Acrylamide Binding to Its Cellular Targets: Insights from Computational Studies

Emmanuela Ferreira de Lima¹ and Paolo Carloni²

¹Scuola Internazionale Superiore di Studi Avanzati – SISSA - Statistical and Biological Physics Sector, via Bonomea, Trieste
²German Research School for Simulation Sciences, Jülich Research Center and RWTH-Aachen University, Italy

To the memory of Ernesto Illy (1925 – 2008), who suggested this project.

1. Introduction

Acrylamide (AC, CH₂=CH–CONH₂, Chart 1), present in heated foodstuffs [Rice, 2005], has been classified by the International Agency for Research on Cancer as “probably carcinogenic to humans” (group 2A) (IARC, 1994). The Scientific Committee on Toxicity, Ecotoxicity and the Environment (CSTEE) in the European Union (EU) demonstrated that AC exposure to humans should be controlled as low as possible because of its inherently toxic properties [Zhang et. al., 2009; Dearfield et. al., 1995]. AC is a low molecular-weight, odorless, and colorless compound. It is readily soluble in water and can rapidly polymerize from a monomer state to a polymer form [Klaunig, 2008]. AC is biotransformed in vivo to its epoxide glycidamide (GDE) by cytochrome P450 2E1 (CYP2E1) [Ghanayem et. al., 2005]. GDE has genotoxic properties in both in vitro and in vivo test systems [Kurebayashi & Ohno, 2008].

In spite of the possible carcinogenic nature of AC, no consistent effect of AC exposure on cancer incidence in humans could be identified [Rice, 2005]. This strikingly contrasts with AC submistration in both mice and rats, which may cause tumors at multiple sites [Besaratinia & Pfeifer, 2005; 2007]. A plausible hypothesis is this might be caused, at least partially, by the fact that AC interacts differently with the mouse and human proteins. AC may interact with cysteine residues not engaged in S-S bridges. Indeed, the double bond of conjugated vinyl compounds has strong affinity with SH groups [Friedman, 2003; Carere, 2006].

![Fig. 1. AC structure.](http://www.intechopen.com)
There are six human proteins which are known to bind to AC. For all of them, structural information is available. These are five enzymes [Howland et. al., 1980; Sakamoto & Hashimoto, 1985] and the serum albumin protein [Ferguson et al., 2010] (Tab. 1). All of them are dimers. Inspection of the X-ray structures of all of these targets shows Cys residues may be present in the enzymatic active sites and/or solvent exposed (Tab. 1).

2. Objective

This work has two main goals. On one hand, using bioinformatics, molecular docking, and molecular simulation procedures, we aim at predicting the structural determinants of AC in complex with its cellular targets reported in Tab. 1. On the other hand, we aim at detecting the difference (if any) of binding these molecules to the correspondent proteins in rodents. These differences might contribute to the carcinogenic features of these molecules in rodents as opposed to humans.

<table>
<thead>
<tr>
<th>Protein</th>
<th>PDB ID of human proteins</th>
<th>Sequence Identity (mouse)</th>
<th>Cys not engaged in disulfur bridges</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topoisomerase IIα</td>
<td>1ZXM</td>
<td>88%</td>
<td>Cys216, Cys104, Cys392, Cys405, Cys170, Cys300</td>
</tr>
<tr>
<td>Creatine Kinase (CK)</td>
<td>3B6R</td>
<td>96%</td>
<td>Cys74, Cys283*, Cys254, Cys141, Cys146</td>
</tr>
<tr>
<td>Aldolase</td>
<td>1QO5</td>
<td>95%</td>
<td>Cys72, Cys239, Cys289, Cys338, Cys201, Cys149, Cys177, Cys134</td>
</tr>
<tr>
<td>Serum albumin</td>
<td>1AO6</td>
<td>72%</td>
<td>Cys34</td>
</tr>
<tr>
<td>glyceraldehyde-3-phosphate dehydrogenase (GAPDH)</td>
<td>1ZNQ</td>
<td>93%</td>
<td>Cys152*, Cys156, Cys247</td>
</tr>
<tr>
<td>Enolase</td>
<td>2PSN</td>
<td>97%</td>
<td>Cys118, Cys356, Cys336, Cys338, Cys398, Cys388</td>
</tr>
</tbody>
</table>

Table 1. Structural information of AC targets. For human proteins, for which the experimental structures are available, the PDB codes are included. For mouse proteins, as there is no structure available, the sequence identities with human proteins are reported. The cysteines present in the active sites of the enzymes are shown with an asterisk while the solvent-exposed ones are highlighted in bold.

3. Methodology

3.1. Molecular docking

Docking methods attempt to find the “best” matching between two molecules, typically a receptor and a ligand [Halperin et. al., 2002]. Hence, they make the prediction of ligand conformation and orientation within a targeted binding site [Kitchen et. al., 2004; Halperin
et. al., 2002]. This prediction is carried out by performing a conformational space search based on an ad hoc potential energy function. [Halperin et. al., 2002].

The accuracy of the method can be investigated by docking ligands into the protein from which they were extracted (self-docking) and by docking them into its target protein in a different conformation (usually taken from different protein/ligand complexes) (cross-docking) [Kawatkar et. al., 2009; Sutherland et. al., 2007].

The increasing availability of protein three-dimensional structures combined with continuing computational advances have made docking a very important tool for small-molecule lead discovery [Campbell et. al., 2003].

Here we use the Genetic Optimisation for Ligand Docking (GOLD) program [Jones et. al., 1995; 1997]. GOLD is an automated ligand docking program that uses a genetic algorithm to explore the full range of ligand’s conformational flexibility with partial flexibility of the protein. It satisfies the fundamental requirement that the ligand must displace loosely bound water on binding [Jones et. al., 1995; 1997].

GOLD uses a genetic algorithm (GA) to optimize a variety of structural parameters: (a) dihedrals of ligand’s rotatable bonds; (b) ligand’s ring geometries; (c) dihedrals of protein OH groups and NH$_3^+$ groups; and (d) the position of the ligand in the binding site [Verdonk et. al., 2003].

GOLD uses a genetic algorithm (GA) to optimize a variety of structural parameters: (a) dihedrals of ligand’s rotatable bonds; (b) ligand’s ring geometries; (c) dihedrals of protein OH groups and NH$_3^+$ groups; and (d) the position of the ligand in the binding site [Verdonk et. al., 2003].

GOLD uses a genetic algorithm (GA) to optimize a variety of structural parameters: (a) dihedrals of ligand’s rotatable bonds; (b) ligand’s ring geometries; (c) dihedrals of protein OH groups and NH$_3^+$ groups; and (d) the position of the ligand in the binding site [Verdonk et. al., 2003].

In GOLD, one minimizes a molecular mechanics-like function with four terms, the so called GOLD Fitness (Eq. 1) [Jones et al., 1995; 1997]:

$$\text{GOLD Fitness} = S_{\text{hb,ext}} + S_{\text{vdw,ext}} + S_{\text{hb, int}} + S_{\text{vdw, int}}$$

The scoring function is taken as the negative of the GOLD Fitness [Jones, 1995; 1997]. The larger the scoring function of a pose, the higher its rank.

We have performed GOLD-based docking for the human proteins in Tab. 1. These are: human topoisomerase IIα [Sciandrello et al, 2010]; human-brain creatine kinase [Lü Z-R. et al, 2009]; human aldolase [Dobryszycki et al., 1999a; 1999b]; human serum albumin [Ferguson et al., 2010]; human glyceraldehyde-3-phosphate dehydrogenase [Sakamoto & Hashimoto, 1985] and human enolase [Howland et. al., 1980]. Since alkylation by acrylamide is limited to cysteine SH groups, the best cys residues putatively binding AC will be those ones with the shortest distances between S atom and AC.

For rodents, there is no experimental structure available. Fortunately, the sequence identity (SI) between the mouse proteins and the human enzymes is always greater than 88%, except for serum albumin, for which SI= 72% (Tab. 1 and Fig. S1, SI). The structure of the latter was built by homology modeling using the server HHpred [Soding, 2005; Soding et. al., 2005; Hildebrand et. al., 2009].

The active site was defined here as a sphere with radius of 10 Å centered at the sulfur atoms of the Cys residues of Tab. 1. The cavity detection algorithm, LIGSITE [Hendlich et. al., 1997], was used to restrict the region of interest to concave, solvent-accessible surfaces. To be sure that most of the high-affinity binding modes were explored, the genetic algorithm was run 200 times for each complex examined.

The poses were ranked using the GOLD Fitness [Jones et al., 1995; 1997] as well as in terms of the number of contacts between AC and the target cysteine residues.
4. Results and discussion

Here we report the results for AC docking to all its known human targets [Friedman, 2003]. We also discuss the putative binding of AC to the correspondent mouse proteins, for which binding has not been shown. We consider here Cys residues not involved in S-S bridges (Target cysteines hereafter, see Tab. 1).

Table 2 shows the docking score values for the best AC poses in all targets, the values are in Kcal/mol. It can be seen that the best score obtained was for creatine kinase, -46.2, this result shows the great protein-ligand interaction provided by the docking calculations, this result is in good agreement with experiments, which have shown that CK is alkylated by AC.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Docking score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topoisomerase IIα</td>
<td>-23.3</td>
</tr>
<tr>
<td>Creatine Kinase (CK)</td>
<td>-46.2</td>
</tr>
<tr>
<td>Aldolase</td>
<td>-21.3</td>
</tr>
<tr>
<td>Serum albumin</td>
<td>-22.6</td>
</tr>
<tr>
<td>glyceraldehyde-3-phosphate dehydrogenase (GAPDH)</td>
<td>-28.1</td>
</tr>
<tr>
<td>Enolase</td>
<td>-28.3</td>
</tr>
</tbody>
</table>

Table 2. Docking score values. The lower the score, the better the binding.

In the topoisomerase enzyme (Fig. 2), the cysteine(s) that react with AC were not identified experimentally. The X-ray structure [Wei et. al., 2005] features six target cysteines in each subunit, four of which are solvent-exposed (Tab. 1). Our docking procedure suggests that, in the most likely pose, AC interacts with Cys 405 (Fig. 3). In all figures, distances are reported in Angstroms.

![3D structure of topoisomerase enzyme](Fig. 2. 3D structure of topoisomerase enzyme [Sciandrello et al, 2010]. The four exposed cysteines (Cys104, Cys216, Cys392 and Cys405) are named and indicated in licorice representation.)
Fig. 3. AC interactions with Lys287, Met303 and Cys405 residues inside topoisomerase enzyme.

The SI between human and mouse proteins is as high as 88% (Fig. S2, SI). In particular, all of the residues surrounding the target cysteines are conserved. Inspection of the structure and of the sequence alignment in Fig. S2 strongly suggests that the chemical environment of the target cysteines in the mouse proteins is basically the same as in the human ones. The same argument applies to all of the enzymes considered here. Hence, AC might bind to mouse topoisomerase in a similar way as it does to the human protein.

The human enzyme creatine kinase (Fig. 4) has five target cysteines in each subunit (Tab. 1). One of them (Cys283) is located in each of the active sites (Fig. 5) and it is known experimentally to bind to AC [Matsuoka et. al., 1996; Meng et. al., 2001]. Consistently with experiment, we find that in the best pose AC interacts with Cys283 (Fig. 5). AC is not found to bind to any other target cysteine.

The SI between human and mouse proteins is considerably high (96%) and hence also in this case AC might bind to the mouse creatine kinase in a similar fashion.

The human enzyme enolase (Fig. 6) has 6 target cysteines in each binding site. Our docking procedure suggests that Cys388 may bind to AC (Fig. 7). Because of the large SI between human and mouse protein (97%), AC might bind to the correspondent Cys residue (Cys388, see Fig. S3, SI) also in the mouse protein.

The human enzyme glyceraldehyde-3-phosphate dehydrogenase (GAPDH) (Fig. 8) has three target cysteines in each dimer. One of them is present in each active sites, Cys152 (Fig. 9). Experiments show that Cys152 in fact interacts with AC [Campian et. Al., 2002]. Our docking is consistent with the experimental evidence (Fig. 9).

The protein human aldolase (Fig. 10) have eight and one target cysteine (Cys239, about 100 times more reactive than remaining exposed groups) in each subunit. Human serum albumin (Fig. 11) has 35 cysteine residues with 17 disulphide bonds. Sulphhydryl residue in position 34 is left free to react with thiols of the environment [Candiano et al., 2009]. Our docking procedure could identify binding poses for AC in neither proteins. This suggests that conformational rearrangements, which are not taken into account in the docking procedure, might allow AC to bind to one or more Cys residues. For mouse aldolase, the SI
Fig. 4. 3D structure of creatine kinase enzyme [Lü Z-R. et al, 2009]. The Cys283, which is located in the active side is named and indicated in licorice representation.

Fig. 5. AC interactions with Leu201, Thr59 and Cys283 residues inside creatine kinase active site.
Fig. 6. 3D structure of enolase enzyme [Howland et al., 1980]. The Cys388, which binds AC is named and indicated in licorice representation.

Fig. 7. AC interactions with Val142, Lys421, Ser140 and Cys388 residues, inside enolase enzyme.
Fig. 8. 3D structure of GAPDH enzyme [Sakamoto & Hashimoto, 1985]. The Cys152, which binds AC is named and indicated in licorice representation.

Fig. 9. AC interactions with Gly212, Thr211, Ser151 and Cys152 residues inside GAPDH active site.
is as high as 95%. This suggests similar consideration for this species. Instead, for mouse serum albumin, SI = 72% (see Fig. S1, SI). In the structure of the mouse serum albumin, which was built by homology modeling, there are two target cysteines (at position 58 and 603). Therefore, we may expect rather different binding in the two species. Because of the limitations of the docking procedure with homology models, we did not proceed to investigate AC binding poses to this protein [Leach, 2001; McGovern & Shoichet, 2003].

Fig. 10. 3D structure of aldolase enzyme [Dobryszycyki et al., 1999a; 1999b]. All the cysteine residues are in licorice representation with indication of the most reactive one, Cys239.

Fig. 11. 3D structure of serum albumin [Ferguson et al., 2010]. The only cysteine free to react, Cys34 is indicated in licorice representation.
5. Conclusions
By means of molecular docking methodology, we have studied the interactions between AC and its human targets. The investigation is complemented by a study of AC interactions with the mouse protein, for which binding has not been reported so far. The calculations are consistent with the available biochemical data and they provide novel information on putative cysteines to which AC could be bound in both mouse and human protein. In the case of one protein, serum albumin, binding is likely to occur at different locations in the protein. Hence, this difference could contribute to the experimentally known differences in toxic properties of AC in humans and mice.

6. Acknowledgment
One of the authors (Lima, EF) is recipient of a grant from the ‘Fondazione Ernesto Illy’, Trieste-Italy.

7. References
Acrylamide Binding to Its Cellular Targets: Insights from Computational Studies


Nowadays it is difficult to imagine an area of knowledge that can continue developing without the use of computers and informatics. It is not different with biology, that has seen an unpredictable growth in recent decades, with the rise of a new discipline, bioinformatics, bringing together molecular biology, biotechnology and information technology. More recently, the development of high throughput techniques, such as microarray, mass spectrometry and DNA sequencing, has increased the need of computational support to collect, store, retrieve, analyze, and correlate huge data sets of complex information. On the other hand, the growth of the computational power for processing and storage has also increased the necessity for deeper knowledge in the field. The development of bioinformatics has allowed now the emergence of systems biology, the study of the interactions between the components of a biological system, and how these interactions give rise to the function and behavior of a living being. This book presents some theoretical issues, reviews, and a variety of bioinformatics applications. For better understanding, the chapters were grouped in two parts. In Part I, the chapters are more oriented towards literature review and theoretical issues. Part II consists of application-oriented chapters that report case studies in which a specific biological problem is treated with bioinformatics tools.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following: