Biophysical Chemistry 142 (2009) 7–16

Contents lists available at ScienceDirect

Biophysical Chemistry

journal homepage: http://www.elsevier.com/locate/biophyschem

Molecular modeling, dynamics and docking studies of Purine Nucleoside Phosphorylase from *Streptococcus pyogenes*

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

[a] *Faculdade de Biociências, Laboratório de Bioquímica Estrutural, Instituto Nacional de Ciência e Tecnologia em Tuberculose, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre-RS, Brazil*

[b] *Centro de Pesquisas em Biologia Molecular e Funcional, Instituto de Pesquisas Biomédicas, Instituto Nacional de Ciência e Tecnologia em Tuberculose, Pontifica Universidade Católica do Rio Grande do Sul, Porto Alegre-RS, Brazil*

ARTICLE INFO

Article history:
Received 3 December 2008
Received in revised form 17 February 2009
Accepted 17 February 2009
Available online 26 February 2009

Keywords:
Molecular dynamics
Molecular modeling
Bioinformatics
Drug design
Streptococcus pyogenes

ABSTRACT

Purine Nucleoside Phosphorylase (PNP) catalyzes the reversible phosphorolysis of *N*-glycosidic bonds of purine nucleosides and deoxynucleosides, except for adenosine, to generate ribose 1-phosphate and the purine base. PNP has been submitted to intensive structural studies. This work describes for the first time a structural model of PNP from *Streptococcus pyogenes* (SpPNP). We modeled the complexes of SpPNP with six different ligands in order to determine the structural basis for specificity of these ligands against SpPNP. Molecular dynamics (MD) simulations were performed in order to evaluate the overall stability of SpPNP model. The analysis of the MD simulation was assessed mainly by principal component analysis (PCA) to explore the trimeric structure behavior. Structural comparison, between SpPNP and human PNP, was able to identify the main features responsible for differences in ligand-binding affinities, such as mutation in the purine-binding site and in the second phosphate-binding site. The PCA analysis suggests a different behavior for each subunit in the trimer structure.

1. Introduction

Streptococcus pyogenes is a member of Group A beta-hemolytic streptococcus gram-positive spherical bacterium that produces a number of somatic constituents like M-protein and capsule, enzymes like streptolysin O, DNase B, and streptoquinase, and toxins like erythrogenic toxins A to C. The pathogen–host interactions play an important role in the development of diseases [1]. In children, *S. pyogenes* is responsible for 30% of pharyngitis cases children aged from 5 to 15 years, and also in 10% of adult cases. However, the most important complications of infection are acute rheumatic fever, acute glomerulonephritis, and peritonsillar abscess. Antimicrobial therapy alleviates pharyngeal symptoms and sometimes prevents some of the sequelae of infection.

Purine Nucleoside Phosphorylase (PNP) has been proposed as target for development of antibacterial drugs [2]. 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 reaction generates the purine base and ribose(deoxyribose)-1-phosphate [3,4]. PNP is specific for purine nucleosides in the β-configuration and cleaves the glycosidic bond with inversion of configuration to produce α-ribose-1-phosphate [5]. PNP is a ubiquitous enzyme of purine metabolism that functions in the salvage pathway, including those of Apicomplexan parasites [6]. PNP is classified as belonging to the class I of Nucleoside Phosphorylase (NP-I) [7]. Drugs that inhibit human PNP activity have the potential of being utilized as modulators of the immunological system, to treat leukemia, autoimmune diseases, and rejection in organ transplantation [8,9].

In the present work we modeled the structure of PNP from *S. pyogenes* (SpPNP). Since there is no crystallographic for PNP from *S. pyogenes* we used the crystallographic structures available for human PNPs to model SpPNP. We modeled the complexes of SpPNP with six different ligands in order to determine the structural basis for specificity of these ligands against SpPNP. Comparative analysis of the model of SpPNP and the structure of human PNP allowed identification of structural features responsible for differences in the computational determined ligand affinities. It was analyzed three binding sites present in the structures of human PNP and SpPNP. 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 Molecular Dynamics (MD) simulations was performed to evaluate the overall stability of SpPNP model. The analysis of the MD simulations was assessed mainly by principal component analysis (PCA) to better explore the trimeric structure behavior, because this kind of analysis consider only the significant conformations during the simulation. The understanding of

* Corresponding author. Av. Ipiranga, 6681, CEP 90619-900, Porto Alegre, Rio Grande do Sul, Brazil. Tel.: +55 51 33203500/4529.
E-mail addresses: luis@azevedolab.net (L.F.S.M. Timmers), rafael@azevedolab.net (R.A. Caceres), raquel@azevedolab.net (R. Dias), luiz.basso@pucrs.br (L.A. Basso), diogenes@pucrs.br (D.S. Santos), walter@azevedolab.net (W.F. de Azevedo).

0301-4622/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.bpc.2009.02.006

the rules governing the specificity of different ligands against PNP could be used to help in the design of more specific inhibitors, and in the case of SpPNP help in the development of new drugs against *S. pyogenes*.

2. Materials and methods

2.1. Molecular modeling

When the crystallographic structure of a protein target is not available the homology modeling could be a potential method to build its tertiary structure. To use the atomic coordinates of a crystallographic structure (template) the sequences of protein target and template need to share at least 30% of identity. For modeling of the SpPNP complexed with acyclovir, guanine, 7-methyl-6-thio-guanosine, 3-deoxyguanosine, guanosine and inosine the following crystallographic structures were used as templates, 1PWY [10], 1V2H [11], 1YRY [12], 1V45 [13], 1RFG [14] and 1RCT [13], respectively. The web server PARMODEL was used to model the binary complexes [15]. PARMODEL is a parallelized version of the MODELLER [16]. The modeling procedure begins with 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 [17].

The high degree of primary sequence identity between SpPNP (target) and of Human PNP(HsPNP) indicates that these crystallography structures are good models to be used as templates for SpPNP enzyme (target). The alignment of the SpPNP (target) and human PNP is shown in Fig. 1 [18].

A total of 1000 models were generated for each binary complex and the final models were selected based on stereochemical quality. All optimization process was performed on a Beowulf cluster with 16 nodes (BioComp, AMD Athlon XP 2100+, BioComp, Brazil).

2.2. Evaluation of binding affinity

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 [19–21]. We used the program XSCORE [22] to evaluate the binding affinity of the ligands against HsPNP and SpPNP. According to this method, the binding affinity of the ligand can be decomposed to the contribution of individual atoms. Each ligand atom obtains a score, called the atomic binding score, indicating its role in the binding process. The program reads the structure, assigns atom types and parameters, performs the calculation, and gives the dissociation constant of the given protein–ligand complex. The computational results are fed into a text file in which the detailed information of each ligand atom, including the atomic binding score, is tabulated. This data were used to evaluate the correlation coefficient between the affinities against both PNPs to verify possible resemblance in the structural basis for specificity against these enzymes.

2.3. Analysis of the models

The overall stereochemical quality of the final models for each enzyme of the SpPNP was assessed by the program PROCHECK [23]. Atomic models were superposed using the program LSQKAB from CCP4 [24] and the intermolecular hydrogen bonds were assessed by the program LIGPLOT [25].

2.4. Molecular docking protocol

Rigid-body docking simulations (RDS) were performed using ZDOCK 2.3 [26], which is used for the prediction of the three-dimensional (3D) structure of a protein–protein complex from the coordinates of its components structures, it is classified as bound docking or unbound docking [27]. This method is important for the development of new drugs, as summarized in previous works [28–38] because RDS is based on three basic tasks, which are: (1) characterization of the binding site; (2) positioning of the ligand into the binding site; and (3) evaluating the strength of interaction for a specific ligand–receptor complex [39].

The RDS was performed with 3-deoxyguanosine (3DG) against SpPNP. Before running this simulation, it was carried out a validation with the HsPNP crystallographic structure (PDB access code: 1PF7) [40]. In the validation, the ligand of the HsPNP crystallographic structure, was rotated 180° along *z* axis and translated 1 (fractional coordinates) along the three axis, using the program LSQKAB [24]. After this, the protein was kept fixed and the binding site was restricted. The residues Gly115,

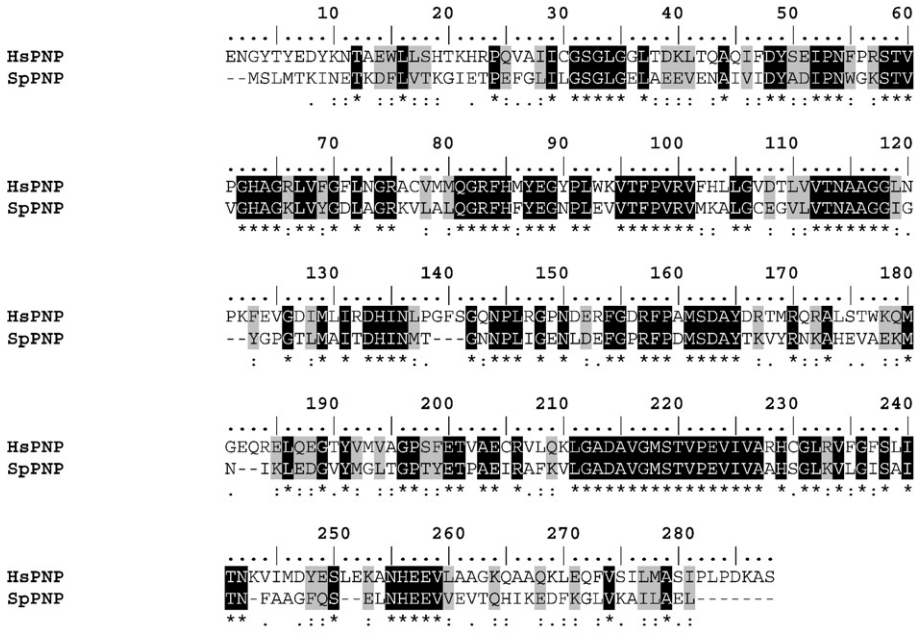

Fig. 1. Sequence alignment for Human Purine Nucleoside Phosphorylase and PNP from *Streptococcus pyogenes*. The multiple alignment was performed using ClustalW [17] and edited with BioEdit [16]. (*) indicates positions which have a single, fully conserved, (:) indicates that one of the following 'strong' groups is fully conserved, and (.) indicates that one of the following 'weaker' groups is fully conserved.

Phe151, Tyr200, Glu201, Val207, Met209 Thr232, Asn233 and His244 were selected to direct molecular docking simulations to the active site. The ligand was allowed to translate around the target. A $128\times128\times128$ point grid with a spacing of 1.2 Å was employed. We used six rotational/translational degrees of freedom. A total of 100 docked structures were generated, and root mean square deviations (RMSD) were calculated against the modeled structure (SpPNP:3DG). The docking of 3DG against SpPNP was performed following the same procedure used for the validation with the crystallographic HsPNP structure.

2.5. Statistical analysis

Spearman ρ correlation coefficients were used to assess the association between the K_d values and RMSD. This method is specific for nonlinear relationship between two continuous variables, which take on values between -1 (uncorrelated or negatively) and $+1$ (positively correlated). The Spearman ρ is the simple correlation coefficients of ranks (the relative order) based on continuous data [41]. The ranked Spearman ρ coefficients correlation is expressed the equation as follow:

$$\rho = 1 - \frac{6\sum_{j-1}^{N}\left[r\left(RMSD_j\right) - r\left(Kd_j\right)^2\right]}{N^3 - N} \quad (1)$$

where n (in our case n = 100) is the number of pairs, $r(x_i)$ and $r(y_i)$ are the rank of the activity and the interaction energy of the ith sample in the testing set [42].

2.6. Molecular Dynamics simulations

Molecular Dynamics (MD) simulations were performed with the GROMACS [43] package using the Gromos 96.1 (53A6) force field [44]. The Guanine (GUN) topology was generated with the PRODRG program [45]. Accurate force fields are essential for reproducing the conformational and dynamic behavior of condensed-phase systems, the Gromos 96.1 force fields well parameterized for proteins but the parameters for small molecules are still limited for simulations of more complicated biological systems so the atomic charges in the 3DG molecule was used GAMESS [46] which were submitted to single-point *ab initio* calculations at RHF 6-31G* level (E) in order to obtain Löwdin derived charges. Manipulation of structures was performed with Swiss-PDBViewer v3.7 program [47]. The first system was composed by apoenzyme SpPNP (system A) and the second by SpPNP enzyme, three sulfate ions and 3DG ligand (system B). The simulations of two systems were performed by a time period of 5 ns. In both systems were added Na^+ counter ions (thirty-nine Na^+ ions on the system A and fifty-seven 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 were placed in the center of a cubic box filled with Extended Simple Point Charge (SPCE) water molecules [48], containing 32526 for the system A and 32465 water molecules for the system B. The initial simulation cell dimensions were 80.13 Å × 80.83 Å × 49.15 Å for the system A and 80.03 Å × 79.38 Å × 50.75 Å for the system B, and had the protein solvated by a layer of water molecules of at least 10 Å in all directions in both systems. During the simulations, bonds lengths within the proteins were constrained by using LINCS algorithm [49]. The SETTLE algorithm was used to constrain the geometry of water molecules [50]. In the MD protocol, all hydrogen atoms, ions, and water molecules were first subjected to 1000 steps of energy minimization by steepest descent followed 500 steps of conjugate gradient to remove close van der Waals contacts. The systems were then submitted to a short molecular dynamic 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 5 steps (50 K to 100 K, 100 K to 150 K, 150 K to 200 K, 200 K to 250 K, 250 K to 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 termalization phase where we used a period of 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 [51]. 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 [52] 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 [53]. All 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 1 ns of dynamics to guarantee a completely equilibrated evolution. The MD simulation and results analysis were performed on in a personal computer with an Intel Core 2 Duo E6300 CPU of 1.86 GHz clock speed and with 4 Gb of RAM. The convergence of the different simulations was analyzed in terms of the secondary structure, RMSD from the initial models structures.

2.7. Essential dynamics analysis

Essential Dynamics (ED), also known as principal component analysis (PCA), is a method commonly used for dissecting the dynamics of proteins and their importance in biological processes, like protein folding or substrate binding. The ED analysis is a technique that reduces the complexity of the data and extracts the concerted motion in simulations that are essentially correlated and presumably meaningful for biological function [54]. In the ED analysis, a variance/covariance matrix was constructed from the trajectories after removal of the rotational and translational movements. A set of eigenvectors and eigenvalues was identified by diagonalizing the matrix. The eigenvalues represented the amplitude of the eigenvectors along the multidimensional space, and the displacements of atoms along each eigenvector showed the concerted motions of protein along each direction. An assumption of ED analysis is that the correlated motions for the function of the protein are described by eigenvectors with large eigenvalues. The movements of protein in the essential subspace were identified by projecting the Cartesian trajectory coordinates along the most important eigenvectors from the analysis.

3. Results and discussion

3.1. Quality of the models

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

The analysis of the Ramachandran diagram $\phi-\psi$ plots for the templates (HsPNP) was used to compare the overall stereochemical quality of the SpPNP structures against those of templates solved by biocrystallography. The homology models present over 89% of the residues in the most favorable regions. The RMSD values for superposition the active binding sites and proteins human PNP and SpPNP are presented in Table 1.

Table 1
The amino acid identity, and RMSD value after superposition of HsPNP (template) and *Sp*PNP (target).

Enzyme	PDB access code	Resolution (Å)	References	Sequence identity (%)	RMSD site/protein (Å)
PNP	1PWY	2.8	[10]	43.40	0.1/0.5
	1RCT	2.8	[13]	43.40	0.2/0.4
	1RFG	2.9	[14]	43.40	1.4/0.5
	1 V45	2.86	[13]	43.40	0.1/0.4
	1 V2H	2.7	[11]	43.60	0.1/0.5
	1YRY	2.8	[12]	43.60	0.1/0.5

3.2. Overall description

The structural model of SpPNP follows the same folding as the human PNP. Fig. 2 [55] shows schematic drawing of the SpPNP structure (monomer). Fig. 3 shows a representation of the active-site of SpPNP. The binary complexes were modeled with six ligands (guanine, guanosine, 3-deoxyguanosine, 7-methyl-6-thio-guanosine, inosine and acyclovir). The high identity between human HsPNP and SpPNP (~46%) classifies SpPNP as a member of the Nucleosides Phosphorylase-I group, more specifically as Low-molecular-mass (low-mm) homotrimers, with M_r ~80–100 kDa, specific for catalysis of 6-oxopurines and their nucleosides. The SpPNP (EC 2.4.2.1) consists of 269 amino acids with a molecular weight of 28881.1 Da and a theoretical pI of 4.98. The HsPNP consists of 288 amino acids with a molecular weight of 32016.7 Da and a theoretical pI of 6.50. Analysis of the structure of both PNPs indicates that, despite 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 [29] were replaced by Tyr190, Ile140 and Asn136 in SpPNP, respectively, these differences are shown in Table 2 (HsPNP) and Table 3 (SpPNP). Furthermore, superposition of the SpPNP onto HsPNP indicates that a small region involving residues 247–251 presents a small helical region not observed in the HsPNP structure (residues 261–263) Fig. 4 shows the superposition of SpPNP against HsPNP. The sequence alignment indicates that this region is not conserved in both PNPs, as shown in Fig. 1. Analysis of the propensities to form helix using Chou and Fasman scale [56] and PsiPred [57] was not conclusive, it indicates only a slightly higher propensity to form turn in the HsPNP sequence.

3.3. Differences in the apoenzyme and the binary complexes

In order to evaluate the possible differences in the SpPNP structure due to ligand binding we superposed the structure of the apoenzyme, modeled using HsPNP as template (PDB access code: 1 M73) [40], against the six binary complexes. The RMSDs for C-α superposition range from 0.8 Å to 0.9 Å. The highest values are observed for the PNP-acyclovir complex 0.9 Å. This region undergoes a conformational change due to the ligand binding. The RMSDs for the atoms from the binding site are presented in Table 1 and they range from 0.1 Å to 1.4 Å. These conformational changes could be by the losses of secondary structure and by loop movement close to the purine binding site. Fig. 5 illustrates the superposition of SpPNP apoenzyme against the complex SpPNP-acyclovir.

3.4. Interactions with ligands

The binding affinity of a protein–ligand complex for human PNP and SpPNP were calculated with the program XSCORE. The affinities of HsPNP and SpPNP are shown in Table 4. There is a low correlation coefficient between the affinity constants for HsPNP and SpPNP. The affinity of the six ligands against HsPNP and SpPNP strongly indicates that there is no correlation between the affinities against both PNPs. The ligand 3-deoxyguanosine presents the highest affinity against both PNPs. Tables 2 and 3 shows the intermolecular hydrogen bonds for all complexes studied in this article. We can clearly see that the intermolecular hydrogen bonds involving Glu201/Glu191 and Asn243/Asn233 are conserved in all complexes. Furthermore, these residues are strongly conserved among several PNPs, as e.g. *Streptococcus agalactiae* [58], *Bacteroides fragilis* [59], *Listeria monocytogenes* [60]. The main rule of these residues is to anchor the ligand into the binding pocket. The lack of correlation between the affinities calculated for all ligands against both enzymes is probably due to differences in the binding pocket, especially at positions Gln144®Asn136 and Arg148®Ile140, and Phe200®-Tyr190, where the first residues are those for human PNP, and the second are those for SpPNP.

3.5. Dynamics simulation

The SpPNP 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 SpPNP was built following the crystallographic HsPNP [61].

To understand the stability and the MD trajectory quality convergence of the trimer, the backbone RMSD from the starting models structures was calculated (Fig. 6). After a rapid increasing during the first 300 ps, the trimeric protein backbone RMSD average and standard deviation over the last 2 ns of the system A trajectory was 2.4 ± 0.4 and the system B 2.6 ± 0.3. The both systems MD simulations achieved a plateau within 1000 ps, suggesting that 5000 ps are sufficient for stabilizing fully relaxed models.

The stability of the quaternary structure during the MD simulation was assessed by the radius of gyration (Rg) (data not shown). The mean values of the Rg averaged over the period from 0 to 5 ns were determined and given (28.5 ± 0.1 Å and 28.3 ± 0.2 Å for the systems 1, 2, respectively), remaining essential constant after 2000 ps for those

Fig. 2. Tertiary structure of the SpPNP (light grey) superimposed with the HsPNP (dark grey). The structure is presented as ribbon diagram. The structure contains an eight-stranded mixed β-sheet a five-stranded mixed β-sheet, which join to form a distorted β-barrel. The image was generated using Pymol [23].

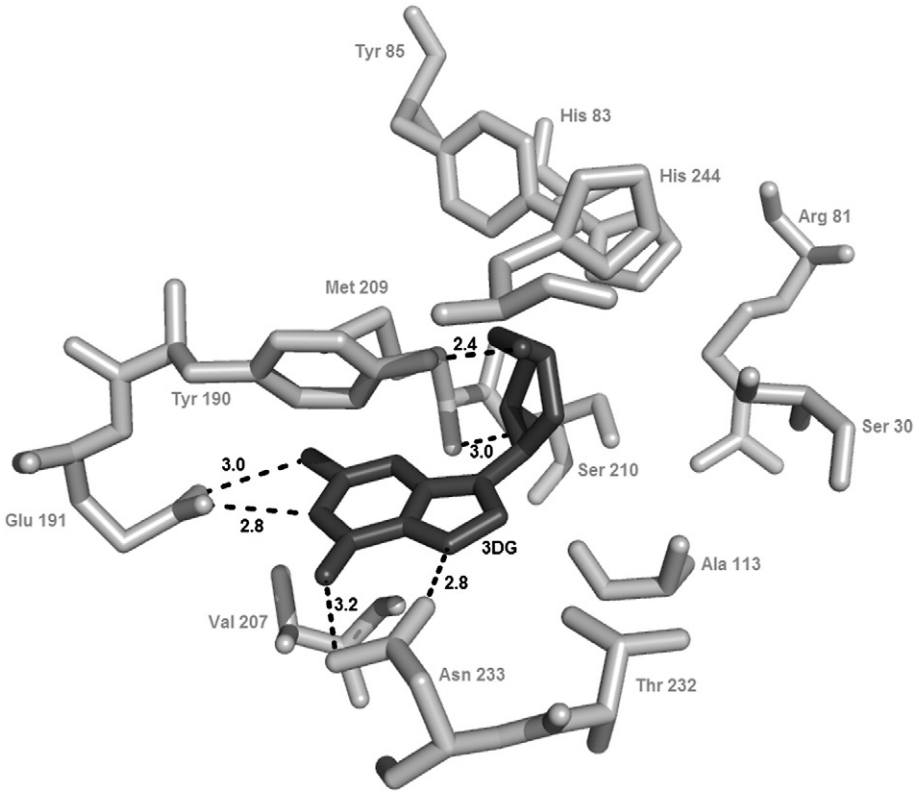

Fig. 3. The image presents the active site of the SpPNP with 3-deoxyguanosine. The residues are presented in light gray and the inhibitor (3-deoxyguanosine) is presented in dark gray. The image was generated using Pymol [23].

systems, suggesting that the molecular conformation was significantly preserved as a whole.

3.6. Principal components analysis

To support our results and investigate the most significant collective modes of motion occurring during the simulations of the uncomplexed and complexed SpPNP, the covariance matrix corresponding to the Cα-atom coordinates was calculated and PCA was performed. By diagonalizing the covariance matrix, the anharmonic and large-scale motions of the protein are isolated from the mostly harmonic and small-scale motions. Because the large-scale anharmonic motions in the essential subspace are often correlated to the vital functions of the protein, we only focus on these movements. The 3*N* eigenvalues (807) of the covariance matrix were ranked in a decreasing order of magnitude. The analysis of the apo simulation has shown that the total positional fluctuations described by the first 50 eigenvectors are 82.1%, 81% and 84.4% for subunits A, B and C respectively. The results of the first 3 eigenvectors alone represent 46% (subunit A), 41.1% (subunit B) and 46.9% (subunit C), of the total motion of the protein. This data are shown in Fig. 7A. For the SpPNP:3DG complex simulation, the analysis has shown that the first 50 eigenvectors total positional fluctuations described 88.8%, 82.4% and 92.3% for subunits A, B and C respectively. The results of the first 3 eigenvectors alone represent 58% (subunit A), 44.6% (subunit B) and 69.2% (subunit C), of the total motion for the protein. This data are shown in Fig. 7B.

The significant fluctuations are shown in Fig. 8. The main fluctuations are presented for the regions R1, R2 and R3. The residues involved in R1 are Ile45 to Ile55 for the monomer B. This region is close to the first phosphate binding site. R2 involving residues from Asn138 to Lys148 for the monomer C that is a region localized in the interface between the subunits B and C, and this region could be responsible to keep the stability of the trimeric structure. The region R3 involves residues from Glu245 to Lys255 for the monomer C. The analysis of these results confirms that in the flexible regions are involved in losses of secondary structures, mainly in substrate entrance and exit (Fig. 8). Furthermore, the PCA results confirm that the analysis with the

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

Residues/ligands	Guanine	Guanosine	3-deoxyguanosine	7-methyl-6-thio-guanosine	Acyclovir	Inosine
Glu201	OE2-N1 -> 2.46 Å	OE2-N1 -> 2.49 Å	OE2-N1 -> 2.78 Å	OE2-N1 -> 2.91 Å	OE2-N1 -> 3.03 Å	OE2-O6 -> 3.33 Å
	OE2-N2 -> 2.80 Å	OE1-N2 -> 2.63 Å	OE1-N2 -> 2.75 Å	OE1-N2 -> 2.70 Å	OE1-N2 -> 2.55 Å	
Asn243	ND2-O6 -> 2.81 Å	ND2-N7 -> 3.16 Å	ND2-O6 -> 3.34 Å	-	ND2-N7 -> 2.91 Å	ND2-O6 -> 2.87 Å
		ND2-O6 -> 2.89 Å	OD1-N7 -> 2.78 Å		ND2-O6 -> 2.65 Å	ND2-N7 -> 3.30 Å
His257	–	–	ND1-O5` -> 2.80 Å	–	–	ND1-O5* -> 2.99 Å
Tyr88	–	OH-O3* -> 3.28 Å	–	OH-O3* -> 2.76 Å	–	OH-O3* -> 3.03 Å
Ala116	–	O-O2* -> 3.19 Å	–	–	–	O-O2* -> 3.42 Å
Met219	–	N-O2* -> 3.08 Å	N-O2` -> 3.07 Å	N-O2* -> 3.05 Å	–	N-O2* -> 3.05 Å

Table 3
Intermolecular contacts of *Sp*PNP with ligands.

Residues/ligands	Guanine	Guanosine	3-deoxyguanosine	7-methyl-6-thio-guanosine	Acyclovir	Inosine
Asn233	OD1-N7 -> 2.61 Å	ND2-O6 -> 2.89 Å	OD1-N7 -> 2.84 Å		ND2-N7 -> 2.88 Å	ND2-O6 -> 2.80 Å
	ND2-O6 -> 2.77 Å	ND2-N7 -> 3.18 Å	ND2-O6 -> 3.29 Å	–	ND2-O6 -> 2.76 Å	ND2-N7 -> 3.28 Å
		OD1-N7 -> 2.77 Å				OD1-N7 -> 2.64 Å
Glu191	OE2-N1 -> 2.59 Å	OE2-N1 -> 2.58 Å	OE2-N1 -> 2.80 Å	OE2-N1 -> 2.96 Å	OE2-N1 -> 2.94 Å	OE2-N1 -> 2.62 Å
	OE2-N2 -> 2.85 Å	OE1-N2 -> 2.90 Å	OE1-N2 -> 2.98 Å	OE1-N2 -> 2.81 Å	OE1-N2 -> 2.71 Å	
Tyr190	–	–	OH-O5' -> 2.44 Å	OH-O5* -> 2.64 Å	–	–
Met209	–	N-O2* -> 3.13 Å	N-O2' -> 3.04 Å	N-O2* -> 3.10 Å	–	N-O2* -> 3.09 Å
Tyr85	–	OH-O3* -> 3.20 Å	–	OH-O3* -> 2.69 Å	–	OH-O3* -> 3.04 Å
His244	–	–	ND1-O5` -> 2.83 Å	–	–	ND1-O5* -> 2.99 Å

trimeric structure is pivotal to better understand the behavior of the SpPNP. This kind of analysis suggests that the MD simulations for a single monomer is not conclusive for PNPs, due to the fact that monomers present a different behavior in the binding pockets and interfaces regions during the simulation.

According to the Fig. 8 (PCA), we may suggest a different behavior for each subunit during the simulation. Analyzing the same figure (PCA) we can observe that the PNP presents different behavior in several regions, mainly in the binding pocket and the interface regions. Therefore, this flexibility is presented in different subunits, if the figure (PCA) is analyzed taking into account, the subunits, two by two, we can clearly observe that the a couple of monomers present a similar behavior (in this simulation the monomers A and B), when compared with the other (in our case monomer C).

3.7. Analysis of average length distances

To corroborate with the PCA results was assessed by classical dynamics the atoms average length distance for each subunit. This analysis could confirm the results obtained with the PCA. Fig. 9 shows the average length distances for each atom which composed each subunit (A, B and C). The data of this analysis suggests that the monomers A and B remain more stable than the monomer C, the same result that was obtained with PCA. We can approximate the trimer as mass-spring system, where each sphere represents a monomer of the system. These results could be interpreted as a normal mode of vibration (NMV), where the both similar monomers (A and B) are closer, suggesting that could be a pattern for this kind of protein. Molecular simulation for longer periods may capture the behavior of alternative modes of vibration.

3.8. Docking simulations and funnel-shaped energy surface

The main goal in the present analysis of the molecular docking results is to identify the lowest RMSDs between the docked structure and the crystallographic structure (validation test). It is expected that the best results will generate the lowest estimated K_d. The concept of a funnel-shaped energy surface, which was originally proposed for protein folding simulation, may be applied to analyze the protein–ligand interaction. It has been proposed that protein–ligand complexation is a similar process, because the ligand can always find the binding pocket on the protein and bind in a unique way [62]. To verify

Fig. 4. This figure illustrates the superposition of complex HsPNP-acyclovir (light gray) against the complex SpPNP-acyclovir (dark gray). The image was generated using Pymol [23].

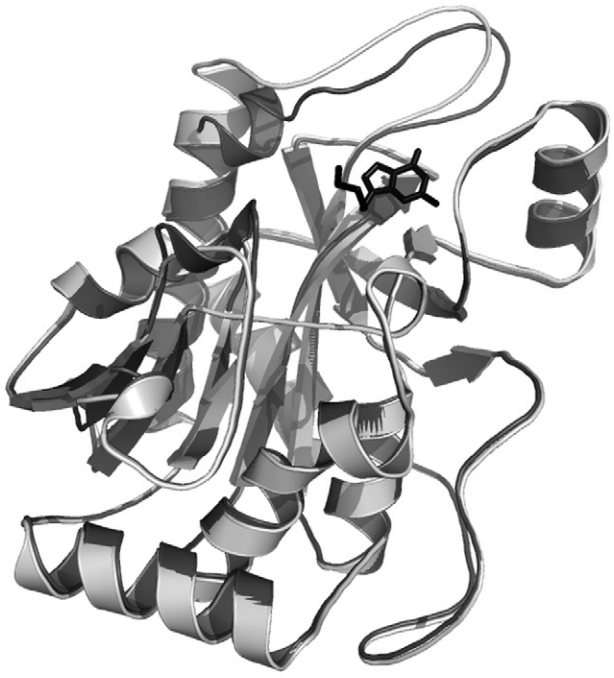

Fig. 5. This figure illustrates the superposition of SpPNP apoenzyme against the complex SpPNP-acyclovir. The image was generated using Pymol [23].

Table 4
pK_d's for HsPNP and SpPNP. The results were generated using XSCORE [22].

PNPs/ligands	HsPNP	SpPNP
Acyclovir	4.04	5.07
Guanine	5.22	4.92
Guanosine	4.92	5.72
3-deoxyguanosine	5.76	5.91
7-methyl-6-thio-guanosine	3.44	5.87
Inosine	4.39	5.72

the correlation between RMSD and estimated K_d's we used Spearman's rank order correlation coefficient (ρ) (Eq. (1)), since there is no direct relationship between both variables (RMSD and pK_d).

The shape of the energy surface is defined by the scoring function, in this study the function employed by XSCORE. In this study, 100 docked conformations of each ligand represent 100 grip points on the protein–ligand complexation energy surface. Considering that the global minimum in the energy surface corresponds to the crystallographic position for ligand, it is expected that the lower K_d is associated with a smaller RMSD and vice versa. The Spearman's rank order correlation coefficient between the predicted K_d and RMSD for XSCORE is 0.39. This value obtained ($\rho = 0.39$) is higher than 0.326, which is the critical value at the 0.0005 level of significance. Our analysis indicates a correlation between RMSD and K_d, throughout the whole conformational ensemble, which suggests a picture of a wide smooth funnel on the protein–ligand complexation energy surface. These results, obtained with XSCORE, by docking simulations, strongly indicate that it could be a good way to predict the binding pocket of SpPNP.

3.9. Cross redocking

In order to assess the goodness of the docked complexes, four different snapshots were selected from the MD complexed with 3-deoxyguanosine (1 ns500 ps, 2 ns, 2 ns500 ps and 3 ns) and compared with the initial model. We performed 100 docked simulations for each moment. The docking simulations were able to predict the correct position for the ligand. The RMSD and k_d values were analyzed employing the Spearman coefficient, as previously described for the initial model. The results for the four snapshots, 1 ns 500 ps, 2 ns, 2 ns 500 ps and 3 ns, from the MD simulation reveals a correlation of the 0.38, −0.28, −0.40 and −0.80 respectively. These results suggest that the conformational changes in the binding pocket of the SpPNP are of pivotal importance in estimating ligand-binding affinity. The accommodation of the ligand into the active site during MD simulation leads to structural changes in the side-chain of residues close to the ligand. One possible reason for this observation may be the poor evaluation of entropic factors in the program employed to estimate ligand-binding affinity. The presence of solvent molecules in the system for MD simulations adds further complexity to the system, which was not well modeled by empirical scoring functions implemented in the program XSCORE.

4. Conclusions

Structural analysis of human PNP and SpPNP have shown three mutations in the binding-pocket, one in the purine-binding site and two in the second phosphate-binding site, which may justify the differences observed for affinity constants against both PNPs. Furthermore, computationally determined affinity constant of the inhibitor 3-deoxyguanosine against SpPNP and HsPNP strongly indicates that this inhibitor presents high affinity against SpPNP. In addition, MD simulation indicated that structural models of SpPNP in apo form and SpPNP:3DG complex display a stable trimer that is fully maintained over the entire simulation time. MD simulations of both systems provide information on structural and dynamical characteristics in term of flexibility. The main differences were observed, when we compared the structural model in apo form with the complexed one. The regions that are involved the binding pocket and interface of the monomers, as expected, were flexible in the protein free form presented higher stability in the complexed form. The PCA analysis corroborated with the data of the flexibility, mainly in the regions R1, R2 and R3. It could be observed with the average length distances analysis a different behavior for each monomers, as a NMV. This NMV is analyzed taking into account, the subunits, two by two, we can clearly observe that a couple of monomers present a similar behavior during the simulation. We suggest that the identical monomers of PNP trimeric structure present different behaviors during the whole

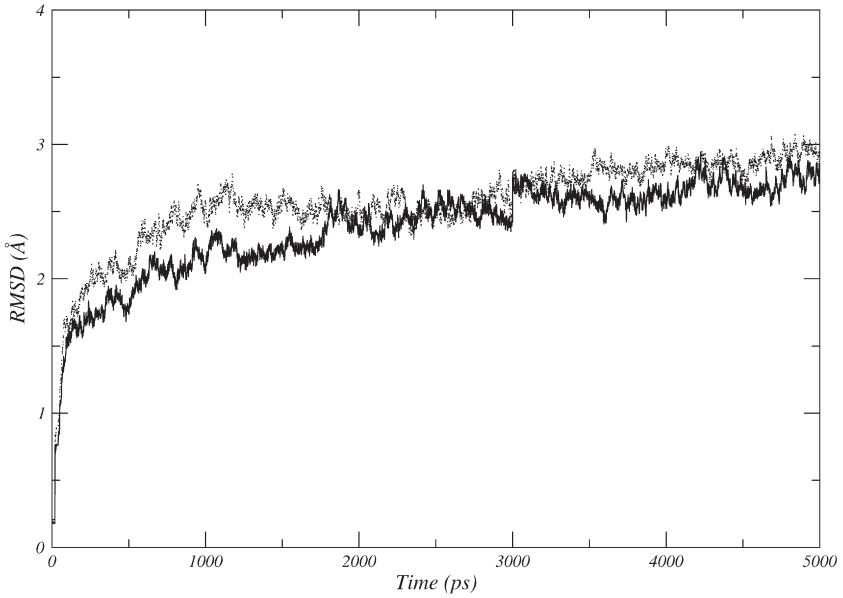

Fig. 6. Graphical representation of root-mean-square deviation (RMSD) of all Cα from starting structure of models as a function of time. The solid black line represents RMSD of SpPNP free form and dotted line SpPNP:3DG.

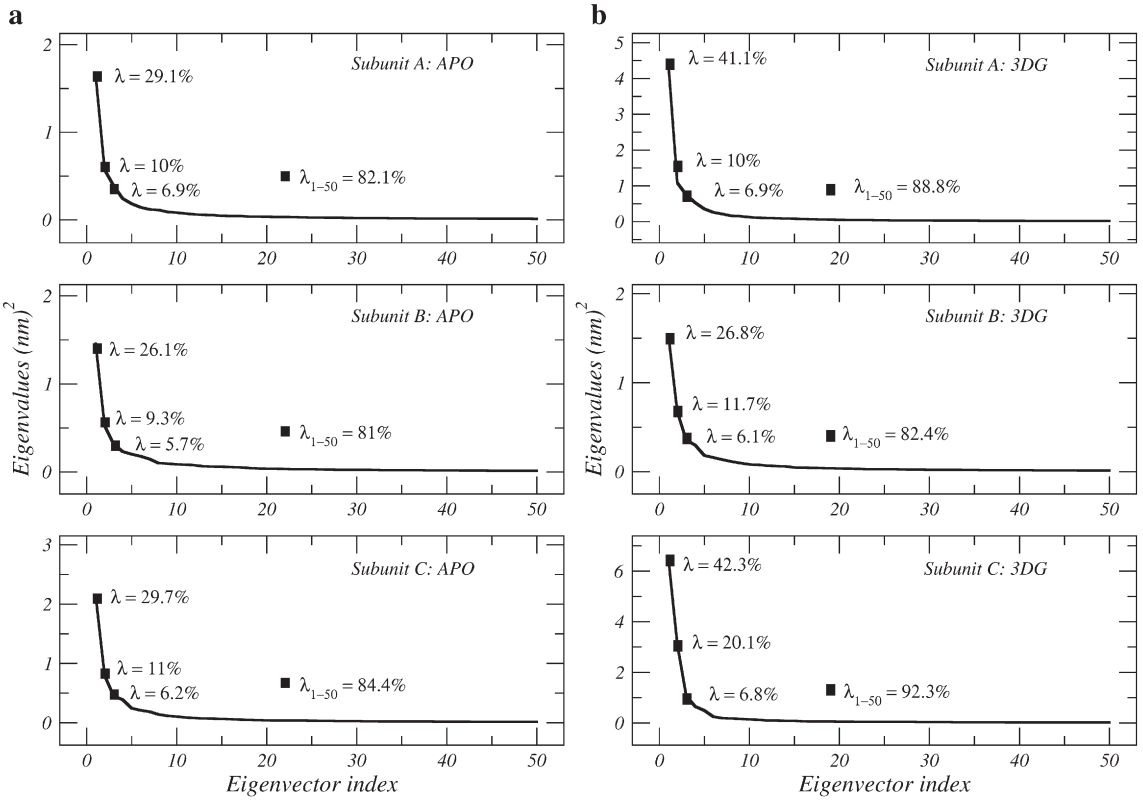

Fig. 7. a. The analysis of the apo simulation has shown that the total positional fluctuations described by the first 50 eigenvectors are 82.1%, 81% and 84.4%for subunits A, B and C respectively. The results of the first 3 eigenvectors alone represent 46% (subunit A), 41.1% (subunit B) and 46.9% (subunit C), of the total motion of the protein. b: For the SpPNP:3DG complex simulation, the analysis has shown that the first 50 eigenvectors total positional fluctuations described 88.8%, 82.4% and 92.3% for subunits A, B and C respectively. The results of the first 3 eigenvectors alone represent 58% (subunit A), 44.6% (subunit B) and 69.2% (subunit C), of the total motion for the protein.

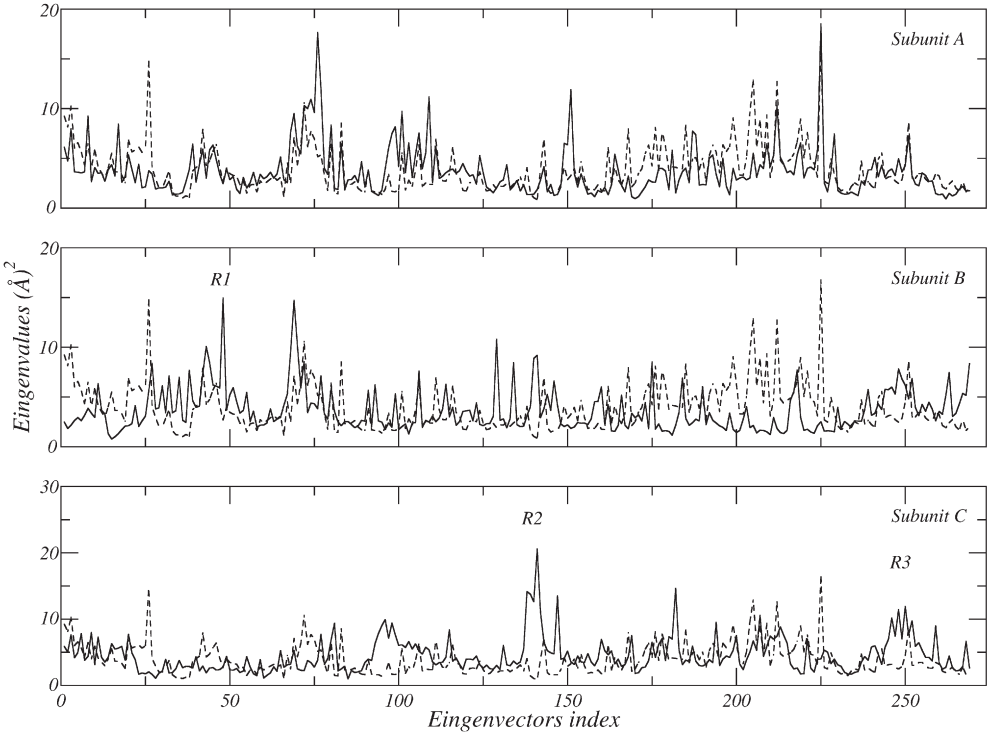

Fig. 8. The most significant fluctuations are shown in this figure. The main fluctuations are presented for the regions R1 (Ile45–Ile55), R2 (Asn138–Lys 148) and R3 (Val245–Ala255).

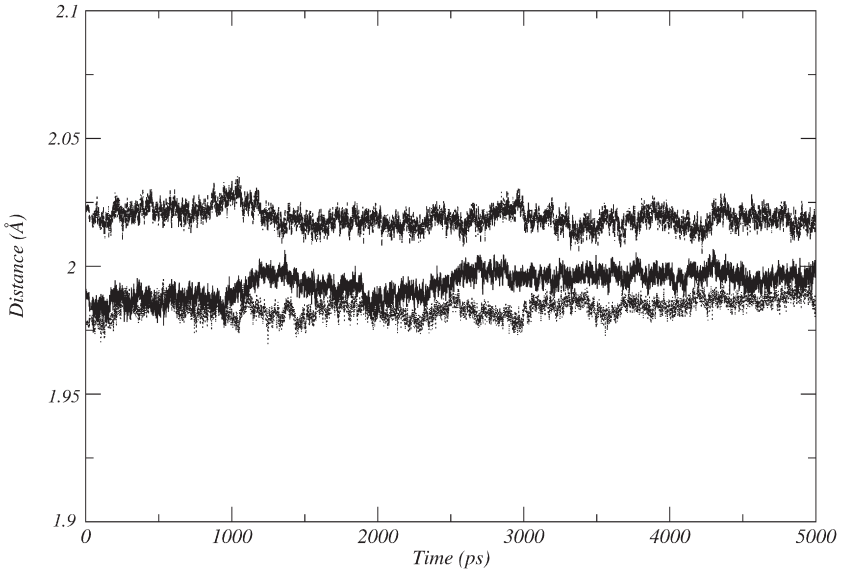

Fig. 9. The average length distances, for each atom which composed each subunit A (solid black line), B (dotted black line) and C (dashed black line).

5 ns simulation which seems to form a pattern of movements. In addition, molecular docking simulations of ligands against the binding site of PNP indicates a RMSD-K_d correlation, which suggests a funnel-shaped energy surface for protein–ligand complexation, and the docking methodology, here described, could be used for virtual screening initiatives.

Acknowledgments

This work was supported by grants from CNPq, CAPES and Instituto Nacional de Ciência e Tecnologia em Tuberculose (INCT-CNPq-MCT). W.F.A., D.S.S. and L.A.B. are senior researchers of CNPq (Conselho Nacional de Pesquisas, Brazil).

References

[1] M. Collin, A. Olsen, Extracellular enzymes with immunomodulating activities: variations on a theme in *Streptococcus pyogenes*, Infect. Immun. 71 (2003) 2983–2992.
[2] L.A. Basso, D.S. Santos, W. Shi, R.H. Furneaux, P.C. Tyler, V.L. Schramm, J.S. Blanchard, purine nucleoside phosphorylase from Mycobacterium tuberculosis. Analysis of inhibition by a transition-state analogue and dissection by parts, Biochem. 40 (2001) 8196–8203.
[3] H.M. Kalckar, Differential spectrophotometry of purine compounds by means of specific enzymes. 1. Determination of hydroxypurine compounds, J. Biol. Chem. 167 (1947) 429–443.
[4] L.F.S.M. Timmers, R.A. Caceres, A.L. Vivan, L.M. Gava, R. Dias, R.G. Ducati, L.A. Basso, D.S. Santos, W.F. de Azevedo Jr., Structural studies of human purine nucleoside phosphorylase. Towards a new specific empirical scoring function, Arch. Biochem. Biophys. 479 (2008) 28–38.
[5] F. Canduri, V. Fadel, L.A. Basso, M.S. Palma, D.S. Santos, W.F. De Azevedo Jr, New catalytic mechanism for human purine nucleoside phosphorylase, Biochem. Biophys. Res. Commun. 327 (2005) 646–649.
[6] A. Bzowska, E. Kulikowska, D. Shugar, Purine nucleoside phosphorylases: properties, functions, and clinical aspects, Pharmacol. Ther. 88 (2000) 349–425.
[7] M.J. Pugmire, S.E. Ealick, Structural analyses reveal two distinct families of nucleoside phosphorylases, Biochem. J. 361 (2002) 1–25.
[8] R.G. Silva, J.E. Nunes, F. Canduri, J.C. Borges, L.M. Gava, F.B. Moreno, L.A. Basso, D.S. Santos, Purine nucleoside phosphorylase: a potential target for the development of drugs to treat T-cell- and apicomplexan parasite-mediated diseases, Curr. Drugs Target 8 (2007) 413–422.
[9] A. Becerra, A. Lazcano, The role of gene duplication in the evolution of purine nucleotide salvage pathways, Orig. Life Evol. Biosph. 28 (1998) 539–553.
[10] D.M. Santos, F. Canduri, J.H. Pereira, M.V.B. Dias, R.G. Silva, M.A. Mendes, M.S. Palma, L.A. Basso, W.F. De Azevedo, D.S. Santos, Crystal structure of human purine nucleoside phosphorylase complexed with acyclovir, Biochem. Biophys. Res. Commun. 308 (2003) 553–559.
[11] W.F. De Azevedo, F. Canduri, D.M. dos Santos, J.H. Pereira, R.G. Silva, M.A. Mendes, L.A. Basso, M.S. Palma, D.S. Santos, Crystal structure of human PNP complexed with guanine, Biochem. Biophys. Res. Commun. 312 (2003) 767–772.
[12] R.G. Silva, J.H. Pereira, F. Canduri, W.F. de Azevedo Jr., L.A. Basso, D.S. Santos, Kinetics and crystal structure of human purine nucleoside phosphorylase in complex with 7-methyl-6-thio-guanosine, Arch. Biochem. Biophys. 442 (2005) 49–58.
[13] F. Canduri, D.M. dos Santos, R.G. Silva, M.A. Mendes, L.A. Basso, M.S. Palma, W.F. De Azevedo, D.S. Santos, Structures of human Purine Nucleoside Phosphorylase complexed with Inosine and ddI, Biochem. Biophys. Res. Commun. 313 (2004) 907–914.
[14] F. Canduri, R.G. Silva, D.M. dos Santos, M.S. Palma, L.A. Basso, D.S. Santos, W.F. De Azevedo Jr., Structure of human PNP complexed with ligands, Acta Crystallogr., D 61 (2005) 856–862.
[15] H.B. Uchoa, G.E. Jorge, N.J. da Silveira, J.C. Camera, F. Canduri, W.F. De Azevedo, Parmodel: a web server for automated comparative modeling of proteins, Biochem. Biophys. Res. Commun. 325 (2004) 1481–1486.
[16] A. Sali, T.L. Blundell, Comparative protein modelling by satisfaction of spatial restraints, J. Mol. Biol. 234 (1993) 779–815.
[17] F. Canduri, H.B. Uchoa, W.F. de Azevedo Jr., Molecular models of cyclin-dependent kinase 1 complexed with inhibitors, Biochem. Biophys. Res. Commun. 324 (2002) 661–666.
[18] T.A. Hall, BioEdit: a user friendly biological sequence alignment editor and analysis program for Windows 95/98/NT, Nucleic Acids Symp. Ser. 41 (1999) 95–98.
[19] W.F. De Azevedo, H-.J. Mueller-Dieckmann, U. Schulze-Gahmen, P.J. Worland, E. Sausville, S.H. Kim, Structural basis for specificity and potency of a flavonoid inhibitor of human CDK2, a cell cycle kinase, Proc. Natl. Acad. Sci. U. S. A. 93 (1996) 2735–2740.
[20] W.F. De Azevedo Jr., F. Canduri, V. Fadel, L.G.V.L. Teodoro, V. Hial, R.A.S. Gomes, Molecular model for the binary complex of uropepsin and pepstatin, Biochem. Biophys. Res. Commun. 287 (2001) 277–281.
[21] J.H. Pereira, F. Canduri, J.S. de Oliveira, N.J.F. Silveira, L.A. Basso, M.S. Palma, W.F. De Azevedo, D.S. Santos, Structural bioinformatics study of EPSP synthase from *Mycobacterium tuberculosis*, Biochem. Biophys. Res. Commun. 312 (2003) 608–614.
[22] R. Wang, L. Lai, S. Wang, Further development and validation of empirical scoring functions for structure-based binding affinity prediction, J. Comput.-Aided Mol. Des. 16 (2002) 11–26.
[23] R.A. Laskowski, M.W. Macarthur, D.S. Moss, J.M. Thornton, PROCHECK — a program to check the stereochemical quality of protein structure, J. Appl. Crystallogr. 26 (1993) 283–291.
[24] Collaborative Computation Project, Number 4, Acta Crystallogr., D 50 (1994) 760.
[25] A.C. Wallace, R.A. Laskowski, J.M. Thornton, LIGPLOT: a program to generate schematic diagrams of protein–ligand interactions, Protein Eng. 8 (1995) 127–134.
[26] R. Chen, L. Li, Z. Weng, ZDOCK: an initial-stage protein-docking algorithm, Proteins 52 (2003) 80–87.
[27] R. Chen, Z. Weng, Docking unbound proteins using shape complementarity, desolvation, and electrostatics, Proteins 47 (2002) 281–294.
[28] W.F. de Azevedo Jr., R. Dias, Computational methods for calculation of ligand-binding affinity, Curr. Drug Targets 9 (12) (2008) 1031–1039.
[29] W.F. De Azevedo Jr., Protein–drug interactions, Curr. Drug Targets 9 (12) (2008) 1030-1030.
[30] R. Dias, W.F. De Azevedo Jr., Molecular docking algorithms, Curr. Drug Targets 9 (12) (2008) 1040–1047.
[31] F. Canduri, W.F. De Azevedo Jr., Protein crystallography in drug discovery, Curr. Drug Targets 9 (12) (2008) 1048–1053.
[32] I. Pauli, L.F.S.M. Timmers, R.A. Caceres, M.B.P. Soares, W.F. De Azevedo Jr., *In silico* and *in vitro*. Identifying new drugs, Curr. Drug Targets 9 (12) (2008) 1054–1061.
[33] R. Dias, L.F.S.M. Timmers, R.A. Caceres, W.F. De Azevedo Jr., Evaluation of molecular docking using polynomial empirical scoring functions, Curr. Drug Targets 9 (12) (2008) 1062–1070.

[34] W.F. De Azevedo Jr., R. Dias, Experimental approaches to evaluate the thermodynamics of protein–drug interactions, Curr. Drug Targets 9 (12) (2008) 1071–1076.
[35] R.A. Caceres, I. Pauli, L.F.S.M. Timmers, W.F. De Azevedo Jr., Molecular recognition models: a challenge to overcome, Curr. Drug Targets 9 (12) (2008) 1077–1083.
[36] G.B. Barcellos, I. Pauli, R.A. Caceres, L.F.S.M. Timmers, R. Dias, W.F. De Azevedo Jr., Molecular modeling as tool for drug discovery, Curr. Drug Targets 9 (12) (2008) 1084–1091.
[37] L.F.S. Timmers, I. Pauli, R.A. Caceres, W.F. De Azevedo Jr., Drug-binding databases, Curr. Drug Targets 9 (12) (2008) 1092–1099.
[38] H.L.N. de Amorim, R.A. Caceres, P.A. Netz, Linear interaction energy (LIE) method in lead discovery and optimization, Curr. Drug Targets 9 (12) (2008) 1100–1105.
[39] P.E. Bourne, Weissig, Structural bioinformatics, Wiley–Liss, Inc., California, 2003.
[40] W.F. De Azevedo, F. Canduri, D.M. Santos, R.G. Silva, J.S. Oliveira, L.P.S. Carvalho, L.A. Basso, M.A. Mendes, M.S. Palma, D.S. Santos, Crystal structure of human purine nucleoside phosphorylase at 2.3 Å resolution, Biochem. Biophys. Res. Commun. 308 (2003) 545–552.
[41] K.H. Zou, K. Tuncali, S.G. Silverman, Correlation and simple linear regression, Radiology 227 (2003) 617–622.
[42] M. Otyepka, V. Krystof, L. Havlicek, V. Siglerova, M. Strnad, J. Koca, Docking-based development of purine-like inhibitors of cyclin-dependent kinase-2, J. Med. Chem. 43 (2000) 2506–2513.
[43] D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A.E. Mark, H.J. Berendsen, GROMACS: fast, flexible, and free, J. Comp. Chem. 26 (2005) 1701–1718.
[44] C. Ooestenbrik, T.A. Soares, N.F. van der Vegt, W.F. van Gunsteren, Validation of the 53A6 GROMOS force field, Eur. Biophys. J. 34 (2005) 273–284.
[45] D.M. van Aalten, R. Bywater, J.B. Findlay, M. Hendlich, R.W. Hooft, G. Vriend, PRODRG, a program for generating molecular topologies and unique molecular descriptors from coordinates of small molecules, J. Comput. Aided Mol. Des. 10 (1996) 255–262.
[46] M.W. Schmidt, K.K. Baldridge, J.A. Boatz, S.T. Elbert, M.S. Gordon, J.H. Jensen, S. Koseki, N. Matsunaga, K.A. Nguyen, S.J. Su, T.L. Windus, M. Dupuis, J.A. Montgomery, General atomic and molecular electronic-structures system, J. Comput. Chem. 14 (1993) 1347–1363.
[47] N. Guex, M.C. Peitsch, SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling, Electrophoresis 18 (1997) 2714–2723.
[48] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, J. Hermans, Intermolecular Forces, D. Reidel Publishing Company, Netherlands, 1981.
[49] B. Hess, H. Bekker, H.J.C. Berendsen, J.G.M.E. Fraaije, LINCS: a linear constraint solver for molecular simulations, J. Comput. Chem. 18 (1997) 1463.
[50] S. Miyamoto, P.A. Kollman, SETTLE — an analytical version of the shake and rattle algorithm for rigid water models, J. Comput. Chem. 13 (1992) 952–962.
[51] S. Chowdhuri, M.L. Tan, T. Ichiye, Dynamical properties of the soft sticky dipole–quadrupole–octupole water model: a molecular dynamics study, J. Chem. Phys. 125 (2006) 144513.
[52] D.M. York, T.A. Darden, L.A. Pedersen, The effect of long-range electrostatic interactions in simulations of macromolecular crystals — a comparison of the Ewald and truncated list methods, J. Chem. Phys. 99 (1993) 8345–8348.
[53] O.N. de Souza, R.L. Ornstein, J. Biomol. Struct. Dyn. 16 (1999) 1205.
[54] A. Amadei, A.B. Linssen, H.J. Berendsen, Essential dynamics of proteins, Proteins 17 (1993) 412–425.
[55] W.L. Delano, J.W. Lam, PyMOL: a communications tool for computational models, Abstr. Pap. -Am. Chem. Soc. 230 (2005) U1371–U1372.
[56] P.Y. Chou, G.D. Fasman, Empirical predictions of protein conformation, Annu. Rev. Biochem. 47 (1978) 251–276.
[57] L.J. McGuffin, K. Bryson, D.T. Jones, The PSIPRED protein structure prediction server, Bioinformatics 16 (2000) 404–405.
[58] R.A. Caceres, L.F. Saraiva Timmers, R. Dias, L.A. Basso, D.S. Santos, W.F. De Azevedo Jr., Molecular modeling and dynamics simulations of PNP from *Streptococcus agalactiae* (2008), Bioorg. Med. Chem. 16 (2008) 4984–4993.
[59] I. Pauli, L.F.S.M. Timmers, R.A. Caceres, L.A. Basso, D.S. Santos, W.F. De Azevedo Jr., Molecular modeling and dynamics studies of Purine Nucleoside Phosphorylase from Bacteroides fragilis, J. Mol. Model., 10 (in press).
[60] L.F.S.M. Timmers, R.A. Caceres, L.A. Basso, D.S. Santos, W.F. De Azevedo Jr., Structural bioinformatics study of PNP from *Listeria monocytogenes*, Prot. Peptide Letters 15 (2008) 843–849.
[61] W.F. de Azevedo, F. Canduri, D.M. dos Santos, J.H. Pereira, M.V.B. Dias, R.G. Silva, M.A. Mendes, L.A. Basso, M.S. Palma, D.S. Santos, Biochem. Biophys. Res. Commun. 309 (2003) 917–922.
[62] P.E. Leopold, M. Montal, J.N. Onuchic, Protein folding funnels: a kinetic approach to the sequence–structure relationship, Proc. Natl. Acad. Sci. U. S. A. 89 (1992) 8721–8725.