Binding of GSH conjugates to π-GST: A cross-docking approach

Fabio Lannutti *, Alessandro Marrone, Nazzareno Re

Dipartimento di Scienze del Farmaco, Università "G d'Annunzio" di Chieti-Pescara, Via dei Vestini 31, I-66100 Chieti, Italy

ARTICLE INFO

ABSTRACT

The high degree of flexibility characterizing the members of the GST protein family is supposed to be an evolution-resolved feature related to the detoxifying role of these enzymes. Many evidences suggest that overexpression of these enzymes may be implicated in the development of acquired resistance to antitumor agents. Among the most effective GST inhibitors, GSH conjugates have been found to be particularly promising because of their low toxicity. Here, we used a cross docking approach based on an ensemble of X-ray structures of GST bound complexes to model the effects of protein flexibility on the binding of GSH conjugates. We showed that our multitarget approach, allows to analyze the impact of protein flexibility and induced fit effects in GSH conjugate docking to GST. Moreover, the inclusion of conserved water molecules in the model allowed to include a further source of target variability and improve the performances in the docking of GSH conjugates through an enhanced description of the GSH moiety interactions. Therefore, a map of ligand–protein interactions reflecting the target variability included in the docking model was retraced and used to gain a thorough insight about the way GSH conjugates bind to GST.

© 2011 Elsevier Inc. All rights reserved.

1. Introduction

Glutathione S-transferases [1] (GSTs, EC 2.5.1.18) are members of a superfamily of multifunctional dimeric proteins involved in the cellular detoxification of cytotoxic and genotoxic compounds and in the prevention of tissue oxidative damages [1–8]. The cytosolic mammalian GSTs are homodimeric proteins classified into eight distinct classes: α, μ, π, θ, σ, κ, ζ and o [9–14]. The sequence homology is typically high within a class (greater than 70%) whereas it is usually low (less than 30%) between different classes [15–17]. GSTs catalyze the nucleophilic addition of the glutathione (GSH, γ-Glu-Cys-Gly) to a wide variety of endogenous and exogenous hydrophobic electrophiles including alkyl and aryl halides, epoxides, quinones, and activated alkenes. In general, the products of S-conjugation catalyzed by GST (hereafter called GSH conjugates) are more polar than the corresponding precursors and, thus, more easily excreted. Among the possible substrates of GST, alkylating agents used in the cytotoxic chemotherapy of many cancers can be targeted thus making GSTs of clinical relevance in oncology [18]. Indeed, many evidences suggest that overexpression of these enzymes may be implicated in the development of acquired resistance to antitumor agents [18]. Hence, the co-administration of potent, selective GST inhibitors as adjuvants to

cytotoxic chemotherapy has emerged as a possible strategy to restore the drug sensitivity of resistant cells [19]. However, the rational basis for the design of therapeutically valid GST inhibitors is still lacking because of the inherent complexity of the mechanism by which GSTs recognize and conjugate their ligands.

There are at least two ligand-binding sites per subunit: the glutathione-binding site (G-site), which is very specific for GSH and the hydrophobic substrate-binding site (H-site), which can bind a large variety of different electrophiles [6]. While the structure diversity of the H-site governs the substrate specificity of a particular GST (generally not very high), the high flexibility of GST enzymes, mainly due to their dimeric structure, is responsible of large number of diverse electrophilic substrates they are able to recognize. Furthermore, several evidences have shown that the dimeric structure of GST induces also large cooperative effects [20].

In recent years, the crystal structure characterization of several co-crystallized π-GST complexes has shed more light on how these enzymes recognize and react with a broad range of substrates and inhibitors [21]. Among the most promising GST inhibitors, the GSH conjugate ligands have been found to be particularly interesting, because of their low toxicity [22]. However, despite several co-crystallized complexes between π-GST and GSH conjugates have been determined in the last years, the use of crystal structure information in the design of new inhibitors has been particularly challenging. These difficulties are probably related to the high degree of flexibility of these enzymes [23] which may determine large induced-fit effects in the GST–ligand binding process and affect the extent and strength of the target–ligand interactions.
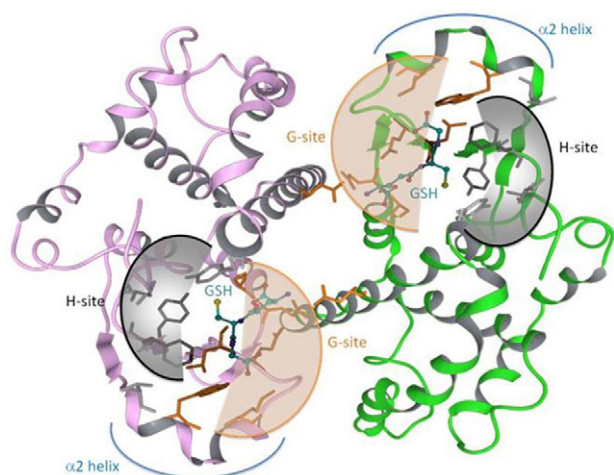
Fig. 1. Structure of the homodimeric human GSTP1-1 (π-GST) shown as a ribbon diagram.

Indeed, in spite of the intense interest in this field, a limited number of docking studies have been conducted on GST inhibitors so far, and most of them are limited to predict the binding mode of specific inhibitors, such as or naphthalene sulfonate, sulphanilamide, or fluorescein derivatives [24,25]. The only general docking study of GST inhibitors considered the binding of 14 ligands but was rather focused on the evaluation of the software SLIDE developed by the authors [25] (Fig. 1).

Due to its flexibility, the inclusion of target structural variability in docking studies to GST is expected to be of fundamental importance to increase the accuracy in the binding pose prediction.

In the present study, cross-docking approaches were employed to model the binding of a set of GSH conjugate inhibitors to human π-GST receptor and to characterize the influence of target flexibility on docking performances. This computational strategy makes use of an ensemble of N crystal structures of a target protein bound with different ligands: each ligand is then extracted from its native crystal structure and docked to each of the protein structures of the ensemble yielding a matrix of $N \times N$ docked complexes. Here, we selected nine π-GST co-crystallized with GSH conjugates to form the protein ensemble for cross-docking. The similarity in both the structure and the binding mode of GSH conjugates reduces the structural variability encoded by the GST ensemble which is limited to side chain conformations and slight backbone deformations. The term target "flexibility", which characterizes the extent of the overall conformational changes between apo and bound states of a protein, is more properly replaced by the term protein "variability", related to structural differences between diverse (by different ligands) bound states.

The docking analysis was conducted by testing the performances of Glide algorithm [26], recently indicated as one of the most effective docking programs [27], in predicting the binding pose in the native proteins. Our GST ensemble was preliminary analyzed through the comparison of the atomic fluctuations obtained by the analysis of B-factor values with those mimicked by our protein ensemble, to define its intrinsic degree of structural variability. In addition, the crystallographic conformations of GST were analyzed through the singular decomposition of atomic coordinates covariance matrix to gain an insight of how the variability encoded by the ensemble is distributed throughout the enzyme structure.

By using the same docking approach, we also studied how the inclusion of explicit water molecules in the target structures can affect the docking pose of GSH conjugates, representing a further source of variability in to their native-proteins. Indeed, the dimeric structure of GST lets its binding site partially exposed to the bulk solvent, suggesting the active participation of some water molecules into the GST–ligand binding process. The inclusion of explicit water molecules in the description of the target binding pocket has been recently shown to be able to increase the docking accuracy [28].

Therefore, we employ the cross-docking results to map the target–ligand interactions for each considered GSH conjugate and to score each target–ligand interaction (either H-bond or steric contacts) according to the probability it could be detected in the ensemble. Because the representation of the GST binding site is done by a series of conformations, this scoring scheme will reflect the structural variability of the target due to both the conformational flexibility and the presence of explicit water molecules in the binding site. Valuable information about the way GSH conjugates bind to GST was gained by our computational approach and can be used in the rational design of new GSH conjugate inhibitors.

2. Results and discussion

The adaptation of enzyme conformation to the bound ligand (induced fit) is crucial for the recognition mechanism and, consequently, for the enzymatic specificity. The high degree of flexibility characterizing the members of the GST protein family [23] is supposed to be an evolution-resolved feature related to the detoxifying role of these enzymes: the more diverse substrates could be bound the more detoxifying function wide-spreads.

In this study we analyzed how the inclusion of GST flexibility through a multitarget scheme can improve the ligand pose prediction for a set of nine GSH conjugated GST inhibitors operated by Glide docking protocol [26].

The structures of nine GST-bound complexes of GSH conjugates (Table 1), characterized by X-ray analyses, were employed to generate an ensemble of nine different GST conformations; the binding poses for the selected GSH conjugates were then calculated by cross-docking to this ensemble. The RMSD between the pose calculated in both native and non-native protein conformations is then used to build up cross-docking matrices which allow to highlighting how the variability in the target conformation affects the ligand binding. A preliminary analysis of our GST ensemble was carried out to estimate the degree of structural variability in our protein ensemble.

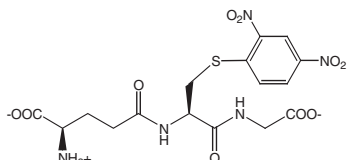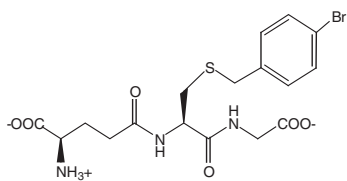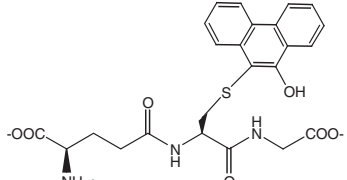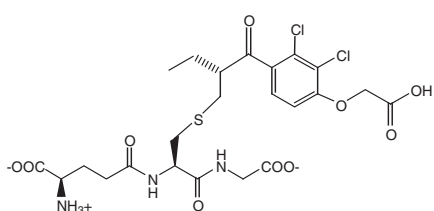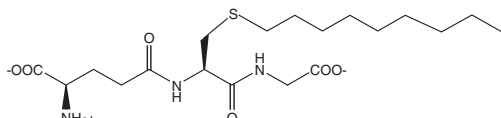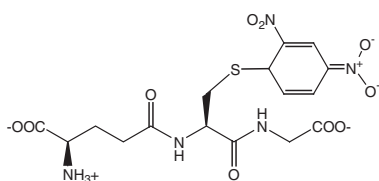### 2.1. Structural variability of the GST structure ensemble

The X-ray structures of the selected GST-bound complexes are characterized by similar crystallographic features, in particular both crystal symmetries and packing contacts per unit cell are almost identical for the considered pdb entries (Table S7).

The degree of GST conformational flexibility of the X-ray structures was analyzed by employing the thermal B-factors available from the crystallographic data. Indeed, important information about the flexibility of a protein can be acquired from the thermal B-factors obtained from the X-ray data [29]. The B-factors represent the smearing of the of atomic electron densities around their equilibrium positions due to the thermal motions and static disorder and are therefore directly related to the atomic fluctuation in the protein through the Debye–Waller equation, see Section 4.

The atomic fluctuations of the GST backbone for the enzyme both in the apo and GSH complexed forms are reported in Fig. 2. The apo and GSH bound protein forms give rise to nearly similar profiles in which segments of reduced fluctuation are clearly distinguished for the latter form. The reduced flexibility in the GSH bound form is expected as a consequence of the ligand binding which significantly reduces the mobility of the protein chain proximal to the binding pocket. Also reported in Fig. 2 is the plot

Table 1

Considered GSH conjugate ligands with pdb codes for the X-ray structures of the corresponding co-crystallized complexes with GSH.

Entry PDB[a]	Ligand	Protein	Ligand structure
10gs[37]	**1**	I	
1aqw[44]	**2**	II	
1pgt[45]	**3**	III	
18gs[46]	**4**	IV	
1aqv[47]	**5**	V	
2pgt[45]	**6**	VI	
3gss[48]	**7**	VII	
12gs[49]	**8**	VIII	
1aqx[44]	**9**	IX	
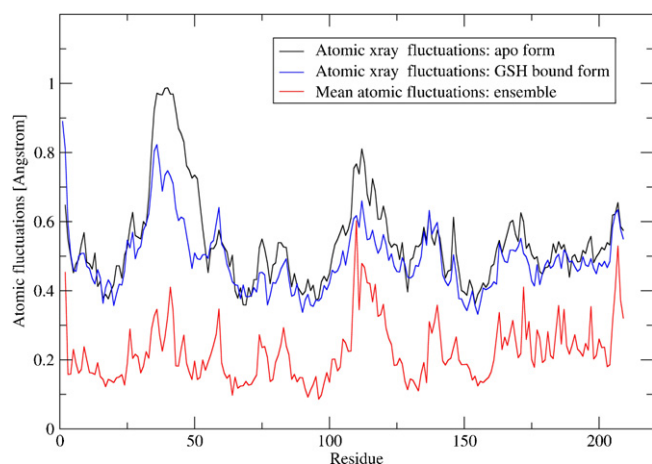
Fig. 2. Backbone Cα atomic fluctuations of the GST in the apo (black) and GSH bound (blue) forms, calculated form the *B*-factors, and mean atomic deviations for the considered X-ray structures ensemble (red). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)
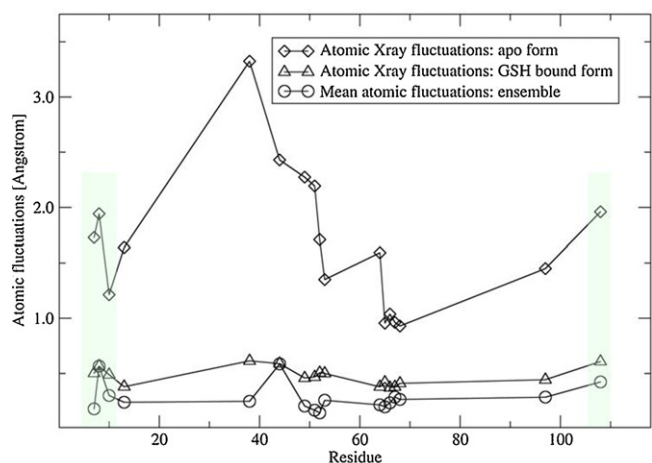
of the mean atomic deviations (MAD) for the ensemble of nine co-crystallized GST–GSH conjugate complexes. These values are calculated by superimposing the GST structures, and thus represent an estimate of the receptor flexibility covered by our protein ensemble. As shown in Fig. 3, the GST backbone fluctuations derived by the *B*-factors for the apo and GSH bound X-ray structures and that covered by the considered structure ensemble share almost similar profile, even though the atomic deviations are markedly lower.

In Fig. 3 the atomic fluctuations for all the atoms of the binding site residues are reported. The plot of the apo form fluctuations range at high values, namely 1.0–3.5 Å, and indicates large atomic fluctuations for the side-chain atoms as a consequence of the higher flexibility of GST in the absence of any ligand. As expected, a markedly reduced mobility of the side-chain atoms is observed in the GSH-bound complex, the corresponding atomic fluctuations ranging within 0.5 Å. It is worth comparing the GSH-bound complex profile with that of the MAD of the GSH conjugates co-crystallized ensemble. Fig. 3 shows these two profiles very close, with significant differences for the Lys44 and Phe8 residues only, indicating that the side chain mobility of GST binding site should be adequately represented by the selected ensemble of crystallographic

Fig. 3. Side chain atomic fluctuations of the GST binding site for the GST in the apo (diamonds) and GSH bound (triangles) forms, calculated form the *B*-factors, and mean atomic deviations for the considered X-ray structures ensemble (circles). Shaded data correspond to the H-site residues 7, 8, 10 and 108.

GST conformations. The analysis of the GST atomic fluctuations also allows to distinguish between two effects within the influence of the ligand binding onto receptor flexibility. The backbone atoms are on overall little sensitive to the binding process, as reflected by the small difference of the thermal fluctuations for the apo and GSH-bond protein. On the other hand, the side-chain residues proximal to the binding pocket undergo a marked reduction of their flexibility as a consequence of the protein–ligand interaction.

To gain an insight of how the structural variability encoded by our protein ensemble is distributed throughout the GST structure, we performed the singular decomposition of Cartesian coordinates covariance matrix (and its transposed) for all protein atoms. This procedure allowed to calculate a set of orthogonal eigenvectors, obtained as linear combinations of the atomic coordinates deviations (see Section 4), and corresponding to structural deformations representative of our GST ensemble. The first five deformations, giving the highest contribution to the coordinates covariance, were visually represented by corresponding motion pictures (see Supplementary Content). Most of the structural changes induced by these deformations can be ascribed to crystal packing because of the involvement of many residues on protein surface. On the other hand, the largest deformations (first three) interesting either single or both protein chains, are also reflected in the size and shape of ligand binding site and can thus contribute to the target variability influencing the docking outcomes (see below).

2.2. Cross-docking in the absence of water molecules

A possible way to increase the accuracy in the docking of GSH conjugates to GST can be the inclusion of target flexibility by the use of multiple receptor structures, whereby the target structure is represented by an ensemble of conformations [30,31]. In this investigation, we adopted a cross-docking procedure to dock each of the considered nine GSH conjugate ligands, **1**–**9**, in each of the corresponding π-GST structures, I–IX, using Glide 3.5 (see Section 4). The cross-docking results are presented in Table 2 in matrix form where each cell gives the minimum RMSD (calculated for all the heavy atoms of the GST ligands) of a ligand (in the column) to a structure (in the row). The columns refer to the different GSH conjugate ligands and the rows to the different GST structures The diagonal cells gives the RMSD for the docking of the ligands in their native proteins, while the off-diagonal cells gives the RMSD for the docking of a ligand to a different protein conformation and encode the information on the protein flexibility. As commonly accepted in the literature [32], a ligand binding mode is considered to be correctly predicted if the RMSD is lower than 2.0 Å and incorrectly predicted if the RMSD is higher than 3.0 Å, while ligand binding modes with RMSD between 2.0 and 3.0 Å can be considered partially docked. The diagonal RMSD entries shows that the considered ligands are correctly docked in their native proteins. Indeed, all RMSD values are below 2.0 Å except **4** and **7** (RMSD of 2.49 and 2.63 Å, respectively) while five of the remaining ligands show RMSD values lower than 1.0 Å, indicating an accurate reproduction of the native X-ray structures. These good values allow to assess the usability of the Glide program in the docking to GST. The off diagonal RMSD were significantly higher than diagonal values in only a few cases (V-**5** and IX-**9**), whereas the predicted binding poses seem to be more influenced by the ligand nature, as indicated by the RMSD values along a column which reflect the accuracy of docking calculations to predict the experimental binding pose of a given ligand to all the protein conformations of the ensemble. In most cases these values are characterized by small variances, showing a small dependence of the binding pose prediction on the employed GST conformation. To provide for a quantification of the pose quality of a ligand to all

Table 2
Cross docking matrix with the RMSDs between experimental and calculated binding poses of ligands **1–9** into structures I–IX of the GST ensemble not including any water molecule.

| | | Ligands | | | | | | | | |
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Protiens	I	1.77	0.56	0.77	1.93	0.72	2.12	2.30	0.78	2.70
	II	0.50	0.31	0.90	2.45	1.13	1.29	2.16	0.94	2.65
	III	0.52	0.36	0.73	2.20	0.74	1.16	1.87	1.62	0.62
	IV	0.69	0.28	0.97	2.49	0.66	1.27	2.63	0.81	0.61
	V	0.43	0.29	0.76	1.56	0.35	1.02	1.98	0.69	0.75
	VI	1.62	0.36	0.99	2.59	2.12	1.22	1.78	0.78	0.64
	VII	1.65	0.20	0.87	1.49	1.17	1.84	2.63	0.99	1.12
	VIII	0.70	0.28	0.86	2.06	0.72	1.89	2.56	0.72	2.75
	IX	1.62	0.32	0.59	2.33	2.09	2.20	2.06	1.89	0.41
	D.A.	1.00	1.00	1.00	0.67	0.89	0.89	0.67	1.00	0.83

proteins, we used the docking accuracy function (*DA*), calculated as:

$$DA = f_{\text{RMSD}<2.0} + \frac{1}{2}(f_{\text{RMSD}<3.0} - f_{\text{RMSD}<2.0})$$

where $f_{\text{RMSD}<\nu}$ is the fraction of RMSD values below ν Å.

Notably, some ligands can be correctly docked into all or almost all protein structures of the ensemble, as reflected by *DA* values of 1.0 or slightly lower. In particular, the GSH (**2**) binding geometry is accurately predicted by docking to all structures (column 2) as tested by RMSD values within 0.6 Å with most values below 0.4 Å. This is not surprising as GSH binds only to the G-site whereas the conjugated ligands bind to both G and H-site and are, thus, subjected to more geometrical restrictions. Also ligands **1**, **3** and **8**, and to a lesser extent **5**, are correctly docked into all structures with RMSD values within 1.0 Å for **3**, 2.0 Å for **1** and **8** and 2.1 Å for **5**. On the other hand, ligands **4**, **6**, **7** and **9** are incorrectly docked to several protein structures, with RMSD above 2.5 Å. Interestingly, all these ligands bear in their conjugated moiety an aryl group that could interact with the target by π–π stacking. Indeed, the graphical inspection of the crystal pose of these four ligands showed the conjugated aryl stacked to the phenolic moiety of Tyr108, **4** and **7** apparently be more stacked than **6** and **9**. This observation suggests that a π–π stacking interaction could take place between the Tyr108 phenol ring and an electron-pour aryl ring conjugated to GSH. As known, these stacking interactions are not commonly included in the employed scoring functions, thus explaining the lack of docking accuracy at least for **4** and **7**. The comparison of X-ray and native docking poses (Fig. S8) clearly shows that the conjugated aryl rings of these ligands are projected approximately towards Phe8 to make favorable hydrophobic contacts rather than stacking with the Tyr108 phenol ring, as experimentally detected.

2.3. Cross-docking in the presence of water molecules

The important role played by water molecules in the protein–ligand binding has been recently recognized [28]. A recent analysis of high-resolution ligand–protein complexes pointed out that more than 80% of them present one or more water molecules stabilizing the complex [33]. The water molecules found in the

crystal structures of protein–ligand complexes tend to occupy conserved positions in structures containing different bound ligands.

The search of a common set of water molecules to explicitly include in the cross-docking calculations was performed by superposing the Cα atoms of all the structures in the ensemble with respect to protein I (PDB code 1Ogs), arbitrarily used as reference. Thus, the water molecules in the binding-site having a constituent role and/or actively participating to the ligand binding were easily determined. As shown in Fig. 4, we identified six conserved water molecules localized around the γ-glutamyl portion of the ligand that can therefore favor the binding by means of strong hydrogen bonds involving the NH_3^+ and COO^- groups of this terminus. The six water molecules (Fig. 4) are not placed exactly the same but, "flexibly" adapted to the ligand they correspond to. Thus, the six conserved water molecules may represent a further source of target variability, in addition to that related to either backbone or sidechains conformation of GST.
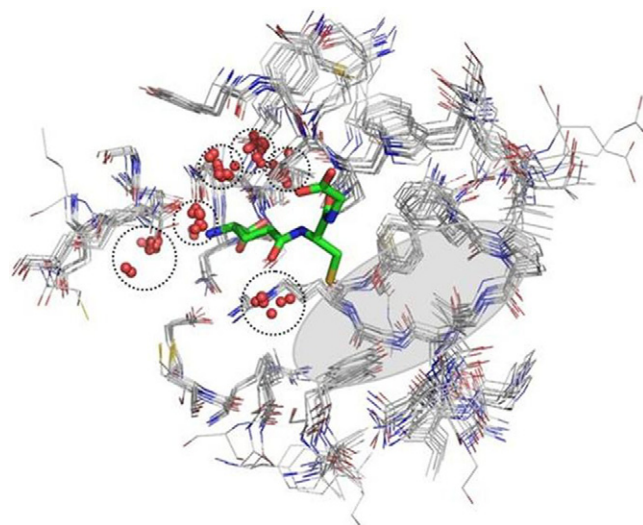
Fig. 4. Conserved water molecules (dotted circles) involved in the GSH conjugates interactions in the I–IX GST ensemble structures (lines). Also shown (grey shadowed) the H-site binding region and the GSH moiety (sticks). This picture has been processed by PyMOL graphical interface [33].

Table 3
Cross docking matrix with the RMSDs between experimental and calculated binding poses of ligands **1–9** into structures I–IX of the GST ensemble, including the six conserved water molecules.

		Ligands								
		1	2	3	4	5	6	7	8	9
Proteins	I	0.20	0.33	0.69	2.86	0.28	0.84	2.6	0.79	1.15
	II	0.40	0.28	0.62	2.52	0.42	2.16	2.64	1.33	1.58
	III	0.35	0.20	0.60	2.39	0.39	1.07	2.32	0.78	1.00
	IV	0.34	0.26	0.56	2.56	0.28	0.55	3.45	0.80	1.44
	V	0.58	0.18	0.61	2.53	0.27	2.89	2.28	0.85	1.03
	VI	1.61	0.25	0.83	2.52	0.84	0.28	2.40	0.60	1.53
	VII	0.44	0.29	1.04	2.54	0.99	0.95	3.37	0.89	1.05
	VIII	0.36	0.17	0.67	2.37	0.43	0.47	2.82	0.66	1.40
	IX	0.73	0.69	0.93	2.48	1.94	2.91	2.27	0.87	0.84
	D.A.	1.00	1.00	1.00	0.50	1.00	0.83	0.39	1.00	1.00

The same cross-docking approach presented above was repeated using the ensemble structures I–IX explicitly including the six conserved water molecules in the docking simulations, and the results are reported in Table 3.

As a first evidence of the inclusion of water molecules in target model, ligands **6** and **9**, two decoys in the cross-docking without explicit waters, were now fairly reproduced.

Consistently with the previous docking model, the comparison of the fraction of diagonal versus off-diagonal RMSD values below 2.0 Å showed that, with the exception of **4** and **7** still representing decoys of our estimations, no significant induced fit effect could be unveiled. On the other hand, we detected a slight improvement in the overall docking performances upon water inclusion as evidenced by the corresponding *DA* values.

Although the comparison of Table 2 and Table 3 values suggests that the inclusion of the six conserved water molecules did only slightly improve the overall docking accuracy, the calculation of cross-docking matrices restricted to only GSH moieties showed a clear and overall improvement in the pose prediction of this ligand portion by the inclusion of explicit water, appreciable from both a decrease of the diagonal values and an increase in the diagonal to off-diagonal contrasts when water molecules are included (Table S3). The visual inspection of calculated poses indicated an overall improvement in the pose of γ-glutamyl fragments due to the additional H-bond interactions between this terminus and three of the explicit waters and to the increased shape complementarity obtained by water inclusion. It is worth noting that the RMSD values for **4** and **7** are still high (Table S3), indicating that the pose of their GSH moieties is somewhat negatively influenced by the incorrect prediction of aryl stacking in the H-site.

2.4. Docking with the GOLD program

For the sake of comparison, the cross-docking calculations were also carried out using the Gold docking and the results are reported in Tables S5 and S6 (see Supporting Information and references therein). The docking algorithm implemented in the latter program is considerably different from that in Glide so that the results of Gold docking give a valuable terms of comparison.

The calculated diagonal and off-diagonal entries, in the absence and in the presence of explicit water molecules, show qualitative trends quite similar to those observed with Glide. However, the RMSD values are on average higher, reflecting a lower accuracy of GOLD in the prediction of the binding pose. This loss of accuracy is even more meaningful considering that, for some ligand, good binding poses could only be found by constraining some of the most conserved hydrogen-bond interactions. These results underline the

superiority of Glide with respect to GOLD to tackle the docking of GSH conjugate ligands to GST.

2.5. Analysis of target–ligand interactions

The above cross-docking results were exploited to perform a detailed qualitative and quantitative analysis of GST–ligand interactions. As discussed above, the cross-docking into the selected ensemble of nine crystallographic GST conformations allows to investigate how the enzyme–ligand interactions are affected by their intrinsic structural flexibility improving the usual enzyme–ligand interaction mapping schemes [34] which are normally performed on only one protein structure. The mapping of the enzyme–ligand interactions was conducted by highlighting and estimating the hydrogen-bond interactions that best describe the binding within the G-site of the GSH moiety of all GSH conjugate inhibitors. Furthermore, we reported the hydrophobic contacts between the residues within the H-site and the conjugates moiety of the ligands considered in this study. Exploiting the cross-docking results, we could thus evaluate (see Section 4) the frequencies each LBS residue is involved in either a hydrogen-bond or hydrophobic contact and use them to score the occurrence of each interaction in the GST ensemble. This analysis was conducted by using the cross-docking results in both the absence and presence of the six conserved water molecules, and the results allowed us to retrace the maps reported in Figs. 5 and 6, respectively. As shown in Fig. 5, the formation of the H-bond interactions between the receptor and the ligand is observed with the maximum frequency for seven residues of the LBS: Gln64, Asp98, Ser65, Leu52, Lys44, Trp38, and Gln51, which form hydrogen-bonds with the ligands through either backbone or side chain functional groups. The high interaction frequencies shown by Gln64, Leu52, Lys44 and Trp38, ranging around 0.95, highlight how these residues are fundamental for the binding of the GSH moiety, common to all ligands, and indicate that this binding mode is barely affected by the nature of the conjugate moiety which is different for each ligand. On the other hand, Gln51, Asp98, and Ser65 show slightly lower interaction frequencies (0.70–0.88) reflecting that the hydrogen-bond interactions of these residues can be more sensitive to the nature of the conjugate moiety. The hydrophobic contacts between the H-site residues and the conjugated moiety of the ligands were also reported in Fig. 5. Seven residues, Trp38, Val35, Gly205, Tyr108, Val10, Phe8 and Tyr7 are significantly involved in this type of contact, with interaction frequencies higher than ca. 0.1. As expected, compared to H-bond, the interaction frequencies for these hydrophobic contacts are spread over a wider range of values, namely between 0.11 for Trp38 and 0.80 for Phe8, as a consequence of both the non-localized
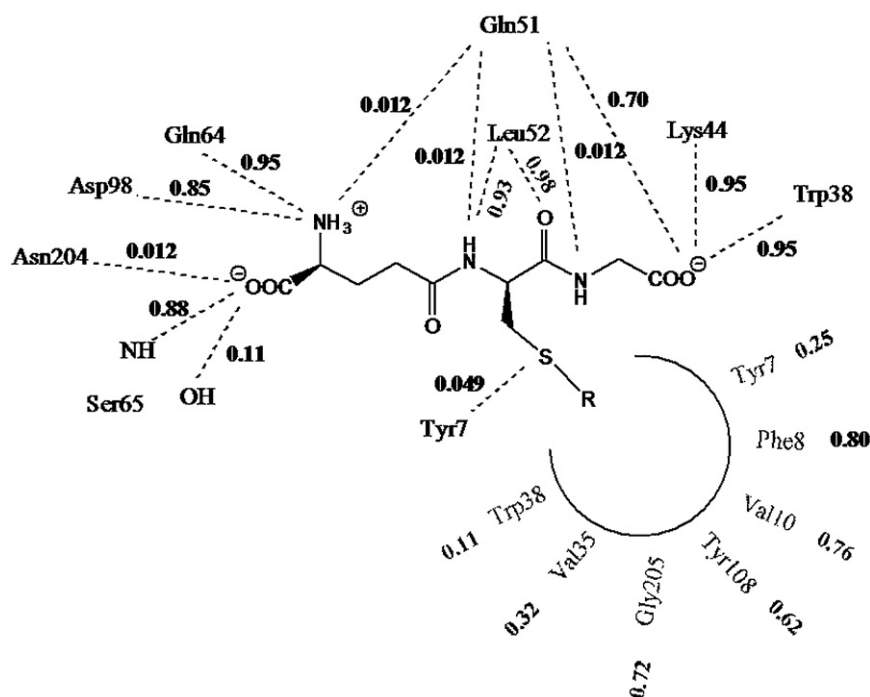
Fig. 5. Map of the GST–ligand interaction frequencies obtained from cross-docking into LBS without any water molecule.

nature of the hydrophobic contacts and the higher flexibility of the H-site. Also, as above stated, the interaction frequency calculated for Tyr108 is expected to be underestimated because of the low accuracy of docking in the prediction of π–π stacking. Notably, the Tyr7 residue, involved in the activation of the thiolate group of GSH in the enzymatic catalysis, showed relatively low interaction frequencies for both H-bond, 0.049, and hydrophobic contact, 0.25.

As shown in Fig. 6, the frequencies of both H-bond interactions and hydrophobic contacts are influenced by the inclusion of the six conserved water molecules. As already stated, the presence of explicit water molecules nearby the site occupied by the γ-glutamyl moiety influences the binding mode of the zwitterionic terminus of GSH conjugates. Moreover, as expected, Fig. 6 indicates the formation of hydrogen-bonds patterns between the ligand and the explicit H_2O molecules. In particular, only three H_2O molecules
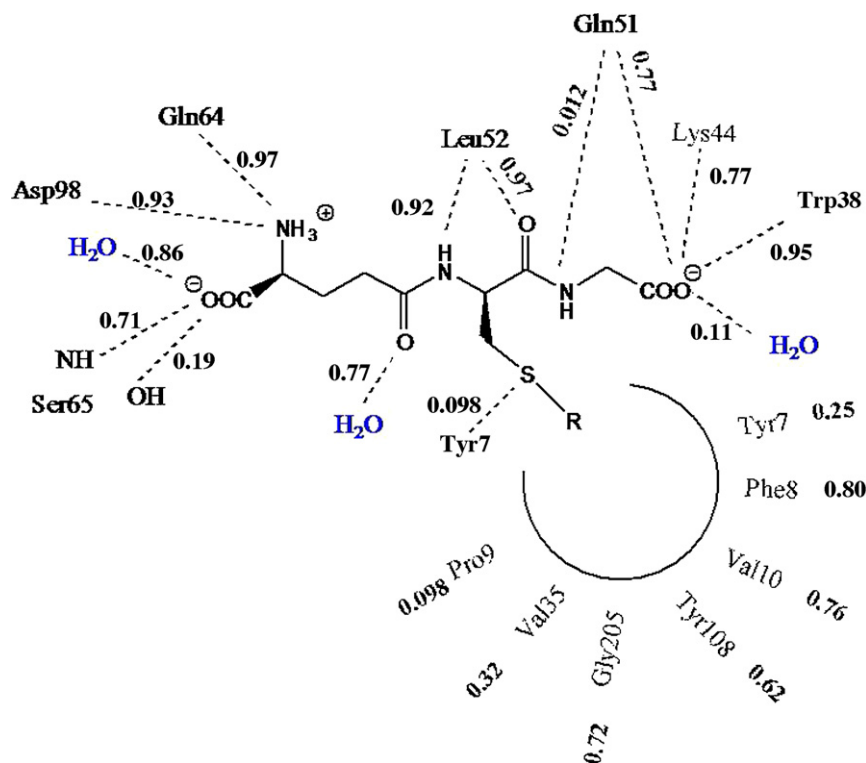
Fig. 6. Map of the GST–ligand interaction frequencies obtained from cross-docking into LBS including six conserved water molecules.

are shown to be explicitly involved in H-bond interactions with the ligand and, among them, two interact with the carboxylate and the carbonyl groups of the γ-glutamyl moiety (with interaction frequencies of 0.86 and 0.77, respectively), and one with the Gly carboxylate (with a lower interaction frequencies of 0.11). On the basis of the high interaction frequencies of two of these water molecules, they can be considered not only structurally conserved but also functionally important for the binding at G-site. The remaining three water molecules are not explicitly involved in receptor–ligand interaction patterns and are thus expected to play only a role in stabilizing the structural configuration of the binding site through inter residues interactions and only indirectly contributing to the target–ligand recognition process. The analysis of hydrophobic interactions in Fig. 6, points out essentially the same pattern of interaction frequencies of Fig. 5, indicating that the binding of the conjugate moiety of the ligands is little affected by the inclusion of explicit water molecules in the target model. The only significant difference is the replacement of Trp38 by Pro9 (with interaction frequency of 0.098), which can be considered as an improvement in the description of conjugate moiety binding mode, since Trp38 is not usually considered part of the H-site.

The map of interactions between GST and GSH conjugates can be useful in the design of new inhibitors with improved binding properties. For instance, the above results suggest that the interactions of both the ionic termini (interacting with Asp98 and Gln64) and the central peptide bond (interacting with Leu52) are highly conserved and any modification of these moieties is expected to be challenging. Our calculations indicated that the binding of γ-glutamyl terminus is also influenced by the presence of conserved water molecules interacting via H-bond with the γ-carboxylate and with α-carbonyl. A possible modification of the γ-glutamyl moiety suggested by our results to enhance the target–ligand affinity could be the attachment of additional hydrophilic groups on this terminus. This chemical modification would allow the fitting of the space region occupied by the water molecules, and thus the exploitation of water accessible binding sites, without substantially affecting the other target–ligand interactions of γ-glutamyl moiety.

On the other hand, changes in the conjugated moiety should be, expectedly, much more tolerated due to the reduced specificity of the interactions they are involved. Our results indicated that GST is able to adapt its H-site to hydrophobic moieties of different extents. In addition, the graphical inspection of crystal poses of ligands **4** and **7**, showed that Tyr108 can be involved in π–π stacking interactions with the aryl moieties of these ligands. It can be noticed that this stacking interaction occupies only a limited portion of H-site, on a space region close to the GSH sulfur, thus leaving unoccupied a large portion of this site. On this basis, the conjugation of GSH with extended moieties composed of an electron-pour aryl ring attached to a hydrophobic tail could be suggested to strengthen the target–ligand affinity on the H-site.

3. Conclusion

In this study we used a cross-docking approach to model the binding at GST of GSH conjugate inhibitors and to characterize the influence of target flexibility on docking performances. We employed an ensemble of nine X-ray structures for the GST complexes with GSH conjugates which was preliminarily analyzed through the comparison of the atomic fluctuations obtained by the analysis of B-factor values with those mimicked by our protein ensemble, and by the analysis of the covariance matrix of atomic coordinates deviations to define the inherent degree of structural variability.

We showed that the considered ligands are correctly docked in their native proteins with RMSD values below 2.0 Å except **4** and **7**,

the only ligands with a conjugate moiety forming π–π interactions with the phenol ring of Tyr108, allowing to assess the usability of the Glide program in the docking of GSH conjugates to GST. On overall, the cross-docking results showed that ligand poses were slightly affected by the target variability in the ensemble, as witnessed by the low variances of in-column RMSD values, whereas they resulted to be more influenced by the ligand nature.

We also accounted for the effect on docking accuracy of including explicit water molecules in the target models as a source of additional target variability. Our results indicate that the inclusion of water molecules improves the docking accuracy of the GSH moieties by essentially enhancing the description of the H-bond patterns involving the γ-glutamyl terminus.

Finally we used the cross-docking results to retrace a map of and to score the target–ligand interactions for GSH conjugates according to the probability they could be detected in the ensemble. Valuable information about the way GSH conjugates bind to GST was gained by our computational approach and can be used in the rational design of new GSH conjugate inhibitors.

4. Computational methods

4.1. Analysis of GST structure and GST ensemble

The crystal structures of nine complexes of the human P1-1 GST enzyme (π-GST) with GSH and with several GSH-conjugate ligands (Table 1) were picked up from the PDB archives. Successive graphical inspection and analysis of each structure active site were performed by using the molecular modeling software Maestro 7.0 [35] to find and fix possible structural mismatches. Besides the Cartesian coordinates of the protein atoms, employed to graphically inspect the 3D structure of the enzyme, the thermal B-factors of the considered set were also retrieved and used to investigate the enzyme flexibility [29]. To this purpose the thermal B-factors or temperature-factors of each PDB entry, which are related to the mean displacement of a vibrating atom $\langle u \rangle$ (in Å) through the Debye–Waller equation, were converted to the corresponding atomic fluctuations

$$B = \frac{8}{3}\pi^2 \langle u \rangle^2$$

The X-ray atomic fluctuations were then used to represent the proteins flexibility in the absence (apo-form) or in the presence of either GSH or the GSH conjugate ligands. The atomic fluctuations for the GST backbone were calculated by using the B-factors corresponding to the Cα of only one monomer (chain A). The mean atomic fluctuations for the GST sidechains were calculated for only the ligand binding site (LBS) residues by averaging the corresponding $\langle u \rangle$ values obtained through the Debye–Waller equation. More precisely, we considered only the GST residues within 4.0 Å from either GSH or conjugated portion which were detected in each of the nine pdb structures and correspond to 7, 8, 10, 13, 38, 44, 49, 51–53, 64–68, 97 and 108 on one subunit (chain A) and 94 and 98 on the other (chain B).

Superimposition of the crystallized proteins was carried out with Maestro assuming protein I, i.e. the protein conformation obtained from the PDB entry 1ogs [36], as reference and superimposing all the others to it. This step allowed the graphical inspection of the conserved water molecules within the active-site of each protein. In particular, we explicitly considered only the water molecules around the G-site, excluding the water molecules far from the binding-site that have no effect on the ligand binding. This simple graphical analysis allowed us to identify six conserved water molecules, among the crystallographic water molecules (Fig. 4).

The conformational variability encoded by the selected protein ensemble was further analyzed by the singular decomposition

of the covariance matrix of atomic coordinates deviations. More specifically, the pdb structures of GST ensemble were processed by an in-home written R script [42] which allowed to (i) rescale and average each atomic coordinates, (ii) build the matrix M of deviations from mean coordinates, (iii) perform the singular decomposition of MM^{-1} (the covariance matrix) and its transposed, and (iv) calculate the atomic coordinates of GST corresponding to deformations along the M covariance matrix eigenvectors. Being sorted in decreasing order of associated eigenvalues, the first five eigenvectors intercepted most of M covariance and are thus considered as the most representative of the structural variability encoded by the GST ensemble. The calculated structural deformations were then applied on the 10gs crystal structure to obtain picture movies of GST (10gs-pmovie#.mpeg files included in the Supplementary Content) fluctuating along these eigenvectors. The movie files were assembled by using the movie maker tool implemented in Maestro.

4.2. Docking calculations

The structures of the target used in docking calculations were prepared in two stages by using the Protein Preparation module implemented in the Maestro program. In the first stage (preparation) hydrogen atoms are automatically added to each protein structure according to the chemical nature of each aminoacid, on the basis of the ionized form expected in physiological condition. This module also controls the atomic charges assignment by neutralizing the total charge of the receptor atoms within a radius of 14 Å from the co-crystallized ligand mass center. In the second stage (refinement) each 3D structure of the protein–ligand complex was relaxed through constrained local minimization using the OPLS force fields [37] in order to remove possible structural mismatches due to the automatic procedure employed to add the hydrogen atoms. The above protein preparation scheme was applied alternatively to receptor models in which the crystallization water molecules are all removed and to receptor models including the six conserved water molecules shown in Fig. 4. The target structure resulting after each protein preparation was then used in the subsequent docking calculations. The 3D structures of the selected ligands were built using Maestro 7.0 and subsequently minimized using MacroModel [38]. The ionization form expected at physiological pH was considered for all ligands, i.e. anionic carboxylic and cationic ammine functions. In addition, since the force field differentiates between the possible double bond localization pattern, two resonance structures were considered for the co-crystallized ligand of the pdb entry 1aqx [39], namely **9** and **9′** ligands in which a negative charge is located on the *para* and *ortho* nitro group, respectively, of the conjugated 2,4,6-tri-nitro-benzene moiety. However, only the former structure, which was better docked to the native protein, was subsequently employed. The local minimization has been performed with the all-atom OPLS force field through the first-derivatives TNCG algorithm applying the gradient threshold criteria at 0.05 kJ mol^{-1} Å^{-1}, 500 maximum iteration and simulating the water (bulk) environment by using the GB/SA polarizable continuum method [40] implemented in the program. The locally minimized structures of GSH and GSH conjugates were used as inputs for subsequent docking calculations. Docking calculations were carried out with Glide 3.5 [41] which uses a series of hierarchical filters to search for possible locations of the ligand in the active-site region of the receptor [26]. The early stage of the docking is represented by the grids calculation in which the molecular fields generated by the binding site residues of the receptor are mapped and stored in archive grid files providing the "binding information" which are then used during the further steps of the docking. Even though apparently more time demanding, the preliminary grids mapping makes the following binding pose research much more

rapid and effective. A $27 \times 27 \times 27$ grid was centered to the center of mass of each co-crystallized ligand. The positional space of the ligand was instead limited to a $14 \times 14 \times 14$ box centered as already mentioned; these settings are default for grid calculation with Glide 3.5 [26]. The final stage of docking is represented by the search of the best binding pose of the ligand in the molecular interactions fields generated by the binding site residues. After generation and selection from an exhaustive enumeration of the minima in the ligand torsion-angle space, the selected conformations are prescreened to locate promising ligand poses and to reduce the region of LBS over which energy and gradient evaluations will later be performed. Subsequently, the prescreened conformations are minimized in the field of the receptor using a standard molecular mechanics energy function, in this case that of the OPLS-AA force field in conjunction with a distance-dependent dielectric model. Finally, the lowest-energy poses obtained in this fashion are subjected to a Monte Carlo procedure to obtain the final set of docking solutions. The evaluation of binding pose quality (fitness) is performed via a combination of the Glide implemented scoring function, GlideScore, the ligand–receptor molecular mechanics interaction energy, and the ligand strain energy, being this composite scoring function, called Emodel, much better at selecting the correct pose than either of single contribution alone. GSH and the considered GSH-conjugates were docked into the binding sites of nine GST conformations in two sessions by using different docking parameters set-ups. In the first session the ligands were docked by using the Single Precision (SP) parameterization set-up [34] by imposing a maximum of 30 docking solutions per ligand. The SP solutions were first graphically inspected and only the pose best resembling that of the corresponding co-crystallized structure was selected for the following session. The second docking session was performed by using the Extra Precision (XP) parameterization set-up [34] by imposing a maximum of 10 docking solutions per docking run. The ten XP solutions for each ligand resulted from docking into a single GST conformation were considered for further structural analysis. After graphical inspection of the gained binding poses, the best docking solutions were compared with the experimental binding pose of the corresponding co-crystallized ligand through RMSD minimization of heavy (non hydrogen) atoms. As a result we gained the cross-docking matrix reported in Table 2, corresponding to the LBS without any water molecule, and that reported in Table 3, corresponding to the LBS including the six conserved water molecules.

4.3. Target–ligand interaction frequencies

The analysis of the target–ligand interactions from cross-docking calculations with Glide was carried out with the graphical tools of Maestro 7.0. We initially confined the region defined as ligand binding site (LBS), for all the considered cross docking solutions. The LBS was further divided into three main sub-regions: (a) binding region of γ-glutamyl (γ-Glu) portion, (b) the binding region of glycine (Gly) portion and (c) the H-site. To achieve this goal, the sulfur atom of each ligand was taken as the center of the LBS, and all the residues of receptor within a radius of 6.0 Å from this center were included in the LBS. All possible H-bond interactions observed in the (a) and (b) regions were graphically highlighted by using a maximum distance cut-off of 2.5 Å for the H-bond formation. The hydrophobic contacts observed in the (c) site were instead assessed by the default distance criteria implemented in the Maestro program that correspond to 1.3 Å for the good contacts for which only those involving carbon atoms were considered. Subsequently, a cross analysis was performed considering the interactions at each ligand with all the considered GST conformations. In this way, we were able to both identify the residues of the target protein involved in such interactions and estimate the corresponding interaction

frequencies expressed by the ratio between the total number of contacts per residues to the total 81 entries, i.e. the 9×9 bound complexes obtained by cross-docking.

Acknowledgement

We thank The Ministry of Education, Universities and Research (MIUR) for the financial support.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmgm.2011.09.006.

References

[1] B. Mannervik, U.H. Danielson, CRC Crit. Rev. Biochem. 23 (1988) 283–337.
[2] T.D. Boyer, Hepatology 9 (1989) 486–496.
[3] C.B. Pickett, A.Y.H. Lu, Annu. Rev. Biochem. 58 (1989) 734–764.
[4] B. Coles, B. Ketterer, CRC Crit. Rev. Biochem. 25 (1990) 47–70.
[5] J.D. Hayes, C.R. Wolf, in: H. Sies, B. Ketterer (Eds.), Glutathione Conjugation: Mechanisms and Biological Significance, Academic Press, London, 1988, pp. 316–356.
[6] D.J. Waxman, Cancer Res. 50 (1990) 6449–6454.
[7] J.D. Hayes, D.J. Pulford, CRC Crit. Rev. Biochem. Mol. Biol. 30 (1995) 445–600.
[8] R.N. Armstrong, Chem. Res. Toxicol. 10 (1997) 2–18.
[9] B. Mannervik, Y.C. Awasthi, P.G. Board, J.D. Hayes, C. Di Ilio, B. Ketterer, L. Listowsky, R. Morgenstern, M. Muramatsu, W.R. Pearson, C.B. Pickett, K. Sato, M. Widersten, C.R. Wolf, Biochem. J. 282 (1992) 305–306.
[10] D.J. Meyer, B. Coles, S.E. Pemble, K.S. Gilmore, G.M. Fraser, B. Ketterer, Biochem. J. 274 (1991) 409–414.
[11] D.J. Meyer, M. Thomas, Biochem. J. 311 (1995) 739–742.
[12] S.E. Pemble, A.F. Wardle, J.B. Taylor, Biochem. J. 319 (1996) 749–754.
[13] P.G. Board, R.T. Baker, G. Chelvanayagam, L.S. Jermiin, Biochem. J. 328 (1997) 929–935.
[14] P.G. Board, M. Coggan, G. Chelvanayagam, S. Easteal, L.S. Jermin, G.K. Schulte, D.E. Danley, L.R. Hoth, M.C. Griffor, A.V. Kamath, M.H. Rosner, B.A. Chrunyk, D.E. Perregaux, C.A. Gabel, K.F. Geoghegan, J. Pandit, J. Biol. Chem. 275 (2000) 24798–24806.
[15] T.M. Buetler, D.L. Eaton, Environ. Carcinog. Ecotoxicol. Rev. C10 (1992) 181–203.
[16] M.C.J. Wilce, M.W. Parker, Biochim. Biophys. Acta 1205 (1994) 1–18.
[17] H. Dirr, P. Reinemer, R. Huber, Eur. J. Biochem. 220 (1994) 645–661.
[18] G. Di Pietro, L.A.V. Magno, F. Rios-Santos, Expert Opin. Drug Metab. Toxicol. 6 (2010) 153–170.
[19] S. Mahajan, W.M. Atkins, Cell. Mol. Life Sci. 62 (2005) 1221–1233.
[20] T.K. Hitchens, B. Mannervik, G.S. Rule, Biochemistry 40 (2001) 11660–11669.
[21] X. Ji, P. Zhang, R.N. Armstrong, G.L. Gilliland, Biochemistry 31 (1992) 10169–10184;
I. Sinning, G.J. Kleywegt, S.W. Cowan, P. Reinemer, H.W. Dirr, R. Huber, G.L. Gilliland, R.N. Armstrong, X. Ji, P.G. Board, B. Olin, B. Mannervik, T.A. Jones, J. Mol. Biol. 232 (1993) 192–212.
[22] S. Mahajan, W.M. Atkins, Cell. Mol. Life Sci. 62 (2005) 1221–1223;
B. Shi, R. Stevenson, D.J. Campopiano, M.F. Greaney, J. Am. Chem. Soc. 128 (2006) 8459–8467.
[23] L. Stella, M. Nicotra, G. Ricci, N. Rosato, E.E. Di Iorio, Proteins: Struct. Funct. Genet. 37 (1999) 1–9.
[24] M.J. de Groot, M.P.E. Vermeulen, D.L.J. Mullenders, G.M. Donné-Op den Kelder, Chem. Res. Toxicol. 9 (1996) 28–40;
R.T. Koehler, H.O. Villar, K.E. Bauer, D.L. Higgins, Proteins: Struct. Funct. Genet. 28 (1997) 202–216;
D.S. Hamilton, X. Zhang, Z. Ding, I. Hubatsch, B. Mannervik, K.N. Houk, B. Ganem, D.J. Creighton, J. Am. Chem. Soc. 125 (2003) 15049–15058;
I.A. Axarli, D.J. Rigden, N.E. Labrou, Biochem. J. 382 (2004) 885–893;
N. Kinsley, Y. Sayed, S. Mosebi, R.N. Armstrong, H.W. Dirr, Biophys. Chem. 137 (2008) 100–104;
R. Artali, G. Beretta, P. Morazzoni, E. Bombardelli, F. Meneghetti, J. Enzyme Inhib. Med. Chem. 24 (2009) 287–295;
I. Axarli, N.E. Labrou, C. Petrou, N. Rassias, P. Cordopatis, Y.D. Clonis, Eur. J. Med. Chem. 44 (2009) 2009–2016;
J. Son, J.-J. Lee, J.-S. Lee, A. Schűller, Y.T. Chang, ACS Chem. Biol. 5 (2010) 449–453.
[25] M.I. Zavodszky, P.C. Sanschagrin, R.S. Korde, L.A. Kuhn, J. Comput. Aided Mater. Des. 16 (2002) 883–902.
[26] R.A. Friesner, J.L. Banks, R.B. Murphy, T.A. Halgren, J.J. Klicic, D.T. Mainz, M.P. Repasky, E.H. Knoll, M. Shelley, J.K. Perry, D.E. Shaw, P. Francis, P.S. Shenkin, J. Med. Chem. 47 (2004) 1739–1749;
T.A. Halgren, R.B. Murphy, R.A. Friesner, H.S. Beard, L.L. Frye, W.T. Pollard, J.L. Banks, J. Med. Chem. 47 (2004) 1750–1759.
[27] J.B. Cross, D.C. Thompson, B.K. Rai, J.C. Baber, K. Yi Fan, Y. Hu, C. Humblet, J. Chem. Inf. Model. 49 (2009) 1455–1474.
[28] C. de Graaf, P. Prospisil, P. Wouter, G. Folkers, N.P.E. Vermeule, J. Med. Chem. 48 (2005) 2308–2318;
M.L. Verdonk, G. Chessari, J.C. Cole, M.J. Hartshorn, C.W. Murray, J.W.M. Nissink, R.D. Taylor, R. Taylor, J. Med. Chem. 48 (2005) 6504–6515;
C. Roberts, R.L. Mancera, J. Chem. Inf. Model. 48 (2008) 397–408;
R. Thilagavathi, R.L. Mancera, J. Chem. Inf. Model. 50 (2010) 415–421.
[29] H. Frauenfelder, G.A. Petsko, D. Tsernoglou, Nature 280 (1979) 558–562;
G.A. Petsko, D. Ringe, Annu. Rev. Biophys. Bioeng. 13 (1984) 331–371;
M. Gerstein, A. Lesk, C. Chothia, Biochemistry 22 (1991) 6739–6749;
Z. Yuan, J. Zhao, Z.X. Wang, Protein Eng. 16 (2003) 109–114.
[30] S.J. Teague, Nat. Rev. Drug Discov. 2 (2003) 527–541;
P. Cozzini, G.E. Kellogg, F. Spyrakis, D.J. Abraham, G. Constantino, A. Emerson, F. Fanelli, H. Gohlke, L.A. Kuhn, G.M. Morris, M. Orozco, T.A. Pertinhez, M. Rizzi, C.A. Sotriffer, Proteins: Struct. Funct. Bioinform. 66 (2007) 399–421.
[31] S.-Y. Huang, X. Zou, J. Med. Chem. 51 (2007) 6237–6255;
K.L. Damm, H.A. Carlson, J. Am. Chem. Soc. 129 (2007) 8225–8235.
[32] See for instance: J.C. Cole, C.W. Murray, J.W.M. Nissink, R.D. Taylor, R. Taylor, Proteins 60 (2005) 325–332.
[33] Y. Lu, R. Yang, C.Y. Yang, S. Yang, J. Chem. Inf. Model. 47 (2007) 668–675.
[34] Z. Deng, C. Chuaqui, J. Singh, J. Med. Chem. 47 (2004) 337–344;
R.K. Nandigam, S. Kim, J. Singh, C. Chuaqui, J. Chem. Inf. Model. 49 (2009) 1185–1192.
[35] Maestro, Version 7.0, Schrödinger, LLC, New York, NY, 2005.
[36] A.J. Oakley, M.L. Bello, A. Battistoni, G. Ricci, J. Rossjohn, H.O. Villar, M.W. Parker, J. Mol. Biol. 274 (1997) 84–100.
[37] W.L. Jorgensen, D.S. Maxwell, J. Tirado-Rives, J. Am. Chem. Soc. 118 (1996) 11225–11236;
G.A. Kaminski, R.A. Friesner, J. Tirado-Rives, W.L. Jorgensen, J. Phys. Chem. B 105 (2001) 6474–6487.
[38] MacroModel, Version 9.0, Schrödinger, LLC, New York, NY, 2005.
[39] L. Prade, R. Huber, T.H. Manoharan, W.E. Fahl, W. Reuter, Structure 5 (1997) 1287–1295.
[40] W. Clark Still, A. Tempczyk, R.C. Hawley, T. Hendrickson, J. Am. Chem. Soc. 112 (1990) 6127–6129.
[41] Glide, Version 3.5, Schrödinger, LLC, New York, NY, 2005.
[42] R Development Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2011, http://www.R-project.org/.