J Mol Model (2001) 7:80–89
DOI 10.1007/s008940000001

ORIGINAL PAPER

Nuria Campillo · Hampapathalu A. Nagarajaram
Indira Ghosh

Phosphoribosyltransferase superfamily: A comparative structural analysis

Received: 14 August 2000 / Accepted: 24 October 2000 / Published online: 21 April 2001

Abstract Purine phosphoribosyltransferases are members of a group of enzymes that are responsible for the formation of purine, pyrimidine and pyridine nucleotides. One feature of this family is a flexible loop, which is involved in the catalytic mechanism. This loop is variable both in sequence and structure in the phosphoribosyltransferase family. This paper describes the modelling and validation of a model of *Plasmodium falciparum* hypoxanthine-guanidine- phosphoribosyltransferase in an open conformation. A comparison of this model with 3D-structures of other members of the phosphorybosyltransferase family has allowed an assessment of the role of the open and closed conformations of the loop in the catalytic mechanism.

Keywords Comparative modelling · Sequence alignment · Phosphoribosyltransferase superfamily · *Plasmodium falciparum*

Abbreviations *GPAT* Glutamine PRPP amidotransferase · *GMP* Guanosine-5'-monophosphate · *GPRT* Guanine phosphoribosyltransferase · *HPRT* Hypoxanthine phosporibosyltransferase · *HGPRT* Hypoxanthine-guanine-phosphoribosyltransferase · *HGXPRT* Hypoxanthine-guanine-xanthine-phosphoribosyltransferase · *IMP* Inosine-5'-monophosphate · *IMU* ImmucillinGP · *OPRT* Orotate phosphoribosyltransferase · *PDB* Protein data base · *Pf Plasmodium falciparum* · *POP* 2-Pyrophosphate · *PRPP* 5-Phospho-a-D-ribosyl-1-pyrophosphate · *PPi* Pyrophosphate · *PRTase* Phosphoribosyltransferase · *RMSD* Root mean square deviation · *SCR* Structurally conserved region · *SVR* Structurally variable region · *TS* Transition state · *XGPRT* Xanthine-guanine phosphoribosyltransferase · *XPRT* Xanthine phosphoribosyltransferase · *XMP* Xanthine-5′-monophosphate

N. Campillo (✉) · H.A Nagarajaram
Department of Biochemistry, University of Cambridge,
80 Tennis Court Road, Cambridge CB2 1GA, United Kingdom
e-mail: nuria@cryst.bioc.cam.ac.uk
Tel.: +44-1223 766031, Fax: +44-0-1223-766082

I.Ghosh
AstraZeneca, R & D, Bangalore, India

Introduction

Purine PRTases are members of a group of enzymes that are responsible for the formation of purine, pyrimidine and pyridine nucleotides. This family of enzymes uses PRPP and a nitrogenous base to form a nucleoside monophosphate with liberation of PPi [1].

The PRTase enzymes can be classified into two unrelated groups, types I and II [2]. The type I PRTases are characterised by a PRPP binding site motif, which features two adjacent acidic residues surrounded by hydrophobic residues. All known type I PRTases have a canonical fold characterised by a sheet of five β-strands surrounded by three or four α-helices [3]. The other feature of this family is the flexible loop, which is involved in the catalytic mechanism [4]. This loop is variable both in sequence and structure in the PRTase family. The type II PRTases are less well described than type I. Members of this class of enzyme do not have the binding-site motif and their structures comprise a mixed α/β N-terminal domain and an α/β barrel-like C-terminal domain [2].

The type I PRTase superfamily consists of XGPRTs, HGXPRTs, GPATs, HGPRTs and OPRTs. The first four PRTase families are involved in purine salvage, whereas members of the fifth family are involved in the salvage of pyrimidine (Table 1).

The PRTase enzymes follow an S_N1-type mechanism that results in an oxocarbonium-like TS [5]. This TS cannot occur in an active site that is exposed to solvent because the TS complex would be hydrolysed immediately, rather than going on to product formation. It has been proposed that the flexible loop plays a role in oxocarbonium stabilisation by moving to cover the active site [4]. Thus, the enzymes adopt two different conformations, one described as open or “inactive”, where the flexible loop is far away from the active site, and the second as

Table 1 Summary of PRTase structures available in the Brookhaven Protein Data Bank (*PDB*). *a* active, *i* inactive

PDB code	Resolution (Å)	Source	PRTase	Form
1cjb	2.0	*P. falciparum*	HPRT	a
pfm[a]	–	*P. falciparum*	HPRT	i
1bzy	2.0	Human	HPRT	a
1hmp	2.5	Human	HPRT	i
1dbr	2.4	*Toxoplasma gondii*	HPRT	a
1qk3	1.6	*Toxoplasma gondii*	HPRT	i
1nul	1.8	*E. coli*	XPRT	a
1ecc	2.4	*E. coli*	GAT	a
1ecb	2.7	*E.coli*	GAT	i
1oro	2.4	*E.coli*	OPRT	a
1opr	2.6	*S. tryphimurium*	OPRT	i

[a] Model of Pf HGPRT

Table 2 Pf HGPRT homologues of known 3D structure

Homologue	PDB code	Sequence identity (PfHGPRT) (%)	References
HGPRT *(Homo sapiens)*-Chain A	1hmp	49.1	[4]
HGPRT *(Toxoplasma gondii)*-Chain A	1dbr	49.3	[5]
HGPRT *(T. foetus)*-Chain A	1hgx	27.4	[8]

Table 3 Options used in MODELLER program

Option	
Set starting_model	Refine
Set deviation	4.0
Set hetatm_IO	Off
Set watr_IO	Off
Set hydrogen	Off
Call routine	Model

closed or "active", where the flexible loop covers the active site.

Most of the HGPRTs follow a ternary complex mechanism where both the substrates bind to the enzyme together to form a TS complex. The mechanism has been studied for the cases involving human [6], *Schistosoma mansoni* [7] and *Tritrichomonas foetus* [8], where the PRPP binds to the enzyme first followed by the purine base. PPi release is followed by the release of the respective monophosphate. In comparison, the Pf HGPRT follows a ping-pong type of mechanism (D. Sarkar, I. Ghosh, S. Datta manuscript in preparation). This might indicate different dynamics for Pf HGPRT.

Recently, several crystal structures of the open and closed conformations of the PRTases have been solved. Structures with the open conformation are available for human HGPRT (PDB code: 1hmp [4]), *Toxoplasma gondii* HXGPRT (PDB code: 1dbr [5]), *Escherichia coli* OPRT (PDB code: 1oro [9]) and *E.coli* GPRT (PDB code: 1ecc [10]) and the structures of closed forms include human HGPRT (PDB code: 1bzy [11]), *Toxoplasma gondii* HXGPRT (PDB code: 1qk3 [12],) and *Plasmodium falciparum* HXGPRT (PDB code: 1cjb [13]). The availability of these structures prompted us to make a detailed comparative structural analysis of the flexible loop in both conformations and its interactions with the substrate. Since there is no experimental structure available for *P. falciparum* HGPRT in open conformation, we built a model using comparative modelling techniques.

Materials and methods

Comparative modelling

Template structures and sequence alignment

The modelling of the Pf HGPRT in its open conformation was initiated before the crystal structure of its closed conformation was published. Structural homologues that show appreciable sequence similarity with Pf HGPRT are given in Table 2. Of these, human HGPRT and *Toxoplasma gondii* HGPRT show significant sequence similarity with Pf HGPRT and both have open conformation crystal structures available. We therefore, chose these as the template structures for comparative modelling of Pf HGPRT.

In order to obtain an accurate model, a correct sequence alignment of the target sequence with the homologues used as basis structures is essential. To achieve an accurate sequence alignment, we have used a structure-assisted approach [14] consisting of three steps.

1. A structure-based sequence alignment of the template structures was obtained using the COMPARER [15,16] suite of programs, and formatted using the program JOY [17].
2. From the alignment, a sequence profile was constructed using the sequence alignment program CLUSTALW [18,19].
3. The target sequence was then aligned to this profile. The alignment is shown in Fig. 1.

Model building

We adopted a recursive approach comprising sequence alignment and model building [14]. From the best alignment of template structures to target sequences, 15 3D models containing all non-hydrogen atoms were obtained automatically using the method implemented in MODELLER (version 4.0) [20]. Minimisation of the models was performed automatically by the program. The models had to satisfy most restraints used to calculate them, particularly those restraining to their stereochemistry. The one corresponding to the lowest value of the objective function was used for further analysis. The cycle of realignment, modelling and structure validation was repeated until no further improvement on the structure was observed. Table 3 shows the options used for running the MODELLER program.

Results

Pf HGPRT model

The alignment used for the final model is shown in Fig. 2. Note that region 100–139 of 1dbr has not been included in the final alignment as it is very poorly defined in the struc-

Fig. 1 Alignment of Pf HGPRT with 1hmp (Chain A) and 1dbrm as the structural templates. 1dbrm is a truncated version of 1dbra with the poorly defined section from residues 100–139 removed. Alignment was generated using COMPARER [15,16] and formatted with JOY [17]. Residue numbers are shown in parentheses

Key to JOY alignments		
Solvent inaccesible	UPPER CASE	Z
Solvent accesible	Lower case	z
Positive ϕ	*Italic*	*z*
cis-Peptide	Breve	
Hydrogen bond to other sidechain	Tilde	z̄
Hydrogen bond to mainchain amide	**Bold**	**z**
Hydrogen bond to mainchain carbonyl	Underline	z
Disulphide bond	Cedilla	ç
α-helix	Red	z
β-helix	Blue	z
3_{10}-helix	Maroon	z

ture. Target sequence to structural template alignments was used as input to the program MODELLER. The resulting output was a comparative model of the target sequence. It is worth mentioning the improvement of the final model with respect to the rest of the models. Table 4 shows the values of the Ramachandran plot [21] for the initial model, for an intermediate model that only used 1hmp as template and for the final model. As can be seen, the final model shows the best values.

Table 4 Values of Ramachandran plot of the initial, intermediate and final models of HGPRT

Residues	Initial (%)	Intermediate (%)	Final (%)
In most favoured region	75.7	75.2	81.1
In additional allowed region	18.4	18.9	15.0
In generously allowed region	3.9	4.4	3.4
In disallowed region	4.0	1.5	0.5

From the alignment in Fig. 2, the sequence similarity in the SCRs is strong, and all important residues in the active site are well conserved. Of the SVRs, the length of segments in the target enzyme varies from 3 to 27, of which the longest are flexible loops.

The final model was built and the variable regions were modelled. Three variable regions were found to be unsatisfactory; these were the N-terminal region (1–16), the loop (97–106) between an α-helix and β-strand and the previously mentioned flexible loop (112–127). Loop 1–16 was remodelled using the equivalent segment from 1dbr. Loop 97–106 is characterised by three insertions. To model this loop, no suitable structural fragment was available from the family of PRTases. Therefore, we searched for a fragment outside the family. The search was made using the LOOPSEARCH option in the SYBYL software suite [22]

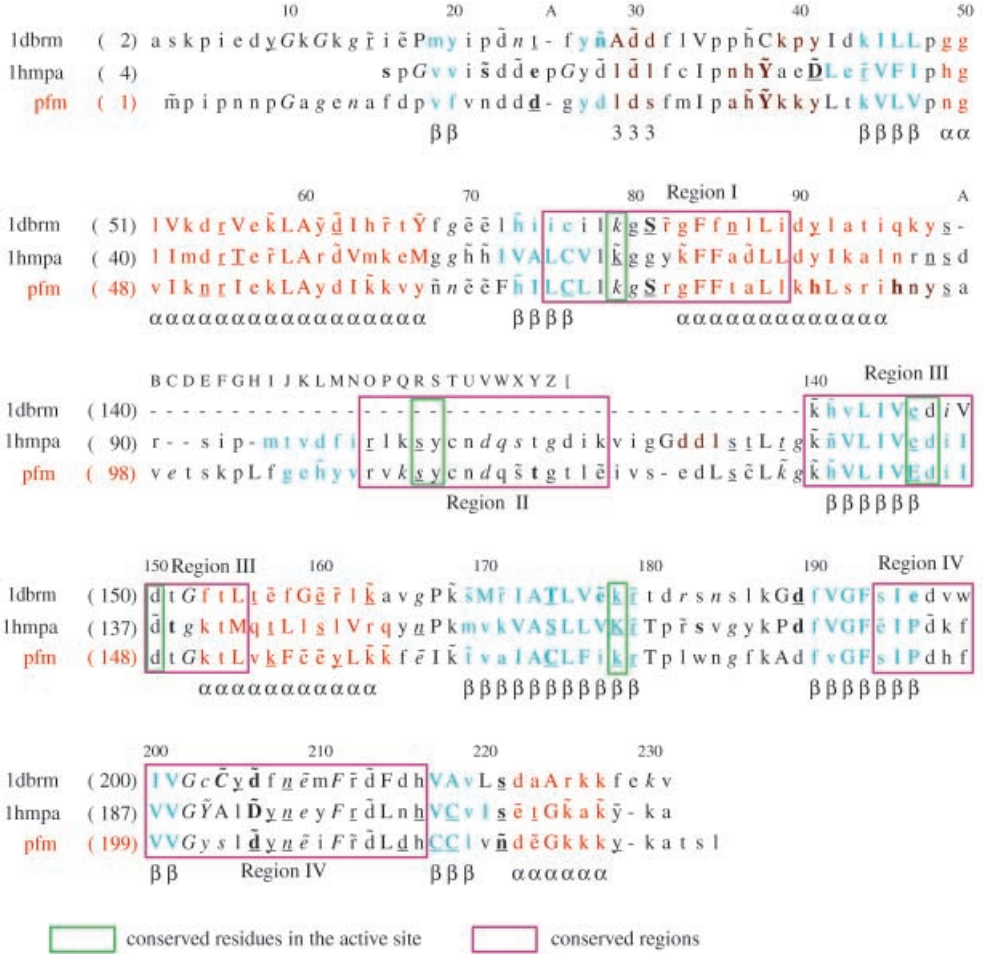


Fig. 2 Structure-based sequence alignment of the final model of Pf HGPRT(pfm) with 1hmp and 1dbr. See Fig. 1 for definitions

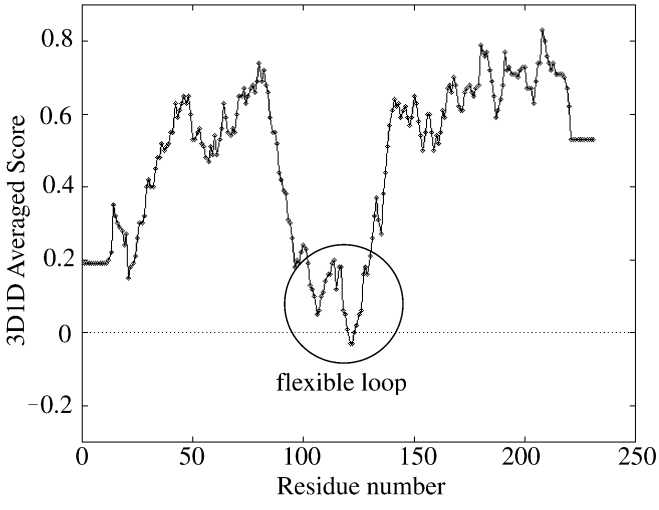


Fig. 3 Energy profile of Pf HGPRT model, obtained from VERIFY3D program [23]. A window of 10 residues was used to compute the energy profile

Table 5 Candidate structural fragments for modelling the loop (97–106)

	Source	Local sequence similarity	The RMSDs (Å) of the anchor regions
1	1the_A	42.71	0.22
2	1the-B	42.71	0.23
3	1huc_B	42.71	0.26
4	3pga_4	42.71	0.29
5	3pga_4	45.83	0.29
6	1cg2_C	51.04	0.28
7	2gbp	45.83	0.29
8[a]	1cgb	61.46	0.34
9	1huc	42.71	0.34
10	1ego	61.46	0.35

[a] The fragment selected for modelling the loop

coupled database. This database has been built from a large set of high-resolution protein structural fragments. The best 25 fragments from the search were selected and fitted to the model and visually inspected for stereochemical compatibility. Table 5 gives the list of the ten best segments. Loop 112–127 corresponds to the flexible loop. In 1hmp and 1dbr it is a poorly defined region (in addition, 1dbr has three residues,116–119, completely missing). However, the 1hmp structure provides a crude framework for the entire loop and this was used as the basis for its construction. To model the N-terminal region (1–16) we used

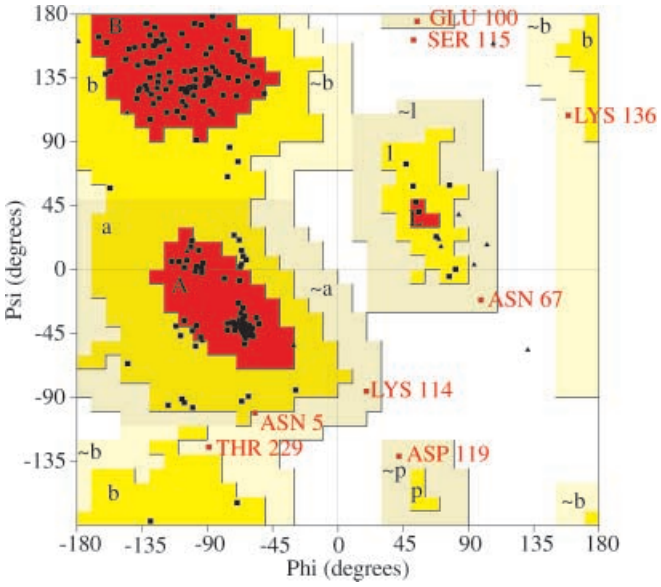


Fig. 4 Ramachandran plot of Pf HGPRT model, obtained from PROCHECK program [21]

Plot statistics		
Residues in most favoured regions (A, B, L)	167	81.1%
Residues in additional allowed regions [*a*, *b*, *l*, *p*]	31	15.0%
Residues in generously allowed regions [~*a*, ~*b*, ~*l*, ~p]	7	3.4%
Residues in disallowed regions	1	0.5%
Number of non-glycine and non-proline residues	206	100.0%
Number of end-residues (excluding Gly and Pro)	2	
Number of glycine residues (shown as *triangles*)	14	
Number of proline residues	9	
Total number of residues	231	

1dbr as a template because in this region the target sequence has a high sequence similarity with 1dbr. Thus, of the three loops only one was modelled using a fragment selected from outside the family.

After the refinement process the model was validated using the VERIFY 3D [23] and PROCHECK [21] programs. Figure 3 shows the energy profile, obtained with VERIFY3D. The stereochemical quality of the final model was examined with PROCHECK. The Ramachandran plot (Fig. 4) shows that only a few residues (0.5%) are in disallowed regions.

Sequence and structural variability in the PRTase family

We investigated the 3D-structures of several PRTases enzymes: HGPRT from human, *T. foetus*, and *T. gondii*, HXGPRT from *P. falciparum*, XPRT from *E. coli* as well as OPRT and GPRT, both from *E. coli*. The structural alignment of these structures generated using COMPARER is shown in Fig. 5.

Only 17 residues that are common to all these proteins are identified in the sequence alignment. However, high local sequence similarity exists between residues 140–150 (1cjb numbering). This region corresponds to the PRPP binding site motif.

Figure 6 shows schematically the alignment of the secondary structures of known PRTase family sequence structures. The canonical core derived from their consensus consists of eight β-strands (β_1-β_8) and four α-helices (α_1-α_4). It is interesting to note that each individual structure is characterised by a few insertions of secondary structure with respect to the canonical core. For example, the HGPRTs (PDB codes: 1cjb, pfm, 1hmp, 1bzy, 1dbr and 1qk3) have insertions of two β-strands and one α-helix, the OPRTs (PDB codes: 1oro and 1opr) have an insertion of one α-helix and the GPRTs (PDB codes: 1ecc and 1ecb) have insertions of two α-helices.

Structural changes between open and closed forms of PRTases

The major structural change during enzymatic activity occurs in the flexible loop, which moves to cover the active site in order to shield the substrate in the TS from the surrounding aqueous environment (Fig. 7). Comparison of the open and closed forms indicates that the loop might sweep out at an angle of about 400 in order to cover the active site. In most of the PRTases the loop covers the active site of the same subunit; however, in OPRT the loop covers the active site of the adjacent subunit. This is necessary since the loop is considerably shorter, compared to the other PRTases, and cannot cover the active site of the same subunit.

The structural change that occurs in the loop involves large conformational changes [24] in the component residues. However, it is still not understood how exactly such large changes are being effected. Furthermore, the flexible loop consists of polar residues, which demand hydrogen-bonding partners. One may postulate that in the inactive enzyme the loop swings between the fully open conformation (where all the side chains are fully accessible to the surrounding water) to the near-closed conformation (some of the side chains becoming inaccessible to the water). When the substrate is in the TS, new interactions, probably involving the side chains of the inaccessible residues of the flexible loop, must stabilise the near-closed conformational state. In order to test this we computed the interactions between the inhibitor IMU300, POP and the two magnesium ions together representing the TS analogue of the substrate and the flexible loop in human HXGPRT. There are indeed three such hydrogen bonds (Table 6) formed between the flexible loop and the TS analogue. However, it is difficult to say whether these three hydrogen bonds are sufficient for the stabilisation of the closed conformation.

In addition to the flexible loop structural changes other changes might occur in some other regions of the enzyme. To examine this, we superposed the active and

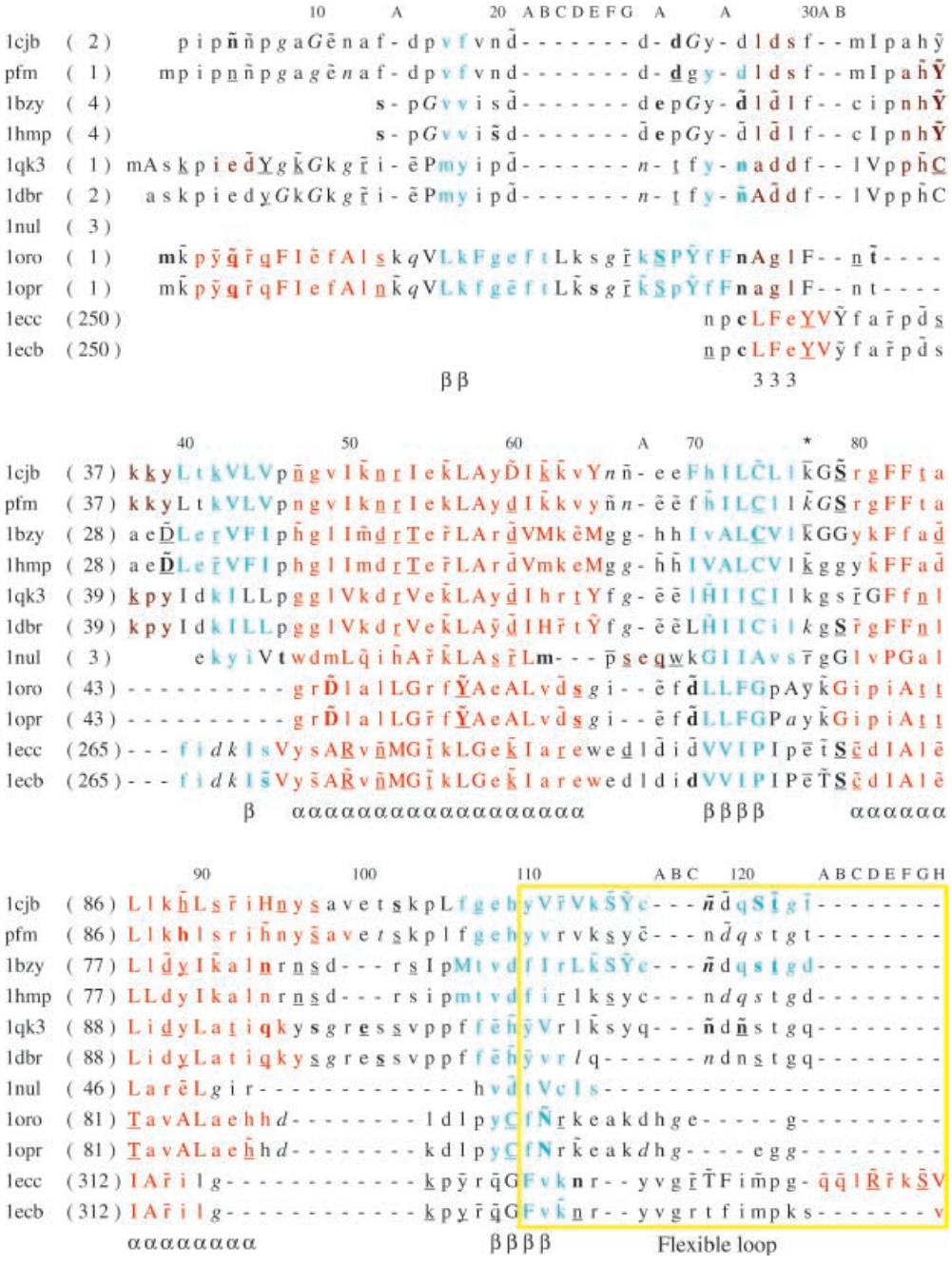


Fig. 5 Structural alignment of the PRTase family. The flexible loop and the conserved binding site motif are indicated with *yellow* and *green* boxes respectively.* The residues involved in the *cis*-peptide

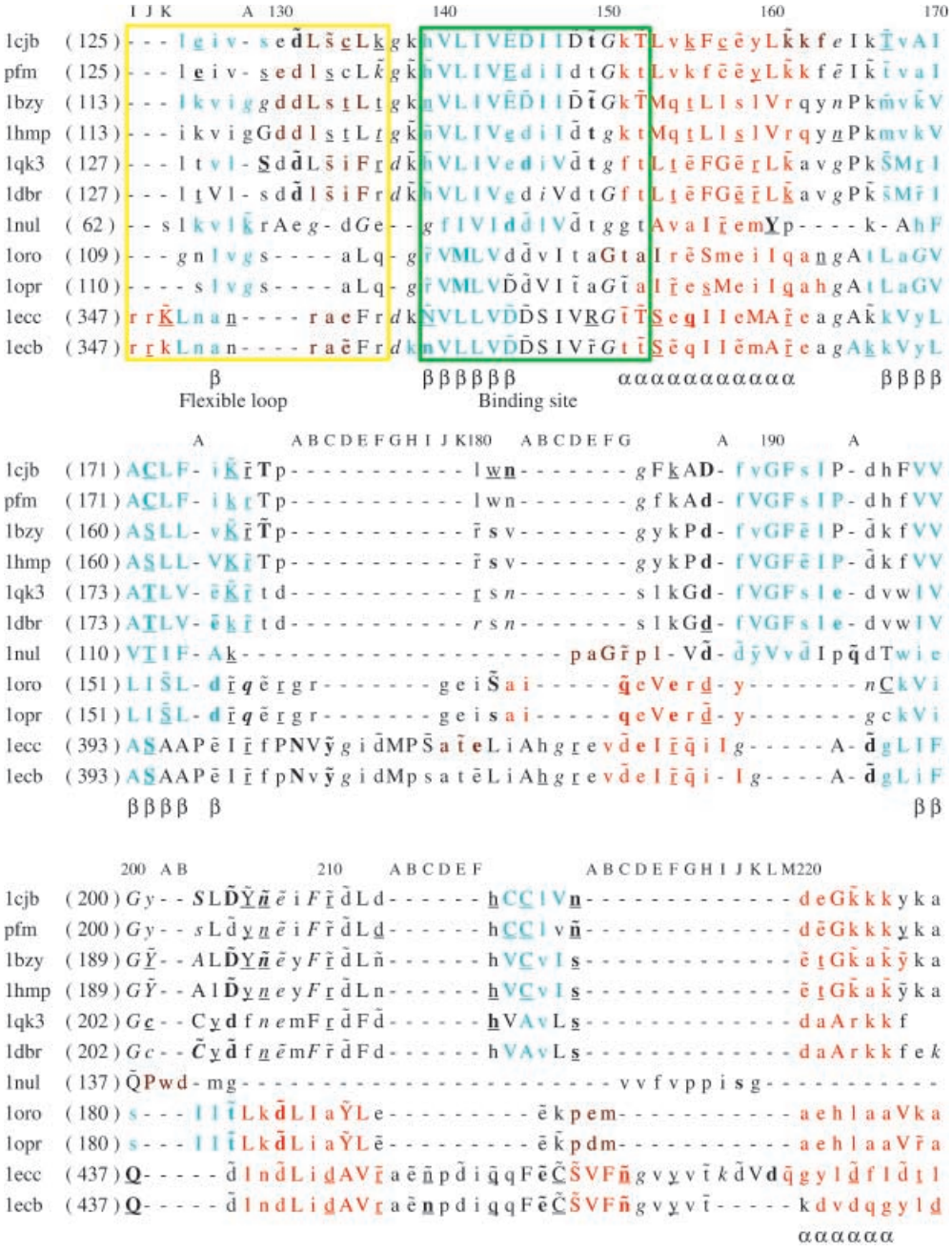
Table 6 Hydrogen bonds between residues of the flexible loop and transition state of human enzyme

Residue	Atom	Ligand atom	Distance (Å)
Tyr 104	OH	O_3^P (IMU)	2.8
Tyr 104	OH	N_7 (IMU)	3.9
Ser 103	N	O_6' (POP)	2.7
Tyr 104	N	O_6 (POP)	2.9

Table 7 RMS of the human enzyme

Region	RMS (Å)
Binding site	1.03
61–71	2.04
80–95	3.35
185–202	1.08

Fig. 5 (continued)



inactive forms of the human PRTase and listed only those regions (flexible loop excluded) which show RMSD values greater than 1.0 Å (Table 7). It is interesting to note that some of the residues undergo appreciable structural changes and these regions are located in the vicinity of the active site. These are due to the contributions not only of local conformational changes but also of the rigid body shifts of some of the secondary structural elements.

Of the conformational changes, one striking feature is the change of peptide bond conformation from *cis* to *trans* [25] (Fig. 5). In not one case is there a proline on the C-terminal side of the *cis* peptide. In most of the open structures, the exception being GPRT, the peptide bond is in the *trans* conformation, while in the corresponding closed forms it is in the *cis* conformation. This association of a particular peptide bond conformation with the active and inactive forms of the enzyme points to *cis-trans* isomerisation of the peptide bond as the enzyme changes between the two forms. If this is so, then how is the energy cost for the peptide bond isomerisation compensated? One possibility is the stabilisation of the TS by interactions in the *cis* peptide conformation. In fact there are extensive hydrogen bonding interactions between the TS analogue and the enzyme (Table 8). In particular, as can be seen in Fig. 8a, the amide nitrogen

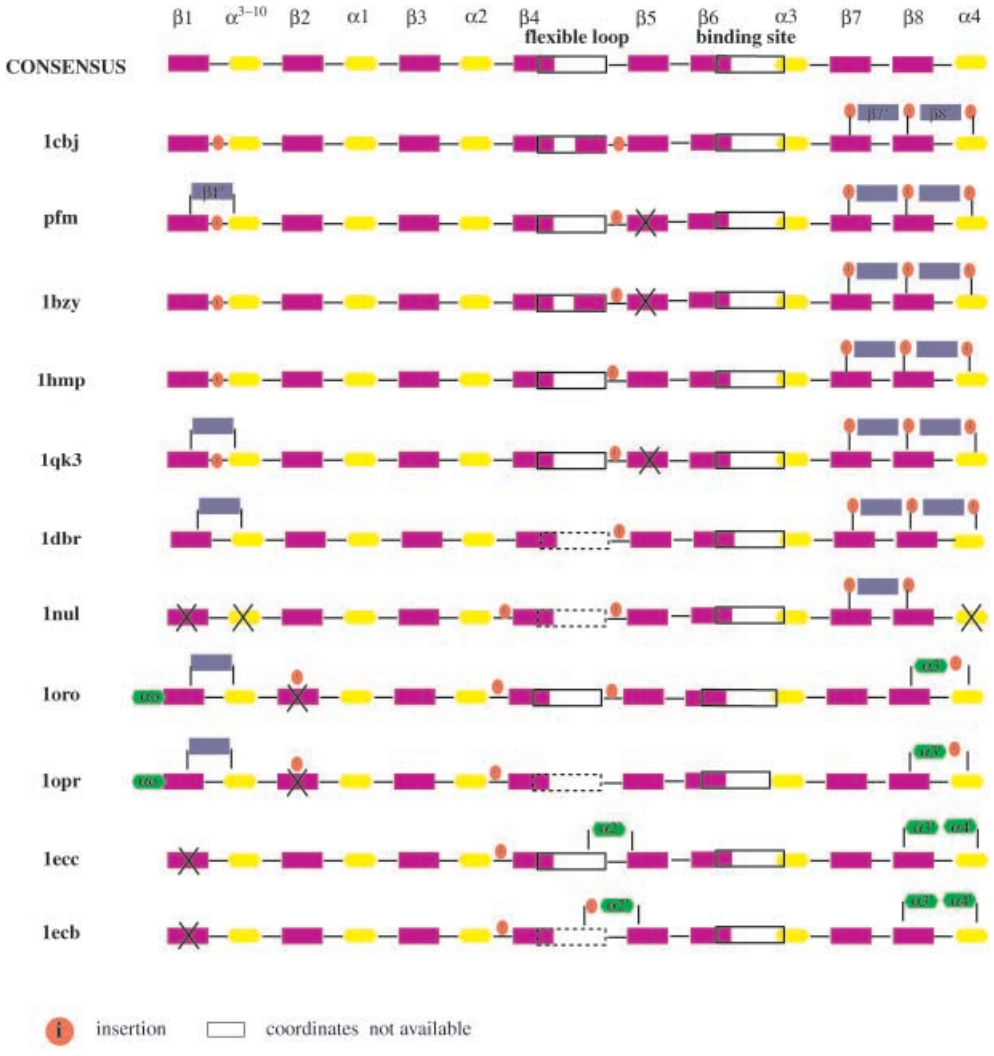


Fig. 6 Schematic representation of the secondary structure of the PRTase family

Table 8 Hydrogen bonds between residues in the TS analogue

Residue	Atom	Ligand atom	Distance (Å)
Gly 139	N	O_1P (IMU)	2.9
Asp 137	N	O_1P(IMU)	3.0
Thr 141	N	O_2P(IMU)	2.7
Thr 141	OG1	O_2P(IMU)	2.7
Lys 140	N	O_2P(IMU)	3.4
Thr 138	OG1	O_3P(IMU)	2.6
Glu 133	OE2	O_3*(IMU)	2.8
Asp 134	OD2	O_2*(IMU)	2.5
Asp 137	OD2	N_7(IMU)	2.8
Tyr 104	OH	N_7(IMU)	3.9
Lys 165	NZ	O_6(IMU)	2.7
Val 187	N	O_6(IMU)	3.5
Val 187	O	N_1(IMU)	2.7
Asp 193	O	N_2(IMU)	2.8
Val 187	O	N_2(IMU)	2.8
Asp193	N	N_2(IMU)	3.4
Lys 68	N	O_1(POP)	3.0
Arg 199	NH2	O_1(POP)	3.2
Gly 69	N	O_2(POP)	2.8
Lys 68	N	O_2(POP)	3.2
Arg 199	NH2	O_3(POP)	2.6
Asp193	OD1	O_3(POP)	2.7
Arg 199	NH1	O_3(POP)	2.8
Ser 103	N	O_6(POP)	2.7
Tyr 104	N	O_6(POP)	2.9

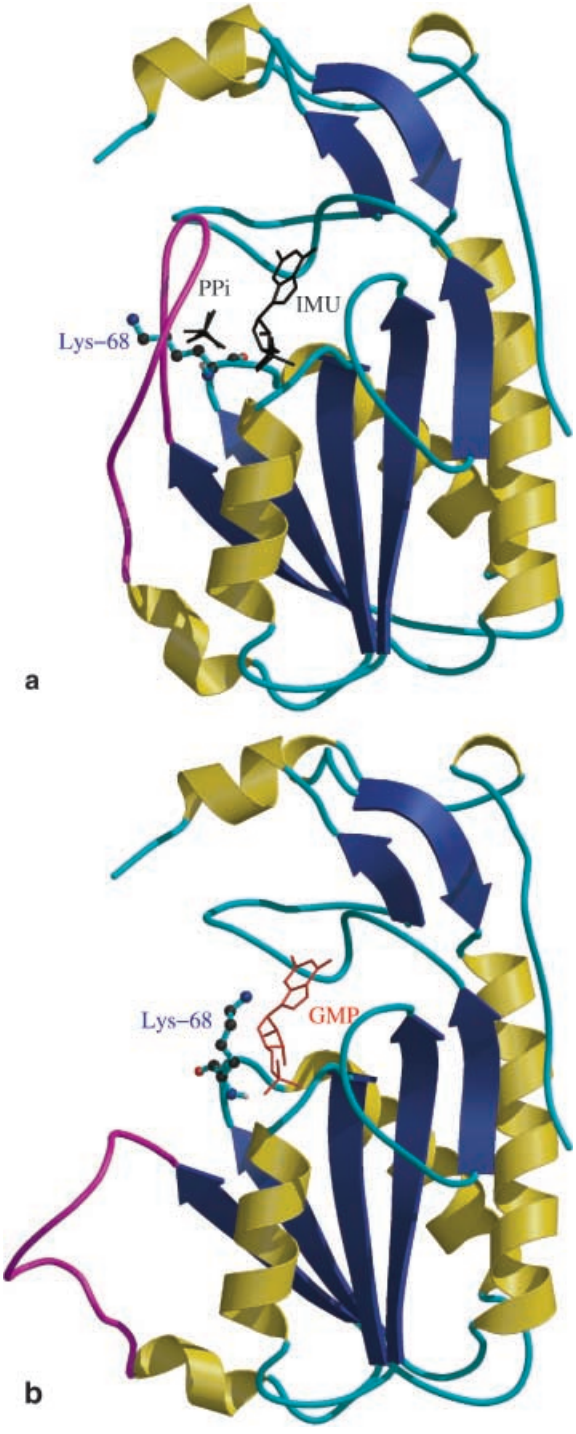


Fig. 7 **a** Human enzyme in closed conformation (1bzy). The flexible loop is shown in magenta. PPi and IMU are shown as *solid bonds* and the residue Lys-68 as a *ball-and-stick model*. **b** Human enzyme in open conformation (1hmp). The flexible loop is shown in magenta. GMP is shown as *solid bond* and the residue Lys-68 as a *ball-and-stick model*. This figure was generated using MOLSCRIPT [26] and RASTER 3D [27]

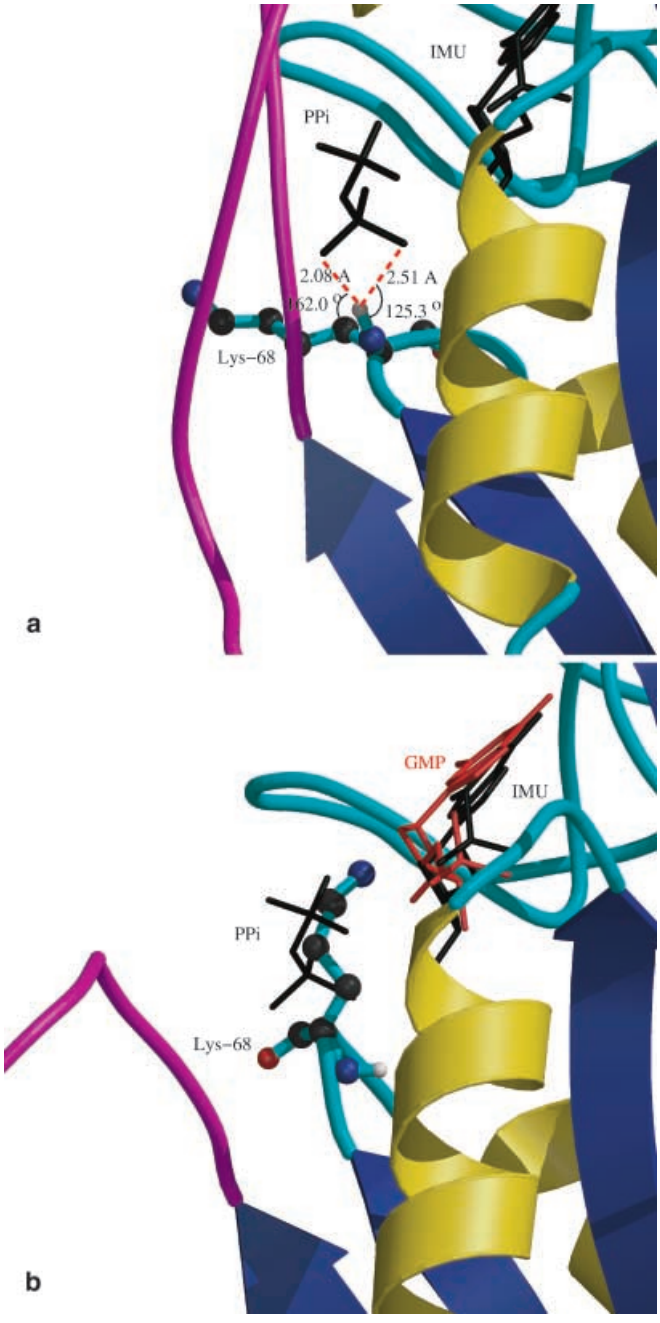


Fig. 8 **a** Interactions of Lys-68 residue of human enzyme (1hmp) with PPi in the closed conformation. The Lys-68 residue is shown as a *ball-and-stick model*, PPi and IMU are shown as *black solid bonds*. The hydrogen bonds formed by Lys-68 and PPi are indicated with *red dashed lines*. **b** Superposition of the enzyme in the open conformation (1bzy) with PPi and IMU. The Lys-68 is shown as a *ball-and-stick model*, PPi and IMU are shown as *black solid bonds* and GMP in *red*. This figure was generated with MOLSCRIPT and RASTER 3D

of Lys-68 forms a hydrogen bond with both oxygens of PPi, whereas in the open conformation (Fig. 8b), this is not possible.

The emphasis in our work, modelling of *Pf* HGPRT and the comparative analysis, was to check the differences and similarity among the HPRT family in order to know a bit more about the structural changes that happen during enzymatic activity and their influence on the rest of the residues. Thus, the modelling of *Pf* HGPRT has helped us to carry out a comparison in both conformations.

In addition to the flexible loop structural changes, other changes occur in the vicinity of the active site. These changes are due to the contributions of local conformation changes and to the rigid body shifts of some of the secondary structure elements.

Acknowledgements We wish to thank Prof. Tom Blundell and Dr. Mark Williams for critical reading of the manuscript.

References

1. Musick WD (1981) CRC Crit Rev Biochem 11:1–34
2. Eads JC, Ozturk D, Wexter JB, Grubmeyer C, Sacchettini JC (1997) Structure 5:47–58
3. Vos S, Parry R, Burns M, Jersey J, Martin JJ (1998) Mol Biol 282:875–889
4. Eads J, Scapin G, Xu Y, Grubmeyer C, Sacchettini C (1994) Cell 78:325–334
5. Schumacher M, Carter D, Ross D, Ullman B, Brennan R (1996) Nature Struct Biol 3:881–887
6. Xu Y, Eads J, Sacchettini JC, Grubmeyer C (1997) Biochem 26:3700–3712
7. Yuan L, Craig SP, Mckerrow JH, Wang CC (1992) Biochem 31:806–810
8. Somoza JR, Chin MS, Focia PJ, Wang CC, Fletterick RJ (1996) Biochem 35:7032–7040
9. Henriksen A, Aghajari N, Jensen F, Gajhede M (1996) Biochem 35:3803–3809
10. Krahn J, Kim J, Burns M, Parry R, Zalkin H, Smith J (1997) Biochem 36:11061–11068
11. Shi W, Li C, Tyler P, Furneaux R, Grubmeyer C, Schramm V, Almo S (1999) Nature Struct Biol 6:588–593
12. Heroux A, White EL, Ross D, Borhani DW (1999) Biochem 38:14485
13. Shi W, Li C, Tyler P, Furneaux R, Cahill SM, Girvin ME, Grubmeyer C, Schramm VL, Almo CS (1999) Biochem 38:9872–9880
14. Burke DF, Deane CM, Nagarajaram HA, Campillo N, Martin-Martinez M, Mendes J, Molina F, Perry J, Reddy BVB, Soares C, Sterward R, Williams M, Carrondo M, Blundell TL, Mizuguchi K (1999) Proteins 3:55–90
15. Sali A, Blundell TL (1990) JMolBiol 212:403–428
16. Zhu ZY, Sali A, Blundell TL (1992) Protein Eng 5:43–51
17. Mizuguchi K, Deane CM, Blundell TL, Johnson MS, Overington JP (1998) Bioinformatics 14:617–623
18. Thompson JD, Higgins DG, Gibson TJ (1994) Nucleic Acids Res 22:4673-4680
19. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) Nucleic Acids Res 25:4876-4882
20. Sali A, Blundell TL (1993) J Mol Biol 234:779-815
21. Laskowski RA, McArthur MW, Moss DS, Thornton JMJ (1993) Appl Crystallogr 26:283–291
22. SYBYL (1997) 65 Tripos, St Louis, Mo., USA
23. Luthy R, Bowie JU, Eisenberg D (1992) Nature 356:83–85
24. Smith J (1999) Nature Struct Biol 6:502–504
25. Focia P, Craig S III, Alicea RN, Fletterick R, Eakin E (1998) Biochem 37:5066–15075
26. Kraulis P (1991) J Appl Crystallogr 24:946–950
27. Bacon DJ, Anderson WFJ (1998) Mol Graphics 6:219–220