles in detail.

Conclusion

Complete understanding of the chemistry and function of the PBPs requires the combination of many approaches. Here, NMR spectroscopy identified the physiological conformation of ApolPBP,

which gave us a novel reasonable model for pheromone uptake and selective interaction with the protein. Our study not only solves the first acetate-binding PBP structure but also provides a framework for future investigation and analysis of substrate recognition and specificity. The key elements of the PBP structure identified in this study are a step forward towards understanding how acetate ligands interact with proteins. That, in turn, will have a far-reaching impact on both insect control, through pheromone-based integrated pest management programs, and pharmacology, through design and development of novel therapeutic drugs for Alzheimer's disease and other acetylcholine-related disorders.

Materials and Methods

Expression, purification, and refolding of ApolPBP

Recombinant ApolPBP was expressed in *Escherichia coli* XA-90 cells using pHN1+ vector. Saturated overnight LB-ampicillin culture was diluted (1:100, v/v) in minimal medium containing 0.1% (w/v) $^{15}NH_4Cl$ and 0.4%(w/v) D-[^{13}C]glucose (Isotec) and grown at 37 °C to an $A_{600\ nm}$ of 0.5–0.6. Expression was induced with 1 mM IPTG and cells were harvested by centrifugation after overnight incubation at 30 °C. The pellet was suspended in B-PER (Pierce) bacterial protein extraction reagent with lysozyme (100 μg/ml) and lysed with a French Press. Inclusion bodies were collected by centrifugation (14,000 rpm, 30 minutes), resuspended by sonication in B-PER reagent with lysozyme (200 μg/ml) and pelleted again (14,000 rpm, 30 minutes). This pellet was washed by sonication in dilute B-PER reagent and centrifugation three times, resulting in over 95% pure protein in the inclusion body as assessed by SDS-PAGE. Refolding of the protein into its native form was achieved by four different refolding protocols: redox refolding,[15] refolding using non-detergent sulfobetaines (NDSB,[23]) refolding using protein disulfide isomerase (PDI,[24]) and refolding by step dialysis, yielding active protein in all cases. Refolded ApolPBP was purified by preparative isoelectric focusing (IEF) using a Rotofor (Bio-Rad) apparatus and gel-filtration on a Sephacryl S-100 column (Amersham Biosciences). Dynamic light-scattering and size-exclusion chromatography measurements revealed that ApolPBP was homogenous and monomeric at pH 6.3. The refolded protein was found to be able to bind pheromone by photoaffinity labeling with a tritium-labeled photoactivable diazoacetate pheromone analog[15] and transfer NOE experiments.[25]

NMR experiments and data analysis

NMR samples contained ~1 mM uniformly ^{15}N or ^{15}N/^{13}C-labeled ApolPBP in 50 mM phosphate buffer (pH 6.33), 1 mM EDTA, 0.1% (w/v) NaN_3, 95% H_2O and 5% ^2H_2O. All NMR data were collected on Bruker DMX500, DMX600, and DRX800 spectrometers at 35 °C. The following NMR experiments[26] were performed for the purpose of sequential assignment, which we have reported earlier,[14] and structure analysis: 2D {^{15}N,^{1}H}-HSQC, 2D {^{13}C,^{1}H}-HMQC, 3D HNCA, 3D HNCO, 3D HNCACB, 3D CC(CO)NH, 3D HC(CC)(CO)NH, 3D CBCA(CO)NH, 3D HCACO, 3D ^{15}N-edited and ^{13}C-edited {^{1}H,^{1}H}-NOESY. The NOE spectra were collected with mixing times of 110 ms (^{13}C-edited NOESY) and 120 ms (^{15}N-edited NOESY). The ^{13}C carrier frequency was set to 63.5 ppm (^{13}C-edited NOESY) and the ^{15}N carrier frequency was set to 117.2 ppm (^{15}N-edited NOESY). All NMR data were processed with NMRPipe[27] and peak-picked interactively using NMRView[28] resulting in 2682 NOE cross-peak intensities. Secondary structure elements were found using the programs CSI[29] and PsiCSI[30] based on the assigned chemical shifts (Figure 3).

Structure calculation and analysis

The 2171 NOE cross-peaks were assigned manually using NMRView. The assignments were confirmed and/or corrected with CANDID,[31] a module of CYANA, using the standard protocol consisting of seven cycles of iterative NOE assignments and structure determination. A total of 243 dihedral angle restraints were derived with the program TALOS[32] based on the ApolPBP primary sequence and chemical shifts. A total of 120 hydrogen bond distance restraints (two restraints per bond) were derived from the CSI output by ARIA.[33] These derived restraints were added to the CANDID restraints for each iteration cycle. During the iterative NOE assignments 419 NOEs were corrected, 66 were removed and 396 were added, yielding a final count of 2501 assigned NOE cross-peaks (Table 1). The experimental upper distance restraints were derived from the intensities of the assigned NOE signals with the program CYANA[34] using two calibration functions: d^{-6} for the backbone of the residues 1–134 and d^{-4} for the rest of the backbone and for all side-chain resonances. The structure of ApolPBP was calculated by torsion angle dynamics simulated annealing as implemented in the program

Table 1. Restraints and structural statistics for ApolPBP

Property	Value
Restraints	
Total experimental distance restraints	2501
Intraresidue ($\|i-j\|=0$)	1552
Sequential ($\|i-j\|=1$)	553
Medium range ($2 \le \|i-j\| \le 4$)	258
Long range ($\|i-j\| \ge 5$)	138
Dihedral angle restraints	243
Hydrogen bond distance restraints[a]	120
Residual restraints violations after simulated annealing	
Distance restraint violations greater than 0.1 Å	22 ± 2
Maximal distance restraint violation (Å)	0.29
Dihedral angle restraint violations greater than 2.5°	17 ± 4
Maximal dihedral angle violation (°)	5.27
rms deviations from the averaged coordinates (Å)	
Backbone of the regular secondary structure	0.73
All heavy atoms of the regular secondary structure	1.12
Backbone of the residues 1–125	0.85
All heavy atoms of the residues 1–125	1.25
Ramachandran plot statistics (%)	
Most favored regions	82.3
Additionally allowed regions	16.1
Generously allowed regions	1.0
Disallowed regions[b]	0.6

[a] Two distance restraints per bond.
[b] All residues found in the disallowed regions belong to the flexibly disordered C terminus.

CYANA. The length of the annealing procedure was increased to 16,000 steps and the starting temperature was set to 20 target function units per degree of freedom. The calculation was started with 200 randomized conformers and the 50 best calculated CYANA structures were used to review and correct the TALOS and CSI-predicted restraints. The structure calculation was repeated and the 50 best structures were energy-minimized in a thin layer of water using the protocol implemented in ARIA.[35] The 20 structures with the lowest potential energy and the best Ramachandran plot scores (Figure 4(a)) were selected for further analysis.

Statistics and visualization

Visualization, root-mean-square distance, hydrogen bond and helix packing angle calculations were performed with the program MOLMOL.[36] Ramachandran plot statistics were calculated by PROCHECK.[37] The internal cavity was analyzed with the program VOIDOO,[38] and visualized with RIBBONS.[39]

PDB and BMRB accession codes

The atomic coordinates have been deposited in the Protein Data Bank† (accession code 1QWV). The assigned chemical shifts have been deposited in the BioMagResBank‡ (accession code 5689).

Acknowledgements

All NMR data were collected in the laboratory of Angela M. Gronenborn, Laboratory of Chemical Physics, NIDDK, NIH. This research was supported by USDA grants 99-35302-8106, 2003-35302-12930 (PECASE program), and NSF grants IBN-0074591 (to S.M.); and in part by the intramural AIDS targeted antiviral program of the office of the Director of NIH (to A.M.G.). We thank Professor Mark Girvin for many useful discussions and kind help, Drs Ramón Campos-Olivas, Bindi Dangi, and Keyang Ding for help with NMR experiment setup, Professor Craig C. Malbon and Professor Steven O. Smith for critical reading of the manuscript.

References

1. Roelofs, W. L. (1995). Chemistry of sex attraction. *Proc. Natl Acad. Sci., USA*, **92**, 44–49.
2. Breer, H., Boekhoff, I. & Tareilus, E. (1990). Rapid kinetics of second messenger formation in olfactory transduction. *Nature*, **345**, 65–68.
3. Vogt, R. G., Prestwich, G. D. & Lerner, M. R. (1991). Odorant-binding-protein subfamilies associate with distinct classes of olfactory receptor neurons in insects. *J. Neurobiol.* **22**, 74–84.
4. Pelosi, P. & Maida, R. (1995). Odorant-binding proteins in insects. *Comp. Biochem. Physiol. ser. B*, **111**, 503–514.
5. Pophof, B. (2002). Moth pheromone binding proteins contribute to the excitation of olfactory receptor cells. *Naturwissenschaften*, **89**, 515–518.
6. Du, G. & Prestwich, G. D. (1995). Protein structure encodes the ligand binding specificity in pheromone binding proteins. *Biochemistry,* **34**, 8726–8732.
7. Maïbèche-Coisné, M., Sobrio, F., Delaunay, T., Lettere, M., Dubroca, J., Jacquin-Joly, E. & Nagnan-Le Meillour, P. (1997). Pheromone binding proteins of the moth *Mamestra brassicae*: specificity of ligand binding. *Insect Biochem. Mol. Biol.* **27**, 213–221.
8. Oldham, N. J., Krieger, J., Breer, H. & Svatos, A. (2001). Detection and removal of an artifact fatty acid from the binding site of recombinant *Bombyx mori* pheromone-binding protein. *Chem. Senses*, **26**, 529–531.
9. Campanacci, V., Krieger, J., Bette, S., Sturgis, J. N., Lartigue, A., Cambillau, C. *et al.* (2001). Revisiting the specificity of *Mamestra brassicae* and *Antheraea polyphemus* pheromone-binding proteins with a fluorescence binding assay. *J. Biol. Chem.* **276**, 20078–20084.
10. Mohl, C., Breer, H. & Krieger, J. (2002). Species-specific pheromonal compounds induce distinct conformational changes of pheromone binding protein subtypes from *Antheraea polyphemus*. *Invert. Neurosci.* **4**, 165–174.
11. Bette, S., Breer, H. & Krieger, J. (2002). Probing a pheromone binding protein of the silkmoth *Antheraea polyphemus* by endogenous tryptophan fluorescence. *Insect Biochem. Mol. Biol.* **32**, 241–246.
12. Vogt, R. G. & Riddiford, L. M. (1981). Pheromone binding and inactivation by moth antennae. *Nature*, **293**, 161–163.
13. Maida, R., Krieger, J., Gebauer, T., Lange, U. & Ziegelberger, G. (2000). Three pheromone-binding proteins in olfactory sensilla of the two silkmoth species *Antheraea polyphemus* and *Antheraea pernyi*. *Eur. J. Biochem.* **267**, 2899–2908.
14. Mohanty, S., Zubkov, S. & Campos-Olivas, R. (2003). ^{1}H, ^{13}C and ^{15}N backbone assignments of the pheromone binding protein from the silk moth *Antheraea polyphemus* (ApolPBP). *J. Biomol. NMR*, **27**, 393–394.
15. Prestwich, G. D. (1993). Bacterial expression and photoaffinity labeling of a pheromone binding protein. *Protein Sci.* **2**, 420–428.
16. Lee, D., Damberger, F. F., Peng, G., Horst, R., Güntert, P., Nikonova, L. *et al.* (2002). NMR structure of the unliganded *Bombyx mori* pheromone-binding protein at physiological pH. *FEBS Letters*, **531**, 314–318.
17. Klusák, V., Havlas, Z., Rulísek, L., Vondrásek, J. & Svatos, A. (2003). Sexual attraction in the silkworm moth. Nature of binding of bombykol in pheromone binding protein–an *ab initio* study. *Chem. Biol.* **10**, 331–340.
18. Sandler, B. H., Nikonova, L., Leal, W. S. & Clardy, J. (2000). Sexual attraction in the silkworm moth: structure of the pheromone-binding-protein–bombykol complex. *Chem. Biol.* **7**, 143–151.
19. Gigant, B., Charbonnier, J. B., Eshhar, Z., Green, B. S. & Knossow, M. (1997). X-ray structures of a hydrolytic antibody and of complexes elucidate catalytic pathway from substrate binding and transition state stabilization through water attack and product release. *Proc. Natl Acad. Sci. USA*, **94**, 7857–7861.
20. Blüml, K., Mutschler, E. & Wess, J. (1994). Functional

† http://www.rcsb.org
‡ http://www.bmrb.wisc.edu

role in ligand binding and receptor activation of an asparagine residue present in the sixth transmembrane domain of all muscarinic acetylcholine receptors. *J. Biol. Chem.* **29**, 18870–18876.
21. Huang, X. P., Nagy, P. I., Williams, F. E., Peseckis, S. M. & Messer, W. S., Jr (1999). Roles of threonine 192 and asparagine 382 in agonist and antagonist interactions with M_1 muscarinic receptors. *Br. J. Pharmacol.* **126**, 735–745.
22. Nishio, M., Umezawa, Y., Hirota, M. & Takeuchi, Y. (1995). The CH/π interaction: significance in molecular recognition. *Tetrahedron*, **51**, 8665–8701.
23. Vuillard, L., Rabilloud, T. & Goldberg, M. E. (1998). Interactions of non-detergent sulfobetaines with early folding intermediates facilitate *in vitro* protein renaturation. *Eur. J. Biochem.* **256**, 128–135.
24. Yao, Y., Zhou, Y. & Wang, C. (1997). Both the isomerase and chaperone activities of protein disulfide isomerase are required for the reactivation of reduced and denatured acidic phospholipase A_2. *EMBO J.* **16**, 651–658.
25. Clore, G. M. & Gronenborn, A. M. (1982). Theory and applications of the transferred nuclear Overhauser effect to the study of conformations of small ligands bound to proteins. *J. Magn. Reson.* **48**, 402–417.
26. Ferentz, A. & Wagner, G. (2000). NMR spectroscopy: a multifaceted approach to macromolecular structure. *Quart. Rev. Biophys.* **33**, 29–65.
27. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfifer, J. & Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, **6**, 277–293.
28. Johnson, B. A. & Blevins, R. A. (1994). NMRView: a computer program for the visualization and analysis of NMR data. *J. Biomol. NMR*, **4**, 603–614.
29. Wishart, D. S. & Sykes, B. D. (1994). The ^{13}C chemical-shift index: a simple method for the identification of protein secondary structure using ^{13}C chemical-shift data. *J. Biomol. NMR*, **4**, 171–180.
30. Hung, L. H. & Samudrala, R. (2003). Accurate and automated classification of protein secondary structure with PsiCSI. *Protein Sci.* **12**, 288–295.
31. Herrmann, T., Güntert, P. & Wüthrich, K. (2002). Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J. Mol. Biol.* **319**, 209–227.
32. Cornilescu, G., Delaglio, F. & Bax, A. (1999). Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR*, **13**, 289–302.
33. Linge, J. P., Habeck, M., Rieping, W. & Nilges, M. (2003). ARIA: automated NOE assignment and NMR structure calculation. *Bioinformatics*, **19**, 315–316.
34. Güntert, P., Mumenthaler, C. & Wüthrich, K. (1997). Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.* **273**, 283–298.
35. Linge, J. P., Williams, M. A., Spronk, C. A., Bonvin, A. M. & Nilges, M. (2003). Refinement of protein structures in explicit solvent. *Proteins: Struct. Funct. Genet.* **50**, 496–506.
36. Koradi, R., Billeter, M. & Wüthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**, 51–55.
37. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK–a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283–291.
38. Kleywegt, G. J. & Jones, T. A. (1994). Detection, delineation, measurement and display of cavities in macromolecular structures. *Acta Crystallog. sect. D*, **50**, 178–185.
39. Carson, M. (1997). RIBBONS. *Methods Enzymol.* **277**, 493–505.

Edited by M. F. Summers

(Received 24 November 2003; received in revised form 23 December 2003; accepted 7 January 2004)