ORIGINAL SCIENTIFIC PAPER

CROATICA CHEMICA

Croat. Chem. Acta **2016**, *89*(2), 197–202
Published online: September 26, 2016
DOI: 10.5562/cca2872

HKD 1926

Can Crystal Symmetry and Packing Influence the Active Site Conformation of Homohexameric Purine Nucleoside Phosphorylases?

Marija Luić,[#] Zoran Štefanić*

Division of Physical Chemistry, Ruđer Bošković Institute, POB 180, HR-10002 Zagreb, Croatia
* Corresponding author's e-mail address: zoran.stefanic@irb.hr
[#] Author's e-mail address: marija.luic@irb.hr

RECEIVED: March 22, 2016 ✶ REVISED: June 17, 2016 ✶ ACCEPTED: June 26, 2016

THIS PAPER IS DEDICATED TO THE LOVING MEMORY OF IVANA WEYGAND-ĐURAŠEVIĆ (1952 – 2014)

Abstract: It is generaly believed that enzymes retain most of their functionality in the crystal form due to the large solvent content of protein crystals. This is facilitated by the fact that their natural environment in solution is not too far from the one found in the crystal form. Nevertheless, if the nature of the enzyme is such to require conformational changes, overcoming of the crystal packing constraints may prove to be too difficult. Such conformational change is present in one class of enzymes (purine nucleoside phosphorylases), that is the subject of our scientific interest for many years. The influence of crystal symmetry and crystal packing on the conformation of the active sites in the case of homohexameric purine nucleoside phosphorylases is presented and analysed.

Keywords: crystal symmetry, crystal packing, active site conformation, enzyme reaction, purine nucleoside phosphorylase.

INTRODUCTION

CATALYTICAL mechanism of homohexameric purine nucleoside phosphorylases (PNPs), key enzymes in the purine salvage pathway, is the subject of our scientific interest for many years. PNPs catalyse the phosphorolytic cleavage of the glycosidic bond of purine (2'-deoxy)nucleosides, generating the corresponding free base and (2'-deoxy)ribose-1-phosphate. The biologically active form of this enzyme is a homohexamer (Figure 1) that can be described as a trimer of dimers.

Allosteric regulation and cooperativity of phosphate and nucleoside binding to PNPs is very complex and still poorly understood process. In *Escherichia coli* PNP phosphate binding induces a segmentation of the helix located at the active site pocket border, leading to a structural change in part of the active sites. In this way, the active site conformation changes from so called open to closed one[1] (Figure 2).

In the open conformation, where the helix (H8) is continuous, the entry into the active site is widely open and

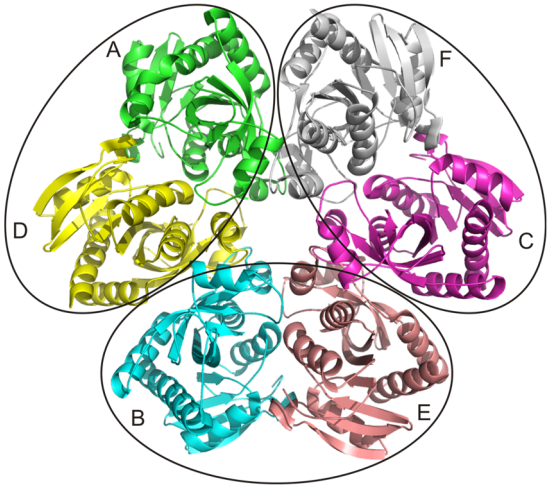

Figure 1. The hexameric structure of PNP with monomers denoted exhibits approximate 32 point group symmetry. Hexamers can be considered as trimers of dimers, denoted by ellipses. Therefore, chains A and D, B and E, and C and F form dimers.

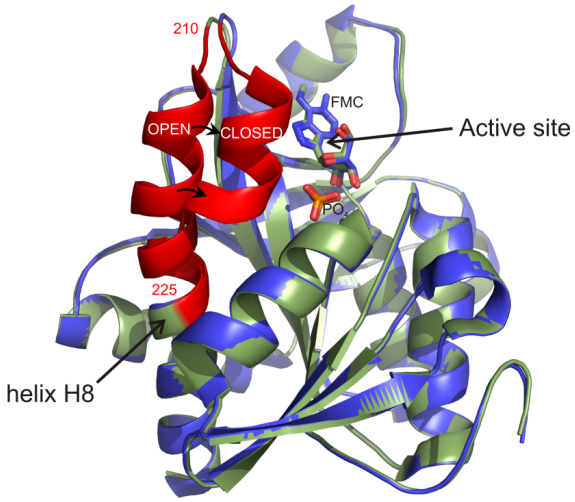

Figure 2. Superimposed open and closed active site conformations in the structure of the *E. coli* purine nucleoside phosphorylase in ternary complex (PDB code 1K9S) with phosphate ion and formycin B (FMC).

the ligands are bound only loosely. By segmentation of the H8, the entry into the active site pocket partially closes and the ligands are bound more tightly. We suppose that from two substrates, phosphate binds first while the active site is in the open conformation. Phosphate binding stabilizes Arg24 (conserved in all hexameric PNPs) and favours breaking of H8 into two segments. Before the binding of the second substrate (nucleoside), catalytically important amino acid Asp204 has to be protonated (Figure 3a).

Catalytic action occurs while the active site is closed. As a consequence of the helix segmentation, two important protein-substrate contacts are established: the H-bond between Arg24 and a main-chain carbonyl oxygen of Arg217 and guanidinium group of Arg217 moves into hydrogen bond distance to Asp204 (Figure 3b). After proton transfer from Asp204 to N7 of purine base, a transition state is formed (Figure 3c).

The two possible conformations of the active sites revealed by X-ray crystallography are in line with solution studies of *E. coli* PNP which also observe strong and weak binding sites for phosphate and nucleoside inhibitor.

In the crystal structures of *E. coli* PNP complexed with its ligands[1–14] the following distributions of the closed and open active sites can be found: 3 open + 3 closed,[1] 4 open + 2 closed sites[2–6] and all six open (see for example reference).[6] It is interesting to stress out, that in the case with 3 open + 3 closed active site conformations they alternate regularly, while in the case of 4 open and 2 closed active sites, the closed sites are always next to each other and belong to two different dimers of one homohexamer.

Recently we focused our interest on PNP from pathogen bacteria *Helicobacter pylori* (HP). Although HP PNP

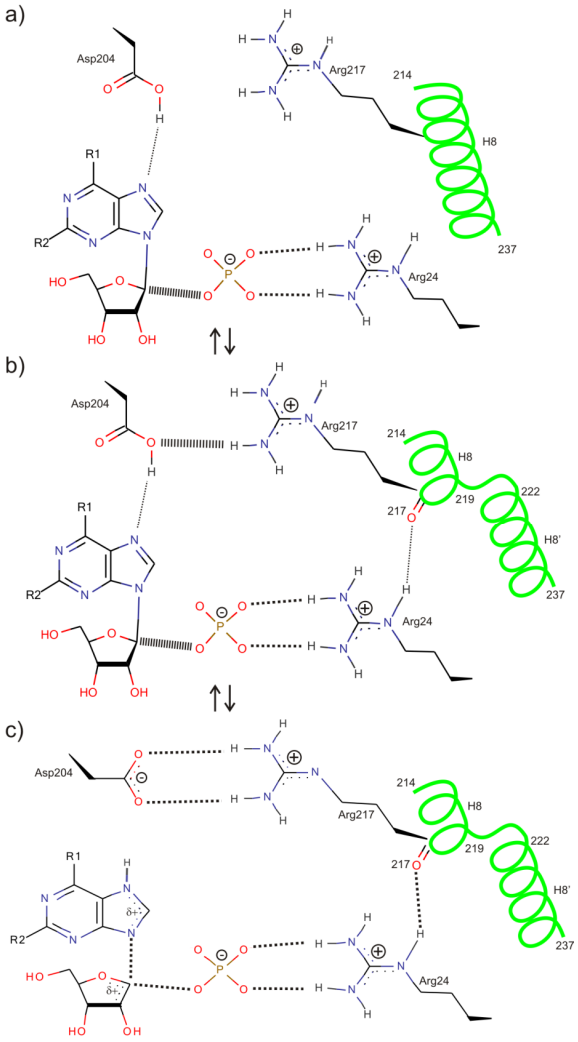

Figure 3. Possible catalytic mechanims of *E. coli* PNP. a) The active site is in the open conformation, helix H8 is continuos and Asp204 in protonated state. The neutral side-chain of Asp204 donates a hydrogen bond to the purine nitrogen N7. b) Phosphate binding stabilizes Arg24 and favors H8 segmentation – the closed conformation is formed. c) The conformational change brings the guanidinum group of Arg217 in contact with Asp204. This triggers proton transfer to the purine base and salt bridge between Arg217 and Asp204 is formed. The positively charged purine base leads to the ribooxocarbenium ion character, representing the transition state.

has 50 % identity and 70 % similarity with *E. coli* PNP, it seems that there are significant differences in their enzymatic activity. In some HP PNP ternary structures we have found unexpected distribution of 5 open and 1 closed active site conformation (data not published). To the best of our knowledge, this is the first such case among homohexameric PNP enzymes.

In order to facilitate the analysis of the distribution of open and closed active site conformations in ever-growing number of PNP crystal structures available in the Protein data bank (PDB), and to insure that all available structures will be taken into account by such an analysis, a custom made Python script was written, which automated the task of recognizing active site conformations. In addition to that, the script extracts the information about crystal contacts between specified parts of the monomers, in particular those that are in proximity of the active site. In this way we tried to identify whether a correlation between the active site conformation on one side, and crystal symmetry and/or crystal packing factors on the other side exists.

Such an analysis would contribute to better understanding of the enzyme mechanism(s) of these complicated two-substrate two-product oligomeric proteins, by identifying a possible sequence of events in ligands binding to distinct subunits of the homohexameric enzyme. At the same time, such an analysis could help us to reveal if there are limitations imposed by crystal symmetry and/or crystal packing which could mask features of the enzyme function when analysed by X-ray crystallographic methods.

EXPERIMENTAL SECTION

The set of protein structures, which are similar in sequence to the referent strain of *H. pylori* (strain 26695) purine nucleoside phosphorylase (HP PNP), was extracted from PDB by sequence alignment. The parameters for sequence alignment were chosen to yield „Significant" entries which corresponded to E value of 0.01.[15] This resulted in 172 structures out of which 149 were hexameric proteins and the rest were dimeric and were not included in further analysis.

The resulting 149 PDB structures were analysed with custom-made Python script (https://www.python.org/). The script uses excellent cctbx library[16] to process a PDB file, and automatically identify whether monomers are in open or closed active site conformation. Due to the difference in numbering and different chain lengths in different PNPs, it was not possible to automate the process of detecting the open and closed conformations by for instance measuring some distances on certain amino acid numbers. Instead, the segmentation of the helix was followed by parsing a secondary structure record and extracting the information on the helix H8 (Figure 2). Namely, this helix is segmented in two helices in the case of closed active site conformation. If another α-helix was found not further than 5 amino acids away from the beginning of H8 helix than the conformation was assigned as closed.

In addition, this script was used to identify all the crystal contacts in the crystal structure. For the crystal contacts, all distances between protein atoms (excluding waters) from different monomers shorter than 3.5 Å were counted. More specifically, the script could narrow down this list of contacts taking into account any protein region of interest. This was used to identify crystal contacts in the vicinity of the closing region of the helix H8. In order to see if there is any correlation between closed conformation of the active site and crystal contacts the following analysis was performed. For the 22 structures that contained at least one closed monomer, the number of crystal contacts of every monomer in that structure was calculated, taking into account only the region close to the part of H8 helix that closes. Due to differences in numbering of amino acids, this could not have been done by simply taking some amino acid range. To overcome this, the number of amino acids which marks the beginning of the helix H8 was identified from the secondary structure. Then the range of ten amino acids forward and five amino acids back from that amino acid was taken into account. Taking this amino acid region insured that part of the protein that undergoes conformational change was taken into account.

RESULTS AND DISCUSSION

The 149 structures retrieved from Protein data bank (PDB) by sequence alignment of HP PNP referent strain (26695) originated from 23 different bacteria. In 127 bacterial PNPs our custom made script identified no active sites in closed conformation (Supplementary materials, Table S1) while in 22 of them (from 7 different bacteria) the closed conformation of the active site was detected (Table 1).

The distribution of the open and closed active site conformations is as follows:

a) The most common distribution between open and closed active site conformations is 4+2, respectively (Table 1) and it occurs in 13 cases. It is interesting to mention that in all of them two closed sites are located close to each other and that they belong to two different dimers forming the same homohexamer. This arrangement is realized in four different space groups ($P6_122$, $P2_12_12_1$, $P6_2$ and $P2_1$) indicating that crystal packing probably does not influence such a distribution. Also, this arrangement is found almost exclusively in *E. coli* with only one exception the PNP structure from *T. vaginalis.*

b) The next by number of occurances is 0+6 arrangement with all active sites closed (6 such cases from 3 different bacteria). Four of them come from a special case of purine nucleoside phosphorylases that are highly specific for 6-aminopurines and therefore called also adenosine phsophorylses. These four cases represent ternary complexes of this enzyme from *Bacillus cereus*

Table 1. The structures of purine nucleoside phosphorylases from the Protein Data Bank that contain at least one active site in closed conformation. Structures can be classified in three groups: 4+2, 0+6 and 3+3

PDB code	Organism	Space group	Chains (closed marked with *)	Open + closed	In active site
1PK9	*Escherichia coli*	$P6_122$[a]	A B C*	4+2	2FA[b]+PO4
1PR1	*Escherichia coli*	$P6_122$	A B C*	4+2	FMB+PO4
1PR5	*Escherichia coli*	$P6_122$	A B C*	4+2	TBN+PO4
2I4T	*Trichomonas vaginalis*	$P6_2$	A B* C	4+2	UA2+PO4
1A69	*Escherichia coli*	$P6_122$	A* B C	4+2	FMB+SO4
3UT6	*Escherichia coli*	$P6_122$	A* B C	4+2	FMC+2PO4
4TS9	*Escherichia coli*	$P6_122$	A* B C	4+2	FMC+PO4
3OOE	*Escherichia coli*	$P2_12_12_1$	A* B C D E F*	4+2	PO4
4TS3	*Escherichia coli*	$P2_12_12_1$	A* B C D E F*	4+2	FMC+PO4 in closed PO4 in open
4TTA	*Escherichia coli*	$P2_12_12_1$	A* B C D E F*	4+2	SO4 in closed FMC+PO4 in 2 open PO4 in other 2 open
4TTI	*Escherichia coli*	$P2_12_12_1$	A* B C D E F*	4+2	FMC+PO4 in closed FMC+PO4 in 2 open PO4 in other 2 open
4TTJ	*Escherichia coli*	$P6_122$	A* B D	4+2	FMC+PO4
3OOH	*Escherichia coli*	$P2_1$	A* B C D* E F* G* H I J K L* M* N O P Q R*	(4+2)x3	PO4
3UAW	*Bacillus cereus*	$P6_322$	A*	0+6	ADN+SO4
3UAX	*Bacillus cereus*	$P6_322$	A*	0+6	NOS+SO4
3UAY	*Bacillus cereus*	$P6_322$	A*	0+6	ADN+SO4
3UAZ	*Bacillus cereus*	$P6_322$	A*	0+6	ADN+SO4
3OCC	*Yersinia pseudotuberculosis*	$P2_12_12_1$	A* B* C* D* E* F*	0+6	DIH+PO4
3OF3	*Vibrio cholerae*	$P2_12_12_1$	A* B* C* D* E* F* G* H* I* J* K* L*	(0+6)x2	DIH+PO4
4D98	*Bacillus subtilis*	$R32$	A B*	3+3	SO4
4M7W	*Leptotrichia buccalis*	$P32$	A B* C	(3+3)+(6+0)	PO4
1K9S	*Escherichia coli*	$P4_12_12$	A* B* C* D E F	3+3	FM2+PO4

[a] In the space groups $P6_122$, $P6_2$ and $P32$ one half of the hexamer is present in the asymmetric unit. In the space group $R32$ two monomers and in the $P6_322$ one monomer is present in the asymmetric unit. The whole hexamer in these space groups is formed by applying crystallographic symmetry.

[b] The abbreviations denote standard PDB ligand codes.

that, comparing with *E. coli* PNP, have more closed the critical part of the active site (residues 207–217). The remainig two structures (PDB codes 3OCC and 3OF3) are the result of Protein structure initiative and are not published yet.

c) In three crystal structures the distribution between closed and open active sites is 3+3. One is of the ternary complex of homohexameric PNP from *Echerichia coli* with formycin A derivatives and phosphate or sulphate ions.[1] This is a very special case of PNP's ternary complex. 6-methyl formycin A is the best know inhibitor of *E. coli* PNP with K_i = 0.3 μM at neutral pH. In aqueous solution this compound undergoes rearrangement to four different formycin derivatives. The main product of hydrolysis is N7-methlyformycin A. Detailed inspection of the active sites of the homohexamer present in the asymmetric unit revealed unambiguously that 6-methylformycin was located in three monomers (closed conformation) and in other three N7-methylformycin (open conformations). Open and closed conformations in the homohexamer alternate

open		closed		open		closed		open		closed	
3ooh				**3of3**				**3ooe**			
B	1	A	9			A	0	B	21	A	4
C	2	D	1			B	0	C	0	F	4
E	3	F	3			C	0	D	0		
H	1	G	0			D	1	E	0		
I	1	L	1			E	0	**2i4t**			
J	0	M	0			F	1	A	1	B	5
K	0	R	9			G	0	C	0		
N	4					H	0	**4tta**			
O	3					I	1	B	6	A	8
P	3					J	1	C	0	F	4
Q	7					K	0	D	0		
4tti						L	1	E	10		
B	15	A	8	**3ut6**				**1k9s**			
C	0	F	3	B	0	A	0	D	3	A	2
D	0			C	19			E	0	B	2
E	6			**4ttj**				F	0	C	1
4ts3				B	0	A	0	**3uaw**			
B	7	A	5	D	14					A	1
C	3	F	4	**4m7w**				**1pr5**			
D	0			A	2	B	0	A	0	C	4
E	1			C	6			B	13		
3uaz								**3uax**			
		A	8							A	3

open		closed	
3ooe			
B	0	A	0
C	0	F	1
D	0		
E	0		
1pk9			
A	0	C	0
B	20		
1pr1			
A	0	C	0
B	18		
4d98			
A	14	B	7
3uay			
		A	1
4ts9			
B	0	A	0
C	17		
1a69			
B	0	A	0
C	16		

Figure 4. This figure displays the number of crystal contacts that particular monomers have in proximity of the active site in 22 PNP structures that have at least one active site in closed conformation. Each bordered cell in the table contains the data for one PDB code given in the top shaded area. Letters indicate the chain identifiers. In each bordered cell the two left columns represent results for the monomers (if any) in open and two right columns for the monomers in closed conformation. The red bars are graphical help to compare values accross the cells.

regularly. In all six active sites one phosphate or sulphate ion (it is not possible to distinguish between them in the electron density maps calculated at the given resolution) is bound. Such a distribution was in agreement with half-the-sites binding for 6-methylformycin determined in solution.[1] The same alternating arrangement is found in the crystal structure of *Bacillus subtilis* (PDB id 4D98), the only structure with one dimer in the asymmetric unit.[17] One other very interesting case in this group is the structure of PNP from *Leptotrichia buccalis* (PDB code 4M7W). Namely, in the crystal packing of this protein there is an equal number of hexamers with 3+3 and 6+0 arrangement.

There is a substantial variability of the space groups found among the members belonging to the same distribution of active sites (for example, 4+2 distribution is found in four different space groups, 0+6 distribution is found in two space groups and one of the space groups has two variations). Furthermore, the same space group alone does not impose similar packing and therefore similar crystal contacts. For this to be true unit cell axis need to be similar too. Therefore, the appearance of the same open and closed site distribution (such as 4+2) in different crystal symmetries implies, that this distribution is not influenced by crystal packing. Although the number of structures with closed active sites is arguably not high enough to make good statistics, this set of structures does not display any regularity which would indicate any correlation between open and closed active site conformations and crystallographic symmetry.

Very recently we have determined several crystal structures of PNP from the pathogen bacteria *H. pylori* (data not published). To our surprise, distribution with five open and one closed conformation was found, which is not present in any structure available in PDB.

The results of detailed analysis of crystal contacts in the crystal structures which contained at least one active site in closed conformation (Table 1) are summarized in Figure 4.

It turns out that active sites in closed conformation have, in most of the cases **less** crystal contacts than those in open conformation. There are quite a few open active sites that have from 15–21 crystal contacts in this region,

while the number of contacts in closed active sites is never higher than 9. This could perhaps be explained by the fact that in open active sites helix H8 is slightly more distant from the central region, and therefore more available for crystal contacts. In any case, it does not support our first hypothesis that closing of the active sites could be influenced by neighbouring homohexamers in the crystal packing (somehow pressing them and thereby forcing them to close).

Therefore, based on our analyses presented here, the straightforward influence of the crystal symmetry and/or crystal contacts on the conformation of the active sites in homohexameric purine nucleoside phosphorylases cannot be inferred.

Acknowledgment. This work has been fully supported by the Croatian Science Foundation under the project number 7423.

Supplementary Information. Supporting information to the paper is enclosed to the electronic version of the article at: http://dx.doi.org/10.5562/cca2872.

REFERENCES

[1] G. Koellner, A. Bzowska, B. Wielgus-Kutrowska, M. Luić, T. Steiner, W. Saenger, J. Stępiński, *J. Mol. Biol.* **2002**, *315*, 351.

[2] G. Koellner, M. Luić, D. Shugar, W. Saenger, A. Bzowska, *J. Mol. Biol.* **1998**, *280*, 153.

[3] E. M. Bennett, C. Li, P. W. Allan, W. B. Parker, S. E. Ea-lick, *J. Biol. Chem.* **2003**, *278*, 47110.

[4] G. Mikleušević, Z. Štefanić, M. Narczyk, B. Wielgus-Kutrowska, A. Bzowska, M. Luić, *Biochimie* **2011**, *93*, 1610.

[5] Z. Štefanić, M. Narczyk, G. Mikleušević, B. Wielgus-Kutrowska, A. Bzowska, M. Luić, FEBS Letters **2012**, *586*, 967.

[6] Z. Štefanić, G. Mikleušević, M. Narczyk, B. Wielgus-Kutrowska, A. Bzowska, M. Luić, *Croat. Chem. Acta* **2013**, *86*, 117.

[7] D. Paul, S. E. O'Leary, K. Rajashankar, W. Bu, A. Toms, E. C. Settembre., J. M. Sanders, T. P. Begle, S. E. Ealick, *Biochemistry* **2010**, *49*, 3499.

[8] W. Bu, E. C. Settembre, M. H. el Kouni, S. E. Ealick, *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2005**, *61*, 863.

[9] T. T. Caradoc-Davies, S. M. Cutfield, I. L. Lamont, J. F. Cutfield, *J. Mol. Biol.* **2004**, *337*, 337.

[10] E. M. Bennett, R. Anand, P. W. Allan, A. E. Hassan, J. S. Hong, D. N. Levasseur, D. T. McPherson, W. B. Parker, J. A. Secrist, E. J. Sorscher, T. M. Townes, W. R. Waud, S. E. Ealick, *Chem. Biol.* **2003**, *10*, 1173.

[11] F. T. Burling, R. Kniewel, J. A. Buglino, T. Chadha, A. Beckwith, C. D. Lima, *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2003**, *59*, 73.

[12] C. Mao, W. J. Cook, M. Zhou, G. W. Koszalka, T. A. Krenitsky, S. E. Ealick, *Structure* **1997**, *5*, 1373.

[13] E. Y. Morgunova, A. M. Mikhailov, A. N. Popov, E. V. Blagova, E. A. Smirnova, B. K. Vainshtein, C. Mao, Sh. R. Armstrong, S. E. Ealick, A. A. Komissarov, *FEBS Lett.* **1995**, *367*, 183.

[14] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, *J. Mol. Biol.* **1990**, *215*, 403.

[15] R. W. Grosse-Kunstleve, N. K. Sauter, N. W. Moriarty, P. D. Adams, *J. Appl. Cryst.* **2002**, *35*, 126.

[16] P. O. de Giuseppe, N. H. Martins, A. N. Meza, C. R. Dos Santos, H. D. Pereira, M. T. Murakami, *Plos One* **2012**, *7*, e44282.

Supplementary materials

Table S1. The structures of purine nucleoside phosphorylases that contain no closed chains in the Protein Data Bank.

PDB id	Organism	Space group	Chains
3uav	*Bacillus Cereus*	P 63 2 2	A
3uaz	*Bacillus Cereus*	P 63 2 2	A
4da8	*Bacillus Subtilis*	P 63 2 2	A
4dab	*Bacillus Subtilis*	P 63 2 2	A
4dae	*Bacillus Subtilis*	P 63 2 2	A
4d8x	*Bacillus Subtilis*	P 63 2 2	A
4da6	*Bacillus Subtilis*	P 63 2 2	A
4da0	*Bacillus Subtilis*	P 63 2 2	A
4dar	*Bacillus Subtilis*	P 63 2 2	A
4da7	*Bacillus Subtilis*	P 63 2 2	A
1sq6	*Plasmodium Falciparum*	R 3 2 :H	A
2bsx	*Plasmodium Falciparum*	P 3 2 1	A
2b94	*Plasmodium Knowlesi*	R 3 2 :H	A
3emv	*Plasmodium Vivax*	R 3 2 :H	A
4lkr	*Shewanella Oneidensis*	P 63 2 2	A
4tym	*Streptococcus Agalactiae Lmg 15084*	P 63 2 2	A
1z34	*Trichomonas Vaginalis*	P 41 3 2	A
1z39	*Trichomonas Vaginalis*	P 41 3 2	A
1z38	*Trichomonas Vaginalis*	P 41 3 2	A
1z37	*Trichomonas Vaginalis*	P 41 3 2	A
1z36	*Trichomonas Vaginalis*	P 41 3 2	A
1z33	*Trichomonas Vaginalis*	P 41 3 2	A
1z35	*Trichomonas Vaginalis*	P 41 3 2	A
4mci	*Vibrio Fischeri*	R 3 2 :H	A
4ldn	*Vibrio Fischeri*	R 3 2 :H	A
4mch	*Vibrio Fischeri*	R 3 2 :H	A
2ac7	*Bacillus Cereus*	P 63	A B
4dan	*Bacillus Subtilis*	P 3 2 1	A B
4dao	*Bacillus Subtilis*	P 3 2 1	A B
4d9h	*Bacillus Subtilis*	P 3 2 1	A B
4d8v	*Bacillus Subtilis*	P 3 2 1	A B
1rxy	*Escherichia Coli*	R 3 :H	A B
1tgy	*Escherichia Coli*	R 3 :H	A B
1lx7	*Escherichia Coli*	R 3 :H	A B
1tgv	*Escherichia Coli*	R 3 :H	A B
1t0u	*Escherichia Coli*	R 3 :H	A B
3fow	*Plasmodium Falciparum*	I 41 3 2	A B
2iq5	*Salmonella Typhimurium*	R 3 :H	A B
2hsw	*Salmonella Typhimurium*	R 3 :H	A B
3mb8	*Toxoplasma Gondii*	P 6	A B
4ny1	*Yersinia Pseudotuberculosis*	R 3 :H	A B
4of4	*Yersinia Pseudotuberculosis*	R 3 :H	A B
1ovg	*Escherichia Coli*	P 61 2 2	A B C

1oum	*Escherichia Coli*	P 61 2 2	A B C
3onv	*Escherichia Coli*	P 61 2 2	A B C
1otx	*Escherichia Coli*	P 61 2 2	A B C
1ov6	*Escherichia Coli*	P 61 2 2	A B C
1oty	*Escherichia Coli*	P 61 2 2	A B C
1ou4	*Escherichia Coli*	P 61 2 2	A B C
1pk7	*Escherichia Coli, Escherichia Coli O157:H7*	P 61 2 2	A B C
1pw7	*Escherichia Coli, Escherichia Coli O157:H7*	P 61 2 2	A B C
1pr0	*Escherichia Coli, Escherichia Coli O157:H7*	P 61 2 2	A B C
1pr6	*Escherichia Coli, Escherichia Coli O157:H7*	P 61 2 2	A B C
1pr2	*Escherichia Coli, Escherichia Coli O157:H7*	P 61 2 2	A B C
1pr4	*Escherichia Coli, Escherichia Coli O157:H7*	P 61 2 2	A B C
1pke	*Escherichia Coli, Escherichia Coli O157:H7*	P 61 2 2	A B C
4mar	*Meiothermus Ruber*	C 2 2 21	A B C
4m3n	*Meiothermus Ruber Dsm 1279*	C 2 2 21	A B C
1je0	*Sulfolobus Solfataricus*	C 2 2 21	A B C
1jdt	*Sulfolobus Solfataricus*	C 2 2 21	A B C
1jdu	*Sulfolobus Solfataricus*	C 2 2 21	A B C
1jdz	*Sulfolobus Solfataricus*	C 2 2 21	A B C
1jp7	*Sulfolobus Solfataricus*	C 2 2 21	A B C
1jpv	*Sulfolobus Solfataricus*	C 2 2 21	A B C
1k3f	*Escherichia Coli*	P 1 21 1	A B C D E F
4ogk	*Salmonella Enterica Subsp. Enterica Serovar Typhimurium*	P 61	A B C D E F
2hrd	*Salmonella Typhimurium*	P 21 21 21	A B C D E F
2i8a	*Salmonella Typhimurium*	P 21 21 21	A B C D E F
1sj9	*Salmonella Typhimurium*	P 61	A B C D E F
2rj3	*Salmonella Typhimurium*	P 21 21 21	A B C D E F
4oeh	*Vibrio Cholerae O1 Biovar El Tor*	P 1	A B C D E F
4r31	*Actinobacillus Succinogenes*	P 21 21 21	A B C D E F
1xe3	*Bacillus Anthracis*	P 21 21 21	A B C D E F
4d8y	*Bacillus Subtilis*	P 21 21 21	A B C D E F
1u1c	*Escherichia Coli*	P 21 21 21	A B C D E F
1ecp	*Escherichia Coli*	P 1 21 1	A B C D E F
1u1f	*Escherichia Coli*	P 21 21 21	A B C D E F
1u1e	*Escherichia Coli*	P 21 21 21	A B C D E F
3kvv	*Escherichia Coli*	P 32	A B C D E F
4rj2	*Escherichia Coli*	P 1 21 1	A B C D E F
1u1g	*Escherichia Coli*	P 21 21 21	A B C D E F
1u1d	*Escherichia Coli*	P 21 21 21	A B C D E F
3enz	*Plasmodium Falciparum*	I 41 2 2	A B C D E F
1q1g	*Plasmodium Falciparum*	P 21 21 21	A B C D E F
3phc	*Plasmodium Falciparum*	P 1	A B C D E F
1y1q	*Salmonella Typhimurium*	P 21 21 21	A B C D E F
3fwp	*Salmonella Typhimurium*	P 21 21 21	A B C D E F
2oec	*Salmonella Typhimurium*	P 21 21 21	A B C D E F
2pga	*Salmonella Typhimurium*	P 21 21 21	A B C D E F
1zl2	*Salmonella Typhimurium*	P 21 21 21	A B C D E F
2hwu	*Salmonella Typhimurium*	P 21 21 21	A B C D E F
3ddo	*Salmonella Typhimurium*	P 21 21 21	A B C D E F

2qdk	*Salmonella Typhimurium*	P 21 21 21	A B C D E F
3c74	*Salmonella Typhimurium*	P 21 21 21	A B C D E F
1y1r	*Salmonella Typhimurium*	P 21 21 21	A B C D E F
2hn9	*Salmonella Typhimurium*	P 21 21 21	A B C D E F
1y1s	*Salmonella Typhimurium Lt2*	P 21 21 21	A B C D E F
4r2x	*Shewanella Oneidensis*	P 1 21 1	A B C D E F
4yjk	*Shewanella Oneidensis (Strain Mr-1)*	P 1 21 1	A B C D E F
4r2w	*Shewanella Oneidensis Mr-1*	P 1 21 1	A B C D E F
1jds	*Sulfolobus Solfataricus*	P 1 21 1	A B C D E F
1je1	*Sulfolobus Solfataricus*	P 1 21 1	A B C D E F
1jdv	*Sulfolobus Solfataricus*	P 1 21 1	A B C D E F
2isc	*Trichomonas Vaginalis*	P 21 21 21	A B C D E F
4u2k	*Vibrio Cholerae*	P 31	A B C D E F
4h1t	*Vibrio Cholerae*	P 1	A B C D E F
4g8j	*Vibrio Cholerae*	P 1 21 1	A B C D E F
4k6o	*Vibrio Cholerae*	P 1 21 1	A B C D E F
4lzw	*Vibrio Cholerae*	P 1 21 1	A B C D E F
4ip0	*Vibrio Cholerae*	P 1	A B C D E F
1vhj	*Vibrio Cholerae*	P 1 21 1	A B C D E F
1vhw	*Vibrio Cholerae*	P 1 21 1	A B C D E F
4ogl	*Vibrio Cholerae O1 Biovar El Tor*	P 1 21 1	A B C D E F
4jp5	*Yersinia Pestis*	P 32	A B C D E F
4i2v	*Yersinia Pseudotuberculosis*	P 3	A B C D E F
4e1v	*Salmonella Enterica Subsp. Enterica Serovar Typhimurium*	C 1 2 1	A B C D E F G H I
1rxc	*Escherichia Coli*	P 1 21 1	A B C D E F G H I J K L
3opv	*Escherichia Coli*	C 1 2 1	A B C D E F G H I J K L
1rxu	*Escherichia Coli*	P 1 21 1	A B C D E F G H I J K L M N O P Q R
3qpb	*Streptococcus Pyogenes Serotype M6*	P 1	A B C D E F G H I J K L M N O P Q R
1rxs	*Escherichia Coli*	P 1 21 1	A B C D E F G H I J K L M N O P Q R a b c d e h i j k l m o
2oxf	*Salmonella Typhimurium*	R 3 :H	A F
3dps	*Salmonella Typhimurium*	R 3 :H	A F
1y1t	*Salmonella Typhimurium Lt2*	R 3 :H	A F
1odk	*Thermus Thermophilus*	P 43 21 2	A? B? C? D? E? F?
1odi	*Thermus Thermophilus*	P 43 21 2	A? B? C? D? E? F?
1odl	*Thermus Thermophilus*	P 43 21 2	A? B? C? D? E? F?
1odj	*Thermus Thermophilus*	P 43 21 2	A? B? C? D? E? F?

Python script that was used in crystal contacts calculations

```
#!/bin/env cctbx.python_dev-546
import sys
import pickle
import iotbx.pdb
import glob
from mmtbx.lattice import *
from scitbx.array_family import flex

organisms = pickle.load(open('organisms.pkl','rb'))

def open_or_closed(f):
    inpdb = iotbx.pdb.input(f)
    h1 = inpdb.construct_hierarchy()
    chains = [c.id for c in h1.chains()]
    res = ""
    for chain in sorted(set(chains)):
        res += chain_open_closed(inpdb,h1,chain) + " "
    return res

def open_and_closed_chains(f):
    inpdb = iotbx.pdb.input(f)
    h1 = inpdb.construct_hierarchy()
    chains = [c.id for c in h1.chains()]
    opn, clsd= [],[]
    for chain in sorted(set(chains)):
        if '*' in chain_open_closed(inpdb,h1,chain):
            clsd.append(chain)
        else:
            opn.append(chain)
    return opn, clsd

def average_chain_length(f):
    from numpy import array
    inpdb = iotbx.pdb.input(f)
    h1 = inpdb.construct_hierarchy()
    chains = list(h1.chains())
    l = [len(chain.as_sequence()) for chain in chains]
    l = array(l)
    return int(l[l>5].mean())

def chain_open_closed(inpdb,h,chain):
    """This tries to extract the secondary structure and see if the helix is open or closed"""
    try:
        secstr = inpdb.extract_secondary_structure()
    except:
        return chain + '?'
    h8 = filter(lambda h: h.start_chain_id == chain, secstr.helices)[-1]
    h7 = filter(lambda h: h.start_chain_id == chain, secstr.helices)[-2]
    start_h8 = h8.get_start_resseq_as_int()
    end_h7 = h7.get_end_resseq_as_int()
    if start_h8 - end_h7 < 5:
        oc = "*"
    else:
        oc = ""
    return chain+oc

def get_resseq(f,chain):
    """Calculates resseq around helix is open or closed"""
    return 210,225
    inpdb = iotbx.pdb.input(f)
    try:
        secstr = inpdb.extract_secondary_structure()
    except:
```

```
        return 210,225
    h8 = filter(lambda h: h.start_chain_id == chain, secstr.helices)[-1]
    start_h8 = h8.get_start_resseq_as_int()
    return start_h8 - 10, start_h8 + 5

def crystal_contacts_from_selection(pdb,selection="all",distance_cutoff=3.5):
    inpdb = iotbx.pdb.input(pdb)
    h = inpdb.construct_hierarchy()
    x = inpdb.xray_structure_simple()
    #x.show_summary()
    atoms = list(h.atoms_with_labels())

    sc = h.atom_selection_cache()
    selected = sc.iselection(selection)

    sites = x.sites_frac()
    unit_cell = x.unit_cell()

    pat = x.pair_asu_table(distance_cutoff)
    pst = pat.extract_pair_sym_table()
    contacts = []
    selected_sites = list(selected)

    pairs = {}
    for i, psd in enumerate(pst):
        for j in psd.keys():
            pairs[(i,j)] = psd[j]

    selected_pairs = sorted(filter(lambda p: p[0] in selected_sites or p[1] in selected_sites,
pairs.keys()))

    for pair in selected_pairs: # (i_seq, j_seq)
        i_seq, j_seq = pair
        site_i = sites[i_seq]
        atom_i = atoms[i_seq]
        chain_i = atom_i.parent().parent().parent()
        res_i = atom_i.parent()
        site_j = sites[j_seq]
        atom_j = atoms[j_seq]
        chain_j = atom_j.parent().parent().parent()
        res_j = atom_j.parent()
        sym_op = pairs[pair][0]
        if res_j.resname == "HOH" or res_i.resname == "HOH" :
            continue
        if not sym_op.is_unit_mx():
            site_ji = sym_op * site_j
            distance = unit_cell.distance(site_i, site_ji)
            contacts.append((i_seq, j_seq, sym_op, distance))
            #print "%s %s %-30s %8.4f" % (res_i.id_str(), res_j.id_str(), sym_op, distance)
            #print "%s %s %-30s %8.4f" % (chain_i.id, chain_j.id, sym_op, distance)

    #print 'Number of contacts: %d' % len(contacts)
    return contacts

def get_chains(pdb):
    h = iotbx.pdb.input(pdb).construct_hierarchy()
    return set(sorted([c.id for c in list(h.chains())]))

def get_xray_structures():
    pdbs = glob.glob('*.pdb')
    xray_structures = {}
    for pdb in pdbs:
        xray_structures[pdb] = iotbx.pdb.input(pdb).xray_structure_simple()
    return xray_structures


def doit(f):
```

```
    oc = open_and_closed_chains(f)
    print f
    print 'Open'
    for c in oc[0]:
        start,stop = get_resseq(f,c)
        ncont = crystal_contacts_from_selection(f,'chain %s and resseq %d:%d' %(c,start,stop))
        print c, len(ncont)
    print 'Closed'
    for c in oc[1]:
        start,stop = get_resseq(f,c)
        ncont = crystal_contacts_from_selection(f,'chain %s and resseq %d:%d' %(c,start,stop))
        print c, len(ncont)
    print '-'*80

def print_table():
    xray_structures = get_xray_structures()
    def similars(x):
        res = []
        for k, v in xray_structures.items():
            if x.is_similar_symmetry(v):
                res.append(k)
        return res
    for k, v in xray_structures.items():
        #print '%s\t%-50s\t%-20s\t%-20s\t%20s%5s\t%s' %(k, v.unit_cell(),
v.space_group().info(),
v.space_group().crystal_system(),open_or_closed(k),average_chain_length(k),similars(v))
        #print '%s\t%-50s\t%-10s\t%30s\t%s' %(k[:4], organisms[k.upper()[:4]],
v.space_group().info(),open_or_closed(k),similars(v))
        print '%s\t%-50s\t%-10s\t%-30s\t%5.2f' %(k[:4], organisms[k.upper()[:4]],
v.space_group().info(),open_or_closed(k),len(crystal_contacts_from_selection(k))/len(get_chain
s(k)))

def run1():
    f = sys.argv[1]
    print f
    selection = sys.argv[2]
    contacts = crystal_contacts_from_selection(f,selection)
    if not contacts:
        print 'No contacts'
    print open_or_closed(f)
    print "-"*80

if __name__ == "__main__":
    run1()
    #print_table()
```