research papers

Supramolecular insight into the substitution of sulfur by selenium, based on crystal structures, quantum-chemical calculations and biosystem recognition

Ivana S. Đorđević,[a] Marko Popadić,[b] Mirjana Sarvan,[c] Marija Petković-Benazzouz[c]* and Goran V. Janjić[a]*

[a]SI Institute of Chemistry, Technology and Metallurgy, National Institute, University of Belgrade, Njegoševa 12, Belgrade, Serbia, [b]Faculty of Chemistry, University of Belgrade, Studentski trg 12-16, Belgrade, Serbia, and [c]Faculty of Physics, University of Belgrade, Studentski trg 12, Belgrade, Serbia. *Correspondence e-mail: marijapetkovic@ff.bg.ac.rs, goran.janjic@ihtm.bg.ac.rs

Statistical analysis of data from crystal structures extracted from the Cambridge Structural Database (CSD) has shown that S and Se atoms display a similar tendency towards specific types of interaction if they are part of a fragment that corresponds to the side chains of cysteine (Cys), methionine (Met) seleno-cysteine (Sec) and selenomethionine (Mse). The most numerous are structures with C—H···Se and C—H···S interactions (∼80%), notably less numerous are structures with Se···Se and S···S interactions (∼5%), and Se···π and S···π interactions are the least numerous. The results of quantum-chemical calculations have indicated that C—H···Se ($\sim$−0.8 kcal mol^{-1}) and C—H···S interactions are weaker than the most stable parallel interaction ($\sim$−3.3 kcal mol^{-1}) and electrostatic interactions of σ/π type ($\sim$−2.6 kcal mol^{-1}). Their significant presence can be explained by the abundance of CH groups compared with the numbers of Se and S atoms in the crystal structures, and also by the influence of substituents bonded to the Se or S atom that further reduce their possibilities for interacting with species from the environment. This can also offer an explanation as to why O—H···Se ($\sim$−4.4 kcal mol^{-1}) and N—H···Se interactions ($\sim$−2.2 kcal mol^{-1}) are less numerous. Docking studies revealed that S and Se rarely participate in interactions with the amino acid residues of target enzymes, mostly because those residues preferentially interact with the substituents bonded to Se and S. The differences between Se and S ligands in the number and positions of their binding sites are more pronounced if the substituents are polar and if there are more Se/S atoms in the ligand.

1. Introduction

Selenium and sulfur are basic elements with closely related properties, found in nature in a vast array of active compounds, and also used as reactants of numerous chemical and biochemical processes. Selenium is an essential trace element found in selenoproteins in the form of the selenium analogue of cysteine (selenocysteine, Sec). Although it is toxic at higher concentrations, recent research indicates that it might also be effective in preventing different types of disease (Vinceti *et al.*, 2019; Li *et al.*, 2019; Raygan *et al.*, 2019), mostly because of its high chemical reactivity in metabolism. As these elements are very similar when it comes to electronegativity (2.58 for sulfur and 2.55 for selenium), number of oxidation states and ionic radius, they also share many properties (Yoshizawa & Böck, 2009) However, a few important differences between these amino acids should be pointed out. The pK_a of Sec is much lower (5.3) than that of Cys (8.3) (Allmang
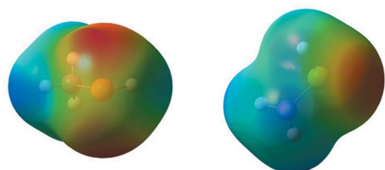
& Krol, 2006; Muttenthaler & Alewood, 2008) which makes Sec a considerably more potent nucleophile under neutral and acidic conditions. In addition, selenium is softer than sulfur, with a polarizability volume of 3.8 Å (compared with just 2.9 Å for S) (Steinmann *et al.*, 2010).

Non-covalent interaction has been identified as a driving force for crystal packing (Caracelli *et al.*, 2012), self-assembly of (macro)molecules (Voth *et al.*, 2009), stabilization of different systems (Wang *et al.*, 2019; Breugst *et al.*, 2017) and biological recognition (Daze & Hof, 2016). Hydrogen bonds (HBs), a class of non-covalent interaction, are generally the most abundant and the most studied.

It is also worth mentioning that the importance of various classes of non-covalent interaction, such as those involving the tetrels (Group 14 in the periodic table), pnictogens (Group 15), chalcogens (Group 16), halogens (Group 17) and the noble gases (Group 18), has been recognized over the years (Bauzá *et al.*, 2013; Bauzá & Frontera, 2018; Metrangolo & Resnati, 2001; Brammer, 2017; Benz *et al.*, 2018; Politzer *et al.*, 2014; Bauzá & Frontera, 2015; Frontera & Bauzá, 2017). In the oxygen family, the non-covalent interactions involving O and S have been extensively investigated. By comparison, interactions that involve Se and Te have only recently gained more attention. In organoselenium compounds, selenium is most frequently present in the divalent state, and as such displays two σ-holes, electropositive regions located along the bonding axis and perpendicular to the two lone pairs of valence electrons (Murray *et al.*, 2009).

Intermolecular $N-H\cdots$Se interactions were observed more than 50 years ago by Hope (1965), while intramolecular $C-H\cdots$Se interactions were characterized by Iwaoka & Tomoda (1994). Further spectroscopic and crystal structure investigations confirmed the presence and significance of selenium hydrogen-bonding interactions within biomacromolecular systems, the crystal packing, and the synthesis of chiral selenones (Mukherjee *et al.*, 2010). More recently, a modest number of theoretical or combined studies involving $X-H\cdots$Se have been carried out. The influence of various electron-donating and -withdrawing substituents (Q) in the *para* position of cationic phenol [4-Q–C_6H_4–OH$\cdots$SeH$_2$]$^+$ was investigated by Das *et al.* (2017). Madzhidov and Chmutova analysed the nature of $X-H\cdots$Se hydrogen-bonded complexes of several organoselenium compounds using quantum-chemistry methods. They pointed out the unusual electrostatic repulsion between Se and H atoms located opposite to classical $H\cdots$N or $H\cdots$O contacts, but the greater stabilization effect of charge transfer and covalence in the former (Madzhidov & Chmutova, 2010). The implications of $S-H\cdots$S and $Se-H\cdots$Se in determining the molecular arrangement in the crystal packing of thiophenol and selenophenol were reported by Thomas *et al.* (2015).

Non-covalent interactions of Se(lone pair)$\cdots$(aryl) or Se(lone pair)$\cdots$(lone pair) type are probably the most studied and well documented. Small molecules retrieved from the Cambridge Structural Database (CSD; Groom *et al.*, 2016) have been used as model systems to study the interactions between an Se atom and aromatic molecules. In their paper,

Hartman *et al.* (2006) came to the conclusion that these non-covalent interactions may participate in the stabilization of Se-containing drugs and protein active sites. Caracelli and co-workers have pointed out that these Se(lone pair)$\cdots$(aryl) interactions lead to zero- or one-dimensional aggregation patterns, based on a crystal structure database search of selenium/aryl-containing compounds (Caracelli *et al.*, 2012). In the last few years, Saberinasab and co-workers have studied the effects of electron-donating and electron-withdrawing substituents R = H, F, Cl, CH$_3$ on R_2Se$\cdots$benzene interactions and the effect of a strong cation$\cdots\pi$ interaction in the R_2Se$\cdots$C$_6$H$_6\cdots M^+$ model system (M^+ = Li$^+$, Na$^+$ and R = H, CH$_3$) (Saberinasab *et al.*, 2016a,b). A theoretical study published recently explored the Se$-$H$\cdots\pi$(aryl) and Se(lone pair)$\cdots\pi$(aryl) interactions between an H$_2$Se group and unsubstituted or monosubstituted benzene molecules, and these results were further compared with an analogous series of sulfur–benzene dimers. The study found that selenium interacts more strongly with aromatic systems than sulfur, which could be important in the stabilization and inhibition of proteins, indicating that the Se$-$H group is a better donor than the S$-$H group when it comes to hydrogen bonding (Senćanski *et al.*, 2017).

Non-covalent interactions of Se$\cdots X$ type (X = heteroatom) have also attracted a great deal of attention. Sanz and co-workers have published several articles on the competition in strength between O$-$H$\cdots$Se or S$-$H$\cdots$Se intramolecular hydrogen bonds and O$\cdots$Se and S$\cdots$Se interactions (Sanz *et al.*, 2002, 2003a,b). The analysis of crystal structures with an Se$-$Se$\cdots X$ fragment (X = Se, S, O) has shown the correlation between the Se$-$Se bond length and the strength of the Se$\cdots X$ interaction. The trend in Se$-$Se distances for different Se$-$Se$\cdots X$ fragments follows the sequence of X as Se > S > O (Linden *et al.*, 2014). Iwaoka and co-workers made an important contribution to the study of Se$\cdots$O, Se$\cdots$N and Se$\cdots$halogen interactions (Iwaoka *et al.*, 2002a,b, 2004; Komatsu *et al.*, 1999; Iwaoka & Tomoda, 1996). So far, several important papers have been published concerning chalcogen–chalcogen Se$\cdots$S, Se$\cdots$Se and Se$\cdots$Te types of interaction. In one systematic approach, Tiecco and co-workers reported that Se$\cdots$S interactions have a high directionality influence on asymmetric syntheses. They came to the conclusion that Se$\cdots$S contacts have a more pronounced selectivity in comparison with the influence originating from Se$\cdots$O or Se$\cdots$N (Tiecco *et al.*, 2000, 2002, 2003, 2004, 2006a). The same authors also offered experimental evidence for the presence of Se$\cdots$S interactions in the crystal form, as well as in CDCl$_3$ solution (Tiecco *et al.*, 2006b). Also, a few scientific papers deal with Se$\cdots$Se and Se$\cdots$Te interactions in an attempt to give a better insight into the nature of these interactions (Ibrahim & Safy, 2019; Bleiholder *et al.*, 2007; Sánchez-Sanz *et al.*, 2011).

Theoretical characterization of the impact of sulfur-to-selenium substitution on hydrogen-bonding potential and photophysical properties in an emissive RNA alphabet has shown that the replacement of sulfur with selenium in derived nucleobases has a minimal effect on the geometries and energies of base pairs in the classical Watson–Crick model,

while base pairs in the Hoogsteen model are destabilized, compared with natural pairs (Chawla *et al.*, 2018).

The concept of the present work includes a statistical analysis of the interactions of S and Se sites from Cys and Met residues and their Se derivatives (Sec and Mse) in crystal structures retrieved from the CSD. Three model systems were used to search the CSD and examine whether the replacement of an S with an Se atom provides a significant change in the number and variety of certain types of interactions with surrounding species. Density functional theory (DFT) calculations on hydrogen bonds between CH_4, NH_3 or H_2O molecules as donors for hydrogen bonding, and CH_3SeCH_3 and CH_3SeH molecules with both donor and acceptor roles, have been carried out. We also investigated the parallel interaction and σ-type interactions between CH_3SeCH_3 and CH_3SeH molecules. Finally, we assessed the interactions of S and Se analogues with biochemical systems based on results from molecular docking, in order to determine whether our prediction, based on the CSD results and our calculations, fits the interpretation for binding of compounds that contain Se and/or S atoms to biomolecules.

2. Methodology

The crystallographic study is based on structures archived in the Cambridge Structural Database (CSD, Version 5.36; Groom *et al.*, 2016). A particular structure was considered a 'hit' if the distance between the Se and *A* (any atom) was less than 4.5 Å (for Se interactions) or the distance between the S and *A* (any atom) was less than 4.0 Å (for S interactions) (Fig. 1). This distance was denoted the *d* parameter.

The program *ConQuest* (Version 1.10; Bruno *et al.*, 2002) was used to retrieve all structures from the CSD satisfying the following criteria: (i) error-free coordinates according to the criteria used in the CSD, (ii) not disordered structures, (iii) H-atom positions normalized using the CSD default bond lengths, (iv) no powder structures, (v) no polymer structures, (vi) determined 3D coordinates, and (vii) a crystallographic *R* factor less than 0.10.

All quantum-chemical calculations were performed using the *GAUSSIAN09* program (Frisch *et al.*, 2013). The geometries of isolated methaneselenol (CH_3SeH), dimethylselenide (CH_3SeCH_3), water, ammonia and methane molecules were optimized and further used to calculate interaction energies in all dimer model systems. Surface investigations were performed on the MP2/aug-cc-pVDZ level. The CCSD(T)/CBS method offered information on interaction energies and limits for calculated minima. The CCSD(T)/CBS interaction energies were evaluated using the extrapolation scheme of Makie & DiLabio (2011), including calculations at the MP2/aug-cc-pVDZ, MP2/aug-cc-pVTZ, MP2/aug-cc-pVQZ and CCSD(T)/aug-cc-pVDZ levels. The basis set superposition error (BSSE) was corrected for all calculated energies based on the well known counterpoise method (Boys & Bernardi, 1970).

By comparing selected computational levels with the CCSD(T)/CBS energy values, the desired accuracy can be achieved. However, the choice of the best method (the best match to calculated limits and the most time-saving) for a particular property is realized through a rigorous benchmark study of a broad set of quantum-chemical methods and basis sets, along with the LC-WPBE, TPSS, TPSS-D3, M062X-D3, CAM-B3LYP-D3, CAM-B3LYP-D3BJ, WB97XD and PBE0 functionals. The basis sets aug-cc-pVDZ, aug-cc-pVTZ and def2-TZVP were tested for each functional. The calculations were also performed for interactions between two molecules of CH_3SeCH_3 and CH_3SeH along the C—Se and H—Se bonds.

The molecular electrostatic potential surface (MEPS) has been shown to be an effective method for predicting non-covalent interactions, helping us to understand them better in our model systems. As a preliminary study, MEPS analysis was performed for both molecules, CH_3SeCH_3 and CH_3SeH. Using *GAUSSIAN09*, the MP2 method and the aug-cc-pVQZ basis set to calculate wavefunctions, we were able to investigate electrostatic potential maps for both molecules.

All enzyme structures used in the docking studies were extracted from the Protein Data Bank (PDB; Berman *et al.*, 2000). In the subsequent step of enzyme preparation from that structure, all water molecules and ligands were removed. The compounds of S and Se, treated as ligands for binding to biomolecules, were optimized at the B3LYP/aug-cc-pVDZ level using *GAUSSIAN09*. The *AutoDockTools* program was applied to the process of docking preparation, while the docking study was achieved with the *AutoDock Vina* program (Trott & Olson, 2010). All enzyme residues were kept rigid during the calculation, while the single bonds of the ligands were set to rotate. The whole enzyme was placed in a grid box large enough to accommodate the ligands and allowing it to move freely during the docking run.

According to the father of 'supramolecular chemistry' J. M. Lehn (Lehn, 1994), it is defined as the chemistry of the intermolecular bond, covering the structures and functions of the entities formed by the association of two or more chemical species. In addition to that, it is predominantly based upon tailor-made intermolecular interactions. Biological systems offer valuable inspiration and motivation that guide supramolecular research. The wide spectrum of scientific research in supramolecular chemistry is strongly supported by crystal structure data. The CSD contains data for more than 1 000 000 crystal structures and provides good insight into inter-
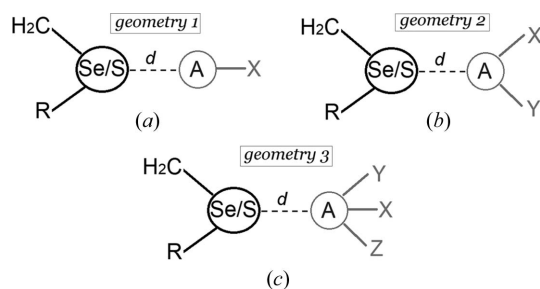

Figure 1
The model systems and geometric parameter *d* of (*a*) geometry 1, (*b*) geometry 2 and (*c*) geometry 3, used for searching the CSD.

Table 1
The results of our statistical analysis of selenium and sulfur interactions in crystal structures.

CSD analysis	Selenium (Se)			Sulfur (S)		
	Geometry 1	Geometry 2	Geometry 3	Geometry 1	Geometry 2	Geometry 3
Number of contacts	7714	829	166	135 882	13 910	5 889
A	H (7114)	Se (432)	Csp^2 (148)	H (125 288)	S (10 066)	Csp^2 (5161)
Type of interaction	C—H···Se (81.7%)	Se···Se (5.0%)	Se···π (1.9%)	C—H···S (80.5%)	S···S (6.5%)	S···π (3.3%)
Total number of contacts	8709			155 681		
R	Csp^2 (3411), Csp^3 (2522), Csp (937), Se (782), M (608), H (53)			Csp^2 (90901), Csp^3 (39824), Csp (1550), S (3725), M (14563), H (2140)		

molecular interactions and information, conformational preferences, drug design and delivery. One other productive database we have used in our research is the PDB. This is the global archive of 3D structural models for many biological macromolecules and their complexes. Most structures in this database were determined using X-ray crystallography, NMR or 3D cryo-electron microscopy (cryo-EM). Our docking results and accompanying information on binding energies were obtained on protein structures derived from the PDB.

3. Results

3.1. CSD analysis of Se and S interactions

Our search of the CSD yielded 552 structures for geometry 1, 1507 structures that correspond to geometry 2 and 2563 structures for geometry 3 (Fig. 1). After simple elimination of all contacts that had Se/S···X, Se/S···Y and Se/S···Z distances shorter than the Se/S···A distance, only 7714 contacts remained in the search results for geometry 1, 829 contacts for geometry 2 and 166 contacts for geometry 3 (Table 1). All parameter distributions used to describe the geometry of the analysed contacts are presented in Fig. 2. The distribution of the parameter d for Se interactions reveals that all contacts from geometry 1 appear in the lower values compared with the other two geometries. The contacts of geometries 2 and 3 show distribution maxima between 3.5 and 3.9 Å, but the contacts of geometry 1 do not show a similar strong tendency to have parameter d in a narrow range of values.

Different trends are observed for the distribution of angles in the P_1 plane (formed by a C–Se–R fragment) with vector V_1 (X–A vector) for geometry 1, in the P_2 plane (formed by an X–A–Y fragment) for geometry 2 or in the P_3 plane (formed by atoms X, Y and Z) for geometry 3. Although the distribution of the d parameter is very similar in geometries 2 and 3, their angle distributions are quite different. Angles between 0° and 20° are attributed to parallel orientation, angles between 70° and 90° are connected with orthogonal orientation, and all remaining angle values (between 20° and 70°) are attributed to slope orientation. The angle distributions (Fig. 2) reveal a strong tendency of all three groups to slope orientations. Additionally, orthogonal orientations are more abundant than parallel ones in the first and third geometries. The contacts from geometry 2 have an equal contribution from both orientations (∼25%).

All important differences in the distances and mutual orientations of all fragments that interact are connected to the natures of those interactions. When the influence of the nature of an atom in position A is considered (Table 1), C—H···Se interactions become the most numerous in group 1 (2859 Csp^2—H···Se contacts, 2294 Csp^3—H···Se contacts and 759 Csp—H···Se contacts). Additionally, the most abundant interactions in group 2 are those between two Se atoms (432 Se···Se contacts) and Se···S (163 contacts), while Se···π interactions are observed in group 3 (148 contacts). Compared with these, σ interactions (contacts with a C—Se···A, R—Se···A, X—A···Se, Y—A···Se or Z—A···Se angle higher than 160°) are rare: 168 interactions in the first group, 182 in the second and only 28 interactions in the third.

Based on statistical results, selenium has a tendency to build hydrogen bonds of the C—H···Se type. These structures are 150 times more numerous than those where classical hydrogen-bond donors (O—H and N—H groups) interact with Se (see the supporting information). When all contacts where the Se···X, Se···Y or Se···Z distances are shorter than the Se···A distance are eliminated, the main database of 2563 structures is reduced to 166 contacts. This obviously indicates that the Se atom prefers to participate in interactions with aryl C—H groups in C—H···Se interactions rather than with π systems in Se···π interactions.

A similar tendency to build supramolecular structures is observed with sulfur derivatives (Table 1). The majority of contacts involve C—H···S interactions (125 288 contacts), S···S interactions (10 065 contacts) and S···π interactions (5161 contacts).

Although our goal was to gain insight into the supramolecular structure for the Cys and Met side residues and their Se derivatives Sec and Mse, we did not restrict the atoms in the R position during the CSD search. In this way we have obtained a more complete view of the substitution of S with Se as a function of the various substituents at the R position. Structures with Csp^2 and Csp^3 atoms in position R form the largest group of contacts (Table 1). An interesting result is the abundance of structures with an Se—Se bond (782 contacts). There are 608 contacts with a metal ion M in the R position and 53 contacts with an H atom in the R position. A statistical analysis of all contacts with an H atom in the R position (corresponding to a Sec side chain) or a Csp^3 atom (corresponding to an Mse side chain) indicated that C—H···Se interactions are many times more numerous than the other types of interaction. The second most common type of inter-

action is the Se···Se interaction. The number and distribution of certain types of interaction in these two groups of contacts are consistent with the overall trend for all contacts from our CSD search.

The most numerous are the structures in which a Csp^2 or Csp^3 atom is bound to sulfur in the R position. Less numerous are structures with a metal ion M (14 563 contacts) or S (3728 contacts) in the R position. Structures with an H atom in the R position (2140 contacts) were also found. Similar to selenium, these structures including S and H atoms (corresponding to a Cys side chain) and a Csp^3 atom (corresponding to a Met side chain) in the R position were analysed separately. A statistical analysis showed that C—H···S interactions are drastically more numerous than the others. For structures with a frag-

ment corresponding to a Met side chain, the S···O interactions (1855 contacts) and S···S interactions (1082 contacts) are the second most abundant type of interaction. In the case of a Cys fragment, the second most common are the S···O contacts, as a consequence of the S—H···O interactions (285 contacts). Interestingly, the number of structures with electronegative atoms (O, N or a halogen element) in the R position is negligible. This suggests that the view of the supramolecular structures we have obtained here refers to structures containing nonpolar groups bonded to an S or Se atom.

The results obtained by a statistical analysis of geometric parameters can be illustrated by analysing packages in the crystal structure of [(4-methoxybenzyl)selanyl] acetic acid
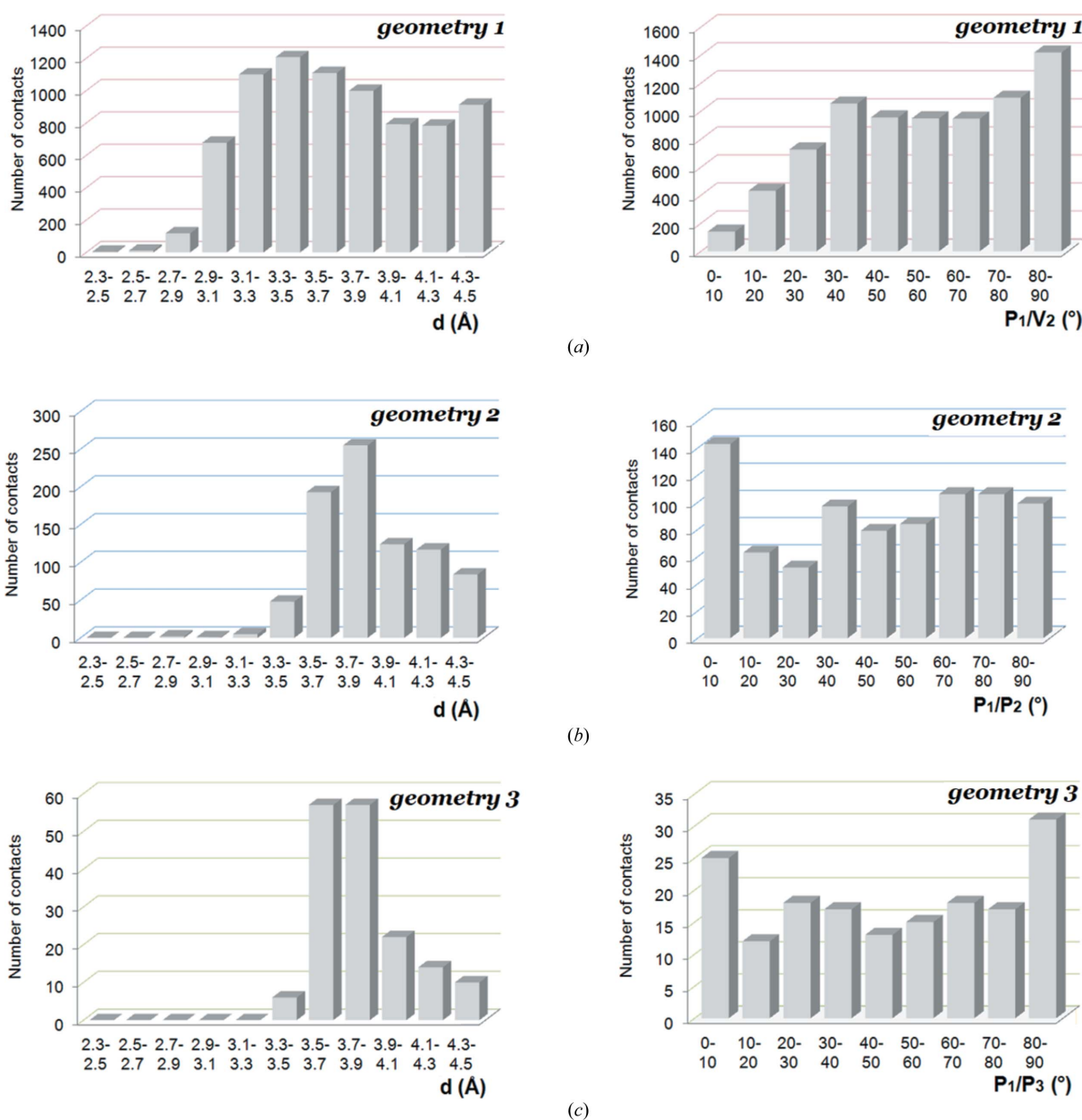

Figure 2
Distributions of d, P_1/V_2, P_1/P_2 and P_1/P_3 geometric parameters of (*a*) geometry 1, (*b*) geometry 2 and (*c*) geometry 3, describing the selenium interactions in crystal structures archived in the CSD.
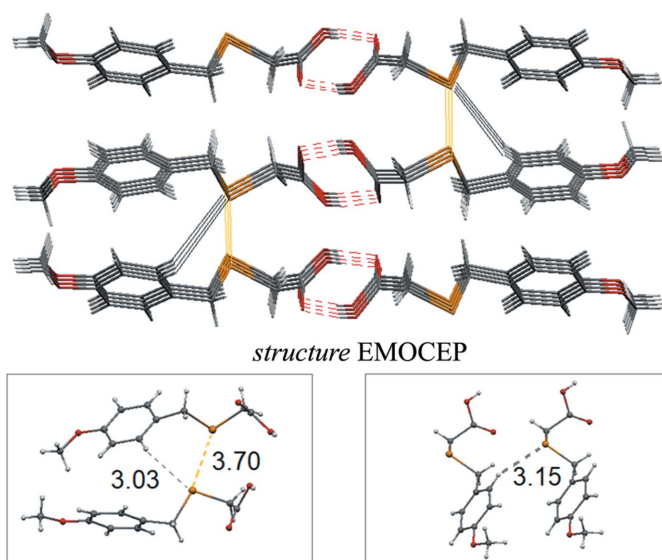
structure EMOCEP

Figure 3
The crystal packing of [(4-methoxybenzyl)selanyl] acetic acid, illustrating the interactions of selenium (red dashed lines) (Bhuyan & Mugesh, 2011). Distances are in Å.

(refcode EMOCEP, Fig. 3). In this structure, the considered molecule contains Se, carboxyl, aryl and methoxy groups. In this crystal structure the main synthon is a dimer that contains bifurcated hydrogen bonds formed by carboxyl groups (red dashed lines in Fig. 3). These dimers are the main components of lamellae, where each molecule of acid interacts differently with the ones under and above it. In one of these orientations, the Se atom participates in a C—H···Se bond with the aryl group of an adjacent acid molecule (the length of this bond is 3.15 Å). In another orientation, acid molecules form interactions that include Se. The acid groups simultaneously form Se···Se interactions (length 3.70 Å) and C—H···Se interactions with an aryl group (length 3.03 Å).

When Se···Se interactions are observed, the angle between two contact C—Se—C planes is 62.3°. Based on this example, it can be concluded that the Se atom prefers to participate in interactions with C—H groups rather than aryl or O—H groups, which are reserved to form π–π aromatic interactions

and classical hydrogen bonds, respectively. It is also very interesting that the Se atom does not interact with alkyl C—H groups, even if they are more numerous than the aryl C—H groups.

3.1.1. Se···Se interactions. These have been treated as a sub-subsection until now and their geometry has not been described in detail. These interactions belong to the second group of contacts. The length distribution of Se···Se interactions is similar to that of contacts retrieved from the CSD search of geometry 2 (the majority of interaction lengths are between 3.5 and 3.9 Å). The main geometric difference is connected with the distribution of the P_1/P_2 angle (Fig. 4). This distribution shows that Se···Se contacts prefer to form interactions with parallel orientations of contact R—Se—C planes.

A typical example of a parallel interaction is the crystal structure with refcode AKANAB (Fig. 5). This structure contains bis(ethylenediseleno)tetrathiafulvalene molecules, where the five-membered ring with two S atoms is condensed with a six-membered ring with two Se atoms. Fulvalene derivatives form lamellae where two different types of packing are found. In one packing, the six-membered rings of neighbouring fulvalenes are in contact through bifurcated C—H···Se interactions (lengths of 2.96 and 3.16 Å). In the second packing, the six-membered rings are placed above the five-membered rings, forming multiple Se···S interactions (lengths between 3.7 and 3.9 Å). Fulvalene derivatives from adjacent lamellae simultaneously form Se···Se and Se···S interactions. In these interactions all the Se···Se distances (3.54 and 3.62 Å) are shorter than the Se···S ones (3.76 and 4.04 Å). All these results point to stronger interactions between Se and another Se atom compared with an S atom. Supporting this, the number of Se···Se contacts is several times higher than the number of Se···S contacts in the CSD (Table 1).

3.2. Quantum-chemical calculations

Our analysis of statistical data on all contacts has pointed to a greater preference of Se and S atoms for hydrophobic regions in order to participate in C—H···Se and C—H···S interactions. The substitution of S with Se does not bring a major change in the supramolecular structure of its deriva-
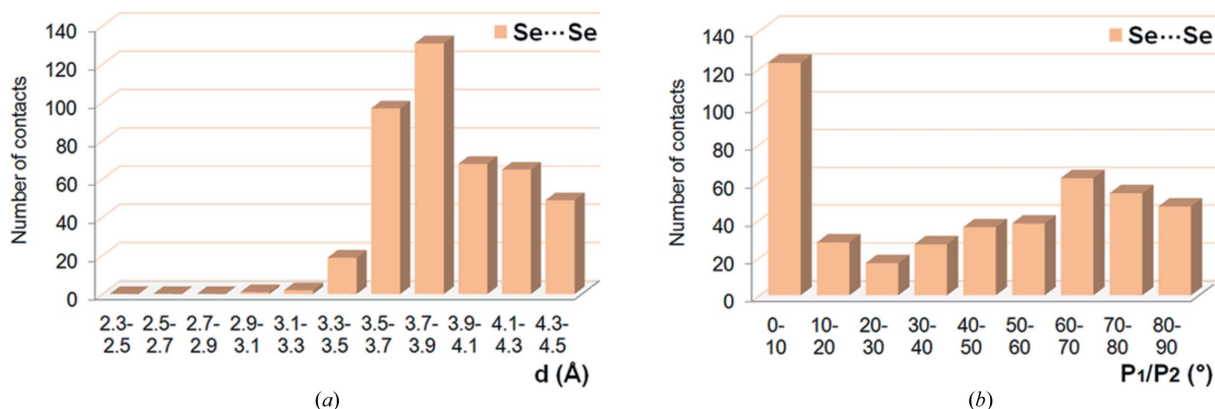

Figure 4
Distributions of (*a*) parameter *d* and (*b*) the P_1/P_2 angle, used for describing Se···Se interactions in crystal structures extracted from the CSD.

tives. The main difference could be expected to be in the strength of their interactions. In some cases, this strength plays a crucial role in the chemistry and function of their compounds. The strengths of hydrogen, parallel and sigma interactions were calculated for both molecules, CH_3SeCH_3 and CH_3SeH. These two molecules represent models for Se derivatives of Cys and Met residues.

3.2.1. Electrostatic potential maps. Se atoms have a region of negative and a region of positive potential (Fig. 6), enabling electrostatic attraction between Se atoms. Similar behaviour is observed with S atoms (Antonijević *et al.*, 2016). Red areas in the electrostatic potential maps indicate an increase in the electron density, while a positive potential is denoted by blue regions.

As can be seen from the maps, there is a positive electrostatic potential area (σ region) at the end of the Se atom along the C—Se and H—Se bond vectors of both molecules, which can act as electron acceptors (yellow and green regions). The most positive potentials are localized on the H atoms. In addition, the maps show a negative region above the Se atom (π hole) in the direction of the free electron pairs. This negative region on one CH_3SeCH_3 or CH_3SeH molecule can be paired with a more positive region on the Se atom of the other molecule through σ–π-hole interactions. The MEPS results show that electrostatic interaction could play an important role in the formation of these σ–π interactions.

Figure 6
Electrostatic potential maps of (*a*) CH_3SeCH_3 and (*b*) CH_3SeH.

3.2.2. Hydrogen bonding. An evaluation of the strength of Se hydrogen bonds was done on eight model systems, numbered HB1–HB8 (Fig. S1 in the supporting information). In the first six model systems (Table 2), CH_4, NH_3 or H_2O molecules were used as donors of H atoms, while CH_3SeCH_3 or CH_3SeH molecules were used as hydrogen acceptors. For each model system (Fig. 7) the CCSD(T)/CBS limit was
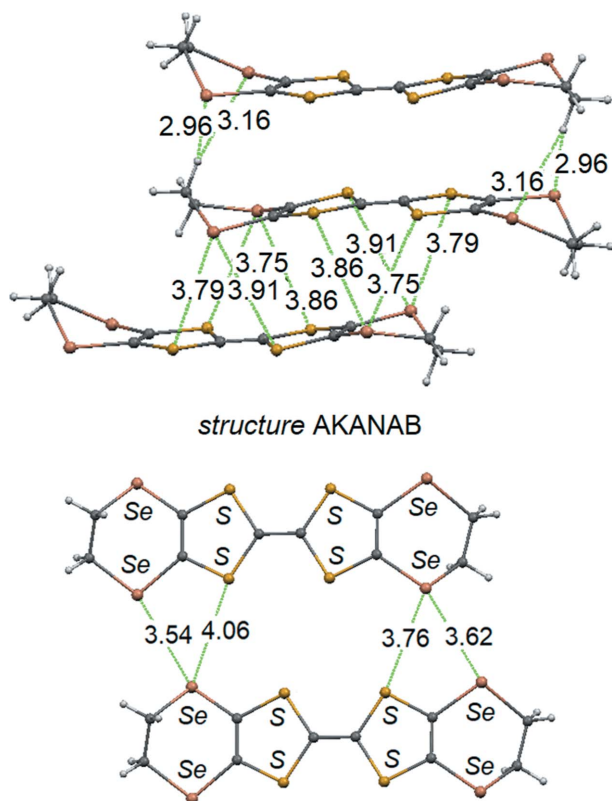

Figure 5
The crystal packing of bis(ethylenediseleno)tetrathiafulvalene, illustrating the Se···Se and Se···S interactions (Clemente-León *et al.*, 2002). Distances are in Å.
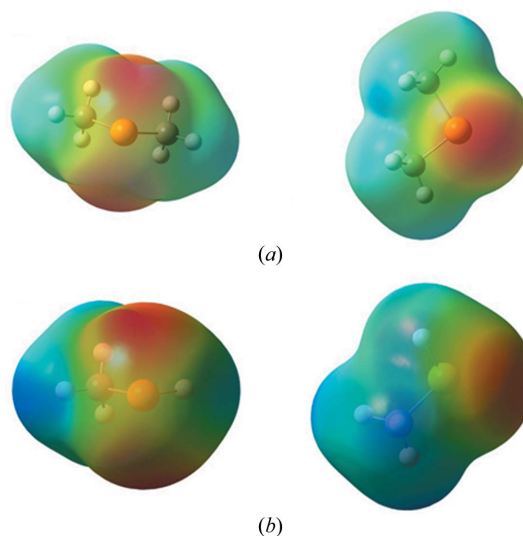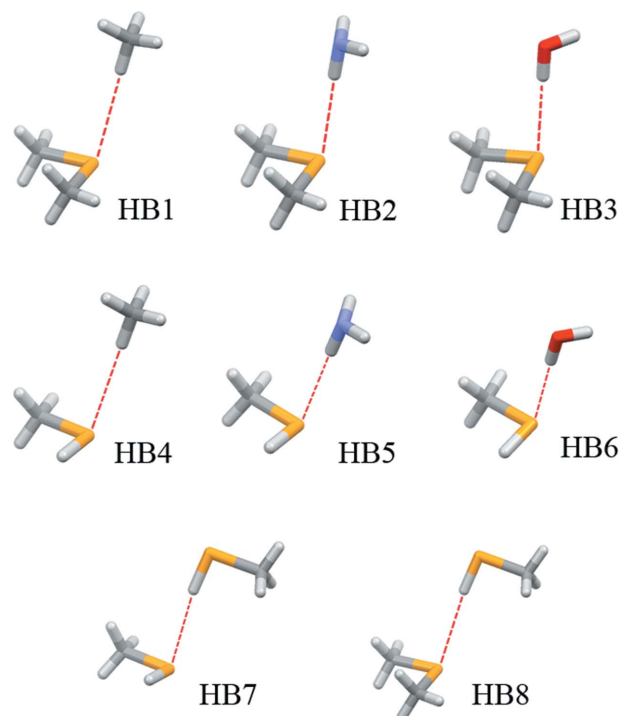

Figure 7
The eight model systems used to evaluate the strength of Se hydrogen bonds: CH_4···CH_3SeCH_3 (HB1), NH_3···CH_3SeCH_3 (HB2), H_2O···CH_3SeCH_3 (HB3), CH_4···CH_3SeH (HB4), NH_3···CH_3SeH (HB5), H_2O···CH_3SeH (HB6), CH_3SeH···CH_3SeH (HB7) and CH_3SeH···CH_3SeCH_3 (HB8).

Table 2
Energies of the Se hydrogen bonds HB1–HB8 (in kcal mol^{-1}), calculated in the procedure for evaluating CCSD(T)/CBS limits and obtained in the benchmark studies.

Functional	Basis set	HB1	HB2	HB3	HB4	HB5	HB6	HB7	HB8
MP2	aug-cc-pVDZ	−0.57	−1.74	−3.57	−0.48	−1.38	−2.81	−1.55	−1.99
	aug-cc-pVTZ	−0.66	−1.93	−3.95	−0.56	−1.57	−3.05	−1.81	−2.25
	aug-cc-pVQZ	−0.69	−1.99	−4.09	−0.60	−1.64	−3.14	−1.90	−2.34
CCSD(T)	aug-cc-pVDZ	−0.56	−1.68	−3.43	−0.46	−1.32	−2.73	−1.27	−1.67
	CBS	−0.84	−2.15	−4.37	−0.79	−1.85	−3.42	−2.08	−2.47
LC-WPBE	aug-cc-pVDZ	−0.68	−2.00	−4.17	−0.58	−1.63	−3.20	−0.60	−0.90
	aug-cc-pVTZ	−0.69	−1.99	−4.13	−0.58	−1.61	−3.16	−0.55	−0.83
	def2-TZVP	−0.71	−2.01	−4.03	−0.60	−1.64	−3.16	−0.58	−0.85
TPSS-D3	aug-cc-pVDZ	−0.85	−2.28	−4.66	−0.74	−1.88	−3.54	−1.99	−2.50
	aug-cc-pVTZ	−0.86	−2.27	−4.63	−0.75	−1.87	−3.52	−1.99	−2.48
	def2-TZVP	−0.87	−2.27	−4.55	−0.75	−1.87	−3.51	−2.00	−2.49
TPSS-D3BJ	aug-cc-pVDZ	−0.86	−2.30	−4.38	−0.74	−1.89	−3.62	−0.82	−1.17
	aug-cc-pVTZ	−0.86	−2.29	−4.35	−0.75	−1.89	−3.60	−0.82	−1.15
	def2-TZVP	−0.87	−2.28	−4.34	−0.75	−1.89	−3.59	−0.83	−1.16
TPSS	aug-cc-pVDZ	−0.20	−1.37	−3.73	−0.15	−1.07	−2.81	−2.12	−2.63
	aug-cc-pVTZ	−0.21	−1.37	−3.70	−0.16	−1.07	−2.79	−2.11	−2.61
	def2-TZVP	−0.22	−1.36	−3.62	−0.17	−1.06	−2.78	−2.12	−2.62
M062X-D3	aug-cc-pVDZ	−0.47	−1.75	−3.97	−0.38	−1.40	−3.07	−1.28	−1.66
	aug-cc-pVTZ	−0.46	−1.72	−3.99	−0.37	−1.38	−3.06	−1.16	−1.52
	def2-TZVP	−0.50	−1.75	−3.93	−0.41	−1.42	−3.07	−1.20	−1.55
CAM-B3LYP-D3	aug-cc-pVDZ	−0.73	−2.17	−4.51	−0.62	−1.79	−3.48	−1.74	−2.18
	aug-cc-pVTZ	−0.72	−2.15	−4.46	−0.62	−1.77	−3.44	−1.70	−2.12
	def2-TZVP	−0.74	−2.14	−4.41	−0.64	−1.77	−3.45	−1.73	−2.14
CAM-B3LYP-D3BJ	aug-cc-pVDZ	−0.74	−2.15	−4.19	−0.63	−1.77	−3.48	−1.82	−2.27
	aug-cc-pVTZ	−0.74	−2.12	−4.15	−0.63	−1.74	−3.44	−1.77	−2.21
	def2-TZVP	−0.75	−2.12	−4.16	−0.63	−1.75	−3.44	−1.80	−2.23
WB97XD	aug-cc-pVDZ	−0.65	−1.94	−4.12	−0.53	−1.55	−3.12	−1.56	−2.01
	aug-cc-pVTZ	−0.66	−1.93	−4.07	−0.55	−1.54	−3.07	−1.51	−1.94
	def2-TZVP	−0.70	−1.95	−3.97	−0.59	−1.57	−3.06	−1.54	−1.96
PBE0	aug-cc-pVDZ	−0.37	−1.62	−4.01	−0.31	−1.31	−3.06	−1.09	−1.30
	aug-cc-pVTZ	−0.37	−1.61	−3.99	−0.31	−1.31	−3.03	−1.06	−1.26
	def2-TZVP	−0.39	−1.62	−3.94	−0.33	−1.32	−3.05	−1.09	−1.28
Geometry d (Å)		3.4	3.0	2.6	3.4	3.0	2.8	3.0	3.0

calculated, with the calculation done on ten levels of theory. To find the level of theory that would reduce the time cost compared with calculations at the CCSD(T)/CBS level, an extensive benchmark study was carried out using a broad set of quantum-chemical methods and basis sets (LC-WPBE, TPSS, TPSS-D3, TPSS-D3BJ, M062X-D3, CAM-B3LYP-D3, CAM-B3LYP-D3BJ, WB97XD and PBE0 functionals, in combination with aug-cc-pVDZ, aug-cc-pVTZ and def2-TZVP basis sets). As the criterion that the results would be accepted as good, the energy values should not vary by more than 10% of the CCSD(T)/CBS energies.

For calculations on the HB1 model system, acceptable results are only given with the TPSS-D3 and TPSS-D3BJ functionals. If we expand the criterion for accuracy of calculations to 20% of the CCSD(T)/CBS energy, then all the functionals are acceptable for this model system, except for those at the TPSS, M062X-D3 or PBE0 levels. Similar results are obtained for the HB2 model system, while for the HB3 model system, in addition to those three mentioned functionals, WB97XD should also be strongly avoided. For calculations based on the HB4 model system, acceptable results are only given with the TPSS-D3 and TPSS-D3BJ functionals.

For the HB5 and HB6 model systems all the functionals are accepted, except for the TPSS, M062X-D3, WB97XD and PBE0 functionals. The best functionals for the HB7 and HB8

model systems are TPSS-D3 and, surprisingly, TPSS, while acceptable results were also obtained with the CAM-B3LYP-D3 and CAM-B3LYP-D3BJ functionals. Functionals PBE0 and M062X-D3 should be avoided because they yielded no accurate results. The interaction energies for the HB7 and HB8 model systems are slightly higher [$\Delta E_{\mathrm{CCSD(T)/CBS}} = -2.08$ and -2.47 kcal mol^{-1}, respectively; 1 kcal mol^{-1} = 4.184 kJ mol^{-1}] than the energies for the corresponding systems with an NH$_3$ molecule as acceptor [$\Delta E_{\mathrm{CCSD(T)/CBS}}$ = -1.85 and -2.15 kcal mol^{-1}, respectively], indicating that CH$_3$SeH is a better H-atom donor than NH$_3$.

Based on the given criteria one can conclude that the TPSS and M062X-D3 functionals should not be used for most of the systems [except for HB7 and HB8, where TPSS provides strong agreement with the CCSD(T)/CBS values]. Also, it can be noted that significantly fewer functionals can be used to calculate weak interactions (in the HB1 and HB4 model systems) than moderate (HB2 and HB5 model systems) or strong interactions (HB3 and HB6 model systems). The energies of the hydrogen bonds increase with increasing hydrogen donor capacity (CH$_4$ < NH$_3$ < CH$_3$SeH < H$_2$O). The results of the calculations show that hydrogen bonds of CH$_4$, NH$_3$ and H$_2$O with CH$_3$SeCH$_3$ [$\Delta E_{\mathrm{CCSD(T)/CBS}}$ = -0.84, -2.15 and -4.37 kcal mol^{-1}, respectively] are stronger than the corresponding hydrogen bonds with CH$_3$SeH as acceptor [$\Delta E_{\mathrm{CCSD(T)/CBS}}$ = -0.79, -1.85 and -3.42 kcal mol^{-1},

Table 3
Interaction energy values (in kcal mol^{-1}) and geometric parameters (in Å) for the most stable parallel and σ–π interactions.

Functional	Basis set	Parallel interactions					σ–π interactions		
		P1	P2	P3	P4	P5	ES1	ES2	ES3
MP2	aug-cc-pVDZ	−0.31	−1.58	−0.91	−1.08	−2.62	−1.12	−1.64	−1.99
	aug-cc-pVTZ	−0.46	−1.87	−1.16	−1.37	−3.01	−1.30	−1.85	−2.32
	aug-cc-pVQZ	−0.51	−1.97	−1.25	−1.46	−3.14	−1.38	−1.95	−2.44
CCSD(T)	aug-cc-pVDZ	−0.38	−1.39	−0.97	−0.81	−2.24	−0.94	−1.39	−1.63
	CBS	−0.74	−2.21	−1.52	−1.75	−3.27	−1.51	−2.06	−2.55
LC-WPBE	aug-cc-pVDZ	−0.40	−1.88	−1.27	−1.30	−2.99	−1.23	−1.79	−2.02
	aug-cc-pVTZ	−0.43	−1.84	−1.23	−1.29	−2.93	−1.11	−1.66	−1.91
	def2-TZVP	−0.40	−1.85	−1.24	−1.26	−2.92	−1.13	−1.68	−1.89
TPSS-D3	aug-cc-pVDZ	−0.51	−1.93	−1.33	−1.37	−2.98	−1.28	−1.85	−2.13
	aug-cc-pVTZ	−0.56	−1.97	−1.36	−1.43	−3.00	−1.25	−1.80	−2.07
	def2-TZVP	−0.51	−1.94	−1.34	−1.37	−2.96	−1.22	−1.79	−2.03
TPSS-D3BJ	aug-cc-pVDZ	−0.64	−2.02	−1.46	−1.54	−3.18	−0.22	−0.42	−0.31
	aug-cc-pVTZ	−0.70	−2.06	−1.48	−1.60	−3.20	−0.19	−0.37	−0.26
	def2-TZVP	−0.65	−2.03	−1.46	−1.54	−3.16	−0.16	−0.36	−0.21
TPSS	aug-cc-pVDZ	0.25	−0.24	0.09	0.63	−0.16	−1.49	−2.06	−2.49
	aug-cc-pVTZ	0.19	−0.27	0.07	0.57	−0.17	−1.46	−2.01	−2.44
	def2-TZVP	0.25	−0.24	0.09	0.63	−0.14	−1.43	−2.01	−2.39
M062X-D3	aug-cc-pVDZ	−0.27	−1.81	−1.11	−1.17	−2.93	−1.14	−1.71	−2.09
	aug-cc-pVTZ	−0.28	−1.66	−1.02	−1.07	−2.75	−0.91	−1.48	−1.88
	def2-TZVP	−0.24	−1.69	−1.03	−1.04	−2.75	−0.96	−1.49	−1.87
CAM-B3LYP-D3	aug-cc-pVDZ	−0.31	−1.71	−1.08	−1.08	−2.71	−1.15	−1.70	−1.88
	aug-cc-pVTZ	−0.35	−1.74	−1.10	−1.13	−2.73	−1.13	−1.67	−1.85
	def2-TZVP	−0.30	−1.74	−1.10	−1.09	−2.72	−1.13	−1.68	−1.84
CAM-B3LYP-D3BJ	aug-cc-pVDZ	−0.51	−1.88	−1.29	−1.37	−2.97	−1.31	−1.87	−2.17
	aug-cc-pVTZ	−0.56	−1.97	−1.31	−1.42	−3.05	−1.29	−1.84	−2.14
	def2-TZVP	−0.53	−1.96	−1.44	−1.38	−3.02	−1.29	−1.86	−2.13
WB97XD	aug-cc-pVDZ	−0.40	−1.75	−1.17	−1.20	−2.73	−1.17	−1.67	−1.92
	aug-cc-pVTZ	−0.42	−1.71	−1.13	−1.18	−2.68	−1.02	−1.53	−1.77
	def2-TZVP	−0.40	−1.73	−1.14	−1.16	−2.69	−1.05	−1.56	−1.77
PBE0	aug-cc-pVDZ	0.36	−0.74	0.15	0.22	−0.80	−1.48	−2.02	−2.43
	aug-cc-pVTZ	0.30	−0.77	0.11	0.17	−0.82	−1.43	−1.96	−2.37
	def2-TZVP	0.35	−0.74	0.13	0.22	−0.79	−1.42	−1.97	−2.33
Geometry, r (Å), R (Å), d (Å)		$r = 0$, $R = 4.4$	$r = 0$, $R = 4.2$	$r = 0$, $R = 4.4$	$r = 0$, $R = 4.2$	$r = 0$, $R = 4.0$	$d = 4.0$	$d = 4.0$	$d = 3.8$

respectively]. The calculations also show that CH$_3$SeCH$_3$ is a better acceptor of hydrogen bonding than CH$_3$SeH.

All these calculations offered us the possibility of also using MP2/aug-cc-pVTZ and MP2/aug-cc-pVQZ because of their good agreement with the CCSD(T)/CBS level for the majority
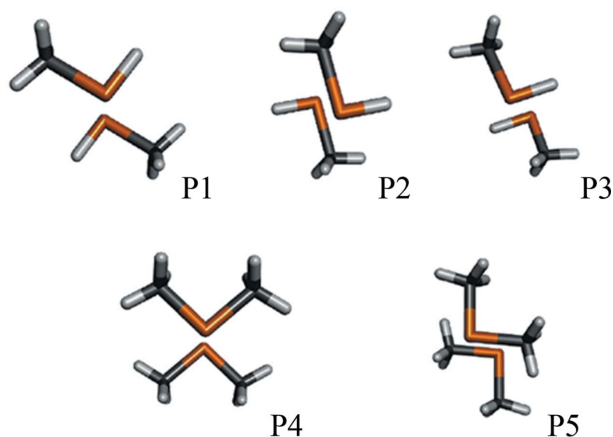

Figure 8
The model systems of methaneselenol dimers (P1, P2 and P3 model systems) and dimethyl selenide dimers (P4 and P5 model systems) for parallel interactions.

of HB models. However, these levels of calculation are time costly so we do not recommend them.

3.2.3. Parallel interactions. The results of our statistical analysis of crystal structures indicate that Se$\cdots$Se interactions with parallel orientation of the C—Se—R planes are the most numerous. The calculations were performed on five model systems (Fig. 8): three model systems with methane–selenol dimers (P1, P2 and P3 model systems) and two model systems with dimethylselenide dimers (P4 and P5 model systems). Scheme 1 shows the parameters describing the geometries of parallel interactions in the studied Se derivatives, and the directions of molecular movement during the examination of the strength of the parallel interactions. In the P1 and P4 model systems, the Se atom of the first molecule is moved along direction 1 containing the height of the C—Se—X triangle (X = C or H) of the second molecule (Scheme 1). In other systems, the Se atom is moved along direction 2 passing through the centre of the Se—H (P3 model system) or Se—C bond (P2 and P5 model systems) (Scheme 1). The starting point for moving is the Se atom of the P1 and P4 model systems, the centre of the Se—C bond for the P2 and P5 model systems, or the centre of the Se—hydrogen bond for the P3 model system. The results of interaction energy calculations for different values of the parameter r are presented in the supporting information, Fig. S2
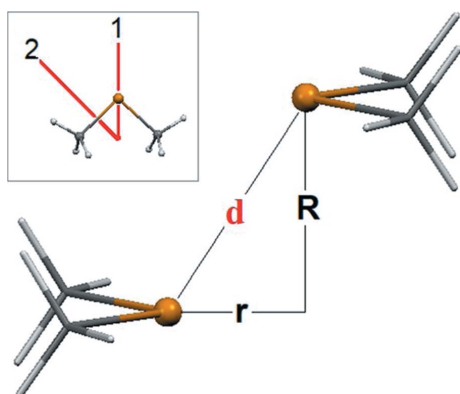
Scheme 1

Our calculations have shown that the movement of parallel fragments from the starting point leads to a reduction in interaction energies. Orientations with the Se atom above the centre of the C—Se bond (P2 and P5 model systems) are more stable than those of the corresponding species with the Se atom above the centre of the Se—H bond (P3 model system). The least stable orientations (P1 and P4 model systems) have the Se atom located above the Se atom of the second molecule (Table 3). The energy of the most stable parallel orientation of two dimethylselenide molecules is considerably greater [$\Delta E_{CCSD(T)/CBS}$ = −3.27 kcal mol^{-1}] than the energy of the most stable parallel interaction between two methaneselenols [$\Delta E_{CCSD(T)/CBS}$ = −2.21 kcal mol^{-1}]. The results of these calculations show that the substitution of an H atom by an alkyl group leads to a strengthening of Se···Se interactions with parallel orientation.

The benchmark study revealed that the TPSS-D3 and CAM-B3LYP-D3BJ functionals are suitable to obtain a good match with calculated energy values at the CCSD(T)/CBS level (Table 3). To get excellent matches, a useful method is TPSS-D3BJ, while the rest are not recommended.

The analysis of the interactions of cysteine residues (R—CH$_2$SH) in the crystal structures from the CSD showed that S···S interactions with a parallel orientation of the C—S—H planes are the most numerous (Antonijević *et al.*, 2016). Quantum-chemical calculations at the CCSD(T)/CBS level were performed on model systems of methanethiol dimers with parallel orientations and with starting geometries similar to the geometries presented here. In the most stable geometry of a methanethiol dimer with parallel orientation [$\Delta E_{CCSD(T)/CBS}$ = −1.80 kcal mol^{-1}], the S atom is located

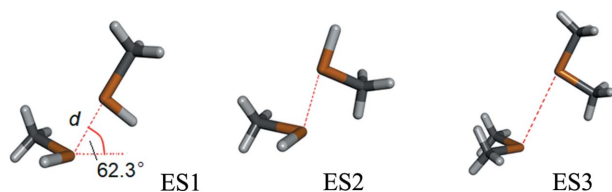above the centre of the S—C bond (Antonijević *et al.*, 2016). Based on these findings and on our calculations, it can be concluded that the replacement of an S atom with Se leads to an increase in the strength of parallel interactions.

3.2.4. σ–π-hole interactions. Model systems used for investigating σ–π interactions (electrostatic models, ES) are based on dimers of CH$_3$SeCH$_3$ or CH$_3$SeH (Fig. 9). The main parameter used to describe the geometries of these interactions is the distance between the Se atoms (d). The angle between the C—Se vector (or H—Se vector) and the C—Se—X plane of the second molecule (X = C or H) is constant and takes a value of 62.3°. During the calculation the distance d was changed, while the value of the C—Se···Se angle (ES1 and ES3 model systems) or H—Se···Se angle (ES2 model system) was kept at 180°. Interaction energies are calculated for all geometries with parameter d between 3.4 and 4.5 Å. Results for different values of parameter d are offered as part of the supporting information, Fig. S3.

The interaction energies calculated at the CCSD (T)/CBS level for each model system minimum are presented in Table 3. The whole benchmark study for three model systems (ES1, ES2 and ES3) is also presented in Table 3. The results of quantum-chemical calculations show that the σ–π-hole (Se···Se) interaction is strongest in the case of the model system with two CH$_3$SeCH$_3$ molecules [ES3, $\Delta E_{CCSD(T)}$ = −2.55 kcal mol^{-1}]. This interaction also has the shortest distance between Se atoms (d = 3.8 Å); this distance corresponds to the sum of the Se van der Waals radii (1.9 Å). Systems with CH$_3$SeH molecules (ES1 and ES2) have weaker interactions [$\Delta E_{CCSD(T)}$ = −1.51 and −2.06 kcal mol^{-1}, respectively].

The strongest S···S interaction for an electrostatic model with a methanethiol dimer (−2.20 kcal mol^{-1}) (Antonijević *et al.*, 2016) is slightly higher than the strongest Se···Se interaction for an electrostatic model with methaneselenol dimers (−2.06 kcal mol^{-1}), suggesting a slightly greater tendency of sulfur to form σ–π interactions.

The benchmark study revealed the functionals that are suitable to obtain good matches with the calculated CCSD(T)/ CBS limits (the TPS and PBE0 functionals). To get very good matches, useful ones are CAM-B3LYP-D3BJ and MP2, while the rest should be avoided. In addition, the more positive effect of including empirical dispersions (D3 and D3BJ) on the matching with limits can not be denied.

3.3. Docking study

The effect of substitution of S by Se on interactions with biosystems will be illustrated through several examples, and the obtained results will be analysed by molecular docking. Molecular docking is a very important tool when it comes to drug discovery and design. It is used to model the interactions between a small molecule and a protein at the atomic level, and it allows us to characterize the behaviour of small molecules on the binding sites of target proteins. It consists of prediction of the ligand's conformation, position and orientation within these sites and assessment of the binding affinity.

Figure 9
The model systems of methaneselenol dimers (ES1 and ES2 model systems) and dimethyl selenide dimer (ES3 model system) for σ–π-hole interactions.

research papers

One important advantage of molecular docking is the possibility of interpreting the substitution of S by Se in all those cases where there are no direct interactions between amino acid residues and those two atoms. Rather, the interactions are indirect, via groups attached to them. When a ligand binds to a biomolecule, it can locate in such a manner that the S and Se atoms are oriented away from the biomolecule and therefore not participating in interactions with amino acid residues. In contrast, in crystal structures generally, crystal packing rules dictate that all the fragments present should participate in interactions with different species from the surrounding environment, although there is a possibility that fragments form only intramolecular interactions and approaching species in the wider environment are sterically hindered.

Validation of the docking procedure by the *AutoDock Vina* program was confirmed by redocking of ITU (iodotubercidin) at the iNOS enzyme (inducible nitric oxide synthase), extracted from the crystal structure with PDB code 4nos (Fischmann *et al.*, 1999). More details are available in the supporting information (Fig. S4). The docking study indicates that ITU (a well known inhibitor for 4nos) binds on the same site as in the crystal structure (Fig. S4), in the narrow cleft inside the larger active-site cavity that also contains a haem complex. The cleaned enzyme structure was further docked with four different compounds, corresponding to selenium and sulfur analogues (PBIT, PBISe, PEIT and PEISe, Fig. 10). The statistical analysis indicated that crystal structures with a Csp^2 atom in the *R* position are the most numerous (Table 1). This residue is also present in the above-mentioned structures, implying their good fitting qualities for the chosen model for our CSD search. PBIT is a well known inhibitor for iNOS, implicated as a potential therapeutic target in malignant melanoma, the most deadly form of skin cancer (Madhunapantula *et al.*, 2008). However, PBIT requires high concentrations for clinical efficacy and at the same time causes systemic toxicity. PBISe is its selenium derivative and a more potent agent, effective at significantly lower concentrations (Madhunapantula *et al.*, 2008). The monosubstituted forms of PBIT and PBISe (PEIT and PEISe) were also examined (Madhunapantula *et al.*, 2008).

Our results from this docking study suggested that PBISe could be the best targeting inhibitor for the active site of iNOS (Fig. 11). This i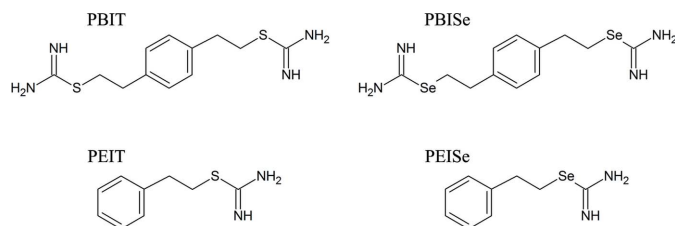s in strong agreement with experimental results showing that PBISe kills melanoma cells more than ten times more effectively than other iNOS inhibitors (PBIT, PEISe and PEIT) (Madhunapantula *et al.*, 2008). All obtained orientations of PBISe are clustered at the same site (active site, position 1, Fig. 11), leading to the much better efficiency of this molecule compared with PBIT, PEISe and PEIT where numerous binding sites can be observed.

The influence of the nature of surrounding amino acids was investigated by analysing the most stable orientation of each ligand (PEISe, PEIT, PBISe and PBIT) for all binding sites. In addition, all observed orientations with PEISe and PBISe (containing Se atoms) were considered as one group and those with PEIT and PBIT (containing S atoms) as the other. From the results of this statistical analysis (Figs. S5–S15) it is evident that the amino acid environment of the presented compounds on their binding sites is non-polar in both cases and independent of the type of atom (sulfur or selenium). By comparison, all interactions with negative amino acids are less abundant. We observed a significantly larger number of these interactions with clusters that contain Se atoms compared with those with S atoms (the same tendency as observed with non-polar acids). One possible explanation for this trend is the larger radius of selenium compared with sulfur, leading to better electron correlation for contacts of Se derivatives with negative amino acid residues. The analysed compounds achieved the fewest contacts with polar and positive amino acid residues. In all the investigated systems there is no clear evidence of a direct interaction between an amino acid residue and an S or Se atom. Differences in the binding affinities of the analysed compounds (determined by the number of binding sites, their position and their binding energy) are attributed to electron-induced effects, caused by replacement of an S with an Se atom. The results show that there is an effect of the substitution of S by Se on the activity of compounds that
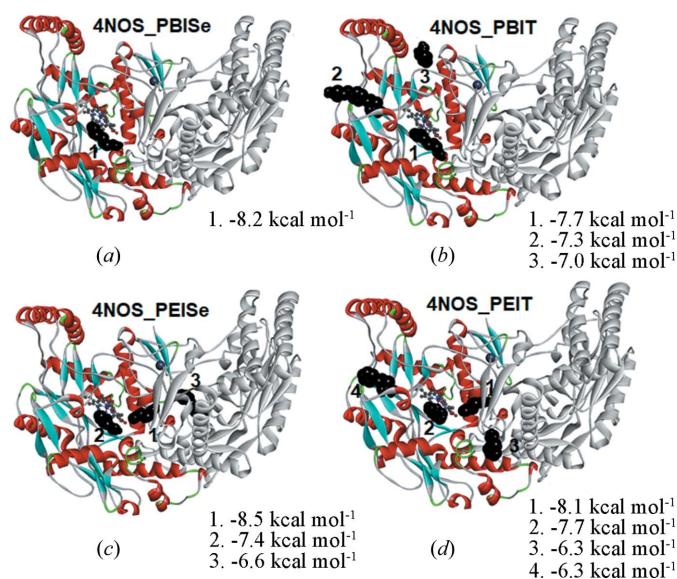

Figure 11
The binding sites and corresponding binding energies of (*a*) PBISe, (*b*) PBIT, (*c*) PEISe and (*d*) PEIT on the iNOS enzyme.

Figure 10
The chemical structures of the active S and Se derivatives, PBIT, PBISe, PEIT and PEISe, used for the docking study of the iNOS enzyme (Madhunapantula *et al.*, 2008).

contain S and/or Se, even when there are no direct interactions of these atoms with amino acid residues (Figs. S5–S15). The investigation of this effect cannot be conducted based on crystal structure analysis, since these atoms in the crystal structures have interacted with some chemical species in their vicinity, as a consequence of the crystal packing. Docking allows the binding of molecules and biosystems, but not all fragments of a molecule may be involved in interactions with the biomolecule.

Similar results were obtained for docking studies that involved eurothiocin A and its Se derivative (Se-eurothiocin A) which, according to the structures of their S and Se centres, are similar to the Met and Mse residues. Specifically, it was shown that the sulfur-containing benzofuran derivative (eurothiocin A), exhibits more potent α-glucosidase inhibitory effects than the clinical inhibitor acarbose (Liu et $al.$, 2014). Eurothiocin A and Se-eurothiocin A are docked at the structure of α-glucosidase. The results obtained by this docking study have shown that both compounds have a tendency to bind on the enzyme active site (located near the catalytic residues Glu277 and Asp352, Fig. 12). Eurothiocin A has a slightly lower binding energy (-7.3 kcal mol^{-1}) than the corresponding Se derivative (-7.4 kcal mol^{-1}). As neither S nor Se atoms participate directly in interactions with amino acid residues on the most stable binding site (Fig. 12), the main differences in binding energies could also be attributed to electron-induced effects. The Se derivative also interacts with a larger number of negative amino acid residues compared with the S derivative, as in the previous system.

To examine whether the substitution of S by Se in drugs which contain a fragment that corresponds to the side chains of Cys or Sec has any effect on the supramolecular structures of these derivatives with biosystems, docking studies of captopril and its Se derivative (Se-captopril) on angiotensin I-converting enzyme (ACE) were undertaken. ACE is a zinc-dependent dipeptidyl carboxypeptidase, and a well known target for the treatment of hypertension and related cardiovascular diseases. Captopril is a well known clinical inhibitor

of ACE, but it is known that the selenium analogue (Se-captopril) also inhibits ACE (Bhuyan & Mugesh, 2011). These two compounds have an SH or SeH group (monosubstituted S/Se atom) bonded to a non-polar chain (Fig. 13). This is probably the reason why, in this case, the S or Se atom interacts with an amino acid residue (His367) and a Zn^{2+} ion from the active site of the enzyme (Fig. S16). In the most stable orientations, captopril and Se-captopril bind at the active site of ACE, in an analogous fashion to that observed in the crystal structures of the captopril/ACE (PDB code 2x8z; Akif et $al.$, 2010) and Se-captopril/ACE complexes (PDB code 3zqz; Akif et $al.$, 2011). These two structures present the zinc coordinating to the sulfur- and selenium-based ACE complexes, with significant antioxidant activity. Se-captopril is an excellent inhibitor of ACE, although its IC$_{50}$ value is twice as high as that for captopril (Bhuyan & Mugesh, 2011). The results of the docking studies have shown that these two compounds bind at the active site of the enzyme with similar binding energies (-5.8 and -5.7 kcal mol^{-1}), indicating that the coordination of these compounds is a decisive factor in their different activities. Although the formation of a selenol–Zn complex is slightly favoured over the formation of a thiolate–Zn complex (Pearson, 1990; Salter et $al.$, 2005), the possible reason for the relatively high IC$_{50}$ value of Se-captopril is partial oxidation of the selenol to the corresponding diselenide derivative. This is further supported by the assumption that diselenides do not inhibit enzyme activity (Bhuyan & Mugesh, 2011) and by the results of molecular docking which showed that the disulfide and diselenide derivatives of captopril and Se-captopril are bonded on the active site of ACE (Fig. S17), although in this case a carboxyl group is in contact with a Zn^{2+} ion, but not an S or Se atom.

The docking results indicate that the Cys and Sec fragments of the above-mentioned drugs have a tendency to interact with the same amino acid residues of the ACE enzyme (Fig. S16). This is consistent with the results of our CSD analysis, which showed that these two fragments have a similar tendency for certain noncovalent interactions.
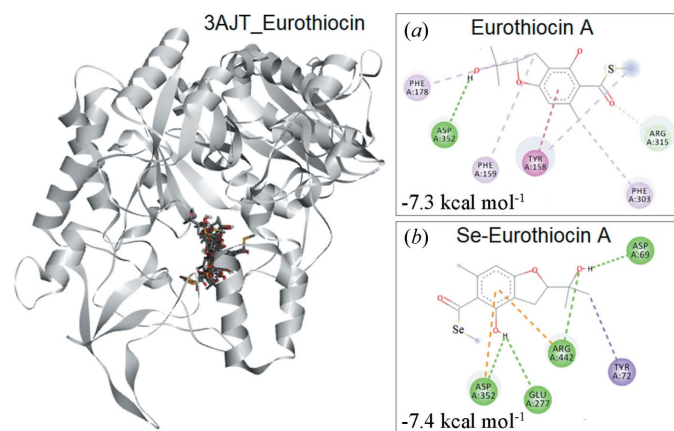

Figure 12
The binding sites of (a) eurothiocin A and (b) Se-eurothiocin A on α-glucosidase (PDB code 3aj7; Yamamoto et $al.$, 2010), and the distributions of amino acid residues for the most stable orientations.
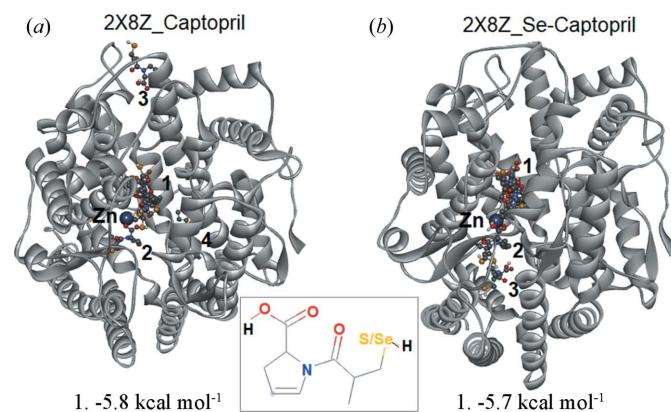

Figure 13
The binding sites and binding energies of (a) captopril and (b) Se-captopril on the active site of angiotensin I-converting enzyme (ACE) (PDB code 2x8z; Bhuyan & Mugesh, 2011).

research papers

4. Conclusions

Based on statistical data analysis of a large number of crystal structures, sulfur and selenium display similar tendencies towards specified interactions. Our crystallographic analysis reveals that the most numerous are structures with C—H$\cdots$Se and C—H$\cdots$S interactions ($\sim$80%). Notably fewer analysed structures exhibit Se$\cdots$Se and S$\cdots$S interactions ($\sim$5%), while Se$\cdots\pi$ and S$\cdots\pi$ interactions are the least numerous.

These results also indicate that both C—H$\cdots$Se and C—H$\cdots$S interactions are weaker than parallel, Se$\cdots\pi$ (Senćanski *et al.*, 2017) and electrostatic σ–π-type interactions. Although C—H$\cdots$Se and C—H$\cdots$S interactions are weaker than the rest, their significant presence can be explained because they accommodate a very large number of CH groups compared with the numbers of Se and S atoms. In addition, all substituents bonded with Se and S atoms influence the crystal packing of these molecules and have a tendency to reduce possibilities for Se$\cdots$Se and S$\cdots$S contacts. This can also explain why O—H$\cdots$Se ($\sim$ −4.4 kcal mol^{-1}) and N—H$\cdots$Se interactions ($\sim$ −2.2 kcal mol^{-1}) are less present than C—H$\cdots$Se interactions ($\sim$ −0.8 kcal mol^{-1}).

The docking results revealed that both S and Se atoms rarely participate in interactions with amino acid residues, mostly because those residues preferentially interact with groups that are bonded to Se or S atoms. Also, in docking, all ligands are bonded on the surface and they are not completely surrounded by amino acids. That is another reason why they do not participate in the mentioned interactions. However, crystal packing dictates demands the densest packing of all species in a crystal. This is the main explanation of why interactions with sulfur and selenium are so numerous in crystal structures. Our statistical results reveal that, in the group of 769 structures that contain Se in our CSD search fragment (Fig. 1), interactions are present in 552 structures. A similar tendency is present taking an S atom into account: in the group of 19 355 structures with this search fragment in the CSD, interactions are present in 4090.

Analysing all groups bonded to an S or Se atom in the ligands and the docking results, there is no meaningful difference in the bonding of their derivatives if S or Se atoms are substituted by non-polar residues. This is the main reason why the two ligands eurothiocin A and Se-eurothiocin A bind to the active site in α-glucosidase with similar binding energies (−7.3 and −7.4 kcal mol^{-1}). All collected results were connected with data obtained from the CSD. There is no difference in trends for the sulfur and selenium compounds. In the CSD search fragments, sulfur or selenium are bonded to CH$_2$ and R groups. In the R position only Csp^3 and sp^2 atoms are mainly found (non-polar groups). It is evident that the search results only offer a better insight into compounds that bind only non-polar residues. Sulfur and selenium derivatives of N,N-diethylcarbamylcholine have shown similar and only marginal inhibitory activity towards electric eel acetylcholinesterase (0.2–0.3 m*M*) (Lindgren *et al.*, 1985). Residues directly bonded to an S or Se atom in this case (alkyl and carbonyl group) do not have the ability to make classical hydrogen bonds as an H-atom donor (Fig. S18). The nature of directly bonded residues at carbamyl compounds (Se and S derivatives of 1-naphthyl-*N*-methyl carbamates, Fig. S18) is similar to previous carbamyl derivatives, which is probably the reason why these two derivatives have similar inhibitory activity towards electric eel acetylcholinesterase (0.2–0.3 m*M*) (Lindgren *et al.*, 1985).

It is important to point out that the top docking position is conditioned by several different criteria, among them the lowest binding energy and the number of ligand interactions with other sites of the enzyme. If polar groups, suitable for interaction with other sites of the enzyme (possible donors of a hydrogen bond), are bonded to an S or Se atom in the R position, the number and positions of possible binding sites and their binding energies could vary. This difference is more pronounced if there is more than one S or Se atom in the ligand. Both PEIT and PEISe have only one S and one Se atom, respectively, substituted with one non-polar and one polar (amidine) group (Fig. 10). This amidine group has the possibility of being a donor for hydrogen bonding. PEIT has four binding sites for iNOS, while its selenium derivative PEISe has only three, but the latter's binding energy to the active site is slightly higher compared with PEIT (−8.5 kcal mol^{-1} for PEISe and −8.1 kcal mol^{-1} for PEIT). As a ligand with two Se atoms, PBISe has only one binding site, while its sulfur derivative PBIT has three. This is the major reason why PBISe is the best inhibitor for the iNOS enzyme (only one binding site and a notable binding energy compared with all the others).

The theoretical conclusions from our docking study on the iNOS enzyme have shown that inhibitory activity rises with increasing number of S atoms, and that the replacement of S with Se also leads to an increase in inhibitory activity. Although the mechanism of inhibition of iNOS is based on non-covalent interactions, this effect is observed for inhibition in which the mechanism is based on the reaction of corresponding S and Se inhibitors. In particular, one *in vitro* study considered the capacity of S- and Se-modified phenols from hydroxytyrosol to inhibit lipid peroxidation (Rodríguez-Gutiérrez *et al.*, 2019). The sulfur compound with the strongest ability to inhibit lipid peroxidation (Fig. S19, compound 2) at the lowest concentration is the disulfide hydroxytyrosol derivative (percentage inhibition 50.78%). This derivative is more than twice as good an inhibitor as a similar compound with only one S atom (Fig. S19, compound 1, percentage inhibition 21.15%). Even better results were obtained in hydroxytyrosol, where two S atoms are replaced with Se atoms (Fig. S19, compound 3, percentage inhibition 80.95%). The study of the antioxidant effects of sulfur and selenium derivatives against disease is focused on several mechanisms, including radical scavenging, enzymes and metal-binding antioxidant mechanisms (Battin & Brumaghim, 2009). Therefore, the activity of these compounds will largely depend on their molecular structure, because that determines which of the above-mentioned mechanisms will have the greatest influence on their activity. In the case of lipid peroxidation, the inhibition by S and Se compounds is most likely attributed to their non-

enzymatic antioxidant capacity. Given how molecular recognition and supramolecular structure have an important role in this, it is clear why conclusions about the activities of compounds, obtained from non-covalent models, can be applied to the understanding of reaction mechanisms.

There is one group of derivatives (monosubstituted derivatives) that express little possibility of interaction between S or Se atoms and amino acid residues. A typical example of this group is the binding of captopril and Se-captopril to the active site of angiotensin I-converting enzyme (ACE). These ligands have the same binding site and similar binding energies (-5.8 kcal mol^{-1} for captopril and -5.7 kcal mol^{-1} for Se-captopril) as they both possess a non-polar fragment directly bonded to an S or Se atom. However, the activity of these compounds demands the existence of Zn—S and Zn—Se coordination bonds and that leads to a higher chemical affinity for sulfur. This explains why captopril is a better inhibitor than Se-captopril.

To conclude, the replacement of an S with an Se atom does not provide significant changes in molecular structure (defined only by the types of interaction). On the other hand, small changes in the strength of those interactions cannot be ignored. The nature of the group directly bonded to these two atoms can provoke differences in binding to an enzyme or biosystems and leads directly to their different activities.

Acknowledgements

Funding information

References

Akif, M., Georgiadis, D., Mahajan, A., Dive, V., Sturrock, E. D., Isaac, R. E. & Acharya, K. R. (2010). *J. Mol. Biol.* **400**, 502–517.
Akif, M., Masuyer, G., Schwager, S. L., Bhuyan, B. J., Mugesh, G., Isaac, R. E., Sturrock, E. D. & Acharya, K. R. (2011). *FEBS J.* **278**, 3644–3650.
Allmang, C. & Krol, A. (2006). *Biochimie*, **88**, 1561–1571.
Antonijević, I. S., Janjić, G. V., Milčić, M. K. & Zarić, S. D. (2016). *Cryst. Growth Des.* **16**, 632–639.
Battin, E. E. & Brumaghim, J. L. (2009). *Cell Biochem. Biophys.* **55**, 1–23.
Bauzá, A. & Frontera, A. (2015). *Angew. Chem. Int. Ed.* **54**, 7340–7343.
Bauzá, A. & Frontera, A. (2018). *Molecules*, **23**, 1770–1781.
Bauzá, A., Mooibroek, T. J. & Frontera, A. (2013). *Angew. Chem. Int. Ed.* **52**, 12317–12321.
Benz, S., Poblador-Bahamonde, A. I., Low-Ders, N. & Matile, S. (2018). *Angew. Chem. Int. Ed.* **57**, 5408–5412.
Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.
Bhuyan, B. J. & Mugesh, G. (2011). *Org. Biomol. Chem.* **9**, 1356–1365.
Bleiholder, C., Gleiter, R., Werz, D. B. & Köppel, H. (2007). *Inorg. Chem.* **46**, 2249–2260.
Boys, S. F. & Bernardi, F. (1970). *Mol. Phys.* **19**, 553–566.
Brammer, L. (2017). *Faraday Discuss.* **203**, 485–507.
Breugst, M., von der Heiden, D. & Schmauck, J. (2017). *Synthesis*, **4**, 3224–3236.
Bruno, I. J., Cole, J. C., Edgington, P. R., Kessler, M., Macrae, C. F., McCabe, P., Pearson, J. & Taylor, R. (2002). *Acta Cryst.* B**58**, 389–397.
Caracelli, I., Zukerman-Schpector, J. & Tiekink, E. R. (2012). *Coord. Chem. Rev.* **256**, 412–438.
Chawla, M., Poater, A., Besalú-Sala, P., Kalra, K., Oliva, R. & Cavallo, L. (2018). *Phys. Chem. Chem. Phys.* **20**, 7676–7685.
Clemente-León, M., Coronado, E., Galán-Mascarós, J. R., Giménez-Saiz, C., Gómez-Garciá, C. J., Fabre, J. M., Mousdis, G. A. & Papavassiliou, G. C. (2002). *J. Solid State Chem.* **168**, 616–625.
Das, B., Chakraborty, A. & Chakraborty, S. (2017). *Comput. Theor. Chem.* **1102**, 127–138.
Daze, K. & Hof, F. (2016). *Molecular Interaction and Recognition.* In *Encyclopedia of Physical Organic Chemistry*, edited by Z. Wang, U. Wille & E. Juaristi, pp. 1528–1580. New York: John Wiley & Sons Inc.
Fischmann, T. O., Hruza, A., Niu, X. D., Fossetta, J. D., Lunn, C. A., Dolphin, E., Prongay, A. J., Reichert, P., Lundell, D. J., Narula, S. K. & Weber, P. C. (1999). *Nat. Struct. Biol.* **6**, 233–242.
Frisch, M. J. *et al.* (2009). *GAUSSIAN09.* Gaussian Inc, Wallingford, Connecticut, USA.
Frontera, A. & Bauzá, A. (2017). *Phys. Chem. Chem. Phys.* **19**, 30063–30068.
Groom, C. R., Bruno, I. J., Lightfoot, M. P. & Ward, S. C. (2016). *Acta Cryst.* B**72**, 171–179.
Hartman, I., Raia, C. A. & Zauhar, R. J. (2006). *Biopolymers*, **83**, 595–613.
Hope, H. (1965). *Acta Cryst.* **18**, 259–264.
Ibrahim, M. A. A. & Safy, M. E. A. (2019). *Phosphorus Sulfur Silicon*, **194**, 444–454.
Iwaoka, M., Katsuda, T., Tomoda, S., Harada, J. & Ogawa, K. (2002*b*). *Chem. Lett.* **31**, 518–519.
Iwaoka, M., Komatsu, H., Katsuda, T. & Tomoda, S. (2002*a*). *J. Am. Chem. Soc.* **124**, 1902–1909.
Iwaoka, M., Komatsu, H., Katsuda, T. & Tomoda, S. (2004). *J. Am. Chem. Soc.* **126**, 5309–5317.
Iwaoka, M. & Tomoda, S. (1994). *J. Am. Chem. Soc.* **116**, 4463–4464.
Iwaoka, M. & Tomoda, S. (1996). *J. Am. Chem. Soc.* **118**, 8077–8084.
Komatsu, H., Iwaoka, M. & Tomoda, S. (1999). *Chem. Commun.* pp. 205–206.
Lehn, J. M. (1994). *Supramolecular Chemistry: Concepts and Perspectives.* Weinheim: VCH Verlagsgesellschaft mbH.
Li, J., Shen, B., Nie, S., Duan, Z. & Chen, K. (2019). *Carbohydr. Polym.* **206**, 163–173.
Linden, A., Zhou, Y. & Heimgartner, H. (2014). *Acta Cryst.* C**70**, 482–487.
Lindgren, B., Lindgren, G., Artursson, E., Puu, G., Fredriksson, J. & Andersson, M. (1985). *J. Enzyme Inhib.* **1**, 1–11.
Liu, Z., Xia, G., Chen, S., Liu, Y., Li, H. & She, Z. (2014). *Marine Drugs*, **12**, 3669–3680.
Mackie, I. D. & DiLabio, G. A. J. (2011). *J. Chem. Phys.* **135**, 134318–10.
Madhunapantula, S., Desai, D., Sharma, A., Huh, S. J., Amin, S. & Robertson, G. P. (2008). *Mol. Cancer Ther.* **7**, 1297–1308.
Madzhidov, T. I. & Chmutova, G. A. (2010). *J. Mol. Struct. Theochem*, **959**, 1–7.
Metrangolo, P. & Resnati, G. (2001). *Chem. Chem. Eur. J.* **7**, 2511–2519.
Mukherjee, A. J., Zade, S. S., Singh, H. B. & Sunoj, R. B. (2010). *Chem. Rev.* **110**, 4357–4416.
Murray, J., Lane, P. & Politzer, P. (2009). *J. Mol. Model.* **15**, 723–729.
Muttenthaler, M. & Alewood, P. F. (2008). *J. Pept. Sci.* **14**, 1223–1239.

Pearson, R. G. (1990). *Coord. Chem. Rev.* **100**, 403–425.

Politzer, P., Murray, J. S., Janjić, G. V. & Zarić, S. D. (2014). *Crystals*, **4**, 12–31.

Raygan, F., Ostadmohammadi, V. & Asemi, Z. (2019). *Clin. Nutr.* **38**, 1594–1598.

Rodríguez-Gutiérrez, G., Rubio-Senent, F., Gómez-Carretero, A., Maya, I., Fernández-Bolaños, J., Duthie, G. G. & de Roos, B. (2019). *Eur. J. Nutr.* **58**, 1847–1851.

Saberinasab, M., Salehzadeh, S., Maghsoud, Y. & Bayat, M. (2016*b*). *Comput. Theor. Chem.* **1078**, 9–15.

Saberinasab, M., Salehzadeh, S. & Solimannejad, M. (2016*a*). *Comput. Theor. Chem.* **1092**, 41–46.

Salter, M. H. Jr, Reibenspies, J. H., Jones, S. B. & Hancock, R. D. (2005). *Inorg. Chem.* **44**, 2791–2797.

Sánchez-Sanz, G., Alkorta, I. & Elguero, J. (2011). *Comput. Theor. Chem.* **974**, 37–42.

Sanz, P., Mó, O. & Yáñez, M. (2003*a*). *Phys. Chem. Chem. Phys.* **5**, 2942–2947.

Sanz, P., Yáñez, M. & Mó, O. (2002). *J. Phys. Chem. A*, **106**, 4661–4668.

Sanz, P., Yáñez, M. & Mó, O. (2003*b*). *Chem. Eur. J.* **9**, 4548–4555.

Senćanski, M., Djordjević, I. & Grubišić, S. (2017). *J. Mol. Model.* **23**, 162–175.

Steinmann, D., Nauser, T. & Koppenol, W. (2010). *J. Org. Chem.* **75**, 6696–6699.

Thomas, S. P., Sathishkumar, R. & Guru Row, T. N. (2015). *Chem. Commun.* **51**, 14255–14258.

Tiecco, M., Testaferri, L., Bagnoli, L., Marini, F., Santi, C., Temperini, A., Scarponi, C., Sternativo, S., Terlizzi, R. & Tomassini, C. (2006*a*). *Arkivoc*, **7**, 186–206.

Tiecco, M., Testaferri, L., Bagnoli, L., Marini, F., Temperini, A., Tomassini, C. & Santi, C. (2000). *Tetrahedron Lett.* **41**, 3241–3245.

Tiecco, M., Testaferri, L., Santi, C., Tomassini, C., Bonini, R., Marini, F., Bagnoli, L. & Temperini, A. (2004). *Org. Lett.* **6**, 4751–4753.

Tiecco, M., Testaferri, L., Santi, C., Tomassini, C., Marini, F., Bagnoli, L. & Temperini, A. (2002). *Chem. Eur. J.* **8**, 1118–1124.

Tiecco, M., Testaferri, L., Santi, C., Tomassini, C., Marini, F., Bagnoli, L. & Temperini, A. (2003). *Angew. Chem. Int. Ed.* **42**, 3131–3133.

Tiecco, M., Testaferri, L., Santi, C., Tomassini, C., Santoro, S., Marini, F., Bagnoli, L., Temperini, A. & Costantino, F. (2006*b*). *Eur. J. Org. Chem.* **2006**, 4867–4873.

Trott, O. & Olson, A. J. (2010). *J. Comput. Chem.* **31**, 455–461.

Vinceti, M., Michalke, B., Malagoli, C., Eichmüller, M., Filippini, T., Tondelli, M., Bargellini, A., Vinceti, G., Zamboni, G. & Chiari, A. (2019). *J. Trace Elem. Med. Biol.* **53**, 154–162.

Voth, A. R., Khuu, P., Oishi, K. & Ho, P. S. (2009). *Nat. Chem.* **1**, 74–79.

Wang, L., Gong, C., Yuan, X. & Wei, G. (2019). *Nanomaterials*, **9**, 285–312.

Yamamoto, K., Miyake, H., Kusunoki, M. & Osaki, S. (2010). *FEBS J.* **277**, 4205–4214.

Yoshizawa, S. & Böck, A. (2009). *Biochim. Biophys. Acta*, **1790**, 1404–1414.