2010

Amino acid transport in thermophiles: characterization of an arginine-binding protein in Thermotoga maritima. 2. Molecular organization and structural stability

Andrea Scirè

Anna Marabotti

Maria Staiano

Luisa Iozzion

Matthew S. Luchansky

See next page for additional authors

Recommended Citation

Scirè, Andrea, Anna Marabotti, Maria Staiano, Luisa Iozzino, Matthew S. Luchansky, Bryan S. Der, Jonathan D. Dattelbaum, Fabio Tanfani, and Sabato D'auria. "Amino Acid Transport in Thermophiles: Characterization of an Arginine-binding Protein in Thermotoga Maritima. 2. Molecular Organization and Structural Stability." *Molecular BioSystems* 6, no. 4 (2010): 687. doi:10.1039/b922092e.

Amino acid transport in thermophiles: characterization of an arginine-binding protein in *Thermotoga maritima*. 2. Molecular organization and structural stability

Andrea Scirè,[a] Anna Marabotti,[b] Maria Staiano,[c] Luisa Iozzino,[c] Matthew S. Luchansky,[d] Bryan S. Der,[d] Jonathan D. Dattelbaum,[d] Fabio Tanfani[a] and Sabato D'Auria*[c]

Received 22nd October 2009, Accepted 12th January 2010
First published as an Advance Article on the web 22nd February 2010
DOI: 10.1039/b922092e

ABC transport systems provide selective passage of metabolites across cell membranes and typically require the presence of a soluble binding protein with high specificity to a specific ligand. In addition to their primary role in nutrient gathering, the binding proteins associated with bacterial transport systems have been studied for their potential to serve as design scaffolds for the development of fluorescent protein biosensors. In this work, we used Fourier transform infrared spectroscopy and molecular dynamics simulations to investigate the physicochemical properties of a hyperthermophilic binding protein from *Thermotoga maritima*. We demonstrated preferential binding for the polar amino acid arginine and experimentally monitored the significant stabilization achieved upon binding of ligand to protein. The effect of temperature, pH, and detergent was also studied to provide a more complete picture of the protein dynamics. A protein structure model was obtained and molecular dynamic experiments were performed to investigate and couple the spectroscopic observations with specific secondary structural elements. The data determined the presence of a buried β-sheet providing significant stability to the protein under all conditions investigated. The specific amino acid residues responsible for arginine binding were also identified. Our data on dynamics and stability will contribute to our understanding of bacterial binding protein family members and their potential biotechnological applications.

Introduction

A large variety of ligand-binding proteins have been utilized for the design platforms of non-consuming optical biosensors capable of targeting many naturally-occurring ligands, including sugars, anions, and amino acids.[1–6] The *Escherichia coli* family of periplasmic binding proteins (PBPs) possesses many members with diverse ligand affinity coupled to a highly conserved three-dimensional structural organization. PBPs are typically composed of a single polypeptide chain that folds into two easily distinguishable domains that are connected by a hinge region. Binding of ligand results in a large rotational-bending movement of the two protein domains.[7,8] Using optical techniques such as environmental-sensitivity, FRET, or plasmonic interactions, rational placement of fluorescent probes allows transduction of the binding event into a quantifiable optical signal that varies as a function of ligand concentration.[9,10] In addition to detecting the natural set of ligands associated with native forms of ligand-binding proteins, re-engineering the binding site of these proteins as a design scaffold has significantly expanded the number of small molecule analytes for which sensors may be constructed.[11,12] However, there is a significant cost in thermal stability for these re-engineered biosensors as a result of the large number of mutations required to alter the binding specificity.[13] As a consequence it appears clear that proteins isolated from thermophilic organisms possess added intrinsic value in the design of new biosensing technology that features enhanced stability.

Thermotoga maritima is a hyperthermophilic eubacterium whose genome[14] contains a number of ABC transport systems which typically involve the presence of a soluble, ligand binding protein to accomplish nutrient uptake.[15] Recently, a thermostable arginine-binding protein (ArgBP) from *T. maritima* has been expressed as recombinant protein in *E. coli*.[16] ArgBP has been purified to homogeneity and was found to occur as a monomer with a molecular mass of 27.7 kDa that binds Arg with micromolar affinity. Arginine (Arg) is present in human bodily fluids, such as serum and urine, and is derived from the catabolism of proteins containing arginine and methylated arginine residues.[17] Bodily production of nitric oxide (NO), the critical modulator of blood flow and blood pressure,[18] occurs through metabolism of arginine by the specific enzyme nitric oxide synthase (NOS). Levels of available arginine are very important for NO synthesis in patients with hypercholesterolemia or atherosclerosis.

[a] *Department of Biochemistry, Biology, and Genetics, Università Politecnica delle Marche, Ancona, Italy*
[b] *Institute of Food Sciences, CNR, Avellino, Italy*
[c] *Laboratory for Molecular Sensing, IBP-CNR, Via Pietro Castellino, 111, 80131 Napoli, Italy. E-mail: s.dauria@ibp.cnr.it; Fax: +39-0816132277; Tel: +39-0816132250*
[d] *Department of Chemistry, University of Richmond, Richmond, VA 23173, USA*

Thus, detection of arginine levels in bodily fluids is useful for diagnosis and treatment of these diseases.

Before utilizing a protein as the basis for a sensing device, it is of high importance to fully characterize the biomolecule with respect to its stability in the potential operating conditions. In this study we performed a structural characterization of ArgBP in a wide range of temperature, at different pH values, and in the absence and in the presence of the detergent SDS using Fourier-transform infrared (FT-IR) spectroscopy. In addition, a model of the protein structure was realized and a series of simulation experiments were performed in different physicochemical conditions.

Materials and methods

L-arginine and L-asparagine were purchased from Sigma, deuterium oxide (99.9% D_2O), DCl and NaOD were purchased from Aldrich. All other chemicals were commercial samples of the best available quality.

Preparation of samples for FTIR analysis

T. maritima arginine-binding protein (ArgBP) was purified as previously described.[16] About 1.5 mg of protein, dissolved in the buffer used for its purification, was concentrated to a volume of approximately 50 μl using a "10 K Centricon" micro concentrator (Amicon) at 3000 × g and 4 °C. Afterwards, 250 μl of 25 mM Hepes/NaOD pD 7.5 (buffer A), or 250 μl of 25 mM Hepes/NaOD/2.4 mM arginine pD 7.5 (buffer B), or 250 μl of 25 mM Hepes/NaOD/2.4 mM asparagine pD 7.5 (buffer C), were added and the protein solution was concentrated again. The pD corresponds to the pH meter reading +0.4.[19] The concentration-dilution procedure was repeated several times in order to completely replace the original buffer with buffer (A) or (B) or (C). Altogether the washings took 24 h, which is the time of contact of the protein with the D_2O medium prior infrared analysis. In the last washing step, the protein solution was concentrated to a final volume of 40 μl and used for FT-IR measurements.

ArgBP at pD 7.5 was extremely heat-resistant as noted by the incomplete denaturation (loss of secondary structure) at 99.5 °C. For this reason and to further characterize the structural properties of the protein, we analyzed ArgBP also at pD 10.5 (buffer D, 50 mM CAPS pD 10.5), or at pD 7.5 in the presence of SDS. In particular, two additional buffers were prepared for experiments in the presence of SDS including 25 mM Hepes/NaOD containing 1% SDS (w/v) pD 7.5 (buffer E), and 25 mM Hepes/NaOD containing 2.5% SDS (w/v) pD 7.5 (buffer F). The concentration-dilution procedure using buffer (A) was applied to these additional two protein samples. In the last washing phase, the protein solution was concentrated to approximately 100 μL and then 100 μL of buffer (E) or (F) was added. Finally, the protein solutions containing SDS were concentrated to a volume of 40 μl and used for FT-IR measurements. Because of the formation of micelles, during the concentration process the majority of SDS remained in the micro concentrator. However, part of the detergent probably passed through the pores of the filter. Hence, in the final concentrated protein samples the SDS concentration was checked using a calibration curve obtained by monitoring the intensity of the SDS symmetric methylene stretching vibration band (2854 cm^{-1})[20] as a function of SDS concentration. This analysis found that the actual SDS concentration was 1.5% or 3.5% (w/v) in samples where buffer (E) or (F), respectively, was used.

FT-IR measurements

The concentrated ArgBP solutions were injected into a thermostatted Graseby-Specac 20500 cell (Graseby-Specac Ltd, Orpington, Kent, UK) fitted with CaF_2 windows and a 25 μm Teflon spacer. FT-IR spectra were recorded by means of a Perkin-Elmer 1760-× Fourier transform infrared spectrometer using a deuterated triglycine sulfate detector and a normal Beer-Norton apodization function. At least 24 h before as well as during data acquisition, the spectrometer was continuously purged with dry air at a dew point of −70 °C. Spectra of samples and buffers were acquired at 2 cm^{-1} resolution under the same temperature and scanning conditions. In the thermal denaturation experiments, the temperature was raised in 5 °C increments from 20 °C to 95 °C using an external bath circulator (Haake F3). Additional spectra were recorded at 98 °C, 99 °C and 99.5 °C. The actual temperature in the cell was controlled by a thermocouple placed directly onto the CaF_2 windows. Before spectrum acquisition, samples were maintained at the desired temperature until thermal equilibrium was reached which typically occurred in 6 min. Spectra were collected and processed using the "Spectrum" software from Perkin-Elmer. Subtraction of the signal due to the D_2O bending absorption close to 1220 cm^{-1} was performed as previously described.[21] The deconvoluted parameters were set with a gamma value of 2.5 and a smoothing length of 60. Second derivative spectra were calculated over a 9-data-point range (9 cm^{-1}).

Comparative modelling of the structure of ArgBP and ArgBP/Arg

The structures of the ligand-free and ligand-bound forms of ArgBP were predicted by homology modeling using the program MODELLER version 9.5.[22] For ArgBP/Arg, the structure of Arg-, Lys-, His-binding protein ArtJ from *Geobacillus stearothermophilus* bound to Arg[23] available in PDB database[24] (PDB code: 2Q2A) was selected as suitable template after a BLAST search.[25] The same homologous structure was identified by the fold recognition servers SAM-T08[26] and FUGUE.[27] In addition to the ArtJ template, information was taken from the structure of the open unliganded form of the Gln-binding protein from *E. coli* (PDB code: 1GGG).[28]

To take common features of the PBP superfamily into account, a preliminary alignment was performed with the program T-Coffee[29] using sequences from all structures of similar proteins found with a BLAST search against the PDB database. The PredictProtein server[30] was then used to perform predictions of the type and position of secondary structure elements in ArgBP, which were compared with the structural information present for each PDB file in PDBsum database (http://www.ebi.ac.uk/pdbsum/). A few manual adjustments to optimize the placement of gaps were performed

before extracting the final alignment of target and template to be used for the following steps.

Using MODELLER we created ten different models of ArgBP (residues 18–246 of the sequence deposited in UniProt database[31]), for both the open and closed forms. To choose the best model, we assessed their quality with the discrete optimized protein energy (DOPE) score from MODELLER, and also with the programs PROCHECK[32] and ProsaII.[33] The stereochemical parameters of the best model obtained for ArgBP (open form) showed 91.2% of the residues in the most favored regions of the Ramachandran plot, and only 2 residues (1%) in disallowed regions (the analysis performed on the template showed 91.9% and 0% of residues for most favored and disallowed regions, respectively). ProsaII z-score (−11.25) is also in the range of scores typically found in proteins of similar sequence length[33] and is similar to that of template (−12.13). We also analyzed the energetic profile calculated by ProsaII on the whole structure, and the chosen model displayed the optimal profile, with no positive peaks indicating errors in the structure (data not shown). Similar results were obtained for the model of ArgBP in close form: 89.2% of the residues were found in the most favored regions of the Ramachandran plot, only 2 residues (1%) in disallowed regions, and a ProsaII *Z*-score of −11.23 with a completely negative energetic profile. These data demonstrate that we obtained high quality models of the open and closed conformations of ArgBP.

The model ArgBP/Arg structure was created with the aid of InsightII tools (Version 2000.1, Accelrys, Inc.; 2000) by merging the structure of the ligand, taken from the template, into the structure of the close form of the protein. Then, a mild optimization was applied to reduce steric clashes using 500 steps of Steepest Descent method, with a final gradient of 0.01 kcal mol^{-1}. CVFF force field developed for InsightII was used to assign potentials and charges. All atoms were allowed to relax with no constraints. This procedure represents the best compromise between the need for relieving steric clashes and the risk of distorting the geometry of the protein with deep and extensive minimization. The control of the final quality of the model was performed again with PROCHECK and ProsaII. The quality of the model of ArgBP/Arg is slightly worse than that of the original close unliganded model (85.8% and 2.0% of residues in most favored and disallowed regions of the Ramachandran plot, respectively; ProsaII z-score −10.96). However, the ArgBP/Arg model is still acceptable, with no major distortion of the protein's structure.

Molecular dynamics simulations and analysis of the results

Molecular dynamics (MD) were carried out using the program GROMACS version 4.0.5[34,35] running in parallel (MPI) on the supercomputer "CRESCO", formed by 300 nodes in SMP (total number of cores: 2700) and 34 multiprocessor servers for specific functions, all interconnected each others by an Infiniband net. The GROMOS96 force field[36] was used throughout the simulations. To simulate the variation of pH, the pK_a values of acidic and basic residues were evaluated by using the web server H^{++}[37] (http://biophysics.cs.vt.edu/H++) based on the work by Bashford and Karplus.[38] Then, the ionizable residues were protonated according to their state at the specified pH. Each system was then included in a triclinic box with a distance of 1 nm per side from the protein, filled with water molecules (SPC model)[39] and Na^+ ions added to neutralize the net negative charge of the whole system, replacing the corresponding number of water molecules. Periodic boundary conditions were used to exclude surface effects.

A preliminary energy minimization step with a tolerance of 500 kJ mol^{-1} nm^{-1} was run with the Steepest Descent method. All bonds were constrained using P-LINCS.[40] After minimization, a short MD simulation (20 psec) with position restraints using NVT ensemble, followed by another 20 psec simulation using NPT ensemble, was applied to each system to soak the solvent into the macromolecule. A time step of 2 fs was used in all cases. The systems were coupled to a temperature bath at 27 °C using a velocity rescaling thermostat with a stochastic term[41] and, in the case of the NPT ensemble, to a barostat at a pressure of 1 atm using Berendsen's method.[42] Long-range electrostatics were handled using the PME method.[43] Cut-offs were set at 0.9 nm for Coulombic interactions and at 1.4 nm for van der Waals interactions. Finally, the production MD simulations were carried out with the same settings, but without any position restraints and using NPT ensemble. Five subsequent MD simulations of 1 ns each (the final conformation of each simulation was used as input for the following simulation at higher temperature) were carried out at 27 °C, 60 °C, 75 °C, 80 °C, and 95 °C at a constant pressure of 1 bar using velocity rescaling thermostat[41] and Berendsen's barostat.[42] The final MD simulation at 100 °C was 5 ns long, for a global duration of 10 ns.

Next, several analyses were conducted using programs built within the GROMACS package, and results were visualized and elaborated with the aid of the freeware program Grace (http://plasma-gate.weizmann.ac.il/Grace). The energy components were analyzed to verify the stabilization of the system. For each simulation an "average" structure representative of the trajectory was calculated, not including hydrogen atoms. These "average" structures were saved in .pdb format and were subsequently minimized with the Steepest Descent method as described above. Visualization and analysis of model features was carried out using InsightII facilities. The percentage of residues embedded in secondary structure elements and their variation during the simulations was evaluated using the program DSSP.[44] The analysis of the accessibility of the residues was made using NACCESS.[45] The analysis of cavities in the proteins was made using the program AVP[46] by using a probe of 0.5 Å to assess the packing of the molecules. The presence of salt bridges was inferred using the criteria by Kumar and Nussinov[47] and also taking into account the protonation state of the residues.

Results and discussion

FT-IR measurements

Fig. 1 shows the absorbance (A), the deconvoluted (B), and the second derivative (C) infrared spectra of ArgBP at 20 °C. Within the 1700–1620 cm^{-1} interval (amide I′ band), the deconvoluted and second derivative spectra show a number

of bands assignable to secondary structural elements.[48,49] The presence of β-sheets and α-helices is revealed by the 1638.7 cm^{-1} and 1651.9 cm^{-1} bands, respectively. The absorption at 1673.2, 1678.9, and 1685.8 cm^{-1} may be due to β-sheets and/or turns.[49,50] Bands below 1620 cm^{-1} are caused by absorption of amino acid side chains.[51–53] In particular, the 1515 cm^{-1} band originates from vibrations of tyrosine, the 1583 cm^{-1} band is assigned to ionized carboxyl groups of aspartic acid and/or to arginyl residues, and the 1614 cm^{-1} band may be due to vibration of arginine. The 1550 cm^{-1} band is associated with residual amide II band absorption (1600–1500 cm^{-1} range), *i.e.* the absorption remaining after H/D exchange of the amide hydrogen atoms of the polypeptide chain. Indeed, in H_2O medium, the intensity of amide II band is approximately 2/3 that of the amide I band. In D_2O medium, the intensity of the amide II band decreases as a consequence of the exchange of amide hydrogen atoms with deuterium. The greater is the decrease in the amide II band intensity, the greater is the extent of the H/D exchange. A large H/D exchange indicates a large accessibility of the solvent (D_2O) to the protein.[54–56] The presence of the small 1550 cm^{-1} band indicates that a portion of amide hydrogen atoms were not exchanged with deuterium during the preparation of the protein sample.

The amide I′ band of ArgBP/Arg and ArgBP/Asn spectra are superimposable onto the amide I′ band of ArgBP spectrum (spectra not shown), indicating that the presence of arginine and asparagines residues do not affect the secondary structure of the protein.

Thermal stability of ArgBP

Information on temperature-induced structural changes of ArgBP in the absence and in the presence of asparagine or arginine can be obtained from data reported in Fig. 2. In particular, Fig. 2 shows the deconvoluted spectra of ArgBP (Fig. 2A), ArgBP/Asn (Fig. 2B) and ArgBP/Arg (Fig. 2C). Each panel A, B or C is composed of superimposed spectra in the 20–65 °C range (bottom set of spectra) and of superimposed spectra in the 65–99.5 °C range (upper set of spectra).

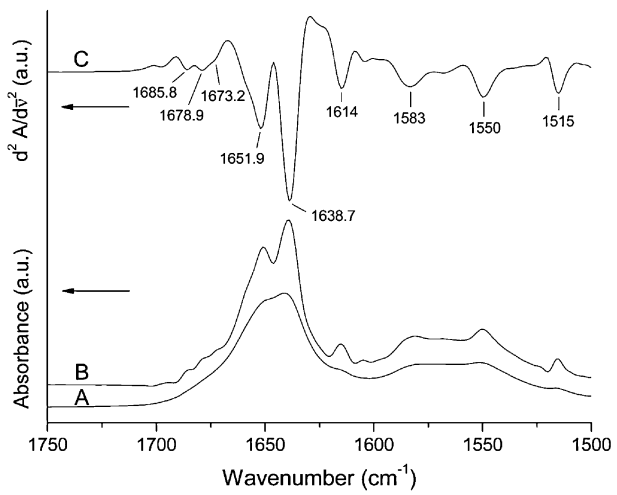


Fig. 1 Absorbance (A), deconvoluted (B), and second derivative (C) infrared spectra of ArgBP at 20 °C and pD 7.5.

Fig. 2D and E display the temperature-dependent changes of α-helix and β-sheet band intensities, respectively. The plotted data were extracted from the deconvoluted spectra showed in the Fig. 2A–C. Comparison of the spectra of ArgBP (Fig. 2A) and of ArgBP/Asn (Fig. 2B) reveals that the amide I′ band outline (1700–1600 cm^{-1}) is almost unchanged from 20 °C to 65 °C (lower set of spectra). The amide I′ band shape remained unchanged until 90 °C (upper set of spectra) for both ArgBP and ArgBP/Asn, indicating no significant loss of secondary structure elements in the temperature range from 20 °C to 90 °C. The absence of evident changes in the amide I′ band of ArgBP and ArgBP/Asn up to 90 °C is indicative of the high thermostability of ArgBP. However, at temperatures higher than 90 °C, a significant decrease in intensity of the α-helix and β-sheet bands and a concomitant appearance of two new bands at 1684 cm^{-1} and at 1616 cm^{-1} indicate the partial loss of secondary structural elements and protein aggregation phenomena, respectively. Indeed, the presence of the two bands at 1684 cm^{-1} and at 1616 cm^{-1} reflect protein intermolecular interactions (aggregation) that usually take place when proteins undergo denaturation.[56,57–60] Since the temperature-induced conformational changes in ArgBP and in ArgBP/Asn are the same, the data reported in Fig. 2A and B indicate that asparagine does not bind to ArgBP and that the protein begins to lose its secondary structural elements at temperatures above 90 °C. This is also shown in Fig. 2D and E, which display the decrease in intensity of the α-helix and β-sheet bands as function of temperature.

In contrast, the spectra of ArgBP in the presence of arginine (Fig. 2C) show no significant changes in the amide I′ band up to 99.5 °C, except a very minor decrease in the α-helix band intensity (see also Fig. 2D and E). These results indicate that arginine binds to ArgBP and that the binding exerts an important stabilizing effect on the protein structure.

The decrease in residual amide II band intensity shown in panels 2A and 2B (upper set of spectra) is due to partial loss of secondary structure that allows the protein to undergo further H/D exchange.[61] On the other hand, the decrease in residual amide II band intensity showed in panels 2A, 2B (bottom set of spectra) and in panel C is due to increase in molecular dynamics since no significant loss of secondary structure occurs in the range of temperature reported.[61]

The behavior of ArgBP in alkaline and denaturing conditions was also studied. Fig. 3 shows the deconvoluted spectra of ArgBP (control), ArgBP/10.5, ArgBP/1.5SDS, and ArgBP/3.5SDS, at 20 °C. Comparison of ArgBP/10.5, ArgBP/1.5SDS, and ArgBP/3.5SDS spectra with control spectrum reveals differences in the relative intensities of α-helix and β-sheet bands, indicating that high pD (spectrum A) or the presence of SDS (spectrum B and C) affect the secondary structure of the protein. An additional evident difference registered is the lower intensity of the 1550 cm^{-1} band in the ArgBP/10.5, ArgBP/1.5SDS, and ArgBP/3.5SDS spectra with respect to the control spectrum. This indicates that high pD or the presence of SDS allow a deeper contact of D_2O with the protein, probably as a consequence of changes in secondary structure and/or as a consequence of a relaxed tertiary structure.[61]

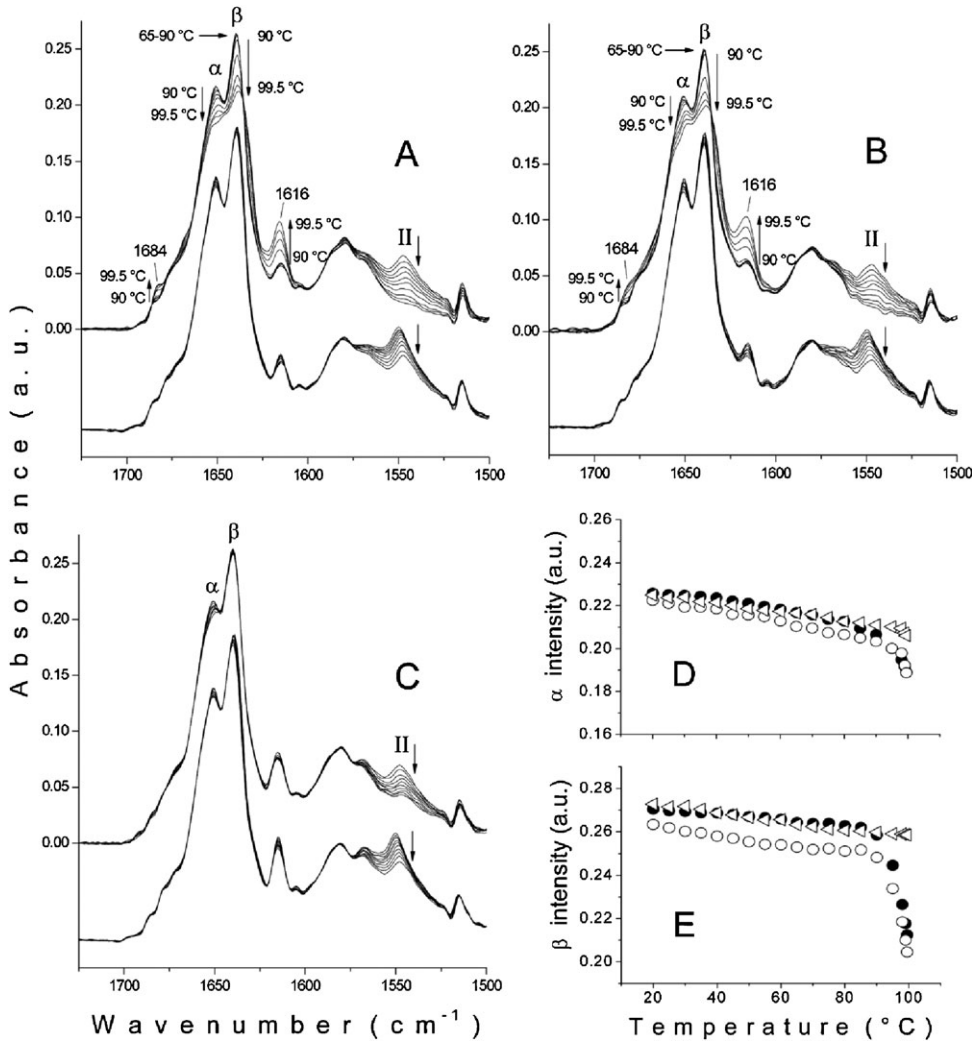


Fig. 2 Temperature-dependent changes in the deconvoluted spectra of ArgBP, ArgBP/Asn, and ArgBP/Arg at pD 7.5 ArgBP (panel A), ArgBP/Asn (panel B) and ArgBP/Arg (panel C). Overlayed spectra are reported in panels A–C in 5 °C steps from 20 °C to 65 °C (bottom set of spectra) and from 65 °C to 99.5 °C (upper set of spectra). Panel D and panel E show the changes in α-helix and β-sheet band intensity for ArgBP (full circle), ArgBP/Asn (open circle) and ArgBP/Arg (open triangle) plotted *versus* temperature. Data are extracted from the deconvoluted spectra shown in panels A–C. The symbols α, β and II stand for α-helices, β-sheets and residual amide II band, respectively.

Thermal stability of ArgBP, ArgBP/10.5, ArgBP/1.5SDS, and ArgBP/3.5SDS

Fig. 4 shows the deconvoluted spectra of ArgBP (control), ArgBP/10.5, ArgBP/1.5SDS, and ArgBP/3.5SDS as a function of temperature. The amide I′ band shape undergoes changes with the increase in temperature in all protein samples. The spectra indicate that the protein at pD 10.5 or in the presence of SDS is less thermostable with respect to the protein under control conditions (see previous data on ArgBP). Indeed, the spectra of ArgBP/10.5, ArgBP/1.5SDS, and ArgBP/3.5SDS do not show significant changes in the 20–65 °C temperature range (spectra not shown). On the other hand, at temperature higher than 65 °C, the spectra indicate an important loss in the secondary structural elements consistent with the presence of protein denaturation phenomena. In particular, Fig. 4B (ArgBP/10.5) shows that the β-sheet band intensity decreases until 90 °C without marked changes in the position of the band maximum. At 95 °C there is a further decrease of the β-sheet band intensity with a concomitant shift of the band maximum to lower wavenumber. At higher temperatures (till 99.5 °C) the β-sheet band intensity and position do not change significantly suggesting that the remaining secondary structural elements are particularly heat-resistant. It is noteworthy that the loss of β-sheet is not accompanied by the appearance of two new bands at 1684 cm^{-1} and at 1616 cm^{-1} (evidence of protein aggregation) as observed in the control sample.

In the case of ArgBP/1.5SDS (Fig. 4C) and ArgBP/3.5SDS (Fig. 4D) the increase in temperature leads to the complete loss of secondary structure. Indeed, at 99.5 °C the amide I′ band decreased in intensity, but increased in width, and becoming almost featureless. These are typical phenomena observed when proteins undergo to denaturation. In particular, the loss of secondary structure is shown by the absence of the β-sheet band in the spectrum of either ArgBP/1.5SDS or ArgBP/3.5SDS samples, indicating that the presence of the detergent destabilises the whole protein structure at high temperatures.

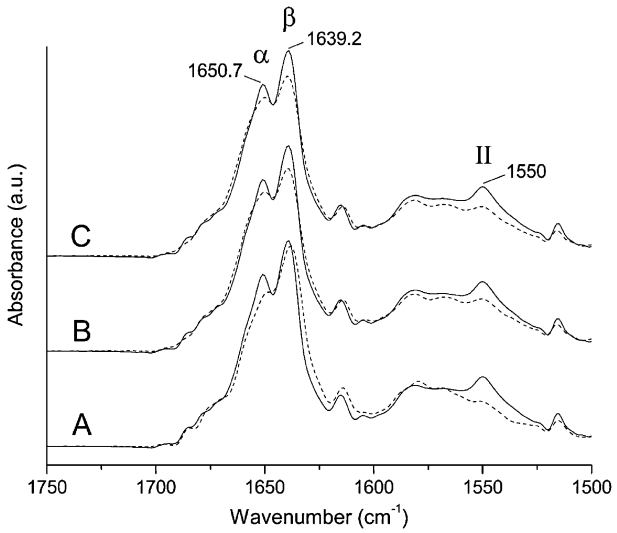


Fig. 3 Deconvoluted spectra of ArgBP (control), ArgBP/10.5, ArgBP/1.5SDS, and ArgBP/3.5SDS, at 20 °C. Continuous line in (A–C) represents the control spectrum of ArgBP. Dashed lines in (A–C) represent spectra of ArgBP/10.5 (A), ArgBP/1.5SDS (B), and ArgBP/3.5SDS (C), respectively. The symbols α, β, and II, stand for α-helix, β-sheet, and residual amide II band absorption, respectively.

The complete unfolding that is only observed in the presence of the detergent at high temperatures suggests that the residual population of β-sheet observed in ArgBP/10.5 at 99.5 °C might be buried in a hydrophobic protein environment which is not affected by alkaline pD but is sensitive to presence of the SDS.

In both ArgBP/1.5SDS and ArgBP/3.5SDS samples, the increase in temperatures causes the disappearance of the small 1614 cm^{-1} band. Amino acid side-chain absorption occurs below 1620 cm^{-1}, and it is possible that the 1614 cm^{-1} band might arise from arginine absorption.[51–53] However, it has been shown that bands around 1617 cm^{-1} may also be due to absorption of β-strands not accessible to the solvent[62] or located in a hydrophobic environment.[63] This finding is in agreement with our hypothesis that a population of β-sheet might be buried in a hydrophobic environment (see above) and that such buried β-sheet (1614 cm^{-1} band) are lost only in the presence of SDS and at high temperature. On the other hand, the absorption of arginine at 1614 cm^{-1} is not excluded. It might be that denaturation of the protein (complete loss of secondary structure) induces changes in the arginine interactions modifying its absorption characteristics.

In conclusion, the FTIR data indicate that (1) ArgBP is very stable until 90 °C at pD 7.5; (2) at temperature higher than

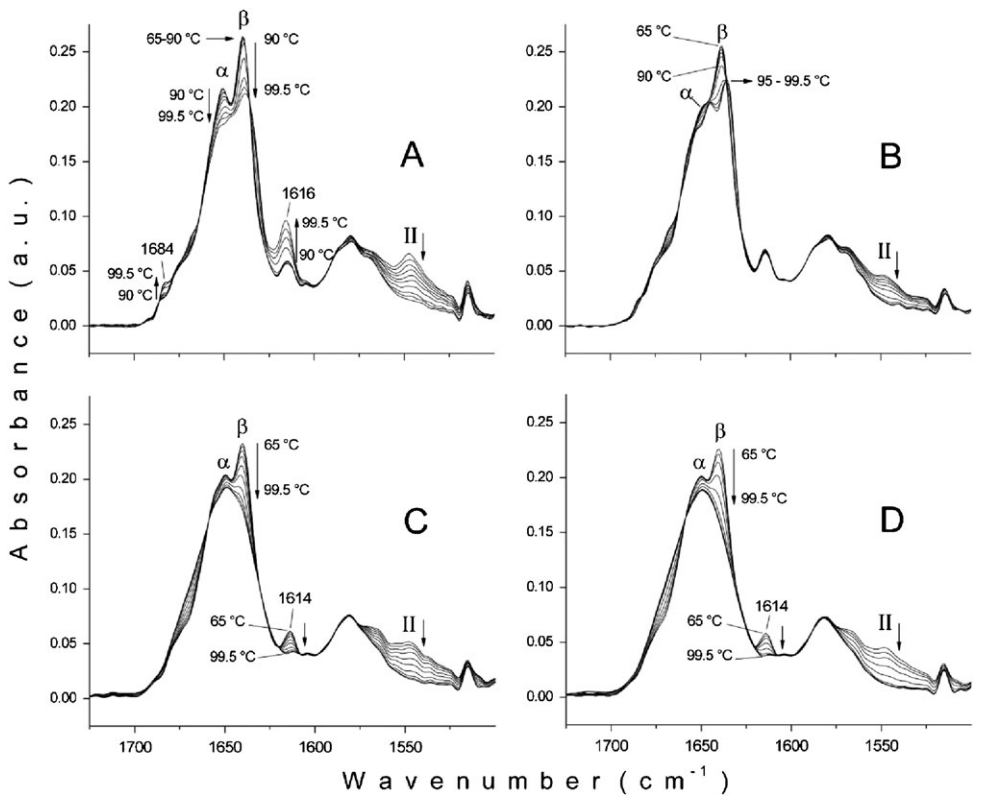


Fig. 4 Temperature-dependent changes in the deconvoluted spectra of ArgBP (A) (control), ArgBP/10.5 (B), ArgBP/1.5SDS (C), and ArgBP/3.5SDS (D). Overlayed spectra are shown in the 65–99.5 °C temperature range in 5 °C steps from 65 °C to 95 °C. Additional spectra recorded at 98, 99 and 99.5 °C are also shown. ArgBP (panel A): no significant spectral (amide I′) changes were detected in the 20–90 °C temperature range, with changes only evident at temperatures greater than 90 °C. ArgBP/10.5 (panel B), ArgBP/1.5SDS (panel C), and ArgBP/3.5SDS (panel D): no significant spectral (amide I′) changes were detected in 20–65 °C temperature range, but temperatures above 65 °C demonstrate significant conformational changes. The symbols α, β, and II, stand for α-helix, β-sheet, and residual amide II band absorption, respectively.

90 °C there is a partial loss of α-helices and β-sheet structures (Fig. 2A); (3) the binding of arginine to ArgBP does not modify the secondary structure of the protein, but dramatically increases the protein stability; (4) the β-sheets structures of ArgBP are not affected by high temperature whilst the α-helices structures are destabilized at 99.5 °C; (5) asparagine does not bind to ArgBP (Fig. 2B).

In addition, we also observed that the secondary structure content of ArgBP is affected by alkaline conditions as well as by the presence of SDS. The high pD value or the presence of SDS allow a deeper contact of D_2O with the protein, probably as a consequence of changes in the secondary structure and/or as a consequence of a relaxed tertiary structure. The protein at pD 10.5 or in the presence of SDS is less stable with respect to the protein under control conditions (ArgBP/7.5). In the presence of SDS the protein undergoes complete thermal denaturation. At pD 10.5 and temperatures up to 90 °C, partial loss of β-sheet was observed; however, at temperatures higher than 95 °C, the β-sheet content remained unaltered which suggests the presence of a population of β-structures located in a protein hydrophobic environment.

Comparison of data obtained at high pD with those obtained in the presence of SDS indicates that both ionic and hydrophobic interactions play an important role in the stabilization of the protein structure being the hydrophobic interactions particularly important for the stabilization of a population of buried β-structures.

Analysis of ArgBP structural features by homology modeling

As presented in the introduction section, ArgBP belongs to the ABC transporter family of proteins whose members possess highly conserved tertiary structure even with only a moderate level of sequence similarity.[64] The architecture of this family is typically composed of a single polypeptide chain that folds into two domains connected by a hinge region. Several X-ray crystal structures of amino acid binding proteins belonging to this family has been solved during the past years, and a BLAST search in the PDB database identified some suitable templates to model the structure of ArgBP. The structure of Arg-, Lys-, His-binding protein, named ArtJ, from *G. stearothermophilus*,[23] which shares 36% sequence identity with ArgBP, was chosen as template. ArtJ serves as a suitable template due to its origin from a thermophilic organism, to its similarity in ligand specificity, and for its higher resolution and lower mean B-factor relative to other structures (the B-factor, or atomic displacement parameter, is a measure of imprecision in the protein coordinates as a result of errors or fluctuations of an atom around its average position). However, since only the close liganded form is available for this template, we took information from the structure of ligand-free Gln-binding protein from *E. coli*[28] to create the open unliganded form of ArgBP. In particular, since the open and closed forms differ in the reciprocal position of the two domains, we used Gln-binding protein as a "scaffold" to align the two domains of ArgBP created by homology model methods using ArtJ as template. Then, this "hybrid" model was used as template for the open unliganded form of ArgBP. In this way, we used the same higher quality template for both models, taking advantage of the lower quality template only to model the hinge region. The assessment of both models confirms that they are reliable and of equal quality.

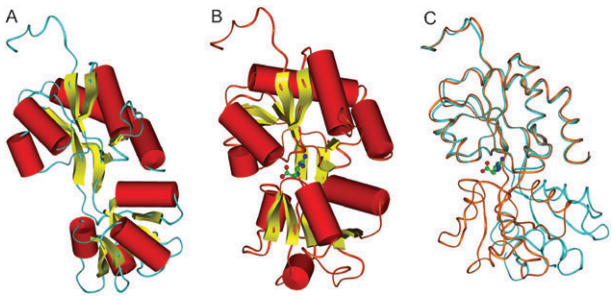


Fig. 5 3D models of ArgBP (A) and ArgBP/Arg (B), and superposition of the two forms (C). The ligand Arg in (B) is displayed in ball & stick mode (color code: red oxygen, blue nitrogen, green carbon). In (A) and (B), secondary structures are shown: helices as red cylinders, strands as yellow arrows. In (C), proteins are color coded: cyan (open form) and orange (close form).

The models of ArgBP and ArgBP/Arg are shown in Fig. 5A and B, respectively. The first domain, containing both the N-terminal and the C-terminal moieties of the protein, includes residues 18–109 and 210–246. The second domain is formed by residues 117–204, and residues 110–116 and 205–209 create the hinge between the two domains. The superposition of the modeled structures of ArgBP and ArgBP/Arg shows a root mean square deviation (RMSD) (the measure of the average distance between the backbones of superimposed proteins) of 0.66 Å, with a perfect coincidence of one of the two domains, whereas the other one is rotated and clamped (Fig. 5C). We identified the positions of the secondary structure elements in the sequence with DSSP. Each domain is formed by a central β-sheet surrounded by helices; the overall content in secondary structure elements is not different in the open and close form of the protein.

We calculated the solvent accessible surface for each residue with NACCESS. We found that three out of the five strands composing the β-sheet in the non-N-, non-C-terminal domain are formed essentially by hydrophobic residues fully shielded from solvent (Table 1).

Another important analysis of the binding protein structure involves characterization of the specific amino acid residues responsible for binding Arg. Fig. 6 shows the ArgBP residues from both domains that contain at least one atom within a 5 Å radius centered on the ligand Arg. From visual inspection and the analysis of interactions, some residues appear to be crucial for the binding of the amino acid into the cavity of the protein. In particular, the oxygens of E42 and S35 appear to be able to form H-bonds with the guanidine moiety of Arg. S93 is able to form an H-bond with Nε, whereas D183 and G94 are able to interact with the Nα atom, *via* the backbone oxygen and nitrogen, respectively. Finally, two Thr residues (T145 and T146) contact the carboxylic moiety of Arg with the oxygen atoms of their side chain, and R101 forms an ionic interaction with the same moiety. Other residues (F38, F76, the methyl group of T96) create a hydrophobic cavity to host the long hydrophobic side chain of the amino acid.

Table 1 Solvent accessibility analysis of β-sheets residues in ArgBP and ArgBP/Arg. The percentage of solvent accessibility for each β-sheet residue is displayed relative to the solvent accessibility of the side chain of the residue. Bold values correspond to residues completely or nearly completely shielded from solvent (< 10% accessibility)

ArgBP			ArgBP/Arg		
Domain	β-sheet	% solvent accessibility	Domain	β-sheet	% solvent accessibility
N + C-terminal	L30	25.2	N + C-terminal	L30	18.6
	L31	24.8		L31	27.6
	V32	**3.9**		**V32**	**1.3**
	G33	**0**		**G33**	**0**
	L34	**1.8**		**L34**	**0**
N + C-terminal	**D37**	**0**	N + C-terminal	D37	31.3
	F38	**7.8**		F38	11.1
N + C-terminal	E42	71.9	N + C-terminal	**E42**	**0.5**
	F43	43.8		F43	16.9
N + C-terminal	L69	12.1	N + C-terminal	L69	14.2
	K70	60.3		K70	56.5
	I71	19.8		I71	12.2
	V72	49.2		V72	43.6
	D73	31.8		D73	47.9
N + C-terminal	A106	15.5	N + C-terminal	A106	18.0
	F107	**4.8**		**F107**	**1.9**
	S108	**1.0**		**S108**	**0.9**
	D109	44.7		D109	65.6
	P110	26.5		P110	46.2
	Y111	**2.1**		**Y111**	0.3
	F112	13.7		F112	21.3
	D113	72.0		D113	85.8
	A114	26.0		**A114**	**7.0**
non-N-, non-C terminal	**Q116**	**2.4**	non-N-, non-C terminal	**Q116**	**0**
	V117	**4.8**		V117	17.0
	I118	**0.4**		**I118**	**0.1**
	V119	**1.9**		**V119**	**2.3**
	V120	**2.5**		**V120**	**2.3**
non-N-, non-C terminal	**V139**	**7.7**	non-N-, non-C terminal	**V139**	**4.1**
	A140	**0**		**A140**	**0**
	V141	**0**		**V141**	**0**
	Q142	38.1		**Q142**	**7.9**
non-N-, non-C terminal	**V159**	**3.1**	non-N-, non-C terminal	**V159**	**1.5**
	V160	41.6		V160	44.5
	R161	34.2		R161	38.4
	F162	18.7		F162	30.7
non-N-, non-C terminal	**A179**	**2.4**	non-N-, non-C terminal	**A179**	**2.0**
	V180	**0**		**V180**	**0**
	V181	**1.4**		**V181**	**0.4**
	L182	**6.6**		**L182**	**2.2**
	D183	62.1		**D183**	**5.2**
non-N-, non-C terminal	L197	23.0	non-N-, non-C terminal	L197	38.7
	V198	20.0		V198	14.0
	I199	12.0		**I199**	**8.3**
N + C-terminal	E207	68.1	N + C-terminal	E207	13.2
	Q208	45.7		Q208	41.5
	Y209	12.3		**Y209**	**0**
	G210	**0**		**G210**	**0**
	I211	13.6		**I211**	**3.0**
	A212	**0.7**		**A212**	**0**
	V213	**1.7**		**V213**	**0.2**

We also analyzed some structural features related to the thermostability of ArgBP. Hydrophobic clusters formed by Ile + Leu residues are considered important to enhance protein thermostability.[65] In our models, there are 10 Ile and 12 Leu residues forming a large cluster that spans over the whole N + C-terminal domain of ArgBP. In the other domain, there is another smaller cluster of hydrophobic residues (2 Ile + 4 Leu) and a couple of isolated Leu. Hydrophobic clusters formed by Ile + Leu residues are also present in the structure of Gln-binding protein from *E. coli*. However the analysis of the cavities in the two proteins revealed that the model of ArgBP is much more compact relative to the structure of the homologous mesophilic Gln-binding protein (the volume of buried voids in ArgBP is only 3 $Å^3$ in the open form and 54 $Å^3$ in the close form, compared to 94 $Å^3$ and 137 $Å^3$, respectively, in the open and close conformation of Gln-binding protein from *E. coli*). The compactness of ArgBP greatly enhances the strength of these hydrophobic interactions and, as a consequence, the thermostability of the protein is increased.

We also explored the ionic interactions that contribute to ArgBP structure and stability. Of the 70 protic residues in ArgBP, only three (E42, D54 and D196) are fully shielded

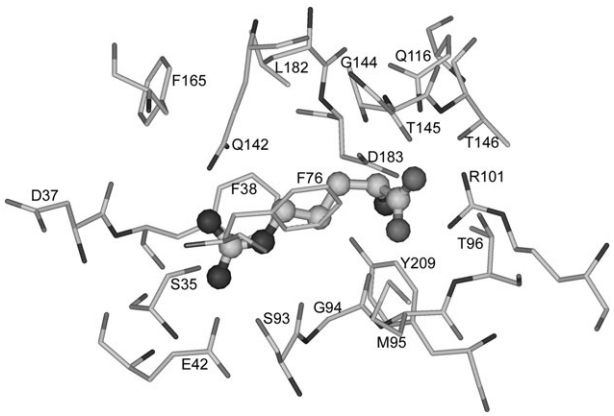


Fig. 6 Binding cleft of ArgBP with bound Arg. Residues with at least one atom within a distance of 5 Å from the center of ligand are shown in stick mode and labeled. The Arg is shown in ball & stick mode.

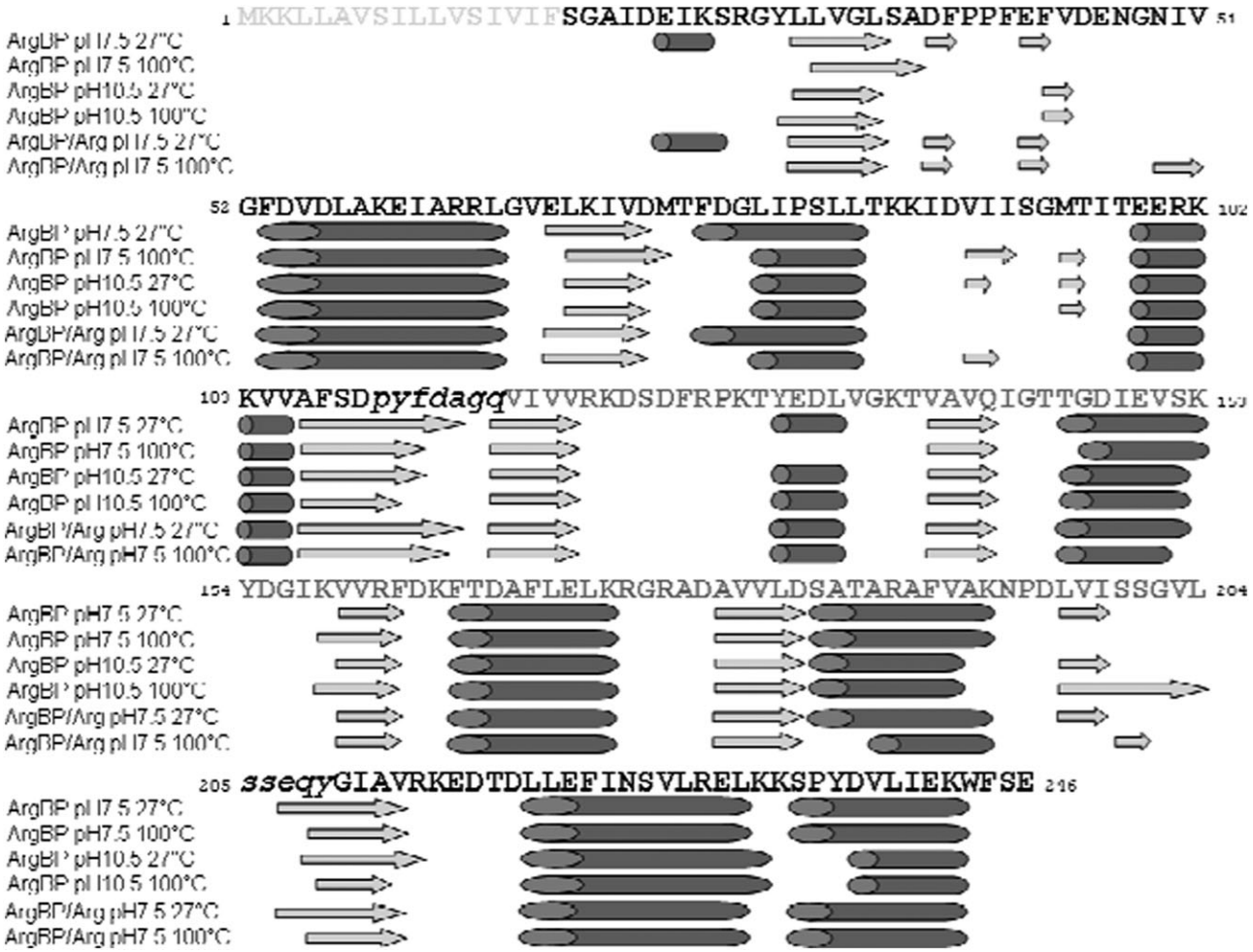
from solvent. The pK_a analysis shows that two acidic residues (E42 and D133) have very high and anomalous values (11.9 and 8.6, respectively), and E42 is largely protonated even at a pH of 10.5. As shown in Fig. 6, this acidic residue is involved in Arg binding, and probably its high pK_a value is likely to be necessary to allow the formation of stable H-bonds with the ligand. Another acidic residue, E207, which is present in one of the two β-sheets forming the hinge region, shows a high pK_a value (9.9) only in the closed form of the protein. This suggests that the conformational changes associated with binding the ligand lead to extensive shielding from the solvent (its relative accessibility changes from 68.1 to 13.2% when Arg is bound). The other protic residues are all at the surface of the protein. Although the models could suffer from potential errors in the placement of these side chains that impair the identification of ionic pairs, these residues are probably able to form a network of transient surface ionic interactions that are highly sensitive to the pH of the solution. The relative importance of these interactions to stabilize the protein structure has been pointed out in some previous works.[66,67]

Based on the structural features identified by homology modeling, the "molecular portrait" of ArgBP is that of a protein in which a tightly packed and structured hydrophobic core in each domain is surrounded by protic residues potentially interacting with each other and with the medium. The non-N-, non-C-terminal domain shows a central β-sheet formed by hydrophobic residues almost completely shielded from solvent; in the N + C-terminal domain Ile and Leu residues are inserted in the strands and around them, creating a large cluster of bulky hydrophobic residues. The ligand is inserted in the cleft between the two domains and is bound to ArgBP *via* interactions with residues belonging to both domains.

Molecular dynamics simulations

The variation of ArgBP behavior in different pH and temperature conditions was simulated using MD. Although the timescale applied (10 ns total) is not sufficient to follow the total unfolding of secondary structures in the protein, these experiments can identify the first structural variations that take place in the protein, suggesting which of these structures are more sensitive to perturbations. The variation of secondary structures was monitored during the entire simulation (data not shown) with the aid of DSSP program,[44] and the same program was also used on the representative average structures obtained for each simulation. Table 2 shows the results of this analysis of structural composition variation. Analysis of the simulated structures reveals that helices appear generally more unstable than β-structures in all cases. However, a more gradual decrease in helical content is observed in ArgBP at pH 7.5 until 100 °C, whereas at pH 10.5, helix lability is already evident in simulations at room temperature. The loss of residues in helices is generally accompanied by an increase in the number of residues in less organized structure, such as H-bonded turns or bends. This is consistent with FT-IR analysis showing that at neutral pH helices start to denature at temperature higher than 90° (Fig. 2, panel A), whereas at basic pH the content in α-helices is already reduced at 20 °C (Fig. 3, spectrum A) and the temperature does not affect significantly these secondary structures (Fig. 4, panel B). In contrast with experimental data, the total percentage of residues involved in β-structures is more or less the same in all conditions, indicating that these secondary structures seem to be more resistant to pH and temperature perturbations. This may be due to the different timescale needed by the population of β-structures to start denaturation: the simulation is too short to show significant variations in the amount of these secondary structures. However, looking at each single segment of secondary structure, we were able to hypothesize which elements are more sensitive to the combined effect of high temperature and pH. Results are shown in Scheme 1. Helices involving residues 23–25, 76–84, 146–153 and 234–243 are those with a major variation in length among the average structures, indicating that these elements unfold first. Among β-structures, it is interesting to note that the three highly hydrophobic and completely buried β-strands forming the central β-sheet of the non-N-, non-C-terminal domain of ArgBP (residues 116–120, 139–142, 179–183) are practically unaltered in every condition of pH and temperature. However, those inserted in the β-sheet of the N + C-terminal domain, particularly those formed by more polar or protic residues (residues 42–43, 106–114, 207–213) show a considerable variation in length during the simulations in different conditions. These observations suggest that the population of buried hydrophobic β-structures resistant to ionic destabilization is the one that forms the central β-sheet of the non-N-, non-C-terminal domain of ArgBP.

In the case of ArgBP/Arg, it is further demonstrated that the presence of ligand is able to substantially increase the stability of the protein. The overall structure and the individual secondary structure elements are conserved at all temperatures, with only a small variation in the total amount of residues inserted in α-helices, in short β-strands and in the two strands forming the hinge region between the two domains. Additionally, the loss of helical residues is generally accompanied by an increase of residues in less organized structures (Table 2 and Scheme 1). The central β-sheets in both domains appear to be formed by stable segments of secondary structures that seem not affected by temperature variations, in perfect agreement with experimental results.



Scheme 1 Schematic representation of the secondary structures variation (helices as cylinders and strands as arrows) during the simulations on ArgBP and ArgBP/Arg at high temperature and pH, shown along the sequence of the protein. Residues belonging to the N + C-terminal domain are in black, and residues belonging to the non-N, non-C-terminal domain are in dark grey. Residues forming the hinge region are in lower case italics. The first 17 residues in light grey were not modeled. Numbers refer to the numbering of the sequence deposited in the UniProt archive. The positions of secondary structure elements are calculated based on the average minimized structures for each simulation using the program DSSP.

From these simulation data it is possible to conclude that ArgBP is extremely resistant to temperature and pH stress. At high temperatures, the first elements of secondary structures that seem to be destabilized are helices, whereas β-strands seem more resistant (Table 2). In particular, it is possible to identify a group of β-structures formed by residues 116–120,

Table 2 Variation in ArgBP and ArgBP/Arg secondary structure composition with pH and temperature Molecular dynamics analyses were performed on the average minimized structures obtained using DSSP as described in the Materials and Methods section, with the percent of each secondary structural element displayed for each temperature/pH/binding combination

Model	% Secondary structure at 27 °C	% Secondary structure at 60 °C	% Secondary structure at 100 °C
ArgBP pH 7.5	Helices[a]: 36.6	Helices[a]: 35.3	Helices[a]: 31.7
	β-structures[b]: 25.9	β-structures[b]: 25.9	β-structures[b]: 23.7
	Other[c]: 21.0	Other[c]: 23.6	Other[c]: 20.5
	Random coil[d]: 16.5	Random coil[d]: 15.2	Random coil[d]: 24.1
ArgBP pH 10.5	Helices[a]: 31.3	Helices[a]: 31.3	Helices[a]: 30.8
	β-structures[b]: 25.4	β-structures[b]: 26.3	β-structures[b]: 25.5
	Other[c]: 23.7	Other[c]: 25.5	Other[c]: 26.3
	Random coil[d]: 19.6	Random coil[d]: 16.9	Random coil[d]: 17.4
ArgBP/Arg pH 7.5	Helices[a]: 36.0	Helices[a]: 33.3	Helices[a]: 33.8
	β-structures[b]: 25.9	β-structures[b]: 26.7	β-structures[b]: 25.9
	Other[c]: 24.6	Other[c]: 23.3	Other[c]: 25.5
	Random coil[d]: 13.6	Random coil[d]: 16.7	Random coil[d]: 14.9

[a] Marked as "H", "G" and "I" in DSSP output. [b] Marked as "E" and "B"in DSSP output. [c] Marked as "S" and "T" in DSSP output. [d] Not classified in DSSP output.

139–142, 179–183, in the interior of the non-N-, non-C-terminal domain of ArgBP, hydrophobic and deeply buried in the protein structure, that are not disrupted either by high temperature and high pH. These results are in agreement with experimental data and depict a portrait of an ideal candidate as a biological component for a biosensor to detect arginine concentration in bodily fluids.

Conclusion

We investigated the structural and molecular dynamic properties of *T. maritima* ArgBP that are important to the stability and the contribution of specific secondary structural elements to the unfolding characteristics of the protein. We developed a homology model and used molecular dynamics simulations to further elucidate the effects of temperature and pH on protein unfolding. This information will be used in the future designs of an Arg biosensor based on this thermophilic protein. The homology model provides a more complete determination of the specific binding pocket interactions that exist between *T. maritima* ArgBP and the arginine ligand. Our detailed model will allow us to make rational predictions of fluorophore attachment sites to not only study arginine binding to this protein, but also to develop a potential thermostable biosensor scaffold. In addition to providing fundamental knowledge concerning the thermal and chemical stability of proteins from extremophiles, this work provides a deeper understanding of structural elements and conformational changes associated with members of the PBP superfamily that are common to a broad spectrum of living organisms.

Abbreviations

Arg	arginine
Asn	asparagine
ArgBP	Arginine-binding protein at pD 7.5
ArgBP/Arg	ArgBP in the presence of arginine at pD 7.5
ArgBP/Asn	ArgBP in the presence of asparagine at pD 7.5
SDS	sodium dodecyl sulfate
ArgBP/10.5	Arginine-binding protein at pD 10.5
ArgBP/1.5SDS	ArgBP in the presence of 1.5% SDS
ArgBP/3.5SDS	ArgBP in the presence of 3.5% SDS
FT-IR	Fourier transform infrared
Amide I′	Amide I band in a D_2O medium

Acknowledgements

This work was in the frame of the CNR Commessa "AGP05 (SD;MS). We thank ENEA (Ente Per le Nuove tecnologie, Energia e Ambiente) and its HPC team (in particular, Dr S. Raia and Dr S. Podda) for the utilization of ENEA-GRID resources and of the HPC facility CRESCO (http://www.cresco.enea.it), located in Portici (Naples), Italy, for the MD simulations (conv. ISA-ENEA 2008). This work was supported by grants from Università Politecnica delle Marche (FT, AS) and was partially supported by CNR-Bioinformatics Project (AM).

References

1 R. M. de Lorimier, J. J. Smith, M. A. Dwyer, L. L. Looger, K. M. Sali, C. D. Paavola, S. S. Rizk, S. Sadigov, D. W. Conrad, L. Loew and H. W. Hellinga, *Protein Sci.*, 2002, **11**, 2655–75.
2 J. S. Marvin, E. E. Corcoran, N. A. Hattangadi, J. V. Zhang, S. A. Gere and H. W. Hellinga, *Proc. Natl. Acad. Sci. U. S. A.*, 1997, **94**, 4366–71.
3 L. Tolosa, I. Gryczynski, L. R. Eichhorn, J. D. Dattelbaum, F. N. Castellano, G. Rao and J. R. Lakowicz, *Anal. Biochem.*, 1999, **267**, 114–20.
4 J. D. Dattelbaum and J. R. Lakowicz, *Anal. Biochem.*, 2001, **291**, 89–95.
5 M. Brune, J. L. Hunter, J. E. Corrie and M. R. Webb, *Biochemistry*, 1994, **33**, 8262–71.
6 L. L. Salins, E. S. Goldsmith, C. M. Ensor and S. Daunert, *Anal. Bioanal. Chem.*, 2002, **372**, 174–80.
7 R. Tam, Jr. and M. H. Saier, *Microbiol. Rev.*, 1993, **57**, 320–46.
8 A. L. Davidson, E. Dassa, C. Orelle and J. Chen, *Microbiol. Mol. Biol. Rev.*, 2008, **72**, 317–64.
9 M. A. Dwyer and H. W. Hellinga, *Curr. Opin. Struct. Biol. Rev.*, 2004, **14**, 494–564.
10 B. S. Der and J. D. Dattelbaum, *Anal. Biochem.*, 2008, **375**, 132–40.
11 M. Allert, S. S. Rizk, L. L. Looger and H. W. Hellinga, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 7907–12.
12 L. L. Looger, M. A. Dwyer, J. J. Smith and H. W. Hellinga, *Nature*, 2003, **423**, 185–90.
13 N. C. Vercillo, K. J. Herald, J. M. Fox, B. S. Der and J. D. Dattelbaum, *Protein Sci.*, 2007, **16**, 362–68.
14 K. E. Nelson, R. A. Clayton, S. R. Gill, M. L. Gwinn, R. J. Dodson, D. H. Half, E. K. Hickey, J. D. Peterson, W. C. Nelson, K. A. Ketchum, L. McDonald, T. R. Utterback, J. A. Malek, K. D. Linher, M. M. Garrett, A. M. Stewart, M. D. Cotton, M. S. Pratt, C. A. Phillips, D. Richardson, J. Heidelberg, G. G. Sutton, R. D. Fleischmann, J. A. Eisen, O. White, S. L. Salzberg, H. O. Smith, J. C. Venter and C. M. Fraser, *Nature*, 1999, **399**, 323–329.
15 C. F. Higgins, *Annu. Rev. Cell Biol.*, 1992, **8**, 67–113.
16 M. S. Luchansky, B. S. Der, S. D'Auria, G. Pocsfalvi, L. Iozzino, D. Marasco and J. D. Dattelbaum, *Mol. BioSyst.*, 2010, **6**, 132–141.
17 J. P. Cooke, *Arterioscler. Thromb. Vasc. Biol.*, 2000, **20**, 2032–37.
18 D. D. Rees, R. M. Palmer and S. Moncada, *Proc. Natl. Acad. Sci. U. S. A.*, 1989, **86**, 3375–78.
19 P. Salomaa, L. L. Schaleger and F. A. Long, *J. Am. Chem. Soc.*, 1964, **86**, 1–7.
20 L. J. Bellamy, *The Infrared Spectra of Complex Molecules*, Chapman and Hall, London, 1975.
21 F. Tanfani, T. Galeazzi, G. Curatola, E. Bertoli and G. Ferretti, *Biochem. J.*, 1997, **322**, 765–69.
22 A. Sali and T. L. Blundell, *J. Mol. Biol.*, 1993, **234**, 779–815.
23 A. Vahedi-Faridi, V. Eckey, F. Scheffel, C. Alings, H. Landmesser, E. Schneider and W. Saenger, *J. Mol. Biol.*, 2008, **375**, 448–459.
24 H. Berman, K. Henrick, H. Nakamura and J. L. Markley, *Nucleic Acids Res.*, 2007, **35**(database), D301.
25 S. F. Altschul, T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller and D. J. Lipman, *Nucleic Acids Res.*, 1997, **25**, 3389–3402.
26 K. Karplus, R. Karchin, J. Draper, J. Casper, Y. Mandel-Gutfreund, M. Diekhans and R. Hughey, *Proteins: Struct., Funct., Genet.*, 2003, **53**(s6), 491–96.
27 J. Shi, T. L. Blundell and K. Mizuguchi, *J. Mol. Biol.*, 2001, **310**, 243–253.
28 C. D. Hsiao, Y. J. Sun, J. Rose and B. C. Wang, *J. Mol. Biol.*, 1996, **262**, 225–242.
29 C. Notredame, D. G. Higgins and J. Heringa, *J. Mol. Biol.*, 2000, **302**, 205–217.
30 B. Rost, G. Yachdav and J. Liu, *Nucleic Acids Res.*, 2004, **32**(web server), W321–W326.
31 The UniProt Consortium, *Nucleic Acids Res.*, 2008, **36**, D190–D195.
32 R. A. Laskowski, M. W. MacArthur, D. S. Moss and J. M. Thornton, *J. Appl. Crystallogr.*, 1993, **26**, 283–291.
33 M. J. Sippl, *Proteins: Struct., Funct., Genet.*, 1993, **17**, 355–362.

34 C. Kutzner, D. van der Spoel, M. Fechner, E. Lindahl, U. W. Schmitt, B. L. de Groot and H. J. Grubmüller, *J. Comput. Chem*., 2007, **28**, 2075–84.
35 B. Hess, C. Kutzner, D. van der Spoel and E. J. Lindahl, *J. Chem. Theory Comput*., 2008, **4**, 435–47.
36 W. F. van Gunsteren, S. R. Billeter, A. A. Eising, P. H. Hunenberger, P. Kruger, A. E. Mark, W. R. P. Scott and I. G. Tironi, *Biomolecular Simulation: The GROMOS96 manual and user guide*, Vdf Hochschulverlag AG an der ETH Zurich, Zurich, Switzerland, 1996, p. 1042.
37 J. C. Gordon, J. B. Myers, T. Folta, V. Shoja, L. S. Heath and A. Onufriev, *Nucleic Acids Res*., 2005, **33**(web server), W368–71.
38 D. Bashford and M. Karplus, *Biochemistry*, 1990, **29**, 10219–10225.
39 H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren and J. Hermans, *Interaction models for water in relation to protein hydration*, in *Intermolecular Forces*, ed. B. Pullman, Reidel D. Publishing Company, Dordrecht, 1981, pp. 331–342.
40 B. J. Hess, *J. Chem. Theory Comput*., 2008, **4**, 116–122.
41 G. Bussi, D. Donadio and M. Parrinello, *J. Chem. Phys*., 2007, **126**, 014101.
42 H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. Di Nola and J. R. Haak, *J. Chem. Phys*., 1984, **81**, 3684–90.
43 U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *J. Chem. Phys*., 1995, **103**, 8577–8593.
44 W. Kabsch and C. Sander, *Biopolymers*, 1983, **22**, 2577–2637.
45 S. J. Hubbard, S. F. Campbell and J. M. Thornton, *J. Mol. Biol*., 1991, **220**, 507–30.
46 A. L. Cuff and A. C. R. Martin, *J. Mol. Biol*., 2004, **344**, 1199–1209.
47 S. Kumar and R. J. Nussinov, *J. Mol. Biol*., 1999, **293**, 1241–55.
48 D. M. Byler and H. Susi, *Biopolymers*, 1986, **25**, 469–87.
49 J. L. Arrondo, A. Muga, J. Castresana and F. M. Goñi, *Prog. Biophys. Mol. Biol*., 1993, **59**, 23–56.
50 S. Krimm and J. Bandekar, *Adv. Protein Chem*., 1986, **38**, 181–364.
51 A. Barth, *Prog. Biophys. Mol. Biol*., 2000, **74**, 141–173.
52 A. Barth and C. Zscherp, *Q. Rev. Biophys*., 2002, **35**, 369–430.
53 A. Barth, *Biochim. Biophys. Acta, Bioenerg*., 2007, **1767**, 1073–1101.
54 E. Pedone, S. Bartolucci, M. Rossi, F. M. Pierfederici, A. Scirè, T. Cacciamani and F. Tanfani, *Biochem. J*., 2003, **373**, 875–883.
55 A. Ausili, A. Scirè, E. Damiani, G. Zolese, E. Bertoli and F. Tanfani, *Biochemistry*, 2005, **44**, 15997–16006.
56 V. Scognamiglio, A. Scirè, V. Aurilia, M. Staiano, R. Crescenzo, C. Palmucci, E. Bertoli, M. Rossi, F. Tanfani and S. D'Auria, *J. Proteome Res*., 2007, **6**, 4119–4126.
57 D. Nanavati, K. M. Noll and A. H. Romano, *Microbiology*, 2002, **148**, 3531–3537.
58 M. Jackson and H. H. Mantsch, *Biochim. Biophys. Acta, Protein Struct. Mol. Enzymol*., 1991, **1078**, 231–235.
59 S. D'Auria, R. Barone, M. Rossi, R. Nucci, G. Barone, D. Fessas, E. Bertoli and F. Tanfani, *Biochem. J*., 1997, **323**, 833–40.
60 A. Ausili, B. Cobucci-Ponzano, B. Di Lauro, R. D'Avino, G. Perugino, E. Bertoli, A. Scirè, M. Rossi, F. Tanfani and M. Moracci, *Proteins: Struct., Funct., Bioinf*., 2007, **67**, 991–1001.
61 A. Sciré, A. Marabotti, V. Aurilia, M. Staiano, P. Ringhieri, L. Iozzino, R. Crescenzo, F. Tanfani and S. D'Auria, *Proteins: Struct., Funct., Bioinf*., 2008, **73**, 839–850.
62 E. Herzyk, D. C. Lee, R. C. Dunn, K. R. Bruckdorfer and D. Chapman, *Biochim. Biophys. Acta, Lipids Lipid Metab*., 1987, **922**, 145–154.
63 S. Bañuelos, J. L. Arrondo, F. M. Goñi and G. Pifat, *J. Biol. Chem*., 1995, **270**, 9192–9196.
64 P. Borst and R. O. Elferink, *Annu. Rev. Biochem*., 2002, **71**, 537–592.
65 M. J. Danson and D. W. Hough, *Trends Microbiol*., 1998, **6**, 307–314.
66 G. R. Grimsley, K. L. Shaw, L. R. Fee, R. W. Alston, B. M. Huyghues-Despointes, R. L. Thurlkill, J. M. Scholtz and C. N. Pace, *Protein Sci*., 1999, **8**, 1843–49.
67 J. M. Sanchez-Ruiz and G. I. Makhatadze, *Trends Biotechnol*., 2001, **19**, 132–135.