

## Handling of Selenomethionines in Macromolecular Refinement

Jarosław Blaszczyk\*

Department of Heteroorganic Chemistry, Sienkiewicza 112, 90-363 Łódź, Poland

\*Corresponding author: Jarosław Blaszczyk, Centre of Molecular and Macromolecular Studies, Polish Academy of Sciences, Department of Heteroorganic Chemistry, Sienkiewicza 112, 90-363 Łódź, Poland, Tel: (48)(42) 6803-216; (48)(42) 6803-261; E-mail: blaszcz8@cbmm.lodz.pl

Received date: October 06, 2014, Accepted date: October 29, 2014, Published date: November 5, 2014

Copyright: © 2014 Blaszczyk J, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

### Abstract

In the Protein Data Bank, we can find X-ray structures which were determined from selenium-derivatized proteins, but the deposited coordinates, to our surprise, contain native methionines. The problem is likely due to not complete substitution of native methionines with selenomethionines during chemical reaction. When the crystal grows from such sample, and later is subjected for X-ray analysis, it may show partial presence of native sulfur.

This paper is a voice in discussion how to handle the refinement of selenomethionines. The purpose of this work is to explain why selenomethionines can't be refined purely as native methionines, particularly in macromolecular structures determined at high resolution. It is important because selenium and sulfur are different chemical elements. To explain the importance of refinement of correct atom type, author provides evidence, collected mainly from his previously reported small organic structures containing sulfur or selenium. Author emphasizes that the X-ray structures, which contain selenium, are never identical with their sulfur-containing analogs. They differ at least in the respective bond lengths. Two available publicly X-ray structures have been re-refined, one small molecule and one macromolecule, in which the atom type was intentionally changed. Refinement of chemically incorrect atom type yielded non-natural behavior of atomic displacement parameters. Author concludes that the structures which contain incorrect atom types can't represent the experimental data but can serve only as models. There are examples, in the PDB, which show that selenomethionines in not fully Se-Met derivatized proteins can be refined with arbitrarily assumed fractional, less than 100%, occupancies. This inspired the author to perform various re-refinements, of the above mentioned Se-Met structure that contained S atoms in coordinates. In these re-refinements, selenomethionines are handled in different manners, and the results are compared. Author suggests refinement of chemical identities, selenium and sulfur, with fractional occupancies, as a method.

**Keywords:** Methionine; Selenomethionine; Sulfur; Selenium; Chemical identity; X-ray crystal structure; Theoretical model; Macromolecular crystallography; Small molecule crystallography; Structure solution and refinement.

### Abbreviations:

PDB: Protein Data Bank; JCSG: Joint Center for Structural Genomics; S-Met: Native Methionine; Se-Met: Selenomethionine; S-Cys: Native Cysteine; Se-Cys: Selenocysteine; B-factor: Atomic Displacement Parameter

### Introduction

The worldwide Protein Data Bank (wwPDB) is an archive of the macromolecular structures, which have been determined by the experimental methods, such as X-ray, NMR, electron microscopy or hybrid methods [1,2]. Since 2006, the Protein Data Bank does not accept theoretical models. The models, which were deposited before 2006, still exist in the PDB, but they are separated from the main archive [3]. In May 2014, the total number of depositions reached 100000, and is growing continuously.

The advances in macromolecular crystallography methodologies allowed to increase significantly the number of structures determined at atomic resolution. The beauty of atomic resolution is that various structural details may be built and refined with much higher confidence. This includes the clear appearance, in electron density, of

multiple conformations of amino acid side chains or entire chain fragments [4]. Another advantage is that high resolution allows determining the chemical identity of the terminal side chain atoms (for example, N versus O atoms in Asn or Gln, or C versus N atoms in His residues). In high-resolution structures, the atom type in these side chains can be determined in similar manner like in small-molecule crystallography, i.e., by simple inspection of the atomic displacement parameters (B-factors) and interatomic distances, without time-consuming and uncertain analyses of the entire systems of intra- or intermolecular contacts. The determination of the atom type, by using the small-molecular approach, has been accomplished, for example, in the PDB structures 1EQO, 1F9H, 1F9Y, 1HQ2, and 2O90 [5,6].

However, even if the methods of crystallization, data collection, computing and visualization become more and more advanced the refinement of protein structures still remains a challenge. The inconveniences include, for example, troubles in handling of refinement of flexible fragments of the molecule. This happens, when the side chains, loops or entire domains lie at the surface of the protein and are solvent-exposed. Such fragments are frequently omitted in the final coordinate file, even if the refinement is performed at atomic resolution. For example, in 1.07 Å entry PDB 2O90, seven C-terminal residues in the final structure of dihydronoopterin aldolase are missing due to lack of electron density [6]. Interestingly, the entire polypeptides are completely visible, for this protein, in two other PDB entries, 2NM2 and 2NM3, even when these structures were

determined at resolution not that high, of 1.50 and 1.68 Å, respectively [7].

Other refinement problems may arise during crystallization. The protein crystals usually grow from the 'soup' which contains many different components such as salts, additives, metal ions, etc. Some of extra entities can also bind to the protein yet at the protein preparation steps such as purification. Very frequently, such extra entities are clearly visible in the electron density, but are very difficult to interpret. An inspection of the X-ray crystal structures, which are deposited in the PDB, reveals many coordinate files, which contain such elements. Unidentified elements are labeled in the PDB entries in three ways: as UNX, UNL, and UNK. UNX is an abbreviation for unknown atom or ion, and UNL stands for unknown ligand. The UNX or UNL elements are present, for example, in PDB entries 4IGV [8], 4IAJ [9], or 3S8S [10]. UNK is an abbreviation for unknown amino acid residue. For example, in PDB entries 4I79 [11] or 4GS0 [12], authors were able to identify only a fragment of the polypeptide, and they are uncertain of the sequence registration.

Next issue is how to handle, during refinement and deposition, the protein residues, which are only partially visible in experimental electron density. Because of the software limitations, the crystallographers usually can't refine the amino acid residues with missing atoms. Instead, they use tricks, which consist in renaming the visible residue fragment to the residue, which in such shape is complete. For example, the residues of longer side chains, such as lysine, glutamine, cysteine, methionine, etc., for which the authors only see the density up to C $\beta$  atom, are usually renamed to alanines, and further refined as alanines.

The PDB policies and deposition procedures do not allow accepting depositions, which contain differences in the amino acid sequence information and in atomic coordinates. Therefore, the longer residues, which were not completely visible in density and are reported as alanines, during PDB deposition and annotation process, have their original names restored, and the information about missing atoms is included in the text of the processed entry. The remark 470 in the PDB entry 4NZK, is an example of such deposition [13].

In the PDB, there are many entries (7839 entries available in September 2014), which contain atomic coordinates obtained from refinement of the data collected from selenium-containing protein. The main purpose of protein derivatization by substitution of native methionines (S-Met) with relatively 'heavier' selenomethionines (Se-Met; 34Se versus 16S) is to facilitate the solution of the phase problem. Selenium can be used for phase determination in multiwavelength anomalous diffraction (MAD) method of crystal structure solving [14]. Incorporation of selenium is the way of producing modified proteins without the structural disturbances that are commonly associated with heavy-atom incorporation. It removes the need for time-consuming and challenging screening for heavy-atom derivatives [15].

Another amino acid residue which contains selenium, is a selenocysteine (Se-Cys). Selenocysteines are naturally present in selenoproteins [16]. Protein derivatization, by replacement of cysteines with selenocysteines, for purpose of structure solution by X-ray crystallography, is not such common approach [17] as the replacement of S-Met with Se-Met. In September 2014, there were only 34 entries available in the PDB which contained Se-Cys residues.

A detailed inspection of the structural information in the PDB archive may reveal the entries, which still (despite the PDB deposition policies) have serious inconsistencies in chemical identity between

author's provided sequence information and the coordinates. The example of such inconsistency is the PDB entry 3B40 [18], which contains the following statement: 'SeMet modeled as Met' (see remark 3, other refinement remarks in the PDB file, or refine. details in cif file). In other words, the PDB entry 3B40 is an X-ray structure of selenomethionine-derivatized protein. Surprisingly, there is no presence of Se-Met residues in the sequence information and in coordinate section. Instead, the sequence and coordinates contain S-Met residues.

This report is an attempt to answer, whether the X-ray structure, which contains different atom type in coordinates (S-Met) than in the crystal (Se-Met), still could be considered as true structure, or only as representative model of investigated macromolecule.

It is widely reported, that the replacement of methionines with selenomethionines, usually does not affect protein properties such as folding [15,19,20]. Therefore, the native, S-Met containing, proteins, are frequently derivatized with selenium to facilitate the structure solution process. After successful determination of the structure of Se-derivatized protein, the obtained Se-Met containing coordinate set is then used as a starting model for structure solution and refinement of the native, S-Met, data. This approach has been utilized, for example, in the X-ray structure determination of 6-hydroxymethyl -7,8-dihydropterin pyrophosphokinase [21] or ESCRT-II eap-45 Glue domain [22].

Even if the main purpose of using the Se-Met data is only to yield a starting model for solving the native (S-Met) structure, we can find many cases when both Se-Met and native data are deposited. Why? - Because the authors found that the native and Se-Met structures were non-isomorphous, and had some interesting differences in the three-dimensional shape (conformation). The examples of such situation include both the native and Se-derivative structures, reported for orotidine-5'-monophosphate decarboxylase [23], the disulfide-bond protein dsbG [24], the RNA-dependent RNA polymerase QDE-1 [25], or phosphoinositide phosphatase MTMR2 [26].

It also happens that the Se-derivatized protein gives much better diffracting crystals than native protein, or, the structure of native protein can't be solved by using, as starting model, the Se-Met structure obtained from collected Se-Met data. In such situation, the Se-Met structure, which is determined from the Se-Met data, becomes the only structure to report. Such phenomenon took place, for example, in structure determination of alpha-isopropylmalate synthase [27], XLF/Cernunnos protein [28], chorismate synthase [29], Rho-associated kinase coiled-coil domain [30] or bacterioferritin A [31].

Because of the difference in chemical identity, the Se-Met structure (which is determined from the Se-Met derivatized protein), can't simply represent the structure of the native protein (and vice versa, the S-Met structure which is determined from the native S-Met data, can't represent the structure of the Se-Met derivatized protein). The Se-Met structure can only serve as model of the native protein, and vice versa, the S-Met native structure can only serve as model of the Se-Met derivatized protein.

Crystallographers very frequently use models to solve the new structures by the molecular replacement method. The starting models, which are used in molecular replacement, contain usually the chain fragments which have only similar folding to that in the investigated protein, with many differences in the amino acid sequence. In many cases, the starting models represent only the fragments of the newly investigated structures. Therefore, starting models, at the stage right

after successful rotation and translation, without any refinement, can't represent the final structures. At least, the missing fragments need to be built, the sequence in the rotated models has to be corrected, and some side chains or larger fragments of the chain need to be repositioned to fit the experimental electron density.

Similarly like the molecular replacement model, the Se-model which is used to solve the native, S-containing protein (and vice versa), needs later to be refined. The atoms with different chemical identity have to be corrected first.

It is not necessary to be an experienced chemist, to say, that selenium- and sulfur-containing compounds, from chemical point of view, are not the same. It is as simple as saying that the brass is not the gold, even if both shine in yellow. There are many examples which could explain why selenium-containing compounds are chemically not the same as sulfur-containing compounds. Not speaking about the differences in physical, biological, physiological, etc. properties, the sulfur-containing compounds and their selenium-containing analogs, even if they are isostructural, i.e., their three-dimensional structures share the same three-dimensional shape (conformation), they most likely will not show the same chemical features such as reactivity [32-34].

Small-molecule crystallography is much more powerful, than macromolecular crystallography, in the judgment of correctness or incorrectness of the refined atom type. Very high resolution of the majority of determined small-molecular structures, allows figuring out the correct atom type, simply by the analysis of interatomic distances and atomic displacement parameters.

Table 1 contains collection of interatomic distances, which involve S or Se atoms, and shows that the distances which involve Se atoms are much longer from the distances which involve S atoms. All examples, which are provided in Table 1, have been selected from author's previous work on X-ray structure determination of small-molecular organic compounds such as phosphoroorganic disulfides and diselenides [35-41], selenides [42], dithianes, dithiolanes, oxathiaphospholanes, dithiaphospholanes and oxaselenaphospholanes [43-53], phosphoramidothioates [44], phosphoramidoseenoates [54], sulfoxides [55-59] and thiophosphoryl compounds [60]. For full listing of compounds 1-42 included in that comparison, please see supplementary material.

Compound	P-S	P-Se	P=S	S-S	Se-Se	C-S	C-Se	Ref
1a,1b				2.033(3)a	2.298(2)b	1.802(7)a	1.960(6)b	[35]
						1.805(6)a	1.944(5)b	[35]
2					2.325(1)		1.961(7)	[35]
3a,3b	2.112(2)a	2.262(2)b	1.894(2)a	2.036(2)a	2.329(1)b			[36]
	2.118(2)a	2.255(2)b	1.891(2)a					[36]
			1.902(3)b					[36]
			1.908(3)b					[36]
1-42	2.087	2.244	1.917	2.057	2.331	1.808	1.949	[35-60]
			2.090P=Se					

**Table 1:** Interatomic distances, which involve sulfur and selenium atoms, in the X-ray crystal structures of selected organic compounds [35-60]. The values, which are listed in row '1-42', are the average bond lengths calculated from the respective values for all compounds 1-42. The detailed bond length listing for compounds 4-42 is provided in supplementary material.

#### Compound names:

1a: bis((2,3,4,6)-tetra-O-acetyl-beta-D-glucopyranosyl) disulfide [35]

1b: bis((2,3,4,6)-tetra-O-acetyl-beta-D-glucopyranosyl) diselenide [35]

2: bis((2,3,4,6)-tetra-O-acetyl-beta-D-glucopyranosyl) diselenide, crystal form 2 [35]

3a: bis(5,5-dimethyl-2-thioxo-1,3,2-dioxaphosphorinan-2-yl) disulfide [36]

3b: bis(5,5-dimethyl-2-thioxo-1,3,2-dioxaphosphorinan-2-yl) diselenide [36]

4-42: for compound names, please see supplementary material [37-60].

The crystal structures of bis (2,3,4,6)-tetra-O-acetyl-beta-D-glucopyranosyl) disulfide 1a and diselenide 1b are not identical (Table 1), even if 1a and 1b are crystallographically isomorphous [35]. The same is the crystallographic symmetry, space group, and three-dimensional shape (conformation) of both molecules 1a and 1b [35]. However, the bond lengths, which involve selenium, are much longer than the respective bonds which involve sulfur (Table 1). More, the unit-cell dimensions a,b,c of diselenide 1b are all longer than the unit cell constants of respective disulfide 1a [35].

Bis (5,5-dimethyl-2-thioxo-1,3,2-dioxaphosphorinan-2-yl) disulfide 3a and diselenide 3b are the examples of the analogs, which are totally non-isomorphous. They do not share even such crystallographic features like crystal symmetry or space group [36]. The overall three-dimensional shape of both molecules (conformation) is also quite different [36]. The bond lengths, which involve selenium, are, similarly

like in the previous observation for compounds 1a and 1b, much longer than the respective bonds that involve sulfur (Table 1).

The bonds, which involve selenium atoms, are longer than the bonds, which involve sulfur atom. This phenomenon is very clear and consistent in all compared compounds 1-42 in Table 1 [35-60], despite the differences in the chemical structure, and, despite the fact that the S or Se atoms in some of these compounds are involved in exo-or/and endocyclic connections. For details, please see the Table in supplementary material.

The average length of the single P-Se bond, which is calculated for the compounds 1-42 (Supplementary material), is 2.244 Å (Table 1) and is much longer than the average length of the single P-S bond (2.087 Å, Table 1). Similarly, the average length of double P=Se bond (2.090 Å) is much longer than respective average P=S bond (1.917 Å, Table 1).

The average length of the C-Se bond, calculated from values for the compounds 1-42 (Table 1 and Supplementary material), is 1.949 Å, and is much longer than the average length of the C-S bond (1.808 Å). Similarly, the average length of the diselenide bond (Se-Se; 2.331 Å) is also much longer than respective disulfide bond (S-S; 2.057 Å; Table 1).

From the presented above analysis of bond lengths, which involve Se and S atoms in small molecules 1-42, it is clear and easy to conclude that the three-dimensional structures of S- and Se-containing compounds, which differ only in chemical identity, S versus Se, will never align well. Thus, the selenomethionine side chain: Ca-C $\beta$ -C $\gamma$ -Se $\delta$ -C $\epsilon$ , will align well, with respective methionine chain: Ca-C $\beta$ -C $\gamma$ -S $\delta$ -C $\epsilon$ , only up to the C $\gamma$  atom. The average linear difference between the  $\gamma$  and  $\epsilon$  positions in such alignment is about 0.29 Å. Similarly, the side chains of selenocysteine, Ca-C $\beta$ -SeY, will not align well with native cysteines: Ca-C $\beta$ -Sy. At atom position  $\gamma$ , the average difference in length is about 0.15 Å. If the cysteines in the native protein form disulfide bridges: Ca1-C $\beta$ 1-Sy1-Sy2-C $\beta$ 2-Ca2, the alignment of the entire disulfide bridges with respective diselenide bridges in the Se-derivatized proteins: Ca1-C $\beta$ 1-SeY1-SeY2-C $\beta$ 2-Ca2, will be much worse than the alignment of single (free) cysteines with free selenocysteines. The alignment of disulfide with Diselenide Bridge, which starts from Ca1 atom, will end yet at C $\beta$ 1 atom, and will give the linear difference between distant  $\beta$ 1 and  $\alpha$ 2 positions of about 0.57 Å.

The evidence, which is provided above, is, hopefully, a convincing argument that the selenomethionines can't be refined in the macromolecular structure as native methionines.

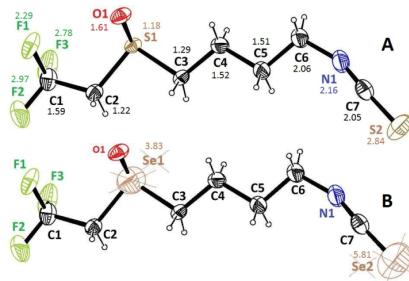
More proper way of handling the Se-Met residues in macromolecular refinement, has been established, as the research protocol, by the Joint Center for Structural Genomics (JCSG). This protocol employs some features from the small-molecular refinement. In each entry, which is deposited by this Center (and which contains Se-Met), we can find the following information: 'A met-inhibition protocol was used for selenomethionine incorporation during protein expression. The occupancy of the SE atoms in the MSE residues was reduced to 0.75 for the reduced scattering power due to partial S-Met incorporation'. For example, please see that information (in remark 3 in pdb file, or, in refine.details in cif file) in the PDB entries 4Q6K [61], 4R7S [62], 2OGI [63], or 2OU5 [64], or in any other deposited Se-Met entry by the JCSG Center. In September 2014, there were total of 1334 such JCSG depositions in the PDB archive. However, in these entries,

authors do not mention anything about handling of the remaining 25 percent of the atom type in these residues.

## Experimental, Results and Discussion

### Refinement of incorrect atom type, Se instead of S, in small molecule

The crystallographic refinement of small molecules, which is usually performed at high resolution, is very powerful in judgment whether the atom type in refined structure is correct or wrong. As an example, in this work, author re-refines one recently reported structure, of 4-isothiocyanato-1-butyl 2',2',2'-trifluoroethyl sulfoxide [57]. For the view of the molecule, Figure 1A. For bond lengths, see supplementary material (compound 36) [57]. For purpose of this work, an incorrect atom type of the two original sulfur atoms has been intentionally assumed (Figure 1B). The two sulfurs, S1 and S2, have been renamed to Se1 and Se2, respectively, and such model has been refined, with the use of the software available in author's facility [65,66].



**Figure 1:** (A) The X-ray crystal structure of 4-isothiocyanato-1-butyl 2',2',2'-trifluoroethyl sulfoxide (compound number 36 in supplementary material) [57]. The values near atom names are the equivalent isotropic atomic displacement parameters (B-factors) refined for these atoms. (B) Re-refined model of the same structure with assumed incorrect atom types (Se instead of S) [this work]. The model shows artificial increase of the size of thermal ellipsoids, and increased values of equivalent isotropic B-factors, due to wrong atom type assignment. The Se1 and Se2 atoms are crossed out, to indicate their wrong chemical identity, and to show that this model was incorrectly refined and can't represent the real structure.

The refinement, which has been performed with assumed wrong atom type (Se instead of S), yielded unreasonably enlarged thermal ellipsoids of these atoms, and significant increase of the value of crystallographic R-factor. For the values of refined equivalent isotropic atomic displacement parameters (B-factors), please see Figure 1. The R-factor of refined 'wrong' structure (Se instead of S, Figure 1B) is dramatically higher: R=0.201 [this work], than the R-factor of 'correct' structure (with S atoms, Figure 1A), which is R=0.016 [57]. The refined B-factors for seleniums Se1 and Se2 are unreasonably high. While the B-factors of sulfurs S1 and S2 were comparable to the B-factors of atoms in the closest vicinity (Figure 1A) [57], the B-factors of wrongly assumed selenium atoms Se1 and Se2 are unreasonably increased (Figure 1B) [this work]. The incorrect Se1 and Se2 atom types, are in Figure 1B crossed out, to indicate their wrong chemical identity.

## Handling of two different atom types, Se and S, in refinement of Se-Met proteins

It is obvious that an X-ray structure, which is usually displayed in the form of a set of atomic coordinates, is an average structure of all molecules, which were present in the investigated, i.e., exposed to X-ray, crystal. Therefore, it is very likely that the crystal which grew from the buffer containing of Se-Met derivatized protein, may include molecules, in which the Se-derivatization process did not proceed with 100% yield. In consequence, some of the individual molecules in that crystal may still contain native, S-Met, residues.

The publicly available, from the Protein Data Bank, entry 3B40 [18], which is under extensive investigation in this work, is most likely an example of that case. The partial presence of S-Met residues, or, in other words, not complete Se-Met derivatization of the protein, is possible [15]. To investigate this problem in detail, the six different re-refinement paths of the entry 3B40 have been carried out.

For purpose of this work, the coordinates (file 3B40.pdb) and structure factors (file 3b40-sf.cif) have been downloaded from the PDB site, and converted to appropriate formats. The downloaded coordinate file (file 3B40.pdb) is an S-Met containing model, and the downloaded structure factor file (file 3b40-sf.cif) contains X-ray intensities collected from the Se-Met crystal.

For purpose of further comparisons, the downloaded model 3B40 (the coordinates which contain S atoms) has been initially re-refined [65,66]. Four rounds of re-refinement were carried out. The coordinates and isotropic atomic displacement parameters (B-factors) were refined for all atoms. During this refinement, some side chains were repositioned in the electron density, and some new, well visible water molecules, added to the solvent network.

A separate coordinate file has been created, from the re-refined model, in which all S atoms have been renamed to Se. Then, the two separate refinement paths, A and B, against the same experimental data set (3b40-sf.cif), have been carried out. In refinement path A, the S-containing coordinates have been used, and refined in same way, as the authors originally did in their deposited structure. In refinement path B, the atom type was changed from S to Se. In both refinement paths A and B, the isotropic atomic displacement parameters (B-factors) were initially set to the arbitrary values, and then, refined individually, together with the B-factors of other atoms. The occupancies of S atoms were fixed, at 100%, in path A, and the occupancies of Se atoms were fixed, also as 100%, in path B, and were not refined. The B-values obtained of both refinement paths A and B, are collected in Table 2.

		Path A	Path B	Path C	Path D	Path E	Path F
Residue number	Average B side-chain, excluded	B-factor 100% S	B-factor 100% Se	B-factor 75% Se	Refined occupancy Se only	Global occupancy 84%S,16%Se	Refined occupancy S/Se
43	(15.4)	26.8	46.7	35.4	0.29	(15.4)	0.87/0.13
104	(35.7)	26.0	45.8	37.0	0.41	(35.7)	0.81/0.19
136	(24.3)	15.6	33.6	27.2	0.52	(24.3)	0.63/0.37
158	(27.2)	35.2	57.5	45.1	0.33	(27.2)	0.86/0.24
166	(15.8)	9.1	26.9	21.0	0.46	(15.8)	0.73/0.27
168	(18.5)	14.7	32.9	26.0	0.41	(18.5)	0.72/0.28
191	(15.9)	4.3	23.2	16.6	0.58	(15.9)	0.64/0.36
245	(16.5)	13.2	31.5	24.6	0.41	(16.5)	0.71/0.29
283	(18.3)	18.8	37.2	29.6	0.46	(18.3)	0.70/0.30
333	(31.4)	25.5	46.5	37.9	0.43	(31.4)	0.74/0.26

**Table 2:** Atomic displacement parameters (B-factors, Å<sup>2</sup>) and occupancies for the S and Se atoms in re-refinement of the PDB entry 3B40 [18]. Path A: refined B-factors, the presence of 100% sulfur was assumed. Path B: refined B-factors, the presence of 100% selenium was assumed. Path C: refined B-factors, the presence of selenium was assumed with 75% occupancy (JCSG approach). Path D: individually refined occupancies for Se atoms. Path E: refined value of global occupancy ratio of partial presence of S or Se atom type. Path F: individually refined values of occupancy ratio of partial presence of S or Se atom type in the residues. In paths D, E, and F, the B-values were set as equal to average values in the chain (values shown in parenthesis) and not refined.

Table 2 contains values of refined isotropic atomic displacement parameters (B-factors) of atoms, which in path A were refined as sulfurs (column 'Path A'), and in path B as seleniums (column 'Path B'). The residues, which contain these heteroatoms, are in the following ten sequence positions: 43, 104, 136, 158, 166, 168, 191, 245, 283, and 333. Refinement path B, in which the heteroatoms have correct chemical identity (with the experimental data), should yield

the B-factors which have comparable values with the average B-factors of the entire side chains. From the values collected in column 'Path B' in Table 2, it is difficult to find even single such regularity. Only for residues 166 and 191, and less evident for 104 and 136, the refined B values (in path B) may speak for the presence of selenium. For residues 283, 168, 245, the B-factors, which are closer to average side-chain value, are rather for S-atom refinement (path A). This speaks rather

for the presence of sulfur in these residues. The B-factors of remaining residues 43,158,333, likely indicate the partial presence of both sulfur and selenium (in fractional, less than 100%, occupancy). Such distribution of B-factor values in refinement paths A and B (Table 2) may speak for partial presence of sulfur and selenium in originally refined residues as S-Met by the authors. Worse, for the ten sequence positions listed in Table 2, the percentage of partial occupancy of S versus Se seems to be different in each residue. This is an indication of possibility, that the investigated protein 3B40 [18], was not fully Se-Met derivatized. Both refinement paths A and B, carried out in this work, did not allow answering the question what atom type, Se or S, is correct in the refined structure.

In order to investigate the problem whether sulfur or selenium was present in the ten sequence positions listed in Table 2, the four additional refinement paths: C, D, E, and F, have been carried out. Refinement path C followed exactly the refinement protocol, which has been established by the JCSG Center. In starting coordinate set, all sulfurs were renamed to seleniums, and the occupancies for the selenium atoms have been arbitrarily set to 0.75 and not refined. The refinement yielded the values of atomic displacement parameters (B-factors), which are collected in Table 2 in column 'Path C'. Comparison of B-values from refinement path C (75% Se) with B-values from refinement path B (100% Se) shows, that the B-values in path C are getting closer to the average values for the respective side chains. However, all refined B-values in Path C are still too large from the respective average side chain B-values.

It became interesting to find out, to which values will refine the individual occupancies of the Se atoms, when their B-values are set as equal to the average values for the respective side chains. Therefore, in refinement path D, the B-values of the ten selenium atoms were set as equal to the average values for the respective side chains, and not refined. Instead, the fractional occupancies for these atoms were refined. The obtained values are collected in Table 2 in column 'Path D'. All values of refined occupancies are lower from the value of 0.75, which was arbitrarily used in the JCSG-followed approach (Path C).

The refinement paths C and D, showed that the occupancies of the selenium atoms are lower than 100%. It may be an indication, that the ten methionines in the native protein were not completely Se-derivatized. However, the question, what happened to the missing 25% of the atom type after refinement path C, and the missing higher percentages in Path D, still remains open.

It seemed that the best way to investigate that problem was to employ directly the small-molecule approach. In paths E and F, the heteroatom positions were set to involve both Se and S atom types with initially equal occupancies (50% Se and 50% S). The B-values for these atoms were arbitrarily set as equal to the average values of the entire side chains, and not refined. In path E, the global occupancy ratio for the Se/S identity, was refined (as the same value for all ten residues). In path F, the occupancy ratio values for the Se/S identity, were refined for each residue individually.

Refinement path E yielded the value of occupancy ratio, for Se/S identity, equal to 16%/84%, respectively. This may indicate that the overall protein derivatization with selenium was successful only in about 16 percent. Refinement path F showed, that the occupancy ratio values were different in each of the ten residues. For details, see Table 2. However, it is too far to conclude, from the values given in column 'Path F' in Table 2, that the methionine number 43 underwent the Se-derivatization in exactly 13%, methionine 104 in 19%, methionine 136

in 37%, etc. The resolution of the collected data is only 2.0 Å [18], which is not high enough to draw such precise conclusions. Similarly, the refined value of the global occupancy ratio of Se/S atom type in path E, and refined individual occupancies for Se-'only' atoms in path D, can also be considered, as only very approximate.

## Conclusion

Handling of selenomethionines in crystallographic refinement of the selenium-derivatized proteins seems to be not as straightforward as, for example, the refinement of native methionines in native proteins. The problem is due to frequent, not complete substitution, of native methionines, with their selenium-containing equivalents, during chemical reaction. The Se-Met derivatized protein samples, used for crystallization, may contain macromolecules, in which the native methionines are still, partially present. In result, the crystal which grows from such sample, and is later subjected for X-ray analysis, may contain, partially, native sulfur.

Selenomethionines can't be refined purely as native methionines, as it was done, for example, by authors of the PDB-deposited entry 3B40 [18]. The reason is obvious, because selenium and sulfur represent different chemical identity. The presented above analysis of bond lengths, which involve Se and S atoms in selected small molecules 1-42 [35-60], shows clearly that the three-dimensional structures of the S-containing compounds and their Se-analogs, can never be identical. The reason is that the bonds, which involve selenium, are much longer than the respective bonds, which involve sulfur (Table 1). Even if the S- and Se-analogs are crystallographically isomorphous, they will never align well. The unit-cell dimensions of the isomorphous Se- and S-analogs, will also never be identical, since the Se-containing molecules are relatively larger than their S-containing counterparts. For example, see crystal structures of compounds 1a and 1b (Table 1) [35]. In fact, the Se- and S-containing analogs do not have any obstacles preventing them from being structurally non isomorphous. For example, see crystal structures of compounds 3a and 3b (Table 1) [36]. The performed refinement of a small molecule with intentionally altered atom type (Figure 1) yielded strange thermal mobility and very bad refinement statistics. This means that the atom coordinate sets which contain different atom types than in the investigated crystal, could serve, at most, only as models, but definitely can't represent directly the structure of investigated compound.

If selenomethionines can't be handled, in refinement, as 100% methionines, would it be better to handle them, during refinement, as UNX (unknown atom types) or UNK (unknown amino acid residues)? - Most likely not, because we already know that the atom type is either Se or S, and the unknown amino acid is either Se-Met or S-Met.

If, again, the Se atoms can't be refined as 100% S, would it be better to delete these atoms from the coordinates and to refine the remaining residue fragments as alanines? Most likely also not, because the entire side chains were already visible in the density, so there was no reason to 'shorten' them.

Refinement of selenomethionines with the complete (100%) occupancies of the Se atoms, in the structures of not-fully SeMet derivatized proteins, is also not the best way, since it yields unnaturally enlarged values of the atomic displacement parameters.

The way of handling the selenomethionines, which is close to reality, has been suggested by the researchers from the Joint Center for

Structural Genomics, who refine the Se-Met residues with the occupancies of the Se atoms reduced to 75%.

In reality, the partial derivatization of native methionines with selenium does not have to proceed with the same yield for each individual methionine in the entire polypeptide chain. The experiment performed in this work (Table 2, path D) suggests that the occupancies of the Se atoms can be refined individually for each Se-Met residue.

The approach, which is commonly employed in small-molecule crystallography, and can mimic the real situation in the closest-possible way, is the partial refinement of both Se and S identities, with fractional occupancy. It means that the residues of interest can be refined partially as selenomethionines and partially as methionines. The experiments performed in this work showed that this method has, however, essential limitations. Macromolecular crystals usually do not diffract to such high resolution ranges as small molecules. Low resolution of macromolecular data is the limitation for use of the small-molecular approach. Therefore, handling the residues, partially as Se-Met and partially as S-Met, can't be used as general approach for macromolecular refinement of every Se-Met derivatized protein. Handling the selenomethionines in macromolecular refinement depends on the individual case, i.e., depends mainly on the crystallographic resolution of the data collected. The small-molecule approach can apply only to the macromolecular data, which were collected from very well ordered crystals, which diffracted to high resolution.

## Acknowledgements

This paper is dedicated to Professor Helen M. Berman on the occasion of her birthday. Financial support (Grant No. DEC-2012/05/B/ST4/00075) by the Polish National Science Center is gratefully acknowledged.

## References

1. Berman H, Henrick K, Nakamura H (2003) Announcing the worldwide Protein Data Bank. *Nat Struct Biol* 10: 980.
2. Berman HM, Kleywegt GJ, Nakamura H, Markley JL (2013) The future of the protein data bank. *Biopolymers* 99: 218-222.
3. Berman HM, Burley SK, Chiu W, Sali A, Adzhubei A, et al. (2006) Outcome of a workshop on archiving structural models of biological macromolecules. *Structure* 14: 1211-1217.
4. Dauter Z (2003) Protein structures at atomic resolution. *Methods Enzymol* 368: 288-337.
5. Li Y, Wu Y, Blaszczyk J, Ji X, Yan H (2003) Catalytic roles of arginine residues 82 and 92 of Escherichia coli 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase: site-directed mutagenesis and biochemical studies. *Biochemistry* 42: 1581-1588.
6. Blaszczyk J, Lu Z, Li Y, Yan H, Ji X (2014) Crystallographic and molecular dynamics simulation analysis of Escherichia coli dihydroneopterin aldolase. *Cell Biosci* 4: 52.
7. Blaszczyk J, Li Y, Gan J, Yan H, Ji X (2007) Structural basis for the aldolase and epimerase activities of Staphylococcus aureus dihydroneopterin aldolase. *J Mol Biol* 368: 161-169.
8. Chruszcz M, Ciardiello MA, Osinski T, Majorek KA, Giangrieco I, et al. (2013) Structural and bioinformatic analysis of the kiwifruit allergen Act d 11, a member of the family of ripening-related proteins. *Mol Immunol* 56: 794-803.
9. (2012) Crystal structure of a conserved domain protein from Streptococcus pneumoniae TIGR4 at 1.91 Å resolution. Joint Center for Structure Genomics entry PDB 4IAJ.
10. Chao X, Tempel W, Bian C, Cerovina T, Walker JR, et al. (2011) Crystal structure of the RRM domain of human SETD1A. Structural Genomics Consortium (SGC) entry PDB 3S8S.
11. Xu C, Tempel W, Li Z, He H, Wernimont AK, et al. (2012) Crystal structure of human NUP43. Structural Genomics Consortium (SGC) entry PDB 4I79.
12. Alicea-Velázquez NL, Jakoncic J, Boggon TJ (2013) Structure-guided studies of the SHP-1/JAK1 interaction provide new insights into phosphatase catalytic domain substrate recognition. *J Struct Biol* 181: 243-251.
13. (2013) Crystal structure of a hypothetical protein from Eubacterium siraeum DSM 15702 at 1.49 Å resolution. Joint Center for Structure Genomics entry PDB 4NZK.
14. Hendrickson WA (1999) Maturation of MAD phasing for the determination of macromolecular structures. *J Synchrotron Radiat* 6: 845-851.
15. Walden H (2010) Selenium incorporation using recombinant techniques. *Acta Crystallogr D Biol Crystallogr* 66: 352-357.
16. Hatfield DL, Gladyshev VN (2002) How selenium has altered our understanding of the genetic code. *Mol Cell Biol* 22: 3565-3576.
17. Salgado PS, Taylor JD, Cota E, Matthews SJ (2011) Extending the usability of the phasing power of diselenide bonds: SeCys SAD phasing of CsgC using a non-auxotrophic strain. *Acta Crystallogr D Biol Crystallogr* 67: 8-13.
18. Bonanno JB, Patskovsky Y, Dickey M, Bain KT, Mendoza M, et al. (2007), Crystal structure of the probable dipeptidase PvdM from *Pseudomonas aeruginosa*. Protein Structure Initiative, New York SGX Research Center for Structural Genomics (NYSGXRC) entry PDB 3B40.
19. Budisa N, Minks C, Medrano FJ, Lutz J, Huber R, et al. (1998) Residue-specific bioincorporation of non-natural, biologically active amino acids into proteins as possible drug carriers: Structure and stability of the per-thiaproline mutant of annexin V. *Proc Natl Acad Sci USA* 95: 455-459. PMID: 9435213.
20. Budisa N, Pifat G (1998) Probing Protein Stability with Non-natural Amino Acids. *Croat Chem Acta* 71: 179-187.
21. Xiao B, Shi G, Chen X, Yan H, Ji X (1999) Crystal structure of 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase, a potential target for the development of novel antimicrobial agents. *Structure* 7: 489-496.
22. Alam SL, Langelier C, Whitby FG, Koirala S, Robinson H, et al. (2006) Structural basis for ubiquitin recognition by the human ESCRT-II EAP45 GLUE domain. *Nat Struct Mol Biol* 13: 1029-1030.
23. Poulsen JCN, Harris P, Jensen KF, Larsen S (2001) Selenomethionine substitution of orotidine-50-monophosphate decarboxylase causes a change in crystal contacts and space group. *Acta Crystallogr D* 57: 1251-1259.
24. Heras B, Edeling MA, Byriel KA, Jones A, Raina S, et al. (2003) Dehydration converts DsbG crystal diffraction from low to high resolution. *Structure* 11: 139-145.
25. Laurila MR, Salgado PS, Makeyev EV, Nettleship J, Stuart DI, et al. (2005) Gene silencing pathway RNA-dependent RNA polymerase of *Neurospora crassa*: yeast expression and crystallization of selenomethionated QDE-1 protein. *J Struct Biol* 149: 111-115.
26. Begley MJ, Taylor GS, Kim SA, Veine DM, Dixon JE, et al. (2003) Crystal structure of a phosphoinositide phosphatase, MTMR2: insights into myotubular myopathy and Charcot-Marie-Tooth syndrome. *Mol Cell* 12: 1391-1402.
27. Koon N, Squire CJ, Baker EN (2004) Crystallization and preliminary X-ray analysis of alpha-isopropylmalate synthase from *Mycobacterium tuberculosis*. *Acta Crystallogr D Biol Crystallogr* 60: 1167-1169.
28. Li Y, Chirgadze DY, Bolanos-Garcia VM, Sibanda BL, Davies OR, et al. (2008) Crystal structure of human XLF/Cernunnos reveals unexpected differences from XRCC4 with implications for NHEJ. *EMBO J* 27: 290-300.

29. Quevillon-Cheruel S, Leulliot N, Meyer P, Graille M, Bremang M, et al. (2004) Crystal structure of the bifunctional chorismate synthase from *Saccharomyces cerevisiae*. *J Biol Chem* 279: 619-625.
30. Tu D, Li Y, Song HK, Toms AV, Gould CJ, et al. (2011) Crystal structure of a coiled-coil domain from human ROCK I. *PLoS One* 6: e18080.
31. Gupta V, Gupta RK, Khare G, Salunke DM, Tyagi AK (2009) Crystal structure of Bfr A from *Mycobacterium tuberculosis*: incorporation of selenomethionine results in cleavage and demetallation of haem. *PLoS One* 4: e8028.
32. Oae S, ed. (1977) Organic Chemistry of Sulfur. Plenum Press, New York.
33. Sunoj RB (2012) Theoretical aspects of organoselenium chemistry. In Rappoport Z, ed. The Chemistry of Organic Selenium and Tellurium Compounds. Wiley 3: 1-48.
34. Woollins JD, Laitinen R, eds. (2011) Selenium and Tellurium Chemistry. From Small Molecules to Biomolecules and Materials.
35. Potrzebowksi MJ, Michalska M, Blaszczyk J, Wieczorek MW, Ciesielski W, et al. (1995) Molecular modeling, X-ray diffraction and <sup>13</sup>C, <sup>77</sup>Se CP/MAS NMR studies of bis((2,3,4,6)-tetra-O-acetyl-beta-D-glucopyranosyl) diselenide and disulfide. *J Org Chem* 60: 3139-3148.
36. Potrzebowksi MJ, Grossmann G, Blaszczyk J, Wieczorek MW, Sieler J, et al. (1994) X-ray and solid state NMR studies of bis (5,5-dimethyl-2-thioxo-1,3,2-dioxaphosphorinan-2-yl) disulfide and diselenide. *Inorg Chem* 33: 4688-4695.
37. Knopik P, Luczak L, Potrzebowksi MJ, Michalski J, Blaszczyk J, Wieczorek MW (1993) Structure and dynamics of bis(organothiophosphoryl) disulfides in the solid state. X-ray diffraction and crosspolarization magic angle spinning nuclear magnetic resonance studies. *J Chem Soc Dalton*, 2749-2757.
38. Potrzebowksi MJ, Blaszczyk J, Wieczorek MW (1995) Studies of solvate and inclusion complexes of bis [6-O,6'-O'(1,2;3,4-diisopropylidene-alpha-D-galactopyranosyl)-thiophosphoryl] disulfide in the solid state by NMR and X-ray methods. *J Org Chem* 60: 2549-2562.
39. Potrzebowksi MJ, Blaszczyk J, Wieczorek MW, Klinowski J (1997) Solid-state NMR and X-ray diffraction studies of the geometry and intramolecular dynamics of bis (organothiophosphoryl) dichalcogenides. *J Phys Chem-US* 101: 8077-8084.
40. Potrzebowksi MJ, BÅ,aszczysk J, Majzner WR, Wieczorek MW, Baraniak J, et al. (1998) X-ray and high resolution selenium-77 solid state NMR spectroscopy as complementary probes to structural studies of organophosphorus diselenides. *Solid State Nucl Magn Reson* 11: 215-224.
41. Potrzebowksi MJ, Ganicz K, Ciesielski W, Skowronska A, Wieczorek MW, Blaszczyk J, et al. (1999) Investigation of solute-solvent interactions in a dithiophosphoroorganic carbohydrate derivative by means of X-ray analysis and solid state NMR. *J Chem Soc Perk T 2*, 2163-2170.
42. Wieczorek MW, Blaszczyk J, Potrzebowksi MJ, Skowronska A, Dembinski R (1995) Crystal and molecular structure and solid state NMR studies of bis(5,5-dimethyl-2-oxo-1,3,2-dioxaphosphorinan-2-yl) selenide. *Phosphorus Sulfur* 102: 15-24.
43. Mikolajczyk M, Luczak T, Graczyk PP, Wieczorek MW, Blaszczyk J, et al. (1997) Solid state conformation of the anomeric effect in 2-phosphoryl, 2-thiophosphoryl and 2-selenophosphoryl substituted 1,3-dithiolanes. *J Organomet Chem* 536: 355-360.
44. Uznanski B, Grajkowski A, Krzyzanowska B, Kazmierkowska A, Stec WJ (1992) Stereochemistry of base-catalyzed ring opening of 1,3,2-oxathiaphospholanes. Absolute configuration of 2-[N-[(Rc)-1-(alpha-naphthyl) ethyl]-amino]-2-thiono-1,3,2-oxathiaphospholanes and O,S-dimethyl N-[(Rc)-1-(alpha-naphthyl)-ethyl] phosphoramidothioates. *J Am Chem Soc* 114: 10197-10202.
45. Mikolajczyk M, Graczyk PP, Potrzebowksi M, Wieczorek MW, Blaszczyk J (1995) Solid state conformation and 31P CP/MAS NMR studies of (5,5-dimethyl-1,3-dithian-2-yl) triphenyl-phosphonium chloride and (5,5-dimethyl-1,3-dithian-2-yl) trimethyl-phosphonium chloride. *Phosphorus Sulfur* 103: 111-124.
46. Blaszczyk J, Wieczorek MW, Okruszek A (1996) Conformation of oxathiaphospholane and dithiaphospholane rings in the solid state. *J Mol Struct* 374: 85-96.
47. Blaszczyk J, Wieczorek MW, Okruszek A, Sierzchala A, Kobylanska A, et al. (1996) The synthesis and X-ray structural studies of 2-cholesteryl-2-thio-1,3,2-oxathiaphospholane and 2-cholesteryl-2-thio-1,3,2-dithiaphospholane. *J Chem Crystallogr* 26: 33-42.
48. Stec WJ, Karwowski B, Guga P, Misiura K, Wieczorek M, Blaszczyk J (1996) Stereochemistry of DBU-assisted reaction of nucleoside 3'-O-(2-thiono-1,3,2-oxathiaphospholanes) with 5'-hydroxynucleosides. *Phosphorus Sulfur* 109-110: 257-260.
49. Potrzebowksi MJ, Blaszczyk J, Wieczorek MW, Misiura K, Stec WJ (1997) The synthesis, X-ray and solid state NMR studies of 2-N,N-diisopropylamino-1,3,2-oxaselenaephospholane-2-selone. *J Chem Soc Perk T 2*, 163-168.
50. Stec WJ, Karwowski B, Boczkowska M, Guga P, Koziolkiewicz M, et al. (1998) Deoxyribonucleoside 3'-O-(2-Thio- and 2-Oxo-'spiro'-4,4-pentamethylene-1,3,2-oxathiaphospholane)s: Monomers for Stereocontrolled Synthesis of Oligo(deoxyribonucleoside phosphorothioate)s and Chimeric PS/PO Oligonucleotides. *J Am Chem Soc* 120: 7156-7167.
51. Mikolajczyk M, Graczyk PP, Wieczorek MW, Blaszczyk J (1992) Crystal and molecular structure of 5,5-dimethyl-1,3-dithian-2-ylcarboxylic acid. Is the axial conformer stabilized by hydrogen bonds? *Heteroatom Chem* 3: 625-630.
52. Krzyzanowska B, Stec WJ, Wieczorek MW, Blaszczyk J (1994) Synthesis and assignment of absolute configuration of enantiomeric dicyclohexylammonium 2-oxo-2-thioxo-1,3,2-oxathiaphospholanes. *Heteroatom Chem* 5: 533-539.
53. Blaszczyk J, Wieczorek MW, Okruszek A, Olesiak M, Karwowski B (1994) Geometry and conformation of crystalline forms of beta-naphthoxy-substituted 2-thio-1,3,2-oxathiaphospholanes and 2-thio-1,3,2-dithiaphospholanes. *Heteroatom Chem* 5: 519-527.
54. Wieczorek MW, Blaszczyk J, Krzyzanowska B (1993) Crystal and molecular structures and the absolute configuration of O-Se-dimethyl-N-(1-phenylethyl)phosphoramidoselenoates. *Heteroatom Chem* 4: 79-83.
55. Mikolajczyk M, Krysiak JA, Midura WH, Wieczorek MW, Blaszczyk J (1996) Alpha-phosphoryl sulfoxides. X. A general synthesis of optically active alpha-chlorovinyl sulfoxides. *Tetrahedron-Asymmetr* 7: 3513-3520.
56. Mikolajczyk M, Midura WH, Wladislaw B, Biaggio FC, Marzorati L (1997) Alpha-phosphoryl sulfoxides. XI. Sulfonylation of alpha-phosphoryl sulfoxides and a general synthesis of optically active ketene dithioacetal mono-S-oxides. *Tetrahedron* 53: 2959-2972.
57. Kielbasinski P, Luczak J, Cierpial T, Blaszczyk J, Sieron L, et al. (2014) New enantiomeric fluorine-containing derivatives of sulforaphane: Synthesis, absolute configurations and biological activity. *Eur J Med Chem* 76: 332-342.
58. Wieczorek MW, Blaszczyk J, Mikolajczyk M, Kielbasinski P, Zurawinski R (1994) Crystal and molecular structure and determination of absolute configuration of two enantiomers of methoxycarbonylmethyl carboxymethyl sulfoxide. *Heteroatom Chem* 5: 167-173.
59. Wieczorek MW, Blaszczyk J, Dziuba B, Kielbasinski P, Mikolajczyk M, et al. (1995) Crystal and molecular structure of cyclic sulfoxides: 2-cyano-2-ethoxycarbonyl-3,6-dihydro-4,5-dimethyl-2H-thiapyran 1-oxide and 2-phenyl-2-methoxy-carbonyl-3,6-dihydro-2H-thiapyran 1-oxide. *Heteroatom Chem* 6: 631-638.
60. Mikolajczyk M, Luczak J, Wieczorek MW, Blaszczyk J (1993) Stereochemistry of organophosphorus cyclic compounds - XVII. Synthesis of N-methyl-2-methoxy-1,3,2-oxazaphosphorinan-2-thione and dicyclohexylammonium salt of N-methyl-2-oxo-2-thio-1,3,2-oxazaphosphorinan and their crystal and molecular structures. *Pol J Chem* 67: 1087-1097.
61. Crystal structure of a putative neuraminidase from *Bacteroides caccae* ATCC 43185 at 1.90 Å resolution. Joint Center for Structure Genomics entry PDB 4Q6K (2014).
62. (2014) Crystal structure of a tetratricopeptide repeat protein from *Parabacteroides merdae* ATCC 43184 at 2.39 Å resolution. Joint Center for Structure Genomics entry PDB 4R7S.

63. (2007) Crystal structure of a putative metal dependent phosphohydrolase (sag1661) from streptococcus agalactiae serogroup v at 1.85 Å resolution. Joint Center for Structure Genomics entry PDB 2OGI (2007).
64. (2007) Crystal structure of a pyridoxamine 5'-phosphate oxidase-related FMN-binding protein (jann\_0254) from Jannaschia sp. ccs1 at 1.60 Å resolution. Joint Center for Structure Genomics entry PDB 2OU5.
65. Sheldrick GM (2008) A short history of SHELX. *Acta Crystallogr A* 64: 112-122.
66. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66: 486-501.