

# Vibrational spectroscopy and computer modeling of proteins: solving structure of $\alpha_1$ -acid glycoprotein

Vladimír Kopecký Jr.<sup>a,b,\*</sup>, Rüdiger Ettrich<sup>c</sup>, Kateřina Hofbauerová<sup>b</sup> and Vladimír Baumruk<sup>a</sup>

<sup>a</sup> *Institute of Physics, Charles University, Ke Karlovu 5, CZ-12116 Prague 2, Czech Republic*

<sup>b</sup> *Department of Biochemistry, Faculty of Science, Charles University, Albertov 2030, CZ-12840 Prague 2, Czech Republic*

<sup>c</sup> *Laboratory of High Performance Computing, Institute of Physical Biology USB and Institute of Landscape Ecology AS CR, Zámek 136, CZ-37333 Nové Hradky, Czech Republic*

**Abstract.** This work introduces a new approach connecting vibrational spectroscopy with homology and energetic molecular modeling of proteins. Combination of both methods can compensate their disadvantages and result in realistic three-dimensional protein models. The approach is most powerful for membrane proteins or glycoproteins with high carbohydrate content where X-ray or NMR analysis is not always successful. Nevertheless, it can also serve as a tool of preliminary analysis of any protein with unknown structure. Power of the approach is demonstrated on human  $\alpha_1$ -acid glycoprotein. Its predicted structure published in [V. Kopecký Jr. et al., *Biochem. Biophys. Res. Commun.* **300** (2003), 41–46] is discussed in detail with respect to the approach and its general employment.

**Keywords:** Raman spectroscopy, Fourier transform infrared spectroscopy, molecular modeling, orosomucoid, binding site, thermal dynamics

## 1. Introduction

### 1.1. Determination of protein structures

Determination of the protein structure represents one of the key tasks of present molecular biology. There is an urgent need for three-dimensional (3D) structures of proteins for understanding their role in organisms and efficient drug design. Although X-ray and NMR spectroscopy can yield a great number of 3D protein structures, these methods are not always successful and certain types of proteins can hardly be approached by the techniques. For instance, membrane proteins or glycoproteins with high carbohydrate content belong to this category. Crystallography demands excellently purified proteins and special physico-chemical conditions and besides membrane or carbohydrate parts of proteins do not tend to constitute periodical structures. NMR spectroscopy requires highly concentrated samples of isotopic labeled proteins that in a case of glycoproteins or membrane proteins cannot be simply prepared and have to be obtained only by direct isolation from organisms.

---

\*Corresponding author: V. Kopecký Jr., Institute of Physics, Charles University, Ke Karlovu 5, CZ-12116 Prague 2, Czech Republic. Fax: +420 224 922 797; E-mail: kopecky@karlov.mff.cuni.cz.

For that reasons other spectroscopic methods are used as a simpler alternative although they cannot provide complete information about the 3D protein structure. Fluorescence and circular dichroism spectroscopy are most widely used for proteins. Another choice represent Raman and infrared (FTIR) spectroscopy – two complementary techniques of vibrational spectroscopy – that can provide good wealth of information about protein structure in comparison to other methods of optical spectroscopy. Nevertheless, due to its low resolution, vibrational spectroscopy data alone are not sufficient for building of 3D structural model and thus, as we propose and demonstrate here, they have to be coupled with theoretical methods to improve predictive ability with respect to protein 3D structure.

### 1.2. Coupling of vibrational spectroscopy with molecular modeling

If we would like to couple different methods, we should choose methods that can compensate their particular disadvantages and add their advantages. Rapid methods for estimation of protein structure are on the one side represented by all kinds of computer modeling and on the other side by different spectroscopic techniques. The former are criticized for no connection with reality, the latter for low structural resolution. However, combination of both approaches can compensate their disadvantages and result in realistic 3D protein models. To establish a continuous feedback to experimental data, the model is confronted with Raman and FTIR structural markers during the homology modeling process (restraint-based method).

Figure 1 depicts the logical scheme of the proposed approach. As the first step, vibrational spectroscopy enables to measure protein samples without any special requirements on preparation or conditions. Thus, we can easily obtain estimation of the secondary structure content and in some cases even a number of segments with regular secondary structure in the studied protein. Despite the fact that this information is loaded with error, a homology model structure of the protein should not be in steep discrepancy with it. The model should reflect these experimental data and, if not, a set of restrains leading to

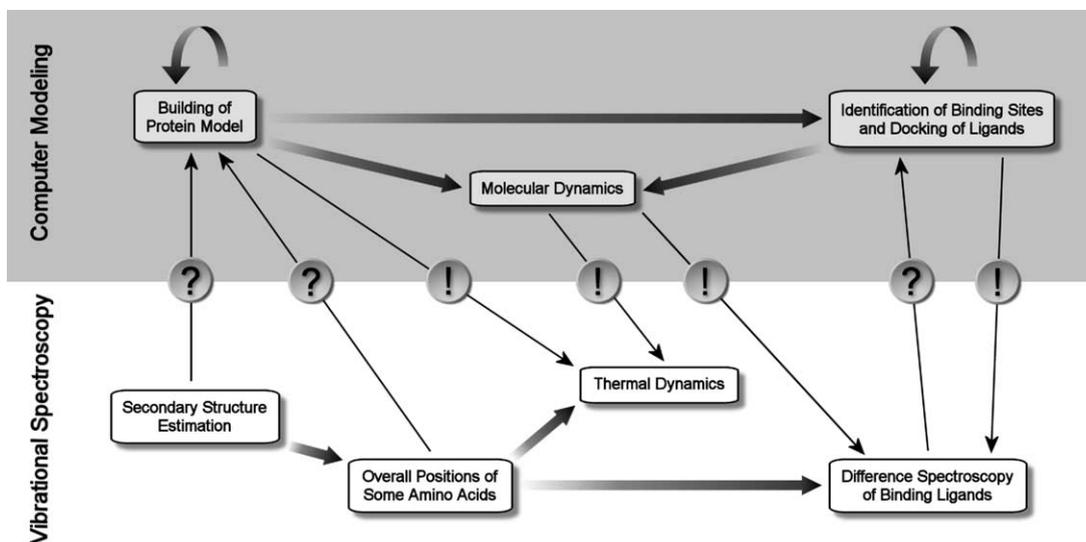


Fig. 1. Scheme of the new strategy for proper determination of protein structure combines vibrational spectroscopy with computer modeling. Interrogation and exclamation marks indicate collation of the models with experimental data and detail explanation of experiments, respectively. For details see introduction.

a different protein fold must be used for remodeling. Subsequently, more detail information concerning about some particular amino acid residues can be obtained from Raman spectra, e.g., local environments and torsion angles of aromatic residues, presence and conformation of S–S bridges, etc. This type of information particularly points to alignment errors or other shortcomings of the computer model and can be used as a starting geometry for its optimization, enlarging the set of constraints used for building of the protein model. After this step the 3D model of the protein consistent with vibrational spectral data is finally gained. Whether in energetically minimized structure only or after a molecular dynamics run, the model can explain many aspects of the behavior observed in thermal dynamics experiments, e.g., identification of amino acids residues surfacing up from the protein core upon heating. One of the most important features is the possible identification of binding sites and docking of ligands into the protein structure, which can lead to a better drug design. Vibrational difference spectroscopy can identify changes in the secondary structure and amino acids influenced by ligand binding. Nevertheless, the situation can be more complicated due to the amino acid composition in the binding site. Only the amino acids with aromatic side chains can be identified by Raman difference spectroscopy upon ligand binding. All structural aspects observed in difference spectra must be present in the binding site identified in the model structure. If not, another docking process must be used. When all experimentally demanded conditions are met, the computer model can identify the composition of the binding site in detail and thus predicts appropriate steps in mutagenesis or modification of ligands. We can conclude that this approach containing several self-corrective steps brings always a valuable contribution to the knowledge of the protein 3D structure, despite of corrections or shortcomings that have to be made, even if X-ray or NMR analysis succeeds in future.

Above mentioned approach has been successfully employed to solve unknown structure of different kinds of proteins, e.g., N-domain of membrane protein  $\text{Na}^+/\text{K}^+$ -ATPase and its ATP-binding site [1, 2], human haptoglobin – a large glycoprotein that naturally tends to oligomerize [3] and human  $\alpha_1$ -acid glycoprotein – a glycoprotein with extremely high carbohydrate content [4]. As an example, power of the approach is demonstrated on human  $\alpha_1$ -acid glycoprotein. Its predicted structure published in [4] is discussed in detail with respect to the approach and its general employment.

### 1.3. Human $\alpha_1$ -acid glycoprotein

Human  $\alpha_1$ -acid glycoprotein (AGP), also known as orosomucoid, represents an interesting example of a heavily studied protein with not yet resolved 3D structure. Despite of many years of crystallization experiments with AGP isolated only from blood serum [5], all attempts at its X-ray analysis failed [6] mainly due to its high carbohydrate content (ca. 42% of its weight of 41 kDa). Five heteropolysaccharide groups are linked via N-glycosidic bonds to asparaginy residues of the protein formed of 183 amino acids [7]. It is known that AGP, as a human blood plasma protein, plays a role under inflammatory or other pathophysiological conditions and is able to bind basic drugs, vanilloids, IgG3, heparin, and certain steroid hormones such as progesterone. Therefore, AGP is widely used in medical blood tests, however its biological function remains unknown [8]. Hence, every piece of structural information can bring valuable contribution to deeper understanding the role of AGP in organisms.

## 2. Materials and methods

All methods and materials discussed in this article were already described in detail [4]. Nevertheless, infrared spectra were recorded under slightly different conditions than described previously. Bruker IFS-

66/S FTIR spectrometer using a standard source, a KBr beamsplitter, and an MCT detector was used as before but samples were measured in a CaF<sub>2</sub> BioCell™ (BioTools) with 10 μm path length placed in a liquid cooling and heating jacket (BioJack, BioTools) connected to a programmable circulating bath (Neslab RTE-111). 1000 scans were collected with 4 cm<sup>-1</sup> spectral resolution and Happ–Genzel apodization. Spectral contribution from a buffer in the carbonyl stretching region was carefully corrected following the standard algorithm [9]. To be highly flexible in editing the restraints used as an input in accordance with the experimental data, the restraint-based approach of Modeller [10] was used for the main homology modeling step.

### 3. Results and discussion

α<sub>1</sub>-Acid glycoprotein (AGP) belongs to the lipocalin family of proteins, a heterogeneous group of extracellular proteins that bind a variety of small hydrophobic ligands. The second derivative of the FTIR spectrum of AGP (see Fig. 2), that allows resolution of overlapping spectral components, validate the high content of extended β-sheets (presence of the strong band at 1636 cm<sup>-1</sup>). Presence of the band at 1692 cm<sup>-1</sup> suggests that β-sheets are antiparallel [11] in agreement with the repeated +1 topology β-barrel presented in the model of the protein moiety (see Fig. 3). Lipocalin family is remarkably diverse at the sequence level, showing low levels of overall sequence conservation with pair-wise comparisons often falling well below 20%. Despite lack of high sequence similarity, lipocalin crystal structures are characterized by the already mentioned and well conserved β-barrel with a root mean square deviation value ( $C_{\alpha}$ ) of only 1.3–1.4 Å for the templates used.

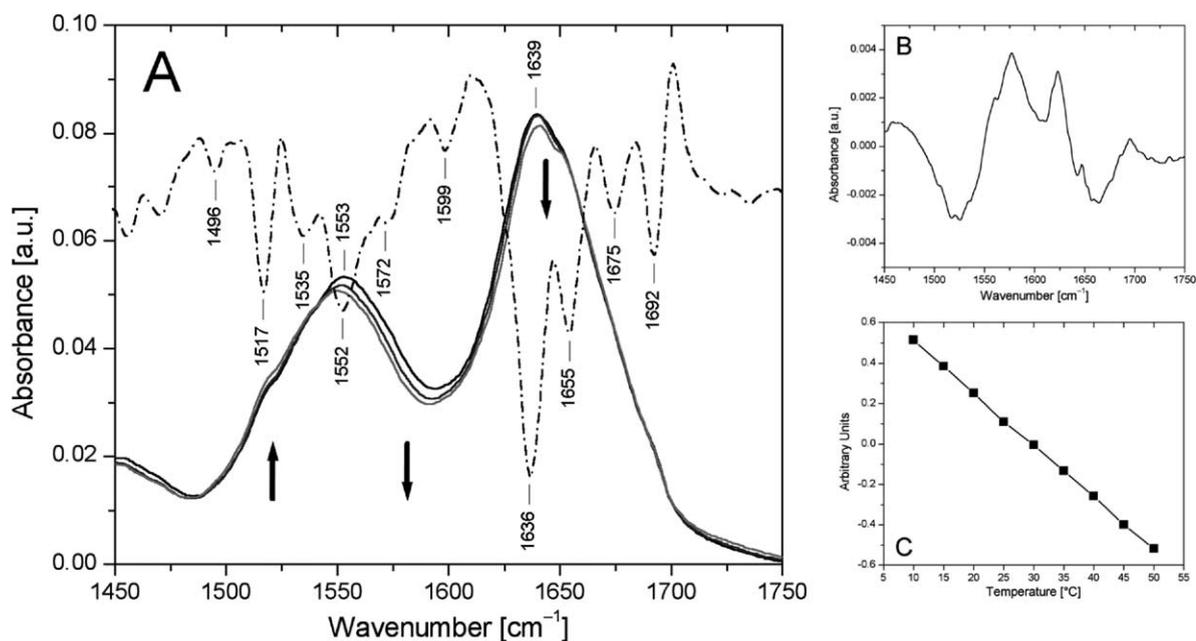


Fig. 2. FTIR spectra of native AGP in the 10–50°C temperature range. (A) Solid lines represent FTIR spectra of AGP recorded at 10, 30 and 50°C while arrows label band changes due to increasing temperature. The dash-dot line is associated with the second derivative (15 pts) of the spectrum at 10°C. (B) The second subspectrum of the PCA of the set of temperature dependent FTIR spectra of AGP. (C) The corresponding PCA coefficient associated with the second subspectrum depicted at (B).

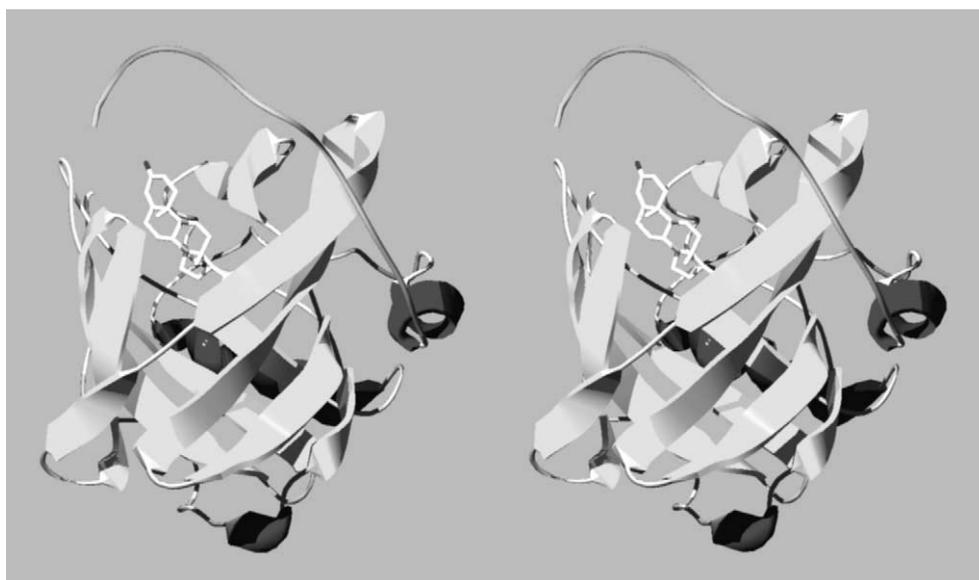


Fig. 3. Three-dimensional stereo picture of AGP. The picture shows the +1 topology  $\beta$ -barrel that AGP has common with other members of the lipocalin family. The hydrophobic pocket in the middle of the barrel holds a progesterone molecule.

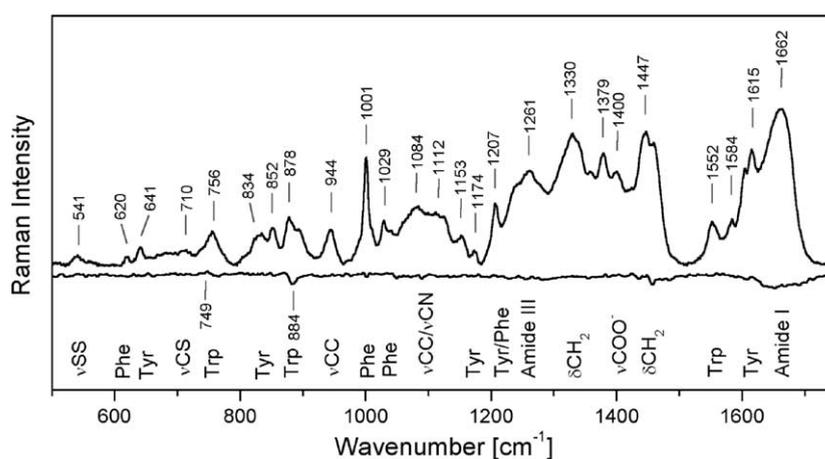


Fig. 4. (Top) Raman spectrum of native form AGP. (Bottom) Raman difference spectrum of native AGP minus AGP with bound progesterone. For band assignment see results and discussion section.

At this point, however, the model can only provide a rough estimation of the protein fold, and cannot serve as a model on the atomic level, e.g., as a basis for docking experiments. Therefore the model of the protein moiety of AGP was verified experimentally in as many details as possible following the scheme depicted on Fig. 1. The secondary structure estimation by least-squares analysis of FTIR amide I and II bands [12], at  $1639\text{ cm}^{-1}$  and  $1553\text{ cm}^{-1}$ , respectively, and of Raman amide I band [13], at  $1662\text{ cm}^{-1}$  (see Fig. 4), leads to a similar percentage representation of the secondary structure content as is presented in our model, i.e., 15% of  $\alpha$ -helix, 41% of  $\beta$ -sheet, 12% of  $\beta$ -turn, 8% of bend and 24% of unordered structure. Interesting feature of our protein model is a presence of unusual *trans-gauche-trans* conformation of S–S bridges, which corresponds to the presence of Raman marker band at  $541\text{ cm}^{-1}$

(Fig. 4) [14]. Positions of amino acids with aromatic side chains, with respect to their environment or surface of the protein, have been also determined from the Raman spectrum (Fig. 4) [4]. The model of the protein moiety reflects all these aspects but any longer energetic minimization or molecular dynamics of the complete structure leads to worse results regarding stereochemistry. Thus, all molecular dynamics studies of the complete protein should be carried out in the presence of the carbohydrate moiety, which strongly stabilize the structure. Unfortunately, molecular dynamics modeling in the presence of such a huge amount of sugar compound, 42%, does not lead to reliable results at the present state of the art. FTIR experiments with thermal dynamics are in agreement with this observation and support the important role of the sugar compound for the protein stability. Although no simply observable changes of FTIR spectra as a function of temperature can be seen in Fig. 2A, the principle component analysis (PCA) can resolve those changes. As a multivariate mathematical technique, PCA reduces spectra to their lowest dimension, thus each spectrum can be expressed as a linear combination of loading coefficients and orthogonal subspectra. The second subspectrum, which reflects changes in the spectral set, depicted in Fig. 2B corresponds to (a) rearrangement in  $\beta$ -structures – mainly decreasing content of turn structures – by bands at  $1641\text{ cm}^{-1}$  and  $1661\text{ cm}^{-1}$ , and (b) changes of the band at  $1518\text{ cm}^{-1}$  – probably surfacing of Tyr residues – with increasing temperature. The loading coefficients, depicted at Fig. 2C, reflects extreme thermal stability of AGP and, in general, corresponds to the “breathing” of  $\beta$ -barrel.

Finally, according to the scheme in Fig. 1, we explored ligand binding to AGP. Raman difference spectroscopy (Fig. 4) revealed proximity of Trp to the binding site by the shift of W17 mode at  $884\text{ cm}^{-1}$  (in the difference spectrum) indicating changes in Trp NH-bond donation. We found only one – Trp<sup>122</sup> – close to the hydrophobic pocket, thus it very precisely identified the binding site. Both FTIR and Raman spectroscopy showed increase of  $\beta$ -sheet – ca. 4%, and decrease of  $\alpha$ -helix – ca. 3%, upon progesterone binding. It is in an excellent agreement with behavior of  $\alpha$ -helix in the first loop above the  $\beta$ -barrel in the model, which is transformed into antiparallel  $\beta$ -sheet upon progesterone binding to the hydrophobic pocket located inside the  $\beta$ -barrel.

It is still questionable whether vibrational spectroscopy is quite sufficient for experimental verification of considerable aspects of the model. Thus, if the model should bring new valuable results then it must also explain previous experimental results, which were not used for the construction of the model.

Fluorescence measurements determined a distance between Trp and progesterone lower than  $5.5\text{ \AA}$  [15] and influence of progesterone binding on Trp [16] that is in excellent agreement with our model where Trp<sup>122</sup> is placed about  $3\text{ \AA}$  far from progesterone in the binding site. Reduction of progesterone binding upon chemical modification of lysine residues [15] can be explained by the presence of Lys<sup>39</sup> which is in our model located within the short  $\alpha$ -helical segment of the loop partially closing the binding pocket, thus any modification of this part should lead to lowering of the binding. Presence of Tyr<sup>27</sup> and Tyr<sup>65</sup> deep in the binding pocket plausibly explains the experimental fact that progesterone in the binding site protects up to two tyrosine residues from nitration of tetranitromethane [15]. Previous observation of decrease of  $\beta$ -sheet content in AGP upon heating by circular dichroism spectroscopy [17] is in agreement with our findings and is in accordance with a notion of  $\beta$ -barrel “breathing” in the model of the protein moiety. Irreversible motions of the AGP structure were observed upon heating under acidic conditions by UV-absorption spectroscopy [18]. The model shows that these changes can appear only on the periphery of the protein, where  $\alpha$ -helices are located and carbohydrate chains are linked. Last but not least, general physico-chemical properties of the binding pocket in our model are in good agreement with those determined by another theoretical method – quantitative structure-activity relationships method [19].

Our model was able to explain most of the previously determined experimental facts and thus the model structure can represent a valuable contribution for design of future experiments and their explanation. (Models of AGP in PDB-format are available upon request.)

#### 4. Conclusion

In this work usefulness of our approach that combines computer modeling of proteins with vibrational spectroscopy is illustrated on human  $\alpha_1$ -acid glycoprotein as an example. Despite the fact that this approach could be criticized for non-100% reliability of the 3D structures found, its aim is to encourage future work on the 3D protein structure and especially on quantitative structure-activity relationship investigations. We can also conclude as a proof of reliability of the method that the resulting 3D protein structure of  $\text{Na}^+/\text{K}^+$ -ATPase N-domain gained by our approach [1] is in excellent agreement with recently determined structures by NMR-spectroscopy [20] (root mean square deviation for the  $\text{C}_\alpha < 1.5 \text{ \AA}$ ) and X-ray analysis [21] (root mean square deviation for the  $\text{C}_\alpha < 1.4 \text{ \AA}$ ). Our approach can bring the most valuable and trustworthy information about three-dimensional structure and binding sites of proteins, such as membrane proteins and glycoproteins, that cannot be in near future approached by X-ray or NMR analysis. Thus, this approach does not represent only a useful rapid “first guess” method but in special cases it can also serve as a very powerful method.

#### Acknowledgements

The support from the Ministry of Education, Youth and Sports of the Czech Republic (projects No. MSM 113100001, MSM 113200002, and LN 00A141) and from the Grant Agency of the Czech Republic (project No. 206/03/D082) is gratefully acknowledged.

#### References

- [1] K. Hofbauerová, V. Kopecký Jr., R. Ettrich, O. Ettrichová and E. Amler, Secondary and tertiary structure of nucleotide-binding domain of  $\alpha$  subunit of  $\text{Na}^+/\text{K}^+$ -ATPase, *Biopolymers* **67** (2002), 242–246.
- [2] K. Hofbauerová, V. Kopecký Jr., R. Ettrich, M. Kubala, J. Teisinger and E. Amler, ATP-binding is stabilized by a stacking interaction within the binding site of  $\text{Na}^+/\text{K}^+$ -ATPase, *Biochem. Biophys. Res. Commun.* **306** (2003), 416–420.
- [3] R. Ettrich, W. Brandt, V. Kopecký Jr., V. Baumruk, K. Hofbauerová and Z. Pavlíček, Study of chaperone-like activity of human haptoglobin: conformational changes under heat shock conditions and localization of interaction sites, *Biol. Chem.* **383** (2002), 1667–1676.
- [4] V. Kopecký Jr., R. Ettrich, K. Hofbauerová and V. Baumruk, Structure of human  $\alpha_1$ -acid glycoprotein and its high-affinity binding site, *Biochem. Biophys. Res. Commun.* **300** (2003), 41–46.
- [5] K. Schmid, Preparation and properties of serum and plasma proteins XXIX, *J. Am. Chem. Soc.* **75** (1953), 60–68.
- [6] A. McPherson, M.L. Friedman and H.B. Halsall, Crystallization of  $\alpha_1$ -acid glycoprotein, *Biochem. Biophys. Res. Commun.* **124** (1984), 619–624.
- [7] K. Schmid, L.H. Chen, J.C. Occhino, E.A. Foster and K. Sperando, Topography of human plasma  $\alpha_1$ -acid glycoprotein, *Biochemistry* **15** (1976), 2245–2255.
- [8] T. Fournier, N. Medjoubi-N and D. Porquet, Alpha-1-acid glycoprotein, *Biochim. Biophys. Acta* **1482** (2000), 157–171.
- [9] F. Dousseau, M. Therrien and M. Pézolet, On the spectral subtraction of water from the FT-IR spectra of aqueous solutions of proteins, *Appl. Spectrosc.* **43** (1989), 538–542.
- [10] A. Sali and T.L. Blundell, Comparative modelling by satisfaction of spatial restraints, *J. Mol. Biol.* **234** (1993), 779–815.
- [11] N. Yamada, K. Ariga, M. Naito, K. Matsubara and E. Koyama, Regulation of  $\beta$ -sheet structures within amyloid-like  $\beta$ -sheet assemblage from tripeptide derivatives, *J. Am. Chem. Soc.* **120** (1998), 12192–12199.

- [12] F. Dousseau and M. Pézolet, Determination of the secondary structure content of proteins in aqueous solutions from their amide I and amide II infrared bands. Comparison between classical and partial least-squares methods, *Biochemistry* **29** (1990), 8771–8779.
- [13] R.W. Williams, Protein secondary structure analysis using Raman amide I and III spectra, *Methods Enzymol.* **130** (1986), 311–331.
- [14] H.E. Van Wart and H.A. Scheraga, Raman spectra of strained disulfides. Effect of rotation about sulfur-sulfur bonds on sulfur stretching frequencies, *J. Phys. Chem.* **80** (1976), 1823–1832.
- [15] T. Kute and U. Westphal, Chemical modification of  $\alpha_1$ -acid glycoprotein for characterization of the progesterone binding site, *Biochim. Biophys. Acta* **168** (1976), 195–213.
- [16] J.R. Albani, Binding effect of progesterone on the dynamics of  $\alpha_1$ -acid glycoprotein, *Biochim. Biophys. Acta* **1336** (1997), 349–359.
- [17] M. Kodíček, A. Infanzón and V. Karpenko, Heat denaturation of human orosomuroid in water–methanol mixtures, *Biochim. Biophys. Acta* **1246** (1995), 10–16.
- [18] K. Hofbauerová, V. Kopecký Jr., J. Sýkora and V. Karpenko, Thermal stability of the human blood serum acid  $\alpha_1$ -glycoprotein in acidic media, *Biophys. Chem.* **103** (2003), 25–33.
- [19] F. Hervé, G. Caron, J.-C. Douché, P. Gaillard, N.B. Rahman, A. Tsantili-Kakoulidou, P.-A. Carrupt, P. D'Athins, J.-P. Tillement and B. Testa, Ligand specificity of the generic variants of human  $\alpha_1$ -acid glycoprotein: generation of the three-dimensional quantitative structure–activity relationship model for drug binding to the A variant, *Mol. Pharmacol.* **54** (1998), 129–138.
- [20] M. Hilge, G. Siegal, G.W. Vuister, P. Guntert, S.M. Gloor and J.P. Abrahams, ATP-induced conformational changes of the nucleotide-binding domain of  $\text{Na}^+, \text{K}^+$ -ATPase, *Nat. Struct. Biol.* **10** (2003), 468–474.
- [21] K.O. Håkansson, The crystallographic structure of  $\text{Na}^+, \text{K}^+$ -ATPase N-domain at 2.6 Å resolution, *J. Mol. Biol.* **332** (2003), 1175–1182.



**Hindawi**

Submit your manuscripts at  
<http://www.hindawi.com>

