



ELSEVIER

Biochimica et Biophysica Acta 1371 (1998) 317–334



## Secondary structure of P-glycoprotein investigated by circular dichroism and amino acid sequence analysis

Maoqing Dong<sup>1</sup>, Laurent Ladavière, François Penin<sup>\*</sup>, Gilbert Deléage, Loris G. Baggetto<sup>2</sup>

*Institut de Biologie et Chimie des Protéines, UPR 412 CNRS, 7 Passage du Vercors F-69367, Lyon Cedex 07, France*

Received 12 November 1997; revised 17 February 1998; accepted 5 March 1998

### Abstract

P-glycoprotein (Pgp) is a plasma membrane protein known as an ATP-dependent drug-efflux pump that confers multidrug resistance to tumor cells. Structural analysis of Pgp was investigated by circular dichroism (CD) for the first time and in combination with amino acid sequence analysis. CD of highly purified Pgp from human, rat and murine Pgp-overexpressing drug resistant cells revealed slight variations in the spectral shape when recorded in the presence of dodecyl maltoside (DM). These species-dependent variations in CD shapes resulted from the interaction of the oligosaccharidic part with the protein core since they were abolished either in the presence of sodium dodecyl sulfate (SDS) or after deglycosylation, the latter not altering the Pgp ATP-dependent drug transport activity. Whatever the level of Pgp glycosylation and the detergent used (SDS or DM), the content in secondary structure deduced from deconvolution of CD spectra is almost the same for the three sources of Pgp and estimated to 43%  $\alpha$ -helix, 16%  $\beta$ -sheet, 15%  $\beta$ -turn and 26% of other structures. These data, which constitute the first report of Pgp structure analysis by circular dichroism, are consistent with the 48%  $\alpha$ -helix and 16%  $\beta$ -sheets global contents predicted by using recently reported efficient secondary structure prediction methods. This consistency reinforces the reliability of the probable nature and localization of predicted Pgp secondary structure elements. This provides a good framework for precise 3D structure modeling of Pgp by homology with proteins of known 3D structure, as it is illustrated here for the A motifs of the ATP-binding domains of Pgp. © 1998 Elsevier Science B.V. All rights reserved.

**Keywords:** P-glycoprotein; Multidrug resistance; Secondary structure; Nucleotide-binding domain; Circular dichroism; Sequence analysis

Abbreviations: ADR, adriamycin; CD, circular dichroism; COL, colchicin; DM, dodecyl maltoside (1-*O*-*n*-dodecyl- $\beta$ -D-glucopyranosyl(1->4) $\alpha$ -D-glucopyranoside); DTT, dithiothreitol; FTIR: Fourier transform infrared spectroscopy; HAP-HPLC, ceramic hydroxyapatite HPLC; HPLC, high performance liquid chromatography; MDR, multidrug resistance; PA, L- $\alpha$ -phosphatidic acid; PAGE, polyacrylamide gel electrophoresis; PC, L- $\alpha$ -phosphatidylcholine; Pgp, P-glycoprotein; PNGase F, peptide-N (*N*-Acetyl- $\beta$ -glucosaminyl) asparagine amidase; SDS, sodium dodecyl sulfate; VLB, vinblastine

<sup>\*</sup> Corresponding author. Fax: +33-4-72-72-26-26; E-mail: f.penin@ibcp.fr

<sup>1</sup> Present address: Center for Basic Research in Digestive Diseases, Gugg 17, Mayo Clinic, Rochester, MN 55905, USA. Recipient of postdoctoral fellowships from CNRS and Fondation pour la Recherche Médicale.

<sup>2</sup> Also corresponding author: E-mail: lg.baggetto@ibcp.fr

## 1. Introduction

P-glycoprotein (Pgp) is a plasma membrane protein overexpressed in cancer cells which acquired multidrug resistance after prolonged treatment with anticancer drugs. Pgp is encoded by a small family of two genes in human (*MDR1* and *MDR2*), and three genes in rodents (*mdr1a* or *mdr3*, *mdr1b* or *mdr1* and *mdr2*), but only *mdr1* genes encode Pgp that is involved in the MDR phenotype [1,2]. Human Pgp is composed of 1280 amino acids while murine and rat Pgp contain 1277 amino acids. Pgp is an *N*-glycosylated protein but its level of glycosylation depends on the species: murine Pgp is poorly glycosylated (apparent mol.wt. of 140 kDa) in comparison to rat and human Pgp (apparent mol.wt. of 170 kDa, [3]). It is believed that Pgp functions as an ATP-dependent drug efflux pump that extrudes drugs out of cells. This has been evidenced *in vitro*, on either cultured cells or isolated plasma membrane vesicles or reconstituted Pgp liposomes ([3], and the references therein). Glycosylation of Pgp has been generally reported not to be directly required for Pgp functioning [4]. However, a recent study indicates that glycosylation may be involved in drug transportation in resistant cells [5], thus raising again the question of the role of glycosylation. A couple of models for the function of Pgp has been proposed [6], but the mechanisms by which Pgp reduces the intracellular drug concentration and the process of ATP coupling to drug transport remain unclear. The identification of structure/function relationships in Pgp has relied upon the scrutiny of Pgp primary sequence and on biochemical and genetic analysis. Pgp consists of two halves, each containing six predicted hydrophobic membrane-spanning segments followed by a cytoplasmic domain including the nucleotide-binding domain (NBD) with the characteristic A and B motifs described by Walker et al. [7]. Putative structural models for the NBDs of Pgp based on the known crystal structures of other NBDs have also been proposed [8,9]. The sequences and domain organization of Pgp are typical of the ATP-binding cassette (ABC) superfamily of active transporters including yeast [7] and protozoan parasite [10] drug transporters, a series of different members from eukaryotic proteins (like the cystic fibrosis gene product CFTR—cystic fibrosis transmembrane conductance regula-

tor), and the recently found basolateral CFTR counterpart EBCR (epithelial basolateral chloride conductance regulator, [11]) to bacterial transporters [12]. Recent structural analyses by electron microscopy provide the first view of the 3D structure of both detergent-solubilized and lipid-reconstituted purified Pgp [13]. Although the resolution was low (2.5 nm), this initial structure revealed the presence of a large central aqueous chamber which appeared closed to the cytoplasmic face and open to the lipid phase. Moreover, the two nucleotide binding domains have been attributed to two lobes exposed to the cytoplasmic face of the membrane. Recent FTIR investigation of purified Pgp reconstituted into liposomes provides an estimation of Pgp secondary structure, with 32%  $\alpha$ -helix, 26%  $\beta$ -sheet, and 13% unordered structure [14]. In addition, Pgp secondary structure was not significantly altered in the presence of ATP and drugs while these compounds induced conformational changes of Pgp [14]. Although these studies provided important information, more detailed structural data are necessary for understanding the molecular mechanisms of drug transportation by Pgp. A prerequisite for Pgp structural and functional studies is the availability of a large amount of highly purified and active Pgp. We have recently developed an efficient procedure to prepare pure Pgp (> 99%) by combining SDS solubilization and ceramic hydroxyapatite chromatography [3]. After exchange of SDS with dodecyl maltoside and reconstitution into liposomes [15], Pgp exhibited both ATPase and drug transport activities. This preparation of Pgp is virtually devoid of lipids and it is therefore particularly useful for structural studies such as circular dichroism (CD).

CD spectroscopy is an excellent tool for monitoring protein and peptide conformation and the modifications that may occur in response to environmental changes. The far-UV CD of proteins is sensitive to their secondary structure and can therefore be used to determine the relative amounts of the different secondary structural elements. For soluble proteins, CD spectra are typically analyzed by deconvolution using a basic set of spectra to extract the percentages of  $\alpha$ -helix,  $\beta$ -sheet,  $\beta$ -turn and other conformations [for reviews, see Refs. [16,17]]. Analysis of CD spectra of membrane proteins in the presence of lipids is complicated by differential light scattering and absorption flattering. However, membrane proteins solubilized

by detergents generally present no special difficulties and CD data are analyzed as soluble proteins [see reviews [17,18]]. A particular difficulty arises from the presence of protein glycosylation which may change the CD spectra [19] and thus analysis of CD spectra of Pgp should be performed carefully.

CD spectroscopy has been greatly improved in recent years and it is possible to analyze the spectrum of a protein for secondary structure with high reliability. One of the powerful methods of analysis is the variable selection method (VARSELEC [16,20,21]): it uses singular decomposition for deconvolution [22] to systematically analyze the effect that each protein which belongs to a basic set exerts on the secondary structure of the sample protein. Reference proteins that present unwanted CD contributions are discarded so as to reach a total secondary structure close to 100%. This method has been reported to reach correlation coefficients of 0.97 for  $\alpha$ -helix, 0.76 for  $\beta$ -sheet, 0.49 for  $\beta$ -turn and 0.86 for other structures when the estimation of secondary structure from CD data obtained for wavelengths down to 178 nm is compared to X-ray structure data for the sixteen proteins in the data base [21]. Although CD analysis gives an indication of the percentage of secondary structure, it does not provide any evidence about the localization of structured stretches along the amino acid sequence. In this context, the analysis of the structure of a protein can be usefully completed using secondary structure prediction methods.

Secondary structure prediction methods have made significant progresses in recent years by taking into account the multiple alignment of related proteins and thanks to the increasing number of resolved protein structures. The most efficient methods are now reliable at a 70% level of accuracy for three conformational states [for a review, see Ref. [23]]. By using the combination of two methods based either on neural networks (PHD, [24]) or on Self-Optimized Prediction from Alignments (SOPMA), the accuracy rises up to 82% for jointly predicted residues [25]. Thus, the secondary structure predictions are now reliable enough to derive relevant structural properties. Moreover, it is well admitted that a protein, which presents high homologies in sequence and in secondary structure with a protein of known 3D structure, likely adopts the same global fold [26].

Hence, modeling by structure homology can be undertaken.

In the present study, we report for the first time the CD analysis of purified, detergent-solubilized human, rat and murine Pgp. It allows us to specify the effect of glycosylation on Pgp structure and its relationship with Pgp functioning. Moreover, deconvolution of CD spectra allowed the estimation of the secondary structure contents of Pgp which is consistent with that obtained by sequence analyses. This reinforces the reliability of the Pgp secondary structure prediction data that are also reported here and which constitute a good basis for further structure–function relationship studies and molecular modeling.

## 2. Materials and methods

### 2.1. Materials

Cell culture media (DMEM and RPMI-1640) were from BioWHITTAKER and fetal calf serum was from Sigma. Dodecyl maltoside, tunicamycin, egg yolk L- $\alpha$ -phosphatidic acid (PA), egg yolk L- $\alpha$ -phosphatidylcholine (PC) and routine chemicals were from Sigma. Recombinant PNGase F was from Genzyme (Cambridge, MA). Sodium dodecyl sulfate (SDS) was from Serva. Prepacked ceramic hydroxyapatite HPLC column (0.7  $\times$  5.2 cm) and Bio-beads SM2 were from Bio-Rad. HPLC was performed on a Waters apparatus equipped with a 991 Photodiode Array Detector.

### 2.2. Cell culture

The highly drug-resistant murine lymphoid leukemia P388/ADR25 (resistant to 25  $\mu$ g/ml adriamycin), rat hepatoma AS30-D/COL10 (resistant to 10  $\mu$ g/ml colchicin), and human lymphoblastic leukemia CEM/VLB5 (resistant to 5  $\mu$ g/ml vinblastine) cell lines were established and maintained as described previously [3]. To inhibit the N-linked glycosylation of Pgp, AS30-D/COL10 cells were grown in the presence of 2  $\mu$ g/ml tunicamycin for 5 days prior to harvesting.

### 2.3. Protein purification and detergent exchange

Preparation of plasma membranes from highly resistant cell lines and purification of Pgp by SDS solubilization and hydroxyapatite chromatography were described previously [3,15]. This procedure yields large amounts of highly purified Pgp (> 99%) and was also used to prepare pure non-glycosylated Pgp from tunicamycin-treated AS30-D/COL10 cells. The removal of SDS from Pgp was done by exchanging it with dodecyl maltoside using ceramic hydroxyapatite chromatography according to the procedure reported elsewhere [3,15]. In brief, Pgp purified in the presence of 1% SDS was loaded onto the hydroxyapatite column equilibrated with 50 mM sodium phosphate (pH 7.0) containing 0.1% DM and 1 mM DTT. After extensive washing with 0.22 M phosphate for 40 min, Pgp was eluted with 0.7 M phosphate containing 0.1% DM and 1 mM DTT. This procedure allows the complete removal of SDS from Pgp [15]. For the preparation of samples for CD measurements, the eluted Pgp fractions were desalted and concentrated by ultrafiltration using Centricon-30 devices (Amicon, 30 kDa cut-off).

### 2.4. Enzymatic deglycosylation of Pgp

To remove oligosaccharides from purified Pgp, 100  $\mu$ g rat AS30-D/COL10 Pgp (0.2 mg/ml) in 10 mM sodium phosphate (pH 7.0) containing 1 mM DTT, 1 mM phenylmethanesulfonyl fluoride and 0.1% DM, were incubated for 18 h with 10 U/ml PNGase F at 37°C. The protein sample was then applied onto the ceramic hydroxyapatite column equilibrated with 0.22 M phosphate (pH 7.0) containing 0.1% DM and 1 mM DTT. Pgp was bound to the column whereas PNGase F and the cleaved oligosaccharides passed through the column. After extensive washing of the column (40 min), deglycosylated Pgp was eluted as reported above for detergent exchange.

### 2.5. Reconstitution and activity assays

Pgp was reconstituted into liposomes by the method developed by Rigaud et al. [27] as detailed previously [3]. In brief, 1.375 ml of buffer containing 20 mM

Tris-HCl, pH 7.4, 75 mM NaCl, 1 mM DTT and 0.5 mM EDTA was added to a dry lipid film (PC 25 mg, PA 2.5 mg) and the mixture was sonicated on ice until the milky phase became transparent (about 15 min). To a 100  $\mu$ l aliquot of this preparation, 610  $\mu$ l of buffer and 40  $\mu$ l of 10% dodecyl maltoside were sequentially added under constant stirring at room temperature. After 15 min, 250  $\mu$ l of 0.2 mg/ml purified Pgp were slowly added and the mixture was incubated under constant agitation for another 15 min. Bio-beads were added stepwise to remove the detergent: 40 mg for 45 min, 40 mg for 45 min, 40 mg for 30 min and finally, 40 mg for 30 min. Measurements of Pgp activities were done on aliquots of the supernatant containing reconstituted proteoliposomes. The assay of Pgp ATPase activity was measured by the NADH fluorimetric method detailed previously [3]. Drug transport activity of reconstituted Pgp was assayed using the protocol described by Horio et al. [28].

### 2.6. Analytical methods

Protein samples were analyzed by SDS-PAGE on an 8% polyacrylamide gel under reducing conditions as described by Laemmli [29] and detailed previously [30] except that samples were not heated before loading. Gels were silver-stained according to Bloom et al. [31]. Protein concentration was estimated by the colorimetric method of Peterson [32] which allows the measurement of protein in the presence of SDS or DM using bovine serum albumin as the reference protein. Data were cross-checked by UV absorption spectroscopy at 280 nm using molar extinction coefficients of 1536 and 5600 for tyrosine and tryptophan residues, respectively [33], and taking into account the Pgp contents in these amino acids for each species.

### 2.7. CD measurement

CD spectra were recorded on a Jobin-Yvon CD6 spectrometer calibrated with ammonium D-10-camphorsulfonate according to Jobin-Yvon technical specifications. All calculations were made using the CD6 spectrum processing software. Protein samples at a concentration of 0.2–0.8  $\mu$ M were scanned at

room temperature in dichroically neutral quartz cuvettes (Hellma) with a path length of either 0.1 or 0.02 cm. Spectra were recorded in the wavelength range 183–250 nm or 190–250 nm in 0.2 nm increments with a 2 s integration time. Spectra were automatically averaged over eight scans. Sample spectrum without protein was subtracted and the baseline-corrected spectra were smoothed by using a third-order least squared polynomial fit. Following baseline correction, the observed ellipticity was converted into a mean residue molar ellipticity  $[\theta]$  ( $\text{deg cm}^2 \text{ dmol}^{-1}$ ), using the relationship  $[\theta] = \theta / (lcN)$ , where  $\theta$  is the observed ellipticity,  $l$  is the path length in millimeters,  $c$  is the peptide molar concentration, and  $N$  is the number of residues in the protein. All calculations were performed using the CD6 spectrum processing software.

### 2.8. Evaluation of protein secondary structure from CD spectra

The secondary structure percentages were estimated from the measured ellipticities using the Dicroprot software (available by anonymous ftp at ftp.ibcp.fr) into which the Varselec method [16,20,21] has been integrated. The VARSLC1 program was kindly provided by W.C. Johnson Jr. In brief, this deconvolution method assumes that the protein CD spectrum is a sum of the contributions of five secondary structures scaled by the mole fraction:  $\alpha$ -helix, parallel and anti-parallel  $\beta$ -sheets,  $\beta$ -turn and other. To accurately estimate the percentage of each secondary structure in the analyzed protein, reference proteins which exhibit features contributing to the CD that are not found in the sample protein should be eliminated from the basic set of 33 reference proteins of known secondary structure. This is obtained by systematically fitting the sample protein with the subsets of reference proteins by removing three proteins at a time. The results of all combinations are checked for a satisfactory solution that follows selection criteria among which the most important ones are: (i) the total of all five secondary structures should be between 0.96 and 1.05; (ii) no negative secondary structure should be lower than  $-0.05$ ; (iii) the  $\alpha$ -helix fraction should not change significantly from the original analysis with all the 33 reference

proteins; (iv) the calculated CD spectrum should reasonably fit the experimental spectrum (low root-mean-square residual). The combination of proteins that meet these criteria and that eliminate the fewest proteins from the basic set are then averaged to give the reported percentage of each secondary structure.

### 2.9. Sequence analysis and secondary structure prediction

Alignment of mouse, rat and human Pgp sequences were achieved with the CLUSTALW pro-

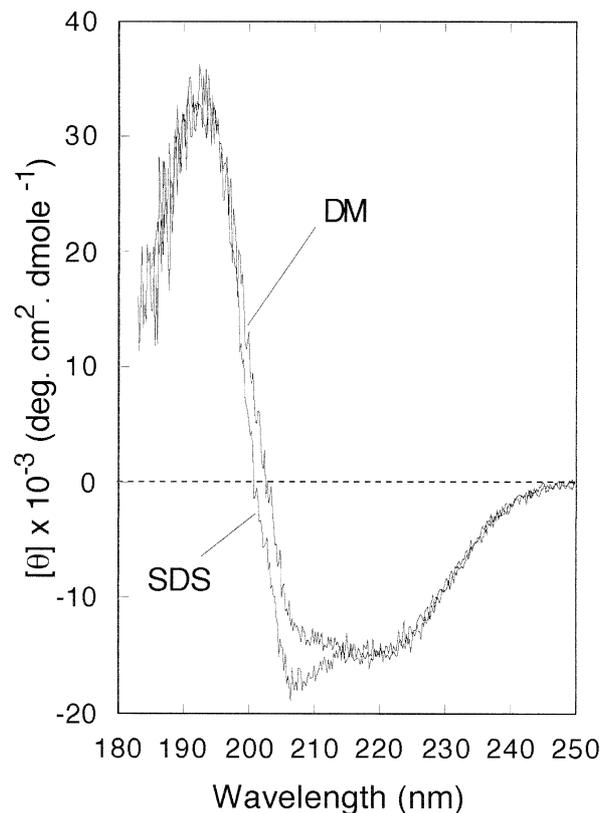


Fig. 1. Circular dichroism spectra of human Pgp in the presence of dodecyl maltoside (DM) or sodium dodecyl sulfate (SDS). Purified human Pgp from CEM/VLB5 cells in 10 mM sodium phosphate buffer (pH 7.0) containing 1% DM or 1% SDS was prepared as described in Section 2. Protein samples at a concentration of  $0.8 \mu\text{M}$  were scanned for 8 times from 183 to 250 nm at 0.2 nm increments, integration time 2 s, in a dichroically neutral quartz cuvette with a path length of 0.2 mm. The baselines were corrected by subtraction of the buffer spectrum. No smoothing was applied. Far-UV circular dichroism spectra are presented in terms of residue mean molecular ellipticity  $[\theta]$  (in  $\text{deg cm}^2 \text{ dmol}^{-1}$ ).

gram [34] with default parameters (fixed and variable gap penalty set to eight, filtering levels set to 1.5 and window width fixed to 10). The putative motifs of nucleotide-binding sites were searched for with the help of the protein sequence analysis server (<http://www.ibcp.fr/predict.html>) and the Prosite dictionary of signatures and potential sites [35]. The secondary structure of mouse, rat and human Pgp sequences (Swissprot code MDR1\_MOUSE, MDR1\_RAT and MDR1\_HUMAN, respectively) was predicted by using both PHD [24] and SOPMA methods [25]. Briefly, both methods scan a sequence database in order to extract homologous proteins and take advantage of the multiple alignment of these related proteins to improve the prediction accuracy. They differ in their predictive algorithm since the PHD method uses neural networks whereas SOPMA uses a similarity matrix and an auto-optimization procedure. The structure superimpositions were per-

formed with the ANTHEPROT 2.0 software [25]. The fit was carried out on the  $\alpha$  carbons of 22 amino acids in the alignment of NBDs with known structures. The averaged root mean square deviation was calculated as follows:

$$\text{RMSD} = \sqrt{\frac{1}{N} \sum (d_i - d_j)^2}$$

where  $N$  is the number of atoms,  $i$  and  $j$  are the number of the superimposed atoms, and  $d$  is the distance between atoms  $i$  and  $j$ .

The graphic display of molecules was done with either Rasmol [36] or ANTHEPROT.

All these analyses were performed through the Protein Sequence Analysis service available on the network (<http://www.ibcp.fr/predict.html>).

Table 1

Percentages of secondary structure of human, murine, and rat Pgp estimated from CD data and predicted from their sequences

| Pgp species                             | Wavelength range (nm) | $\alpha$ -helix | $\beta$ -sheet       |          | $\beta$ -turn | Other  | Total   |
|---|-----------------------|-----------------|----------------------|----------|---------------|--------|---------|
|   |                       |                 | Anti-parallel        | Parallel |               |        |         |
| <i>CD in DM</i>                         |                       |                 |                      |          |               |        |         |
| Human Pgp                               | 183–250               | 43 ± 1          | 10 ± 1               | 6 ± 1    | 15 ± 1        | 26 ± 2 | 101 ± 4 |
| Human Pgp                               | 190–250               | 41 ± 1          | 11 ± 1               | 7 ± 1    | 15 ± 1        | 27 ± 2 | 101 ± 3 |
| Murine Pgp                              | 190–250               | 39 ± 1          | 12 ± 1               | 7 ± 1    | 15 ± 1        | 27 ± 2 | 99 ± 4  |
| Rat Pgp                                 | 190–250               | 39 ± 1          | 11 ± 1               | 7 ± 1    | 14 ± 1        | 26 ± 1 | 98 ± 3  |
| Rat Pgp (non-glycosylated) <sup>a</sup> | 190–250               | 38 ± 1          | 12 ± 1               | 6 ± 1    | 15 ± 1        | 27 ± 1 | 99 ± 4  |
| Rat Pgp (deglycosylated) <sup>b</sup>   | 190–250               | 39 ± 1          | 12 ± 1               | 7 ± 1    | 15 ± 1        | 28 ± 2 | 101 ± 4 |
| <i>CD in SDS</i>                        |                       |                 |                      |          |               |        |         |
| Human Pgp                               | 183–250               | 41 ± 1          | 10 ± 1               | 6 ± 1    | 15 ± 1        | 27 ± 2 | 99 ± 4  |
| Human Pgp                               | 190–250               | 39 ± 1          | 11 ± 1               | 6 ± 1    | 15 ± 1        | 27 ± 2 | 99 ± 4  |
| Murine Pgp                              | 190–250               | 39 ± 1          | 11 ± 1               | 6 ± 1    | 15 ± 1        | 27 ± 2 | 98 ± 4  |
| Rat Pgp                                 | 190–250               | 38 ± 1          | 11 ± 1               | 6 ± 1    | 14 ± 1        | 27 ± 2 | 96 ± 4  |
| Pgp species                             | $\alpha$ -helix       | $\beta$ -sheet  | Remaining structures |          |               |        |         |
| <i>Prediction from sequence</i>         |                       |                 |                      |          |               |        |         |
| Human Pgp                               | 49                    | 16              | 35                   |          |               |        |         |
| Murine Pgp                              | 48                    | 16              | 36                   |          |               |        |         |
| Rat Pgp                                 | 49                    | 16              | 35                   |          |               |        |         |

Deconvolution of CD spectra was done using the Varselec method as detailed in Section 2. For all Pgp spectra, satisfactory analyses were achieved after elimination of the same three proteins from the basic set of 33 reference proteins: poly glutamic acid, triose phosphate isomerase and hemerythin. Data are presented in the top panel as mean percentage of secondary structure with the corresponding standard deviation. The predicted percentages of secondary structure (bottom panel) were calculated on the basis of the average between the data obtained from both SOPMA and PHD methods on the whole sequences.

<sup>a</sup>Rat Pgp purified from tunicamycin-treated AS30-D/COL5 cells.

<sup>b</sup>Rat Pgp digested with PNGase F as described in the text.

### 3. Results

#### 3.1. CD of Pgp

Human Pgp from CEM/VLB5 cells was solubilized and purified in the presence of SDS and the complete removal of SDS was achieved by exchanging with dodecyl maltoside (DM). Fig. 1 shows that the CD spectra of purified human Pgp in the presence of either 1% SDS or 1% DM are both typical of well-folded proteins and the molar ellipticity per residue is in the range that can be expected. These spectra are presented without smoothing to show the accuracy of the measurements even though they were performed down to 183 nm and in the presence of detergent micelles. Both spectra are in the same amplitude range and present a maximum at 193 nm. The spectrum obtained when the protein is in the presence of SDS crosses the zero ellipticity axis at 200 nm and exhibits two minima at 208 nm and 222 nm, which are typical of  $\alpha$ -helical conformation. In the presence of DM, the Pgp spectrum is less characteristic with two blunt minima near 208 and 222 nm, the former being less intense than that for the protein in SDS while the intensity of the latter is identical in the presence of DM or SDS. In addition, relative to that of Pgp in SDS, the spectrum of Pgp in DM is red-shifted by 3 nm at the level of zero ellipticity. Nevertheless, despite the visible differences in the shape of CD spectra in SDS and DM, analysis of the spectra for secondary structure estimation did not reveal significant differences in secondary structure contents (Table 1). These findings suggest a slight effect of SDS on the Pgp secondary structure and the spectral differences may be attributed to some disturbance of the Pgp tertiary structure in SDS.

Estimation of the secondary structure content reported in Table 1 was made using the Varselec method following the guideline and recommendations reported by Johnson and co-workers (see Section 2 and legend of Table 1). It has been pointed out that CD data yields statistically accurate estimates of mul-

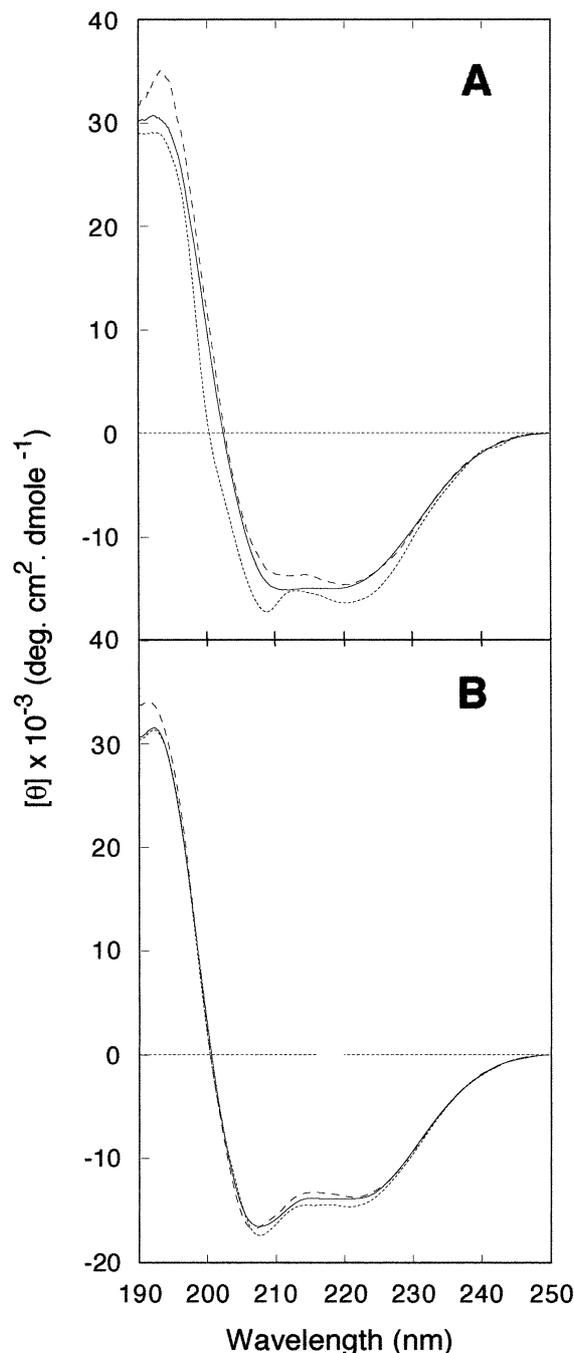


Fig. 2. Comparison of CD spectra of human, rat and murine Pgp in the presence of dodecyl maltoside (A) and SDS (B). Pure Pgp samples were in 10 mM sodium phosphate, pH 7.0, containing 1% dodecyl maltoside (A) or 1% SDS (B) at a protein concentration of 0.5  $\mu$ M. All far-UV CD spectra were recorded from 190 to 250 nm using 1 mm path length cuvette (average of eight scans) and smoothed after subtraction of the buffer spectrum. Dashed line: human Pgp from lymphoblastic leukemia CEM/VLB5 cells; solid line: rat Pgp from hepatoma AS30-D/COL10 cells; dot line: murine Pgp from lymphoid leukemia P388/ADR25 cells.

multiple classes of secondary structure only if the data extend down to 184 nm. To comply with this requirement, narrow CD cells together with concentrated protein solutions were used to obtain a good signal to noise ratio. However, it is well known that membrane proteins tend to aggregate in a time-dependent manner once concentrated. Thus, to avoid any risk of aggregation, CD spectra were routinely recorded using less concentrated Pgp and larger cuvettes, though these conditions allowed the recording of down to 190 nm only to maintain a reasonable signal to noise ratio. In fact, data reported in Table 1 show that for human Pgp recorded in SDS and in DM, the truncation of CD spectra at 190 nm resulted in only slight underestimation of the  $\alpha$ -helix contents while the percentages of other structures remained almost identical. However, these variations are not significant from a statistical point of view. For all these reasons, the Pgp comparisons reported below were made after analysis of CD spectra recorded from 190 to 250 nm.

To address the question of whether there are some differences in the structure of Pgp from various species, we compared the CD spectra of murine and rat Pgp with the human one in the presence of either DM (Fig. 2A) or SDS (Fig. 2B). In the presence of DM, the spectrum of rat Pgp (solid line) is very similar in shape to that of human Pgp (dash line, Fig. 2A). In contrast, although the three spectra exhibit a maximum at 193 nm, the spectrum of murine Pgp (dot line) is rather different from that of rat and human Pgp: (i) both minima at 208 nm and 222 nm are well defined and are lower than those of rat and human Pgp; (ii) the murine Pgp spectrum is blue-shifted relatively to the human and rat ones with zero ellipticity at 200 nm instead of 203 nm for rat and human Pgp. Contrary to what is observed in DM, the spectra obtained for murine, rat and human Pgp are almost identical when recorded in the presence of SDS (Fig. 2B): same amplitudes, a deep minimum at 208 nm and a shoulder at 222 nm, zero ellipticity at 200 nm. Although the spectrum of murine Pgp is different from both rat and human Pgp spectra in the presence of DM, the calculated secondary structure is close to that of the two other sources of Pgp (see Table 1). Moreover, data in Table 1 show that there is also no significant difference between the secondary structure obtained in the presence of DM and that obtained in the presence of SDS for the three

species. It could be concluded at this step that the differences observed in the shape of Pgp CD spectra in DM may be related to some tertiary structure variation.

### 3.2. Effects of glycosylation on Pgp activity

It should be reminded that the main difference between murine Pgp and the rat and human ones lies at the level of glycosylation, the former being much less glycosylated (with an apparent mol.wt. of 140 kDa by electrophoresis) than the two others (apparent mol.wt. of 170 kDa) as evidenced previously [3]. To evaluate the influence of glycosylation on the structure and function of Pgp, oligosaccharide-free Pgp was prepared either by enzymatic treatment with PNGase F or by culturing AS-30D/COL10 cells in the presence of tunicamycin prior to purification (see Section 2). As evidenced from the SDS-PAGE analysis shown in Fig. 3, the native rat Pgp (lane 1) migrated as a 170 kDa band, whereas both PNGase F-treated rat Pgp (called deglycosylated Pgp) and Pgp purified from tunicamycin-treated cells (called non-glycosylated Pgp) moved as 140 kDa proteins. This is

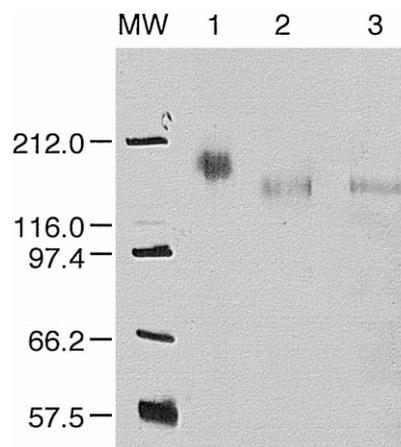


Fig. 3. Analysis of deglycosylated and non-glycosylated rat Pgp by SDS-PAGE. Deglycosylation was performed by PNGase F digestion of native rat Pgp as detailed in Section 2. Non-glycosylated Pgp was purified from tunicamycin-treated AS30-D/COL10 cells, as detailed in Section 2. Samples were run on an 8% SDS-PAGE gel followed by silver staining. Lane 1, purified glycosylated rat Pgp from AS30-D/COL10 cells, 1  $\mu$ g; lane 2, deglycosylated rat Pgp, 0.2  $\mu$ g; lane 3, non-glycosylated rat Pgp, 0.2  $\mu$ g. MW, molecular weight standard, 0.5  $\mu$ g for each band.

in agreement with the expected molecular weight calculated from the amino acid composition of Pgp (141 299 Da) and shows that oligosaccharide-free Pgp can be obtained by either way. From a functional point of view, deglycosylated Pgp reconstituted into liposomes presented a 5-fold increase in ATPase activity when compared to normal Pgp (typically  $1040 \pm 100$  nmol ATP/min/mg deglycosylated Pgp and  $200 \pm 50$  nmol ATP/min/mg Pgp for rat AS-30D/COL10 Pgp), while ATP-dependent [ $^3$ H]-vinblastine uptake activity was slightly lower for deglycosylated Pgp ( $81 \pm 4$  pmol/20 min/mg protein) than normal Pgp ( $120 \pm 10$  pmol/20 min/mg protein). As discussed below, these differences in activities are likely due to the ratio of right-side out and inside out configuration of reconstituted Pgp. Nevertheless, these results demonstrate that normal and deglycosylated Pgp used in the CD experiments were active after reconstitution into liposomes, although they did not exhibit ATPase activity in the detergent-solubilized state because of the absence of lipids [37]. Moreover, when cultured in the presence of tunicamycin, AS-30D/COL10 cells producing non-glycosylated Pgp survive in the presence of  $10 \mu\text{g/ml}$  colchicin (Baggetto and Dong, unpublished observation). These data strongly suggest that the oligosaccharidic part of Pgp is not essential for it to function.

### 3.3. Effect of glycosylation on Pgp structure

CD spectra of oligosaccharide-free rat Pgp recorded in the presence of DM are reported in Fig. 4A in comparison with glycosylated rat Pgp in DM and in SDS. As can be expected, the spectrum of Pgp purified from tunicamycin-treated cells (non-glyco-

lylated Pgp) is almost identical to that of PNGase F-digested Pgp (deglycosylated Pgp), with a minimum at 203 nm and a shoulder at 222 nm, the zero ellipticity being at 200 nm. The shape of these oligosaccharide-free Pgp CD spectra is rather different from that of glycosylated Pgp in the presence of DM,

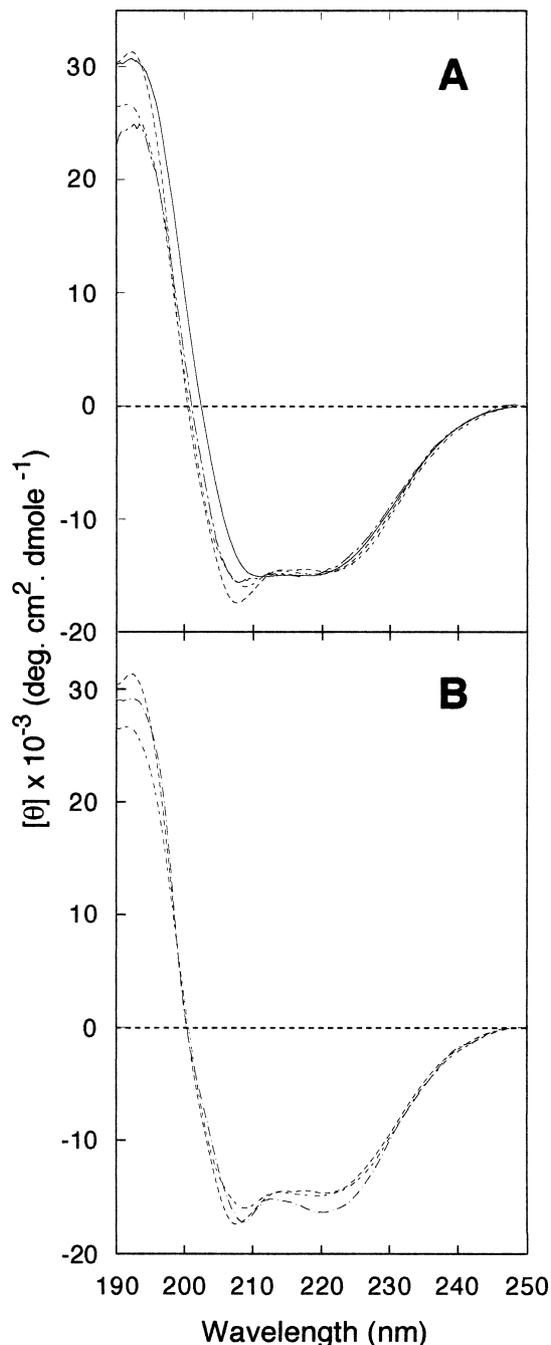


Fig. 4. Effects of glycosylation on CD of Pgp. (A) Comparison of the CD spectra of glycosylated rat Pgp in 1% DM (solid line) and in 1% SDS (dash line) with that of oligosaccharide-free rat Pgp in 1% DM obtained either by enzymatic treatment of purified Pgp with PNGase F (long dash–short dash line) or by Pgp purification from tunicamycin-treated rat AS30-D/COL10 cells (dot–dash line), as detailed in Section 2. (B) Comparison of the CD spectra of murine P388/ADR25 Pgp in 1% DM (dot line) and in 1% SDS (dash line) with that of non-glycosylated rat Pgp purified from tunicamycin-treated AS30-D/COL10 cells (dot–dash line). Protein concentration, buffer used and recording of spectra were as described in the legend of Fig. 2.

although deconvolution analysis by Varselec did not reveal significant differences in the contents of each secondary structure. However, the spectral differences between the glycosylated Pgp and the oligosaccharide-free one remain low, indicating that the contribution of the sugar part of Pgp to the CD spectrum is limited. Taken together, these results indicate that glycosylation of Pgp has little effect on the secondary structure of the protein but certainly induces some local conformational changes in the structure of Pgp. Interestingly, the CD spectrum shape of oligosaccharide-free Pgp is close to that of glycosylated Pgp in the presence of SDS. This indicates that the conformation of Pgp may be modified in the same manner when oligosaccharides are not present or when their interaction with the core protein is abolished by SDS.

As stated above, murine Pgp is naturally much less glycosylated than human and rat Pgp, although some glycosylation level was revealed (Dong and Baggetto, unpublished results). The comparison of the CD spectrum of murine Pgp in DM (dot line in Fig. 4B) with that of non-glycosylated rat Pgp in DM (dot–dash line) obviously shows that the shapes of both spectra are very similar. Here again, both spectra are close to that of rat Pgp in the presence of SDS (dash line) and the calculated secondary structure from these three spectra did not show significant differences (Table 1). It is clear from these data that the particular shape of murine Pgp spectrum in DM evidenced by its comparison with rat and human Pgp (Fig. 2A) is due to its low level of glycosylation. In conclusion, the shape of Pgp CD spectra depends on the presence of glycosylation, indicating that the sugar part of Pgp

interacts with the protein core. As all the Pgp species gave superimposable CD spectra in the presence of SDS while they are rather different in DM, it is concluded that SDS abolishes the interaction between the Pgp sugar part and the protein core.

#### 3.4. Secondary structure of Pgp predicted from its sequence

The secondary structure prediction of the whole Pgp sequence by SOPMA and PHD methods allows the global estimations of Pgp secondary structure (Table 1). The results are very similar for the three Pgp species and yield an average content of 48%  $\alpha$ -helix, 16%  $\beta$ -sheet and 36% coil. These values are rather close to those obtained from CD experiments for three states only (43%  $\alpha$ -helix, 16%  $\beta$ -sheet and 41% others) when considering the most reliable estimation deduced from deconvolution of human-Pgp CD spectra recorded in the 183–250 nm range (Table 1). This consistency of data reinforces the reliability of secondary structure estimation by both approaches and encourages us to further state precisely the localization of stretches of secondary structure along the amino acid sequence. Data presented in Fig. 5 are the result of a careful sequence analysis described as follows. (i) Since secondary structure prediction methods are not very efficient for membrane regions, the 12 transmembrane segments were deduced from previous work [39] where they were predicted as  $\alpha$ -helices. (ii) Intermembrane regions longer than 40 residues were submitted as individual requests to prediction. Thus, eight major regions were distinguished on the human Pgp (region 1, 1–51; region 2,

---

Fig. 5. Secondary structure of human, murine and rat Pgp predicted from their respective sequence. The three sequences were extracted from Swissprot 33 [38] and named with their corresponding code entry: HUM, MDR-HUMAN; MOU, MDR\_MOUSE; RAT, MDR1\_RAT. The sequences were aligned according to the CLUSTALW program [34] with default parameters. A motifs of the nucleotide-binding domains are shown with a heavy solid line. The twelve transmembrane domains defined according to the previous work of Chen et al. [39], are shown under numbered brackets from 1 to 12. Alpha helices are represented with letters in reversed color into a black box, beta sheets are shown in grey boxes, and aperiodic regions in white boxes. Secondary structures are indicated if both prediction methods (PHD and SOPMA, see Section 2) predict the same conformational state. For residues that were predicted as an isolated state within long regular secondary structures, a smoothing was performed as follows: the residue is converted into the surrounding conformational state if (i) it exhibits a good calculated tendency to fit with the surrounding state, and (ii) there are at least two residues at both sides which present the same state. Glycosylation positions reported in the literature are located at N91, N94, N99, and N296 for human Pgp, N91 and N96 for mouse Pgp, and N91, N96, and N295 for rat Pgp, respectively [39,40]. Identical and similar amino acids for the three sequences are symbolized by a star (\*) and a dot (·), respectively.



73–119; region 3, 141–188; region 4, 237–296; region 5, 347–710 containing the first nucleotide-binding domain (NBD1); region 6, 778–832; region 7, 885–936; region 8, 995–1280 containing the second nucleotide binding domain (NBD2). The remaining non-membrane regions (210–215, 318–325, 732–756 and 958–973) were not considered since predictive methods are poorly reliable on sequences shorter than 40 residues. In the case of mouse and rat Pgp, homologous regions to those defined for human Pgp have been considered. (iii) For each sequence, the data are presented in terms of consensus, i.e., the prediction was considered only when a residue is predicted in the same state by both SOPMA and PHD methods. It should be noticed that both PHD and SOPMA methods predict a same conformational state for most residues. The amino acids predicted in  $\alpha$ -helix,  $\beta$ -sheet and coil are presented in black, grey and white boxes, respectively. (iv) Other residues which do not belong to transmembrane regions were not predicted in a consensus way by both SOPMA and PHD methods, i.e., no conformational state was reported for these residues. (v) Human, mouse and rat sequences have been predicted separately and aligned by primary sequence homology to highlight Pgp segments that are predicted in the same consensus state.

Fig. 5 shows that the agreement between the prediction of the secondary structure of Pgp from different species is very good, as can be expected from the high level of sequence homology (78% identity between the three sequences). Secondary structure predictions easily distinguish regular secondary structures from aperiodic ones, and the combination with sequence alignment allows the discrimination of ambiguously predicted regions. Although several regions comprise poorly organized or ambiguously predicted conformational states, Fig. 5 highlights frequent, rather long, and conserved sequences predicted unambiguously and simultaneously for the three sequences, either as  $\alpha$  helices (18 of such helices comprising at least four residues are observed), or as  $\beta$ -sheets (10 segments with at least three residues) or as aperiodic segments (21 segments with at least four residues). Because of the severe limitations described above for the ratification of the predicted conformational state, we assume that the secondary structure predicted for these segments is the most probable one occurring in native Pgp.

It should be noticed that the two major extra membrane segments (347–710 and 995–1280 for human Pgp) containing NBD1 and NBD2 Pgp domains exhibit an alternance of predicted  $\alpha$ -helix and  $\beta$ -sheet and thus appear to belong to the  $\alpha/\beta$  structural class of proteins. This feature is common to proteins that bind nucleotides through the ATP-binding cassette defined by Walker et al., [7]. The position of the nucleotide-binding domain glycine-rich loops (called motif A) and presenting the consensus [AG]-x(4)-G-K-[ST] sequence is indicated by a heavy line on Fig. 5 (segments 427–434 and 1070–1077 for human NBD1 and NBD2 Pgp domains, respectively). These A motifs are both located between a highly probable  $\beta$  sheet and an  $\alpha$  helix, which corresponds to the topology of A motifs observed in four nucleotide-binding proteins for which the 3D structure is known and that are available on Protein Data Bank database: adenylate kinase (3ADK), elongation factor TU (1ETU), guanylate kinase (1GKY), and p21<sup>RAS</sup> (5P21). Fig. 6 shows the A motif sequence alignment for these four proteins together with those of NBD1 and NBD2 domains of



Fig. 6. Primary and secondary structure comparison between A motifs of known 3D structure and of the two Pgp A motifs. The A motif of nucleotide binding domain present the consensus [AG]-x(4)-G-K-[ST] sequence. The four nucleotide-binding proteins taken into account are those for which the 3D coordinates are available on NRL\_3D database: adenylate kinase (3ADK, [41]), elongation factor TU (1ETU, [42]), guanylate kinase (1GKY, [43]), and p21<sup>RAS</sup> (5P21, [44]). The secondary structure has been deduced from 3D data by using the Kabsch and Sander [45] method. The greyed box encircling the two last sequences highlights the two predicted NBDs (motif A) from the human Pgp. Open rectangles represent  $\beta$  sheets; filled black rectangles represent  $\alpha$  helices; dotted rectangles represent  $\beta$  turns. A motifs have been aligned and the number on their left indicates the first amino acid of the corresponding sequence element.

human Pgp. Although the consensus primary sequence homology is limited to only three identical residues, the predicted secondary structures for the two human Pgp A motifs fit well with that observed for the four crystallized proteins. Moreover, Fig. 7A shows that the 3D structure of A motifs of these four

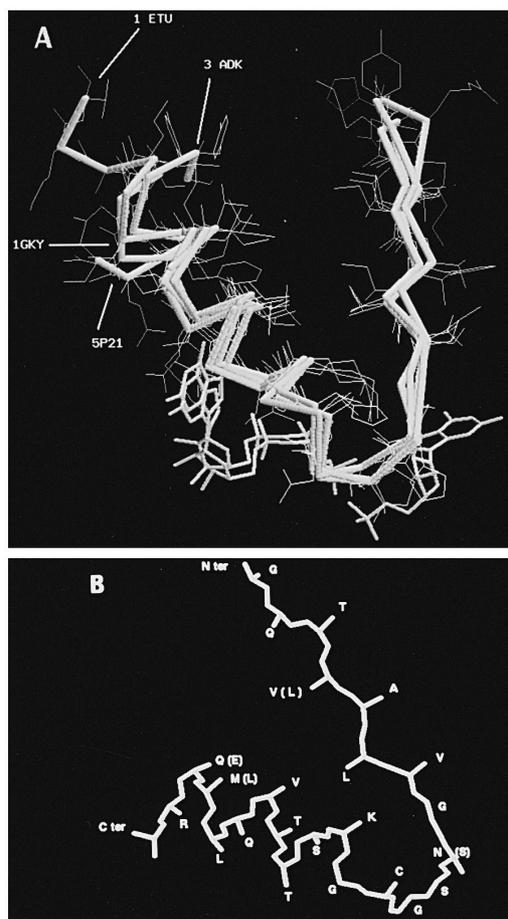


Fig. 7. Conformation of nucleotide-binding A motif. (A), Superimposition of A motif of the four reference crystallized proteins indicated in legend of Fig. 6. 3ADK, 1ETU, 1GKY, and 5P21 are presented with, respectively, no nucleotide, GDP, GMP, and 2-guanosine-5'-( $\beta,\gamma$ -imido) triphosphate. When present, nucleotides are represented with medium-size lines whereas backbones are shown with large-size lines. Side chains are drawn with thin lines. The fit was performed for  $\alpha$ -carbons of 22 amino acids starting with the third position in the alignment. Root mean square deviation at the  $\alpha$ -carbon level is 0.93 Å. (B) Predicted topology for the two A motifs of human Pgp NBD1 and NBD2. The main chain of the represented A motif fold was deduced from the four A motifs presented in (A). The amino acid position is given for NBD1; when different, residues of NBD2 A motif are given in parentheses.

reference proteins can be superimposed and fit with a root mean square deviation of 0.93 Å, i.e., they exhibit a very close conformation. This demonstrates that the structure of A motifs is common for all available nucleotide-binding protein structures and consequently, it appears highly probable that the A motifs of Pgp NBD1 and NBD2 adopt the same global fold. Thus, a model for the topology of the A motifs has been built on the basis of the average conformation of A motifs of known 3D structure. This model is presented in Fig. 7B with the amino acid sequence of human NBD1 and NBD2 A motifs (the sequence of the latter is between parenthesis when no identical residue is observed) and it is also transposable to the homologous rat and mouse A motifs.

#### 4. Discussion

To further understand the functional mechanism of drug transportation, detailed structural data of P-glycoprotein are indispensable. We have therefore investigated the structure of human, rat and murine Pgp by combining circular dichroism analysis and secondary structure prediction from amino acid sequence analysis. The CD study was feasible because of the availability of sufficient amount of highly purified Pgp prepared by the procedure that we reported recently [3], i.e., solubilization and purification of Pgp in the presence of SDS followed by the complete removal and exchange of SDS into DM. The removal of SDS was achieved by the HAP-mediated SDS removal method that we described recently [15] and that combines the effects of phosphate ions and the hydrophobic tail of a non-ionic detergent to displace SDS molecules from their protein-binding sites during ceramic hydroxyapatite chromatography. SDS is well known to delipidate membrane proteins because of its ability to replace phospholipids bound to proteins. Hence, our Pgp preparations are highly depleted of lipids, and thus inactive in this state, as also reported by others [37,46]. This rises the question of whether CD spectra of Pgp solubilized in DM represent those of the native protein. The fact that the reconstitution of DM-solubilized Pgp into liposomes allows the restoration of both ATPase activity and ATP-dependent drug transportation [[3,15], present

work] is a good indication that its global conformation should not be far from the native one, and that its secondary structure should be the same. Moreover, the recent electron microscopy observations of Pgp showed that both DM-solubilized and lipid-reconstituted Pgp displayed similar shapes and dimensions [13], indicating that DM does not disrupt the secondary/tertiary structure of Pgp. The basal ATPase activity of our reconstituted Pgp (about 200 nmol ATP/min/mg Pgp) is in the range or higher of that reported by others [47,14], but about 4 to 8 times lower than that reported for Pgp preparations for which no attempts to delipidate the protein were made [13,46,48]. Although several factors listed previously [3] may account for the explanation of these differences (in particular the lack of specific association with an unidentified phospholipid [37,49]), a very probable explanation comes from the problem of Pgp orientation in reconstituted liposomes. Indeed, it is known that reconstitution into liposomes favors the right-side out configuration for Pgp [13,46]. Furthermore, the reconstitution procedure of Rigaud et al. used in this study is particularly efficient to obtain right-side out orientation for glycoproteins, because of the bulky hydrophilic sugar part ([27], and Rigaud J-L, personal communication). In that configuration, Pgp NBDs are localized on the inner face of the liposome and are thus not directly accessible to ATP. Hence, measured ATPase activity should be mainly due to some Pgp in the inside-out configuration. This ratio between right-side out and inside-out orientation of Pgp can also explain the 5-fold ATPase activity of deglycosylated Pgp (around 1000 nmol ATP/min/mg protein). Indeed, in the absence of the sugar moiety, the hydrophilic/hydrophobic imbalance between the two extramembraneous parts of Pgp is probably not sufficient enough to direct the orientation of Pgp during liposome formation. This yields more Pgp in the inside-out configuration for which NBDs are directly accessible to ATP, yielding a higher ATPase activity than normal Pgp. The above reconstitution features and the activation of Pgp by various compounds render the estimation of Pgp enzymatic activity particularly difficult. Although further careful investigations are needed to better evaluate the reconstitution and the activities of Pgp, it is clear that the sugar part of Pgp does not appear essential for its activity.

#### 4.1. Glycosylation of Pgp

Comparison of Pgp CD spectra recorded in DM showed that the shape differences between the three species are virtually abolished when the sugar part of Pgp is removed by *N*-glycanase (PNGase F) treatment, or not present as for rat Pgp prepared from cells cultured in the presence of tunicamycin, which inhibits glycosylation of newly-synthesized proteins [5], or naturally not abundant like in murine Pgp. This means that the differences in CD shape are actually due to the presence of glycosylation. However, the relative contents in secondary structure deduced from spectral deconvolution is almost the same, indicating that glycosylation has no significant effect on the Pgp secondary structure. Moreover, it is noteworthy that Pgp in SDS exhibited spectral features very close to those of oligosaccharide-free Pgp in DM and that all Pgp spectra from the three species are almost superimposable in SDS. Two main conclusions can be drawn from these observations: (i) the spectrum of naturally low-glycosylated murine Pgp is superimposable to the ones of highly glycosylated rat and human in SDS, meaning that the sugar part of the Pgp exerts a negligible contribution to CD by itself; (ii) SDS probably abolishes the interaction between the sugar part of Pgp with the core protein. It is thus highly probable that the CD shape variations observed for various Pgp in DM can be attributed to some conformational changes due to sugar interactions with the protein core. The above conclusions are in agreement with what has been reported for most glycosylated proteins studied by CD (see for example the anion transport protein of the human erythrocyte membrane, [50,51]; the cobra venom factor, [52]; the water channel CHIP28 protein, [53]), although dramatic changes have been observed for some proteins involved in recognition processes such as Human CD2 [19]. Moreover, it has been reported that mutant human Pgp that lack all the three *N*-linked glycosylation sites in the first extracytoplasmic loop provided drastically reduced drug-resistant transfectants, suggesting that the sugar moiety may contribute to the correct folding and/or proper routing of Pgp, and its stabilization en route to or within the plasma membrane [4]. This feature seems to be common to cell surface proteins [54]. From a functional point of view, it is known that *N*-linked glyco-

sylation can influence many properties of proteins, including intracellular transport, biological activity, stability and antigenicity. In some cases, the biological effect of *N*-linked glycosylation on a glycoprotein depends on the particular site of glycosylation within the protein chain. In the case of Pgp studied here, several results strongly suggest that oligosaccharide-free Pgp is fully functional: (i) the purified rat deglycosylated Pgp reconstituted in liposomes exhibited similar ATP-dependent [<sup>3</sup>H]-vinblastine uptake to that of the native glycosylated Pgp; (ii) rat hepatoma AS30-D/COL10 cells cultured in the presence of tunicamycin are resistant to colchicin although Pgp is actually non-glycosylated; (iii) murine Pgp is naturally poorly glycosylated but it is fully functional, i.e., conferring drug resistance to lymphoid leukemia P388/ADR25 cells. Consequently, all these structural and functional results suggest (i) that the removal of the sugar part of Pgp does not result in a significant destabilization of the secondary and tertiary structure of Pgp, and (ii) that glycosylation of Pgp is not essential for its functioning, although a fine tuning of the Pgp activity by the level of glycosylation is not excluded.

#### 4.2. Effect of SDS on Pgp

It has been reported in the literature that SDS tends to increase the protein  $\alpha$ -helix content [55] and to decrease the  $\beta$ -structure [56–58] while for some proteins, no alteration of secondary structure was observed [59,60]. It is remarkable that in the case of Pgp, the comparison between Pgp CD data in DM and in SDS clearly indicates that any conformation differences are minor. This is particularly obvious in the case of murine Pgp. Moreover, SDS does not induce a significant modification in the percentage of each type of secondary structure deduced from CD spectral analysis. Thus, besides the effect of SDS on the sugar–Pgp core interaction discussed above, these findings suggest that these slight spectral changes are due to limited perturbation of the tertiary structure of Pgp by SDS. According to the proposal for SDS–galactosidase interaction [61], it is thought that SDS molecules bind the outer surface of the Pgp without disrupting the protein core. This is fully compatible with the relative easiness to restore Pgp function upon removal of SDS (exchange into DM, [15]) and

reconstitution into liposomes [3], indicating that SDS does not induce extensive denaturation of Pgp. By the way, this SDS binding, which is limited to the outer surface of the protein, may be a common phenomenon which is not only limited to membrane proteins. Indeed, we have recently shown that proteins such as lysozyme and glutamate dehydrogenase can easily recover their full activities after the complete removal of SDS or its exchange into another mild detergent such as dodecyl maltoside [15].

#### 4.3. Secondary structure of Pgp

The far-UV CD spectra of Pgp are dominated by the contribution of  $\alpha$ -helical structure with prominent bands at 222 and 208 nm. The secondary structure of Pgp deduced from circular dichroism spectra using the Varselec method [21] is similar for the three species and composed of 38–43%  $\alpha$ -helix, 16–19%  $\beta$ -sheets, 14–15%  $\beta$ -turn, and 26–28% of other structures, whatever the Pgp species and its level of glycosylation are (Table 1). As discussed above, both glycosylation and detergents slightly change the shape of Pgp CD spectra, but have no significant effect on the secondary structure content. In addition, it should be mentioned that whatever the Pgp species, either glycosylated or not, and whatever the detergent used, satisfactory analyses of CD spectra were always achieved after elimination of the same three proteins from the basic set of 33 reference proteins: poly glutamic acid, triose phosphate isomerase and hemerythin. This constitutes an indication of the consistency between the various Pgp spectra. In accordance with the statement of Johnson [16], CD spectra should be recorded down to at least 184 nm to yield statistically significant estimates of multiple classes of secondary structure. Moreover, considering that SDS is known to induce some conformational change in protein, the most accurate data were those obtained with human Pgp the CD spectra of which were recorded down to 183 nm in the presence of DM. Although it should be kept in mind that secondary structure estimation was made using soluble non-glycosylated reference proteins of known secondary structure and that the Pgp integrated into micelles may display some unexpected CD contribution, the most reliable estimation of Pgp secondary structure contents is 43%  $\alpha$ -helix, 16%  $\beta$ -sheets (10% anti-

parallel, 6% parallel), 15%  $\beta$ -turn, and 26% other structures. These proportions appear quite different from those estimated by attenuated total reflection FTIR for Pgp reconstituted into liposomes (32%  $\alpha$ -helix, 26%  $\beta$ -sheet, 29%  $\beta$ -turn and 13% unordered structure, [14]). It should be kept in mind that both CD and FTIR methods of secondary structure content estimation used structural parameters deduced from soluble proteins of known structure, and that it could not be ruled out that some unexpected contribution due to the membranous part of Pgp and/or Pgp–lipid interaction may skew these estimations. However, the secondary structure proportions deduced from CD data agree quite well with those predicted from Pgp amino acid sequence combining PHD [24] and SOPMA [25] prediction methods, i.e., 48%  $\alpha$ -helix, 16%  $\beta$ -sheets, and 36% coil structures (see below). This consistency reinforces the reliability of secondary structure prediction which allows the most probable nature and localization of secondary structure elements along the Pgp sequence.

Pgp contains 12 transmembrane regions, the precise topology of which is still a matter of controversy [39,62,40]. Because secondary structure predictive methods are known not to be accurate within membrane regions, we have taken into account the intermembrane regions as proposed by the classical model of Chen et al., [39]. Conversely, prediction methods such as PHD and SOPMA are very effective on globular protein, especially when they are used in combination: a success rate as high as 82% for jointly predicted residues can then be reached [25]. The predictions of non-membrane segments of Pgp reported in Fig. 5 were made very carefully as described in Section 3. In particular, ambiguously predicted residues were not reported in this Figure which thus presents only reliable predictions for  $\alpha$ -helix,  $\beta$ -sheet, and coil conformations. Moreover, a supplementary index of reliability of the predicted structure is given by the three aligned sequences of Pgp. Indeed, the prediction of the same conformational state for a particular residue that occupies the same position in the three sequences increases its probability to be correctly predicted. In our opinion, the present prediction represents an accurate framework of the actual Pgp secondary structure which constitute a good basis for further 3D Pgp structure modeling.

#### 4.4. Modeling of nucleotide binding A motifs of Pgp

Global structural models of ATP-binding domains of periplasmic permeases have been proposed by analogy to the adenylate kinase 3D structure [8,13]. The precise knowledge of Pgp secondary structure allows a better specification of the nucleotide-binding A motifs for which secondary structure exhibits a strong similarity with A motifs of known crystallized proteins (Fig. 6). The 3D structures of A motifs of the four nucleotide-binding proteins exhibit superimposable conformation (Fig. 7A, adenylate kinase (3ADK), elongation factor TU (1ETU), guanylate kinase (1GKY), and p21<sup>RAS</sup> (5P21)), demonstrating that the A motif fold is common for these four proteins and it appears highly probable that the A motifs of Pgp NBD1 and NBD2 adopt the same global fold. A precise model for the folding of the Pgp A motifs have thus been built by analogy with known A motif folds (Fig. 7B). It should be noticed that three of the four nucleotide-binding proteins used for modeling have been crystallized with a nucleotide, as shown in Fig. 7A. It is clear that the three nucleotides are localized near the turn between the  $\beta$ -sheet and  $\alpha$ -helix of the A motif. From these observations, it is not possible to specify the position of ATP in the A motifs of Pgp, but it seems probable that ATP binds closed to the connecting segment between the  $\beta$ -sheet and the  $\alpha$ -helix. This region is thus a good target for punctual mutations in order to localize the residues involved in ATP binding. This approach, combined with fluorescence nucleotide analogues binding studies such as those reported in our recent work concerning the recombinant NBD2 of mouse Pgp [63], would bring further insight in the understanding of the Pgp structure–function relationship.

## 5. Conclusion

The present work is the first report of CD analyses of the Pgp structure and the structural effect of glycosylation on Pgp. By combining the experimental data from circular dichroism and the predicted structure from the amino acid sequence, we have specified the most probable localization of secondary structure elements along the Pgp sequence. These data repre-

sent a useful template for any 3D structure modeling of Pgp and it is well illustrated here by the modeling of Pgp A motifs of ATP binding sites. This is a step towards tertiary structure modeling of Pgp for which more experimental data are needed, such as the knowledge of the tertiary structure of Pgp based on 2D crystals.

## Acknowledgements

This work was supported by Centre National de la Recherche Scientifique (CNRS) and by grants from Association pour la Recherche sur le Cancer (ARC, no. 6756) and from Ligue Nationale Contre le Cancer, Comité du Rhône (to L.G. Baggetto). We wish to thank W.C. Johnson Jr. for kindly providing us with an updated version of the VARSLC1 program with 33 proteins in the basis set. Thanks are also due to J.L. Rigaud for Pgp reconstitution, to B. Lermite for her help in cell culture, C. Van Herrewège for the artwork, A. Bosch for photography, and the Roger Bellon Laboratories for their kind gift of vinblastine.

## References

- [1] J.A. Endicott, V. Ling, *Annu. Rev. Biochem.* 58 (1989) 137–171.
- [2] M.M. Gottesman, I. Pastan, *Annu. Rev. Biochem.* 62 (1993) 385–427.
- [3] M. Dong, F. Penin, L.G. Baggetto, *J. Biol. Chem.* 271 (1996) 28875–28883.
- [4] A.H. Schinkel, S. Kemp, M. Dolle, G. Rudenko, E. Wagenaar, *J. Biol. Chem.* 268 (1993) 7474–7481.
- [5] K. Kramer, T.K. Weber, R. Arceci, N. Ramchurren, W.V. Kastinakis, G. Steele Jr., I.C. Summerhayes, *Br. J. Cancer* 71 (1995) 670–675.
- [6] C.F. Higgins, M.M. Gottesman, *Trends Biochem. Sci.* 17 (1992) 18–21.
- [7] J.E. Walker, M. Saraste, M.J. Runswick, N.J. Gay, *EMBO J.* 1 (1982) 945–951.
- [8] S.C. Hyde, P. Emsley, M.J. Hartshorn, M.M. Mimmack, U. Gileadi, S.R. Pearce, M.P. Gallagher, D.R. Gill, R.E. Hubbard, C.F. Higgins, *Nature* 346 (1990) 362–365.
- [9] C.S. Mimura, S.R. Holbrook, G.F.-L. Ames, *Proc. Natl. Acad. Sci. U.S.A.* 88 (1991) 84–88.
- [10] B. Ullman, *J. Bioenerg. Biomembr.* 27 (1995) 77–84.
- [11] M.A. Van Kuijk, R.A.M.H. van Aubel, A.E. Busch, F. Lang, F.G.M. Russel, R.J.M. Bindels, C.H. van Os, P.M.T. Deen, *Proc. Natl. Acad. Sci. U.S.A.* 93 (1996) 5401–5406.
- [12] C.A. Doige, G.F. Ames L, *Annu. Rev. Microbiol.* 47 (1993) 291–319.
- [13] M.F. Rosenberg, R. Callaghan, R.C. Ford, C.F. Higgins, *J. Biol. Chem.* 272 (1997) 10685–10694.
- [14] N. Sonveaux, A.B. Shapiro, E. Goormaghtigh, V. Ling, J.-M. Ruyschaert, *J. Biol. Chem.* 271 (1996) 24617–24624.
- [15] M. Dong, G. Baggetto, P. Falson, M. le Maire, F. Penin, *Anal. Biochem.* 247 (1997) 333–341.
- [16] W.C. Johnson Jr., *Proteins Struct. Funct. Genet.* 7 (1990) 205–214.
- [17] R.W. Woody, *Methods Enzymol.* 246 (1995) 34–71.
- [18] M.P. Heyn, *Methods Enzymol.* 172 (1989) 575–584.
- [19] D.F. Wyss, J.S. Choi, J. Li, M.H. Knoppers, K.J. Willis, A.R.N. Arulannandam, A. Smolyar, E.L. Reinherz, G. Wagner, *Science* 269 (1995) 1273–1278.
- [20] L.A. Compton, W.C. Johnson Jr., *Anal. Biochem.* 155 (1986) 155–167.
- [21] P. Manavalan, W.C. Johnson Jr., *Anal. Biochem.* 167 (1987) 76–85.
- [22] J.P. Hennessey Jr., W.C. Johnson Jr., *Biochemistry* 20 (1981) 1085–1094.
- [23] F. Eisenhaber, B. Persson, P. Argos, *Crit. Rev. Biochem. Mol. Biol.* 30 (1995) 1–94.
- [24] B. Rost, C. Sander, *J. Mol. Biol.* 235 (1994) 13–26.
- [25] C. Geourjon, G. Deléage, *CABIOS* 11 (1995) 681–684.
- [26] C. Chothia, A.M. Lesk, *EMBO J.* 5 (1986) 823–826.
- [27] J.L. Rigaud, B. Pitard, D. Levy, *Biochim. Biophys. Acta* 1231 (1995) 223–240.
- [28] M. Horio, M.M. Gottesman, I. Pastan, *Proc. Natl. Acad. Sci. U.S.A.* 85 (1988) 3580–3584.
- [29] U.K. Laemmli, *Nature* 227 (1970) 680–685.
- [30] F. Penin, C. Godinot, D.C. Gautheron, *Biochim. Biophys. Acta* 775 (1984) 239–245.
- [31] H. Bloom, H. Beier, H.S. Gross, *Electrophoresis* 8 (1987) 93–99.
- [32] G.L. Peterson, *Anal. Biochem.* 83 (1977) 346–356.
- [33] G.H. Beaven, E.R. Holiday, *Adv. Protein Chem.* 7 (1952) 319–386.
- [34] J.D. Thompson, D.G. Higgins, T.J. Gibson, *Nucleic Acids Res.* 22 (1994) 4673–4680.
- [35] A. Bairoch, *Nucleic Acids Res.* 21 (1993) 3097–3103.
- [36] R. Sayle, M. Saqi, M. Weir, A. Lyall, *CABIOS* 11 (1995) 571–573.
- [37] F.J. Sharom, *J. Bioenerg. Biomembr.* 27 (1995) 15–22.
- [38] A. Bairoch, B. Boeckmann, *Nucleic Acids Res.* 22 (1994) 3578–3580.
- [39] C. Chen, J.E. Chin, K. Ueda, D.P. Clark, I. Pastan, M.M. Gottesman, I.B. Roninson, *Cell* 47 (1986) 381–389.
- [40] J.-T. Zhang, M. Duthie, V. Ling, *J. Biol. Chem.* 268 (1993) 15101–15110.
- [41] D. Dreusicke, P.A. Karplus, G.E. Schulz, *J. Mol. Biol.* 199 (1988) 359–3571.
- [42] T.F. la Cour, J. Nyborg, S. Thirup, B.F. Clark, *EMBO J.* 4 (1985) 2385–2388.
- [43] T. Stehle, G.E. Schulz, *J. Mol. Biol.* 224 (1992) 1127–1141.
- [44] E.F. Pai, U. Krengel, G.A. Petsko, R.S. Goody, W. Kabsch, A. Wittinghofer, *EMBO J.* 9 (1990) 2351–2359.

- [45] W. Kabsch, C. Sander, *Biopolymers* 22 (1983) 2577–2637.
- [46] F.J. Sharom, X. Yu, J.W.K. Chu, C.A. Doige, *Biochem. J.* 308 (1995) 381–390.
- [47] A.B. Shapiro, V. Ling, *J. Biol. Chem.* 269 (1994) 3745–3754.
- [48] S.V. Ambudkar, I.H. Lelong, J. Zhang, C.O. Cardarelli, M.M. Gottesman, I. Pastan, *Proc. Natl. Acad. Sci. U.S.A.* 89 (1992) 8472–8476.
- [49] C.A. Doige, F.J. Sharom, *Biochim. Biophys. Acta* 1109 (1992) 161–171.
- [50] J.R. Casey, C.A. Pirraglia, R.A.F. Reithmeier, *J. Biol. Chem.* 267 (1992) 11940–11948.
- [51] V.E. Sarabia, J.R. Casey, A.F. Reithmeier, *J. Biol. Chem.* 268 (1993) 10676–10680.
- [52] D.C. Gowda, E.C. Petrella, T.T. Raji, R. Bredehorst, C.-W. Vogel, *J. Immunol.* 152 (1994) 2977–2986.
- [53] A.N. Van Hoek, M.C. Wiener, J.-M. Verbavatz, D. Brown, P.H. Lipniunas, R.R. Townsend, A.S. Verkman, *Biochemistry* 34 (1995) 2212–2219.
- [54] C.G. Gahmberg, M. Tolvanen, *Trends Biochem. Sci.* 21 (1996) 308–311.
- [55] W. Parker, P.S. Song, *Biophys. J.* 61 (1992) 1435–1439.
- [56] S.B. Dev, C.K. Rha, F. Walder, *J. Biomol. Struct. Dyn.* 2 (1984) 431–442.
- [57] J.L. Eisele, J.P. Rosenbusch, *J. Biol. Chem.* 265 (1990) 10217–10220.
- [58] S. Hamada, K. Takeda, *J. Protein Chem.* 12 (1993) 477–482.
- [59] C.H. Papavoine, J.M. Aelen, R.N. Konings, C.W. Hilbers, F.J. Van de Ven, *Eur. J. Biochem* 232 (1995) 490–500.
- [60] H.W. Van Den Hooven, C.C. Doeland, M. Van De Kamp, R.N. Konings, C.W. Hilbers, F.J. Van de Ven, *Eur. J. Biochem.* 235 (1996) 382–393.
- [61] A. Muga, J.L. arrondo, T. Bellon, J. Sancho, C. Bernabeu, *Arch. Biochem. Biophys.* 300 (1993) 451–457.
- [62] W.R. Skach, M.C. Calayag, V.R. Lingappa, *J. Biol. Chem.* 268 (1993) 6903–6908.
- [63] H. Baubichon-Cortay, L.G. Baggetto, G. Dayan, A. Di Pietro, *J. Biol. Chem.* 269 (1994) 22983–22989.