

BIOSYNTHETIC PATHWAY AND MOLECULAR DOCKING OF METHYL ISOCYANATE (MIC) - p53 INTERACTION STUDY

Rahul Shrivastava^{1*}, Sumukh Deshpande², Pooja Purohit¹ and Rashi Arora¹

¹ Department of Bioinformatics, Maulana Azad National Institute of Technology, Bhopal, India

²University of Calgary, Department of Biological Science, Calgary, Alberta, Canada

Summary

This paper describes about in-silico analysis of biochemical pathways and inhibition by methyl-isocyanate using bioinformatics approach in human beings. Major cause of diseases or disorders found in inhibition of biochemical pathways by methyl-isocyanate. Bhopal Gas Tragedy marks the toxic effects of methyl-isocyanate (MIC) in humans. Biosynthetic pathways studies reveal p53 protein is present in many cancer causing pathways in humans. Some of them include prostate cancer, small cell lung cancer, bladder cancer and Huntington's disease found in victims of Bhopal Gas Tragedy. Protein ligand docking is performed using Glide docking software. p53-MDM2 interacting residues are selected for docking MIC into the residues of p53. Docking is performed by generating a receptor grid map of p53 onto where MIC can bind and generates energy score. The lower the score, the better the interaction of ligand with the receptor. This information will helpful for MIC interaction with different residues of the proteins and develop a better ligand based pharmacophore model of the complex with chemical nature of the compound such as hydrophobic, aromatic, hydrogen bond acceptor or donor by database search which generate compounds with similar chemical properties used for developing target specific drugs for cancer causing genes in humans.

Keywords: MIC, transactivation domain, connecting proteins, docking, GScore.

*Corresponding Author

Dr. Rahul Shrivastava

Dept. of Bioinformatics,

Maulana Azad National Institute of Technology,

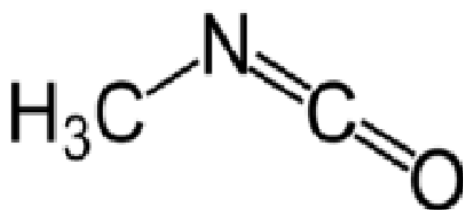
Bhopal-452051, India

E mail – rahulshrivastava_2000@yahoo.com

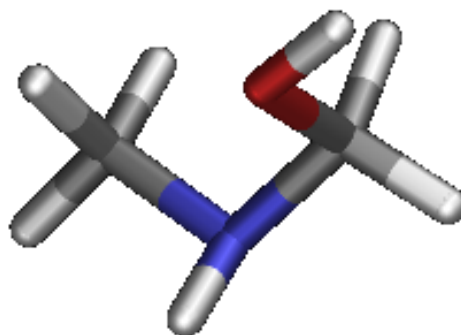
Introduction

The p53 protein is a tumour suppressor protein and is a regulator of cell cycle and one of the frequently altered targets in the majority of human neoplasia.¹ The function of p53 is to maintain DNA integrity and apoptosis of abnormal DNA which is impossible to be repaired. It has been seen that mutation in p53 causes pathogenesis of various cancers including gall bladder cancer and is one of the factor involved in the formation of cancer. Mutations in p53 tumor suppressor is a target of many carcinogens. Mutations causes lung cancer which ranges from 40-70 %. The function of p53 is to act as checkpoint responding to a variety of stress signals that can originate from several external and internal sources. The p53 protein integrates these signals and activate a signalling network that respond to minimize mutation and other errors that can lead to cancer or other pathologies. Tumour suppressor genes resides in p53 network in which p53 forms a central node that receives many stress signals, and transmit these to signalling pathways controlling cell cycle and apoptosis. Tumour suppressor function of p53 is its ability to prevent cell cycle progression and activate apoptosis in response to genotoxic and non-genotoxic stresses, thereby preventing the damaged DNA from cycling into next generation. The tumour suppressor is known as “guardian of genome”. In order to escape the “safeguard” system mediated by p53, human cancers mutate p53 itself.² In tumours, the p53 pathway is mostly inactivated by its negative regulator MDM2. MDM2 is antagonist of p53 and promotes ubiquitination of p53 followed by degradation in proteasome.³

Methyl Isocyanate is a toxic chemical having CID no. 12228 and molecular formula of C_2H_3NO shown in (Fig 1) Methyl Isocyanate (MIC) is prepared industrially by reacting methylamine with phosgene, oxidizing monomethylformamide at high temperatures ($> 550C$), or heating metal methyl isocyanates. Because of its high reactivity, MIC is used as an intermediate in organic synthesis, most notably in the production of carbamate based pesticides. Studies based on MIC showed that victims of Bhopal gas tragedy who were subjected to longer periods of exposure had an acute damage that led to death was mainly to the respiratory system, most likely pulmonary oedema, bronchospasm, and electrolyte imbalance. However extra pulmonary damage, including tissue anoxia, gastrointestinal symptoms, and muscular weakness, Fibrosis of the lungs were also observed. Within a year of the exposure, survivors continued to exhibit damage to the lung and eyes. Apart from these, some of the victims also suffered from HIV infections, appendicitis, Leukemia, Lung cancer, breast cancer, pancreatic cancer, malaria, etc.^{4,5}



(A)



(B)

Fig 1: (A) Structure of methyl isocyanate (MIC), (B) Minimized structure of MIC

Methyl Isocyanate is capable of undergoing bio transfusion reactions and is chief metabolic intermediate in industrial setting. Human exposure to MIC can lead to severe hypersensitive, mutagenic and genomic instability by promoting cell cycle arrest and apoptosis. Genomic instability is the most common molecular abnormality in mammalian cells that accelerates the process of carcinogenesis. Genomic instability initiates an increased rate of alterations, arising either from an excessive amount of damage overwhelming the ability of normal repair systems to restore genomic integrity, or defective repair systems being unable to cope with normal rates of damage generated through normal cellular and environmental mechanism.¹

Methods

MIC affects the expression of many genes/proteins.⁶ These information about the chemical gene interaction can be found out from CTD (Comparative Toxicogenomics Database) in which it tells us which protein/gene is affected by MIC (**Table 1**). These proteins are found in humans. Proteins affected by MIC are present in the biochemical pathways which take part in metabolism of products leading to chain of reactions. Some of these proteins take part in Allograft Rejection Pathway, Cytokine-cytokine receptor interaction, T-cell receptor signalling pathway, Antigen processing and presentation, p53 signalling pathway, etc. The information about biosynthetic pathways is cited from KEGG and REACT database. The information about the affected proteins by MIC and there respected database (**Table 2**). From this table, MIC affects the expression of IFNG, CCNE1, CDK2, CDKN1A, IL2, IL4, IL10, TP53, MDM2 in humans and SLC4A1 in Rats.

MDM2 interacts with transactivation domain of p53 which inhibits transcriptional activity of p53 thus inhibiting cell cycle progression and cell apoptosis giving rise to malignant tumor growth leading to cancer.⁷ Protein-ligand docking is performed in which MIC is used as ligand and p53 complexed with DNA is used as a receptor. Interaction of MIC with p53 is studied and energy score is found out. The residue which interacts well with the ligand attains the lowest score. This amino acid residue will be the target for developing drugs for treatment of cancers.

Table 1: List of proteins affected by MIC and it's pathways.

| Affected protein | Chemical | Affected pathways |
|------------------|----------|---|
| IFNG | MIC | Allograft rejection, Cytokine-cytokine receptor interaction, Jak-STAT signaling pathway, T cell receptor signaling pathway, Chagas disease, Leishmaniasis, Graft-versus-host disease, Systemic lupus erythematosus, Type I diabetes mellitus, Antigen processing and presentation, Natural killer cell mediated cytotoxicity, Proteasome, Regulation of autophagy, TGF-beta signaling pathway |
| IL2 | MIC | Allograft rejection, Cytokine-cytokine receptor interaction, Jak-STAT signaling pathway, T cell receptor signaling pathway, Autoimmune thyroid disease, Chagas disease, Intestinal immune network for IgA production, Graft-versus-host disease, Type I diabetes mellitus |
| IL4 | MIC | Allograft rejection, Cytokine-cytokine receptor interaction, Jak-STAT signaling pathway, T cell receptor signaling pathway, Autoimmune thyroid disease, Intestinal immune network for IgA production, Leishmaniasis, Asthma, Fc epsilon RI signaling pathway, Hematopoietic cell lineage |
| IL10 | MIC | Allograft rejection, Cytokine-cytokine receptor interaction, Jak-STAT signaling pathway, T cell receptor signaling pathway, Autoimmune thyroid disease, Chagas disease, Intestinal immune network for IgA production, Leishmaniasis, Asthma, Systemic lupus erythematosus |

| | | |
|--------|-----|---|
| CDK2 | MIC | Cell cycle, Cell Cycle Checkpoints, p53 signaling pathway, Pathways in cancer, Prostate cancer, Cell Cycle, Mitotic, Small cell lung cancer, DNA Replication, Oocyte meiosis, Progesterone-mediated oocyte maturation |
| CCNE1 | MIC | Cell cycle, Cell Cycle Checkpoints, p53 signaling pathway, Pathways in cancer, Prostate cancer, Cell Cycle, Mitotic, Small cell lung cancer, Oocyte meiosis |
| CDKN1A | MIC | Cell cycle, Cell Cycle Checkpoints, p53 signaling pathway, Pathways in cancer, Prostate cancer, Cell Cycle, Mitotic, Bladder cancer, DNA Replication, Glioma, Melanoma, ErbB signaling pathway, Signalling by NGF |
| TP53 | MIC | Cell cycle, Cell Cycle Checkpoints, p53 signaling pathway, Pathways in cancer, Prostate cancer, Small cell lung cancer, Bladder cancer, Chronic myeloid leukemia, Glioma, Melanoma, Amyotrophic lateral sclerosis (ALS), Apoptosis, Basal cell carcinoma, Colorectal cancer, Endometrial cancer, Huntington's disease, MAPK signaling pathway, Neurotrophin signaling pathway, Non-small cell lung cancer, Pancreatic cancer, Thyroid cancer, Wnt signaling pathway |

Table 2: List of the connecting proteins present in the pathways affected by MIC

| Proteins | Connecting proteins |
|----------------|---|
| IFNG | PD1, CTLA4, CD45, CD4/8, CD3e, TCR α , CD40L, ICOS, CD28, NF κ B, AP1, NFAT, CaN, LCK, CBL, ZAP70, IGINR1, IFNGR2, TNF α , CTL, GM-CSF |
| IL2, IL4, IL10 | IL3, IL5, IL13, TGF β , GM-CSF, TNF α , CDK4 |
| CDK2, CDKN1A | Cyclin A, Cyclin E, P21, P27/Kip1, Rb, Skp2, CKS1, C-Myc, Max, RINGO, CPEB |
| TP53 | p300, DNA-PK, ATMATR, chk1,2, 143-3 σ , GADD45, cip1, p21, MDM2, Kinase, WAF1/CIP1, Reprimo, B99, |

| | |
|--|---|
| | Fas, PIDD, Bax, PIGs, Scotin, IGF-BP3, PAI, P48, PTEN, TSAP6, CBP, TSP1, EAAT2, SOD1, CaN, Bcl2, PUMA, Htt, P38, JNK, p73, Siah-1, SIP, Skp1, APC, Ebil |
|--|---|

Results

IFNG

Interferon Gamma (IFNG) is a member of type –II class of interferons and is a dimerized soluble cytokine. MIC results in increased expression of IFNG protein. The current view is that interferon γ (IFN γ) activates cells via interaction with the extracellular domain of the receptor complex. This in turn results in the activation of the receptor-associated tyrosine kinases JAK1 and JAK2, leading to phosphorylation and dimerization of the transcription factor STAT1 α , which dissociates from the receptor cytoplasmic domain and undergoes nuclear translocation.⁸ There are many pathways through which IFNG is produced and many products are produced by IFNG. The number of proteins/enzymes which are involved in production, manipulation, synthesis, catalysis of various reactions in human body (**Table 3**).

CDK2 and CDKN1A

Cyclin- dependent kinase 2 and cyclin-dependent kinase inhibitor 1A are proteins encoded by cyclin-dependent kinase family of Ser/Thr protein kinases. MIC affects the expression of CDK2 and CDKN1A proteins.⁹ Genetic variants in *CDKN1A* and *CDK2* may modulate breast cancer. In humans, *CDKN1A*, *CDKN1B* and *CDKN1C* (p57) are differentially expressed in normal versus atherosclerotic vessels. The proteins involved with CDK2 and CDKN1A (**Table 3**).

IL2, IL4 and IL10

Interleukin-2 (IL-2) is an interleukin, is pivotal for the generation and regulation of the immune response. A T lymphocyte product, IL-2 also stimulates T cells to undergo cell cycle progression via a finite number of interactions with its specific membrane receptors.¹⁰ Interleukin 2 (IL2) plays a role in B cell responsiveness. B cells could be induced to express receptors for IL 2, a number of functional studies suggest that IL2 might also play a role in B cell activation.¹¹

MIC results in increased expression of IL2, IL4 and IL10 protein. IL2 protein is produced by T-cells in response to antigenic or mitogenic stimulation, this protein is required for T-cell proliferation and other activities crucial to regulation of the immune response. It can stimulate B-cells, monocytes, lymphokine-activated killer cells, natural killer cells, and glioma cells. The proteins involved with interleukins (**Table 3**)

Table 3: Docking results showing best 20 poses for Arg282 by MIC

| S.No. | Title | Score | GScor | Lip | HBon | Met | Rew | vd | Coul | Rot | Site | Emod | Cvd | Intern | Con | Pose |
|-------|---------------|-------|-------|------|------|------|------|------|-------|------|------|-------|-------|--------|-----|------|
| 16 | MIC_minimized | -4.68 | -4.68 | -0.3 | -0.6 | -0.0 | -2.2 | -6.4 | -7.4 | -0.0 | -0.1 | -18.5 | -13.8 | 0.2 | 1 | 347 |
| 17 | MIC_minimized | -4.64 | -4.64 | 0.3 | 0.4 | 0.0 | 0.3 | 7.3 | -7.0 | 0.0 | 0.0 | -18.5 | -14.4 | 1.4 | 1 | 369 |
| 1 | MIC_minimized | -5.49 | -5.49 | -0.2 | -0.8 | -0.0 | -2.7 | -6.8 | -9.9 | -0.0 | -0.0 | -21.6 | -16.7 | 1.6 | 1 | 42 |
| 18 | MIC_minimized | -4.61 | -4.61 | 0.0 | -0.7 | -0.0 | -2.2 | -6.5 | -8.8 | -0.0 | -0.0 | -19.5 | -15.3 | 1.2 | 2 | 276 |
| 2 | MIC_minimized | -5.21 | -5.21 | -0.3 | -0.6 | -0.0 | -2.4 | -4.1 | -10.7 | -0.0 | -0.0 | -19.7 | -14.9 | 1.2 | 1 | 322 |
| 19 | MIC_minimized | -4.56 | -4.56 | 0.0 | -0.6 | -0.0 | -2.3 | -6.6 | -8.7 | -0.0 | 0.0 | -19.3 | -15.4 | 1.6 | 1 | 28 |
| 3 | MIC_minimized | -5.02 | -5.02 | -0.3 | -0.7 | -0.0 | -2.4 | -9.0 | -7.2 | -0.0 | -0.0 | -20.6 | -16.3 | 1.8 | 1 | 92 |
| 20 | MIC_minimized | -4.45 | -4.45 | 0.0 | -0.5 | -0.0 | -2.2 | -7.1 | -8.8 | -0.0 | -0.1 | -19.8 | -15.8 | 1.4 | 1 | 218 |
| 4 | MIC_minimized | -5.02 | -5.02 | -0.2 | -0.6 | -0.0 | -2.4 | -4.5 | -10.4 | -0.0 | -0.0 | -19.4 | -14.9 | 1.3 | 2 | 382 |
| 5 | MIC_minimized | -5.00 | -5.00 | -0.2 | -0.6 | -0.0 | -2.7 | -8.2 | -7.3 | -0.0 | -0.0 | -20.1 | -15.6 | 1.4 | 1 | 373 |
| 6 | MIC_minimized | -4.89 | -4.89 | 0.0 | -0.7 | -0.0 | -2.7 | -7.6 | -7.1 | -0.0 | -0.0 | -18.8 | -14.7 | 2.0 | 2 | 205 |
| 7 | MIC_minimized | -4.86 | -4.86 | -0.1 | -0.4 | -0.0 | -2.7 | -6.5 | -8.9 | -0.0 | -0.0 | -19.9 | -15.4 | 1.1 | 2 | 327 |
| 8 | MIC_minimized | -4.86 | -4.86 | -0.3 | -0.8 | -0.0 | -2.4 | -8.3 | -6.5 | -0.0 | -0.0 | -19.1 | -14.8 | 1.5 | 2 | 60 |
| 9 | MIC_minimized | -4.81 | -4.81 | -0.3 | -0.5 | -0.0 | -2.2 | -5.2 | -9.5 | -0.0 | -0.1 | -18.8 | -14.7 | 1.8 | 2 | 386 |
| 10 | MIC_minimized | -4.77 | -4.77 | 0.0 | -0.8 | -0.0 | -2.2 | -4.7 | -9.8 | -0.0 | -0.1 | -18.8 | -14.5 | 1.4 | 2 | 325 |
| 11 | MIC_minimized | -4.76 | -4.76 | -0.3 | -0.6 | -0.0 | -2.4 | -8.2 | -7.0 | -0.0 | -0.0 | -19.4 | -15.2 | 1.4 | 2 | 348 |
| 12 | MIC_minimized | -4.75 | -4.75 | -0.3 | -0.7 | -0.0 | -2.7 | -9.1 | -4.3 | -0.0 | -0.0 | -18.3 | -13.4 | 0.0 | 1 | 95 |
| 13 | MIC_minimized | -4.72 | -4.72 | 0.0 | -0.7 | -0.0 | -2.7 | -6.7 | -6.9 | -0.0 | -0.0 | -18.5 | -13.6 | 0.0 | 2 | 223 |
| 14 | MIC_minimized | -4.71 | -4.71 | -0.3 | -0.6 | -0.0 | -2.3 | -8.9 | -7.1 | -0.0 | -0.0 | -20.2 | -16.0 | 1.3 | 1 | 358 |
| 15 | MIC_minimized | -4.71 | -4.71 | -0.3 | -0.6 | -0.0 | -2.3 | -8.9 | -7.0 | -0.0 | -0.0 | -20.6 | -15.9 | 0.3 | 1 | 144 |

TP53

P53 is a tetrameric, sequence-specific DNA-binding protein and transcription factor that controls the expression of an array of gene products in response to diverse stress stimuli. This protein acts as a tumour suppressor which regulates cell division by keeping cells from growing and dividing too fast p53 has been implicated for example, in the control of cell growth, cell-cycle progression, apoptosis, DNA repair and angiogenesis. The p53 protein has at least four recognised domains: the N-terminal transcriptional activation, central DNA-binding, tetramerisation and the C-terminal-negative regulatory domains.¹²

The p53 protein is a transcription factor that binds a very loose DNA recognition sequence found in several hundred genes that are differentially activated depending on the cell type, identity, and extent of damage, and various other parameters that have yet to be identified [4–6]. The unique feature of TP53 compared to other tumour suppressor genes is its mode of inactivation. An important feature of the TP53 protein is the extreme flexibility and fragility of the DNA binding domain. Most TP53 mutations are localized in the DNA binding domain of the protein (residues 100–300) leading to a bias of TP53 mutation analysis.¹³

The TP53 (Tumour Protein 53) gene provides instructions for making a protein called tumour protein p53. This protein acts as a tumour suppressor which regulates cell division by keeping cells from growing and dividing too fast. Tumour protein p53 is located in the nucleus of cells throughout the body, where it binds directly to DNA. When the DNA in a cell becomes damaged by agents such as toxic chemicals, radiation, or ultraviolet (UV) rays from sunlight, this protein plays a critical role in determining whether the DNA will be repaired or the damaged cell will undergo apoptosis. If the DNA can be repaired, TP53 activates other genes to fix the damage. If the DNA cannot be repaired, this protein prevents the cell from dividing and signals it to undergo apoptosis. MIC changes the expression of this protein by binding to the core domain active sites and changes the DNA-binding activity of p53 protein. The loops and the loop-sheet-helix motif of p53 protein consist of conserved regions of the core domain and contains majority of p53 mutations identified in tumours. The molecular mechanism underlying IsoCyanate mediated inflammatory responses and their possible role in the onset of genomic instability in cultured IMR-90 cell line using N-Succinimidyl,N-Methylcarbamate surrogate chemical in human lung fibroblast has also been examined.¹

The p53 protein has the structure of a transcription factor and is made up of several domains viz a transactivation domain and a proline-rich domain are found at its NH₂ terminus. A DNA-binding domain is in the middle of the protein. A tetramerization domain and a regulatory region are present at its COOH terminus. This structure allows the tetrameric p53 protein, on binding to DNA, to regulate the expression of several genes with different biological functions, such as cell cycle regulation, apoptosis, DNA repair, and differentiation. In case of tumours like cancer, p53 is inactivated by MDM2. The murine double minute 2 (MDM2) gene is regulated by p53 pathway however the proper level of p53 is maintained by auto regulatory feedback. Mdm2 interacts primarily through its 100 residue N-terminal domain with the N-terminal transactivation

domain of p53. In cancer, overexpression of MDM2 protein takes place. P53 and MDM2 form an auto regulatory feedback loop which causes p53 to stimulate the expression of MDM2 which in turn inhibits p53. MDM2 interacts with p53 in 3 different ways. First, it inhibits the transcriptional activity of p53 by binding to its transactivation domain. Second, it promotes degradation of p53, which MDM2 acts as ubiquitin ligase hence ubiquitination of p53 leading to degradation. Third, MDM2 favours export of p53 by binding because p53 contains nuclear export signal.⁷ The proteins involved with p53 (**Table 3**).

P53-MDM2 Interaction

The regions involved in the interaction between p53 and MDM2 is the MDM2-binding domain on p53 and have residues between 1 and 52. The other is the p53-binding domain having residues between 1 and 118. The MDM2 acts as ubiquitin ligase promoting ubiquitination of p53 followed by degradation of proteasome. Since MDM2 gene is regulated by p53 pathway, the proper level of p53 is maintained by auto regulatory feedback. MDM2 interacts with p53 through its 100 residue N-terminal domain with the N-terminal transactivation domain of p53. MDM2 binds to p53 transactivation domain (**Fig 2**) and inhibits its transcriptional activity by promoting its degradation and favours the export of p53 from the nucleus. Therefore in presence of MDM2, p53 protein is inactivated and does not stimulate expression of genes involved in apoptosis or cell cycle arrest. In some of the tumours, when MDM2 is overexpressed, p53 is constantly inhibited and tumour growth is favoured. The inactivation of MDM2 in these tumours should activate p53 pathway which activates apoptosis.⁷

There are some strategies which are used to target MDM2 in tumours. Antisense oligonucleotide decreases the cellular levels of MDM2. Other strategy involves inhibiting the ubiquitin ligase activity of MDM2 which prevents p53 degradation. One more strategy involves p14ARF which acts by blocking MDM2-dependent degradation and transcriptional silencing of p53. Inhibitors of p53-MDM2 should release p53 from MDM2 and as a consequence p53 tumour suppressor activity.⁷

Mutations in the human p53 oncogene are found with high frequency in a wide variety of cancers. Mutations in p53 are found in tumours with hepatocellular carcinomas, colorectal cancer, lung cancer, breast cancer, brain tumours, bladder cancer, bone and soft tissue sarcomas, and head and neck cancer. Single base pair substitutions (Frame shift mutation) comprise the most common mutational event in p53 gene followed by small deletions. Mutations in p53 in non-small cell lung cancer(12) and head and neck cancer (14) show a distinct preference for GC → TA transversions. In addition to aflatoxin, UV light appears to produce specific mutations in the p53 gene in skin tumours, in particular inducing a high fraction of CC → TT tandem mutations. Most of the p53 point mutations found are GC to AT transitions both in skin and internal tumours. K-ras mutations were also detected in benign, borderline and malignant tumours.⁷

Many residues of p53 are responsible for interaction with MDM2. The following is cited from paper called “crystal structure of p53”. These are the six residues which are crucial for interaction with MDM2 and further eliciting responses. These are Arg248, Arg273, Arg175, Gly245, Arg249 and Arg282. Among these the two mutated residues directly contact the DNA. Arg248 (L-3 loop) makes the minor groove contact. Arg273 from (loop-sheet helix motif) contacts a back bone phosphate and the remaining four residues stabilize the structure of DNA binding surface of p53. Gly245 is the only non-arginine residue and plays a critical role in allowing the L3 loops to assume a conformation not favoured with residue containing side chain, the phi and psi dihedral angles respectively. Other mutated residues include Arg280 on H2 helix which contacts the invariant guanine in the major groove of the DNA (with 2.1% of mutation).¹⁴

The MDM2 residues involved in inhibition of p53 are Phe19, Trp23 and Leu26. Phosphorylation at serine residue in central acidic domain of MDM2 increase its ability to inhibit the p53 and it gets phosphorylated by kinase CK2 in vitro, the two serine residues are Ser260 and Ser269.¹⁴

Arg175 is the best example of a structural p53 mutation because of its critical role in stabilizing the L2 and L3 loops.¹⁴

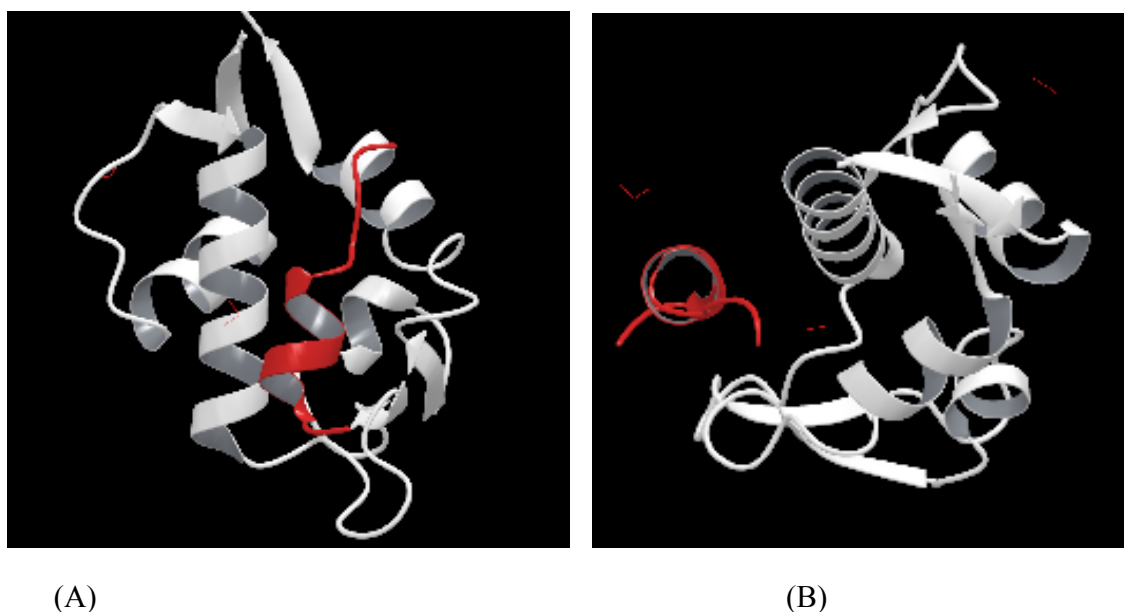


Fig 2: (A) MDM2 bound to p53 (Side view), (B) MDM2 bound to p53 (Top view)

Molecular Docking

Docking is a method which predicts the preferred orientation of one molecule to a second when bound to each other to form a stable complex. Docking is frequently used to predict the binding orientation of small molecule drug candidates to their protein targets in order to in turn predict the affinity and activity of the small molecule. Hence docking plays an important role in the rational design of drugs. There are various types of docking but since we want to find out how methyl isocyanate binds to the receptor and influences it, we focus only on protein-ligand docking.¹⁵

Protein-Ligand docking is a molecular modelling technique. The goal of protein-ligand docking is to predict the position and orientation of a ligand (a small molecule) when it is bound to a protein receptor or enzyme. Pharmaceutical research employs docking techniques for a variety of purposes, most notably in the virtual screening of large databases of available chemicals in order to select likely drug candidates.¹⁶ Protein-Ligand docking is performed by various docking servers and docking software's. There are number of servers, some of them are PatchDock, MEDock, RosettaDock, ClusPro, GRAMM-X, BBCU, etc. The various docking software's are GOLD, AutoDock, FlexX, Glide etc. and for this paper used Glide from Schrodinger Suite.

Glide uses a hierarchical series of filters to search for possible locations of the ligand in the active-site region of the receptor. The shape and properties of the receptor are represented on a grid by several different sets of fields that provide progressively more accurate scoring of the ligand poses. Conformational flexibility is handled in Glide by an extensive conformational search, augmented by a heuristic screen that rapidly eliminates unsuitable conformations, such as conformations that have long-range internal hydrogen bonds.

Protein Preparation

A typical PDB structure file consists only of heavy atoms, can contain waters, cofactors, and metal ions, and can be multimeric. The structure generally has no information on bond orders, topologies, or formal atomic charges. Terminal amide groups can also be misaligned, because the X-ray structure analysis cannot usually distinguish between O and NH₂. Ionization and tautomeric states are also generally unassigned. Glide calculations use an all-atom force field for accurate energy evaluation. Thus, Glide requires bond orders and ionization states to be properly assigned and performs better when side chains are reoriented when necessary and steric clashes are relieved.¹⁷

The first step in Docking is to load the protein molecule '3IGL' (**Fig 3**) and remove the DNA which is complexes with the molecule. This is done by selecting the chains which contain DNA. As the DNA is contained in chain B, so we delete the chain B and keep chain A.¹⁷

The next step is to prepare the protein 3IGL molecule in order to dock it with ligand. This is done by “Protein Preparation Wizard” from Workflows menu in Schrodinger Maestro. The protein is prepared which assigns bond order, adds hydrogen, create zero-order bonds to metals, converts selenomethionines to methionines and delete water beyond 5 Å from hetero groups. After the protein is modified, the structure is refined and minimized. This prepares the protein molecule for docking (**Fig 4**).¹⁷

After the protein is optimized, the ligand is prepared for docking. It is prepared in the same way the protein is prepared. Also partial charges are assigned to the ligand in order to dock with the protein. The ligand file is saved in ‘.mae’ extension which signifies maestro format since Glide only accepts .mae files for ligands.

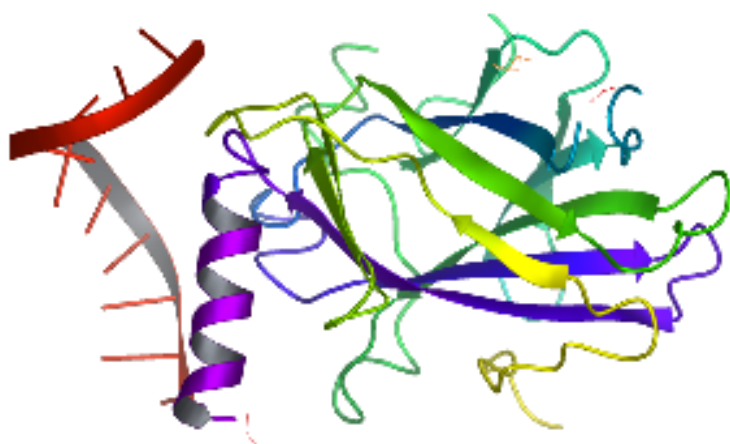


Fig 3: Structure of p53 protein complexes with DNA (3IGL)

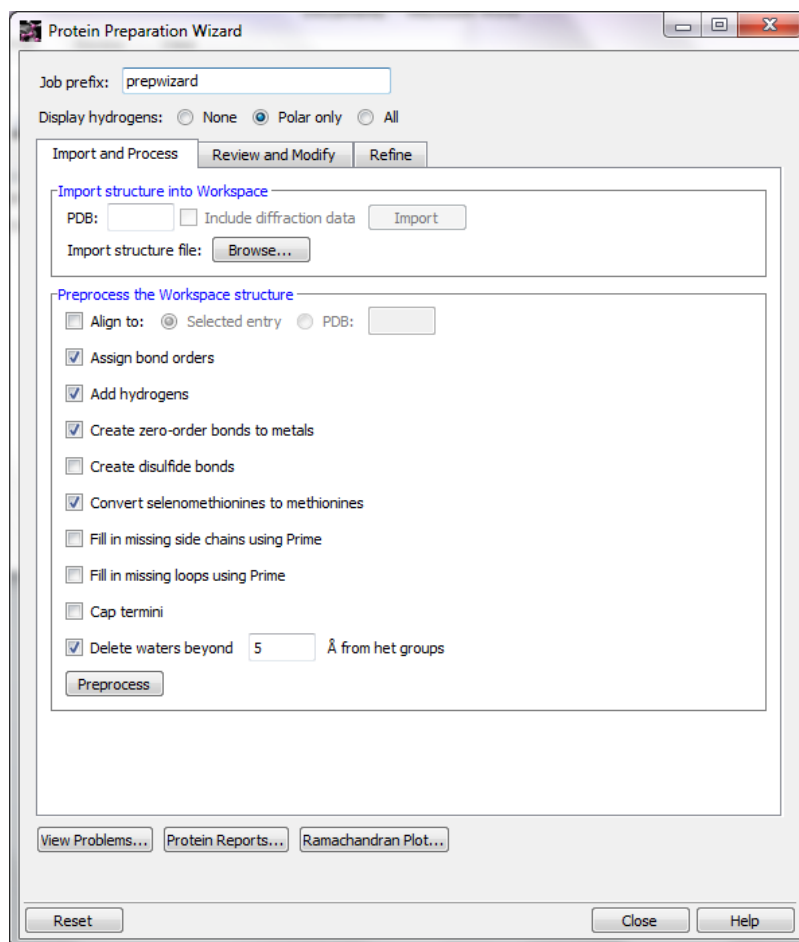


Fig 4: Protein preparation window in Schrodinger maestro software suite

Receptor Grid Generation

In this step a receptor grid map is prepared for the ligand to dock. For the protein 3IGL, a receptor grid map is prepared in which the ligand will come and bind to the active sites.¹⁷

Glide searches for favourable interactions between one or more ligand molecules and a receptor molecule, usually a protein. The shape and properties of the receptor are represented on a grid by several different sets of fields that provide progressively more accurate scoring of the ligand poses.¹⁷

The receptor grid can be set up and generated from the Receptor Grid Generation panel. The options in each tab of this panel allow you to define the receptor structure by excluding any cocrystallized ligand that may be present, determine the position and size of the active site as it will be represented by receptor grids, set up Glide constraints, and set up flexible hydroxyl groups. Ligand docking jobs cannot be performed until the receptor grids have been generated. Receptor grid generation requires a “prepared” structure: an all-atom structure with appropriate bond orders and formal charges. The force field used for grid generation is the OPLS_2005 force field, which allows a proper treatment of metals. Force field cannot be changed from the Receptor Grid Generation panel, however it can be changed by editing the input file.¹⁷

To open the Receptor Grid Generation panel, choose Receptor Grid Generation from the Glide submenu of the Applications menu. The Receptor Grid Generation panel has four tabs (**Fig 5**), which you can use to specify settings for the receptor grid generation job. These are 1) Receptor, 2) Site, 3) Constraints, and 4) Rotatable Groups.¹⁷

In this we define per-atom Van der Waals radius scaling and charge scaling for the active sites and by scaling mean that increase the size of the active sites which we want in order for the ligand to bind to that site. Also we use input partial charges. In the next tab we prepare the site of the receptor i.e. the active site of the receptor and define the coordinates of the active site i.e. X,Y,Z coordinates.¹⁷

This will create an inner box and outer box for the receptor site. Inner box defines how ligand can fit in the receptor and outer box defines where the area for the ligand to bind within that active site. Since we don't want to give any constraints and do not want to specify any rotamer groups, we will start the produce the receptor grid map of the protein. It takes approximately 5-10 minutes to produce this. After the receptor grid is produced, the file is saved as grid-receptor.zip format which contains all the information for docking to ligand. In developing the grid map, we took 6 active sites which are as follows 1) Arg248, 2) Arg273, 3) Arg175, 4) Gly245, 5) Arg249, and 6) Arg282.

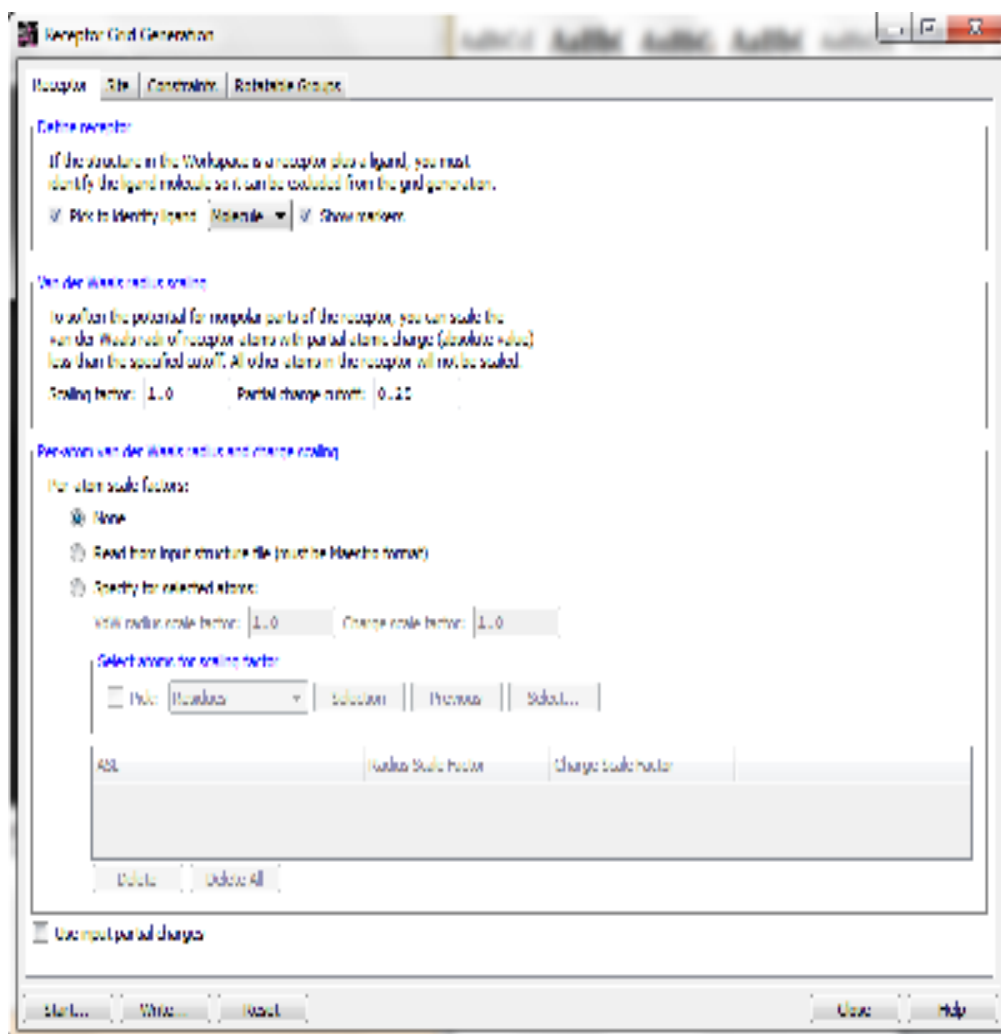


Fig 5: Defining site in receptor grid generation in Schrodinger maestro software suite

Docking

After the receptor grid map is finished, ligand should be docked to the receptor by ligand docking tool in Glide (Fig 7 & 8). In this window we input the grip map in 'zip' format which is already produced in receptor grid generation window in glide. There are two types of precision mode, standard precision (SP) and extra precision (XP). For ligand docking, SP mode is used. In the ligands section, the ligand file along with default values are uploaded.¹⁷

The options in the Selection of initial poses section of the dialog box control the way poses pass through the filters for the initial geometric and complementarity “fit” between the ligand and receptor molecules. The grids for this stage contain values of a *scoring function* representing how favourable or unfavourable it would be to place ligand atoms of given general types in given elementary cubes of the grid. These cubes have a constant spacing of 1 Å. The “rough score” for a given *pose* (position and orientation) of the ligand relative to the receptor is simply the sum of the appropriate grid scores for each of its atoms. By analogy with energy, favourable scores are negative, and the lower (more negative) the better. Poses that pass these initial screens enter the final stage of the algorithm, which involves evaluation and minimization of a grid approximation to the OPLS-AA non bonded ligand-receptor interaction energy.¹⁷

In the output section, we input 20 poses for the molecule and also checkmark the box to write the report file which will give us information about the score and the glide score for each pose. After setting the parameters, docking process is started. This is done for 6 active sites in the protein and each docking run produces 20 poses so total number of poses produced in the docking runs are 120 poses.¹⁷

The output result from docking gives final ranking based on glide score. Also the Coulomb interaction energy and Van der waals energy, hydrogen bond interaction and metal interaction energy should be greater than or equal to zero. If any of the above properties is not defined for a given pose, the corresponding filter is not applied to that pose.¹⁷

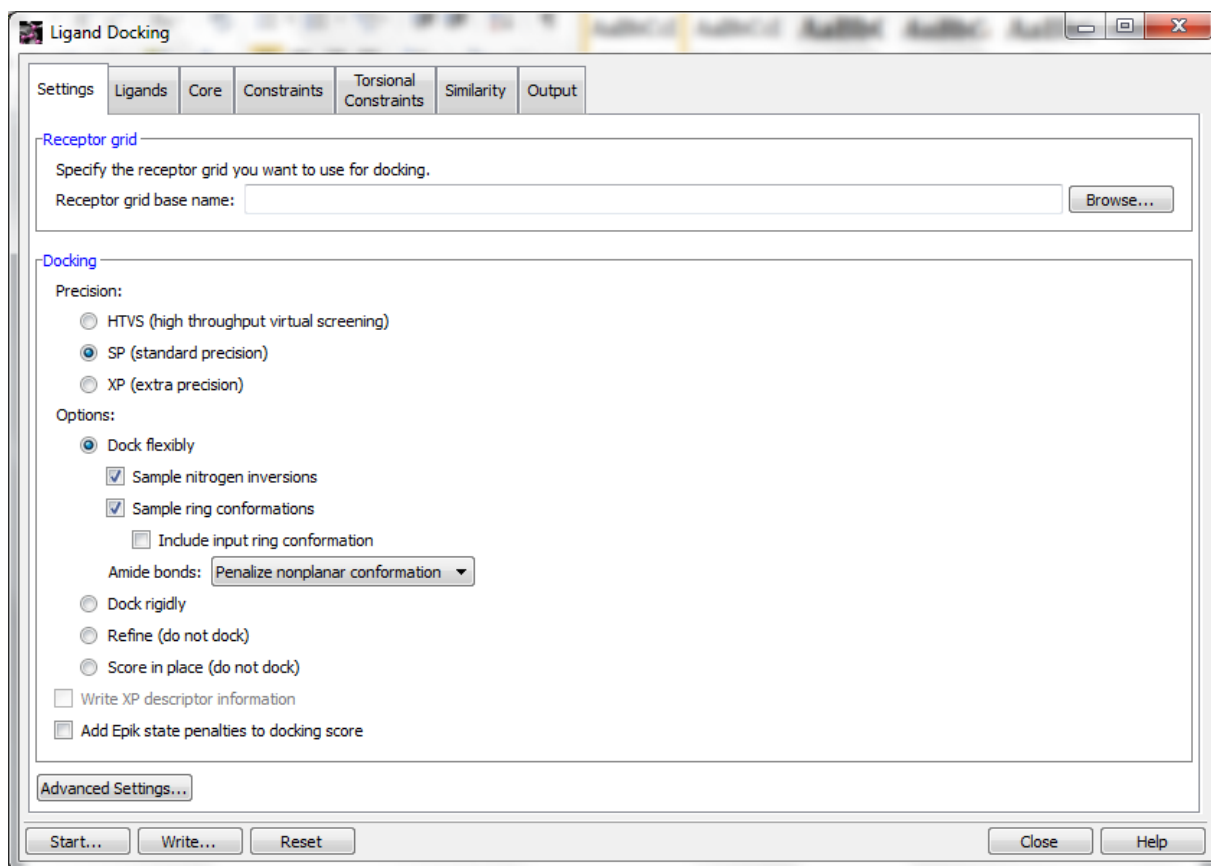


Fig 6: Ligand docking window in Glide Schrodinger maestro

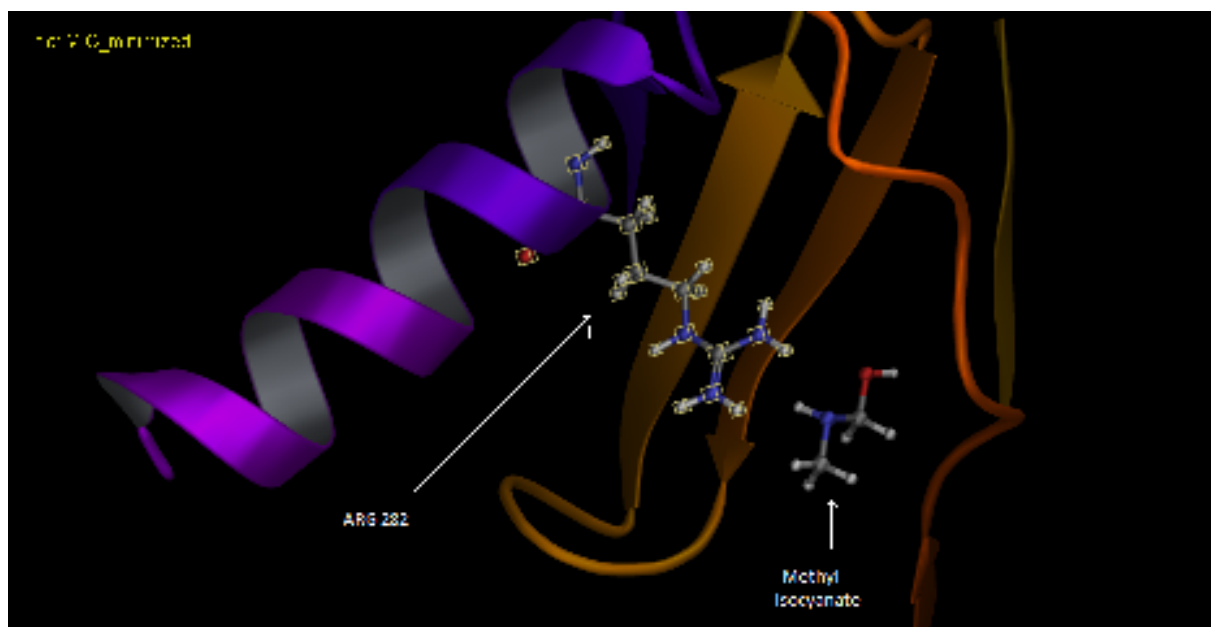


Fig 7: Docking of MIC with ARG 282 residue of p53.

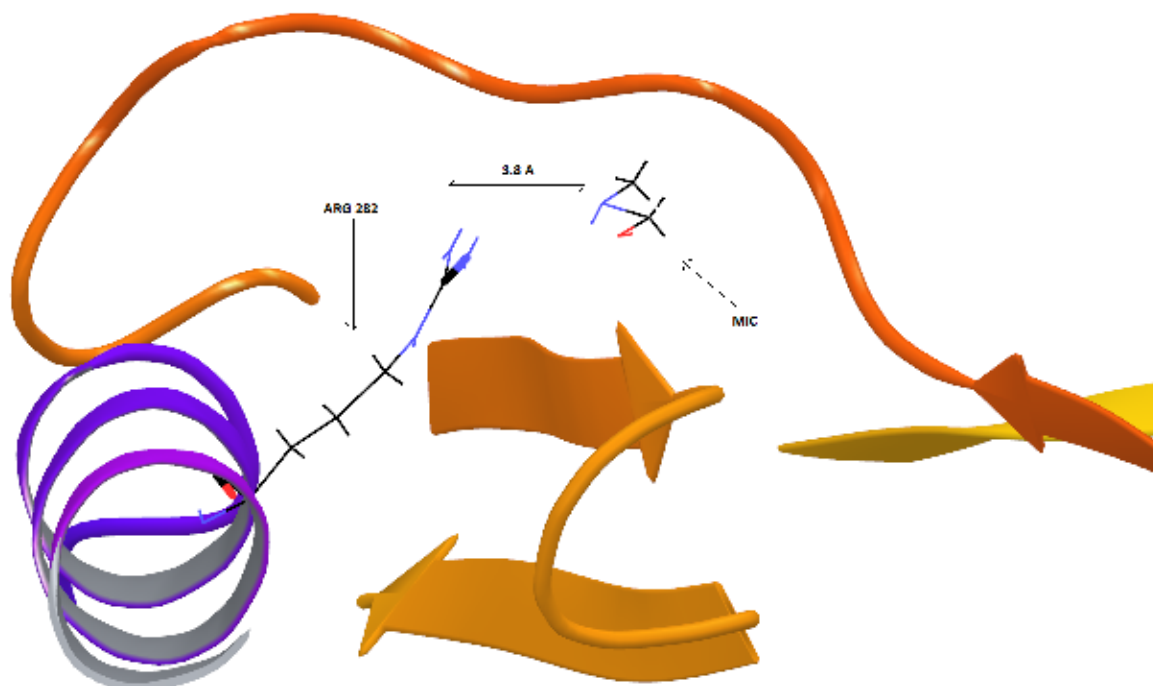


Fig 8: Final docking result showing MIC (ligand) docked to the active site on the receptor (Arg282) in p53.

Discussion

The present analysis of MIC showed that MIC affected many proteins in humans. The analysis of the connecting proteins from the biochemical pathways (KEGG) showed that p53 is the protein which is most susceptible to MIC. p53 is 22 times affected by MIC as compared to different proteins including IFNG, CDK2, CDKN1A, IL2, IL4, IL10. P53 is a tumour protein which acts as a tumour suppressor which regulates cell division. p53 is also inhibited by its connecting protein MDM2 which inhibits its transcriptional activity by binding to its transactivation domain which degrades its function of complexing with the DNA thereby causing unregulated growth of tumour cells in humans leading to cancer. Interaction of p53 with MDM2 was studied in order to determine the active site residues of p53. Six active site residues were determined based on the studies from the paper. Based on the studies, molecular docking was performed using Glide docking software. Glide is used for protein-ligand docking which generates accurate scoring of ligand poses. It first generates a receptor grid map onto which the ligand will bind. This generated glide score.

Result from the docking run performed by Glide gave best 20 poses for every amino acid residue defined as active site in the receptor molecule (3IGL). The result for every active site residue includes rank, title, score, glide score (GScore), lipophilic-contact plus phobic-attractive term (Lipo), hydrogen bonding term (HBond), metal-binding term (Metal), various rewards or penalty term (Rewards), Van der Waals interaction energy (vdW), coulomb interaction energy (Coul), penalty for freezing rotatable bonds (RotB), polar interactions in the active site (Site), estimated conformational energy of the ligand (EModel which is GScore+CvdW+Intern), non-bonded interaction energy between ligand and the receptor (CvdW which is Coul+vdW), Intern, conformations (Conf), pose and RMSD.

The Glide score is calculated by $GScore = a * vdW + b * Coul + Lipo + Hbond + Metal + Rewards + RotB + Site$, where coefficients of vdW and Coul are $a = 0.050$, $b = 0.150$ for Glide 5.0 (the contribution from the Coulomb term is capped at -4 kcal/mol), $a = 0.063$, $b = 0.120$ for Glide 4.5, $a = 0.080$, $b = 0.100$ for Glide 4.0.

Glide gives report for best 20 poses for Arg248. The best glide score for Arg248 is -4.41 having rank 1 for this docking run. Similarly the best glides score for Arg249 is -5.11. For Arg175, the best glide score is -4.55. For Arg282, the best glide score is -5.49 and for Gly245, the best glide score is -4.55. From all the poses Arg282 has the best glide score of -5.49.

Docking results shown in (**Table 3**), Arg282 has the best docking score of -5.49 compared to rest of the active site residues. These results show that Arg282 is the most affected residue by MIC followed by Arg249, Arg175, Gly245 and Arg248.

By performing experiments through bioinformatics approach find out that MIC affected several proteins in humans beings resulting various types of diseases and these proteins participate in biosynthetic pathways in humans and for further analysis of products/catalysts in specific pathways of diseases, specific information will be derived as for the change in expression of genes takes place in the pathways. This information will help pharmaceutical industries to target particular genes involved in the pathway which help for generating target specific drugs to cure the drastic effects of MIC in humans.

References

1. P.K. Mishra, A. Bhargava, G.V. Raghuram, S. Gupta, S. Tiwari, R. Upadhyaya, S.K. Jain and K.K. Maudar. 2009. Inflammatory response to isocyanates and onset of genomic instability in cultured human lung fibroblasts. 2009 ; Genet. Mol. Res. 8 (1): 129-143.
2. Grzegorz M. Popowicz, Anna Czarna and Tad A. Holak. Structure of the human Mdmx protein bound to the p53 tumor suppressor transactivation domain. Cell Cycle 7; 2008:15, 2441-2443; Landes Bioscience.

3. Grzegorz M. Popowicz, Anna Czarna, Ulli Rothweiler, Aleksandra Szwagierczak, Marcin Krajewski, Lutz Weber, Tad A. Holak. Molecular basis for the inhibition of p53 by Mdmx. *Cell Cycle* 6 ,2007;19, 2386-2392, 1 October; 2007 Landes Bioscience.
4. Chronic Toxicity Summary: Methyl Isocyanate. Determination of Noncancer Chronic Reference Exposure Levels Batch 2B; 2001.
5. S. Sriramachari. The Bhopal gas tragedy: An environmental disaster. *Current Science*, 2004; Vol. 86, No. 7
6. Pradyumna K. Mishra, Gorantla V. Raghuram, Hariom Panwar, Deepika Jain, Hemant Pandey and Kewal K. Maudar. Mitochondrial oxidative stress elicits chromosomal instability after exposure to isocyanates in human kidney epithelial cells. *Free Radical Research*, August 2009; 43(8): 718-728.
7. Patrick Che`ne. Inhibition of the p53-MDM2 Interaction: Targeting a Protein-Protein Interface. *Molecular Cancer Research*. 2004;Vol. 2, 20–28,
8. C. M. Iqbal Ahmed, Marjorie A. Burkhart, Mustafa G. Mujtaba, Prem S. Subramaniam and Howard M. Johnson. The role of IFN γ nuclear localization sequence in intracellular function. *Journal of Cell Science* 116, 2003; 3089-3098..
9. Masato Mitsuhashi, David Peel, Argyrios Ziogas and Hoda Anton-Culver. Enhanced Expression of Radiation-induced Leukocyte CDKN1A mRNA in Multiple Primary Breast Cancer Patients: Potential New Marker of Cancer Susceptibility. *Biomarker Insights* 2009;4 201–209.
10. Diane F. Jelinek, Judy B. Splawski and Peter E. Lipsky. The roles of interleukin 2 and interferon- γ in human B cell activation, growth and differentiation. *Eur. J. Immunol.* 1986;16: 925-932.
11. Kendall A. Smith. Interleukin-2: Inception, Impact, and Implications. *Science*, Vol. 240; 1988.
12. Yvonne L Woods and David P Lane. Exploiting the p53 pathway for cancer diagnosis and therapy. *The Hematology Journal* 2003; 4, 233–247.
13. T. Soussi a, G. Lozano. p53 mutation heterogeneity in cancer. *Biochemical and Biophysical Research Communications* 331, 2005; 834–842.
14. Yunje Cho, Svetlana Gorina, Philip D. Jeffrey, Nikola P. Pavietich. Crystal Structure of a p53 Tumor Suppressor-DNA Complex: Understanding Tumorigenic Mutations. *Science*, 1994; Vol. 265.
15. Garrett M. Morris and Marguerita Lim-Wilby. Chapter 19 Molecular Docking. *Methods in Molecular Biology*, vol. 443, *Molecular Modeling of Proteins* Edited by Andreas Kukol. Humana Press, Totowa, NJ.
16. Se´rgio Filipe Sousa, Pedro Alexandrino Fernandes, and Maria Joao Ramos. Protein–Ligand Docking: Current Status and Future Challenges. *PROTEINS: Structure, Function, and Bioinformatics* 65:15–26 (2006).
17. Schrödinger , protein and Ligand Preparation. Chapter 3: Glide 5.6 User Manual. *Glide User Manual* Copyright, 2010, LLC.