Available online at www.sciencedirect.com
ScienceDirect
Bioorganic & Medicinal Chemistry 16 (2008) 4984–4993
Bioorganic & Medicinal Chemistry

Molecular modeling and dynamics simulations of PNP from *Streptococcus agalactiae*

Rafael Andrade Caceres,[a] Luis Fernando Saraiva Timmers,[a] Raquel Dias,[a] Luiz Augusto Basso,[a,b] Diogenes Santiago Santos[b,*] and Walter Filgueira de Azevedo, Jr.[a,*]

[a]*Faculdade de Biociências, Pontifícia Universidade Católica do Rio Grande do Sul, Av. Ipiranga, 6681, CEP 90619-900, Porto Alegre—RS, Brazil*
[b]*Centro de Pesquisas em Biologia Molecular e Funcional, Instituto de Pesquisas Biomédicas, Pontifica Universidade Católica do Rio Grande do Sul, Porto Alegre—RS, Brazil*

Received 6 December 2007; revised 12 March 2008; accepted 14 March 2008
Available online 20 March 2008

Abstract—This work describes for the first time a structural model of purine nucleoside phosphorylase from *Streptococcus agalactiae* (*Sa*PNP). PNP catalyzes the cleavage of *N*-ribosidic bonds of the purine ribonucleosides and 2-deoxyribonucleosides in the presence of inorganic orthophosphate as a second substrate. This enzyme is a potential target for the development of antibacterial drugs. We modeled the complexes of *Sa*PNP with 15 different ligands in order to determine the structural basis for the specificity of these ligands against *Sa*PNP. The application of a novel empirical scoring function to estimate the affinity of a ligand for a protein was able to identify the ligands with high affinity for PNPs. The analysis of molecular dynamics trajectory for *Sa*PNP indicates that the functionally important motifs have a very stable structure. This new structural model together with a novel empirical scoring function opens the possibility to explorer larger library of compounds in order to identify the new inhibitors for PNPs in virtual screening projects.

1. Introduction

Streptococcus agalactiae is a member of Group B streptococcus gram positive spherical bacterium that causes some of the most common infections in humans. The polysaccharide capsule and the secreted hemolysin are of major importance for virulence, and superoxide dismutase and D-alanylated lipoteichic acid play significant roles.[1]

Purine nucleoside phosphorylase (PNP) has been proposed as a target for the development of antibacterial drugs.[2] The catalyzed reaction generates the purine base and ribose(deoxyribose)-1-phosphate.[3] PNP is specific for purine nucleosides in the beta-configuration and cleaves the glycosidic bond with the inversion of configuration to produce α-ribose-1-phosphate.[4] PNP is a ubiquitous enzyme of purine metabolism that functions in the salvage pathway, including those of Apicomplexan parasites.[5] PNP is classified as belonging to the class I of nucleoside phosphorylase (NP-I).[6]

Drugs that inhibit human PNP activity have the potential of being utilized as the modulators of the immunological system, to treat leukemia, autoimmune diseases, and rejection in organ transplantation.[7] PNP has been submitted to intensive studies focused on the identification of new inhibitors, most of them related to human PNP.

In the present work we modeled the structure of PNP from *S. agalactiae* (*Sa*PNP). Three binding sites present in the structures of human PNP and *Sa*PNP were analyzed. The analysis was carried out with different ligands in order to identify the structural basis for the specificity of different ligands against PNPs. Furthermore, the application of a novel empirical scoring function was able to evaluate the specificity of different ligands against PNP that could be used to help in the design

Keywords: Purine nucleoside phosphorylase; Molecular modeling; Bioinformatics; *Streptococcus agalactiae*.
* Corresponding authors. Tel.: +55 51 33203500; e-mail addresses: diogenes@pucrs.br; walter.junior@pucrs.br

0968-0896/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2008.03.044

of more specific inhibitors, and in the case of *Sa*PNP help in the development of a new drug against *S. agalactiae*.

2. Results and discussion

2.1. Quality of the models

There is no crystallographic structure available for *Sa*PNP; however, the similarity between *Sa*PNP and *Hs*PNP sequences makes *Hs*PNP structure a good template for the modeling of *Sa*PNP. Furthermore, there are several binary complexes between human PNP and different ligands, which make the available templates to model binary complexes of *Sa*PNP against these ligands. The atomic coordinates of crystallography structures of templates were used as the basic models for the modeling of the *Sa*PNP. The atomic coordinates of all waters were removed from the templates.

The analysis of the Ramachandran diagram ϕ–ψ plots for the templates (Human PNP–*Hs*PNP) was used to compare the overall stereochemical quality of the *Sa*PNP structures against those of templates solved by biocrystallography. The homology models present over 89.5% of the residues in the most favorable regions.

2.2. Overall description

The structural model of *Sa*PNP contains an eight-stranded mixed beta-sheet and a five-stranded mixed beta-sheet, which join to form a distorted beta-barrel. The residues making up the eight-stranded sheet are 24–28, 43–45, 64–70, 73–79, 106–117, 123–131, 178–185, and 224–232. The five-stranded sheet consists of residues 113–117, 128–131, 182–185, 205–208, and 230–232. Seven alpha-helices surround the beta-sheet structure. The alpha-helices are composed of residues 5–17, 33–37, 90–103, 160–173, 193–202, 212–221, and 249–268.[8] Figure 1 shows a schematic drawing of the *Sa*PNP structure (monomer).

Figure 2 shows a representation of the active site of *Sa*PNP. The binary complexes were modeled with six ligands (guanine, guanosine, 3-deoxyguanosine, 7-methyl-6-thio-guanosine, inosine, and acyclovir).

The high identity between *Hs*PNP and *Sa*PNP (~46%) classifies the *Sa*PNP as a member of the nucleosides phosphorylase-I group, more specifically as low-molecular-mass (low-mm) homotrimers, with $M_r \sim 80$–100 kDa, specific for the catalysis of 6-oxopurines and their nucleosides. The *Sa*PNP (EC 2.4.2.1) consists of 269 amino acids with a molecular mass of 29,091.3 Da, and a theoretical p*I* of 5.12. The *Hs*PNP consists of 288 amino acids with a molecular mass of 32,016.7 Da, and a theoretical p*I* of 6.50. The analysis of the structure of both PNPs indicates that despite the conservation of the ribose-binding site, three mutations were observed in other sites, one in the purine-binding site and two in the second phosphate-binding site. Phe200, Arg148, and Gln144 in the human PNP[9] were replaced by Tyr190, Ile140, and Thr136 in *Sa*PNP, respectively. These differences are shown in Table 1 (*Hs*PNP) and Table 2 (*Sa*PNP).

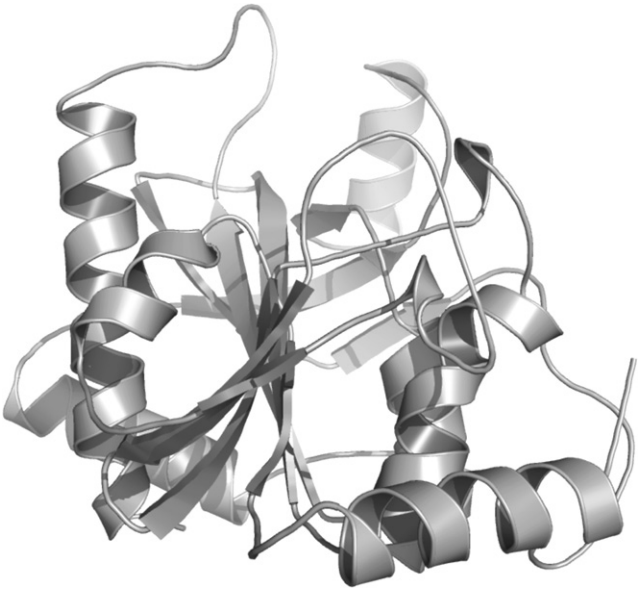

Figure 1. Tertiary structure of the *Sa*PNP. The structure contains an eight-stranded mixed beta-sheet a five-stranded mixed beta-sheet, which joins to form a distorted beta-barrel. The image was generated using Pymol.[8]

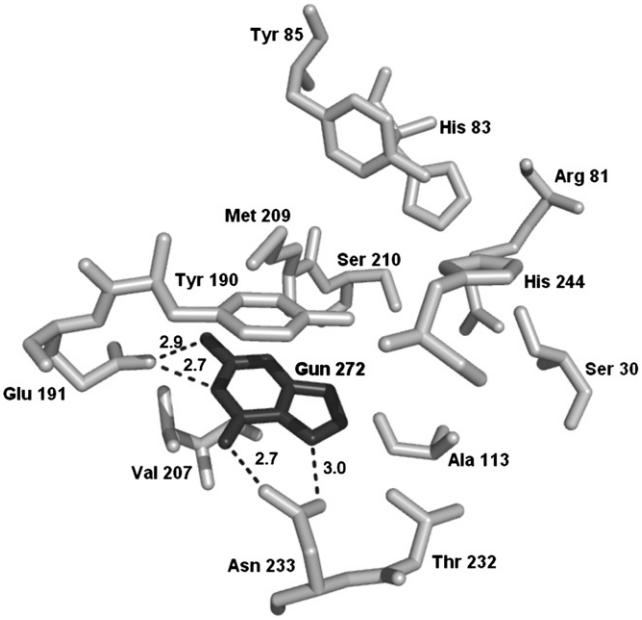

Figure 2. This image presents the active site of the *Sa*PNP with guanine. The residues are presented in light gray and the ligand (guanine) is presented in dark gray. The image was generated using Pymol.[8]

Furthermore, the superposition of the *Sa*PNP onto *Hs*PNP indicates that a small region involving residues 247–250 presents a small helical region, that is, not observed in the *Hs*PNP structure (residues 260–264).

Table 1. Intermolecular contacts of *Hs*PNP with ligands

Ligand	S33	H86	Y88	A116	A117	G118	E201	V217	G218	M219	T242	N243	H257
2-Amino-6-methylthiopurine													
5-Chloro-5-deoxy-8-aminoguanosine		1		1			1				1	1	1
5-Deoxy-5-iodo-9-deazainosine									1	1	1		
6-Hydroxy-9-*p*-aminobenzylpurine	1	1	1									1	
6-Mercaptoguanosine		1			1	1	1			1		2	
6-Mercaptopurine							1			1	1	1	
6-Mercaptoguanosine-riboside	1	1	1				1	1		2			1
8-Aminoguanosine	1				1	1					1	1	2
Acyclovir							1				1	1	
Guanine										1			
Guanosine													
Hypoxanthine			1							1			1
Inosine													
Methylthioinosine	1		2				1	1				1	
*N*7-Acycloguanosine	1			2	2	1	2					1	1

The sequence alignment indicates that this region is not conserved in both the PNPs, as shown in Figure 3. The analysis of the propensities to form helix using Chow and Fasman scale[10] and the PSIPRED (Protein Structure Prediction Server)[11] was not conclusive, it indicates only a slightly higher propensity to form turn in the *Hs*PNP sequence.

2.3. Differences in the apoenzyme and the binary complexes

In order to evaluate the possible differences in the *Sa*PNP structure due to the ligand binding, we superposed the structure of the apoenzyme, modeled using *Hs*PNP as a template (PDB access code: 1M73),[9] against the six binary complexes. The RMSD values for C-alpha superposition range from 0.7 Å to 0.9 Å. In all binary complexes the highest differences are observed in two regions, region 1 from 58 to 61 and region 2 from 247 to 250. Region 1 presents large conformation differences in all the six binary complexes, and region 2 presents the highest differences only for complex *Sa*PNP–acyclovir(*Sa*PNP–ACY). This all together indicates the flexibility of these two regions which allows different ligands to bind to the PNP active site. The highest values are observed for the PNP-3-deoxyguanosine complex 0.9 Å. This region undergoes a conformational change due to the ligand binding. The superposition of *Sa*PNP apoenzyme against the complex *Sa*PNP–ACY structure allows to see clearly both the regions (data not shown).

2.4. Empirical scoring function

A novel set empirical scoring functions have been described in the Section 4. These naïve functions use only two terms to evaluate the binding affinity, involving intermolecular hydrogen bonds and contact surface. A set with 25 polynomial functions yielded a squared correlation coefficient (r^2) up to 0.63, and a standard deviation (s) of 1.9 pK_d units.

The true value of a novel empirical scoring function lies in its predictive capability. In this study we have used 15 binary complexes involving human PNP and different ligands, Table 3, as a test set. The main reason to choose these complexes as test set are the following: (1) we are interested in testing the ability of our novel function to predict the affinity of ligands for PNP; (2) there are experimental information for the affinities of these ligands; and (3) no PNP complexes were used in the training set. Therefore, this test set tends to be a challenge and a validation method for our scoring function. The empirical scoring function presents the following values for c_0, c_1, and c_2, 5.48289, 5.93472×10^{-6}, and 4.07553×10^{-6}, respectively. Table 3 shows these results. A correlation coefficient of 0.65 and a standard deviation of 0.79 in pK_d units were obtained that are better than the ones obtained using XSCORE empirical scoring functions, shown in Table 3. Furthermore, our novel empirical scoring function was able to identify the ligand with highest affinity for human PNP, the inhibitor 5-deoxy-5-iodo-9-deazainosine. We used also this new function to predict the affinity of ligands for *Sa*PNP.

Empirical scoring functions that decompose the binding free energy into a sum of terms present an intrinsic problem in physical sense, since this decomposition is not allowed. Free energy of binding is a state function but its terms are not.[12] Furthermore, additive methods are not able to describe subtle cooperative effects.[13] We partially overcome this problem by introducing cross-terms involving hydrogen bonds and contact surface.

2.5. Interactions with ligands

The binding affinity of protein–ligand complexes for human PNP and *Sa*PNP was calculated using a novel scoring function (Table 3). Correlation coefficient between the affinity constants for *Hs*PNP and *Sa*PNP is 0.32.

The affinity of the 15 ligands against *Hs*PNP and *Sa*PNP strongly indicates that there is a small correlation between the affinities against both the PNPs. Figure 4A–F show the intermolecular contacts between enzyme and ligands for the complexes with highest affinity constants.

The ligands *N*7-acycloguanosine, 5-deoxy-5-iodo-9-deazainosine and 6-mercaptoguanosine present the highest affinities against both the PNPs. Tables 1 and

Table 2. Intermolecular contacts of *Sa*PNP with ligands

Ligand	S30	L32	A113	G122	D153	E191	G208	T232	N233	F234	A236	G237	F238	Q239	S240	E241	L242	H244	V262	L265	L269
2-Amino-6-methylthiopurine																			1	1	1
5-Chloro-5-deoxy-8-aminoguanosine					1						2	2		1				1			
5-Deoxy-5-iodo-9-deazainosine					1				1		2	1	2		1	1					
6-Hydroxy-9-*p*-aminobenzylpurine				1				3													
6-Mercaptoguanosine						2					1			2	1		1	1			
6-Mercaptopurine																				1	1
6-Mercaptoguanosine-riboside	2	1						2										1			
8-Aminoguanosine						2			2			1	1	1	1			1			
Acyclovir																			1		
Guanine																					
Guanosine					1	2					1	2	2	1		2					
Hypoxanthine																			1	1	
Inosine						2				1	1	2	1	1	1	1					
Methylthioinosine						1			2	1	1		1		1	2					
*N*7-Acycloguanosine			3				1		1												

2 show the intermolecular hydrogen bonds for all the complexes studied in this article. We can clearly see that the intermolecular hydrogen bonds involving Glu201/Glu191 and Asn243/Asn233 are conserved in all the complexes. The low correlation coefficient for affinities is probably due to the mutation (Phe200 → Tyr190) in the purine-binding site. The presence of hydroxyl group in the Tyr190 enables the formation of extra intermolecular hydrogen bonds involving the ligands and SaPNP. These intermolecular hydrogen bonds are not observed on the complexes formed by *Hs*PNP.

2.6. Molecular dynamics simulations

To obtain an estimate of the MD trajectory quality convergence, the backbone RMSD from the starting models structures was calculated (Figs. 5 and 6). After a rapid increasing during the first 250 ps, the trimeric protein backbone RMSD average and standard deviation over the last 2 ns of the system A trajectory was 2.7 ± 0.1 Å and the system B 2.5 ± 0.1 Å. A plateau of RMSDs for both the systems were achieved within 150 ps of unrestrained simulation, suggesting that 3880 ps unrestrained simulation was sufficient for stabilizing a fully relaxed models.

The superposition of the average structure of the system A with the initial model (data not shown) does not show major conformational changes from the initial model, which is consistent with the relatively low RMSD value, only the loss of the helix is composed by Gly33-Glu37 and Ser155-Asp156.

The flexibility of the proteins is assessed by the B-factor from MD the trajectory which reflects the flexibility of each atom residue in a molecule. In a typical B-factor pattern, low B-factor values indicate the well-structured regions, while the high values indicate the loosely structured loop regions or domains terminal.[14] Regions R1–R4 in Figure 7 show the four high differences of the B-factor values between the systems. The R1 region is composed by two loops (Gly29-Pro42 and Ala48-Ala62) intermediated by a beta sheet (Ile43-Tyr47), R2 is composed by a big loop formed by Ile134-Thr159, this loop interacts through the subunit interface. The Phe151 is the only residue from neighboring subunit which participates in the catalytic site of purines of an adjacent subunit. It is obvious that the major backbone fluctuations occur in the loop region and in the region surrounding the beta-alpha-beta fold, whereas regions with the low B-factor correspond exclusively to the rigid beta-alpha-beta fold. These facts indicate the stability of our models' structures. The large mobility of these regions R1–R4, and mainly loops R3 composed by Ser186-Gln200 and R4 composed by Asn233-Val248 (Fig. 7), is consistent with the need of this region to undergo conformational changes upon substrate binding. The loop R3 may have a consequential favorable role in capturing the substrate as well as in the chemical step like observed in *Hs*PNP studies[15] as can be observed in a high difference between the B-factors values in these region in ligand free and complexed form. In Figure 7 the differences behavior between system A and system B is

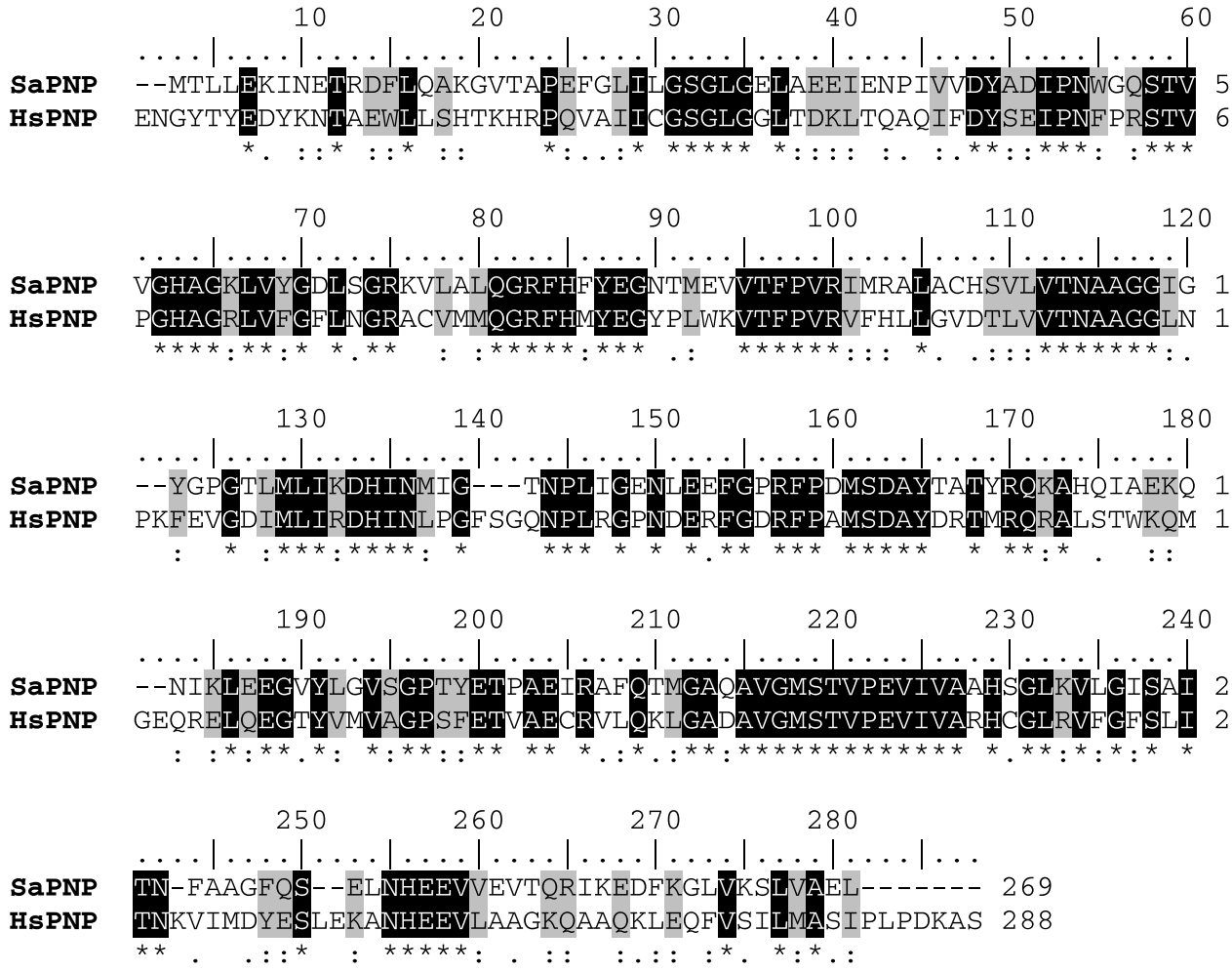

Figure 3. Sequence alignment for Human Purine Nucleoside Phosphorylase and PNP of *Streptococcus agalactiae*. The alignment was performed using ClustalW and edited with BioEdit.[25]

Table 3. 15 binary complexes involving human PNP and different ligands

Ligand	Brenda	XSCORE	PKDT22	H-bond	*A*2	*A*1	*A*1/(*A*2 − *A*1)	*A*1**A*1/(*A*2 − *A*1)
2-Amino-6-methylthiopurine	3.520	4.850	5.552	0	333	153	0.850	130.050
5-Chloro-5-deoxy-8-aminoguanosine	6.400	5.390	5.660	6	435	213	0.959	204.365
5-Deoxy-5-iodo-9-deazainosine	6.740	5.470	5.702	3	425	218	1.053	229.585
6-Hydroxy-9-*p*-aminobenzylpurine	3.690	5.820	5.632	4	422	203	0.927	188.169
6-Mercaptoguanosine	3.970	5.250	5.500	6	442	136	0.444	60.444
6-Mercaptopurine	4.130	4.560	5.496	5	289	100	0.529	52.910
6-Mercaptoguanosine-riboside	4.150	5.510	5.662	3	441	216	0.960	207.360
8-Aminoguanosine	5.150	5.620	5.651	8	424	207	0.954	197.461
Acyclovir	4.040	5.030	5.611	3	377	184	0.953	175.420
Guanine	4.540	4.620	5.569	1	293	146	0.993	145.007
Guanosine	4.850	5.690	5.581	0	440	195	0.796	155.204
Hypoxanthine	4.410	4.450	5.591	3	284	148	1.088	161.059
Inosine	4.390	5.720	5.587	0	429	194	0.826	160.153
Methylthioinosine	4.920	5.460	5.625	6	482	219	0.833	182.361
*N*7-Acycloguanosine	5.300	5.570	5.653	10	453	216	0.911	196.861

evident, demonstrating that the ACY induce a major stability for this trimeric enzyme, as the substrate binding loop and Loops R3 and R4 are involved in substrate entrance and exit.

Finally, from the superposition of the backbone of average last 2 ns structure and the initial structure of system B (data not shown), we observed that it does not change to an appreciable extent, except on the R3 and R4,

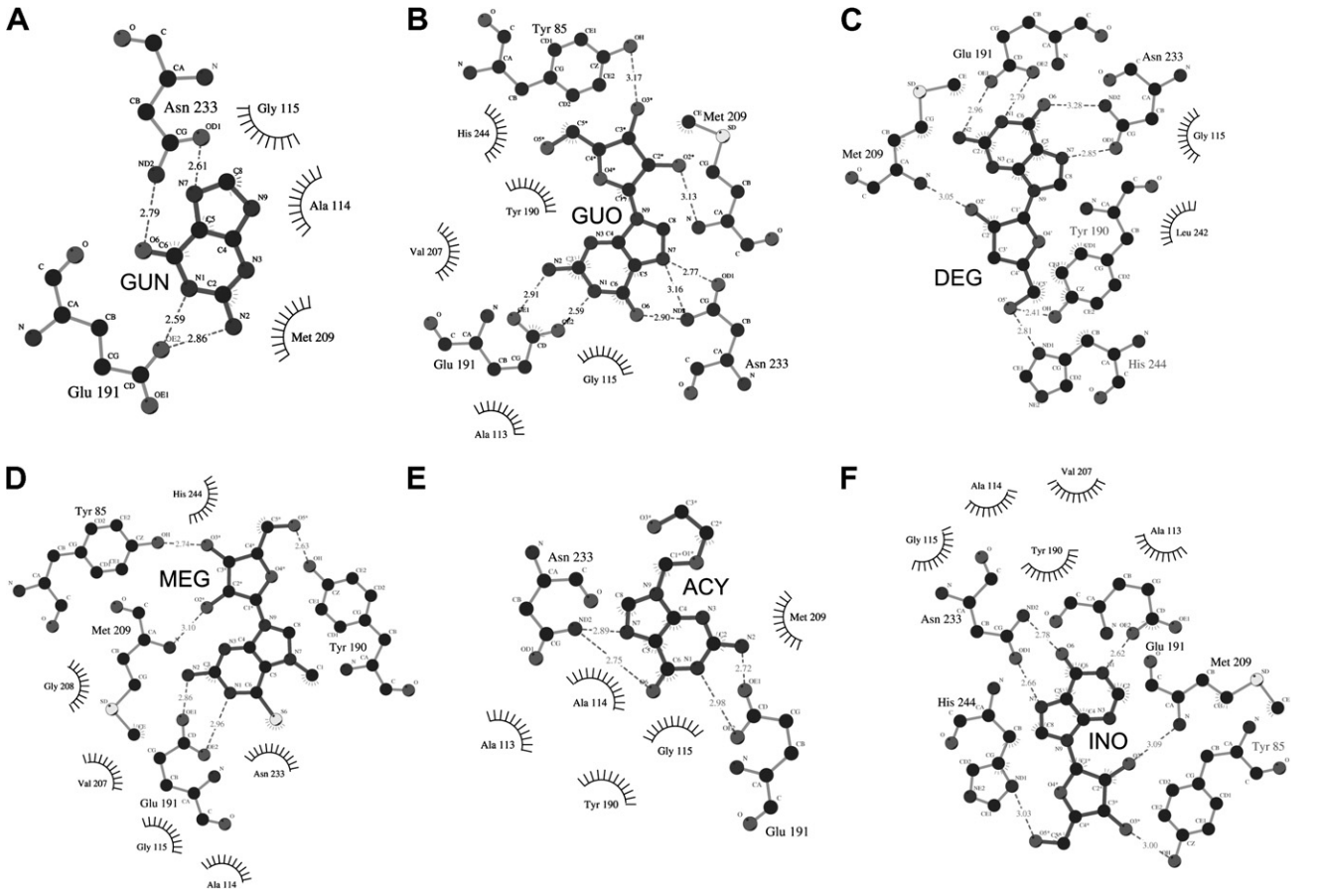

Figure 4. Intermolecular hydrogen bonds contacts of binary complexes. (A) *Sa*PNP with guanine, (B) *Sa*PNP with guanosine, (C) *Sa*PNP with 3-deoxyguanosine, (D) *Sa*PNP with 7-methyl-6-thio-guanosine, (E) *Sa*PNP with acyclovir, and (F) *Sa*PNP with inosine. The images were generated using Ligplot.[36]

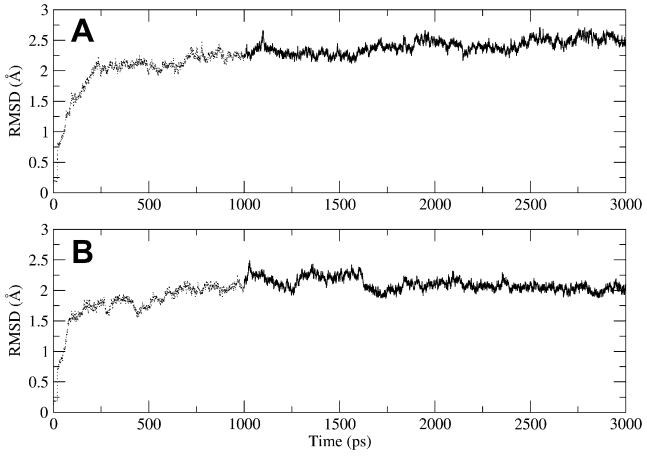

Figure 5. Graphical representation of root-mean-square deviation (RMSD) of all C-alpha from starting structure of models as a function of time. The graphic (A) shows the RMSD of apoenzyme *Sa*PNP, (B) shows the RMSD of *Sa*PNP–ACY complex. The dashed line gives the equilibration phase, the solid line shows the last 2 ns of calculation.

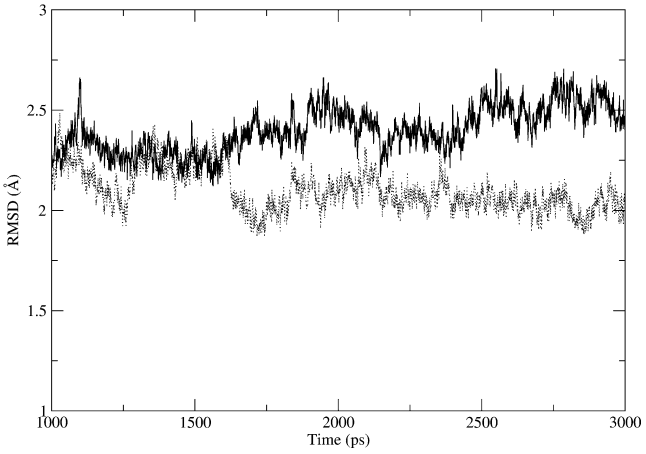

Figure 6. A close-up of RMSD of all C-alpha from the average structure of models of last 2 ns as a function of time. The solid line shows the trajectory of apoenzyme *Sa*PNP and dashed line gives the calculation of complex *Sa*PNP–ACY.

whose backbone moves backwards slightly, and the extinction of three small helices composed by Gly33-Glu37, Val59-Gly60, and Ser155-Asp156.

One of the goals of this study is to characterize the dynamic behavior of the trimeric enzyme and substrate binding mode with the enzyme, thus helping in the de-

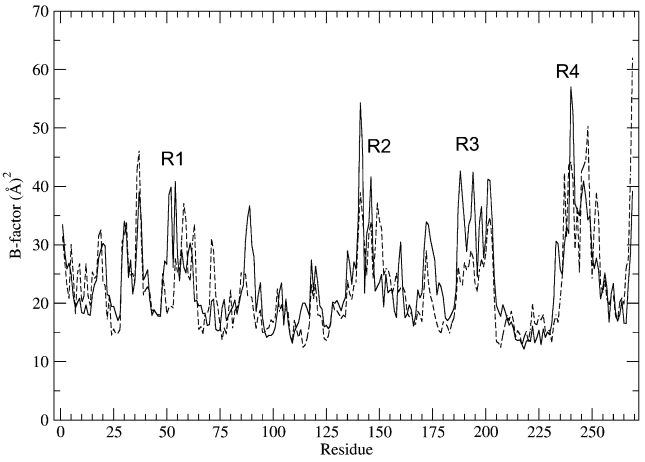

Figure 7. Backbone residue-based B-factors calculated over the last 2 ns time window for *Sa*PNP (solid line) and *Sa*PNP–ACY complex (dashed line). The regions with high fluctuation in the simulation were labeled as R1–R4.

sign of a selective inhibitor of PNP from *S. agalactiae*. The simulation of the *Sa*PNP trimer allowed us to analyze the binding pocket and inter-monomer interactions, since the binding site localized near the inter-monomer regions. Consequently a monomer simulation could create artifacts which could lead to wrong conclusions concerning the behavior of monomer in solution, more specifically the flexibility of regions in the enzyme which modulate the entrance and exit of substrate.

3. Conclusions

Structural analyses of human PNP and *Sa*PNP have shown three mutations in the binding-pocket, one in the purine-binding site and two in the second phosphate-binding site. The mutations present small influences in the affinity constants for both the PNPs. The use of a novel simple scoring function to estimate the pK_d for binary complexes showed good agreement with experimentally determined affinities for human PNP. The scoring capability of our novel scoring function identified 5-deoxy-5-iodo-9-deazainosine as the ligand with the highest affinity, confirmed by the experimental data. The same scoring function was used to estimate the binding affinity of the ligands for *Sa*PNP. Strictly speaking, neither traditional scoring function nor the cross-term function (described here) presents direct physical basis. Nevertheless, the use of the cross-term empirical scoring function was able to predict the binding affinity for human PNPs with results better than previously described empirical scoring functions, such as XSCORE.

Furthermore, computationally determined affinity constant of the inhibitor *N*7-acycloguanosine against *Sa*PNP and *Hs*PNP strongly indicates that this inhibitor presents high affinity against *Sa*PNP, suggesting a possible new lead compound against *S. agalactiae*, future kinetics experiment may confirm this prediction. In addition, we established and validated a novel empirical scoring function that can be used in virtual screening of large databases of small molecules, in order to identify new inhibitors not only for human PNP but also for bacterial PNPs.

The results obtained from MD simulations indicated that structural models of *Sa*PNP in free form and *Sa*PNP–ACY complex display a stable trimer that fully maintained over the entire simulation time. MD simulations of both the systems provide information on structural and dynamical characteristics in terms of flexibility. With reference to the molecular structure model of *Sa*PNP complexed with acyclovir, the presence of acyclovir leads to the stability of R4 and mainly R3, these loops are involved in the substrate entrance and exit.

4. Materials and methods

4.1. Molecular modeling

Homology modeling is usually the method of choice when there is a clear relationship of homology between the sequences of a target protein and at least one experimentally determined three-dimensional structure. This computational technique is based on the assumption that tertiary structures of two proteins will be similar if their sequences were related, and it is the approach most likely to give accurate results.[16]

For modeling of the *Sa*PNP complexed with acyclovir, guanine, 7-methyl-6-thio-guanosine, 3-deoxyguanosine, guanosine, and inosine the following crystallographic structures were used as templates, 1PWY,[17] 1V2H,[18] 1YRY,[19] 1V45, 1RFG,[20] and 1RCT,[21], respectively. The web server PARMODEL was used to model the binary complexes.[22] PARMODEL is a parallelized version of the MODELLER.[23] The modeling procedure begins with the alignment of the sequence to be modeled (target) with related known three-dimensional structures (templates). This alignment is usually the input to the program and the output is a three-dimensional model for the target sequence containing all main-chain and side-chain non-hydrogen atoms.[24]

The high degree of primary sequence identity between *Sa*PNP (target) and of Human PNP (*Hs*PNP) indicates that these crystallography structures are good models to be used as templates for *Sa*PNP enzyme (target). The alignment of the *Sa*PNP (target) and human PNP is shown in Figure 3.[25]

A total of 1000 models were generated for each binary complex, and the final models were selected based on the MODELLER objective function. The model with the lowest value for the MODELLER objective function was selected as the best model. All optimization process was performed on a Beowulf cluster with 16 nodes (BioComp, AMD Athlon XP 2100+, BioComp, Brazil).

4.2. Evaluation of binding affinity

The analysis of the interaction between a ligand and a protein target is still a scientific endeavor. The affinity

and specificity between a ligand and its protein target depend on directional hydrogen bonds and ionic interactions, as well as on shape complementarity of the contact surfaces of both partners.[26] Attempts to use the programs XSCORE[27] to evaluate the binding affinity of the ligands against *Hs*PNP generated unsatisfactory results, with low correlation coefficient between experimental and predicted affinities.

In order to improve the results for computationally determined binding affinities, we developed a novel empirical scoring function. In spite of many problems in the understanding of the structural features important for binding affinity, most of the experimental available data indicate that additive functions for protein–ligand interactions might be good approach for the development of empirical scoring functions. With atomic coordinates (x, y, z) available for protein–ligand complexes, the analysis of the binding can be estimated as a sum of interactions multiplied by weighting coefficients (c_j), as indicated by the following equation:

$$\mathrm{p}K_\mathrm{d} = c_0 + \sum_{j=1}^{N} c_j f_j(x, y, z)$$

$\mathrm{p}K_\mathrm{d}$ is the $-\log K_\mathrm{d}$ where K_d could also be either K_M or K_i, c_0 is a regression constant, f_js are functions that accounts for van der Waals interactions, intermolecular hydrogen bonds, deformation, hydrophobic effect, and others that may be included.

Most of the empirical functions use the size of the contact surface at the protein–ligand interface to estimate the hydrophobic interaction. A reasonable correlation between experimental affinities can be obtained with contact areas alone. We proposed here a novel set of scoring functions which are based on the observation that the major determinants to ligand specificity are intermolecular hydrogen bonds and hydrophobic contacts.[28–30]. The present set of scoring functions contains 25 polynomial functions Table 4, comprising terms up to degree 2, which may involve cross-terms (HB and A). Determination of the intermolecular hydrogen bonds and contact areas of 123 binary complexes available at PDBBIND[31] was used to propose this set of function (supplementary material). No human PNP structures were included in this dataset. This novel function is based exclusively on the hydrogen bonds (HB), contact area ($A1$), and accessible surface area for the ligand ($A2$). The additive nature of the empirical scoring functions generally leads to large ligands obtaining high scores. This effect is undesirable. Due to the overestimation of the contact area term contribution to the empirical scoring function we devise a simple scheme to reduce this contribution. We introduced a penalty term that diminishes the dependence of the score on molecular size. We divided the squared contact area ($A1 * A1$) by the term $A2 - A1$, which presents the highest values for complexes with $A1$ much smaller than $A2$. This scheme reduces the contribution of the contact area term in the empirical function for those ligands with relatively small contact area. The empirical function that presents the highest correlation coefficient between empirical and predict binding affinity is composed of a polynomial function

Table 4. Polinomial functions

pKdT	Function
1	$c_0 + c_1 \cdot x + c_2 \cdot y$
2	$c_0 + c_1 \cdot x + c_2 \cdot y^2$
3	$c_0 + c_1 \cdot x + c_2 \cdot x^2 + c_3 \cdot y$
4	$c_0 + c_1 \cdot x + c_2 \cdot x^2 + c_3 \cdot y^2$
5	$c_0 + c_1 \cdot x + c_2 \cdot x^2 + c_3 \cdot y + c_4 \cdot y^2$
6	$c_0 + c_1 \cdot x + c_2 \cdot y + c_3 \cdot y^2$
7	$c_0 + c_1 \cdot x^2 + c_2 \cdot y + c_3 \cdot y^2$
8	$c_0 + c_1 \cdot x^2 + c_2 \cdot y$
9	$c_0 + c_1 \cdot x^2 + c_2 \cdot y^2$
10	$c_0 + c_1 \cdot x + c_2 \cdot x \cdot y$
11	$c_0 + c_1 \cdot x \cdot y + c_2 \cdot y$
12	$c_0 + c_1 \cdot x + c_2 \cdot \mathrm{x}^2 + c_3 \cdot x \cdot y$
13	$c_0 + c_1 \cdot x + c_2 \cdot x \cdot y + c_3 \cdot y$
14	$c_0 + c_1 \cdot x + c_2 \cdot \mathrm{x}^2 + c_3 \cdot x \cdot y + c_4 \cdot y$
15	$c_0 + c_1 \cdot x^2 + c_2 \cdot x \cdot y + c_3 \cdot y$
16	$c_0 + c_1 \cdot x^2 + c_2 \cdot x \cdot y$
17	$c_0 + c_1 \cdot x + c_2 \cdot \mathrm{x}^2 + c_3 \cdot x \cdot y \ + c_4 \cdot y + c_5 \cdot y^2$
18	$c_0 + c_1 \cdot x + c_2 \cdot x \cdot y \ + c_3 \cdot y + c_4 \cdot y^2$
19	$c_0 + c_1 \cdot x \cdot y + c_2 \cdot y + c_3 \cdot y^2$
20	$c_0 + c_1 \cdot x^2 + c_2 \cdot x \cdot y + c_3 \cdot y^2$
21	$c_0 + c_1 \cdot x^2 + c_2 \cdot x \cdot y \ + c_3 \cdot y + c_4 \cdot y^2$
22	$c_0 + c_1 \cdot x \cdot y + c_2 \cdot y^2$
23	$c_0 + c_1 \cdot x \cdot y$
24	$c_0 + c_1 \cdot x + c_2 \cdot x \cdot y + c_3 \cdot y^2$
25	$c_0 + c_1 \cdot x + c_2 \cdot \mathrm{x}^2 + c_3 \cdot x \cdot y + c_4 \cdot y^2$

x = number of intermolecular hydrogen bonds (HB).
$y = A$.

involving the number of intermolecular hydrogen bonds (HB) and the modified contact area (A) as follows:

$$\mathrm{p}K_\mathrm{d} = c_0 + c_1 \mathrm{HB} \cdot A + c_2 A^2$$

where c_0 is the regression constant, c_1 is the weight for the term HB multiplied by contact area A, and c_2 is the weight for the squared A. The term A is determined by the following equation:

$$A = \frac{A1^2}{(A2 - A1)}$$

The terms $A1$ and $A2$ were calculated using the Lee and Richards algorithm[32] implemented in the CCP4.[33] The intermolecular hydrogen bonds were determined using a method derived from the one described for the program XSCORE.[27] Standard multivariate regression was carried out on the whole training set.

4.3. Molecular docking

In order to generate binary complexes for *Hs*PNP and *Sa*PNP in complexes with different ligands, we used the docking program ZDOCK.[34] We generated binary complexes for complexes between *Hs*PNP and the ligands for which crystallographic structures were not available. The same procedure was used to generate binary complexes for *Sa*PNP structure. In order to validate our methodology we generate binary complexes using ZDOCK for the crystallographic structures of *Hs*PNP in complex with ligands. The docking simulation was able to identify the ligand position with RMSD below 0.5 Å. We used the trimeric structure in all the docking

simulations. The trimers were generated based on the crystallographic symmetry presenting the *Hs*PNP.[9]

4.4. Analysis of the models

The overall stereochemical quality of the final models for each enzyme of the *Sa*PNP was assessed by the program PROCHECK[35] and the objective function supplied by the program MODELLER. Atomic models were superposed using the program LSQKAB from CCP4[33] and the intermolecular hydrogen bonds were assessed by the program LIGPLOT.[36]

4.5. Molecular dynamics simulations protocol

The *Sa*PNP is biologically functional as a trimer so to better understand, the structure and dynamics features were submitted a molecular dynamics simulation (MD). The trimer of *Sa*PNP was built following the crystallographic *Hs*PNP.[9]

MD were performed with the GROMACS[37] package using the Gromos 96.1 (53A6) force field.[38] The acyclovir (ACY) topology was generated with the PRODRG program.[39] Accurate force fields are essential for reproducing the conformational and dynamic behavior of the condensed-phase systems, the Gromos 96.1 force fields well parameterized for proteins but the parameters for small molecules are still limited for the simulations of more complicated biological systems, so for the atomic charges in the ACY molecule was used GAMESS[40] which were submitted to single-point ab initio calculations at RHF 6-31G* level in order to obtain Löwdin derived charges. Manipulation of structures was performed with Pymol program.[8] The first system was composed by apoenzyme *Sa*PNP (system A) and the second by *Sa*PNP enzyme and ACY ligand (system B). The simulations of two systems were performed by a time period of 4 ns. In both the systems were added Na^+ counter ions (33 Na^+ ions on the system A and 45 Na^+ ions on the system B) using *Genion* Program of the GROMACS simulation suite to neutralize the negative charge density of the systems.

Each structure was placed in the center of a truncated cubic box filled with Extended Simple Point Charge (SPC/E) water molecules,[41] containing 34,520 for the system A and 33,690 water molecules for the system B. The initial simulation cell dimensions were 79.54 Å, 83.76 Å and 51.79 Å for the system A and 79.49 Å, 80.17 Å and 48.16 Å for the system B, and had the protein solvated by a layer of water molecules of at least 10 Å in all the directions in both the systems. During the simulations, bonds lengths within the proteins were constrained by using LINCS algorithm.[42] The SETTLE algorithm was used to constrain the geometry of water molecules.[43] In the MD protocol, the binary complex, ions, and water molecules were first subjected to 1500 steps of energy minimization by steepest descent followed 1500 steps of conjugate gradient to remove close van der Waals contacts. The systems were then submitted to a short molecular dynamics with position restrains for period of 20 ps and afterwards performed a full molecular dynamics without restrains. The temperature of the system was then increased from 50 K to 300 K in five steps (50–100 K, 100–150 K, 150–200 K, 200–250 K, 250–300 K), and the velocities at each step were reassigned according to the Maxwell–Boltzmann distribution at that temperature and equilibrated for 10 ps except the last part of thermalization phase, which was for 40 ps. Energy minimization and MD were carried out under periodic boundary conditions. The simulation was computed in the NPT ensemble at 300 K with the Berendsen temperature coupling and constant pressure of 1 atm with isotropic molecule-based scaling.[44] The LINCS algorithm, with a 10^{-5} Å tolerance, was applied to fix all bonds containing a hydrogen atom, allowing the use of a time step of 2.0 fs in the integration of the equations of motion. No extra restraints were applied after the equilibration phase. The electrostatic interactions between nonligand atoms were evaluated by the particle-mesh Ewald method[45] with a charge grid spacing of ~1.0 Å and the charge grid was interpolated on a cubic grid with the direct sum tolerance set to 1.0×10^{-5}. The Lennard–Jones interactions were evaluated using a 9.0 Å atom-based cutoff.[46]

All the analysis were performed on the ensemble of system configurations extracted at 0.5-ps time intervals from the simulation, and MD trajectory collection was initiated after 2 ns of dynamics to guarantee a completely equilibrated evolution. The MD simulation and results analysis were performed on a personal compute Intel Core 2 Duo E6300—1.86 GHz and 4 GB RAM.

The convergence of the different simulations was analyzed in terms of the secondary structure, root mean-square deviation (RMSD) from the initial models structures, and root mean-square fluctuation (RMSF) to estimate the B-factor. For the B-factor calculation, the RMSFs were calculated relative to the last 2 ns averaged backbone structures, and all coordinate frames from the trajectories were first superimposed on the initial conformation to remove any effect of overall translation and rotation. Atomic isotropic B-factors were calculated from trajectories using the equation

$$\text{B-factor}_i = (8\pi^2/3)(\langle r_i^2\rangle - \langle r_i\rangle^2),$$

where $(\langle r_i^2\rangle - \langle r_i\rangle^2)$ is the mean-square positional fluctuation of atom *i*.[47,48]

Acknowledgments

This work was supported by grants from CNPq, CAPES, and Instituto do Milenio (CNPq-MCT). WFA, DSS and LAB are senior researchers of CNPq (Conselho Nacional de Pesquisas, Brazil).

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bmc.2008.03.044.

References and notes

1. Lindahl, G.; Stalhammar-Carlemalm, M.; Areschoug, T. *Clin. Microbiol. Rev.* **2005**, *18*, 102.
2. Montgomery, J. A. *Med. Res. Rev.* **1993**, *13*, 209.
3. Kalckar, H. M. *J. Biol. Chem.* **1947**, *167*, 429.
4. Canduri, F.; Fadel, V.; Basso, L. A.; Palma, M. S.; Santos, D. S.; de Azevedo, W. F., Jr. *Biochem. Biophys. Res. Commun.* **2005**, *327*, 646.
5. Bzowska, A.; Kulikowska, E.; Shugar, D. *Pharmacol. Ther.* **2000**, *88*, 349.
6. Pugmire, M. J.; Ealick, S. E. *Biochem. J.* **2002**, *361*, 1.
7. Silva, R. G.; Nunes, J. E. S.; Canduri, F.; Borges, J. C.; Gava, L. M.; Moreno, F. B.; Basso, L. A.; Santos, D. S. *Curr. Drugs Target* **2007**, *8*, 413.
8. Delano, W. L.; Lam, J. W. *Abstr. Pap. Am. Chem. Soc.* **2005**, *230*, 1371.
9. Azevedo, W. F.; Canduri, F.; dos Santos, D. M.; Silva, R. G.; Oliveira, J. S.; Carvalho, L. P. S.; Basso, L. A.; Mendes, M. A.; Palma, M. S.; Santos, D. S. *Biochem. Biophys. Res. Commun.* **2003**, *308*, 545.
10. Chou, P. Y.; Fasman, G. D. *Adv. Enzymol. Relat. Areas Mol. Biol.* **1978**, *47*, 45.
11. Jones, D. T. *J. Mol. Biol.* **1999**, *292*, 195.
12. Mark, A. E.; van Gunsteren, W. F. *J. Mol. Biol.* **1994**, *240*, 167.
13. Williams, D.; Bardsley, B. *Perspect. Drug Discovery Des.* **1999**, *17*, 43.
14. Alexander, C. S.; Yan, X.; Pei, T. *Biophys. J.* **2005**, *88*, 1009.
15. Nunez, S.; Wing, C.; Antoniou, D.; Schramm, V. L.; Schwartz, S. D. *J. Phys. Chem.* **2006**, *110*, 463.
16. Kroemer, R. T.; Doughty, S. W.; Robinson, A. J.; Richards, W. G. *Protein Eng.* **1996**, *9*, 493.
17. dos Santos, D. M.; Canduri, F.; Pereira, J. H.; Dias, M. V. B.; Silva, R. G.; Mendes, M. A.; Palma, M. S.; Basso, L. A.; de Azevedo, W. F.; Santos, D. S. *Biochem. Biophys. Res. Commun.* **2003**, *308*, 553.
18. de Azevedo, W. F.; Canduri, F.; dos Santos, M. D.; Pereira, J. H.; Dias, M. V. B.; Silva, R. G.; Mendes, M. A.; Basso, L. A.; Palma, M. S.; Santos, D. S. *Biochem. Biophys. Res. Commun.* **2003**, *312*, 767.
19. Silva, R. G.; Pereira, J. H.; Canduri, F.; de Azevedo, W. F., Jr.; Basso, L. A.; Santos, D. S. *Arch. Biochem. Biophys.* **2005**, *442*, 49.
20. Canduri, F.; Silva, R. G.; dos Santos, D. M.; Palma, M. S.; Basso, L. A.; Santos, D. S.; de Azevedo, W. F. *Acta Crystallogr., Sect. D Biol. Crystallogr.* **2005**, *61*, 856.
21. Canduri, F.; dos Santos, D. M.; Silva, R. G.; Mendes, M. A.; Basso, L. A.; Palma, M. S.; de Azevedo, W. F., Jr.; Santos, D. S. *Biochem. Biophys. Res. Commun.* **2004**, *313*, 907.
22. Uchoa, H. B.; Jorge, G. E.; Da Silveira, N. J. F.; Camera, J. C.; Canduri, F.; de Azevedo, W. F. *Biophys. Res. Commun.* **2004**, *325*, 1481.
23. Sali, A.; Blundell, T. L. *J. Mol. Biol.* **1993**, *234*, 779.
24. Canduri, F.; Uchoa, H. B.; de Azevedo, W. F., Jr. *Biochem. Biophys. Res. Commun.* **2004**, *324*, 661.
25. Hall, T. A. *Nucleic Acids Symp. Ser.* **1999**, *41*, 95.
26. De Azevedo, W. F.; Mueller-Dieckmann, J. H.; Schulze-Gahmen, U.; Worland, P. J.; Sausville, E.; Kim, S. H. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 2735.
27. Wang, R.; Liu, L.; Lai, L.; Tang, Y. *J. Mol. Mod.* **1998**, *4*, 379.
28. Boehm, H. –J.; Klebe, G. *Angew. Chem. Int. Ed. Engl.* **1996**, *35*, 2588.
29. Matsumara, M.; Becktel, W. J.; Matthews, B. W. *Nature* **1988**, *334*, 406.
30. Nauchatel, V.; Villaverde, M. C.; Sussman, F. *Protein Sci.* **1995**, *4*, 1356.
31. Wang, R.; Fang, X.; Lu, Y.; Yang, C. Y.; Wang, S. *J. Med. Chem.* **2005**, *48*, 4111.
32. Lee, R.; Richards, F. M. *J. Mol. Biol.* **1971**, *55*, 379.
33. Collaborative Computation Project, Number 4 *Acta Crystallogr., Sect. D Biol. Crystallogr.* **1994**, *50*, 760.
34. Chen, R.; Li, L.; Weng, Z. *Proteins* **2003**, *52*, 80.
35. Laskowski, R. A.; Macarthur, M. W.; Moss, D. S.; Thornton, J. M. *J. Appl. Crystallogr.* **1993**, *26*, 283.
36. Wallace, A. C.; Laskowski, R. A.; Thornton, J. M. *Protein Eng.* **1995**, *8*, 127.
37. van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. *J. Comput. Chem.* **2005**, *26*, 1701.
38. Ooestenbrik, C.; Soares, T. A.; van der Vegt, N. F. A.; van Gunsteren, W. F. *Eur. Biophys. J.* **2005**, *34*, 273.
39. van Aalten, D. M. F.; Bywater, B.; Findlay, J. B. C.; Hendlich, M.; Hooft, R. W. W.; Vriend, G. *J. Comput. Aided Mol. Des.* **1996**, *10*, 255.
40. Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347.
41. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. Interaction models for water in relation to protein hydration. In *Intermolecular Forces*; Pullman, B., Ed.; Reidel D. Publishing Company: Dordrecht, The Netherlands, 1981; p 331.
42. Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. *J. Comput. Chem.* **1997**, *18*, 1463.
43. Miyamoto, S.; Kollman, P. A. J. *Comput. Chem.* **1992**, *13*, 952.
44. Chowdhuri, S.; Tan, Ming-Liang; Ichiye, T. *J. Chem. Phys.* **2006**, *125*, 144513.
45. Darden, T.; York, D.; Pedersen, L. A. *J. Chem. Phys.* **1993**, *98*, 10089.
46. de Souza, O. N.; Ornstein, R. L. *J. Biomol. Struct. Dyn.* **1999**, *16*, 1205.
47. Hünenberger, P. H.; Mark, A. E.; van Gunsteren, W. F. *J. Mol. Biol.* **1995**, *252*, 492.
48. van Gunsteren, W. F.; Mark, A. E. *J. Chem. Phys.* **1998**, *108*, 6109.