ILP-2 modeling and virtual screening of an FDA-approved library: a possible anticancer therapy

Saeed KHALILI1, Hemn MOHAMMADPOUR2, Mahideh SHOKROLLAH BAROUGH3,4, Parviz KOKHAEI3,5,*

1Department of Medical Biotechnology, Faculty of Medical Sciences, Tarbiat Modares University, Tehran, Iran
2Department of Immunology, Faculty of Medical Sciences, Tarbiat Modares University, Tehran, Iran
3Cancer Research Center and Department of Immunology, Semnan University of Medical Sciences, Semnan, Iran
4Students’ Research Committee, Semnan University of Medical Sciences, Semnan, Iran
5Department of Oncology-Pathology, Immune and Gene Therapy Lab, Cancer Center Karolinska (CCK), Karolinska University Hospital Solna and Karolinska Institute, Stockholm, Sweden

* Correspondence: parviz.kokhaei@ki.se

Background/aim: The members of the inhibitors of apoptosis protein (IAP) family inhibit diverse components of the caspase signaling pathway, notably caspase 3, 7, and 9. ILP-2 (BIRC-8) is the most recently identified member of the IAPs, mainly interacting with caspase 9. This interaction would eventually lead to death resistance in the case of cancerous cells. Therefore, structural modeling of ILP-2 and finding applicable inhibitors of its interaction with caspase 9 are a compelling challenge.

Materials and methods: Three main protein modeling approaches along with various model refinement measures were harnessed to achieve a reliable 3D model, using state-of-the-art software. Thereafter, the selected model was employed to perform virtual screening of an FDA approved library.

Results: A model built by a combinatorial approach (homology and ab initio approaches) was chosen as the best model. Model refinement processes successfully bolstered the model quality. Virtual screening of the compound library introduced several high affinity inhibitor candidates that interact with functional residues of ILP2.

Conclusion: Given the 3D structure of the ILP2 molecule, we found promising inhibitory molecules. In addition to high affinity towards the ILP2 molecule, these molecules interact with residues that play pivotal rules in ILP2-caspase interaction. These molecules would inhibit ILP2-caspase interaction and consequently would lead to reactivated cell apoptosis through the caspases pathway.

Key words: BIRC-8, apoptosis, Robetta, modeling

1. Introduction
As a physiological cell death process, apoptosis extensively regulates development, hemostasis, and immune responses of the cells. A balance between proapoptotic and antiapoptotic signals determines the cell’s fate (1). Inhibitors of apoptosis protein (IAP) family members are one of the main apoptosis regulators (2). IAP family members inhibit many types of the caspase signaling pathways, namely caspase 3, 7, and 9 (3). Human IAPs family proteins are composed of eight members, including cellular IAP 1 (c-IAP1), cellular IAP 2 (c-IAP2), IAP-like protein 2 (ILP-2), melanoma IAP (ML-IAP), X-chromosome-linked IAP (XIAP), neuronal apoptosis inhibitory protein (NAIP), BIR repeat-containing ubiquitin conjugating enzyme (4), and survivin. All members are characterized by the presence of a BIR domain (5).

ILP-2 (BIRC-8) is the most recently identified member of the IAP family, containing a RING finger domain and only one BIR domain. ILP-2 expression is restricted to humans and great apes. Various normal tissues such as the testis and lymphoblastoid tissue express ILP-2. This protein inhibits cell death through Bax or caspase 9. More interestingly, instances of increasing expression levels of ILP-2 were recently reported in breast cancer patients. However, Fas ligand and tumor necrosis factor-mediated cell death has not been induced following ILP-2 overexpression (6).

Overexpression of IAP family members is reported in different types of cancers by inhibition of caspases and indirect modulation of NF-kB signaling. Targeting the functions of the IAP family members is an intriguing strategy to overcome cancer cells that are increasingly resistant to standard chemo and radiation therapies (7).
In this regard, modeling of the ILP-2 3-dimensional (3D) structure as a member of the IAP family would be of great significance. A 3D structure would be applicable to find novel inhibitors for ILP2 and caspase 9 interactions. In the present study, we aimed to find such inhibitors capable of solving cancer cells resistance to death.

2. Materials and methods

2.1. Sequence retrieval and BLAST search
Protein sequence of ILP-2 was obtained from UniProt (Universal Protein Resource) knowledgebase at http://www.uniprot.org/. To perform homology modeling predictions, the NCBI protein BLAST tool at http://blast.ncbi.nlm.nih.gov/Blast.cgi was used to arrive at a suitable template structure. The BLAST plan was restricted for Homo sapiens only while Protein Data Bank proteins were set to be the target database; all other parameters were set as default.

2.2. Protein modelling
Since BLAST search did not find any suitable template for ILP-2 homology modeling, we used both fold recognition and initio modeling approaches for model construction. The I-TASSER server (http://zhanglab.ccmb.med.umich.edu/I-TASSER/), building its 3D models based on multiple-threading alignments by LOMETS and iterative template fragment assembly simulations, was employed for the ILP-2 structure prediction. According to the CASP7, CASP8, CASP9, and CASP10 experiments, I-TASSER ranked as the no. 1 server for accurate structure prediction. The other software employed for ILP-2 3D structure prediction was Robetta (http://robetta.bakerlab.org/). This server predicts protein domain structures based on both ab initio and comparative modeling approaches. Using a complete automated method, the Rosetta de novo protocol was used for domains modeling without a detectable PDB homology study while detecting template PDBs, which are used to build comparative models by locally installed versions of HHSEARCH/HHpred, RaptorX, and Sparks-X.

2.3. Model quality assessment
To assess the quality of the obtained files from the predicted models, PDB files of models were input into the QMEAN model quality assessment server (http://swissmodel.expasy.org/qmean/cgi/index.cgi). In order for the composite scoring function to estimate both global and local errors on the basis of one single model, QMEAN could help to determine the best predicted model in further QMEAN assessment and structure validation. The Prosa server (https://prosa.services.came.sbg.ac.at/prosa.php) was used for further structure validation.

2.4. Model refinement and molecular dynamics analyses
As the first step through the model refinement process, the ModLoop server (http://modbase.compbio.ucsf.edu/modloop/) was used to rectify modeling errors on the selected best model. Loop remodeling was executed on a loop spanning residues between 150 and 160 residues of the model that apparently match the QMEAN residue error plot, indicating a high residue error peak. The result model was further modified by a full atomic 3D refinement run employing the server at http://sysbio.rnet.missouri.edu/3Drefine/. This server modifies protein structures by a two-step protocol, initialized by optimizing the hydrogen bonding network and followed by an atomic-level energy minimization. Finally, to further refine the model, the CHARMMING server (http://charmming.org/) was harnessed to conduct a molecular dynamics run on the refined model. The structure was minimized and solvated by the CHARMMING server before the main molecular dynamics run executed. The parameters of the molecular dynamics run were set to the following: 1000 steps, starting temperature of 210.15, final temperature of 310.15, temperature increment of 10, steps between temperature increments of 100, and bath temperature of 310.15.

2.5. Final model validation
To assess the accuracy of prediction, PDB files with PDB ID of 1XB1 were fed to the Protein Data Bank File Editor by Jonas Lee to acquire a single BIR Domain of ILP-2. The identified BIR domains in the predicted model were superimposed onto equivalent atoms to calculate the root mean square deviation (RMSD) and the topology score using the CLICK server (http://mspc.bia.a-star.edu.sg/minhn/pairwise.html) and the iPBA webserver (http://www.dsimb.inserm.fr/dsimb_tools/ipba/index.php). Meanwhile, the TM-align server (http://zhanglab.ccmb.med.umich.edu/TM-align/) was used to compute the TM score and RMSD. The stereochemical quality of the final achieved model was assessed using Procheck software (http://swissmodel.expasy.org/) to evaluate the geometry of the residues in the given protein structure. Moreover, the atomic empirical mean force potential ANOLEA (http://swissmodel.expasy.org/p) was used to check the packing quality of the model, performing energy calculations on the protein chain.

2.6. Compound library preparation and virtual screening
The ZINC12 database (http://zinc.docking.org/), containing about 35 million compounds for structure-based virtual screening, was used to retrieve a compound library. The SDF file of a library containing FDA-approved compounds provided from the DrugBank Database was downloaded from the available prepreared categories. PyRx 0.8, available for free download at http://pyrx.sourceforge.net/downloads, was employed to carry out docking calculations. PyRx 0.8 uses a large body of established open-source software such as AutoDock Vina and AutoDock 4 wizard, AutoDock Tools, Python,
Open Babel. AutoDock Vina software is a new program for molecular docking and virtual screening analysis that brings about approximately 2 orders of magnitude speed-up compared with AutoDock 4. This software significantly improves the accuracy of the binding mode predictions. Therefore, we used the Open Babel tool to import the SDF file of the compound library. Energy minimization runs and conversion to PDBQT format were executed on all imported compounds by the Open Babel tool. The ILP-2 model was prepared for docking analysis, adding hydrogen atoms and merging all nonpolar hydrogen. Calculations were performed with a grid of $28.54 \times 25.15 \times 34.38$ xyz points, at grid center of (xyz) $-7.05, 14.97, -1.74$ to cover the residues of the BIR domain.

2.7. Compound selection and visual inspection
All compounds with binding energy greater than or equal to $-8$ kcal/mol were selected for visual inspection. The compounds were visually inspected for their spatial location regarding the caspase 9 interacting amino acids using 3D structure visualizers like Pymole and Discovery Studio Visualizer. The structurally important residues for ILP2 and caspase 9 were assigned using data reported by Sun et al. (18). The schematic diagram of detailed protein–ligand interactions were formed using the LigPlus program.

3. Results
3.1. Sequence and homology analyses
The protein sequence of Baculoviral IAP repeat-containing 8 (BIRC8) or ILP-2 was retrieved from the UniProt knowledgebase under the Q96P09 ID code. It is a cytoplasmic protein comprising 236 amino acids and containing a BIR domain. Unlike the BLAST search, results from the PDB Database using this sequence as a query return similar sequences, the best of which (PDB ID: 1XB1) belongs to the 3D structure of the ILP-2, covering only 40% of the whole protein length.

3.2. 3D model construction and quality assessment
Full length protein 3D models of ILP-2 were successfully built by the I-TASSER and Robetta servers. Based on their scoring algorithms, each server provides 5 top predicted models. Quality assessment z-scores were calculated for the best models predicted by both QMEAN and Prosa servers (Table 1).

<table>
<thead>
<tr>
<th>Quality assessment</th>
<th>QMEAN z-score</th>
<th>Prosa z-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>I-TASSER</td>
<td>-4.4</td>
<td>-5.95</td>
</tr>
<tr>
<td>Robetta</td>
<td>-2.9</td>
<td>-5.98</td>
</tr>
<tr>
<td>Refined model</td>
<td>-2.5</td>
<td>-5.75</td>
</tr>
</tbody>
</table>

3.3. Model refinement and molecular dynamics
Feeding the best predicted model into the loop remodeling process resulted in resolving the existing high residue error peak, spanning a region containing 150-10 amino acids. Loop modeling together with refinement performed by the 3D refine server improved the quality z-scores for both QMEAN and Prosa servers (Table 1). Ultimately the molecular dynamics analyses, performed on the refined model, formed the final coordinates of the ILP-2 model (Figure 1).

3.4. Final model validation
RMSD calculations following the superimposition between the final model and the BIR domain indicate that the equivalent residues of the predicted model take similar coordinates to the experimentally resolved structure (Figure 2). RMSD values were 1.09, 1.26, and 1.26 for Click, iPBA, and TMalign respectively. Meanwhile, the TM score and topology score were 0.8 and 1, respectively, for superimposed structures. The quality score of the final model was $-0.32$ for the Procheck total G-factor. The Ramachandran plot for the finally achieved structure revealed that more than 90% of residues are in the allowed regions (Figure 3). ANOLEA results for the finally refined model indicated that most of the amino acids are in their favorable energy environment with acceptable QMEAN scores (Figure 4).

3.5. File preparation and virtual screening
The library of FDA-approved compounds contains 2136 molecules, some of which are different conformations.
of the same molecule. The energy of all compounds is minimized and converted to PDBQT format using Open Babel tool. PDBQT file format for the predicted model was generated using PyRx. Performed virtual screening analysis resulted in several predicted conformations of the docked compound and the macromolecule for each compound of the library along with their binding energy.

3.6. Compound selection

Over 2100 compound/protein interactions were predicted using AutoDock Vina software, among which 67 complexes got ≤–8 kcal/mol binding energy. All these complexes were visually inspected and 14 complexes were found to be in a spatial location suitable to interfere with interactions of caspase 9-interacting ILP2 amino acids. Among the 14 compounds, all 14 were found to have actual interactions with caspase 9-interacting ILP2 amino acids. Table 2 lists these compounds and their properties along with a list of caspase 9-interacting ILP2 amino acids. Figure 5 shows spatial location of the 14 selected compounds in

![Figure 2](image1.png)

Figure 2. Superimposed structures of the predicted BIR model with the experimentally resolved structure of the same region.

![Figure 3](image2.png)

Figure 3. A Ramachandran plot for the finally achieved model. Only 2 amino acids (GLU34 and THR 95) are in the disallowed region.
Figure 4. ANOLEA and QMEAN plots for the finally achieved model. Negative values represent a favorable energy environment for a given amino acid, indicating the accuracy of the modelling process. Lower QMEAN values correspond to regions in the model being potentially more reliable.
interaction with the BIR domain of ILP 2. The detailed interaction network of each ligand is depicted in Figure 6. According to these diagrams each selected compound is in interaction with the residues, which are important for ILP2 and caspase 9 interaction.

4. Discussion
Bioinformatics is an evolving field in contemporary biology, aiming at implementing computers and their simulations and calculations to solve biological challenges. Bioinformatics insinuates itself into various fields of biology, including immunology (8–11), structural biology, and molecular interaction studies (12). Especially in the postgenomic era, bioinformatics helps in dealing with overwhelming amounts of produced data, avoiding ethical aspects of animal use, reducing the cost of empirical studies, designing and searching for novel therapeutics, and presenting novel hypotheses. In the present study we employed bioinformatics tools to delve into an FDA-approved compound library in search of potential inhibitors of ILP-2 and caspase 9 interactions. To this end, initially we tried to build a full-length ILP-2 model. As the most reliable modelling approach, homology modelling was considered to model the ILP-2 molecule. However, there was no amenable template for full length model development. Sharing less than 30% identity between query and subject makes the prediction more likely to fail accurate modeling, affected by alignment errors (13). Since the best existing template just covered 40% of the protein and the rest was without any templates, homology approaches failed to build a full length ILP-2 model. However, threading and ab initio protein modelling approaches successfully managed to build ILP-2 models. The QMEAN z-score is an absolute quality score that is independent of protein size. Relating the model’s structural features to experimental structures of similar size, QMEAN z-score could be used to select between alternative structures of a modeled protein. Since the Robetta model gets better z-scores, we decided to perform the following analyses on this model.

Although the best modeling criteria were contemplated during the 3D protein modeling, most of the achieved models are spurious and their coordinates show

<table>
<thead>
<tr>
<th>ZINC ID</th>
<th>Binding energy (kcal/mol)</th>
<th>List of caspase 9-binding amino acids of ILP2</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZINC00020243</td>
<td>–8.9</td>
<td>A90</td>
</tr>
<tr>
<td>ZINC19594557</td>
<td>–8.5</td>
<td>H82</td>
</tr>
<tr>
<td>ZINC02568036</td>
<td>–8.4</td>
<td>R85</td>
</tr>
<tr>
<td>ZINC52955754</td>
<td>–8.3</td>
<td>E85</td>
</tr>
<tr>
<td>ZINC03978005</td>
<td>–8.2</td>
<td>E53</td>
</tr>
<tr>
<td>ZINC33359785</td>
<td>–8.2</td>
<td>D54</td>
</tr>
<tr>
<td>ZINC95862733</td>
<td>–8.2</td>
<td>R25</td>
</tr>
<tr>
<td>ZINC01550477</td>
<td>–8.1</td>
<td>E21</td>
</tr>
<tr>
<td>ZINC01612996</td>
<td>–8.1</td>
<td>V18</td>
</tr>
<tr>
<td>ZINC01996117</td>
<td>–8.1</td>
<td>Y16</td>
</tr>
<tr>
<td>ZINC00057278</td>
<td>–8</td>
<td>M15</td>
</tr>
<tr>
<td>ZINC00538275</td>
<td>–8</td>
<td>I9</td>
</tr>
<tr>
<td>ZINC28240499</td>
<td>–8</td>
<td>W14</td>
</tr>
<tr>
<td>ZINC53073961</td>
<td>–8</td>
<td>Q33</td>
</tr>
<tr>
<td>–</td>
<td>–</td>
<td>N48</td>
</tr>
<tr>
<td>–</td>
<td>–</td>
<td>W49</td>
</tr>
<tr>
<td>–</td>
<td>–</td>
<td>E88</td>
</tr>
<tr>
<td>–</td>
<td>–</td>
<td>E71</td>
</tr>
<tr>
<td>–</td>
<td>–</td>
<td>G3</td>
</tr>
<tr>
<td>–</td>
<td>–</td>
<td>T13</td>
</tr>
</tbody>
</table>
KHALILI et al. / Turk J Med Sci

discrepancies from native protein structure. To arrive at robust models and bolster their associated errors, loop remodeling and model refinement seem inevitable. Molecular dynamics provides an opportunity to simulate atomic motions for a certain period of time. During the dynamics run, the positions of the atoms are changed according to Newton’s laws of motion. All possible forces involved in atom motions would be exerted by a force field and finally the model could get more native-like coordinates (14). After performing all refinement processes, our results reveal that the finally achieved model has high quality scores. Assigning over 90% of its residues in the favored regions of a Ramachandran plot, getting RMSD in the range of closely homologous proteins values (<3 Å) between the experimentally resolved protein and the predicted model, getting a maximum topology score of 1 (which indicates topologically identical structures) (15), a TM score of >0.5 for the superimposition fold (which means the structures share the same SCOP/CATH-two prominent protein structural classification) (16), and favorable energy environment for most amino acids of the predicted structure favors the high quality of the final model. The existence of a high quality region spanning 71–81 amino acids is due to the algorithm used by the Robetta server. This region corresponds to the alpha-Helix of the ILP2 protein, which belongs to the BIR domain according to the Uniport database; since there are crystallographically resolved structures for the BIR domain, Robetta uses homology modelling to model this region. Therefore, due to the existence of a suitable template to model the BIR domain, this region is modeled with high quality. The high quality peak of this region could be rationalized considering this fact.

Virtual screening of a large compound library against a target protein is known to be a useful method to select hits and search for leads from a vast database. Virtual screening, compared with laboratory experiments, is more cost effective, time effectual, labor efficient, and always a sensible option to reduce the initial number of compounds before using high-throughput screening methods (17). Using this strategy would pave the way to find possible agonists and antagonists of target proteins. The BIR domain of the ILP2 protein is sequentially and structurally most relevant to the BIR3 domain of the XIAP protein. Therefore, equivalent residues in the XIAP BIR3 domain and the ILP2 BIR domain would contribute in BIR and Caspase 9 interactions. Given these residues, determined by Sun et al. (18), it would be possible to find ILP2 inhibiting compounds. To consider a compound as a potential inhibitor, it should be in suitable spatial location and contacting with essential residues of the ILP2 and caspase 9 interaction. Meanwhile, the interacting compounds should have a stable interaction with the ILP2 molecule. Chang et al. (19) defined a threshold of –7.0 kcal/mol that works well to discriminate between putative specific and nonspecific bindings with HIV protease. They claim that applying this threshold to data sets may be useful in filtering out noise in weakly binding compounds (19). Since this threshold is defined for AutoDock users, we used a threshold equal to –8.0 kcal/mol for our results to be more restrictive in compound selection. Herein, 14 compounds were screened out of an FDA-approved library, all of which met the main criteria for a binding energy threshold, spatial location suitability, and interaction with caspase 9-interacting ILP2 amino acids. It could be extrapolated that these compounds would stably occupy the caspase 9 interaction hot spots. This would spatially inhibit their interaction, while the caspase 9-interacting ILP2 amino acids are preoccupied interacting with selected compounds.

ILP-2 protects cells against apoptosis induction by the Bax protein. Its interaction with caspases 9, especially in cancerous cells, results in death resistance of tumor cells (20). As a conserved mechanism of IAP family members, the BIR domain of the ILP-2 binds to caspases 3 and 9, and inhibits apoptosis. The IBM interacting groove is the most conserved surface structure in BIR-2 and BIR-3 domains that interacts with caspase 3 and 7 and caspase 9,
respectively (21–23). Therefore, the compounds introduced in Table 2 may efficiently inhibit the caspase interactions of the ILP-2 or other BIR domain-containing IAP family members. Due to their high affinity and their interaction with functionally important residues of the BIR domain, these compounds could be considered for functional inhibition of ILP-2 and subsequently increased apoptosis and elevated cell susceptibility for current treatments. Moreover, since these compounds are screened out of an FDA-approved library, there is less concerns about their clinical applications.

In conclusion, using an integrative method, a 3D model of ILP-2 was constructed and used to screen a compound library. Consequently, exerting restrictive criteria, several

Figure 6. The interaction diagram of all 14 compounds in interaction with the ILP2 molecule. Each compound and the amino acids that have interactions with it are depicted.
potential inhibitors of its interaction with caspase 9 were introduced. Finally, the achieved compounds could efficiently interact with ILP-2 and inhibit ILP-2 functions that may lead to activated cell apoptosis through caspase pathway. The high homology of the ILP-2 model (especially the BIR domain) with other members of the IAP family suggests that these compounds could have the same inhibitory effect on the other members of the family.

References