

**Practical number: 1****Practical title: Chemical File Formats and Properties Identification of Compounds****Introduction:**

The first version of DrugBank was released in 2006. This early release contained relatively modest information about 841 FDA-approved small molecule drugs and 113 biotech drugs. The DrugBank database is a comprehensive, freely accessible, online database containing information on drugs and drug targets created and maintained by the University of Alberta and The Metabolomics Innovation Centre located in Alberta, Canada.

DrugBank is a unique bioinformatics/cheminformatics resource that combines detailed drug (i.e. chemical) data with comprehensive drug target (i.e. protein) information. The database contains >4100 drug entries including >800 FDA approved small molecule and biotech drugs as well as >3200 experimental drugs. Additionally, >14 000 protein or drug target sequences are linked to these drug entries.

**Morphine:**

**Morphine** is an opioid agonist used for the relief of moderate to severe acute and chronic pain. Morphine is a pain medication of the opiate family that is found naturally in a dark brown, resinous form, from the poppy plant (*Papaver somniferum*). It can be taken orally or injected; it is also often smoked. It acts directly on the central nervous system (CNS) to increase feelings of pleasure and warm relaxation and to reduce pain, and is often abused for this purpose.

TARGET	ACTIONS	ORGANISM
<u>AMu-type opioid receptor</u>	agonist regulator	Humans
<u>AKappa-type opioid receptor</u>	agonist	Humans
<u>ADelta-type opioid receptor</u>	agonist	Humans
<u>ULymphocyte antigen 96</u>	activator	Humans

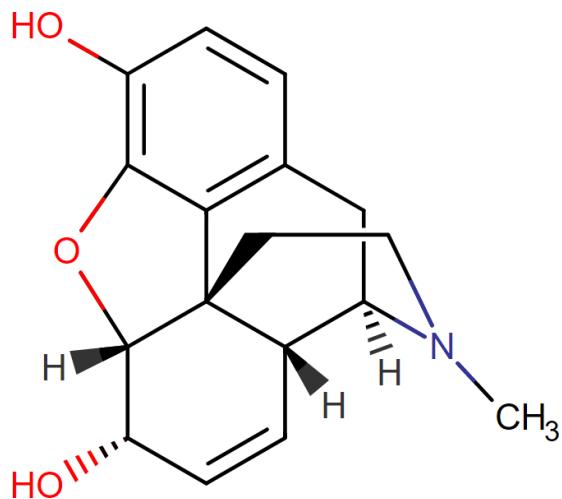
**Drug bank accession number:**  
DB00295

**Weight:**

Average: 285.3377

Monoisotopic: 285.136493479

**Structure:**



**Chemical Formula C17H19NO3**

**Pharmacodynamics**

Morphine binding to opioid receptors blocks transmission of nociceptive signals, signals pain-modulating neurons in the spinal cord, and inhibits primary afferent nociceptors to the dorsal horn sensory projection cells.

Morphine has a time to onset of 6-30 minutes. Excess consumption of morphine and other opioids can lead to changes in synaptic neuroplasticity, including changes in neuron density, changes at postsynaptic sites, and changes at dendritic terminals.

Intravenous morphine's analgesic effect is sex dependent. The EC<sub>50</sub> in men is 76ng/mL and in women is 22ng/mL.

Morphine-6-glucuronide is 22 times less potent than morphine in eliciting pupil constriction.

**Mechanism of action**

Morphine-6-glucuronide is responsible for approximately 85% of the response observed by morphine administration. Morphine and its metabolites act as agonists of the mu and kappa opioid receptors.

The mu-opioid receptor is integral to morphine's effects on the ventral tegmental area of the brain. Morphine's activation of the reward pathway is mediated by agonism of the delta-opioid receptor in the nucleus accumbens, while modification of the respiratory system and addiction disorder are mediated by agonism of the mu-opioid receptor.

**Targets:****1. Mu-type opioid receptor:**

The screenshot shows the DrugBank Online interface with the URL <https://go.drugbank.com/drugs/DB00295>. The main content area displays the "1. Mu-type opioid receptor" profile. Key details include:

- Kind:** Protein
- Organism:** Humans
- Pharmacological action:** Yes
- Actions:** Agonist, Regulator
- General Function:** Voltage-gated calcium channel activity
- Specific Function:** Receptor for endogenous opioids such as beta-endorphin and endomorphin. Receptor for natural and synthetic opioids including morphine, heroin, DAMGO, fentanyl, etorphine, buprenorphin and methadone...
- Gene Name:** OPRM1
- Uniprot ID:** P35372
- Uniprot Name:** Mu-type opioid receptor
- Molecular Weight:** 44778.855 Da

The left sidebar lists various pharmacological categories, with "Targets (4)" highlighted. The bottom status bar shows the weather as 26°C Cloudy, system icons, and the date/time as 14-07-2022 00:19.

**2. Kappa-type opioid receptor**

The screenshot shows the DrugBank Online interface with the URL <https://go.drugbank.com/drugs/DB00295>. The main content area displays the "2. Kappa-type opioid receptor" profile. Key details include:

- Kind:** Protein
- Organism:** Humans
- Pharmacological action:** Yes
- Actions:** Agonist
- General Function:** Opioid receptor activity
- Specific Function:** G-protein coupled opioid receptor that functions as receptor for endogenous alpha-neoendorphins and dynorphins, but has low affinity for beta-endorphins. Also functions as receptor for various synt...
- Gene Name:** OPKR1
- Uniprot ID:** P41145
- Uniprot Name:** Kappa-type opioid receptor
- Molecular Weight:** 42644.665 Da

The left sidebar lists various pharmacological categories, with "Targets (4)" highlighted. The bottom status bar shows the weather as 26°C Cloudy, system icons, and the date/time as 14-07-2022 00:19.

## 3. Delta-type opioid receptor

The screenshot shows the DrugBank Online interface with the URL <https://go.drugbank.com/drugs/DB00295>. The main content area displays the "3. Delta-type opioid receptor" profile. The left sidebar includes links for Pharmacology, Interactions, Products, Categories, Chemical Identifiers, References, Clinical Trials, Pharmacoconomics, Properties, Spectra, Targets (4), Enzymes (8), Carriers (1), and Transporters (1). The right sidebar features a vertical toolbar with icons for various functions. The central panel contains detailed information about the receptor, including its Kind (Protein), General Function (Opioid receptor activity), Organism (Humans), Specific Function (G-protein coupled receptor that functions as receptor for endogenous enkephalins and for a subset of other opioids. Ligand binding causes a conformation change that triggers signaling via guanine n...), Pharmacological action (Yes), Actions (Agonist), Gene Name (OPRD1), Uniprot ID (P41143), Uniprot Name (Delta-type opioid receptor), and Molecular Weight (40368.235 Da). Below this is a "References" section listing a single study by Yamada et al. (2001).

## 4. Lymphocyte antigen 96

The screenshot shows the DrugBank Online interface with the URL <https://go.drugbank.com/drugs/DB00295>. The main content area displays the "4. Lymphocyte antigen 96" profile. The left sidebar includes links for Pharmacology, Interactions, Products, Categories, Chemical Identifiers, References, Clinical Trials, Pharmacoconomics, Properties, Spectra, Targets (4), Enzymes (8), Carriers (1), and Transporters (1). The right sidebar features a vertical toolbar with icons for various functions. The central panel contains detailed information about the antigen, including its Kind (Protein), General Function (Lipopolysaccharide receptor activity), Organism (Humans), Specific Function (Binds bacterial lipopolysaccharide (LPS) (PubMed:17803912, PubMed:17569869). Cooperates with TLR4 in the innate immune response to bacterial lipopolysaccharide (LPS), and with TLR2 in the response ...), Pharmacological action (Unknown), Actions (Activator), Gene Name (LY96), Uniprot ID (Q9Y6Y9), Uniprot Name (Lymphocyte antigen 96), and Molecular Weight (18545.345 Da). Below this is a "References" section listing a single study by Hutchinson et al. (2001).

**Toxicity:**

The LD<sub>50</sub> is 0.78µg/mL in males and 0.98µg/mL in females.

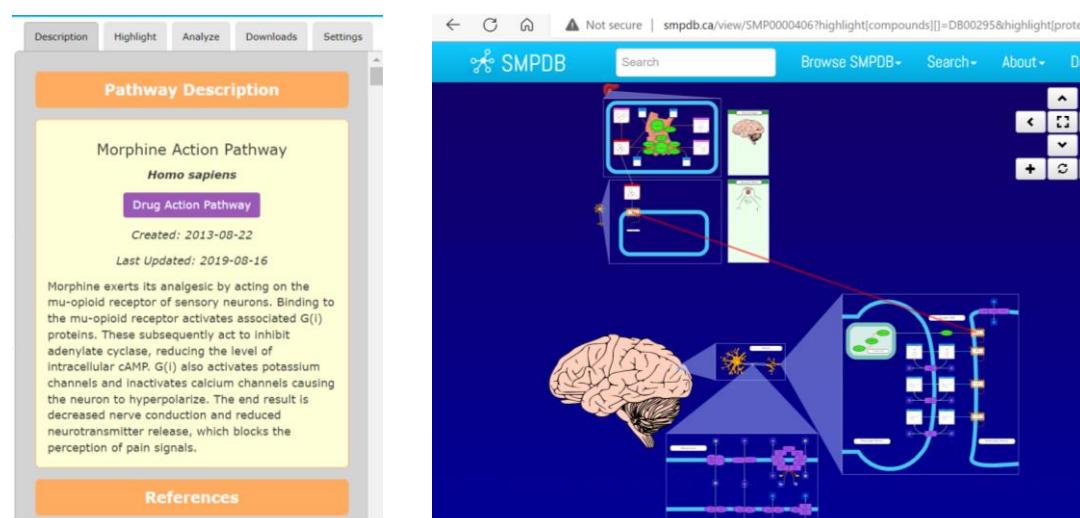
Patients experiencing an overdose present with respiratory depression, somnolence, skeletal muscle flaccidity, cold and clammy skin, miosis, and mydriasis. Symptoms of overdose can progress to pulmonary edema, bradycardia, hypotension, cardiac arrest, and death. Treat overdose with symptomatic and supportive treatment which may include the use of oxygen, vasopressors, and naloxone.

### Pathways:

PATHWAY	CATEGORY
Morphine Action Pathway	Drug action
Codeine Action Pathway	Drug action
Heroin Metabolism Pathway	Drug metabolism
Heroin Action Pathway	Drug action
Codeine Metabolism Pathway	Drug metabolism
Morphine Metabolism Pathway	Drug metabolism

### Morphine Action Pathway:

Morphine exerts its analgesic by acting on the mu-opioid receptor of sensory neurons. Binding to the mu-opioid receptor activates associated G(i) proteins.



### Properties:

**State:** Solid

**Experimental Properties**

LogP = logarithm partition coefficient

**Significance:**

- LogP indicates hydrophobicity of a molecule.
- Increase in value = Increase in logP
- Decrease in value = Decrease in logP
- 0.87 is less hydrophobic.
- -2 to +5 is hydrophobic range
- pKa is experimental value in different ph.

**Experimental Properties**

PROPERTY	VALUE	SOURCE
melting point (°C)	255 °C	MSDS
boiling point (°C)	190°C	Lide DR. CRC Handbook of Chemistry. 2004
water solubility	149 mg/L (at 20 °C)	MSDS
logP	0.87	Agilent Technology. LC/MS/MS Analysis of Opiates.
pKa	8.21 (at 25 °C)	Agilent Technology. LC/MS/MS Analysis of Opiates.

**Predicted Properties**

- logS = logarithm of Solvent
- Rule of 5 :- Lipinsky's rule of 5
- LogP < 5
- Molecule is not < 500 Da
- H bond / Acceptor < 10
- H bond/ Donor < 5

**Predicted Properties**

PROPERTY	VALUE	SOURCE
Water Solubility	10.2 mg/mL	ALOGPS
logP	0.99	ALOGPS
logP	0.9	ChemAxon
logS	-1.4	ALOGPS
pKa (Strongest Acidic)	10.26	ChemAxon
pKa (Strongest Basic)	9.12	ChemAxon
Physiological Charge	1	ChemAxon
Hydrogen Acceptor Count	4	ChemAxon
Hydrogen Donor Count	2	ChemAxon
Polar Surface Area	52.93 Å²	ChemAxon
Rotatable Bond Count	0	ChemAxon

Refractivity	80.12 m <sup>3</sup> .mol <sup>-1</sup>	ChemAxon
Polarizability	29.94 Å <sup>3</sup>	ChemAxon
Number of Rings	5	ChemAxon
Bioavailability	1	ChemAxon
Rule of Five	Yes	ChemAxon
Ghose Filter	Yes	ChemAxon
Veber's Rule	No	ChemAxon
MDDR-like Rule	No	ChemAxon

### Predicted ADMET Features

PROPERTY	VALUE	PROBABILITY
Human Intestinal Absorption	+	0.9971
Blood Brain Barrier	+	0.9882
Caco-2 permeable	+	0.8867
P-glycoprotein substrate	Substrate	0.8787
P-glycoprotein inhibitor I	Non-inhibitor	0.8782
P-glycoprotein inhibitor II	Non-inhibitor	0.956
Renal organic cation transporter	Inhibitor	0.6221
CYP450 2C9 substrate	Non-substrate	0.7451
CYP450 2D6 substrate	Substrate	0.8919
CYP450 3A4 substrate	Substrate	0.7375
CYP450 1A2 substrate	Non-inhibitor	0.5191
CYP450 2C9 inhibitor	Non-inhibitor	0.9046
CYP450 2D6 inhibitor	Non-inhibitor	0.647
CYP450 2C19 inhibitor	Non-inhibitor	0.8155
CYP450 3A4 inhibitor	Non-inhibitor	0.9176
CYP450 inhibitory promiscuity	Low CYP Inhibitory Promiscuity	0.7503
Ames test	Non AMES toxic	0.9132
Carcinogenicity	Non-carcinogens	0.9634
Biodegradation	Not ready biodegradable	0.9944
Rat acute toxicity	2.8989 LD50, mol/kg	Not applicable
hERG inhibition (predictor I)	Weak inhibitor	0.8367
hERG inhibition (predictor II)	Non-inhibitor	0.874

### Significant properties of Predicted ADMET Features:

Human Intestinal Absorption :

Human intestinal absorption (HIA) is one of the most important ADME properties. Utilization of drugs in the human body is such a complicated process that it can hardly be analyzed precisely by statistical models. HIA is also one of the key steps during the drugs' transporting to their targets.

**Blood-brain barrier :**

The blood vessels that vascularize the central nervous system (CNS) possess unique properties, termed the blood-brain barrier, which allow these vessels to tightly regulate the movement of ions, molecules, and cells between the blood and the brain.

**Caco-2 permeable:**

Cyprotex's Caco-2 permeability assay uses an established method that measures the rate of flux of a compound across polarised Caco-2 cell monolayers and from which the data generated can be used to predict in vivo absorption of drugs. The Caco-2 cell line is derived from a human colon carcinoma.

**P-glycoprotein substrate:**

Important substrates of P-glycoprotein include calcium channel blockers, cyclosporin, dabigatran etexilate, digoxin, erythromycin, loperamide, protease inhibitors and tacrolimus. Predicting clinically important interactions is difficult because of interindividual differences in drug transport

**P-glycoprotein inhibitor I:**

P-glycoprotein is an important mediator of drug-drug interactions. The pharmacokinetics of a drug may be altered when co-administered with compounds which inhibit or induce P-glycoprotein. Inhibitors include clarithromycin, erythromycin, ritonavir and verapamil. Inducers include rifampicin and St John's wort.

**Q1- Identify any 1 drug molecule structure and list out the various file formats and give the explanation?**

**Result:**

**Morphine:**

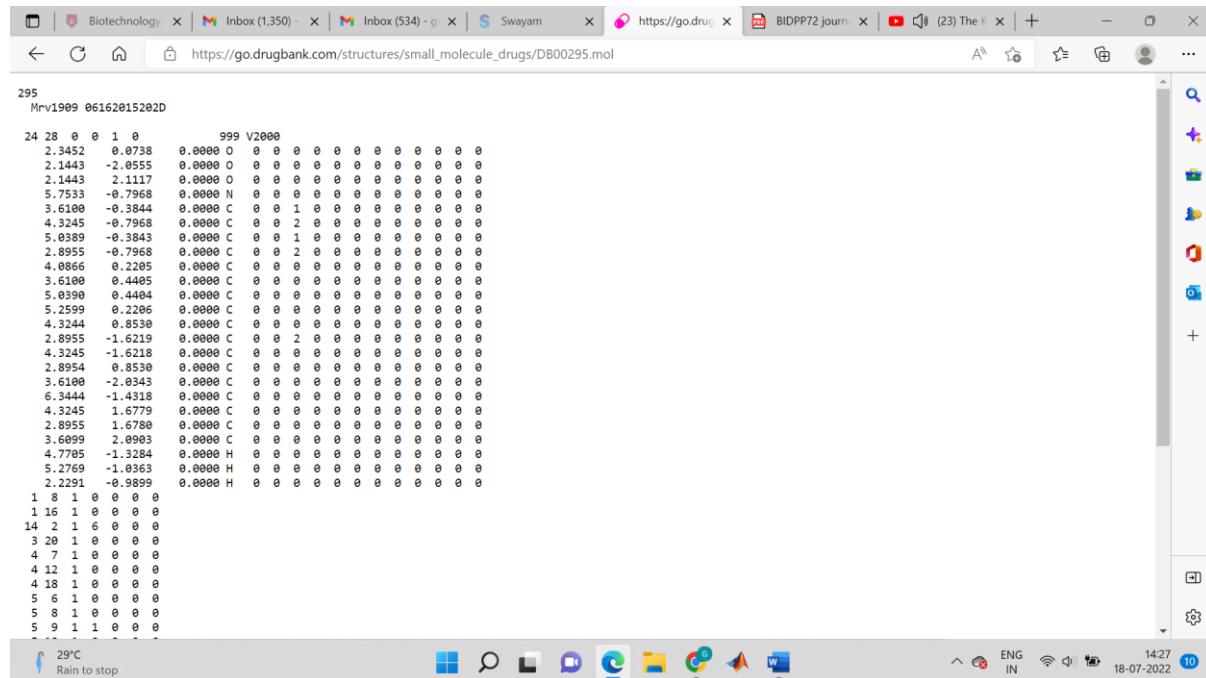
So for Morphine the files can be downloaded in PDB, MOL, SDF, 3D-SDF, SMILES, InChI format.

**File formats:**

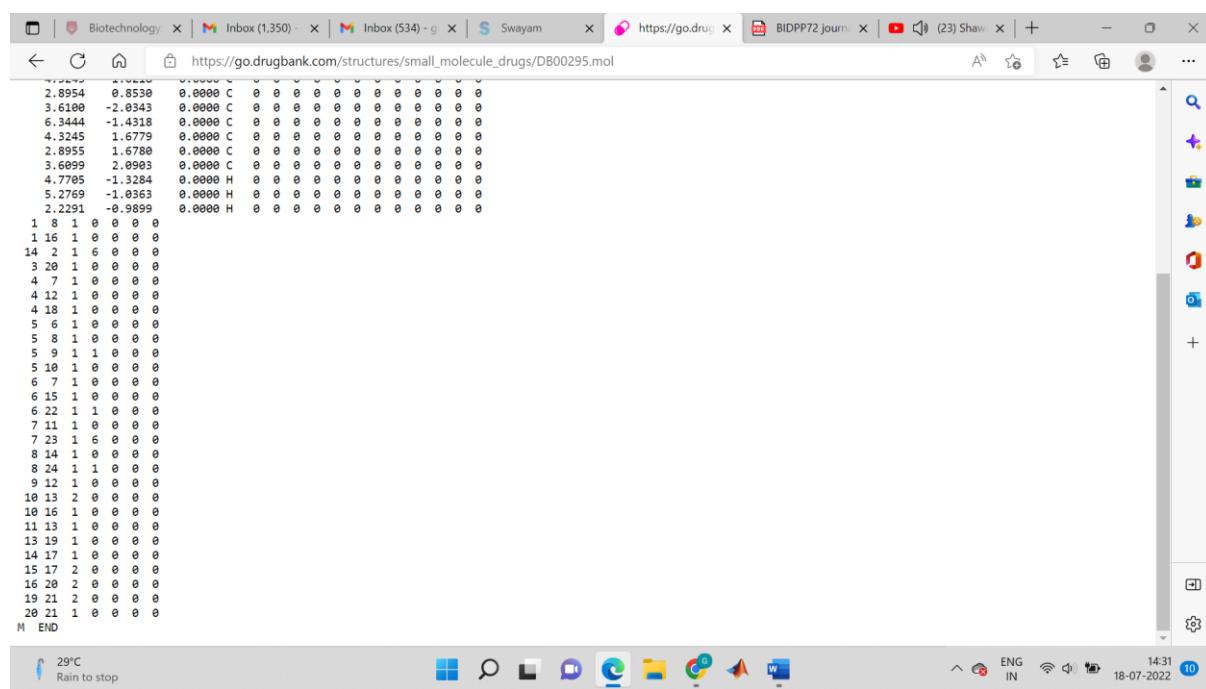
1) .mol file format:

- Only one file can be saved in .mol file format.
- Three coordinates X, Y, Z are shown in the respective columns along with the number of atoms column.
- 295  
Mrv1909 06162015202D
- Connection table is also shown (1 8 1 are co atoms).
- The default number of mol file format will be the last number as 999.
- Termination of .mol file is shown as "M END" at the end.

BID 19006



Termination of .mol file is shown as "M END" at the end. Connection table is also shown (1 8 1 are co atoms).

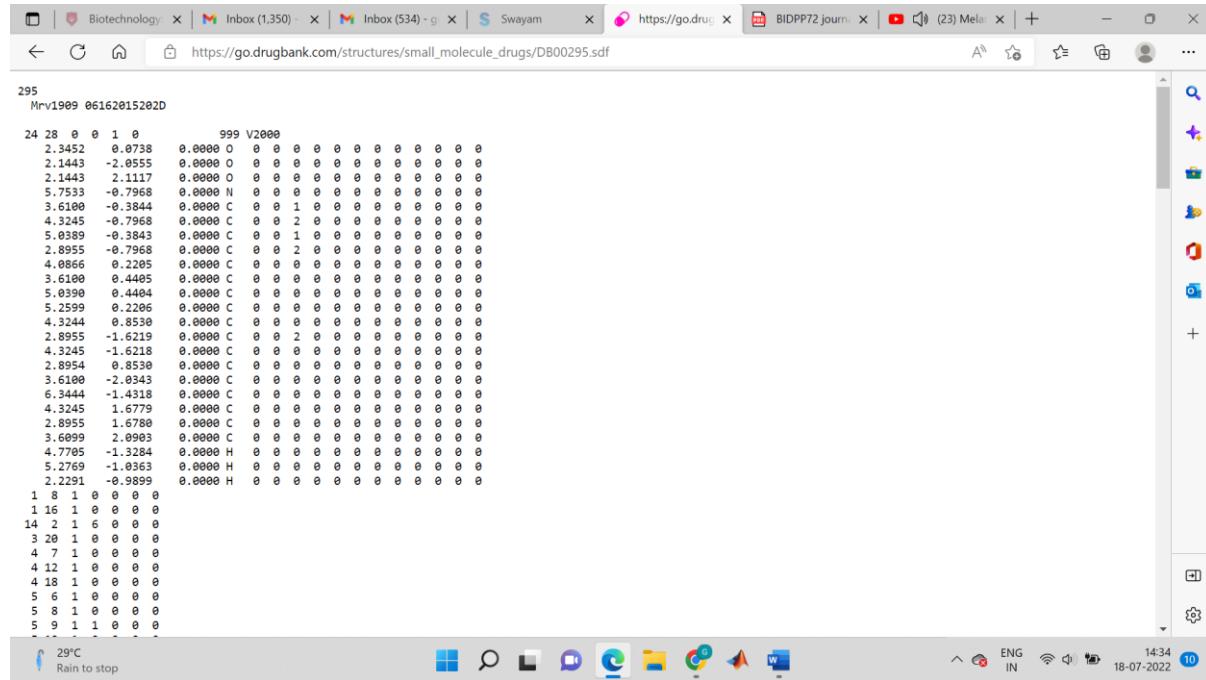


## 2) .SDF file format:

- Structure Data File.
  - Three coordinates X, Y, Z are shown in the respective columns along with the number of atoms column.
  - Connection table is also shown (1 8 1 are co atoms).
  - Termination of .sdf file is shown as “\$\$\$\$” at the end.

BID 19006

- .sdf file format can save multiple files at a time.



**SDF file format saves complete set of properties including :**

Database\_id  
Db00295

Database\_name  
Drugbank

**Smiles**  
[h][c@@]12oc3=c(o)c=cc4=c3[c@@]11ccn(c)[c@]([h])(c4)[c@]1([h])c=c[c@@h]2o

Inchi\_identifier  
Inchi=1s/c17h19no3/c1-18-7-6-17-10-3-5-13(20)16(17)21-15-12(19)4-2-9(14(15)17)8-11(10)18/h2-5,10-11,13,16,19-20h,6-8h2,1h3/t10-,11+,13-,16-,17-/m0/s1

Inchi\_key  
Bqjcrhhnabkaku-kbqpjgbksa-n

## Formula C<sub>17</sub>H<sub>19</sub>NO<sub>3</sub>

## Molecular weight

BID 19006

285.3377

Exact\_mass

285.136493479

The screenshot shows a Microsoft Edge browser window with multiple tabs open. The active tab displays the SDF (Structure Data File) content for compound DB00295. The SDF file starts with a connection table (CARTOON) and includes fields for database ID, name, SMILES, InChI, InChI key, formula, molecular weight, exact mass, and a termination marker (\$\$\$\$). The browser interface includes a search bar, address bar, and various toolbars.

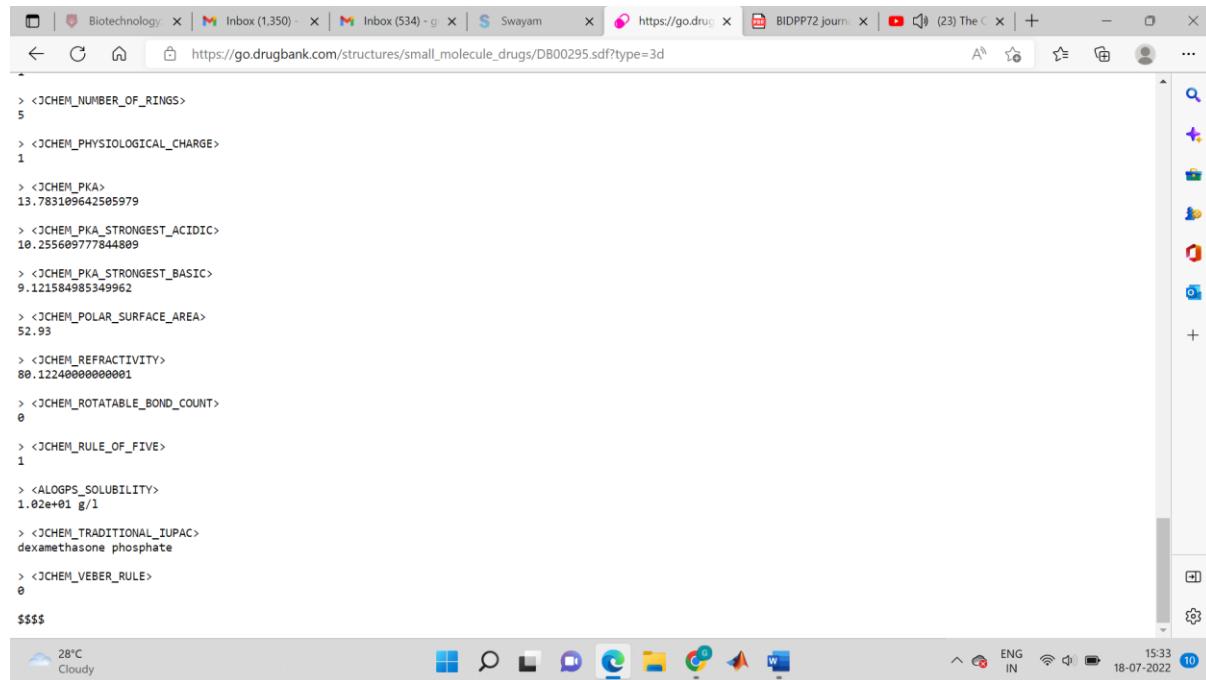
```
6 15 1 0 0 0 0
6 22 1 1 0 0 0
7 11 1 0 0 0 0
7 23 1 6 0 0 0
8 14 1 0 0 0 0
8 24 1 1 0 0 0
9 12 1 0 0 0 0
10 13 2 0 0 0 0
10 16 1 0 0 0 0
11 13 1 0 0 0 0
13 19 1 0 0 0 0
14 17 1 0 0 0 0
15 17 2 0 0 0 0
16 20 2 0 0 0 0
19 21 2 0 0 0 0
20 21 1 0 0 0 0
M END
> <DATABASE_ID>
DB00295
> <DATABASE_NAME>
drugbank
> <SMILES>
[H][C@H]12OC3=C(O)C=CC4=C3[C@H]11CCN(C)[C@H](C4)[C@H]1([H])C=C[C@H]2O
> <INCHI_IDENTIFIER>
InChI=1S/C17H19NO3/c1-18-7-6-17-10-3-5-13(20)16(17)21-15-12(19)4-2-9(14(15)17)8-11(10)18/h2-5,10-11,13,16,19-20H,6-8H2,1H3/t10-,11+,13-,16+,17-/m0/s1
> <INCHI_KEY>
BQJCRHHNABKAKU-KBQPJGBKSA-N
> <FORMULA>
C17H19NO3
> <MOLECULAR_WEIGHT>
285.3377
> <EXACT_MASS>
285.136493479
> <COORDINATES>
29°C
Rain to stop
```

3) 3D – SDF file format:

- Structure Data File.
- We have hydrogens added.
- Three coordinates X, Y, Z are shown in the respective columns along with the number of atoms column.
- 528826  
-OEChem-06162011203D
- Connection table is also shown (1 8 1 are co atoms).
- Termination of 3D .sdf file is shown as “\$\$\$\$” at the end.

BID 19006

Termination of 3D .sdf file is shown as “\$\$\$\$” at the end.



```
> <JCHEM_NUMBER_OF_RINGS>
5
> <JCHEM_PHYSIOLOGICAL_CHARGE>
1
> <JCHEM_PKA>
13.783109642505979
> <JCHEM_PKA_STRONGEST_ACIDIC>
18.255609777844809
> <JCHEM_PKA_STRONGEST_BASIC>
9.121584985349962
> <JCHEM_POLAR_SURFACE_AREA>
52.93
> <JCHEM_REFRACTIVITY>
80.12248000000001
> <JCHEM_ROTATABLE_BOND_COUNT>
0
> <JCHEM_RULE_OF_FIVE>
1
> <ALOGPS_SOLUBILITY>
1.02e+01 g/l
> <JCHEM_TRADITIONAL_IUPAC>
dexamethasone phosphate
> <JCHEM_VEBER_RULE>
0
$$$$
```

**3D SDF file format saves complete set of properties including :**

> <DATABASE\_ID>  
DB00295

> <DATABASE\_NAME>  
drugbank

> <ORIGINAL\_SOURCE>  
PUBCHEM

> <ORIGINAL\_SOURCE\_URL>  
[https://pubchem.ncbi.nlm.nih.gov/rest/pug/compound/inchikey/BQJCRHHNABKAKU-KBQPJGBKSA-N/SDF?record\\_type=3d](https://pubchem.ncbi.nlm.nih.gov/rest/pug/compound/inchikey/BQJCRHHNABKAKU-KBQPJGBKSA-N/SDF?record_type=3d)

> <SMILES>  
[H][C@@]12OC3=C(O)C=CC4=C3[C@@]11CCN(C)[C@]([H])(C4)[C@]1([H])C=C[C@@H]2O

> <INCHI\_IDENTIFIER>  
InChI=1S/C17H19NO3/c1-18-7-6-17-10-3-5-13(20)16(17)21-15-12(19)4-2-9(14(15)17)8-11(10)18/h2-5,10-11,13,16,19-20H,6-8H2,1H3/t10-,11+,13-,16-,17-/m0/s1

> <INCHI\_KEY>  
BQJCRHHNABKAKU-KBQPJGBKSA-N

> <FORMULA>

C17H19NO3

> <MOLECULAR\_WEIGHT>

285.3377

> <EXACT\_MASS>

285.136493479

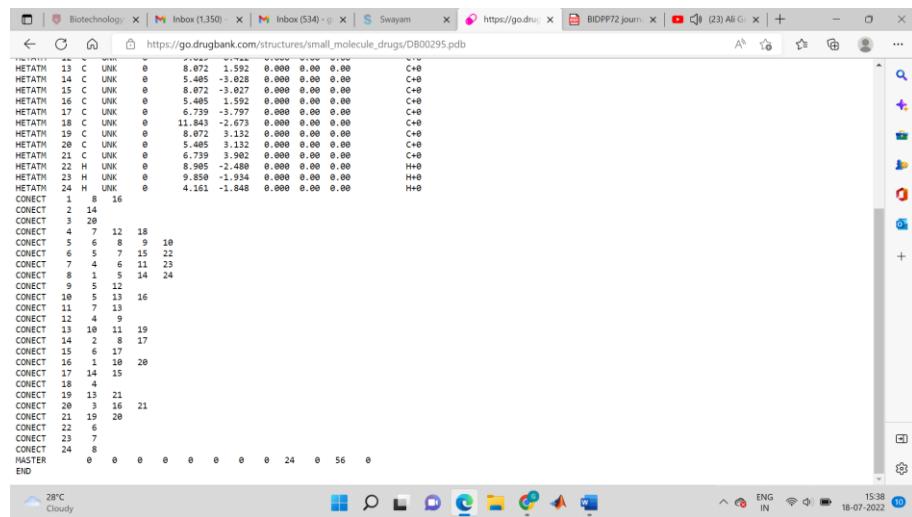
4) PDB file format:

- Protein Data Bank
- U/K. UNK = unknown
- MASTER = Total number of elements
- Total number of atoms = 24
- Total number of connections = 56
- Termination of PDB file = END

```
HEADER PROTEIN          16-JUN-20  NONE
TITLE    NULL
COMPND   MOLECULE: 295
SOURCE   NULL
KEYWDS  NULL
EXPDTA  NULL
AUTHOR  Marvin
REVDAT  1 16-JUN-20      0
HETATOM 1 O  UNK  0   4.378  0.138  0.000  0.00  0.00      O+0
HETATOM 2 O  UNK  0   4.003 -3.837  0.000  0.00  0.00      O+0
HETATOM 3 O  UNK  0   4.003  3.942  0.000  0.00  0.00      O+0
HETATOM 4 N  UNK  0   10.740 -1.487  0.000  0.00  0.00      N+0
HETATOM 5 C  UNK  0   6.739 -0.718  0.000  0.00  0.00      C+0
HETATOM 6 C  UNK  0   8.872 -1.487  0.000  0.00  0.00      C+0
HETATOM 7 C  UNK  0   9.406 -0.717  0.000  0.00  0.00      C+0
HETATOM 8 C  UNK  0   5.405 -1.487  0.000  0.00  0.00      C+0
HETATOM 9 C  UNK  0   7.628  0.412  0.000  0.00  0.00      C+0
HETATOM 10 C UNK  0   6.739  0.822  0.000  0.00  0.00      C+0
HETATOM 11 C UNK  0   9.406  0.822  0.000  0.00  0.00      C+0
HETATOM 12 C UNK  0   9.819  0.412  0.000  0.00  0.00      C+0
HETATOM 13 C UNK  0   8.872  1.592  0.000  0.00  0.00      C+0
HETATOM 14 C UNK  0   5.405 -3.028  0.000  0.00  0.00      C+0
HETATOM 15 C UNK  0   8.872 -3.027  0.000  0.00  0.00      C+0
HETATOM 16 C UNK  0   5.405  1.592  0.000  0.00  0.00      C+0
HETATOM 17 C UNK  0   6.739 -3.797  0.000  0.00  0.00      C+0
HETATOM 18 C UNK  0   11.843 -2.673  0.000  0.00  0.00      C+0
HETATOM 19 C UNK  0   8.872  3.152  0.000  0.00  0.00      C+0
HETATOM 20 C UNK  0   5.405  3.152  0.000  0.00  0.00      C+0
HETATOM 21 C UNK  0   6.739  3.962  0.000  0.00  0.00      C+0
HETATOM 22 H UNK  0   8.905 -2.486  0.000  0.00  0.00      H+0
HETATOM 23 H UNK  0   9.850 -1.954  0.000  0.00  0.00      H+0
HETATOM 24 H UNK  0   4.161 -1.848  0.000  0.00  0.00      H+0
CONECT  1   8   16
CONECT  2   14
CONECT  3   20
CONECT  4   7   12   18
CONECT  5   6   8   9   10
CONECT  6   5   7   15   22
```

Total number of atoms = 24. Total number of connections = 56. Termination of PDB file = END

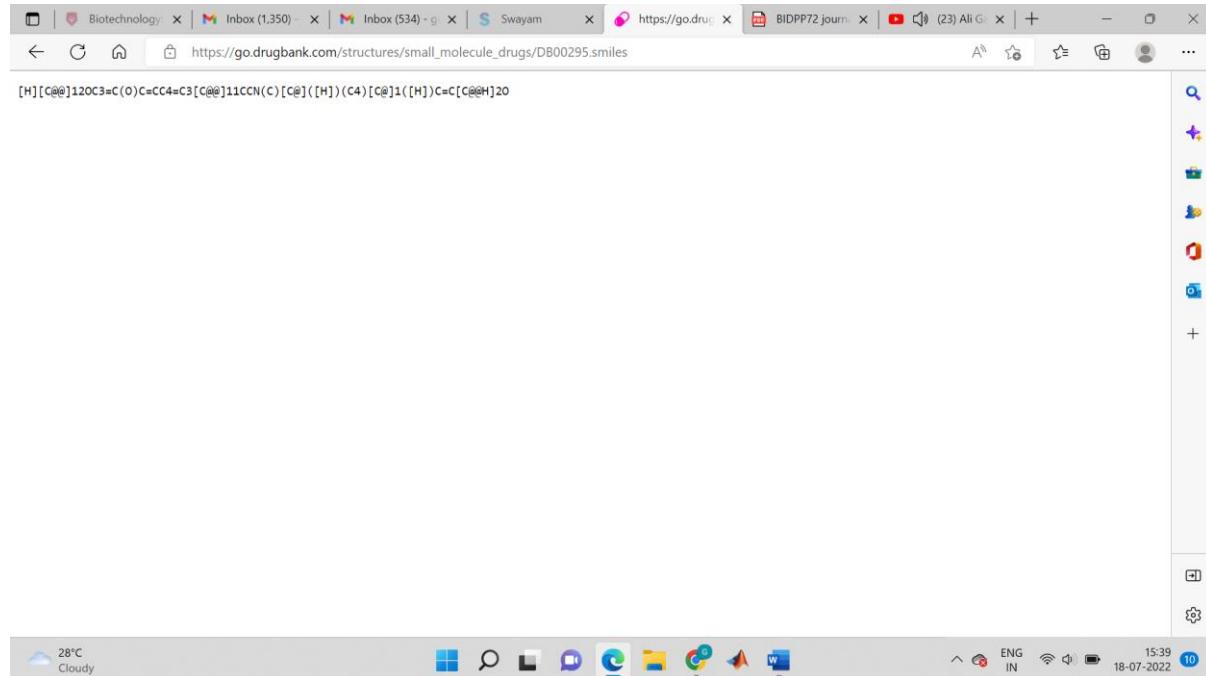
## BID 19006



```
https://go.drugbank.com/structures/small_molecule_drugs/DB00295.pdb
HETATM 1 C UNK 0 6.672 -1.392 0.000 0.00 0.00 C+0
HETATM 14 C UNK 0 5.405 -1.028 0.000 0.00 0.00 C+0
HETATM 15 C UNK 0 8.672 -3.027 0.000 0.00 0.00 C+0
HETATM 16 C UNK 0 5.405 1.592 0.000 0.00 0.00 C+0
HETATM 17 C UNK 0 5.700 -3.759 0.000 0.00 0.00 C+0
HETATM 18 C UNK 0 18.443 -1.773 0.000 0.00 0.00 C+0
HETATM 19 C UNK 0 8.672 3.132 0.000 0.00 0.00 C+0
HETATM 20 C UNK 0 5.405 3.132 0.000 0.00 0.00 C+0
HETATM 21 C UNK 0 6.739 3.960 0.000 0.00 0.00 C+0
HETATM 22 C UNK 0 6.739 3.960 0.000 0.00 0.00 C+0
HETATM 23 H UNK 0 9.850 -1.934 0.000 0.00 0.00 H+0
HETATM 24 H UNK 0 4.151 -1.848 0.000 0.00 0.00 H+0
CONECT 1 8 16
CONECT 2 14
CONECT 20
CONECT 4 7 12 18
CONECT 5 6 8 9 18
CONECT 6 5 7 15 22
CONECT 7 6 11 13
CONECT 8 1 5 14 24
CONECT 9 5 12
CONECT 10 5 13 16
CONECT 11 7 13 15
CONECT 12 4 9
CONECT 13 10 11 19
CONECT 14 2 8 17
CONECT 15 6 17
CONECT 16 18 20
CONECT 17 14 15
CONECT 18 4
CONECT 19 13 21
CONECT 20 9 16 21
CONECT 21 19 28
CONECT 22 6
CONECT 23 7
CONECT 24 8
MASTER
0 0 0 0 0 0 0 0 24 0 56 0
END
```

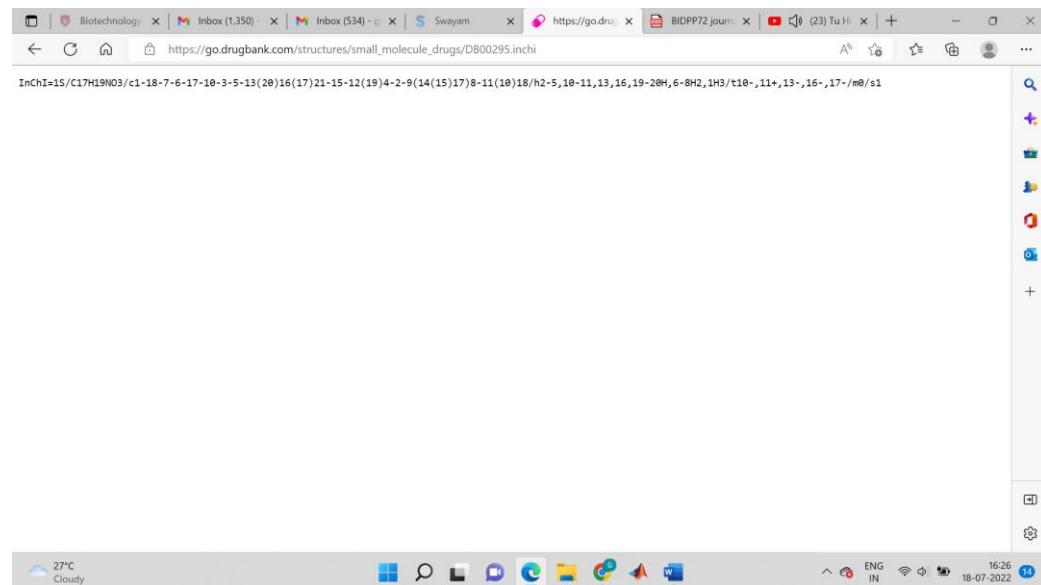
## 5) SMILES file format:

- Linear representation of a compound.
- Through SMILES we can get a PDB structure.
- Basic hydrogen carbons using some rules are shown.
- Oxygen = @.
- Branch / Side chain = ()
- Double bond = = .
- [H][C@@@]12OC3=C(O)C=CC4=C3[C@@@]11CCN(C)[C@](H)(C4)[C@]1([H])C=C[C@H]2O



## 6) InChI file format:

- International Chemical Identifier
- Used for medical chemical structure representation.
- Mixture of alphabets / atomic numbers / elements.
- InChI=1S/C17H19NO3/c1-18-7-6-17-10-3-5-13(20)16(17)21-15-12(19)4-2-9(14(15)17)8-11(10)18/h2-5,10-11,13,16,19-20H,6-8H2,1H3/t10-,11+,13-,16-,17-/m0/s1



### Similar structures:

There are 42 similar structures of morphine

The screenshot shows the DRUGBANK Online search interface. The search bar contains the URL https://go.drugbank.com/structures/search/small\_molecule\_drugs/structure?database\_id=DB00295&search\_type=similarity#results. On the left, there is a chemical structure of morphine labeled "Absolute". To the right, there are search parameters: "Similarity threshold" set to 0.7, "Minimum Weight" set to e.g. 100, "Maximum Weight" set to e.g. 200, and "Maximum Results" set to 100. Below these are filter options for "Drug Types (default all)": Approved, Vet approved, Nutraceutical, Illicit, Withdrawn, Investigational, and Experimental. A "Search" button is at the bottom. The interface has a pink header with the DRUGBANK logo and a navigation menu. The taskbar at the bottom shows various application icons.

C<sub>17</sub>H<sub>19</sub>NO<sub>3</sub>  
Mono mass: 285.136493479

# BID 19006

C<sub>18</sub>H<sub>21</sub>NO<sub>3</sub>

Mono mass: 299.152143543

C<sub>19</sub>H<sub>23</sub>NO<sub>3</sub>

Mono mass: 313.167793607

DRUGBANK Online

Results 1 – 30 of approximately 42 results

Score	Chemical Structure	Drug Name	Chemical Formula	Mono mass
1.0		Morphine 57-27-3	C <sub>18</sub> H <sub>21</sub> NO <sub>3</sub>	285.136495479
0.99		Codeine 76-57-3	C <sub>19</sub> H <sub>23</sub> NO <sub>3</sub>	299.152143543
0.99		Ethylmorphine 76-58-4	C <sub>19</sub> H <sub>23</sub> NO <sub>3</sub>	313.167793607
0.99		Codeine methylbromide 125-27-9	C <sub>19</sub> H <sub>23</sub> BrNO <sub>3</sub>	393.093956286
0.98		Benzylmorphine 14297-87-1	C <sub>19</sub> H <sub>23</sub> NO <sub>3</sub>	375.183443671
0.98		Nalorphine 62-67-9	C <sub>19</sub> H <sub>23</sub> NO <sub>3</sub>	311.15214354
0.96		Etorphine 14321-96-1	C <sub>20</sub> H <sub>25</sub> NO <sub>4</sub>	411.240958549

Cloudy 27°C ENG IN 16:33 18-07-2022

DRUGBANK Online

Results 1 – 30 of approximately 42 results

Score	Chemical Structure	Drug Name	Chemical Formula	Mono mass
0.97		Diamorphine 561-27-3	C <sub>21</sub> H <sub>23</sub> NO <sub>5</sub>	369.157622851
0.85		Dihydrocodeine 125-28-0	C <sub>19</sub> H <sub>23</sub> NO <sub>3</sub>	301.167793607
0.85		Dihydromorphine 509-60-4	C <sub>19</sub> H <sub>23</sub> NO <sub>3</sub>	287.152143543
0.84		Dihydroetorphine 14357-76-7	C <sub>20</sub> H <sub>25</sub> NO <sub>4</sub>	413.256609813
0.83		Hydromorphone 2183-56-4	C <sub>20</sub> H <sub>25</sub> NO <sub>4</sub>	303.147058165
0.83		ORP-101 1820753-68-1	C <sub>20</sub> H <sub>25</sub> N <sub>2</sub> O <sub>5</sub>	460.622767673
0.83		Buprenorphine 52485-79-7	C <sub>20</sub> H <sub>25</sub> NO <sub>4</sub>	467.303558805
		Diprenorphine 13357-78-0	C <sub>20</sub> H <sub>25</sub> NO <sub>4</sub>	

Cloudy 27°C ENG IN 16:33 18-07-2022

### Conclusion:

DrugBank is a unique bioinformatics/cheminformatics resource that combines detailed drug (i.e. chemical) data with comprehensive drug target (i.e. protein) information. DrugBank was designed to serve as a comprehensive, fully searchable in silico drug resource that linked sequence, structure and mechanistic data about drug molecules (including biotech drugs) with sequence, structure and mechanistic data about their drug targets. The DrugBank database is a comprehensive, freely accessible, online database containing information on drugs and drug targets created and maintained by the University of Alberta and The Metabolomics Innovation Centre located in Alberta, Canada.

### Practical number 2:

#### Practical title: Chemical data retrieval and analysis using PubChem

### Introduction:

#### PubChem Compound:

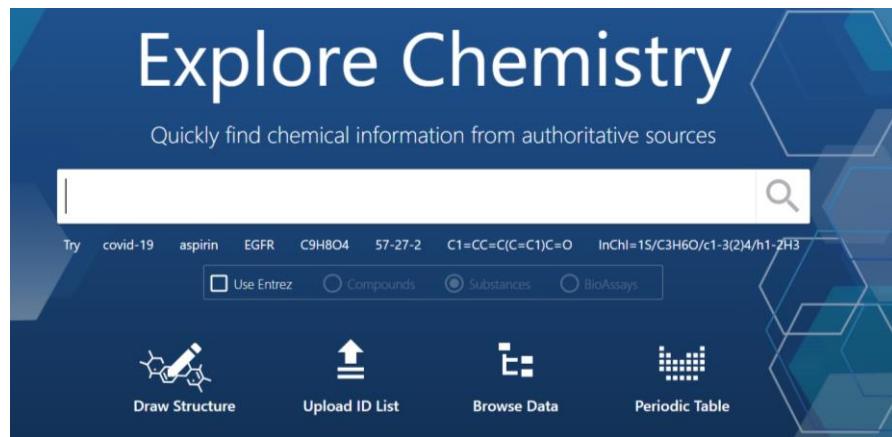
The PubChem Compound Database contains validated chemical depiction information provided to describe substances in PubChem Substance. Structures stored within PubChem Compounds are pre-clustered and cross-referenced by identity and similarity groups.

#### PubChem Substance

The PubChem Substance Database contains descriptions of samples, from a variety of sources, and links to biological screening results that are available in PubChem BioAssay. If the chemical contents of a sample are known, the description includes links to PubChem Compound.

#### PubChem BioAssay:

The PubChem BioAssay Database contains bioactivity screens of chemical substances described in PubChem Substance. It provides searchable descriptions of each bioassay, including descriptions of the conditions and readouts specific to that screening procedure.



PubChem mostly contains small molecules, but also larger molecules such as nucleotides, carbohydrates, lipids, peptides, and chemically-modified macromolecules. We collect information on chemical structures, identifiers, chemical and physical properties, biological activities, patents, health, safety, toxicity data, and many others.

PubChem is the world's largest collection of freely accessible chemical information. Search chemicals by name, molecular formula, structure, and other identifiers. Find chemical and physical properties, biological activities, safety and toxicity information, patents, literature citations and more.

### **Compound: Morphin**

Morphine is one of the natural plant alkaloids found in opium and is the prototype opiate, against which other derivatives are measured in terms of analgesic effects and side effects. Morphine has not been linked to serum enzyme elevations during therapy or to clinically apparent liver injury.

Morphine is a morphinan alkaloid that is a highly potent opiate analgesic psychoactive drug. Morphine acts directly on the central nervous system (CNS) to relieve pain but has a high potential for addiction, with tolerance and both physical and psychological dependence developing rapidly. Morphine is the most abundant opiate found in *Papaver somniferum* (the opium poppy).

It has a role as an opioid analgesic, a mu-opioid receptor agonist, a plant metabolite, an environmental contaminant, a xenobiotic, a vasodilator agent, an anaesthetic, a drug allergen and a geroprotector. It is an organic heteropentacyclic compound, a tertiary amino compound and a morphinan alkaloid. It is a conjugate base of a morphine(1+). It derives from a hydride of a morphinan.

### **IUPAC Name**

(4R,4aR,7S,7aR,12bS)-3-methyl-2,4,4a,7,7a,13-hexahydro-1H-4,12-methanobenzofuro[3,2-e]isoquinoline-7,9-diol

### **InChI**

InChI=1S/C17H19NO3/c1-18-7-6-17-10-3-5-13(20)16(17)21-15-12(19)4-2-9(14(15)17)8-11(10)18/h2-5,10-11,13,16,19-20H,6-8H2,1H3/t10-,11+,13-,16-,17-/m0/s1

### **Canonical SMILES**

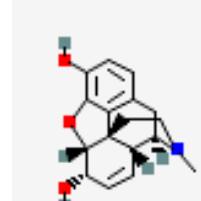
CN1CCC23C4C1CC5=C2C(=C(C=C5)O)OC3C(C=C4)O

### **Molecular Formula**

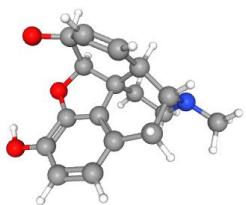
C<sub>17</sub>H<sub>19</sub>NO<sub>3</sub>

## Substance Morphine:

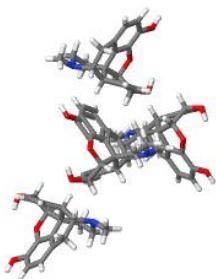
## Structure: 2D



## Interactive Chemical Structure Model:



**Crystal Structure Depiction:**



**BioAssays:**

Antinociceptive activity in sc dosed mouse assessed as inhibition of morphin response by tail flick assay

BioAssay AID: 320416

BioAssay Type: Literature-derived

Tested Compounds Count: 11

Tested Substances Count: 11

Data Source: ChEMBL

External ID: CHEMBL929785

Data Source Category: Curation Efforts; Research and Development

Description: Title: The influence of esters and carboxylic acids as the N-substituent of opioids. Part 1: Benzomorphans.  
|| Abstract: To investigate the effects of carboxylic ester and acid moieties as the N-substituent of opioids, a short series of racemic N-substituted normetazocines was prepared. The introduction of both groups as the normetazocine N-substituent produced compounds which displayed low potency in vitro and in vivo, with the esters displaying the greater activity. The pharmacology of the compounds is discussed with implications resulting from potential in vivo metabolic hydrolysis.

**Properties:**

**Odor:**

Odorless

**Color/Form:**

Prisms

**Physical Description:**

Solid

**Boiling Point:**

190°C

**Melting Point:**

255 °C

**Computed Properties**

Property Name	Property Value	Reference
Molecular Weight	285.34	Computed by PubChem 2.1 (PubChem release 2021.05.07)
XLogP3	0.8	Computed by XLogP3 3.0 (PubChem release 2021.05.07)
Hydrogen Bond Donor Count	2	Computed by Cactvs 3.4.8.18 (PubChem release 2021.05.07)
Hydrogen Bond Acceptor Count	4	Computed by Cactvs 3.4.8.18 (PubChem release 2021.05.07)
Rotatable Bond Count	0	Computed by Cactvs 3.4.8.18 (PubChem release 2021.05.07)
Exact Mass	285.13649347	Computed by PubChem 2.1 (PubChem release 2021.05.07)
Monoisotopic Mass	285.13649347	Computed by PubChem 2.1 (PubChem release 2021.05.07)
Topological Polar Surface Area	52.9 Å <sup>2</sup>	Computed by Cactvs 3.4.8.18 (PubChem release 2021.05.07)
Heavy Atom Count	21	Computed by PubChem
Formal Charge	0	Computed by PubChem

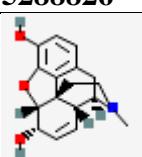
**Q1- Identify a drug molecule from PubChem and show compound substance, BioAssay, and Patent information.**

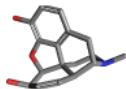
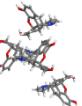
**Molecule:**

Morphine

**Compound Substance:**

Morphine

<b>PubChem CID</b>	<b>5288826</b>
Structure	

	 
Chemical Safety	Laboratory Chemical Safety Summary (LCSS) Datasheet
Molecular Formula	C <sub>17</sub> H <sub>19</sub> NO <sub>3</sub>
Synonyms	morphine Morphia Morphium Morphinum Morphin More...
Molecular Weight	285.34

## BioAssay:

pubchem.ncbi.nlm.nih.gov/#query=morphin&tab=bioassay

**PubChem**

[Compounds \(28\)](#) [Substances \(69\)](#) [BioAssays \(1\)](#) [Literature \(57\)](#) [Patents \(48\)](#)

Searching descriptions and metadata of bioassay records submitted by PubChem's contributors. [Read More...](#)

1 result

[Antinociceptive activity in sc dosed mouse assessed as inhibition of morphin response by tail flick assay](#)

BioAssay AID: 320416 BioAssay Type: Literature-derived  
 Tested Compounds Count: 11 Tested Substances Count: 11  
 Data Source: ChEMBL External ID: CHEMBL929785  
 Data Source Category: Curation Efforts; Research and Development  
 Modified Date: 2018-10-23  
 Description: Title: The influence of esters and carboxylic acids as the N-substituent of opioids. Part 1: Benzomorphans. II. Abstract: To investigate the effects of carboxylic ester and acid moieties as the N-substituent of opioids, a short series of racemic N-substituted normetazocines was prepared. The introduction of both groups as the normetazocine N-substituent produced compounds which displayed low potency in vitro and in vivo, with the esters displaying the greater activity. The pharmacology of the compounds is discussed with implications resulting from potential in vivo metabolic hydrolysis.

[Download](#) [Search in Entrez](#)

ACTIONS ON RESULTS WITH ID TYPE  
 BioAssays  
 Substances  
 Compounds  
[Push to Entrez](#)  
[Save for Later](#)  
[Linked Data Sets](#)

## Patent information:

## 15 Patents

US5378474	US5807572	US7815934
CA2065210	US5891467	US7682634
CA128591	US6171613	US8877247
US6066339	US8685443	US9192608
US5931809	US8158156	US9072781
US5962016	US7682633	US9248229
US6193998	US8623418	US7955619
US5997899	US8685444	US9044402
US6241999	US8846104	US9549899
		US10314788

Process for the preparation of quaternary N alkyl morphin or morphinan alkaloid derivatives

Publication Number: US-8101756-B2

Patent Family: CA-2638195-A1; CA-2638195-C; DK-2039696-T3; EP-2019105-A1; EP-2039696-A1; EP-2039696-B1; ES-2628060-T3; US-2009054651-A1; US-8101756-B2

Priority Date: 2007-07-19

Grant Date: 2012-01-24

Inventor(s): EIPERT MARTIN; LAUTERBACH ERIK HEINZ; HELLER SABRINA; DINKEL THOMAS; HAKE STEPHANIE

Assignee(s): EIPERT MARTIN; LAUTERBACH ERIK HEINZ; HELLER SABRINA; DINKEL THOMAS; HAKE STEPHANIE; ...

Classification: C07D489/02; C07D489/08

Abstract: The present invention relates to a process for the preparation of quaternary N-alkyl morphin or morphinan alkaloid derivatives. This is achieved by using a nucleophilic nitrogen, phosphor or sulfur containing base in the reaction mixture.

Linked Compounds Count: 26

Linked Substances Count: 39

Publication Number	Title	Priority Date	Grant Date
US-11033208-B1	Fixed operation time frequency sweeps for an analyte sensor	2021-02-05	2021-06-15
US-11033516-B1	Combination therapies with disulfiram	2020-09-18	2021-06-15
US-10907163-B1	Aptamers that bind to natural and synthetic cannabinoids	2020-06-15	2021-02-02
US-10940133-B1	Methods of providing solriamfetol therapy to subjects with impaired renal function	2020-03-19	2021-03-09
US-10980756-B1	Methods of treatment	2020-03-16	2021-04-20

WhatsApp App x | PubChem x | New Tab x | +

pubchem.ncbi.nlm.nih.gov/#query=morphin&tab=patent

PubChem morphin

Compounds (28) Substances (69) BioAssays (1) Literature (57) Patents (48)

Searching patent abstracts and metadata. [Read More...](#)

48 results Filters SORT BY Relevance Download

**Process for the preparation of quaternary N-alkyl morphin or morphinan alkaloid derivatives**

Publication Number: US-8101756-B2

Patent Family: CA-2638195-A1; CA-2638195-C; DK-2039696-T3; EP-2019105-A1; EP-2039696-A1; EP-2039696-B1; ES-2628060-T3; US-2009054651-A1; US-8101756-B2

Priority Date: 2007-07-19 Grant Date: 2012-01-24

Inventor(s): EIPERT MARTIN; LAUTERBACH ERIK HEINZ; HELLER SABRINA; DINKEL THOMAS; HAKE STEPHANIE

Assignee(s): EIPERT MARTIN; LAUTERBACH ERIK HEINZ; HELLER SABRINA; DINKEL THOMAS; HAKE STEPHANIE; ...

Classification: C07D489/02; C07D489/08

Abstract: The present invention relates to a process for the preparation of quaternary N-alkyl morphin or morphinan alkaloid derivatives. This is achieved by using a nucleophilic nitrogen, phosphor or sulfur containing base in the reaction mixture.

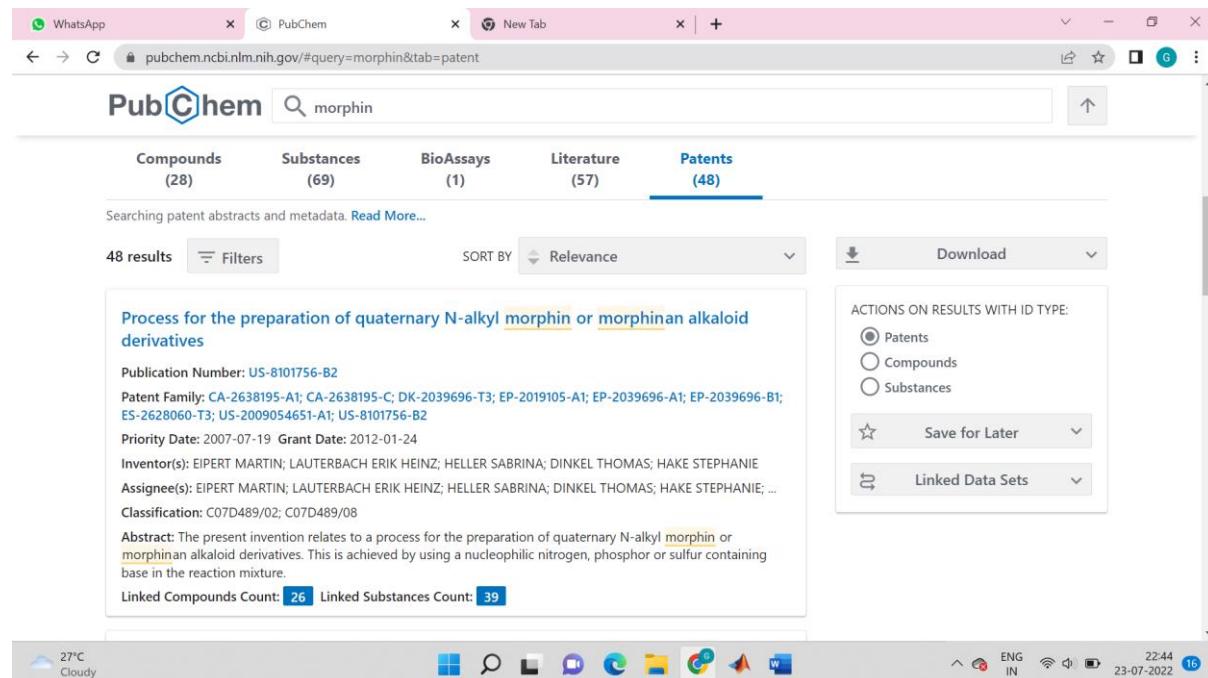
Linked Compounds Count: 26 Linked Substances Count: 39

ACTIONS ON RESULTS WITH ID TYPE:

Patents  
 Compounds  
 Substances

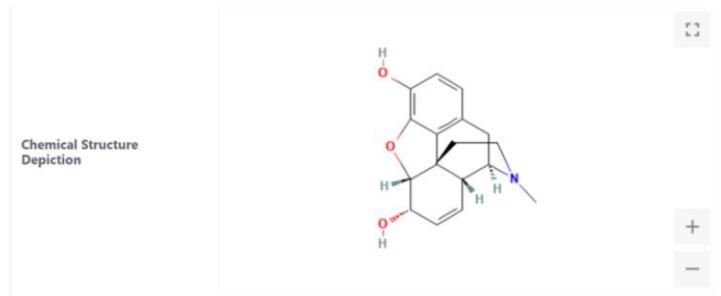
Save for Later

Linked Data Sets



**Q2-Identify the 2D structure and 3D conformer of the molecule and list out the similar 3D structure compounds. Visualize the 3D structure of compounds using DS (Discovery Studio) software.**

## 2D structure



**Visualizing 3D structure using DS (Discovery Studio) software:**

BID 19006

Step 1: Open the morphine 3D structure in pubchem:

WhatsApp x | Meet - fdn-yxq-trk x | PubChem x | Inbox (536) - gun.des.bt1 x | (no subject) - gunjandesha x | +

pubchem.ncbi.nlm.nih.gov/?query=CID129850019%20structure&tab=similarity\_3d

# PubChem

CID129850019 structure

624 results (incomplete) Filters SORT BY Shape Then Feature

**Morphine-D6**

Compound CID: 129850019

MF:  $C_{17}H_{19}NO_3$  MW: 291.37g/mol

IUPAC Name: (4R,4aR,7S,7aR,12bS)-4,4a,5,6,7,7a-hexadeuterio-3-methyl-2,13-dihydro-1H-4,12-methanobenzofuro[3,2-e]isoquinoline-7,9-diol

Isomeric SMILES: [2H]C1=C([C@]([C@]2[C@@]3[C@]([C@]1[C@]([C@@]2CC5=C3C(=C(C=C5)O)O2)(N(CC4)C)[2H])[2H])[2H])O[2H]

InChIKey: BQJCRHHNABKAKU-XSCBNIJMSA-N

InChI: InChI=1S/C17H19NO3/c1-18-7-6-17-10-3-5-13(20)16(17)21-15-12(19)4-2-9(14(15)17)8-11(10)18/h2-5,10-11,13,16,19-20H,6-8H2,1H3/t10-,11+,13-,16-,17-/m0/s1/i3D,5D,10D,11D,13D,16D

Create Date: 2017-09-13

Download

ACTIONS ON RESULTS WITH ID TYPE:

Compounds

Push to Entrez

Save for Later

Linked Data Sets

Summary Similar Structures Search Related Records

Compound CID: 129631816

MF:  $C_{17}H_{19}NO_3$  MW: 288.36g/mol

IUPAC Name: (4R,4aR,7S,7aR,12bS)-4,4a,5,6,7,7a-hexadeuterio-3-methyl-2,13-dihydro-1H-4,12-methanobenzofuro[3,2-e]isoquinoline-7,9-diol

meet.google.com is sharing your screen. Stop sharing Hide

28°C Cloudy ENG IN 13:04 24-07-2022

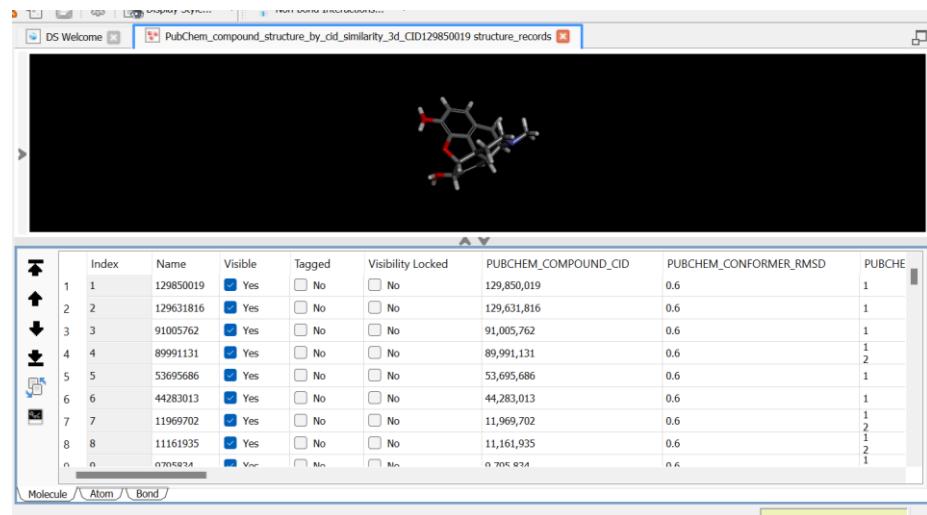
Step 2: Download the 3D chemical structure in .sdf format:

## BID 19006

Step 3: Select the top 10 molecules and click on show in graphic view format:

	Visibility	Locked	PUBCHEM_COMPOUND_CID	PUBCHEM_CONFORMER
1	<input type="checkbox"/>	No	129,850,019	0.6
2	<input type="checkbox"/>	No	129,631,816	0.6
3	<input type="checkbox"/>	No	91,005,762	0.6
4	<input type="checkbox"/>	No	89,991,131	0.6
5	<input type="checkbox"/>	No	53,695,686	0.6
6	<input type="checkbox"/>	No	44,283,013	0.6
7	<input type="checkbox"/>	No	11,969,702	0.6
8	<input type="checkbox"/>	No	11,161,935	0.6
9	<input type="checkbox"/>	No	9,795,834	0.6
10	<input type="checkbox"/>	No	6,912,286	0.6
11	<input type="checkbox"/>	No	5,288,826	0.6
12	<input type="checkbox"/>	No	86,587,414	0.6
13	<input type="checkbox"/>	No	23,253,564	0.6
14	<input type="checkbox"/>	No	10,850,629	0.6
15	<input type="checkbox"/>	No	54,394,266	0.6
16	<input type="checkbox"/>	No	44,540,051	0.6
17	<input type="checkbox"/>	No	15,677,952	0.6

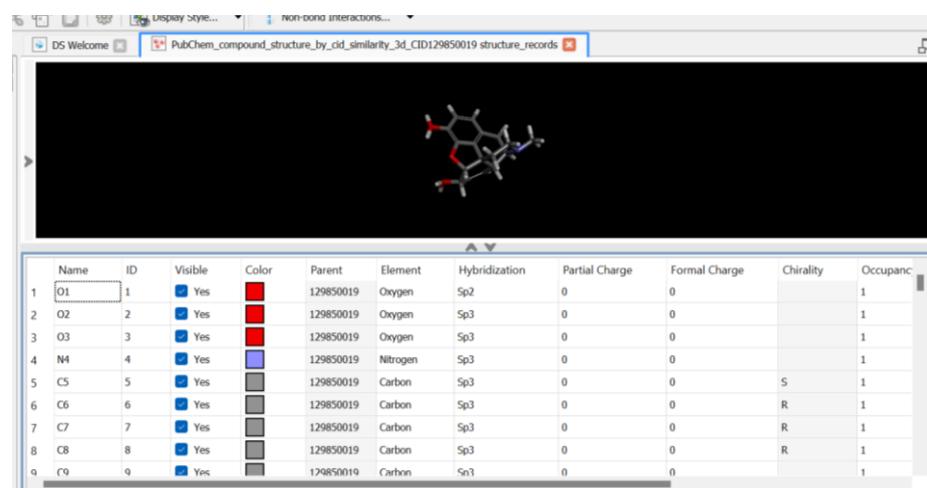
## BID 19006



Grey colour rings = hydrophobic

Red colour rings = negatively charged pharmacophore features

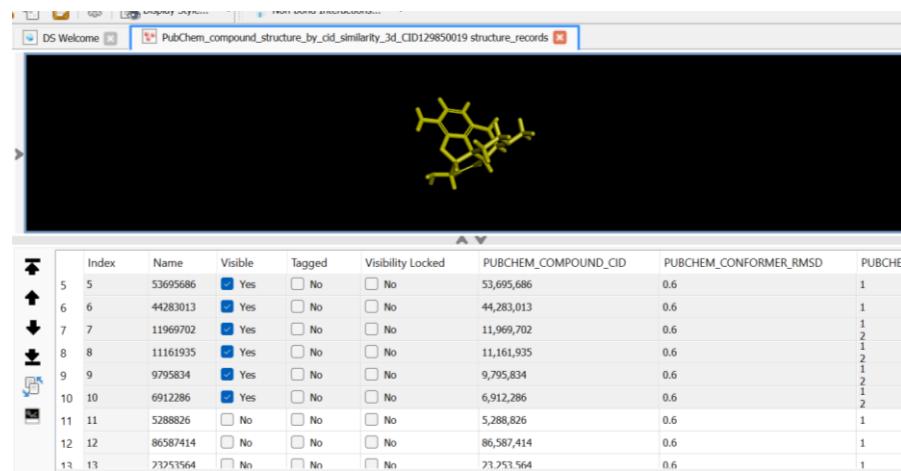
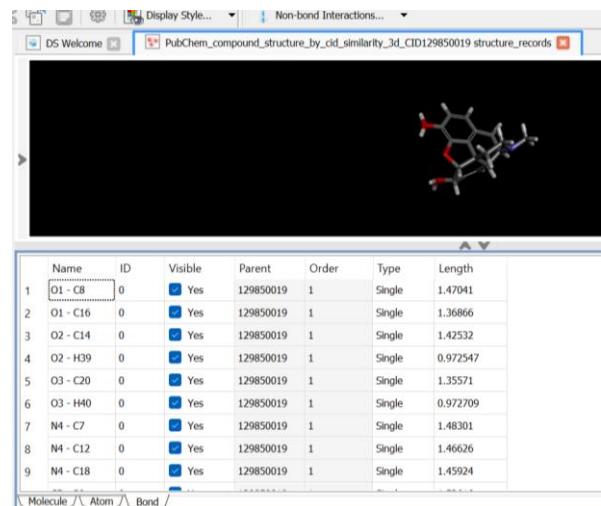
Blue colour rings = positively charged



Grey colour rings = hydrophobic

Red colour rings = negatively charged pharmacophore features

Blue colour rings = positively charged



### Q3-List out the chemical and physical properties of the drug.

#### Computed Properties

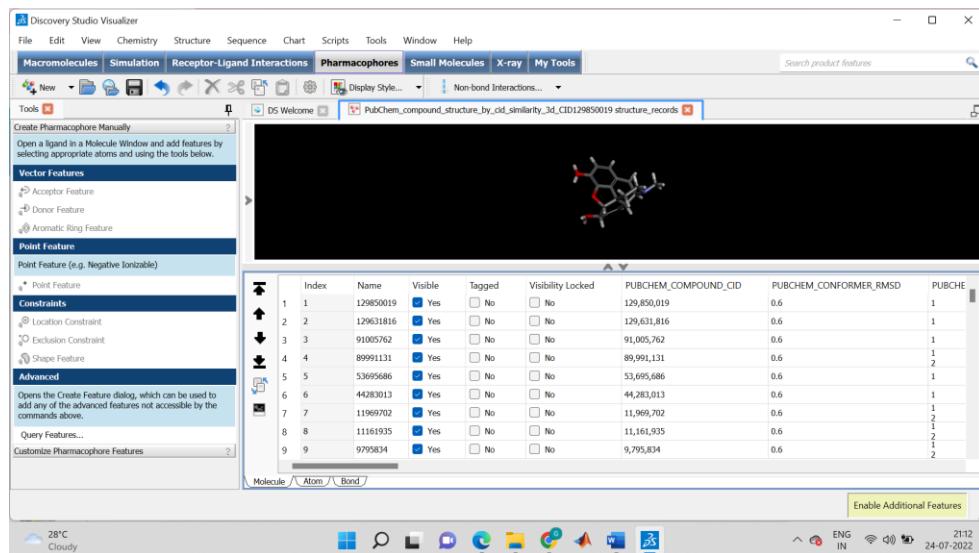
Property Name	Property Value	Reference
Molecular Weight	285.34	Computed by PubChem 2.1 (PubChem release 2021.05.07)
XLogP3	0.8	Computed by XLogP3 3.0 (PubChem release 2021.05.07)
Hydrogen Bond Donor Count	2	Computed by Cactvs 3.4.8.18 (PubChem release 2021.05.07)
Hydrogen Bond Acceptor Count	4	Computed by Cactvs 3.4.8.18 (PubChem release 2021.05.07)
Rotatable Bond Count	0	Computed by Cactvs 3.4.8.18 (PubChem release 2021.05.07)

Exact Mass	285.13649347	Computed by PubChem 2.1 (PubChem release 2021.05.07)
Monoisotopic Mass	285.13649347	Computed by PubChem 2.1 (PubChem release 2021.05.07)
Topological Polar Surface Area	52.9 Å <sup>2</sup>	Computed by Cactvs 3.4.8.18 (PubChem release 2021.05.07)
Heavy Atom Count	21	Computed by PubChem
Formal Charge	0	Computed by PubChem
Complexity	494	Computed by Cactvs 3.4.8.18 (PubChem release 2021.05.07)
Isotope Atom Count	0	Computed by PubChem
Defined Atom Stereocenter Count	5	Computed by PubChem
Undefined Atom Stereocenter Count	0	Computed by PubChem
Defined Bond Stereocenter Count	0	Computed by PubChem
Undefined Bond Stereocenter Count	0	Computed by PubChem
Covalently-Bonded Unit Count	1	Computed by PubChem
Compound Is Canonicalized	Yes	Computed by PubChem (release 2021.05.07)

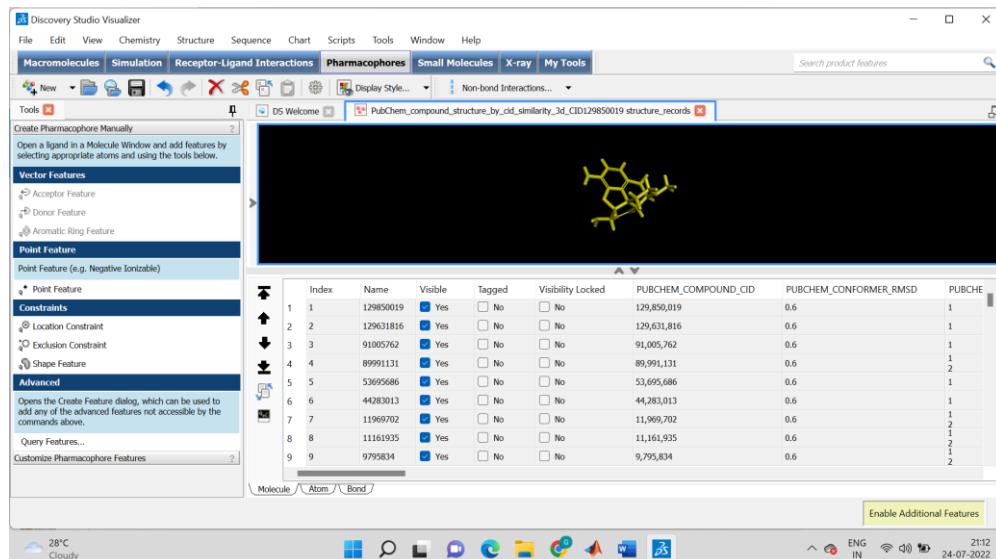
**Q4-Visualize the key pharmacophore features of the top 10 molecules/compounds using DS (Discovery Studio) software.**

Step 1: Go to “create pharmacophore manually” :

## BID 19006

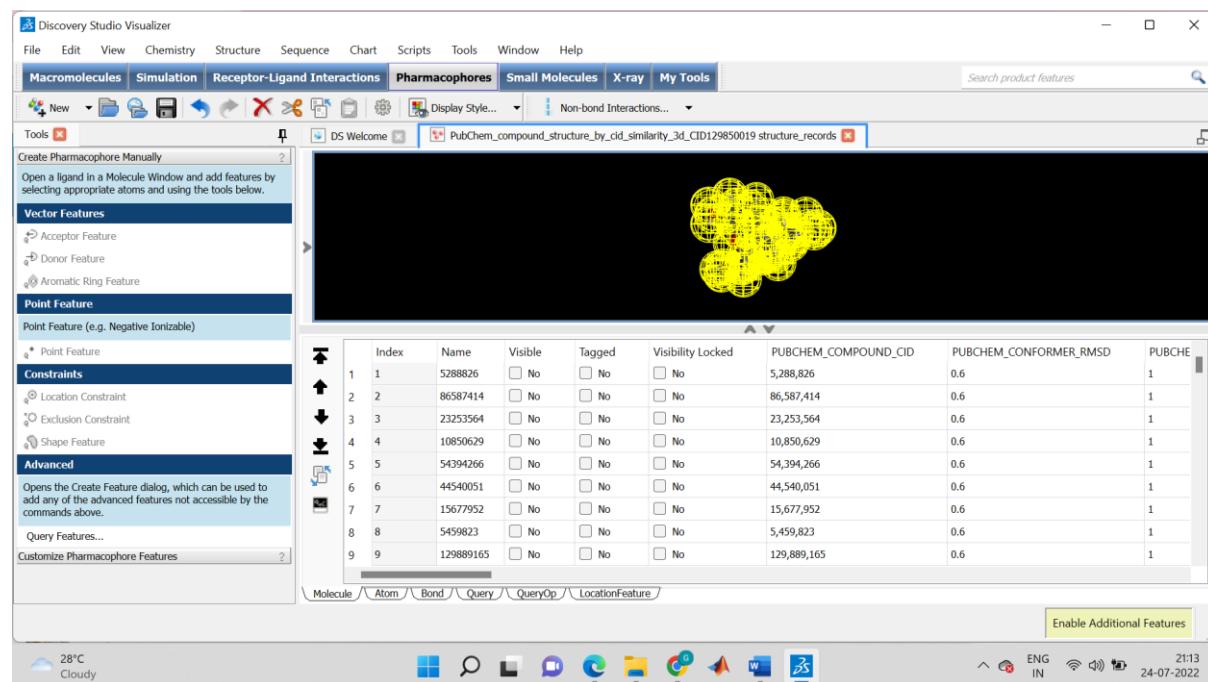


Step 2: Using cursor draw line around the structure and see the yellow colour:



Step 3: Now click on Location constraint:

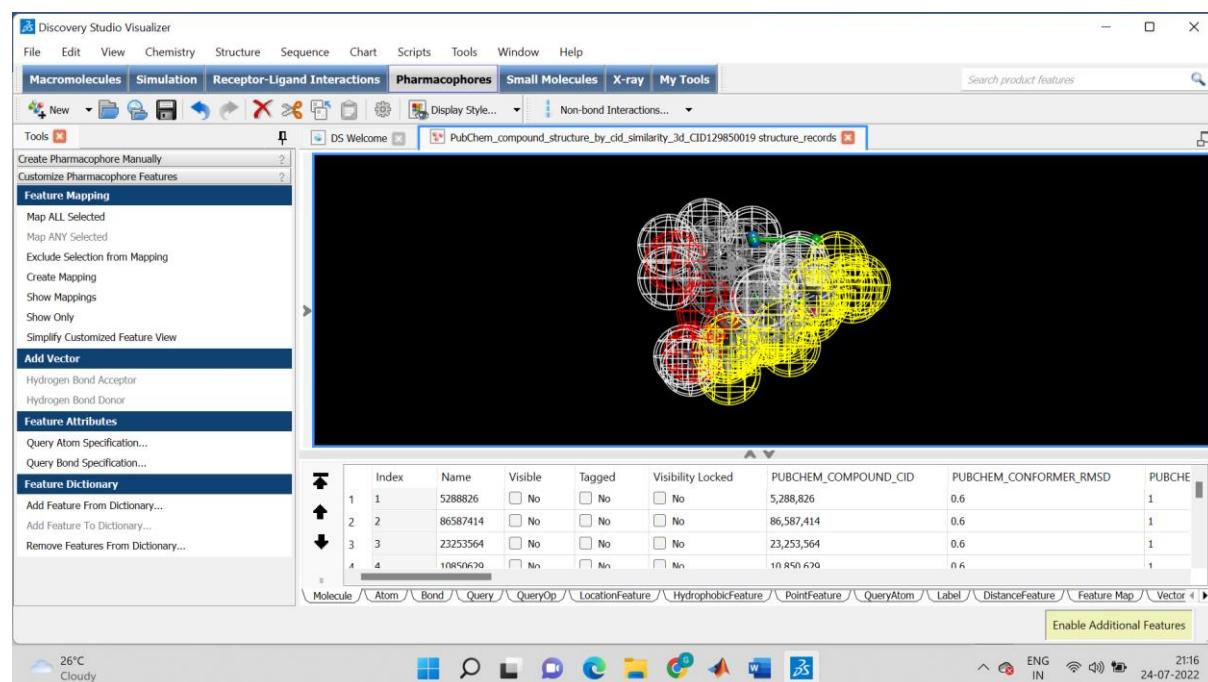
## BID 19006



Grey colour rings = hydrophobic

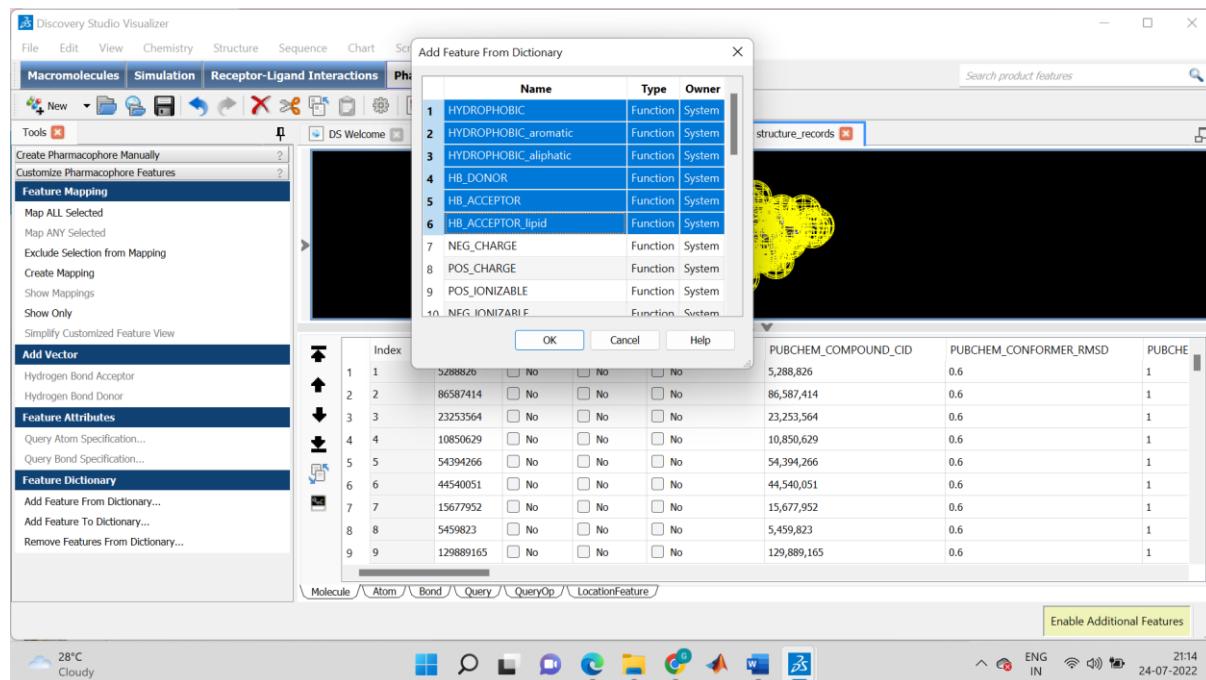
Red colour rings = negatively charged pharmacophore features

Blue colour rings = positively charged



Step 4: Now add features from dictionary:

## BID 19006

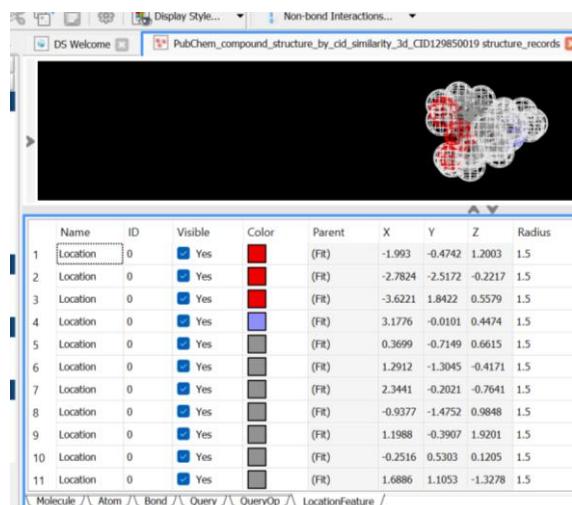


### Step 5: Location features:

Grey colour rings = hydrophobic

Red colour rings = negatively charged pharmacophore features

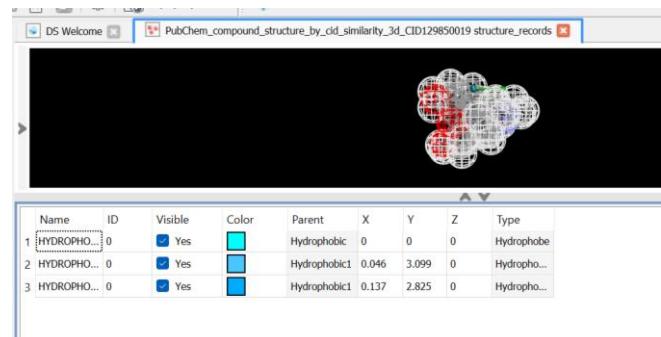
Blue colour rings = positively charged



BID 19006

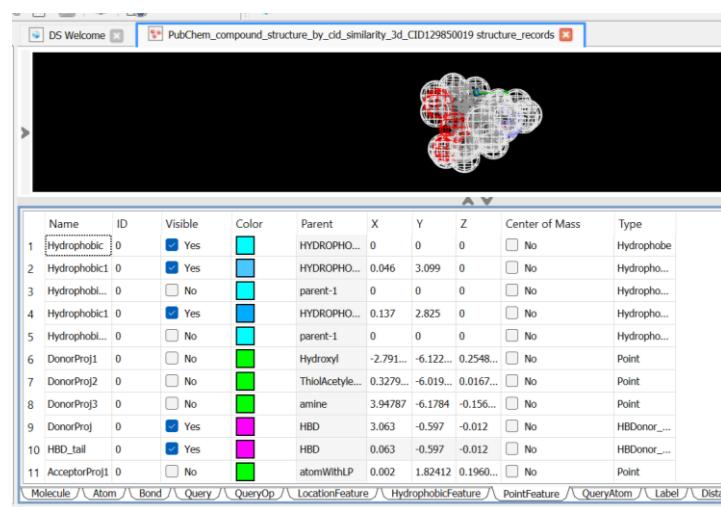
### Step 6: Hydrophobic features

Hydrophobic features- 10 columns (1 column hidden)



### Step 7: Point features

11columns (1 column hidden)

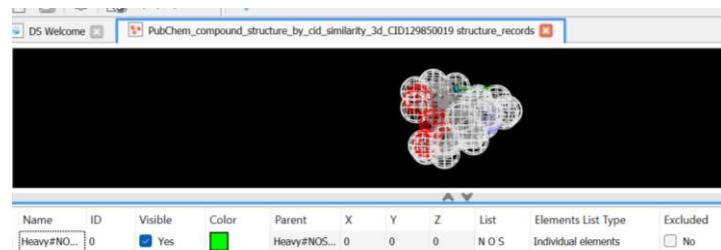


### Step 8: Query Atom

Parent : Heavy NOS

Colour : Green

Element List Type : Individual elements



Step 9: Vector Features:  
11 columns (1 column hidden)

Pink : Hydroxyl

Pink : ThiolAcetylene

Pink : amine

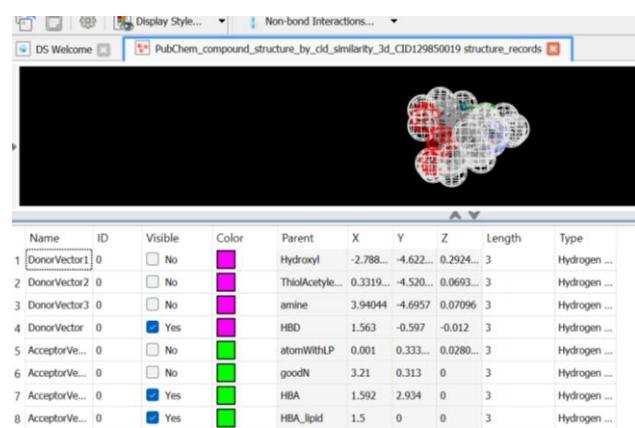
Pink : HBD

Green : AtomWithLP

Green : GoodN

Green :HBA

Green : HBA\_lipid



### Practical Number: 3

## Practical title: Molecular Interactions Studies using Autodock

### Introduction:

Download Autodock

Save both the files in a separate folder on desktop:

- autogrid.eve
- autodock.exe

### Objective:

To study molecular interaction between protein and respective ligand using genetic algorithm based approach.

Step 1: Go to RCSB PDB and search for 1BDR (HIV protease complex):

The screenshot shows the RCSB PDB interface for entry 1BDR. The main content area displays a 3D ribbon model of the HIV-1 protease complexed with inhibitor SB203386. Below the model, there is an 'Experimental Data Snapshot' section and a 'wwPDB Validation' section. The validation section includes a table with metrics like Method (X-RAY DIFFRACTION), Resolution (2.80 Å), and R-Value Free (0.255). On the right side of the page, there is a sidebar with various download options for the PDB file in different formats.

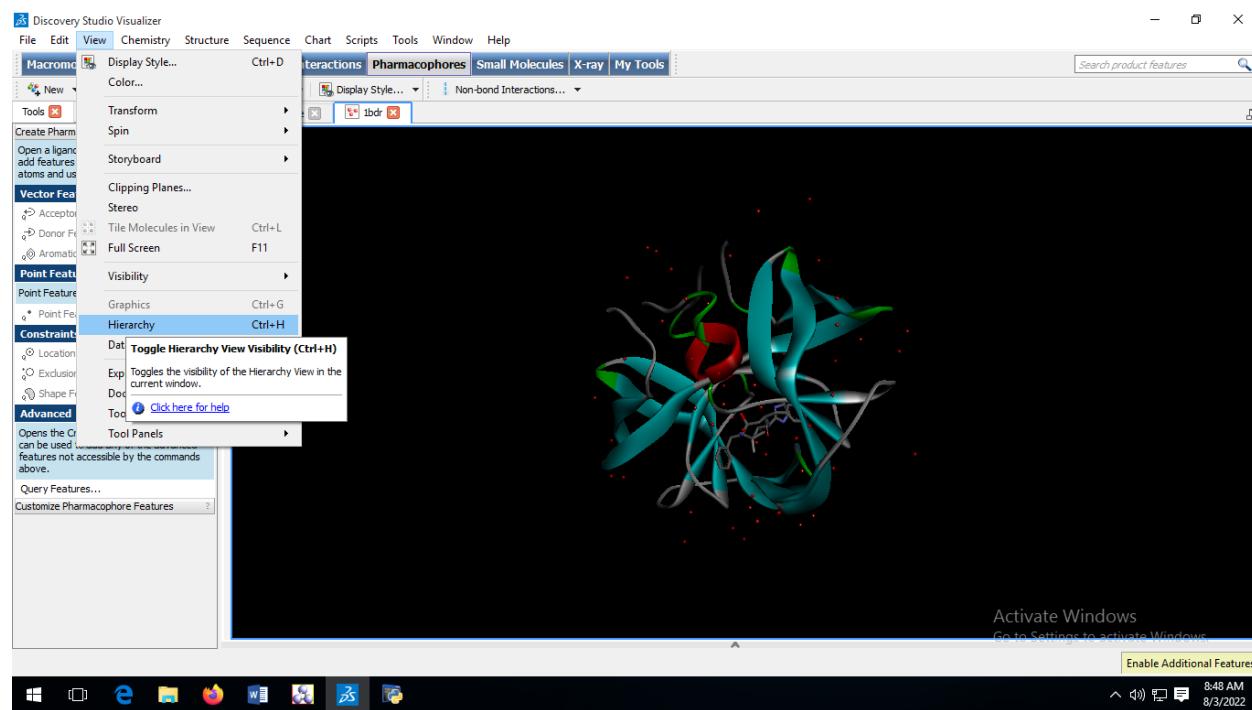
Download the .pdb format

The screenshot shows the same RCSB PDB interface for entry 1BDR. A dropdown menu is open on the right side, listing various file formats for download. The options include FASTA Sequence, PDB/mmCIF Format, PDB/mmCIF Format (gz), PDB Format, PDB Format (gz), PDBML/XML Format (gz), Structure Factors (CIF), Structure Factors (CIF - gz), Validation Full PDF, Validation XML, Validation CIF, Biological Assembly 1 (CIF - gz), and Biological Assembly 1 (PDB - gz).

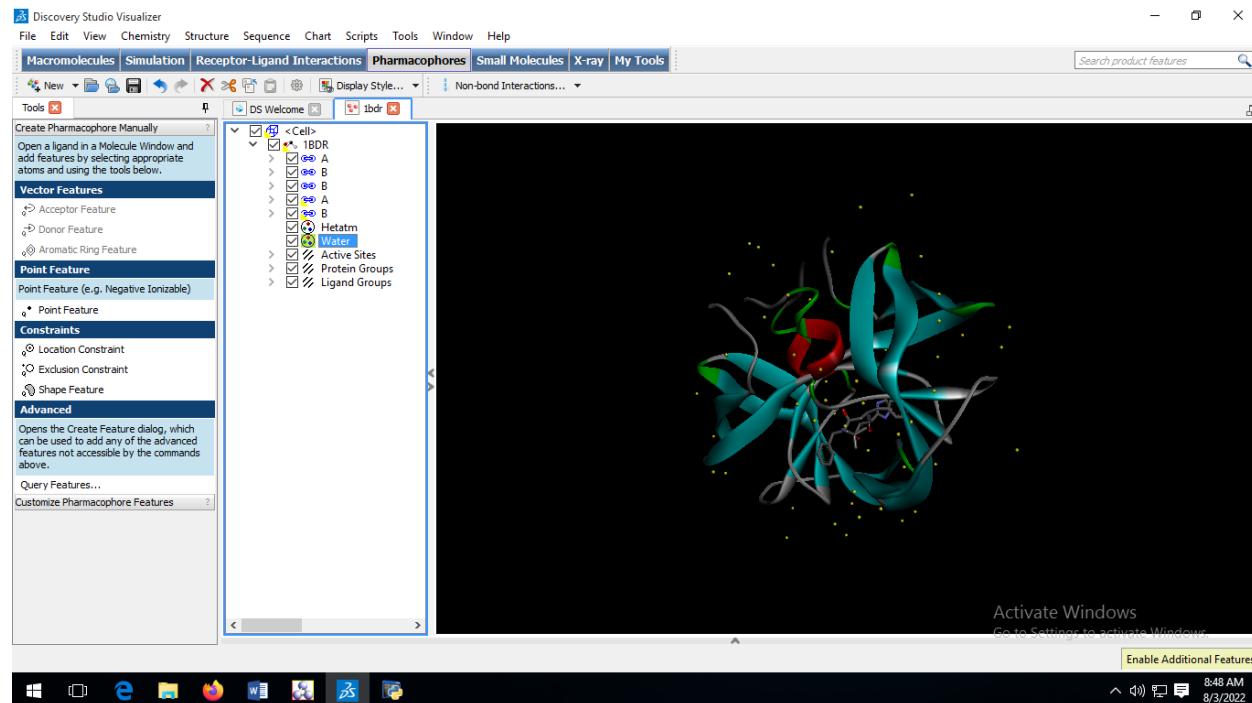
Step 2: Open with discovery studio visualizer

# BID 19006

Go to view, hierarchy:

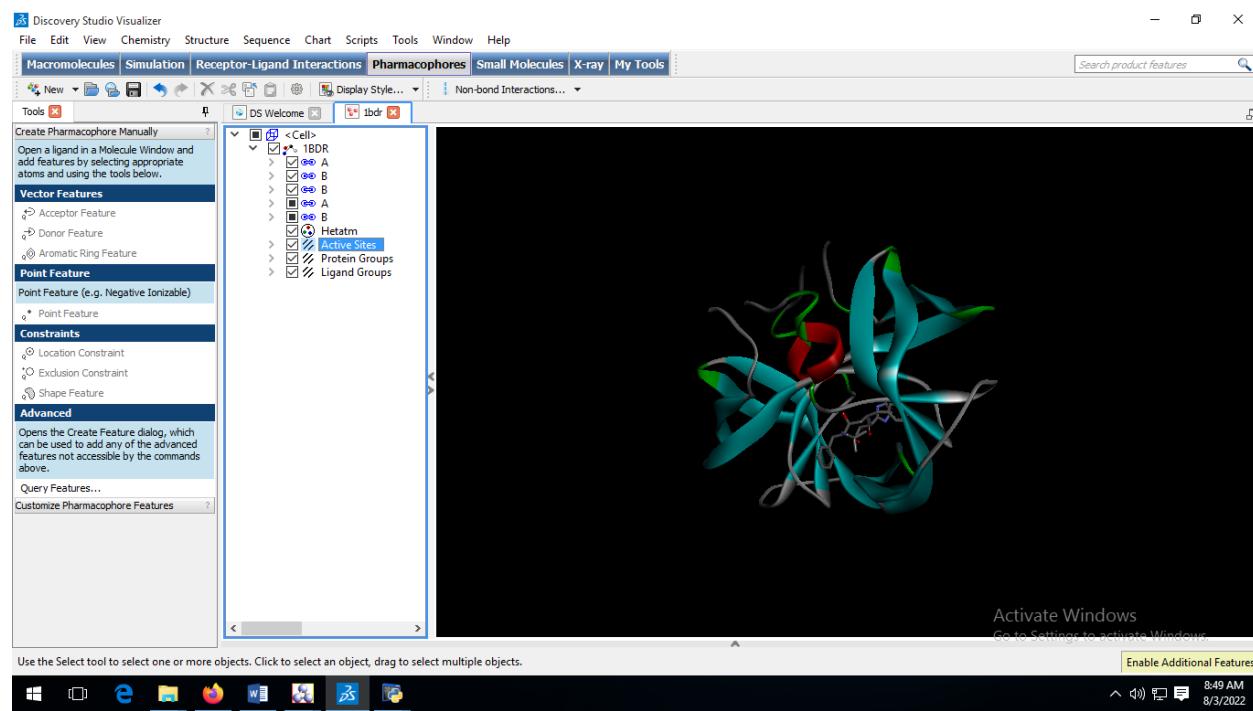


Step 3: Remove water molecules:

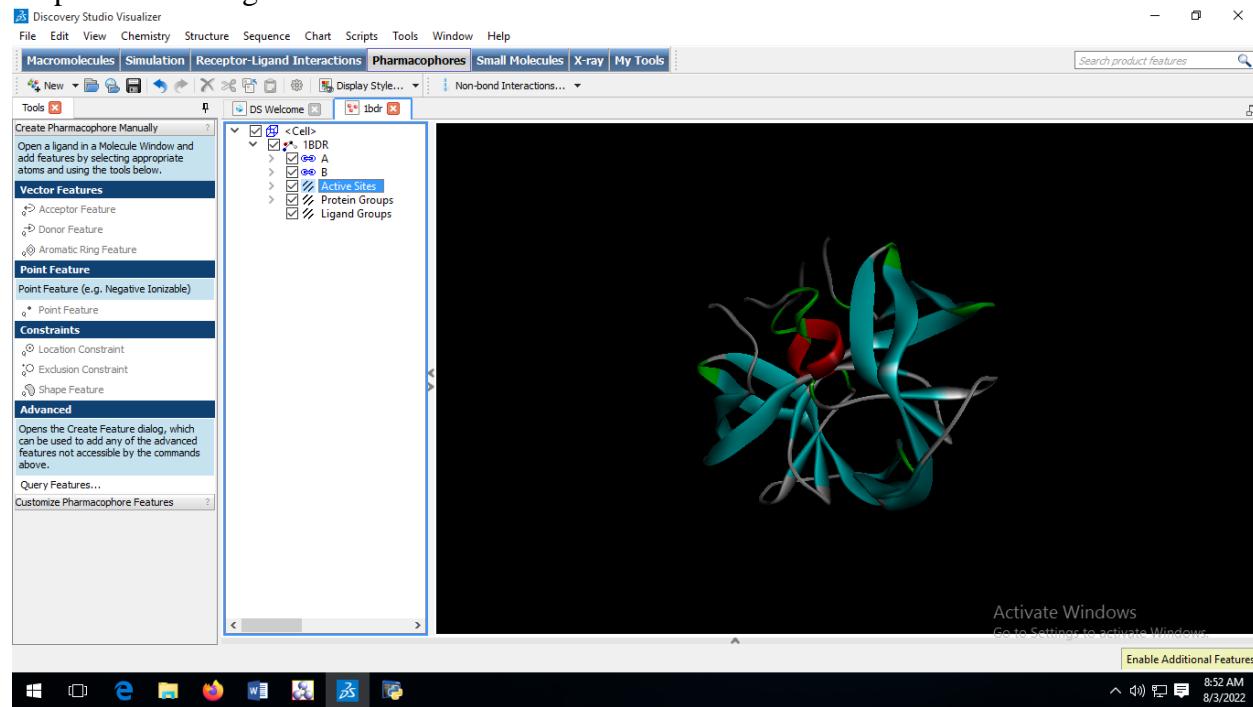


Step 4: Water molecules removed:

## BID 19006



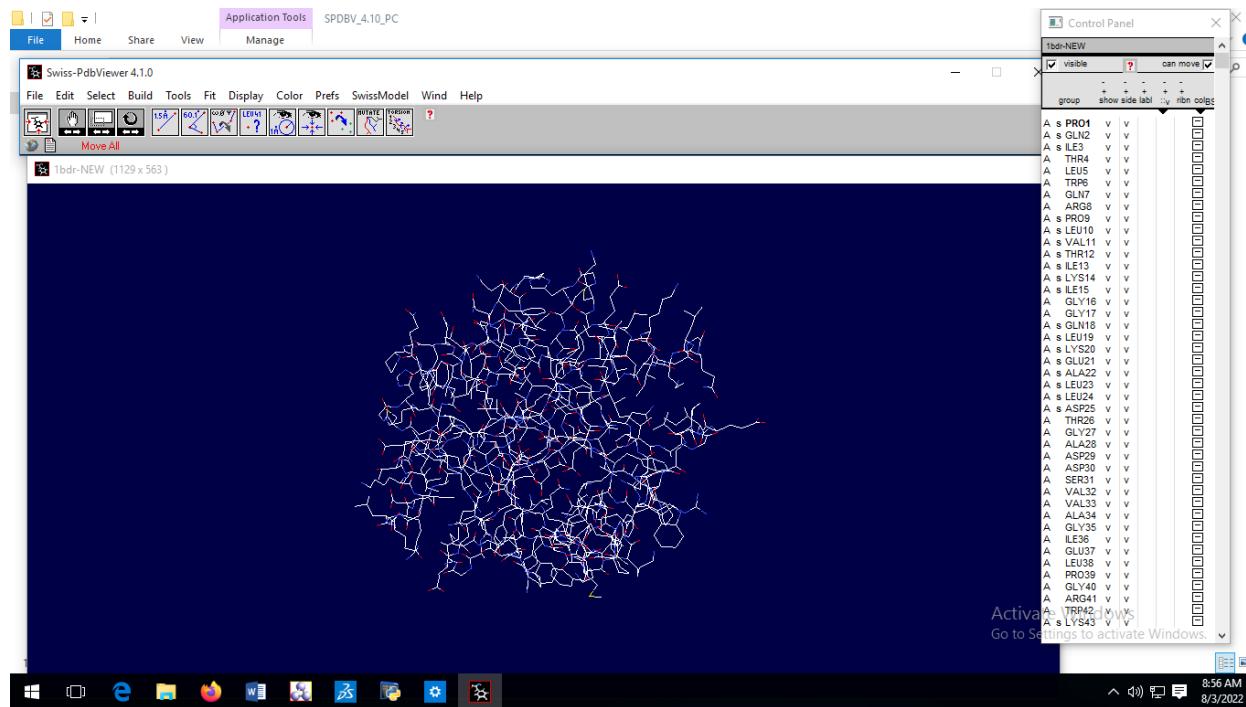
### Step 5: Remove ligand:



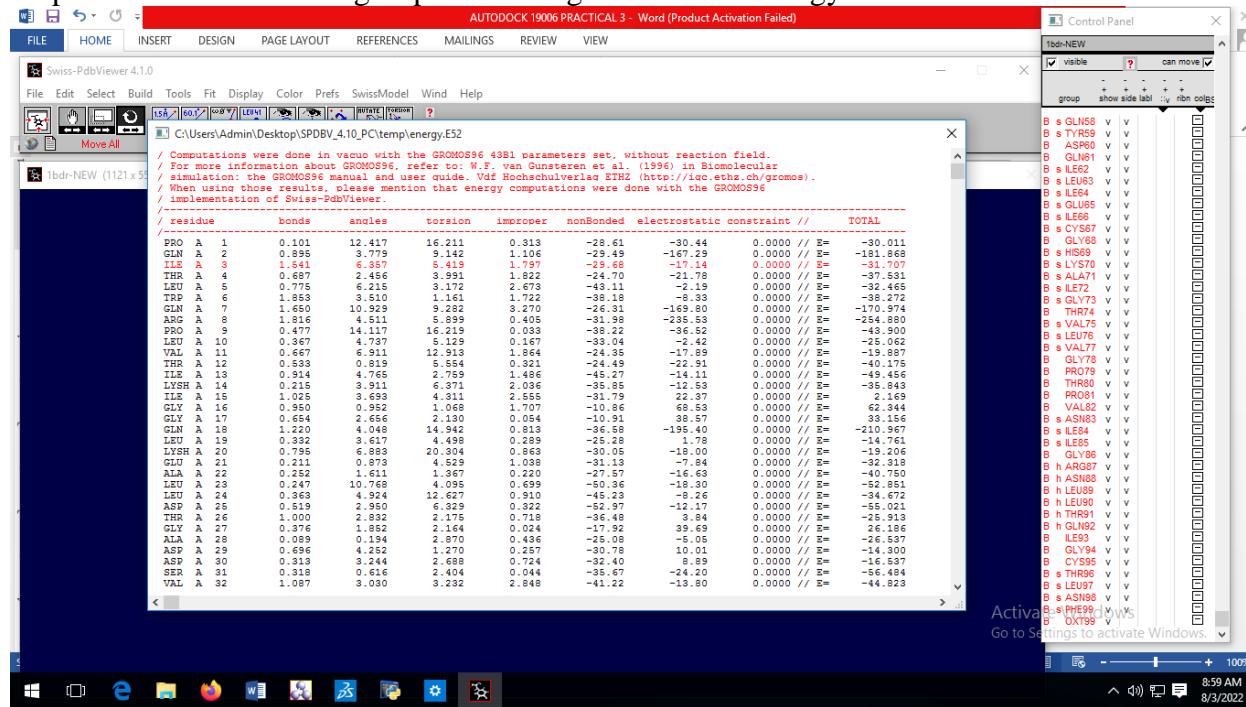
### Step 6: Save protein , go to file, save as, protein data :

Now open it in Swiss PDB viewer:

## BID 19006



Step 7: Now select all the groups and than go to tools for energy minimization:



Step 8: Search for Indinavir in Pub Chem:

The screenshot shows the PubChem homepage with a search bar at the top containing "Indinavir". Below the search bar, there are two main sections: "Compound" and "Gene". The "Compound" section lists several related entries: indinavir, Indinavir sulfate, Indinavir-d6, Indinavir hydrate, and INDINAVIR SULPHATE. The "Gene" section lists: inhibitor of DNA binding 2, inhibitor of DNA binding 3, inhibitor of DNA binding 4, Insulator binding factor 1, and Insulator binding factor 2. The background features a blue hexagonal geometric pattern.

This screenshot shows the detailed compound page for Indinavir (C<sub>36</sub>H<sub>47</sub>N). At the top, it says "PubChem Indinavir (Compound)". Below that is a header for "1.2 3D Conformer". On the left, there's an "Interactive Chemical Structure Model" with options for "Ball and Stick", "Sticks", "Wire-Frame", "Space-Filling", "Show Hydrogens" (which is checked), and "Animate". The main area displays a 3D ball-and-stick model of the indinavir molecule. To the right, there are buttons for "Find Similar 3D Structures", "Get Image", and "Download". A "CONTENTS" sidebar on the right lists various sections: Title and Summary, 1 Structures (which is expanded), 2 Names and Identifiers, 3 Chemical and Physical Properties, 4 Spectral Information, 5 Related Records, 6 Chemical Vendors, 7 Drug and Medication Information, 8 Pharmacology and Biochemistry, 9 Use and Manufacturing, 10 Toxicity, and 11 Associated Disorders and Diseases. A message at the bottom right says "Activate Windows Go to Settings to activate Windows".

## BID 19006

PubChem Indinavir (Compound)

1.2 3D Conformer

Interactive Chemical Structure Model

Ball and Stick  
Sticks  
Wire-Frame  
Space-Filling  
Show Hydrogens  
Animate

Find Similar 3D Structure

Download

SDF Save Display

JSON Save Display

XML Save Display

ASNT Save Display

Cite Download

CONTENTS

Title and Summary

1 Structures

2 Names and Identifiers

3 Chemical and Physical Properties

4 Spectral Information

5 Related Records

6 Chemical Vendors

7 Drug and Medication Information

8 Pharmacology and Biochemistry

9 Use and Manufacturing

10 Toxicity

11 Associated Disorders and Diseases

Activate Windows Go to Settings to activate Windows.

Step 9: Open in graphics view in Discovery Studio Visualizer

Discovery Studio Visualizer

File Edit View Chemistry Structure Sequence Chart Scripts Tools Window Help

Macromolecules Simulation Receptor-Ligand Interactions Pharmacophores Small Molecules X-ray My Tools

New... Open... Save... Display Style... Non-bond Interactions...

Create Pharmacophore Manually

Open a ligand in a Molecule Window and add features by selecting appropriate atoms and using the tools below.

Vector Features

Acceptor Feature  
Donor Feature  
Aromatic Ring Feature

Point Feature

Point Feature (e.g. Negative Ionizable)  
Point Feature

Constraints

Location Constraint  
Exclusion Constraint  
Shape Feature

Advanced

Opens the Create Feature dialog, which can be used to add any of the advanced features not accessible by the commands above.

Query Features...  
Customize Pharmacophore Features

DS Welcome Conformer3D\_CID\_5362440

Molecule / Atom / Bond

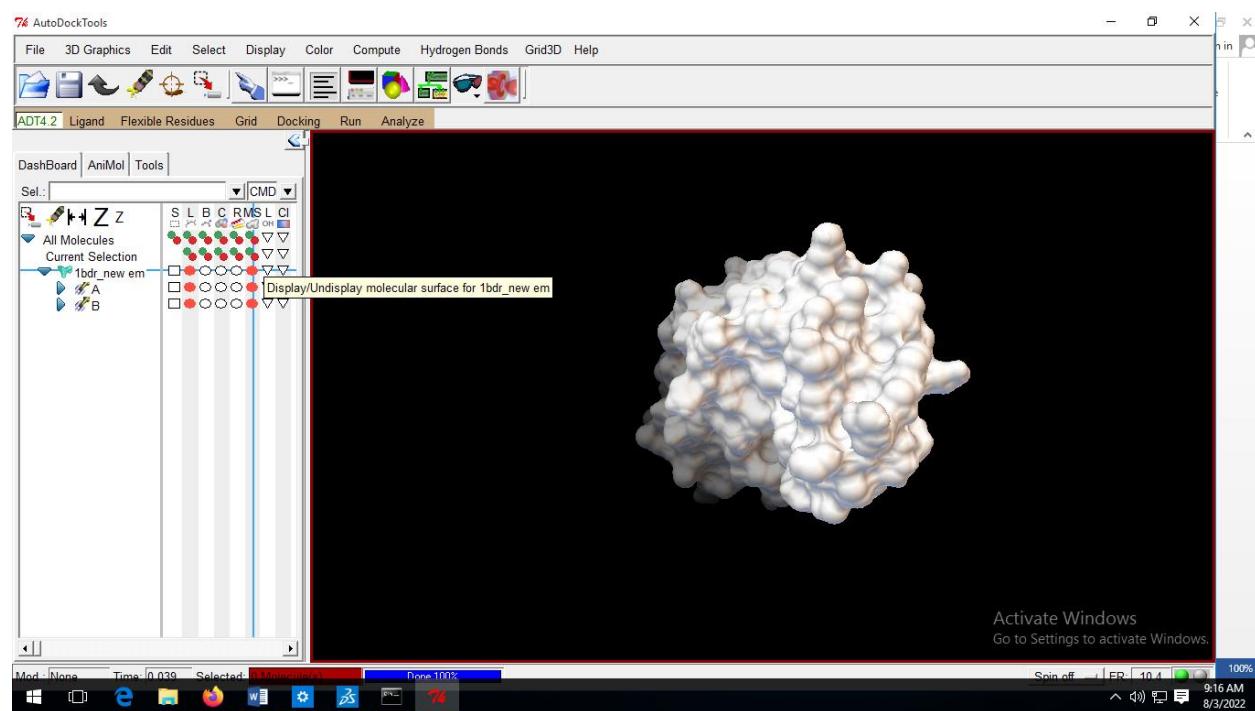
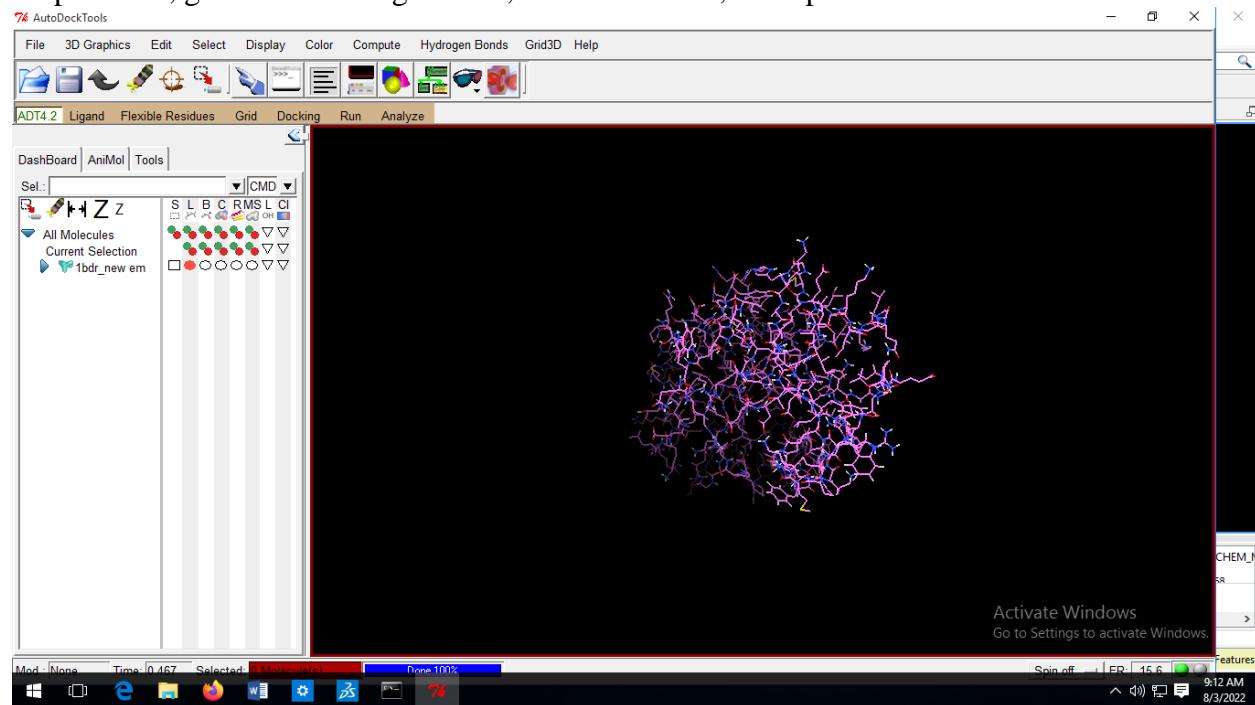
	Index	Name	Visible	Tagged	Visibility Locked	PUBCHEM_COMPOUND_CID	PUBCHEM_CONFORMER_RMSD	PUBCHEM_CONFORMER_DIVERSEORDER	PUBCHEM_N
1	1	5362440	<input checked="" type="checkbox"/> Yes	<input type="checkbox"/> No	<input type="checkbox"/> No	5,362,440	1.6	14	1-0.68

Activate Windows Go to Settings to activate Windows.

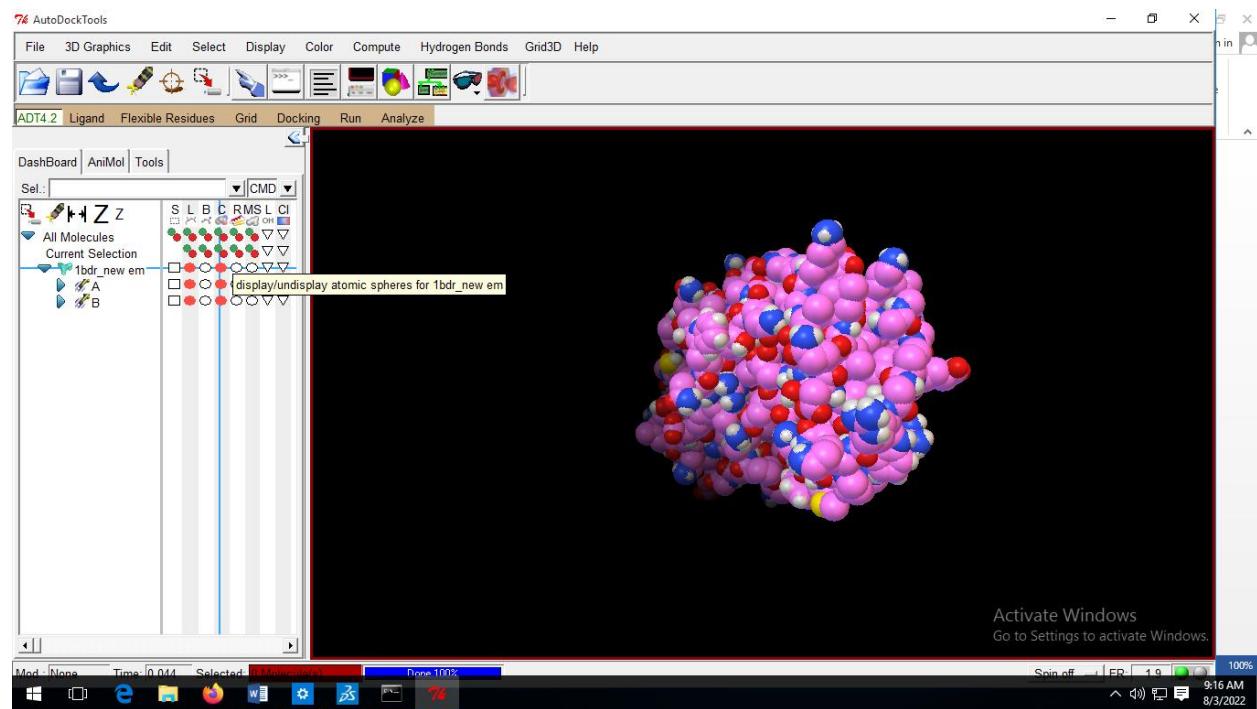
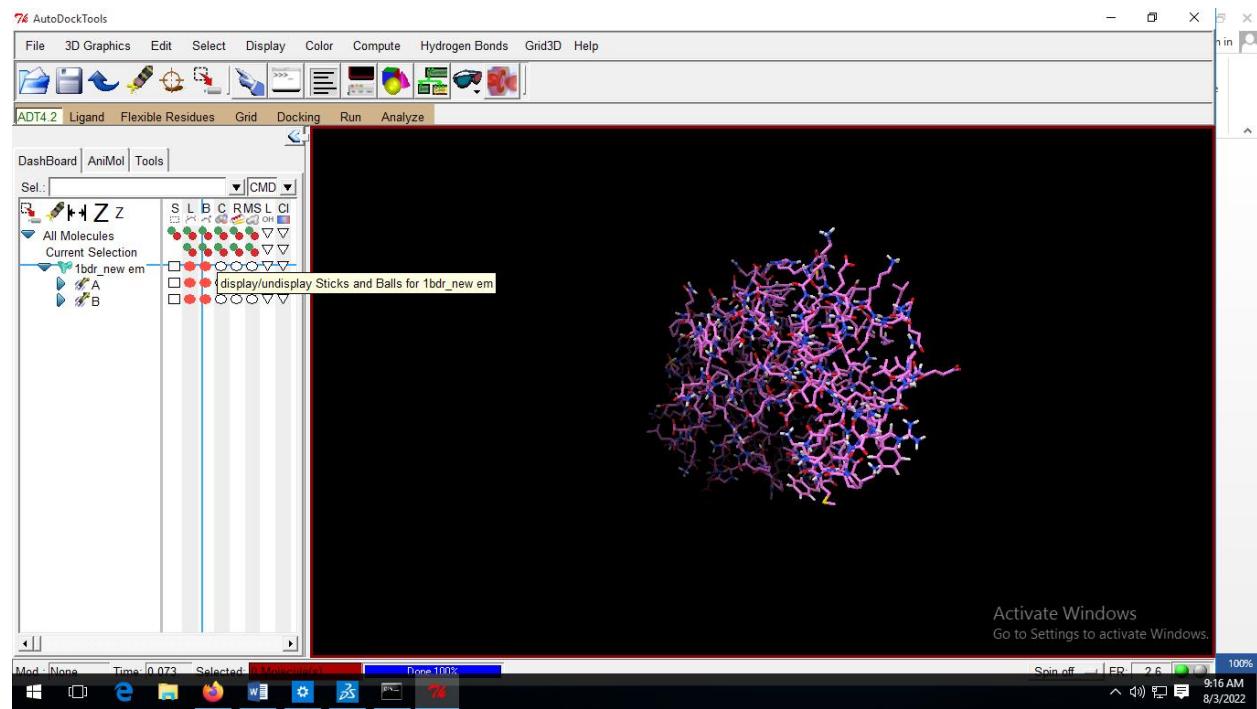
Enable Additional Features

## BID 19006

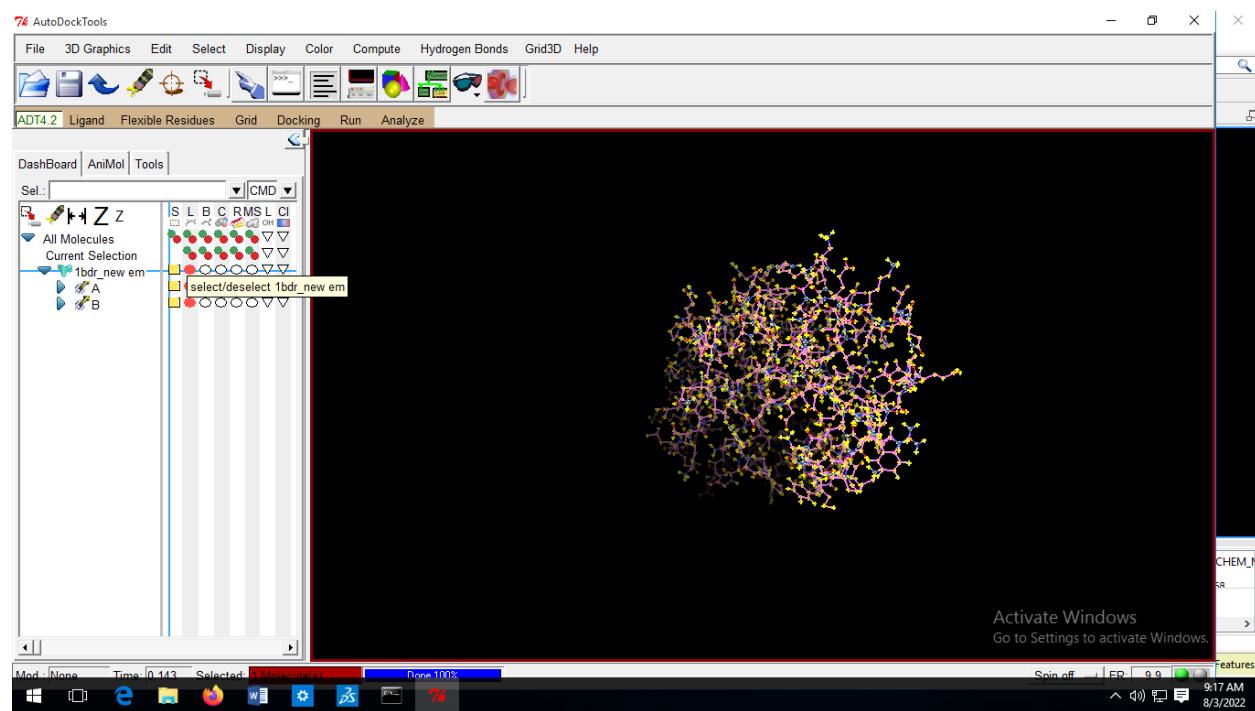
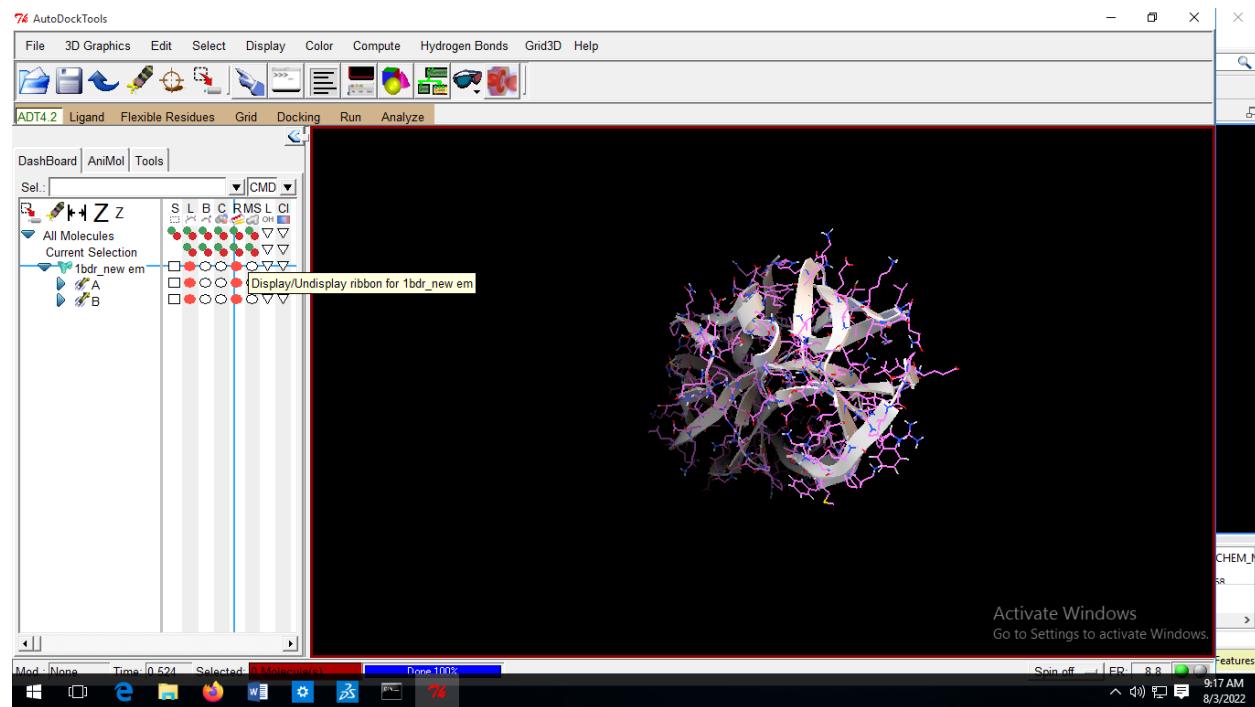
Step 10: So , go to Autodock go to file, read molecule , and open the molecule indinavir



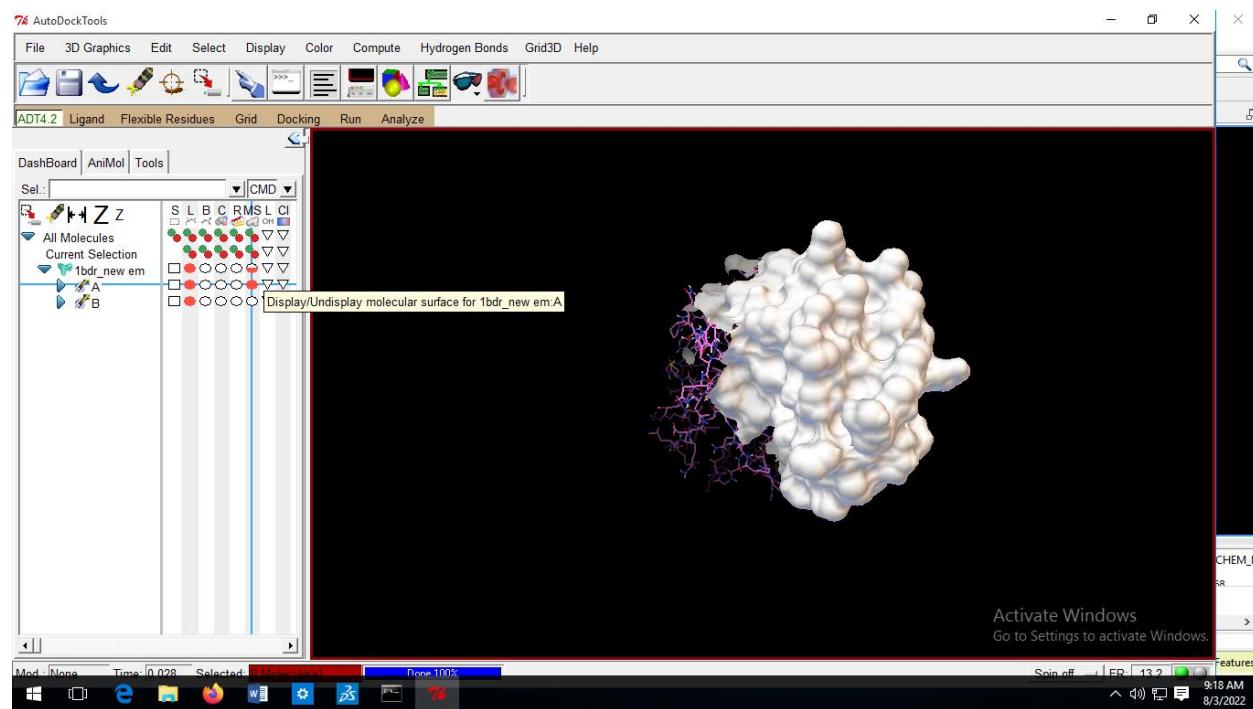
## BID 19006



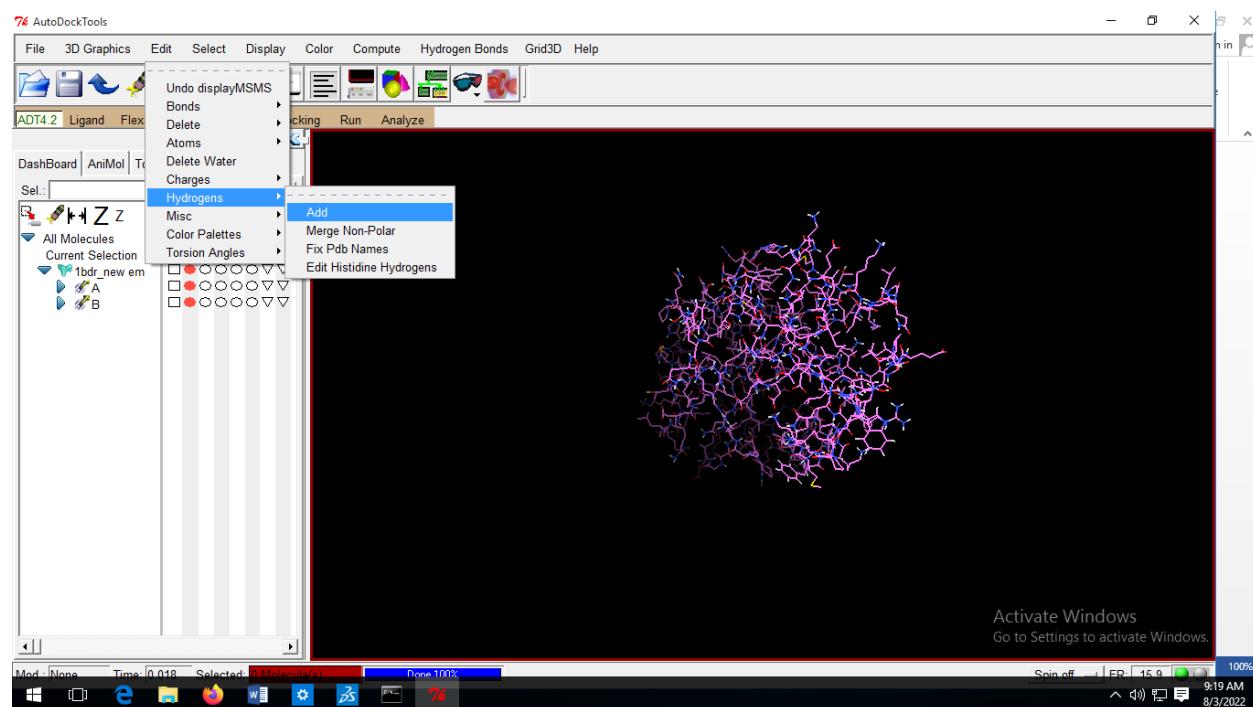
## BID 19006



## BID 19006

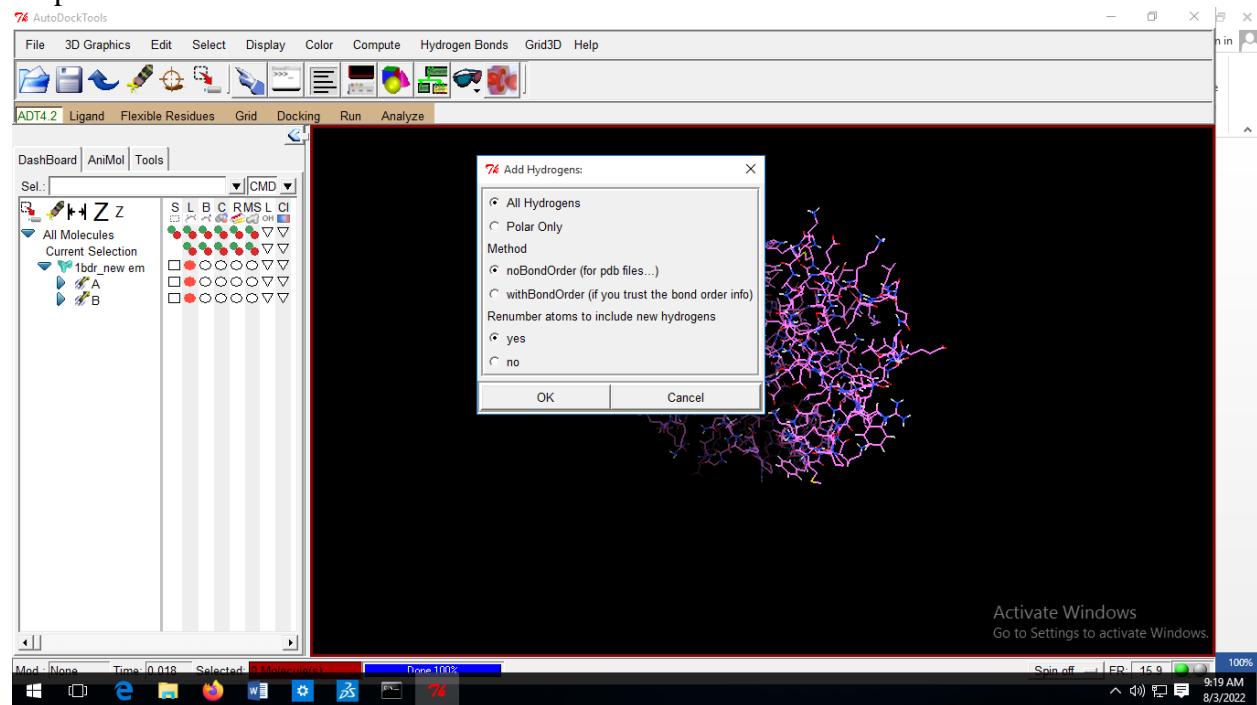


Step 11: Go to hydrogens, and click on add

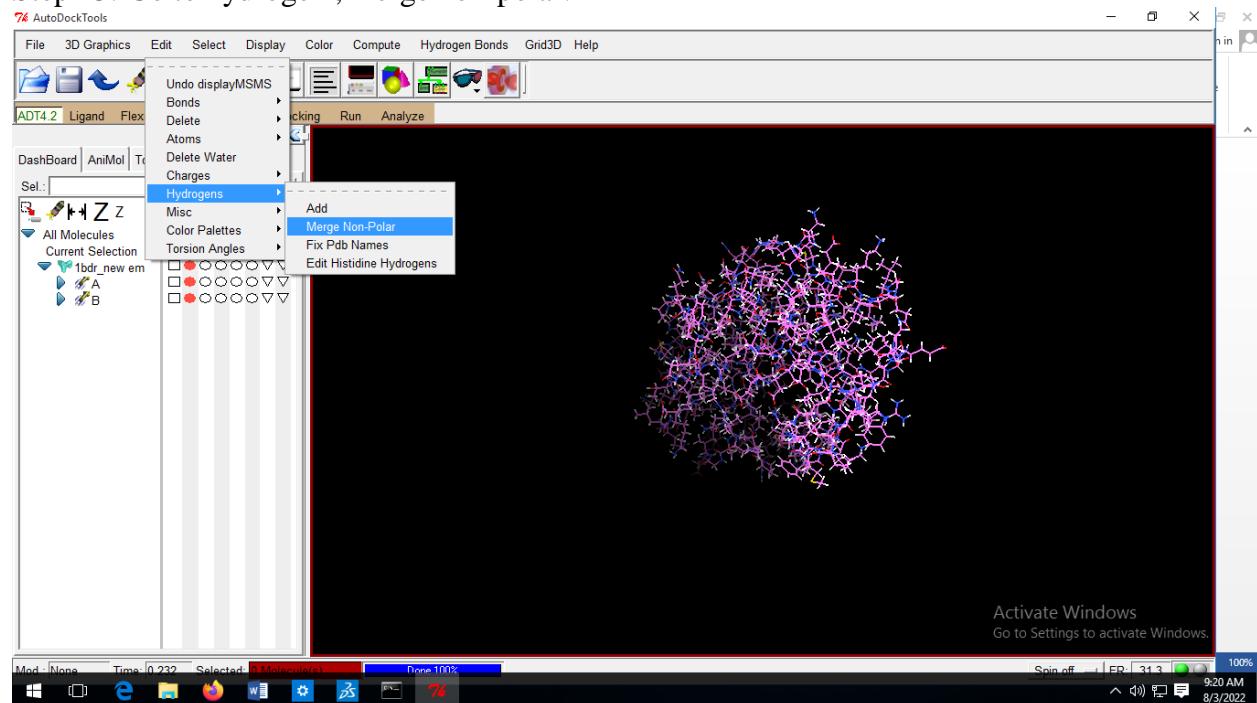


BID 19006

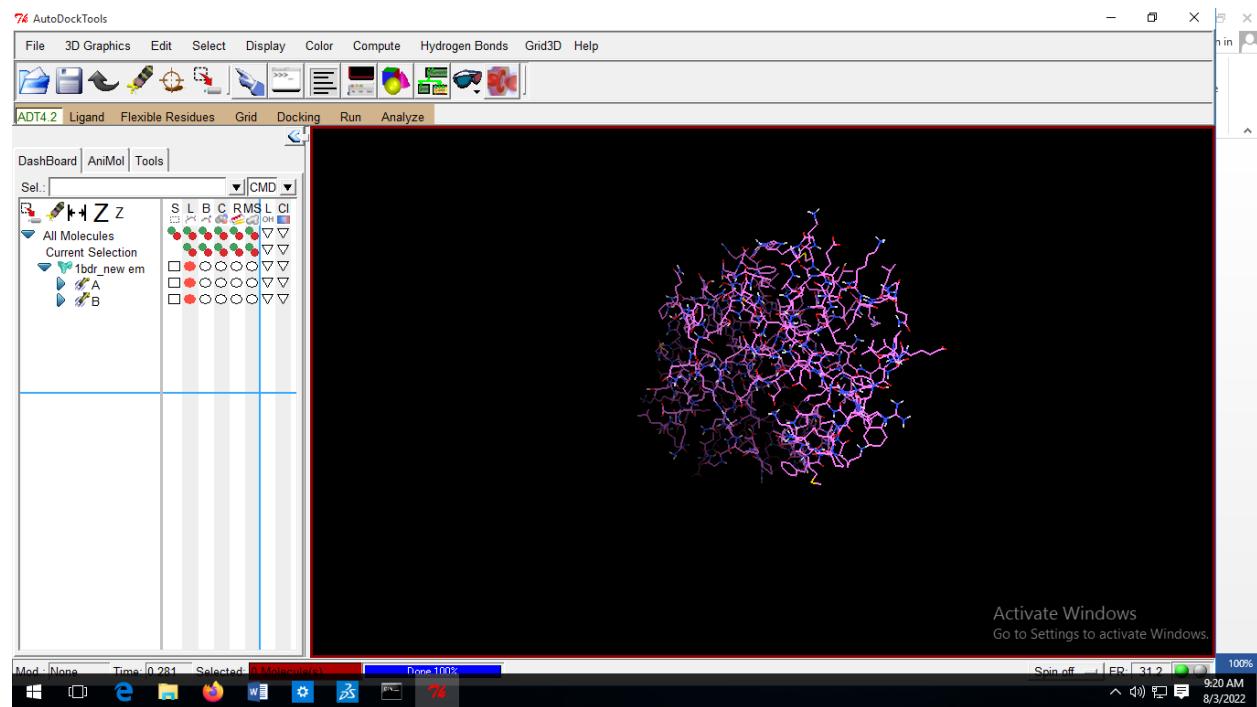
Step 12: Click on ok



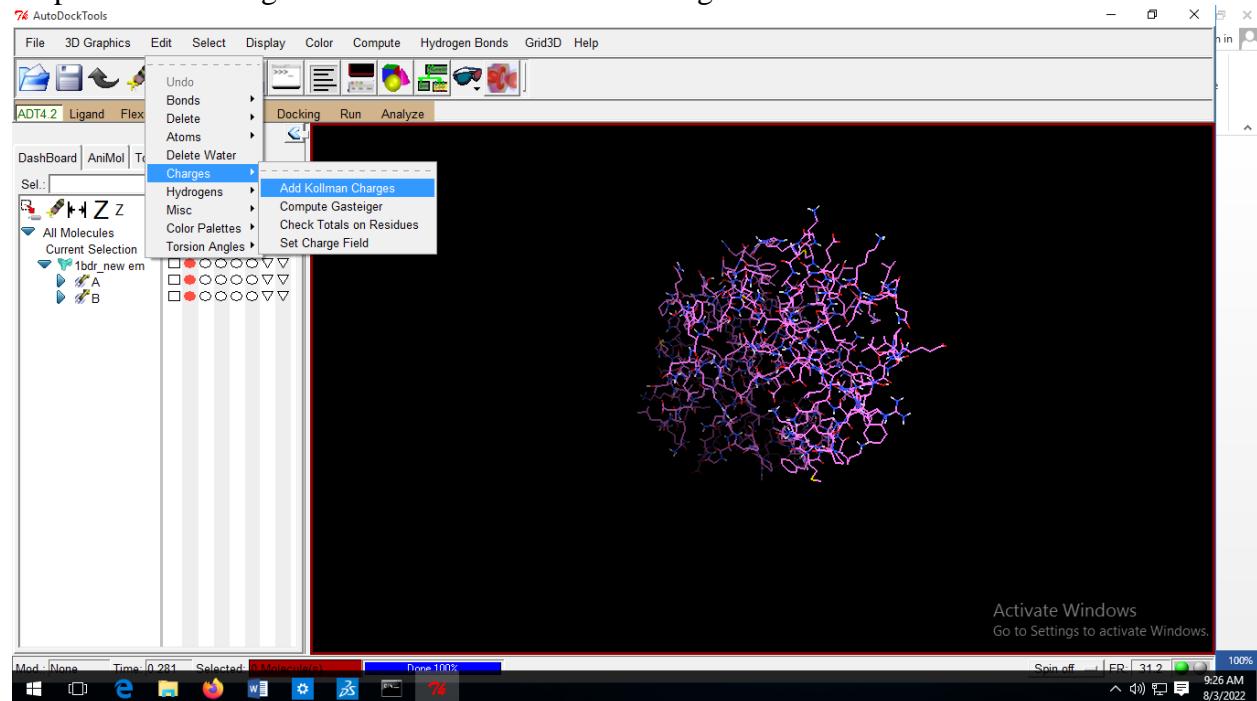
Step 13: Go to hydrogen , merge non-polar:



## BID 19006

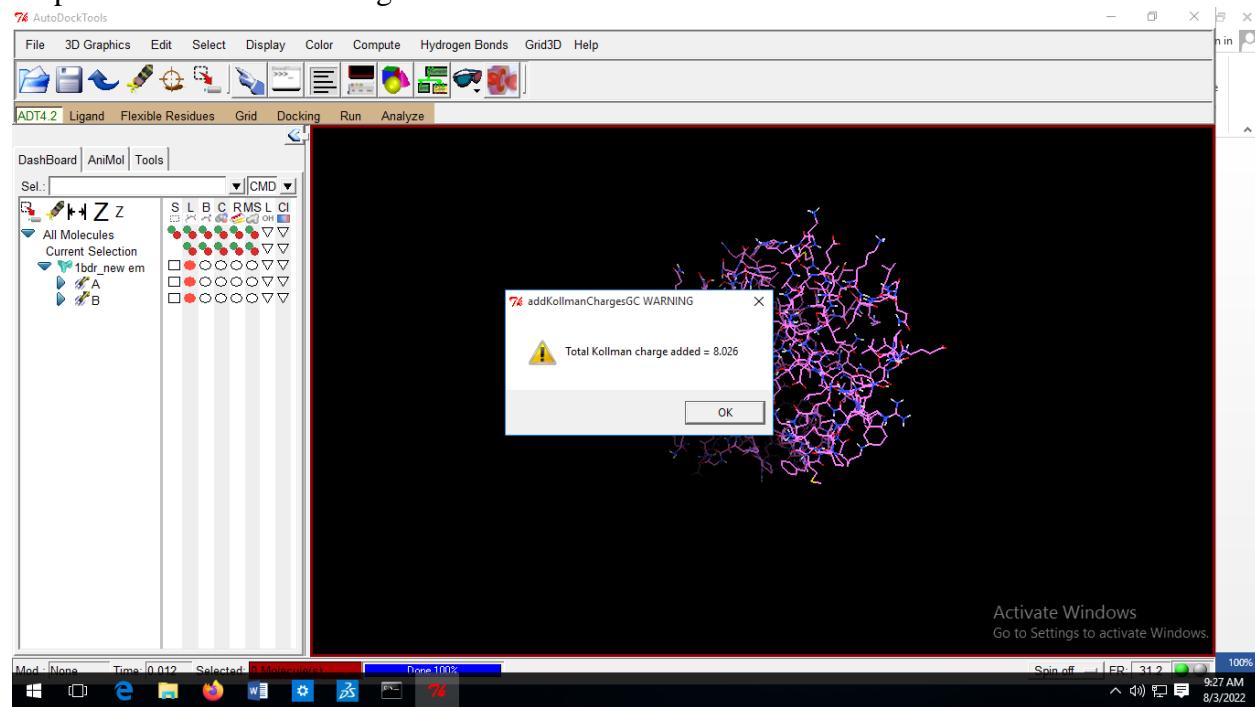


Step 14: Go to charges and click on add kollman charge:

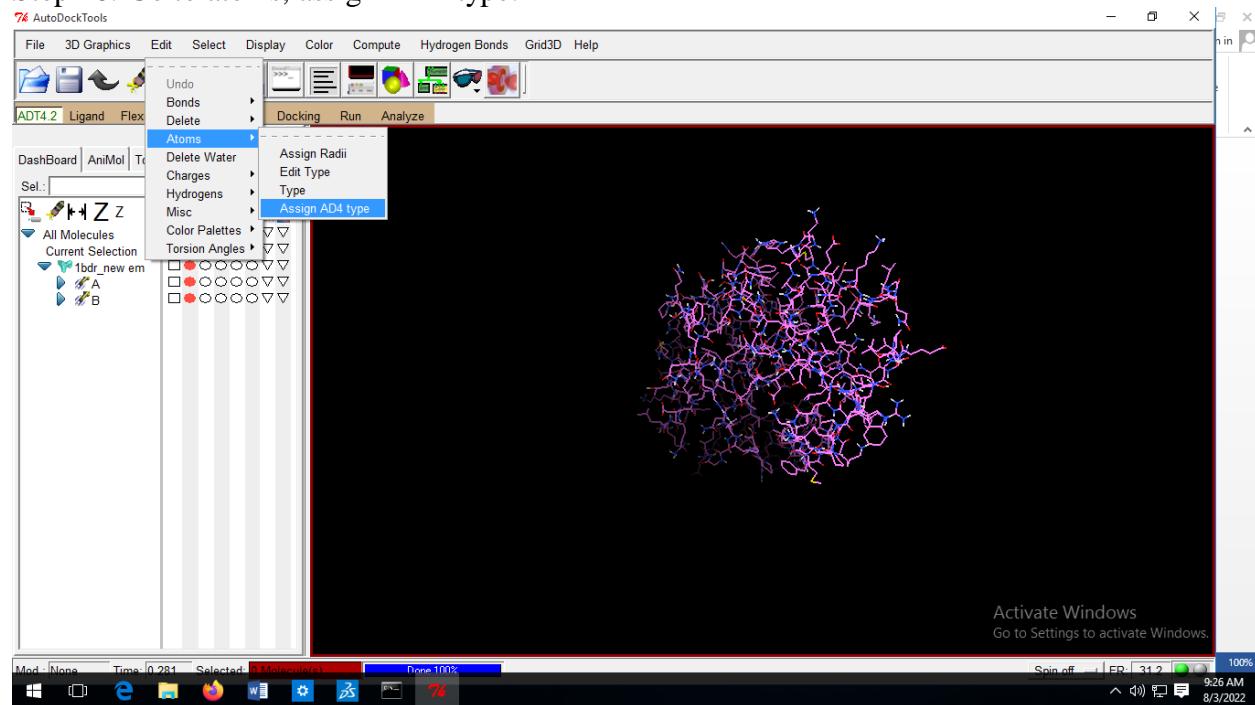


BID 19006

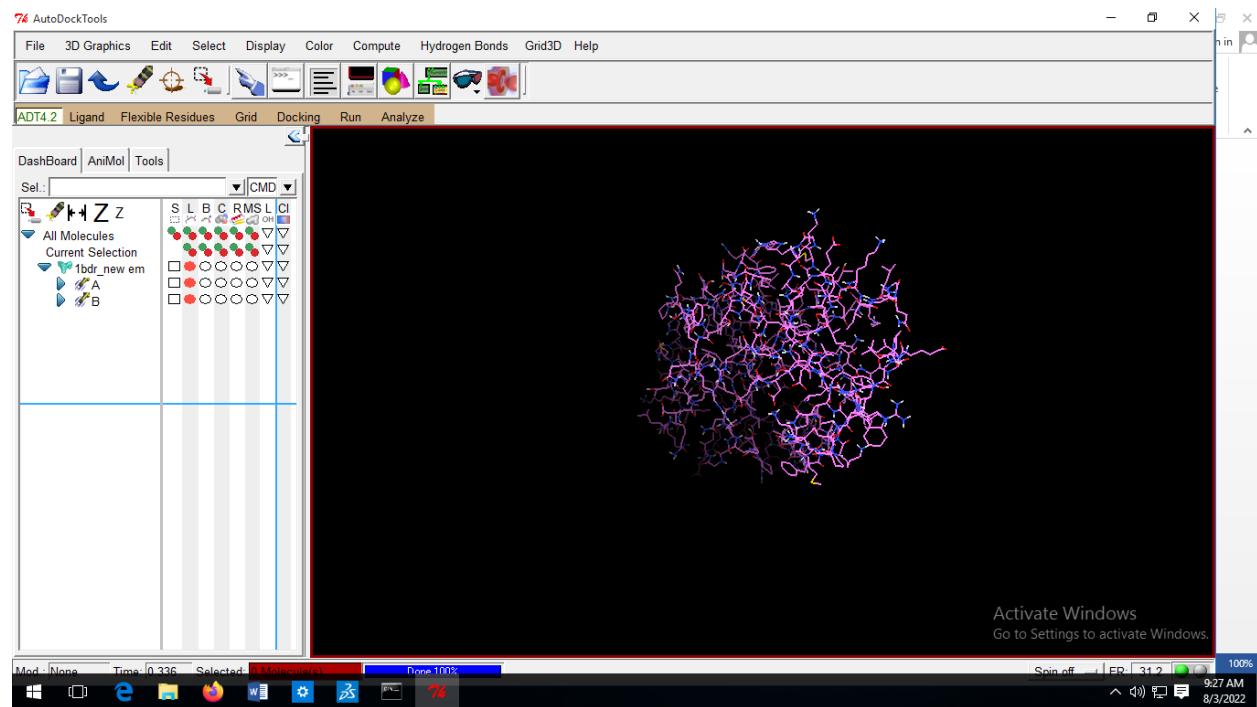
Step 15: Total Kollman charge added = 8.026



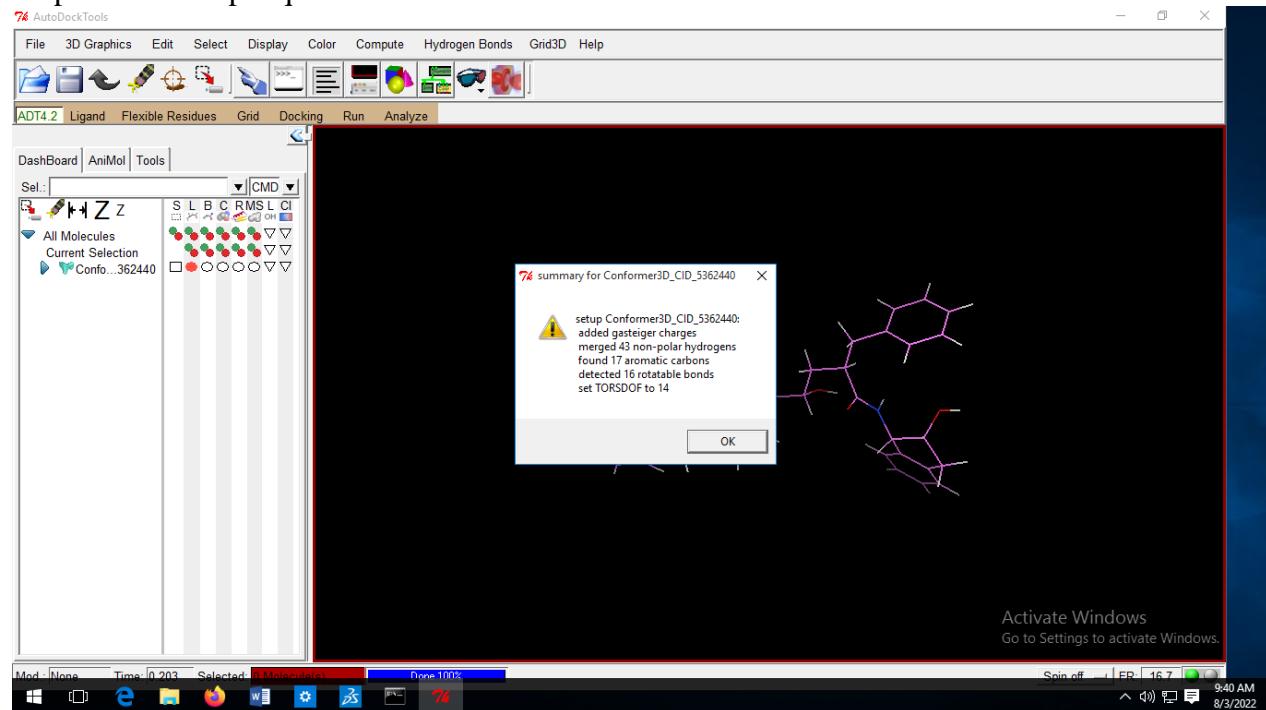
Step 16: Go to atoms, assign AD4 type:



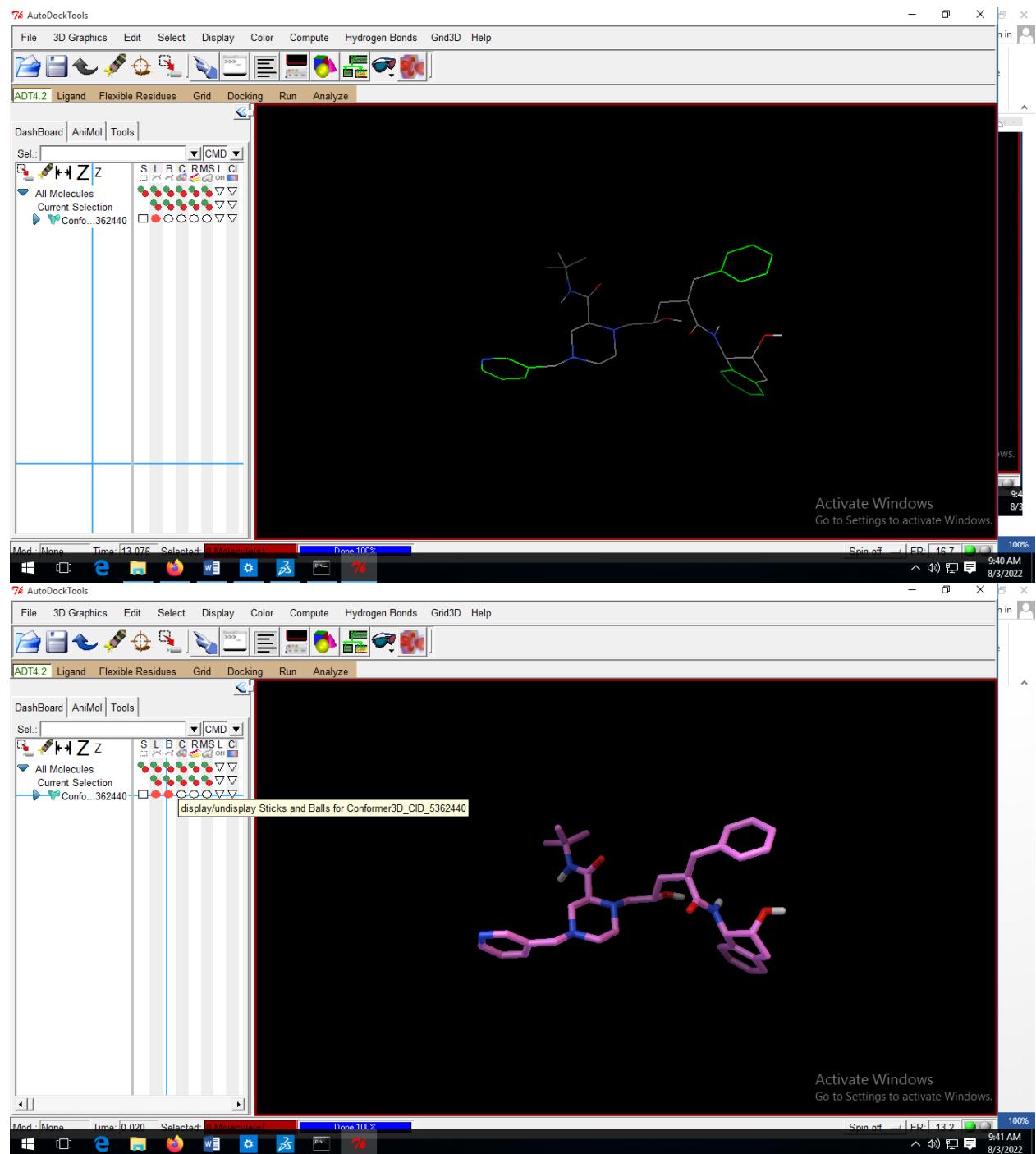
# BID 19006



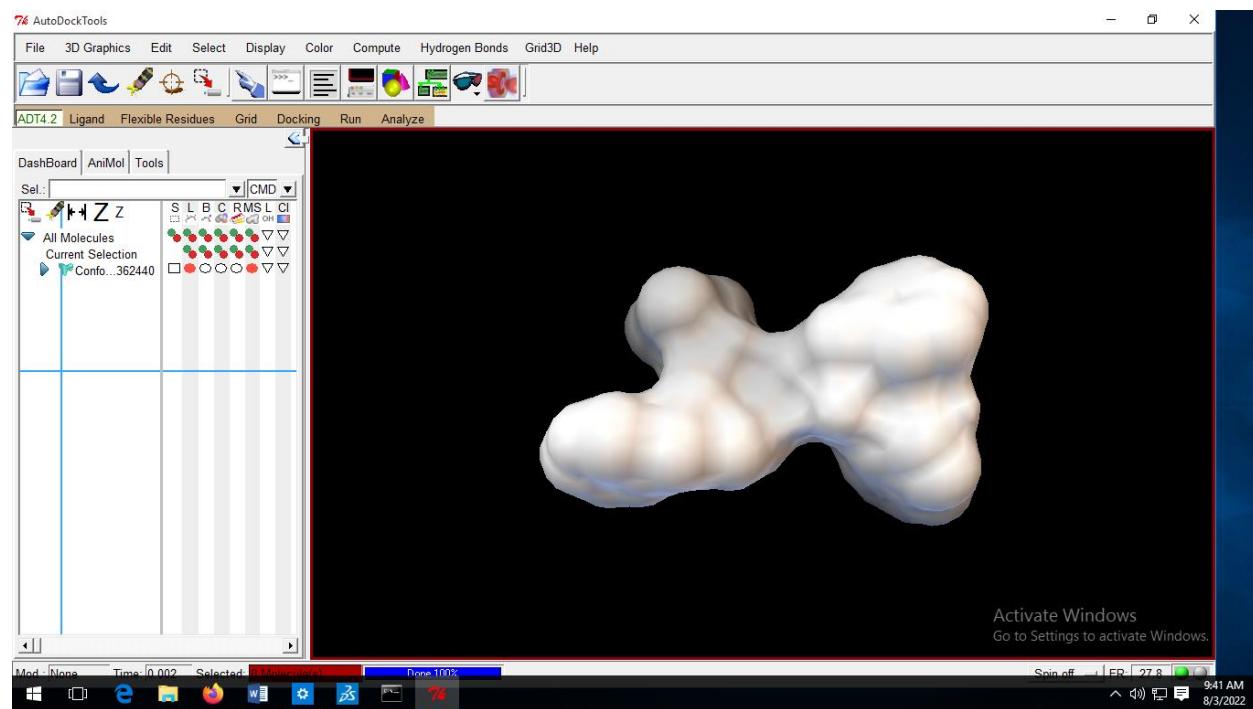
## Step 17: Save in pdbqt format



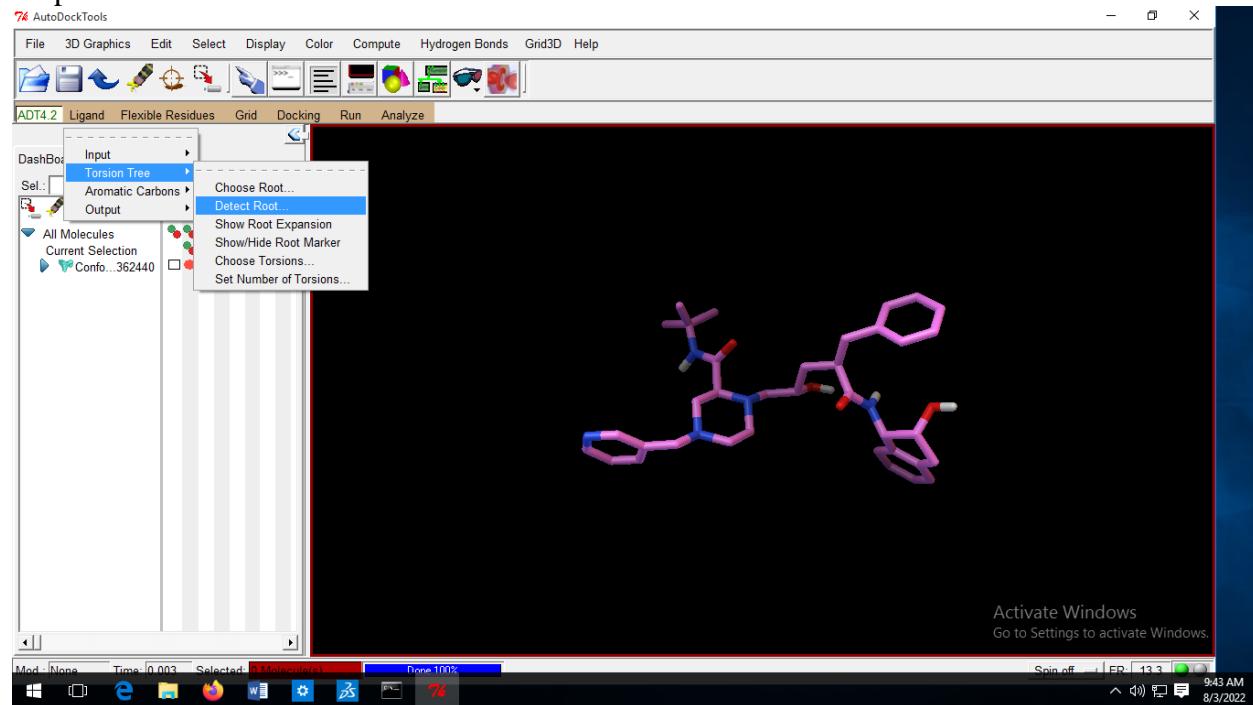
# BID 19006



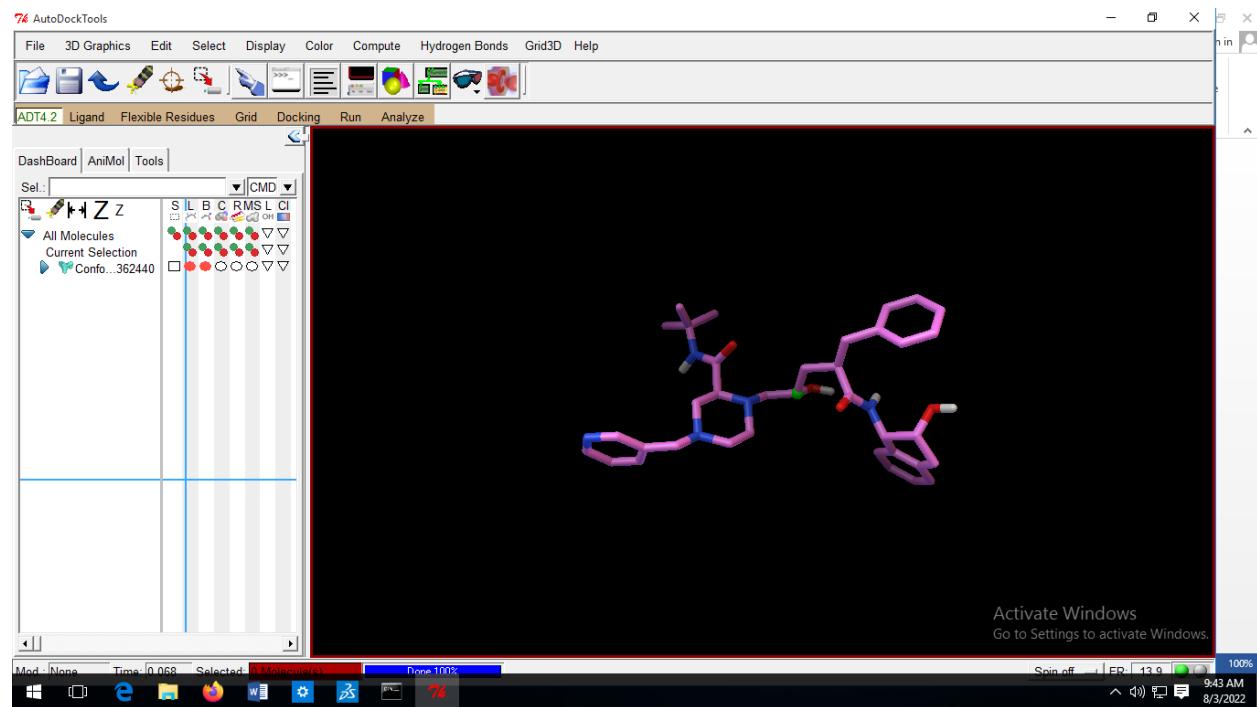
## BID 19006



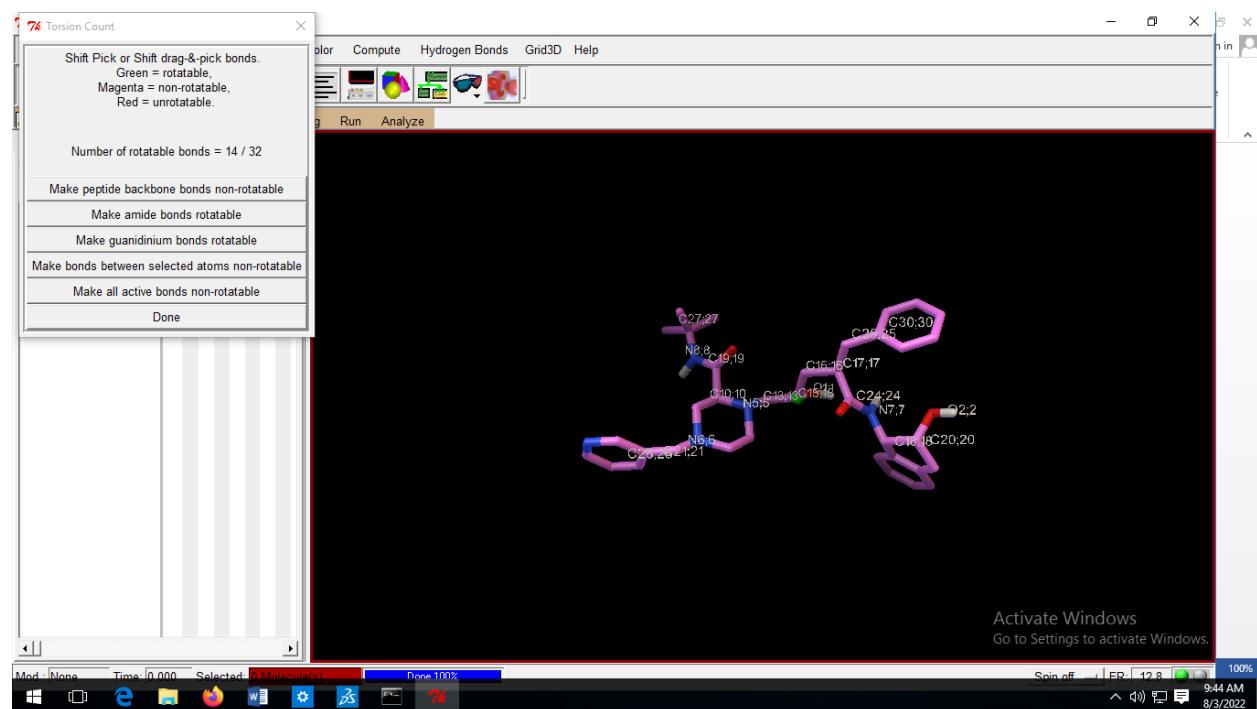
### Step 18: Go to Torsion Tree and detect root:



# BID 19006

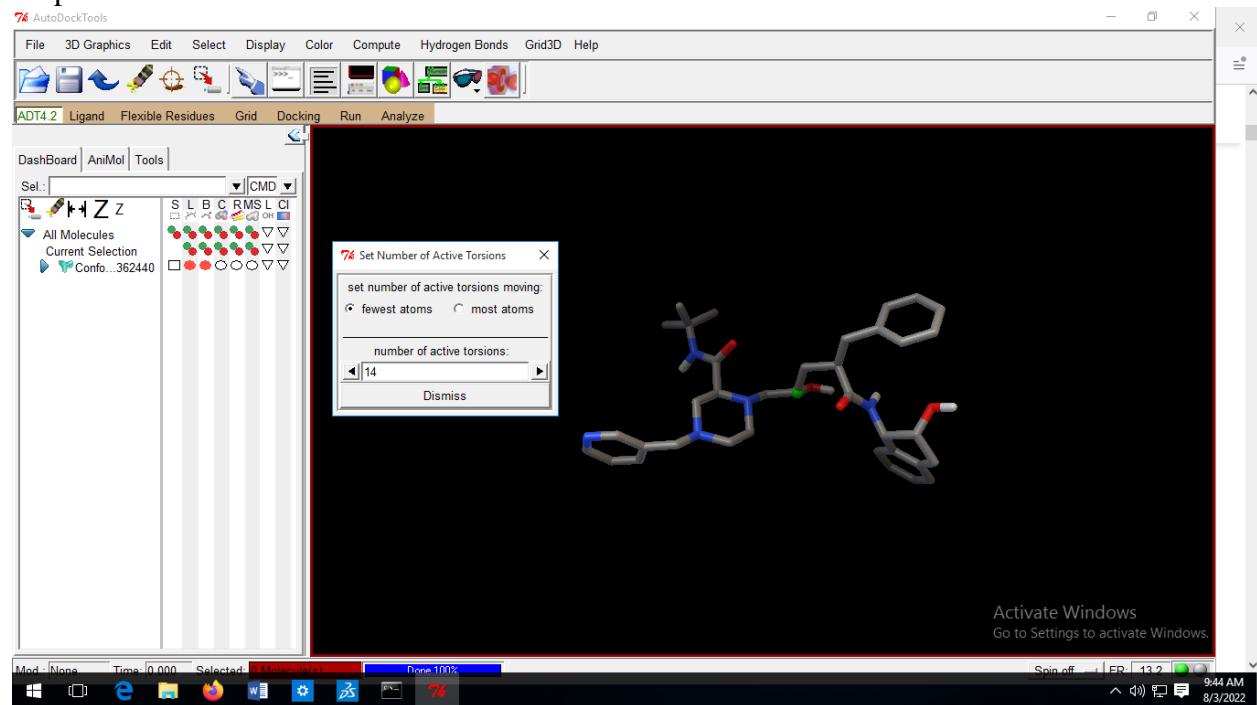


Step 19: Torsion shows no of rotatable bonds = 14/32

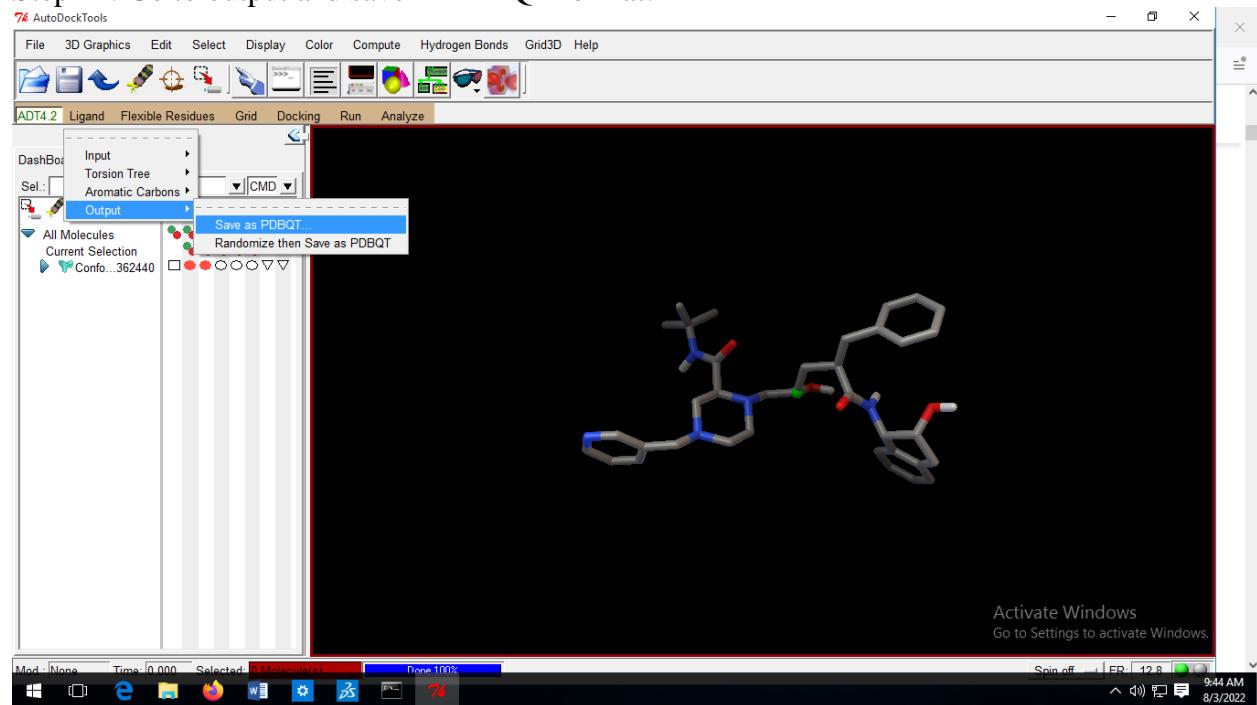


BID 19006

Step 20: Number of active torsions=14

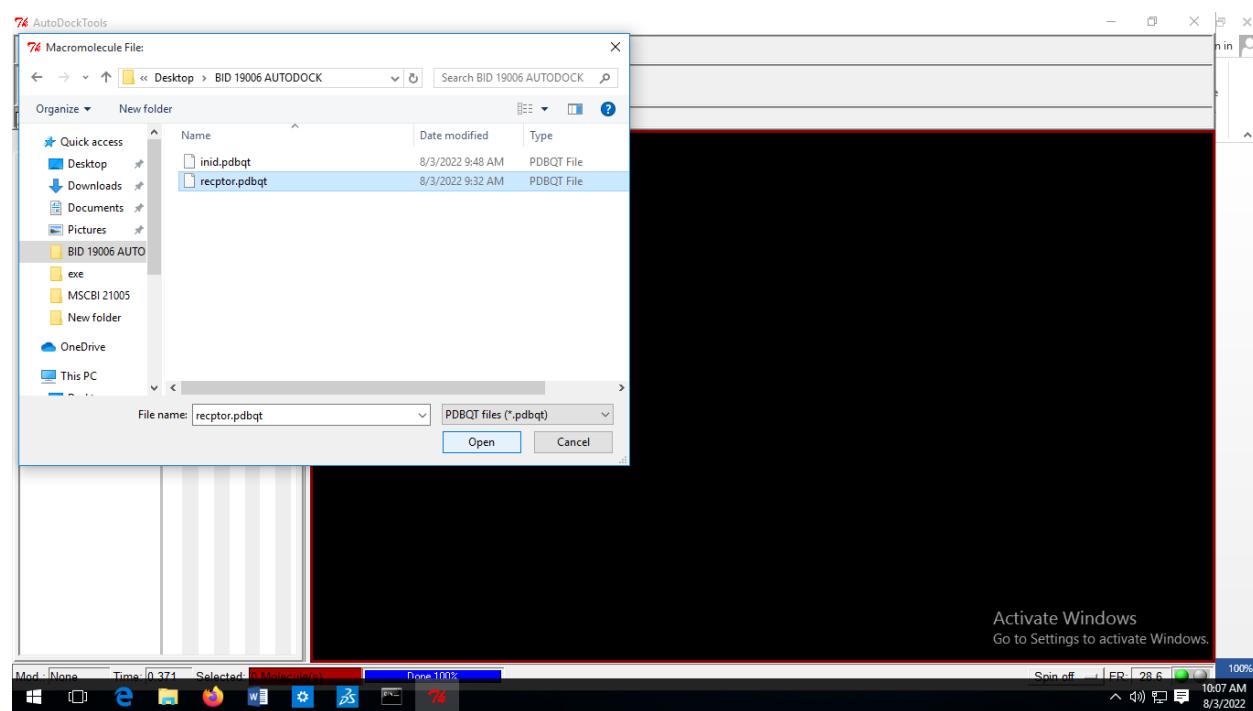
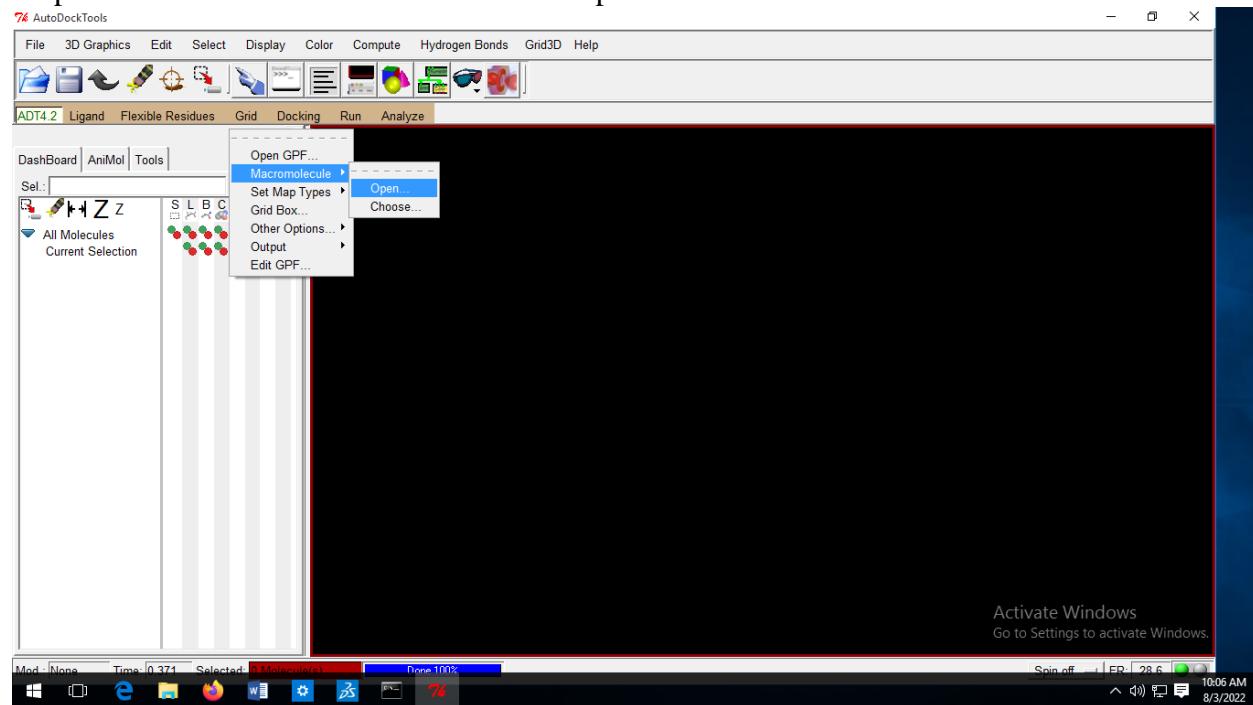


Step 21: Go to output and save in PDBQT format:



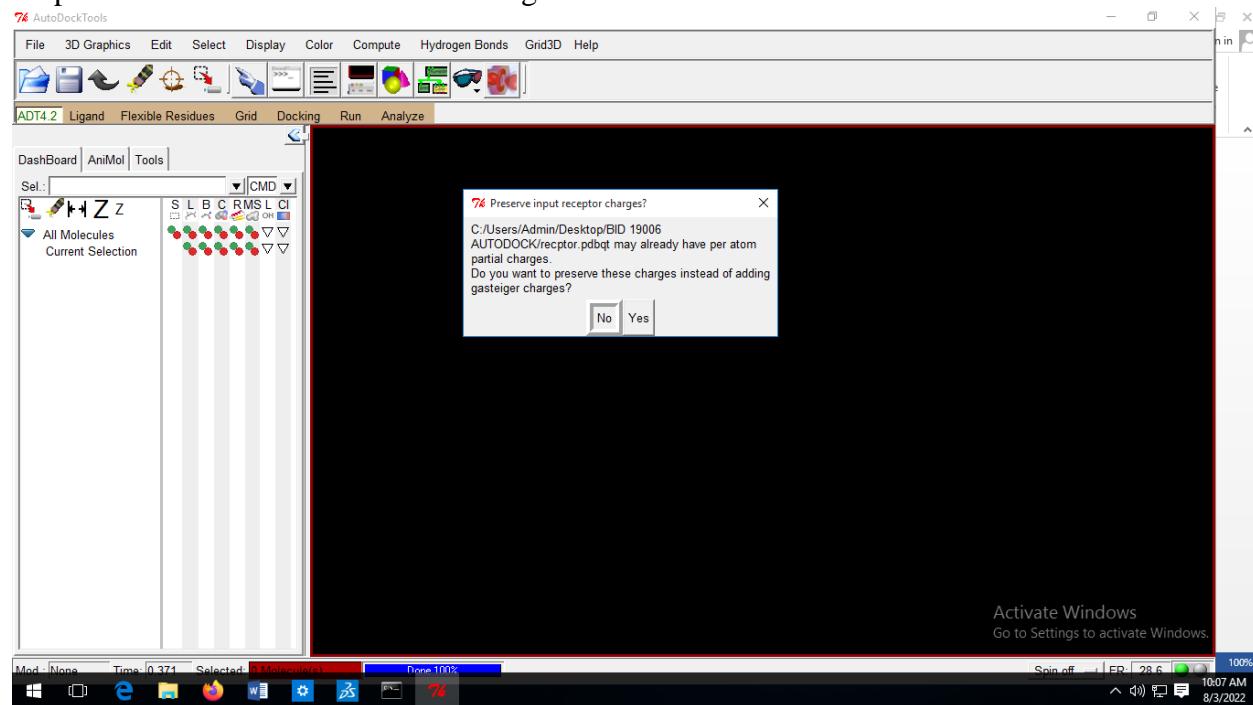
## BID 19006

Step 22: Go to Macromolecule and click on open:

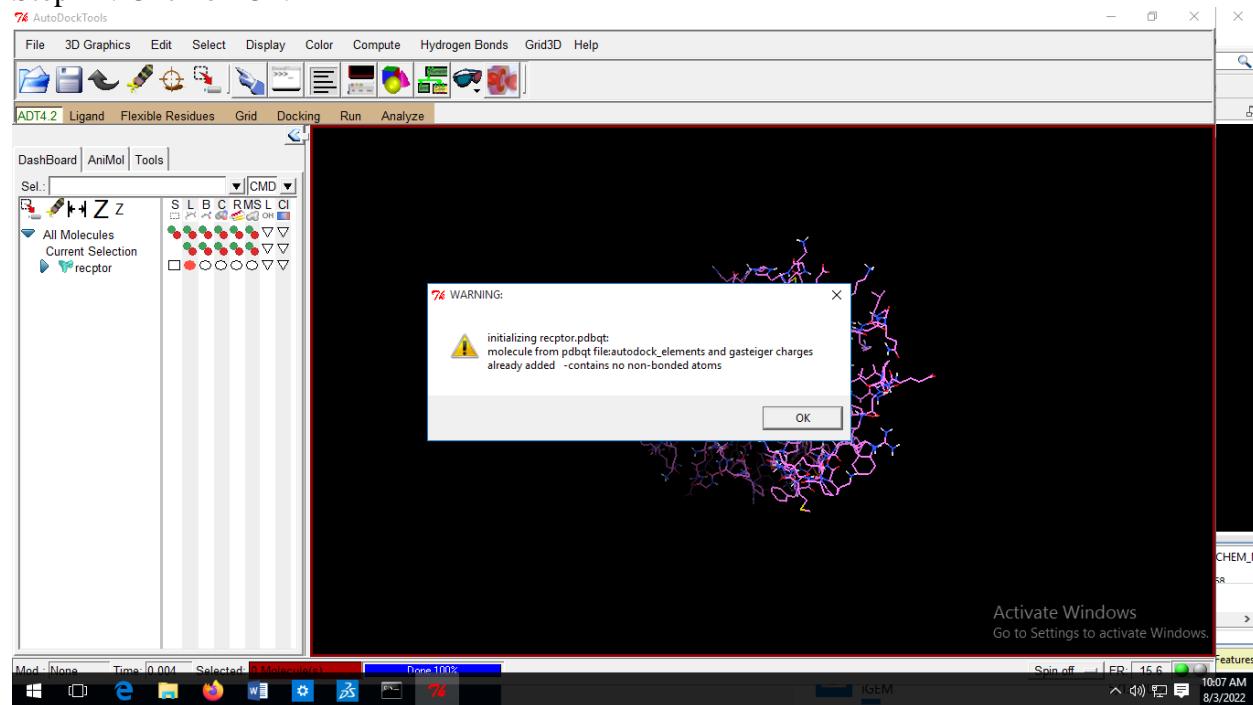


## BID 19006

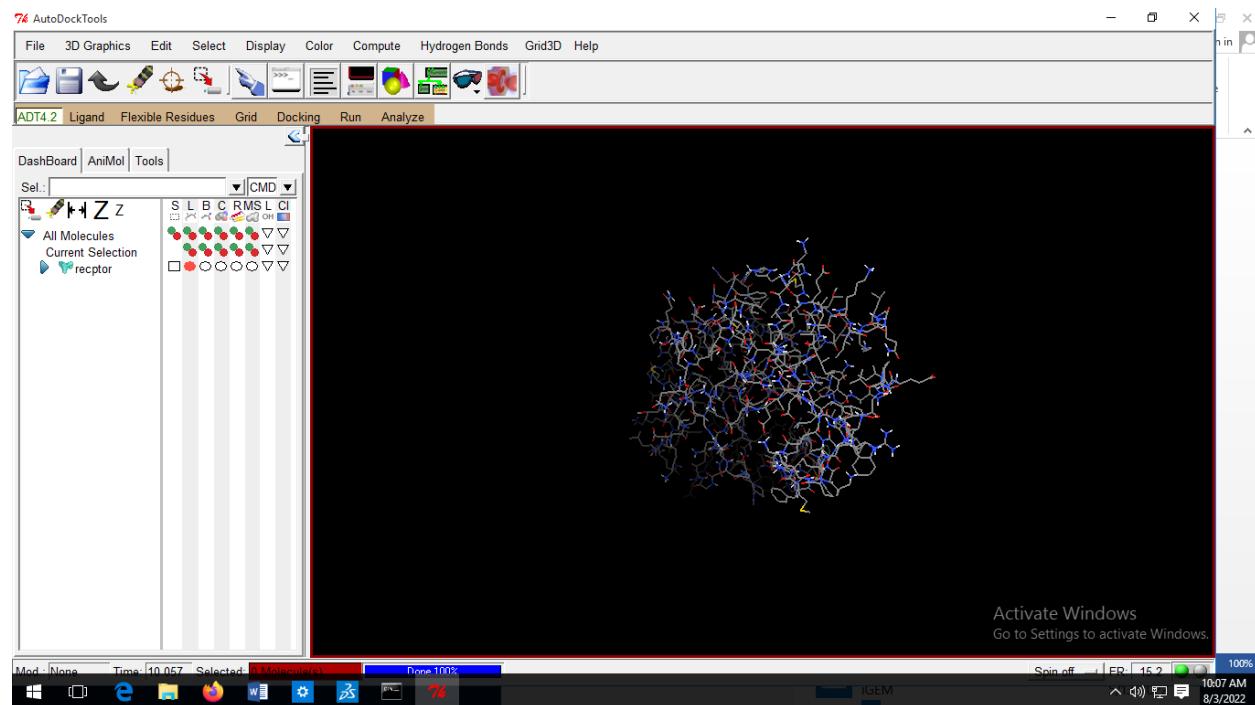
### Step 23: Click on Yes Preserve the charge amount



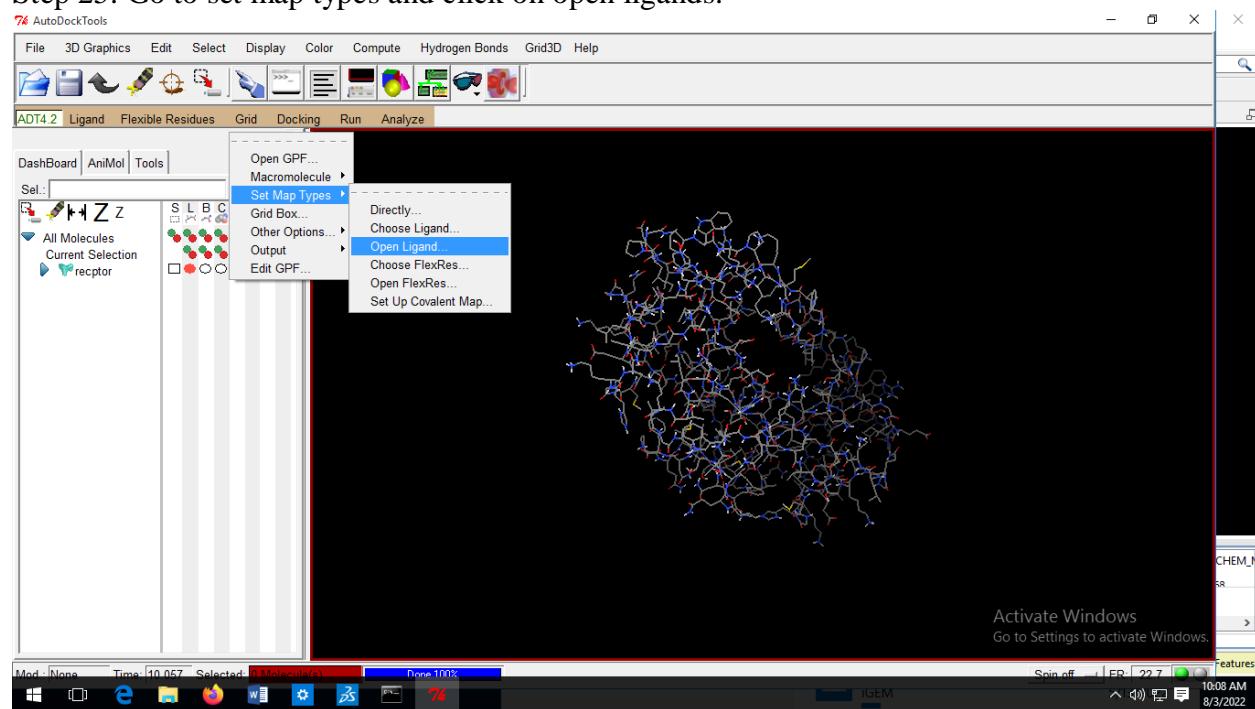
### Step 24: Click on Ok:



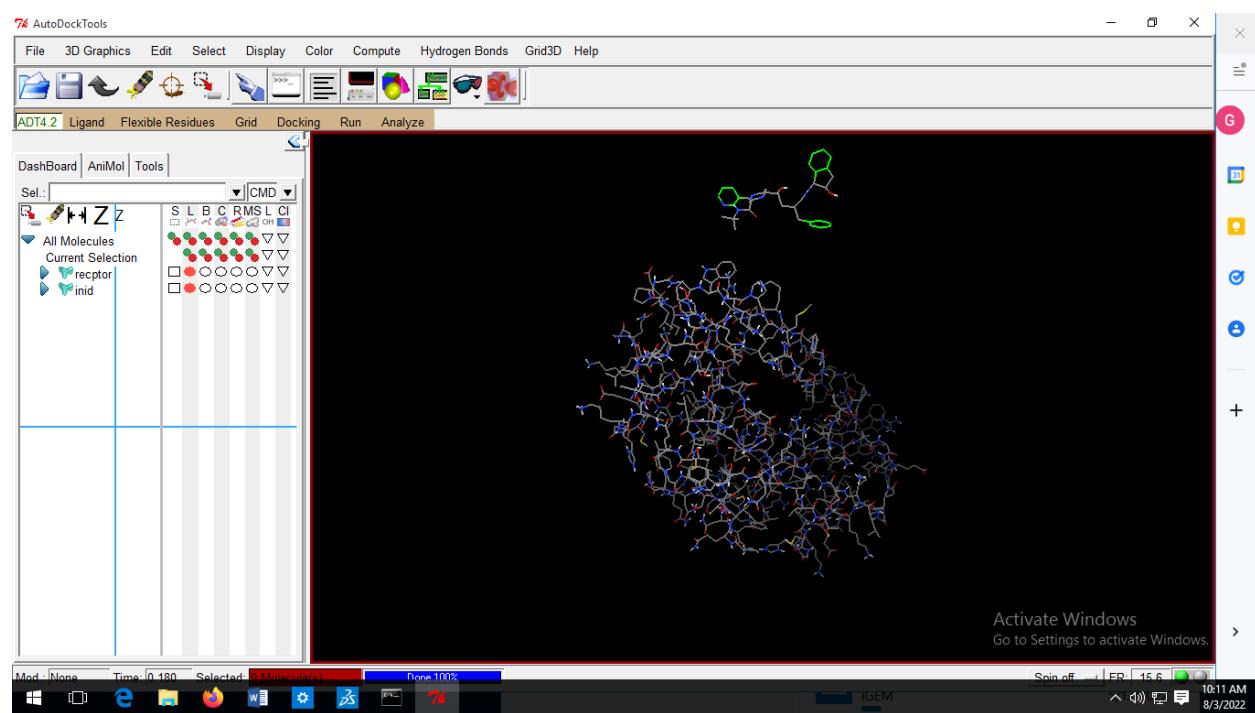
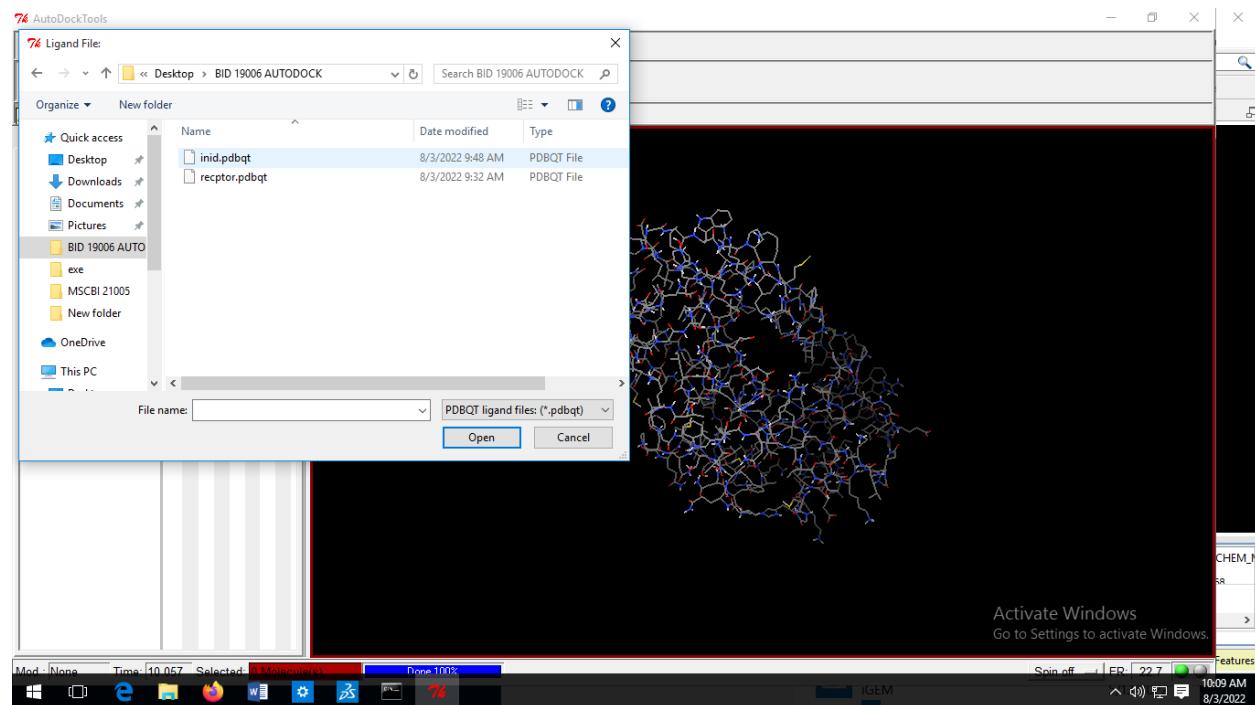
## BID 19006



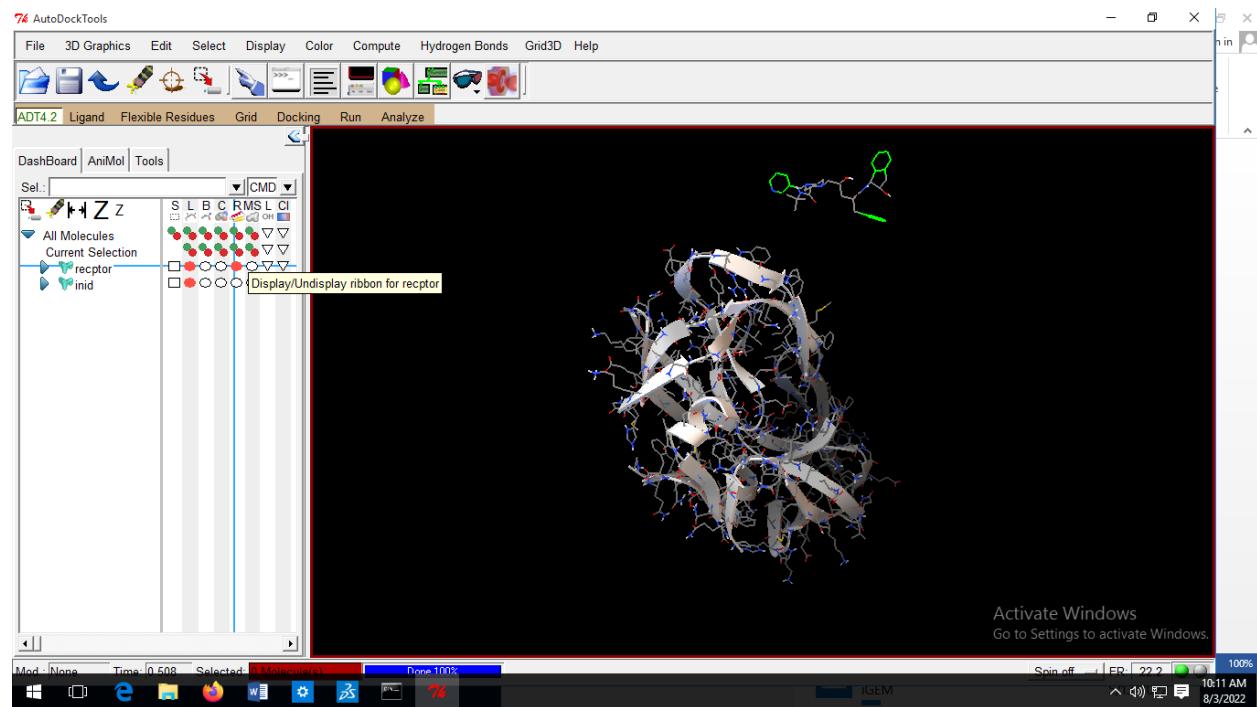
Step 25: Go to set map types and click on open ligands:



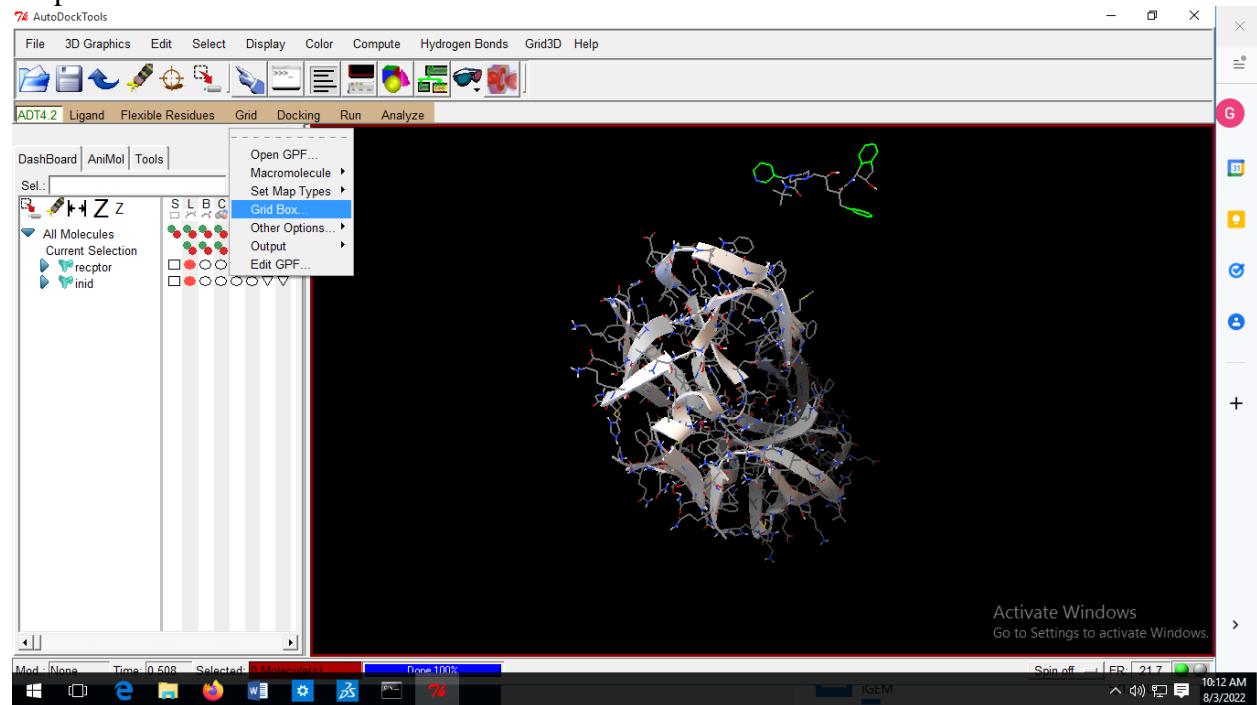
## BID 19006



## BID 19006

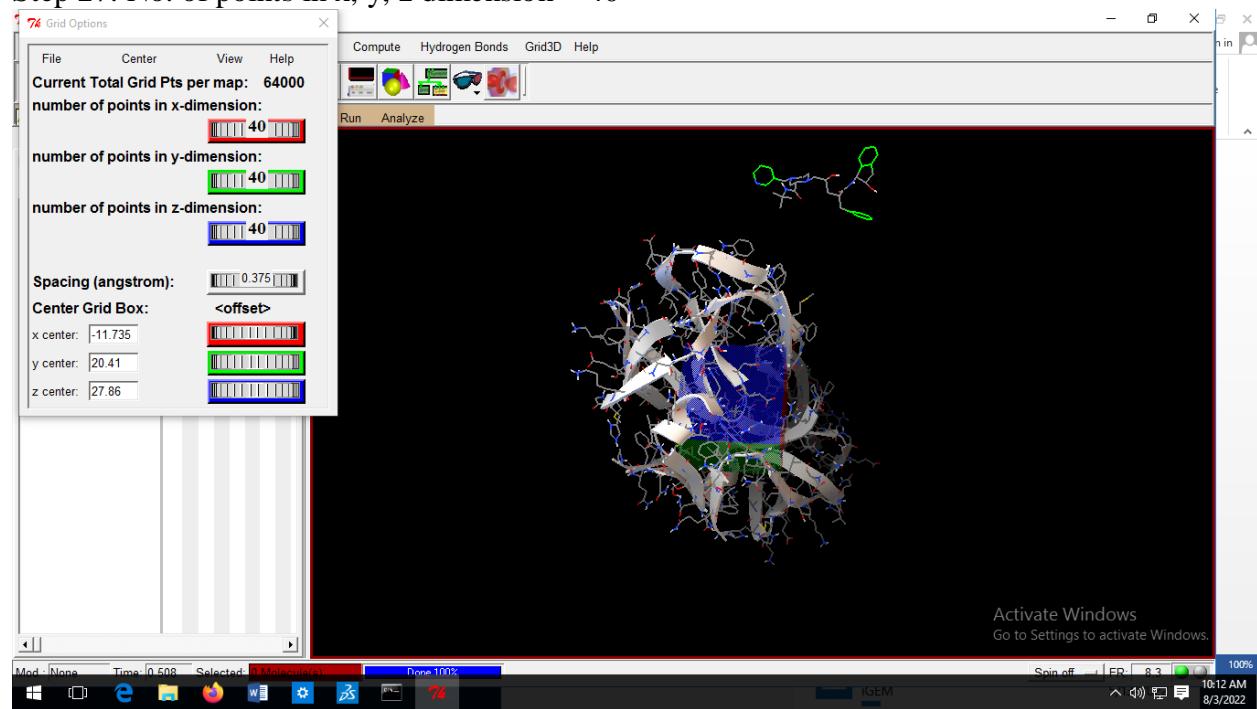


### Step 26: Go to Grid box :

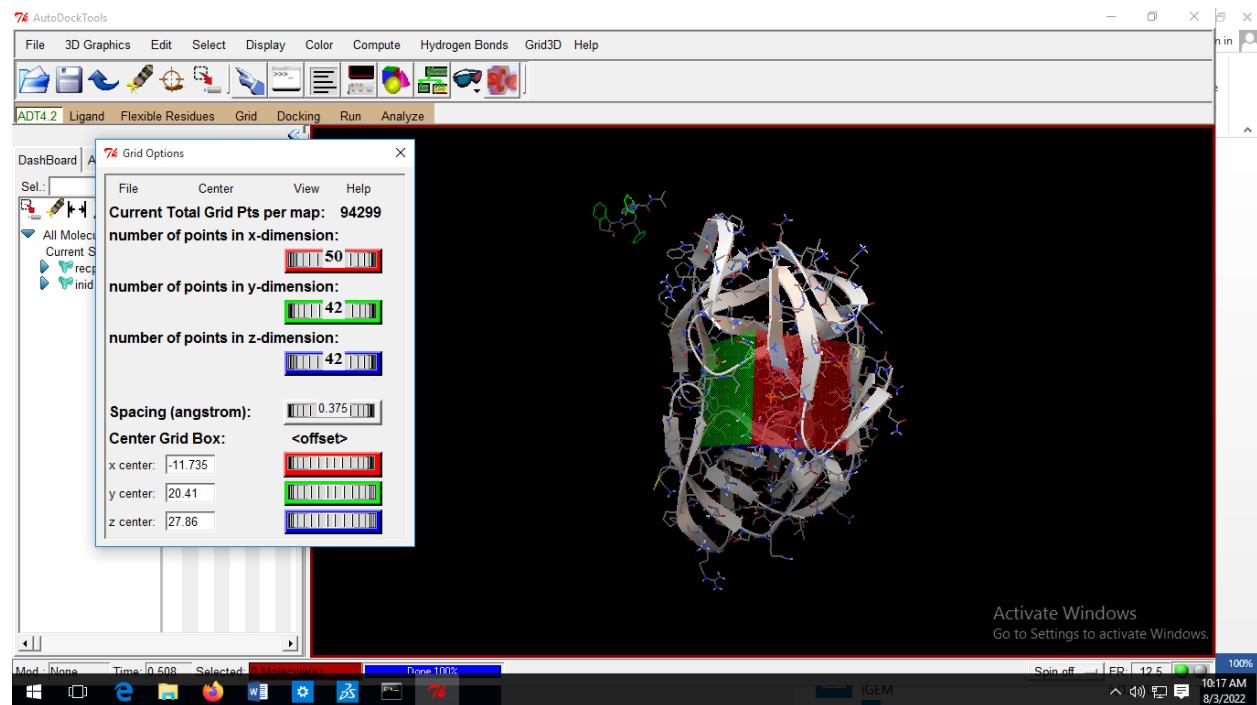


BID 19006

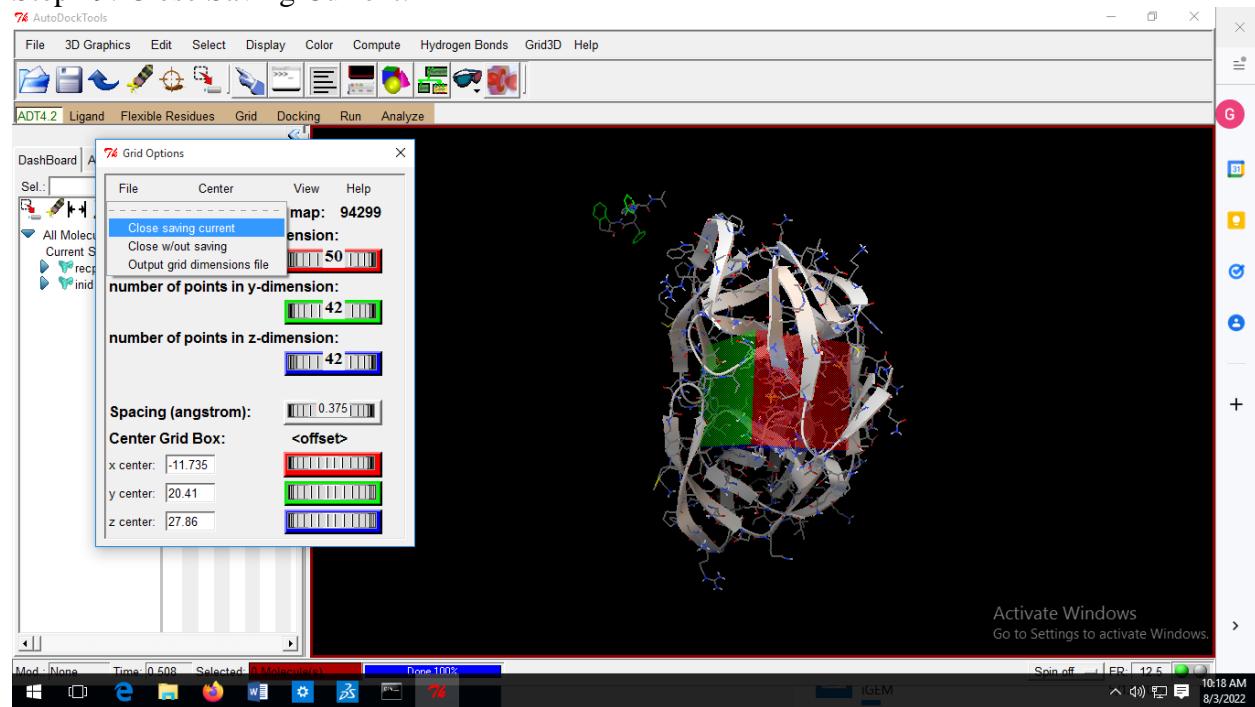
Step 27: No. of points in x, y, z dimension = 40



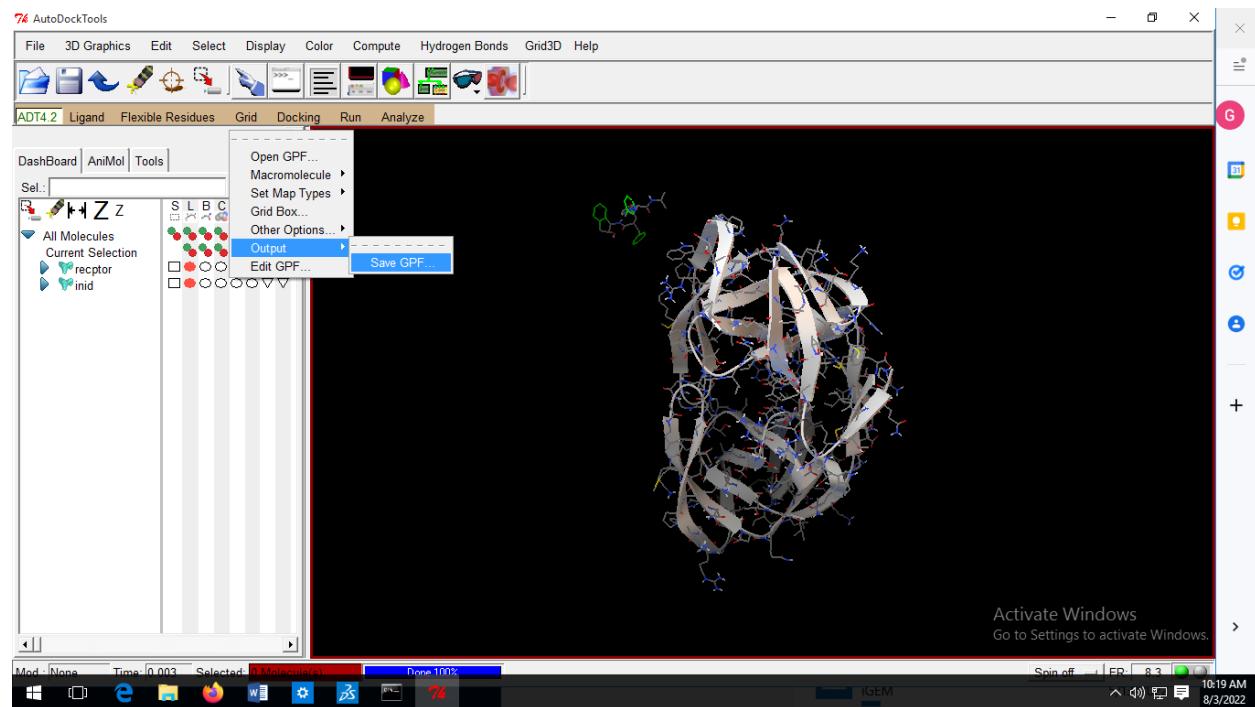
Step 28: No. of points in x dimension = 50,  
No. of points in y dimension = 42,  
No. of points in z dimension = 42,



Step 29: Close Saving Current:

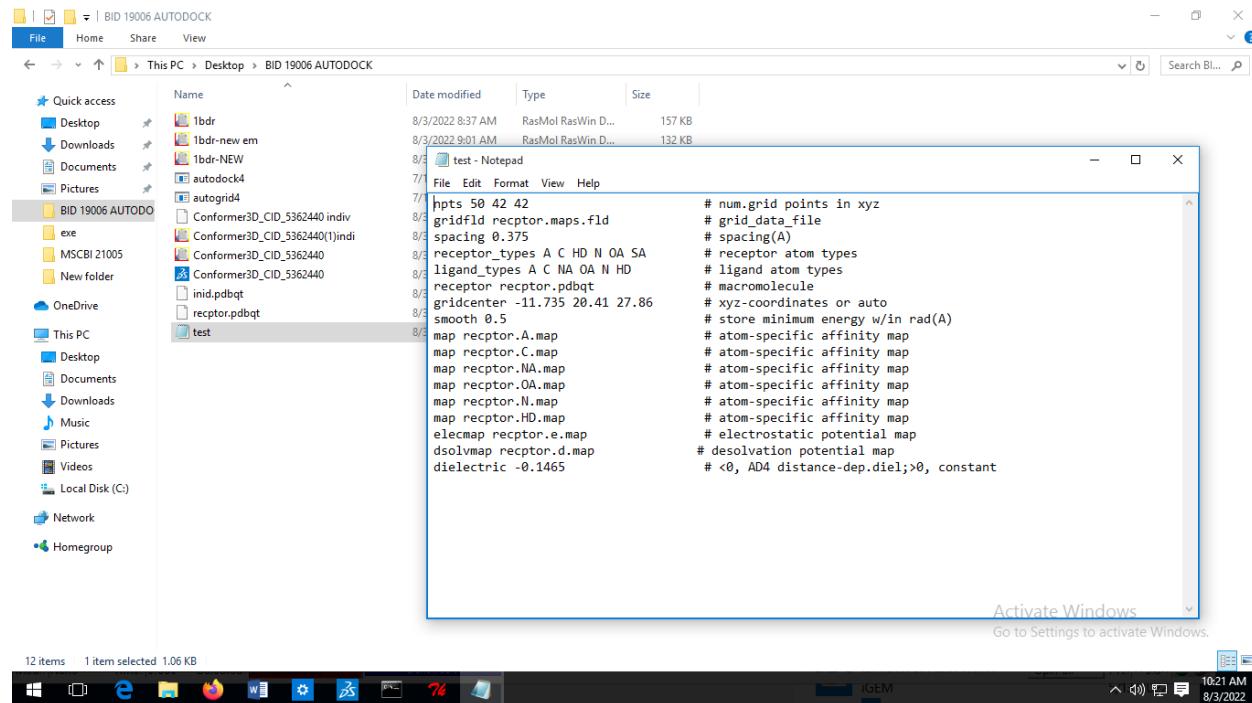


Step 30: Click on output, save GPF= Grid parameter file



## BID 19006

Step 31: Open in notepad  
Map - mapping files  
dsolvmap receptors- solvation parameters  
Dielectric- charge distribution



## Step 32: Command Prompt:

```
C:\Windows\system32\cmd.exe - autogrid4.exe -p test.gpf -l text.glg
'BID' is not recognized as an internal or external command,
operable program or batch file.

C:\Users\Admin\Desktop>cd BID 19006 AUTODOCK
C:\Users\Admin\Desktop\BID 19006 AUTODOCK>dir
 Volume in drive C has no label.
 Volume Serial Number is DC13-E7CC

 Directory of C:\Users\Admin\Desktop\BID 19006 AUTODOCK

08/03/2022  10:20 AM    <DIR>        .
08/03/2022  10:20 AM    <DIR>        ..
08/03/2022  09:01 AM        134,193 1bdr-new.em.pdb
08/03/2022  08:54 AM        121,711 1bdr-NEW.pdb
08/03/2022  08:37 AM        160,704 1bdr.pdb
07/19/2014  04:23 AM        619,379 autodock4.exe
07/19/2014  04:23 AM        1,927,292 autogrid4.exe
08/03/2022  09:45 AM        5,431 Conformer3D_CID_5362440.indv
08/03/2022  09:26 AM        7,540 Conformer3D_CID_5362440(1).indi.pdb
08/03/2022  09:07 AM        7,540 Conformer3D_CID_5362440.pdb
08/03/2022  09:05 AM        11,434 Conformer3D_CID_5362440.sdf
08/03/2022  09:48 AM        4,014 inid.pdbqt
08/03/2022  09:32 AM        148,495 receptor.pdbqt
08/03/2022  10:20 AM        1,092 test.gpf
                           12 File(s)     3,148,825 bytes
                           2 Dir(s)    73,567,358,976 bytes free

C:\Users\Admin\Desktop\BID 19006 AUTODOCK>autogrid4.exe -p test.gpf -l test.glg
autogrid4.exe: unknown switch -1

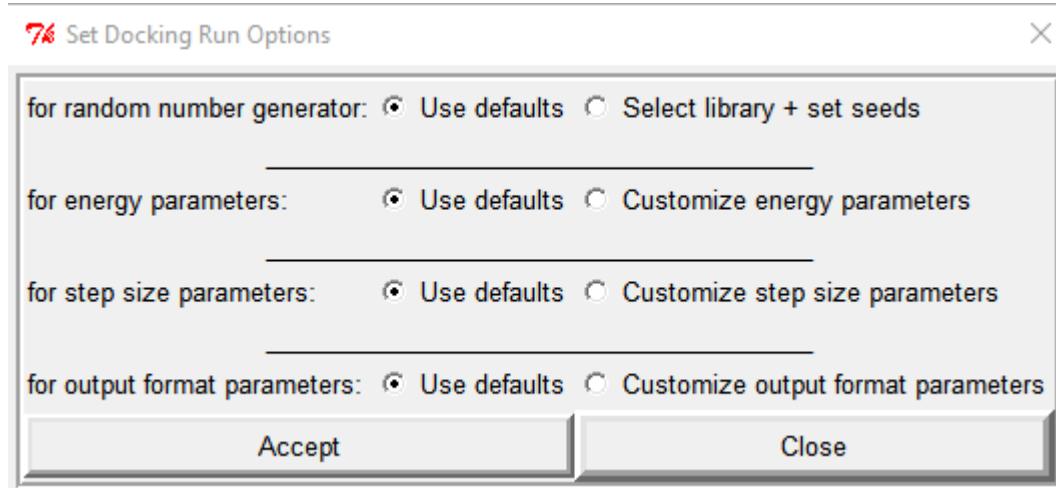
C:\Users\Admin\Desktop\BID 19006 AUTODOCK>autogrid4.exe -p test.gpf -l text.glg
-
```

## BID 19006

Step 33: For docking the protein and ligand go to docking go to macromolecule and than we have to set rigid filename so just select the protein name and open it.

Now go to docking ligand and just choose, select the ligand.

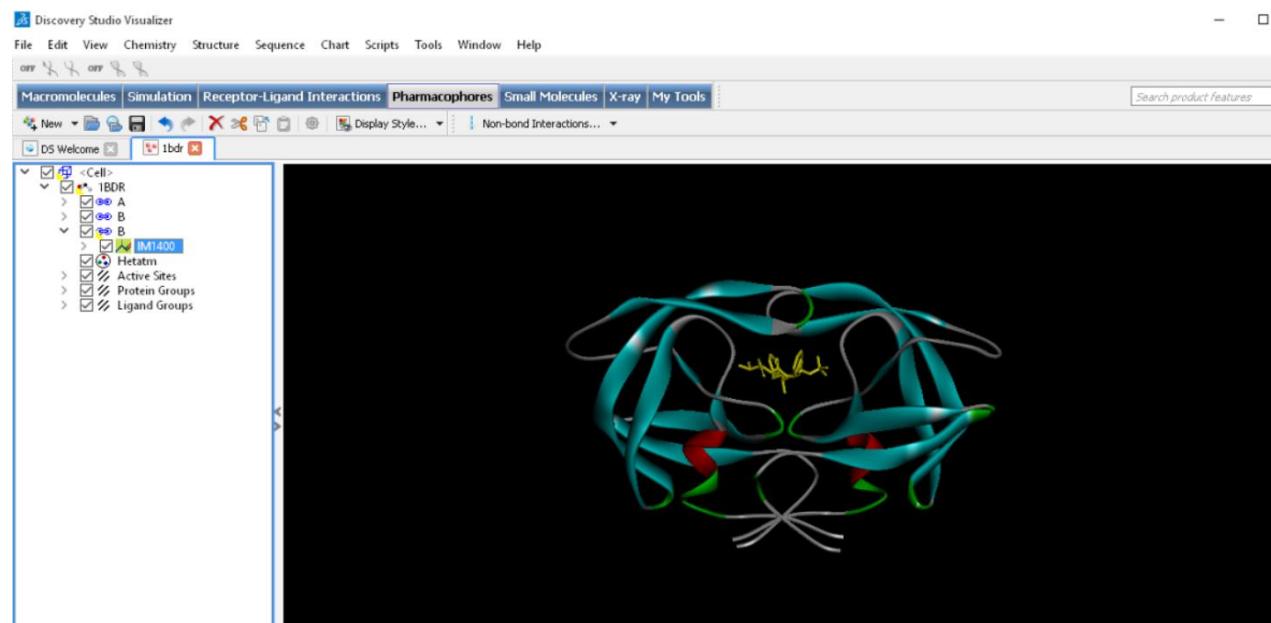
Now go to docking and just set parameters and go to genetic algorithm and keep default parameters and now click on OK.



## Step 34:

Now go to docking and than go to docking parametaers click on accept.

Now, go to Docking , click on output than go to Lamarckian GA and you will now save the file and click on OK.



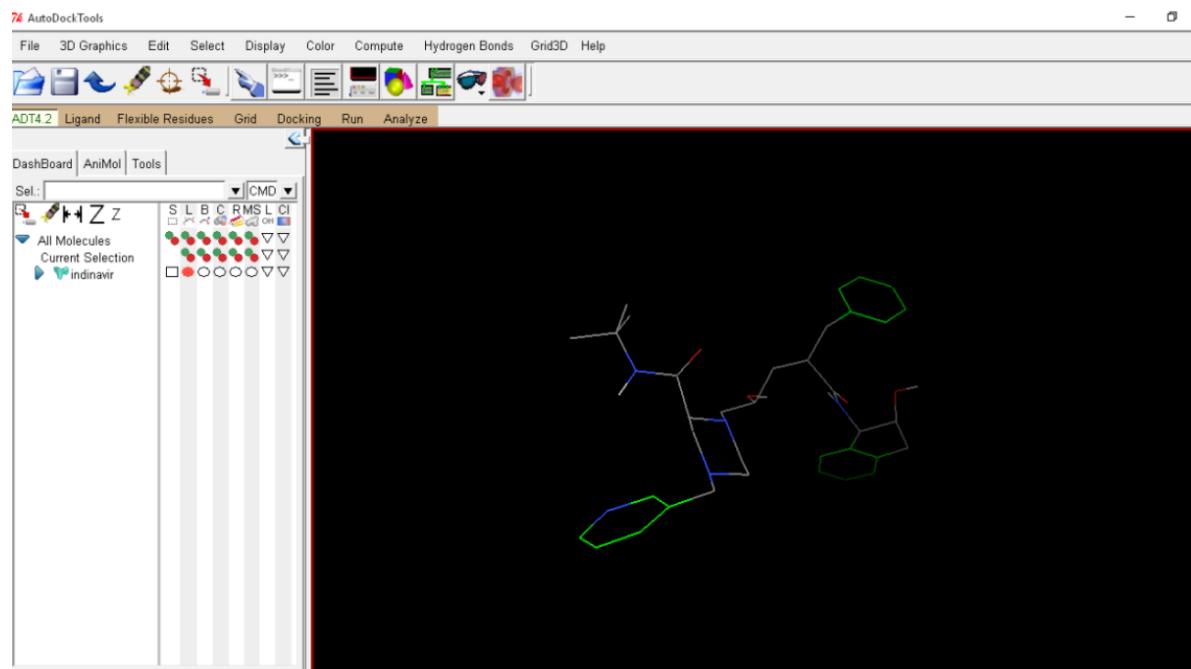
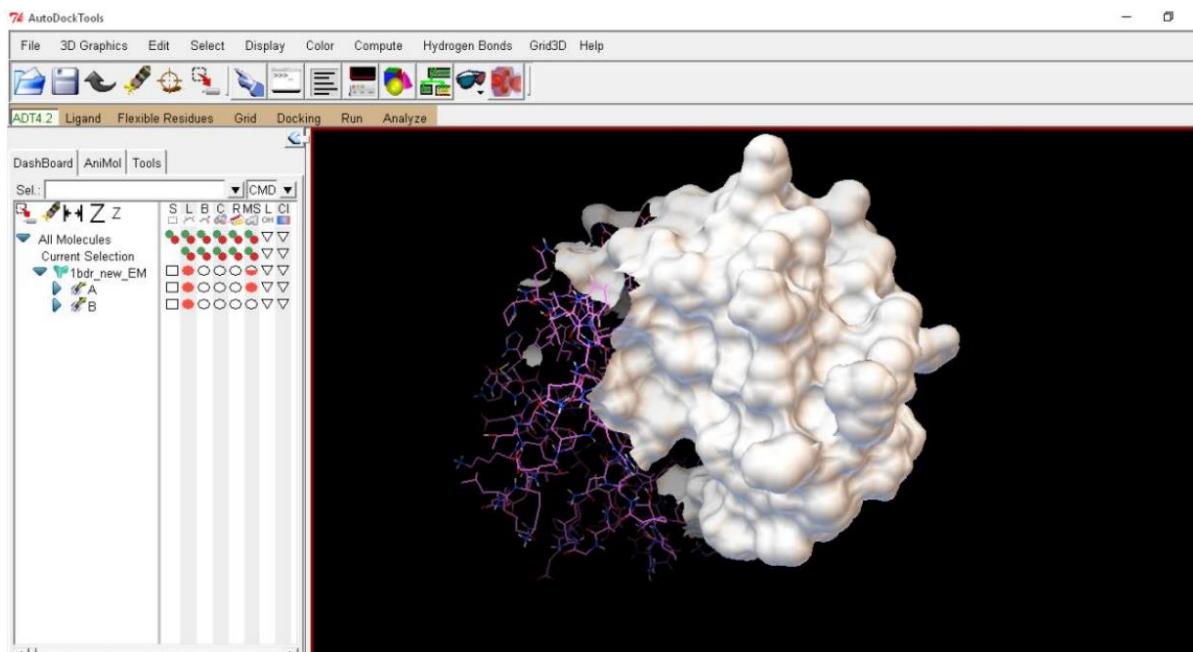
## Step 35:

## BID 19006

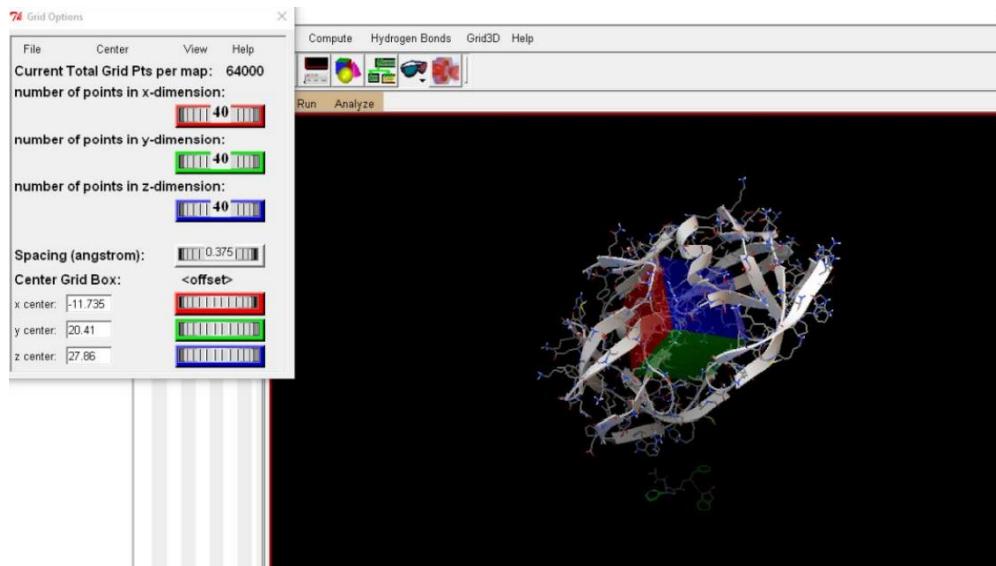
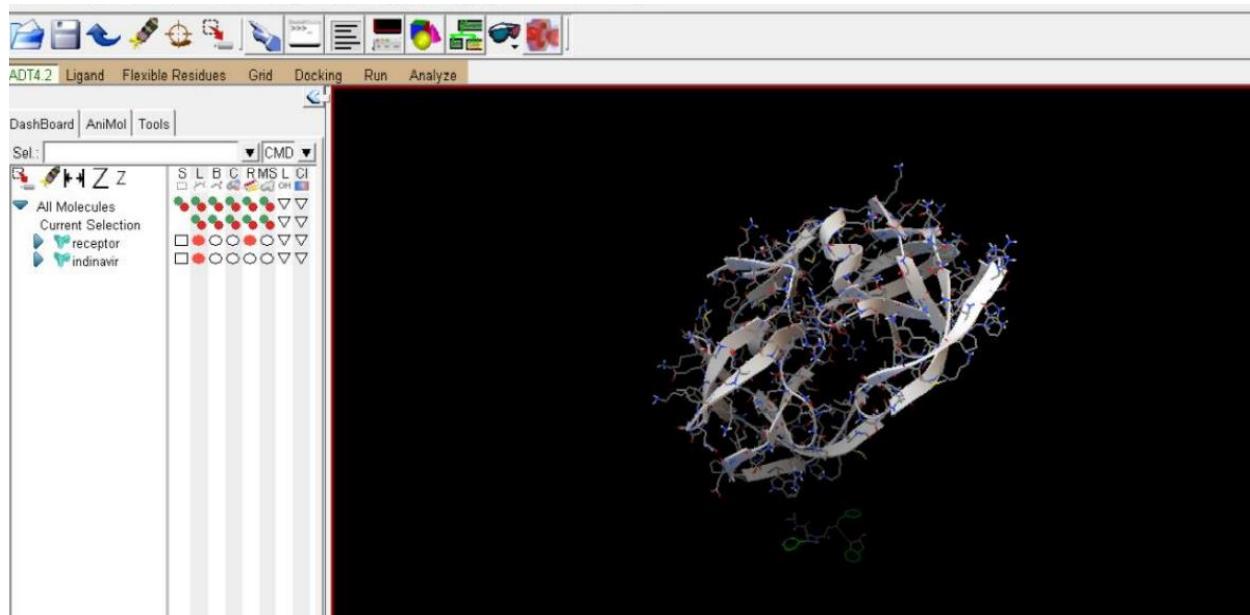
Now, the 4 best conformations are displayed in the table with their binding energy and number of clusters. So, once that is done, after the file is generated proceed with analysis. So, for that to analyse the docking results go to analyse go to docking click on open and than select the dlg file click on OK. Now go to analyse and go to conformations and than play ranked by energy.

NET	B	46	0.227	2.284	23.834	1.791	-36.08	-1.21	0.0000 // E=	-9.155
ILE	B	47	0.562	1.798	9.887	0.864	-38.63	28.50	0.0000 // E=	2.974
GLY	B	48	0.341	3.819	1.275	0.151	-9.09	71.99	0.0000 // E=	68.494
GLY	B	49	0.458	0.455	2.559	0.294	-15.25	27.29	0.0000 // E=	15.805
ILE	B	50	0.664	3.133	12.013	1.556	-28.77	43.13	0.0000 // E=	31.724
GLY	B	51	0.742	3.164	4.820	0.390	-11.37	75.26	0.0000 // E=	73.001
GLY	B	52	0.181	8.849	3.789	0.211	-17.08	27.43	0.0000 // E=	23.375
PHE	B	53	1.263	11.017	14.696	2.214	-30.66	5.62	0.0000 // E=	4.159
ILE	B	54	0.598	4.770	5.048	1.982	-36.31	-9.66	0.0000 // E=	-33.569
LYSH	B	55	0.589	4.563	14.558	1.973	-34.00	-0.20	0.0000 // E=	-12.518
VAL	B	56	0.632	1.729	2.315	1.403	-34.46	-7.00	0.0000 // E=	-35.380
ARG	B	57	0.883	1.726	18.292	0.472	-46.26	-266.06	0.0000 // E=	-290.944
GLN	B	58	0.978	8.128	3.685	0.732	-42.52	-172.48	0.0000 // E=	-201.478
TYR	B	59	0.241	2.405	6.611	1.258	-72.37	-60.33	0.0000 // E=	-122.185
ASP	B	60	0.073	2.836	3.562	0.203	-31.37	6.94	0.0000 // E=	-17.762
GLN	B	61	1.105	4.565	8.711	0.702	-18.59	-158.08	0.0000 // E=	-161.584
ILE	B	62	0.795	4.443	3.184	1.350	-33.51	-5.19	0.0000 // E=	-28.922
LEU	B	63	0.201	2.574	3.392	1.007	-33.34	-0.82	0.0000 // E=	-26.988
ILE	B	64	0.910	7.280	9.147	1.530	-34.06	-10.10	0.0000 // E=	-25.292
GLU	B	65	0.117	2.842	4.648	2.065	-41.25	-9.72	0.0000 // E=	-41.303
ILE	B	66	0.505	2.461	5.178	1.408	-33.22	-8.34	0.0000 // E=	-32.012
CYSH	B	67	1.027	1.699	2.221	1.499	-15.50	29.18	0.0000 // E=	20.127
GLY	B	68	0.484	3.653	0.086	0.119	-11.80	43.18	0.0000 // E=	35.719
HISA	B	69	1.112	4.591	3.679	0.515	-37.45	-18.46	0.0000 // E=	-46.013
LYSH	B	70	0.283	2.146	6.269	0.621	-23.83	1.36	0.0000 // E=	-13.148
ALA	B	71	0.145	0.922	2.015	0.378	-26.57	-12.84	0.0000 // E=	-35.951
ILE	B	72	1.367	8.372	3.818	2.713	-14.05	33.69	0.0000 // E=	35.908
GLY	B	73	0.401	0.464	1.331	0.107	-20.00	26.12	0.0000 // E=	8.427
THR	B	74	0.203	0.787	6.215	0.788	-26.19	-20.74	0.0000 // E=	-38.937
VAL	B	75	0.484	5.300	3.451	0.841	-36.50	-10.75	0.0000 // E=	-37.170
LEU	B	76	0.290	2.963	4.370	0.894	-47.87	-11.75	0.0000 // E=	-51.104
VAL	B	77	0.531	2.424	1.725	1.078	-35.31	22.49	0.0000 // E=	-7.059
GLY	B	78	1.279	3.083	1.561	0.229	-11.20	48.16	0.0000 // E=	43.111
PRO	B	79	1.968	24.146	13.805	0.194	-15.35	-24.65	0.0000 // E=	0.110
THR	B	80	1.568	4.163	1.385	2.564	-23.71	0.71	0.0000 // E=	-13.321
PRO	B	81	0.451	16.308	14.483	0.266	-21.27	-24.36	0.0000 // E=	-14.125
VAL	B	82	0.664	1.836	0.671	1.594	-21.81	-8.66	0.0000 // E=	-25.708
ASN	B	83	0.900	3.076	3.252	0.779	-39.34	-188.47	0.0000 // E=	-219.800
ILE	B	84	0.717	3.065	6.058	0.656	-36.97	-13.72	0.0000 // E=	-40.198
ILE	B	85	0.828	7.016	2.668	2.805	-39.63	22.23	0.0000 // E=	-4.080
GLY	B	86	0.753	1.452	3.918	1.394	-26.93	11.27	0.0000 // E=	-8.148
ARG	B	87	1.520	6.876	4.093	1.722	-52.06	-279.68	0.0000 // E=	-317.524
ASN	B	88	1.536	3.476	4.262	1.112	-35.60	-175.38	0.0000 // E=	-200.589
LEU	B	89	0.161	2.866	4.169	0.809	-50.04	-0.75	0.0000 // E=	-42.791
LEU	B	90	0.253	3.923	1.017	0.276	-55.06	-9.81	0.0000 // E=	-59.407
THR	B	91	0.322	1.343	1.279	0.575	-31.38	-29.21	0.0000 // E=	-57.063
GLN	B	92	0.788	4.631	2.318	0.625	-34.51	-166.88	0.0000 // E=	-193.025
ILE	B	93	1.668	3.263	4.619	1.278	-31.25	41.29	0.0000 // E=	20.872
GLY	B	94	0.329	2.872	0.496	0.188	-12.92	33.50	0.0000 // E=	24.468
CYSH	B	95	0.219	2.181	3.138	0.090	-41.18	-8.00	0.0000 // E=	-43.552
THR	B	96	0.446	2.307	2.973	0.778	-35.24	-28.56	0.0000 // E=	-57.291
LEU	B	97	0.139	13.558	8.252	1.137	-47.06	-11.02	0.0000 // E=	-34.997
ASN	B	98	1.114	2.810	4.742	0.269	-40.99	-184.35	0.0000 // E=	-216.407
PHE	B	99	1.361	6.669	3.172	1.230	-51.56	69.46	0.0000 // E=	30.335
OXT	B	99	0.000	0.000	0.000	0.000	0.06	9.31	0.0000 // E=	9.368

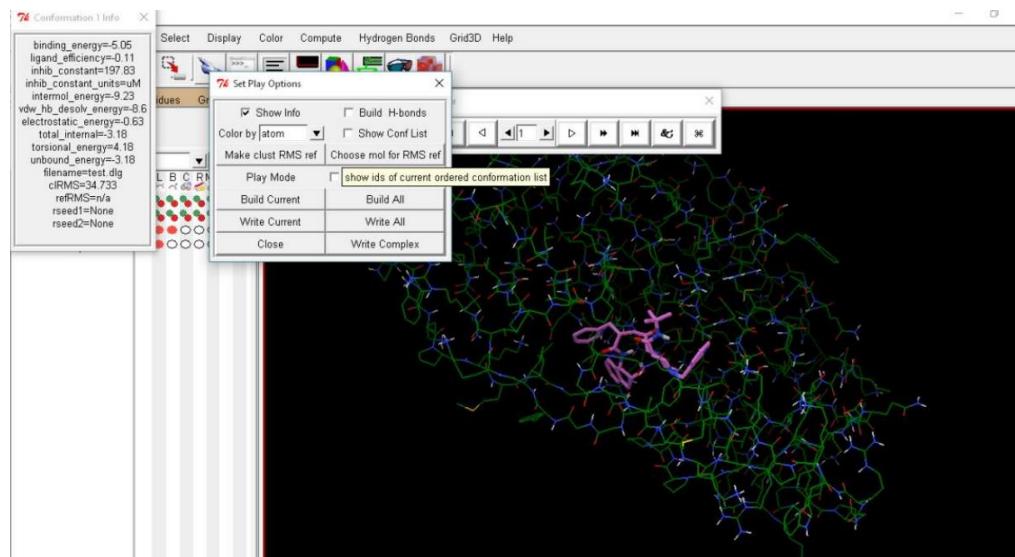
# BID 19006



# BID 19006

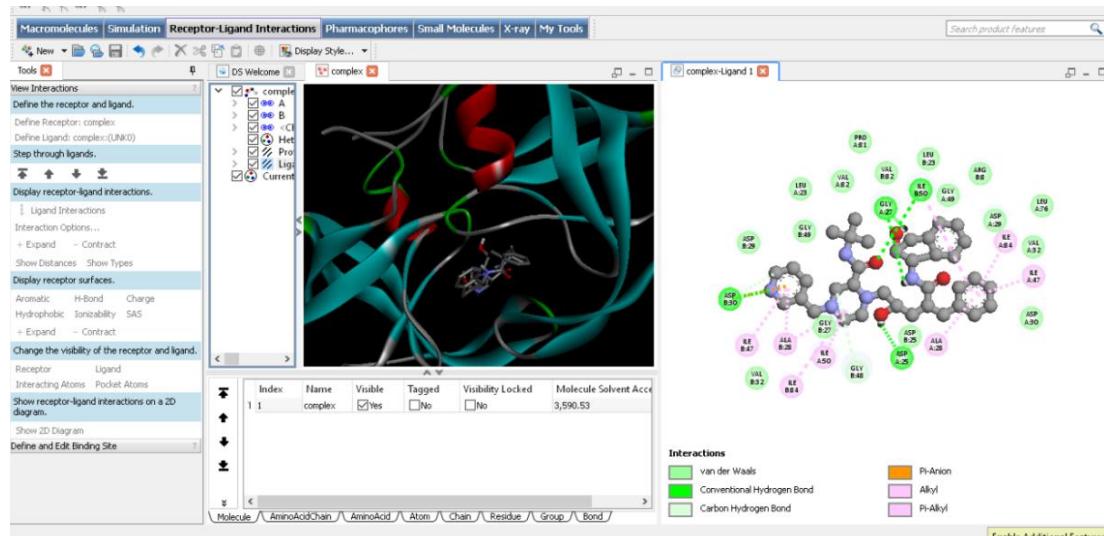


## BID 19006

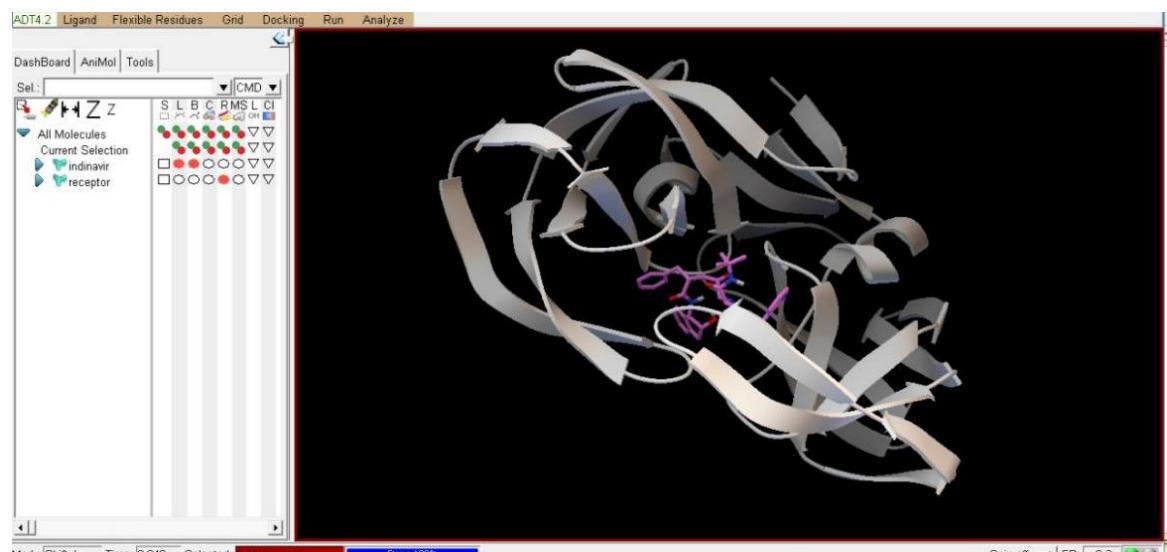
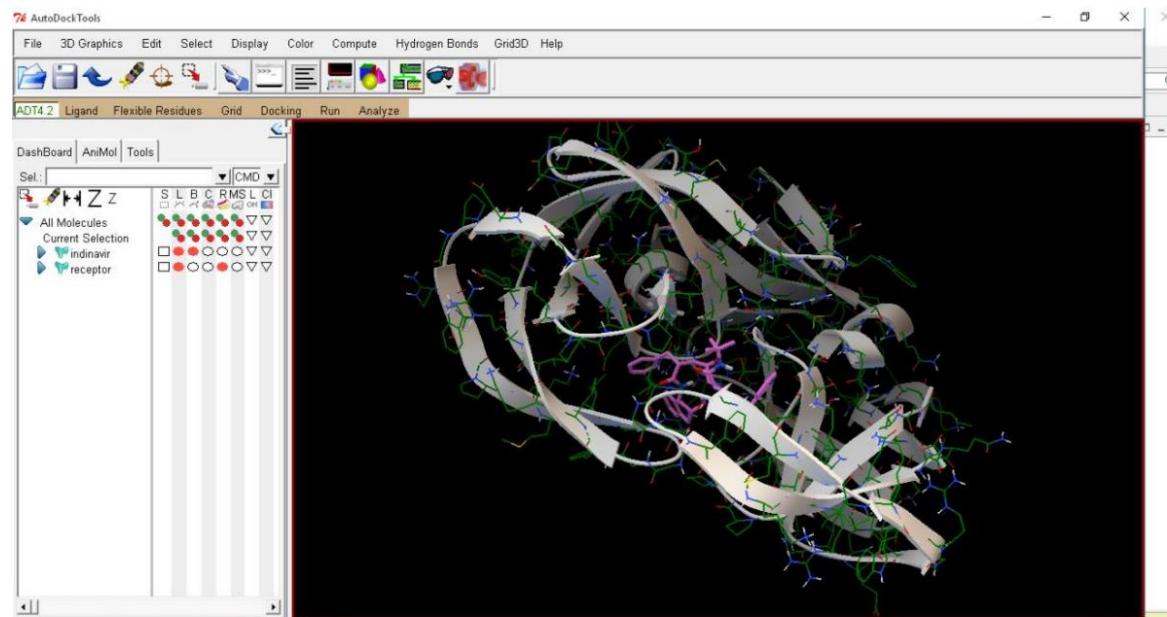


### Step 36:

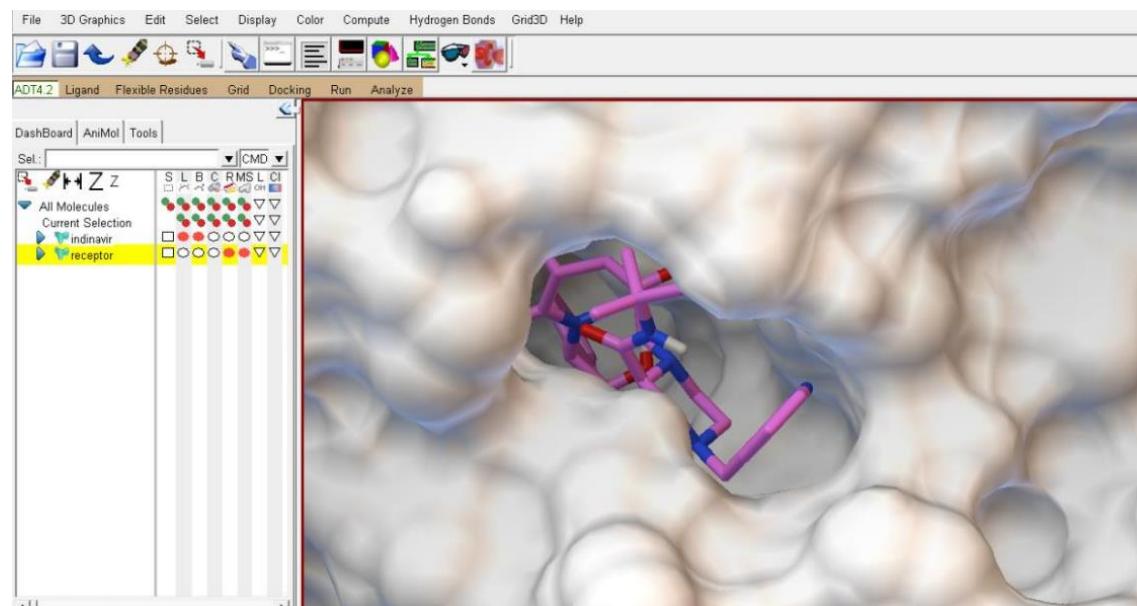
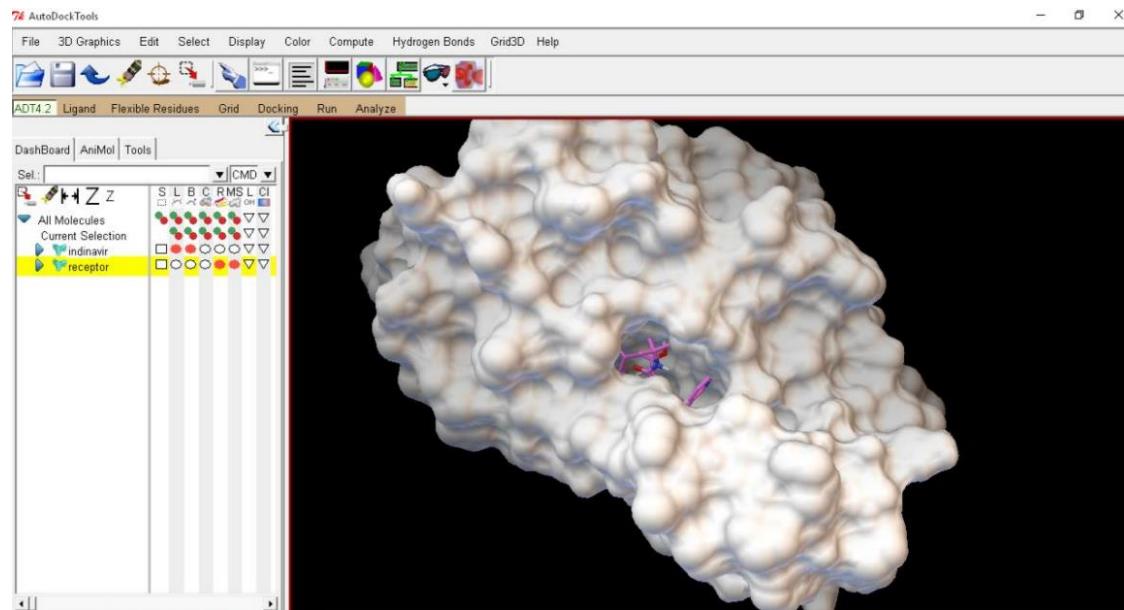
First we have to write the complex for first two conformations and save in it the pdbqt format and now we have to save both the complexes in pdb format for viewing the interactions, now to view the interactions open discovery studio visualizer. Nowe have to go to file and than click on open and than open the complex. Just select the receptor ligand interactions after that we have to change the display style to sticks and ball and sticks once that is done we have to select the ligand and now click on define ligand. Select the chain and than click on show 2D diagram. After that view the interactions and than just do the same for the second conformation.



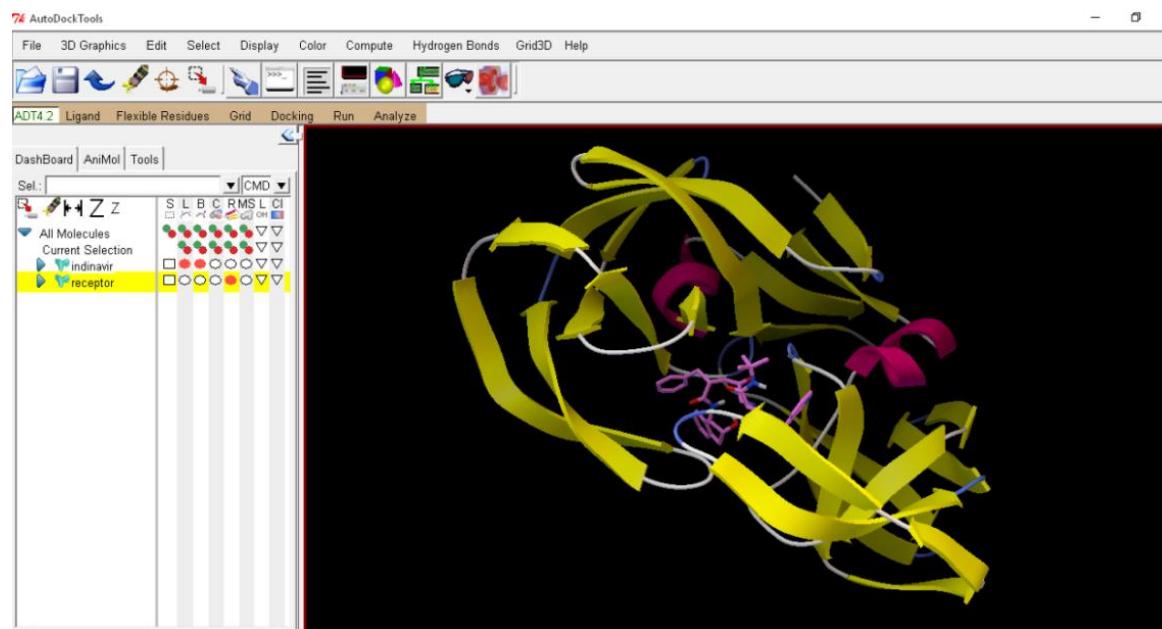
## BID 19006



# BID 19006



# BID 19006



CLUSTERING HISTOGRAM

Clus	Lowest	Run	Mean	Num	Histogram
-ter	Binding		Binding	in	
Rank	Energy		Energy	Clus	5 10 15 20 25 30 35
1	-5.51	9	-3.73	3	####
2	-3.66	1	-3.18	2	##
3	-2.99	4	-2.99	1	#
4	-2.89	7	-2.89	1	#
5	-2.70	6	-2.70	1	#
6	-1.95	10	-1.95	1	#
7	-1.47	5	-1.47	1	#

Number of multi-member conformational clusters found = 2, out of 10 runs.

RMSD TABLE

Rank	Sub-	Run	Binding	Cluster	Reference	Grep
	Rank		Energy	RMSD	RMSD	Pattern
1	1	9	-5.51	0.00	31.69	RANKING
1	2	3	-5.05	1.62	31.96	RANKING
1	3	2	-0.64	1.79	31.88	RANKING
2	1	1	-3.66	0.00	31.80	RANKING
2	2	8	-2.71	1.38	31.65	RANKING
3	1	4	-2.99	0.00	32.28	RANKING
4	1	7	-2.89	0.00	31.89	RANKING
5	1	6	-2.70	0.00	32.04	RANKING
6	1	10	-1.95	0.00	31.80	RANKING
7	1	5	-1.47	0.00	30.13	RANKING

## Practical-4

**Title:** Virtual screening of compounds with AutoDock Vina.

### **AutoDock Vina:**

It is a new program for molecular docking and virtual screening, is presented. AutoDock Vina achieves an approximately two orders of magnitude speed-up compared to the molecular docking software previously developed in our lab (AutoDock 4), while also significantly improving the accuracy of the binding mode predictions, judging by our tests on the training set used in AutoDock 4 development. Further speed-up is achieved from parallelism, by using multithreading on multi-core machines. AutoDock Vina automatically calculates the grid maps and clusters the results in a way transparent to the user. The **AutoDock Vina** tool allows running ligand-receptor docking calculations with AutoDock Vina, using either a web service provided by the National Biomedical Computation Resource (NBCR) or a locally installed copy of the program.

The scoring function used in Vina was derived using the PDBbind data set, and the performance of Vina has been compared to that of AutoDock 4.0.1 (referred to as AutoDock here) on a set of 190 protein-ligand complexes that had been used as a training set for the AutoDock scoring function.

In this set, the receptors are treated as rigid, and the ligands are treated as flexible molecules with the number of active rotatable bonds ranging from 0 to 32.

Both Vina and AutoDock require a specification of the "search space" in the coordinate system of the receptor, within which various positions of the ligand are to be considered.

Vina uses a sophisticated gradient optimization method in its local optimization procedure. The calculation of the gradient effectively gives the optimization algorithm a "sense of direction" from a single evaluation. By using multithreading, Vina can further speed up the execution by taking advantage of multiple CPUs or CPU cores

### **Molecular docking:**

Molecular docking is a computational procedure that attempts to predict noncovalent binding of macromolecules or, more frequently, of a macromolecule (receptor) and a small molecule (ligand) efficiently, starting with their unbound structures, structures obtained from MD simulations, or homology modeling, etc. The goal is to predict the bound conformations and the binding affinity.

Exercise: Identify an amylase protein and its inhibitor from database. Perform a Lamarckian based Genetic Algorithm search using AutoDock tools. Identify the most favourable conformation and the interacting amino acids. Prepare a table and explain the properties.

Step 1: Go to homepage of PubChem and search for Curcumin:

**BEST MATCH**

**Curcumin; 458-37-7; Diferuloylmethane; Natural Yellow 3; Turmeric Yellow; Turmeric; Indian Saffron; Curcuma; ...**

Compound CID: 969516

MF:  $C_{21}H_{20}O_6$  MW: 368.4g/mol

IUPAC Name: (1E,6E)-1,7-bis(4-hydroxy-3-methoxyphenyl)hepta-1,6-diene-3,5-dione

Isomeric SMILES: COC1=C(C=CC=C1)/C=C/C(=O)CC(=O)/C=C/C2=CC(=C=C2O)OC2=O

InChIKey: VFLDPWPHBUODDF-FCXRPNKRSA-N

InChI: InChI=1S/C21H20O6/c1-26-20-11-14(5-9-18(20)24)3-7-16(22)13-17(23)8-4-15-6-10-19(25)21(12-15)27-2/h3-12,24-25H,13H2,1-2H3/b7-3+,8-4+

Create Date: 2004-09-16

[Summary](#) [Similar Structures Search](#) [Related Records](#) [PubMed \(MeSH Keyword\)](#)

Compounds (110)	Substances (506)	Genes (1)	Proteins (4)	Pathways (60)	BioAssays (5,325)	Literature (19,925)
<a href="#">Compounds</a>	<a href="#">Substances</a>	<a href="#">Genes</a>	<a href="#">Proteins</a>	<a href="#">Pathways</a>	<a href="#">BioAssays</a>	<a href="#">Literature</a>

Patents

27°C Haze 02:39 ENG IN 10-11-2022

Step 2: Now you will get to see Structure , molecular formula etc for curcumin:

**Curcumin**

PubChem CID: 969516

**Structure**

Find Similar Structures

**Chemical Safety**

! Irritant

[Laboratory Chemical Safety Summary \(LCSS\) Datasheet](#)

**Molecular Formula**

$C_{21}H_{20}O_6$

curcumin  
458-37-7  
Diferuloylmethane  
Natural yellow 3

**Synonyms**

**CONTENTS**

- Title and Summary
- 1 Structures
- 2 Names and Identifiers
- 3 Chemical and Physical Properties
- 4 Spectral Information
- 5 Related Records
- 6 Chemical Vendors
- 7 Drug and Medication Information
- 8 Food Additives and Ingredients
- 9 References and...

27°C Haze 02:40 ENG IN 10-11-2022

-Name of compound : Curcumin

-Protein (pdb id): PDBe Ligand Code: CC9

PDBe Structure Code: 4K58

Step 3: After that we will download the first 10 structures in sdf file format.

Refine your results • Subsets of your results

Chemical Properties  
Rule of 5 (691)

BioActivity Experiments  
BioAssays, Active (258)  
BioAssays, Tested (351)  
Protein 3D Structures (4)  
Human protein kinase CK2 alpha in complex with curcumin degradation products (1)  
Human protein kinase CK2 alpha in complex with feruloylmethane (1)  
Crystal structure of lignostilbene bound to Co-LSD4 from *Sphingobium* sp. strain SYK-6 (1)  
... All 6 Structures

BioMedical Annotation  
Pharmacological Actions (6)  
Anti-Inflammatory Agents, Non-Steroidal (2)  
Antioxidants (1)  
Indicators and Reagents (1)  
... All 12 Pharmacological Actions

BioSystems (6)

Depositor Category  
Chemical Vendors (189)  
NIH Molecular Libraries (42)

Summary (Search Results)

CSV JSON XML

COMPRESSION:  
 None  GZip

Chemical Structure Records

SDF JSON XML ASNT

COORDINATE TYPE:  
 2D  3D

NUMBER OF CONFORMERS: 1

COMPRESSION:  
 None  GZip

Chemical Structure Images

Step 4: Go to command prompt in obabel :

```
C:\Users\liliol\Desktop\library\obabel
No input file or format spec or possibly a misplaced option,
most options must come after the input files. (-i -o -m can be anywhere.)
OpenBabel 3.1.1 -- May 16 2020 -- 11:57:55
Usage: obabel [-i<input-type>] [<output-type>] -o<filename> [Options]
Try -h option for more information.

C:\Users\liliol\Desktop\library\obabel 1structure.sdf -O struct.sdf -m
10 molecules converted
10 files output. The first is struct1.sdf
C:\Users\liliol\Desktop\library\obabel -h
C:\Users\liliol\Desktop\library\obabel
Usage: obabel [options] <filename>
options:
    description:
        -c crit      set convergence criteria (default=1e-6)
        -cg          use conjugate gradients algorithm (default)
        -sd          use steepest descent algorithm
        -newton     use NewtonRaphm lineage search (default=Simple)
        -ff ffid    select a forcefield:
            -h          add hydrogen atoms
        -n steps    specify the maximum number of steps (default=1500)
        -cut         use cut-off (default=don't use cut-off)
        -rvdw rvdw   specify the van der Waals cut-off distance (default=6.0)
        -rele relc  specify the electrostatic cut-off distance (default=10.0)
        -pf freq    specify the frequency to update the non-bonded pairs (default=10)
        -aff        General Amber Force Field (AMPP).
```

```
Command Prompt
-rvdw rvdw specify the VDW cut-off distance (default=6.0)
-rele rele specify the Electrostatic cut-off distance (default=10.0)
-pf freq specify the frequency to update the non-bonded pairs (default=10)

GAFF General Amber Force Field (GAFF).
Chemical Chemical force field.

MMFF94 MMFF94 force field.

MMFF94s MMFF94s force field.

UFF Universal Force Field.

C:\Users\BioLab\Desktop\library>obabel -isdf *.sdf -opdbqt -O*.pdbqt
10 molecules converted

C:\Users\BioLab\Desktop\library>dir
Volume in drive C has no label.
Volume Serial Number is D2B6-C569

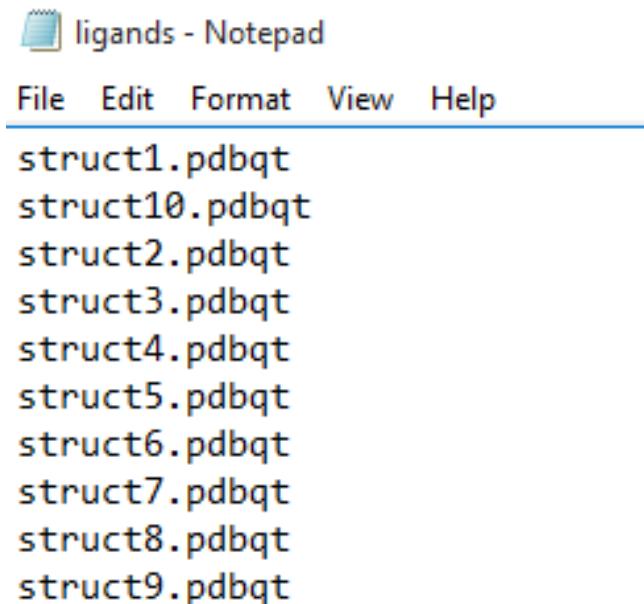
Directory of C:\Users\BioLab\Desktop\library

09/13/2022  10:53 AM    <DIR>          .
09/13/2022  10:53 AM    <DIR>          ..
09/13/2022  10:53 AM           31,325 10structure.pdbqt
09/13/2022  10:22 AM           71,218 10structure.sdf
09/13/2022  10:46 AM           102,543 backup
                           2 File(s)   102,543 bytes
                           3 Dir(s)  262,192,951,296 bytes free

C:\Users\BioLab\Desktop\library>obabel -isdf *.sdf -opdbqt -O*.pdbqt
20 molecules converted
11 files output. The first is 10structure.pdbqt

C:\Users\BioLab\Desktop\library>dir /B >ligands.txt
```

Step 5: Ligands in notepad:

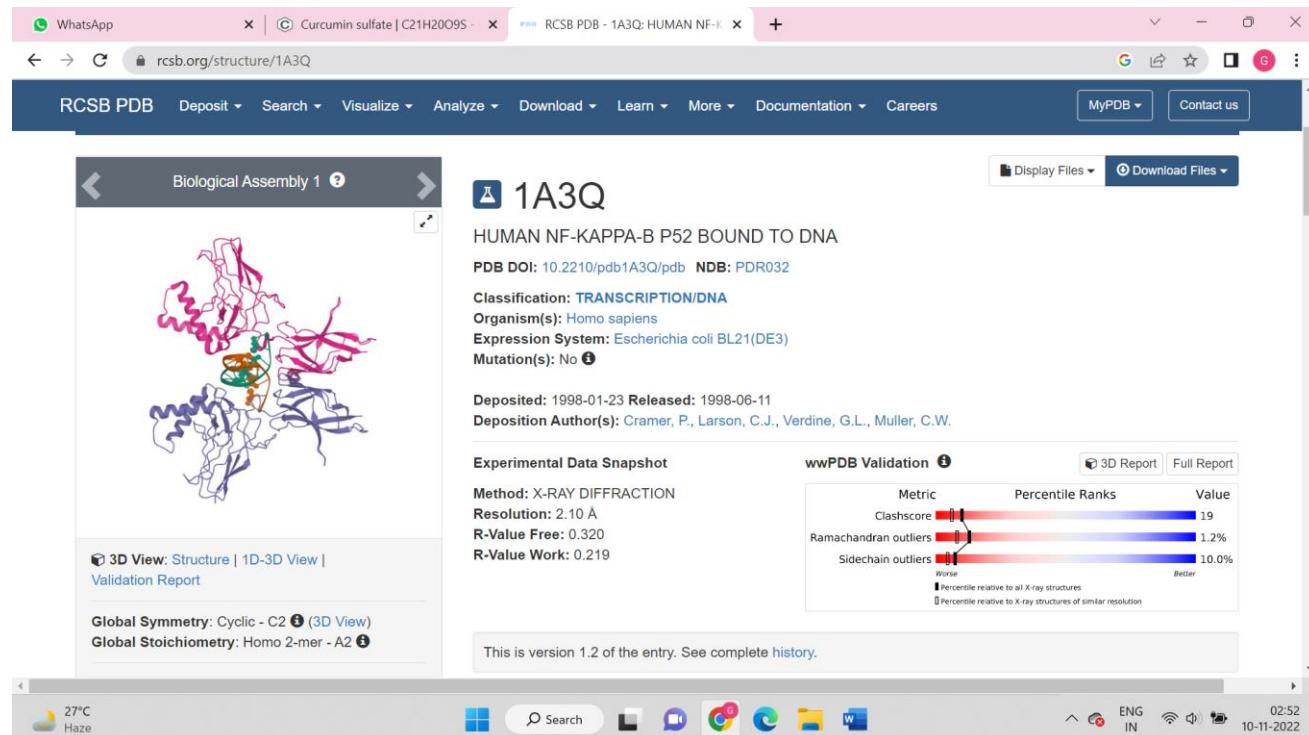


ligands - Notepad

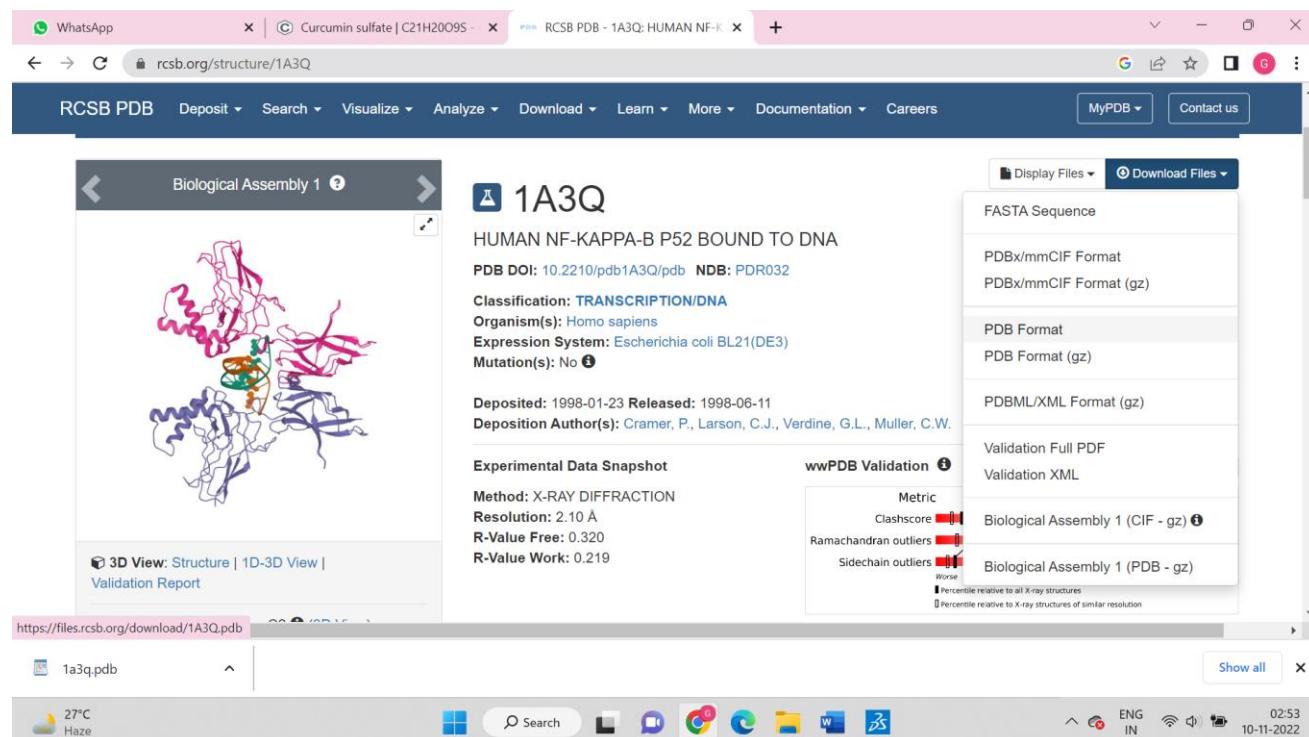
File Edit Format View Help

struct1.pdbqt  
struct10.pdbqt  
struct2.pdbqt  
struct3.pdbqt  
struct4.pdbqt  
struct5.pdbqt  
struct6.pdbqt  
struct7.pdbqt  
struct8.pdbqt  
struct9.pdbqt

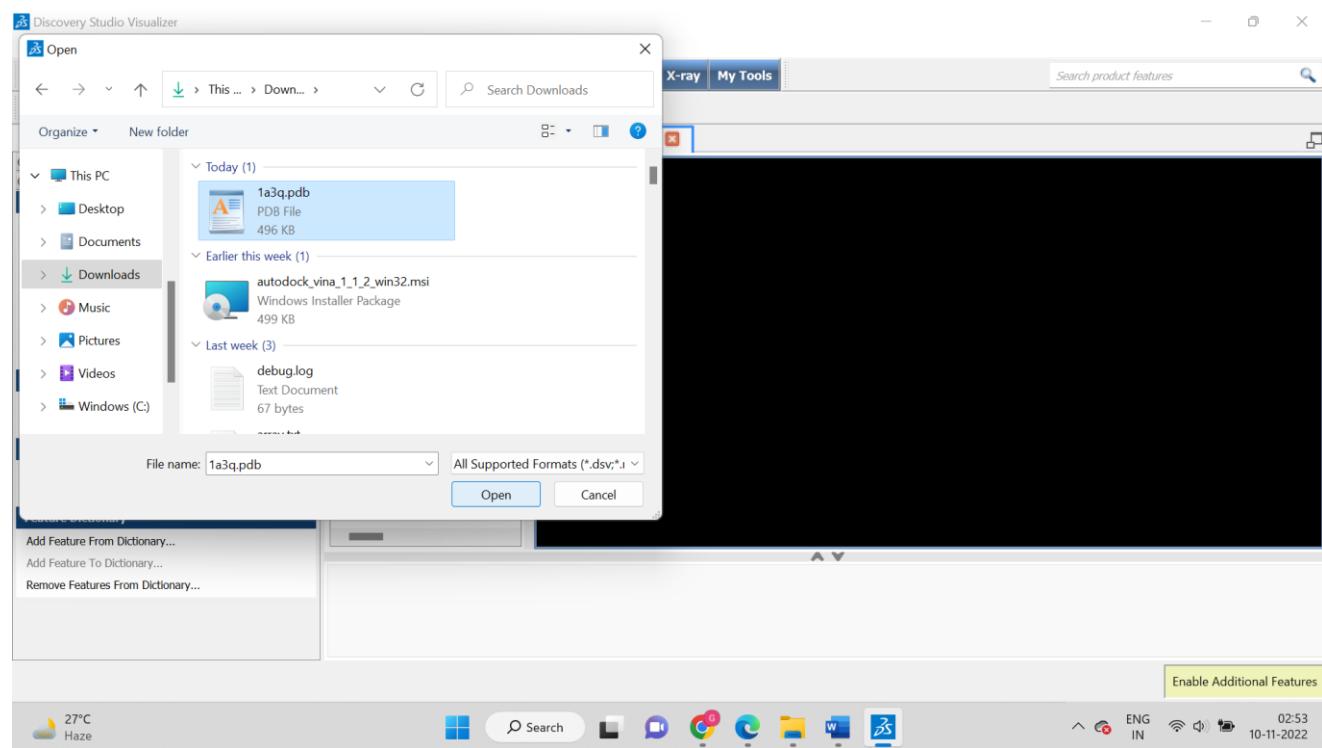
Step 6: Now we have to go to homepage of PDB and search for 1A3Q :



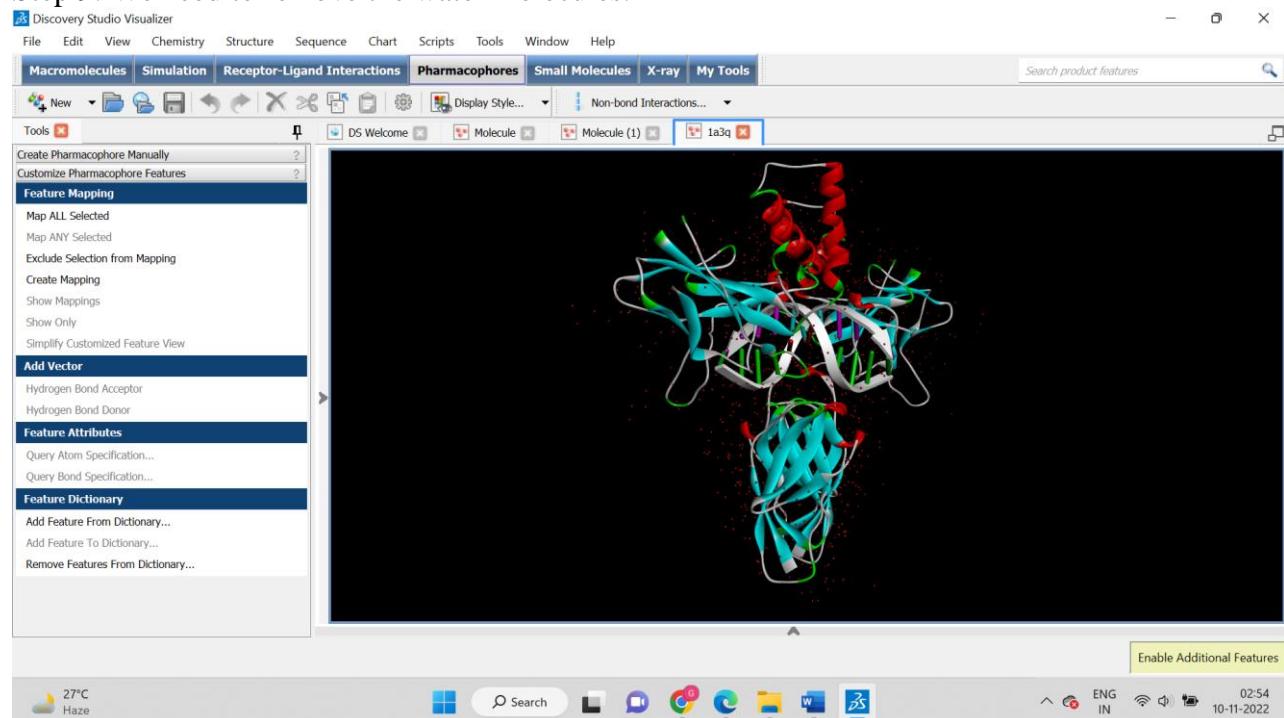
Step 7: Download the file as shown below:



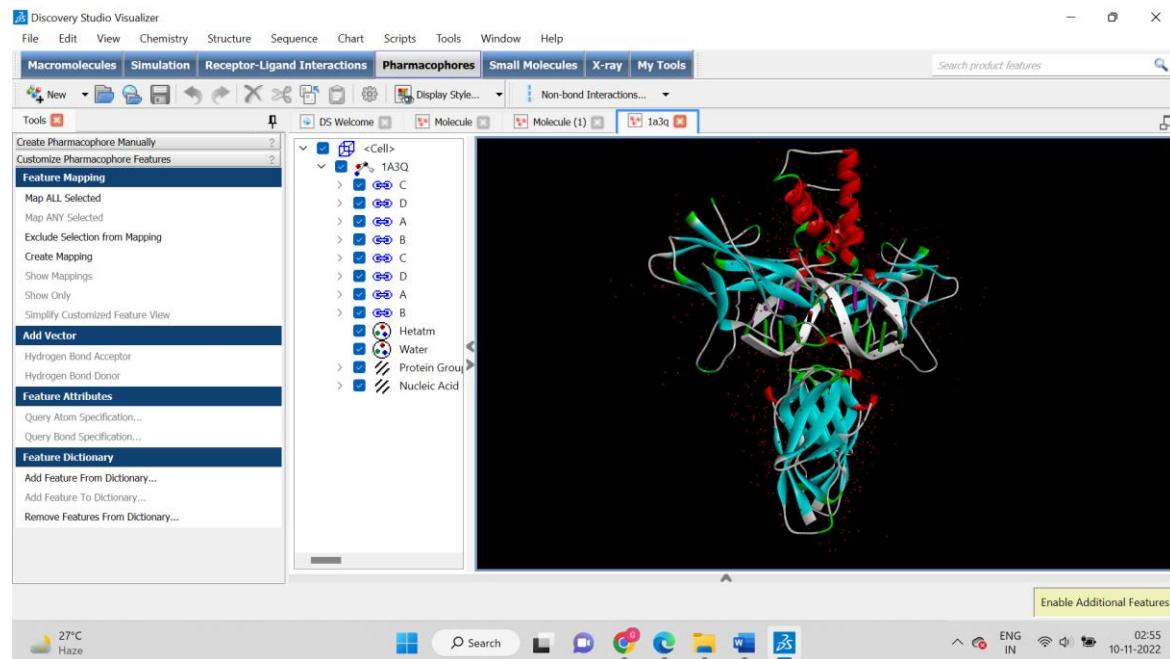
Step 8: Now we have to open the file in Discovery Studio Visualizer and run it as shown :



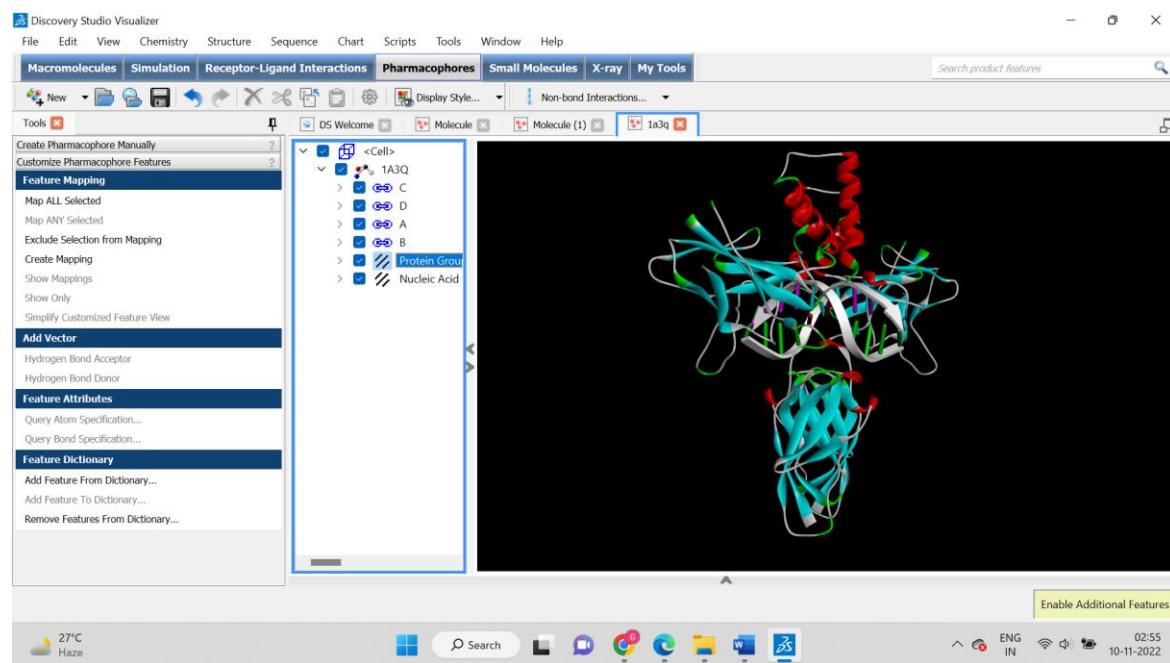
Step 9: We need to remove the water molecules:



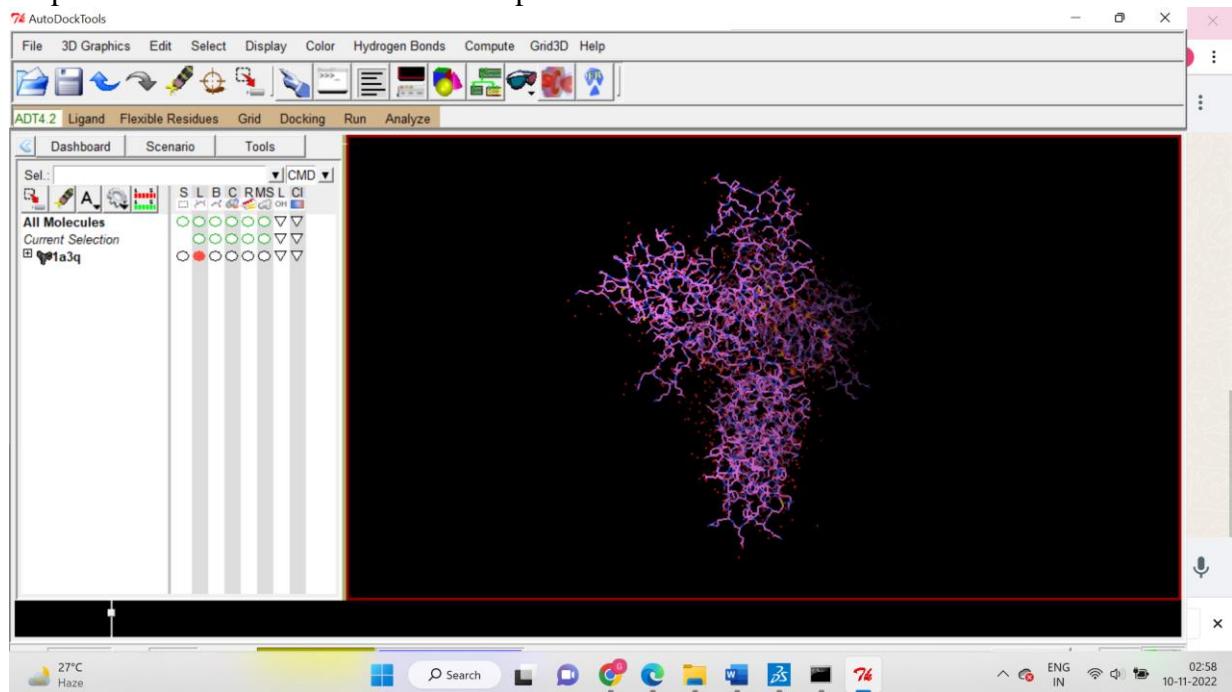
Step 10: Now click on water and delet it:



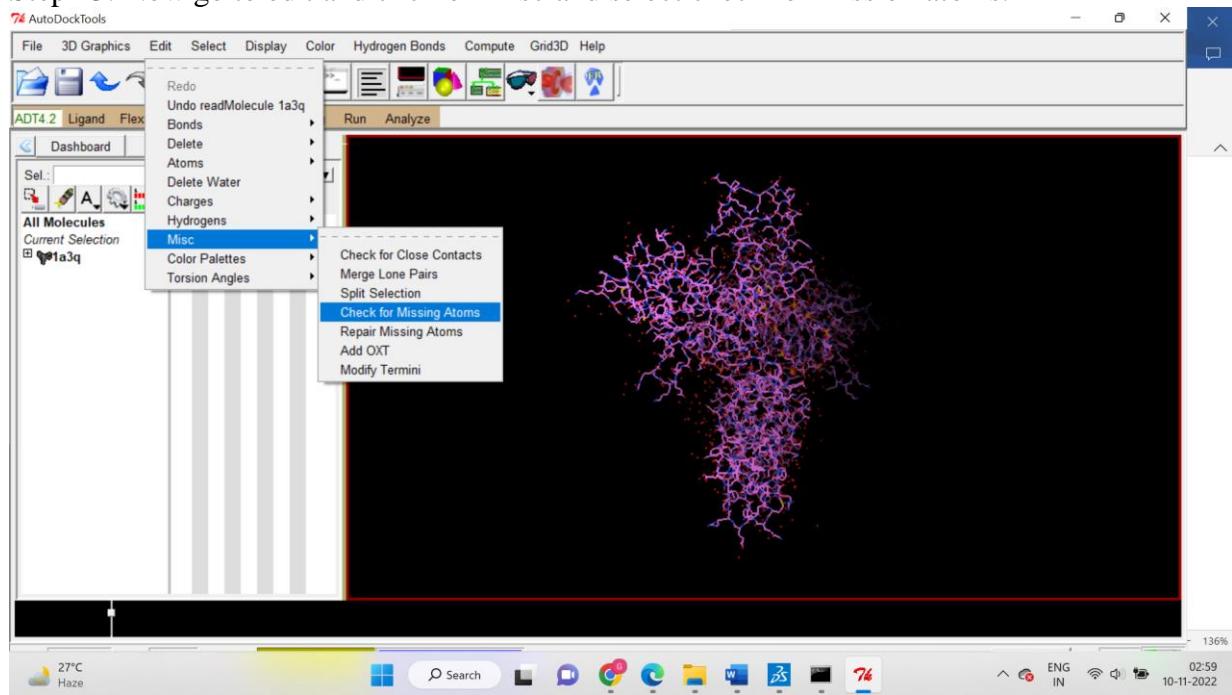
Step 11: Water removed:



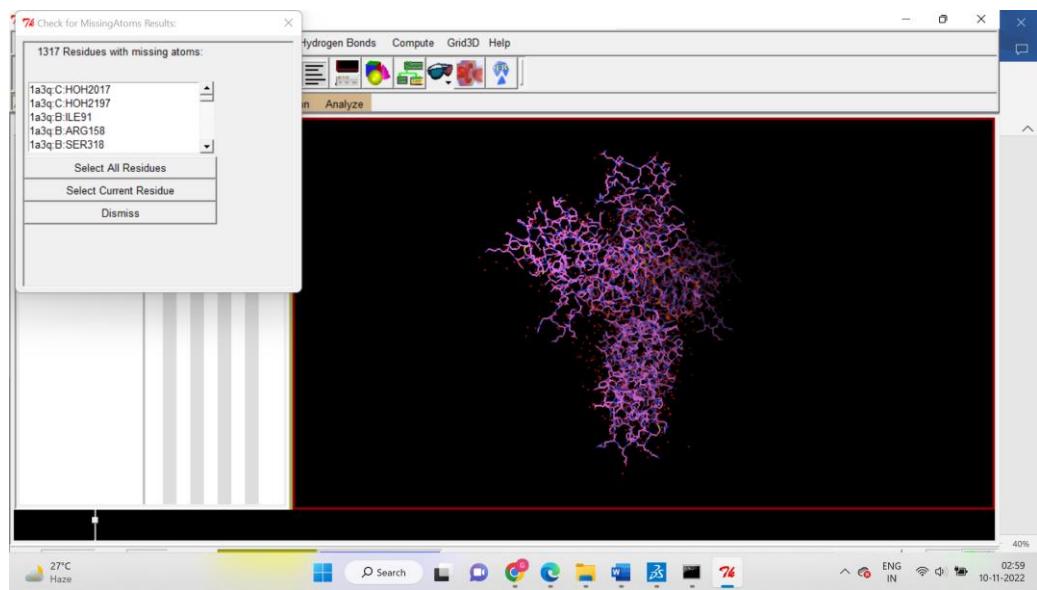
Step 12: Save the above structure and open it in auto dock tool as shown:



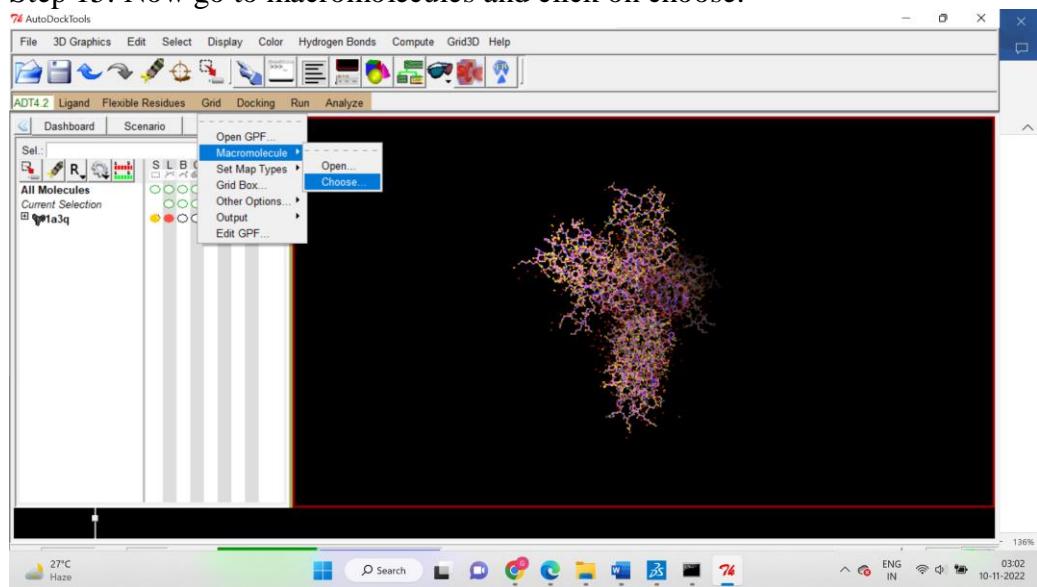
Step 13: Now go to edit and click on misc and select check for mission atoms:

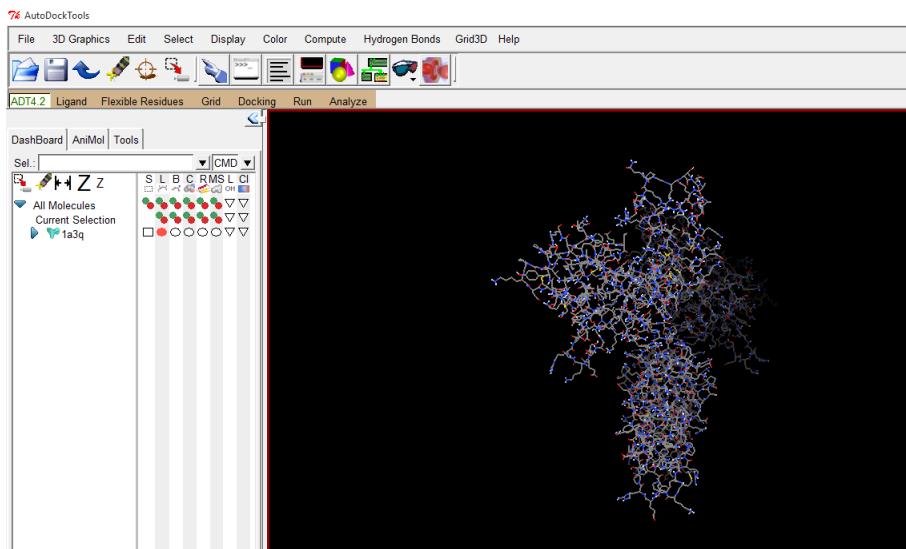


Step 14: It shows 1317 residues with missing atoms:

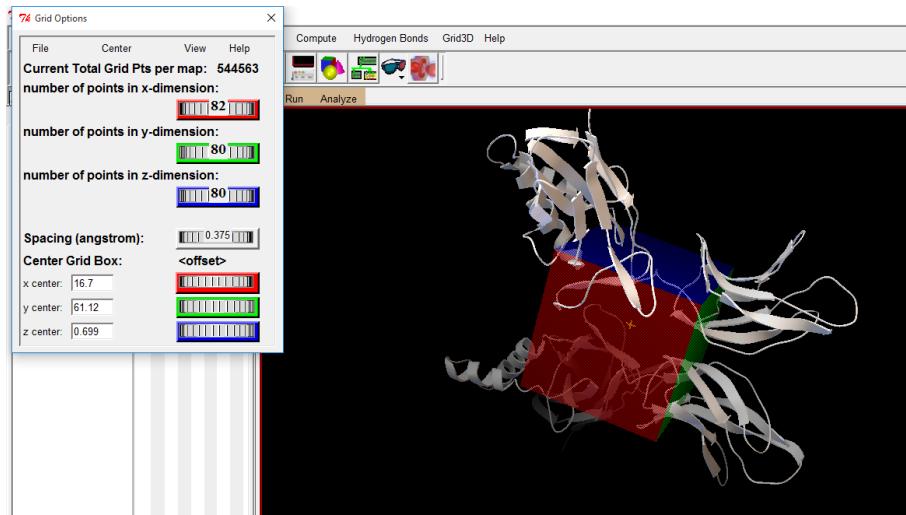


Step 15: Now go to macromolecules and click on choose:





Step 16: We can see grid options: the x , y, z center:



```

File Edit Format View Help
receptor = 1a3q.pdbqt

center_x = 16.7
center_y = 61.12
center_z = 0.699

size_x = 80
size_y = 80
size_z = 80

num_modes = 10
energy_range = 4

```

## The Output:-

```

[cmd] Command Prompt
09/13/2022 10:55 AM      3,517 struct1.pdbqt
09/13/2022 10:55 AM      3,760 struct10.pdbqt
09/13/2022 10:55 AM      3,904 struct2.pdbqt
09/13/2022 10:55 AM      3,228 struct3.pdbqt
09/13/2022 10:55 AM      2,873 struct4.pdbqt
09/13/2022 10:55 AM      3,679 struct5.pdbqt
09/13/2022 10:55 AM      3,624 struct6.pdbqt
09/13/2022 10:55 AM      3,680 struct7.pdbqt
09/13/2022 10:55 AM      2,628 struct8.pdbqt
09/13/2022 10:55 AM      4,003 struct9.pdbqt
09/11/2022 12:01 PM      781 Vina.Windows.pl
09/12/2022 10:49 PM      413 Vina.Windows.pl
19 File(s)   2,684,688 bytes free
3 Dir(s)  261,880,311,888 bytes free

C:\Users\BioLab\Desktop\library>perl -v
This is perl 5, version 12, subversion 3 (v5.12.3) built for MSWin32-x86-multi-thread
Copyright 1987-2010, Larry Wall

Perl may be copied only under the terms of either the Artistic License or the
GNU General Public License, which may be found in the Perl 5 source kit.

Complete documentation for Perl, including FAQ lists, should be found on
this system using "man perl" or "perldoc perl". If you have access to the
Internet, point your browser at http://www.perl.org/, the Perl Home Page.

C:\Users\BioLab\Desktop\library>

[cmd] Command Prompt - perl Vina.Windows.pl
C:\Users\BioLab\Desktop\library>perl Vina.Windows.pl
Ligand file: ligands.txt
struct1.pdbqt
struct10.pdbqt
struct2.pdbqt
struct3.pdbqt
struct4.pdbqt
struct5.pdbqt
struct6.pdbqt
struct7.pdbqt
struct8.pdbqt
struct9.pdbqt
struct1.pdbqt
#####
# If you used AutoDock Vina in your work, please cite: #
# O. Trott, A. J. Olson, #
# AutoDock Vina: improving the speed and accuracy of docking #
# with a new scoring function, efficient optimization and #
# multithreading, Journal of Computational Chemistry 31 (2010) #
# 455-461 #
# DOI 10.1002/jcc.21334 #
# Please see http://vina.scripps.edu for more information. #
#####

WARNING: The search space volume > 27000 Angstrom^3 (See FAQ)
Output will be struct1_out.pdbqt
Detected 2 CPUs
Reading input ... done.
Setting up the scoring function ... done.
Analyzing the binding site ... done.

[cmd] Command Prompt
# If you used AutoDock Vina in your work, please cite: #
# O. Trott, A. J. Olson, #
# AutoDock Vina: improving the speed and accuracy of docking #
# with a new scoring function, efficient optimization and #
# multithreading, Journal of Computational Chemistry 31 (2010) #
# 455-461 #
# DOI 10.1002/jcc.21334 #
# Please see http://vina.scripps.edu for more information. #
#####

WARNING: The search space volume > 27000 Angstrom^3 (See FAQ)
Output will be struct9_out.pdbqt
Detected 2 CPUs
Reading input ... done.
Setting up the scoring function ... done.
Analyzing the binding site ... done.
Using random seed: -1727851404
Performing search ...
0% 10 20 30 40 50 60 70 80 90 100%
|----|----|----|----|----|----|----|----|
*****done.
Refining results ... done.

mode | affinity | dist from best mode
     | (kcal/mol) | rmsd l.b. | rmsd u.b.
-----+-----+-----+
 1    -6.7    0.000    0.000
 2    -6.5    28.492   34.359
 3    -6.2    28.788   34.640
 4    -6.1    26.840   30.863
 5    -6.1    39.497   42.762
 6    -6.0    31.802   35.399
 7    -6.0    31.404   35.286
 8    -5.9    2.965    7.659
 9    -5.9    25.293   30.141
10   -5.9    25.765   30.101

Writing output ... done.

C:\Users\BioLab\Desktop\library>

```

## Practical 5

### Title: Computer Simulation of molecules using GROMACS

#### Introduction:

GROMACS is one of the most widely used open-source and free software codes in chemistry, used primarily for dynamical simulations of biomolecules. It provides a rich set of calculation types, preparation and analysis tools. Several advanced techniques for free-energy calculations are supported.

Molecular dynamics (MD) is a computer simulation method for analyzing the physical movements of atoms and molecules. The atoms and molecules are allowed to interact for a fixed period of time, giving a view of the dynamic "evolution" of the system. In the most common version, the trajectories of atoms and molecules are determined by numerically solving Newton's equations of motion for a system of interacting particles, where forces between the particles and their potential energies are often calculated using interatomic potentials or molecular mechanics force fields. The method is applied mostly in chemical physics, materials science, and biophysics.

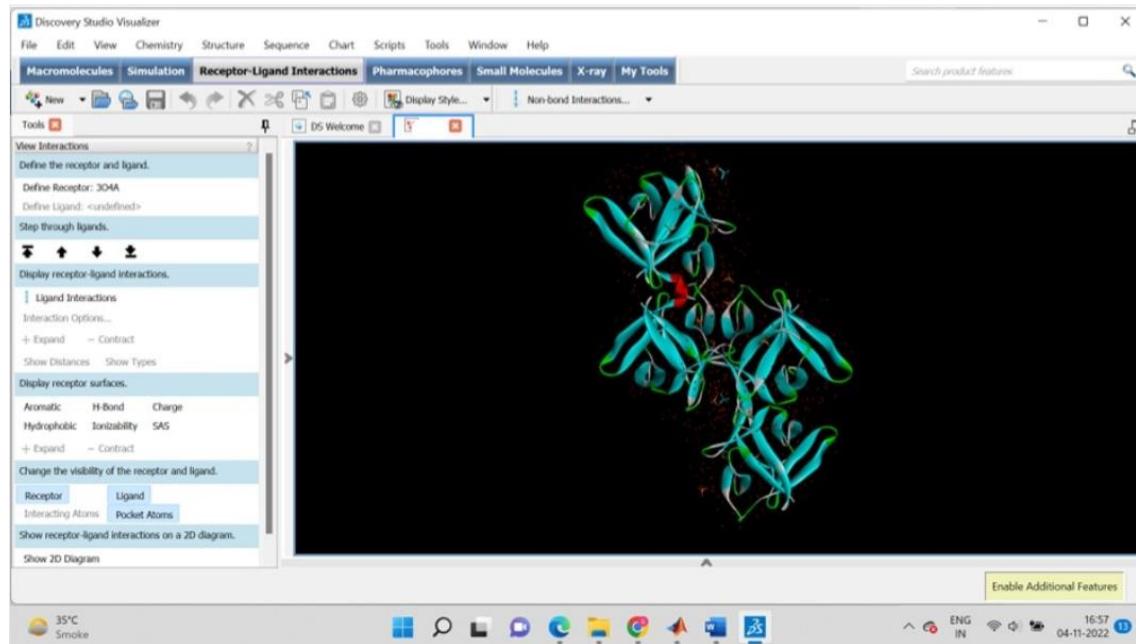
Tools for using MD are like GROMACS:

**Q1: Perform MD simulation of a protein structure (3VH5 or 1AKI) using GROMACS algorithm. Analyze the stability, conformation and geometry of the protein**

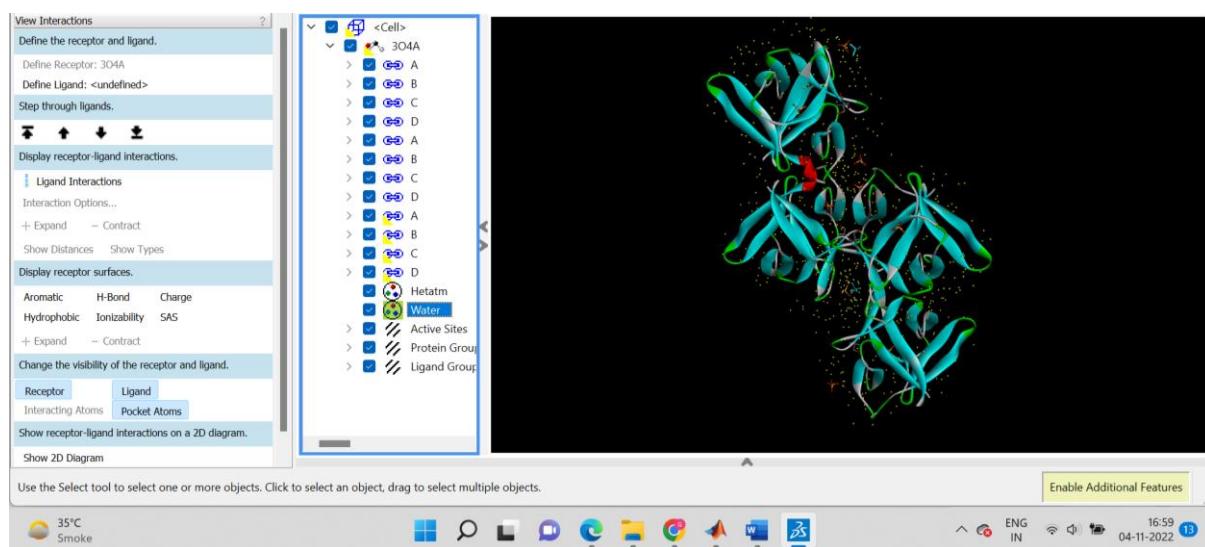
Step 1: Open RCSB PDB and download crystal structure of 3VH5:

The screenshot shows the RCSB PDB homepage with the search term '3VH5' entered. The results page for 3VH5 displays the title 'Crystal structure of the chicken CENP-T histone fold/CENP-W/CENP-S/CENP-X heterotetrameric complex, crystal form I'. It includes the PDB DOI (10.2210/pdb3VH5/pdb), classification as a DNA BINDING PROTEIN, and organism as Galus gallus. The experimental data snapshot shows X-RAY DIFFRACTION at 2.40 Å resolution. The wwPDB Validation section indicates a metric value of 0.242. Navigation links include '3D View', 'Structure', and '1D-3D View'.

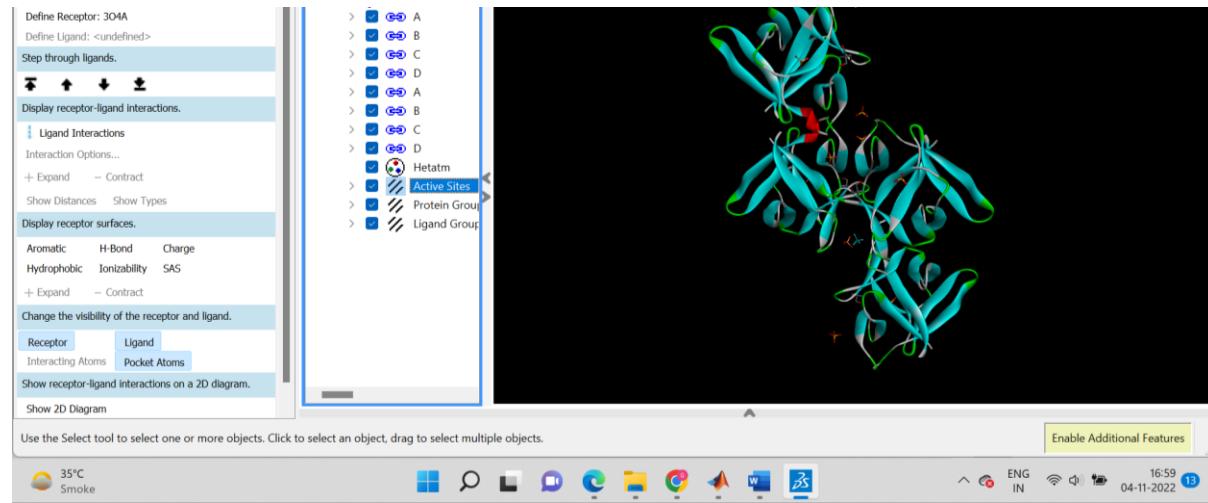
Step 2: Open the peptide in discovery studio visualizer and clear the water molecules.



Step 3: Open hierarchy → select water and right click on it → remove water.



Step 4: Save the file in pdb format.

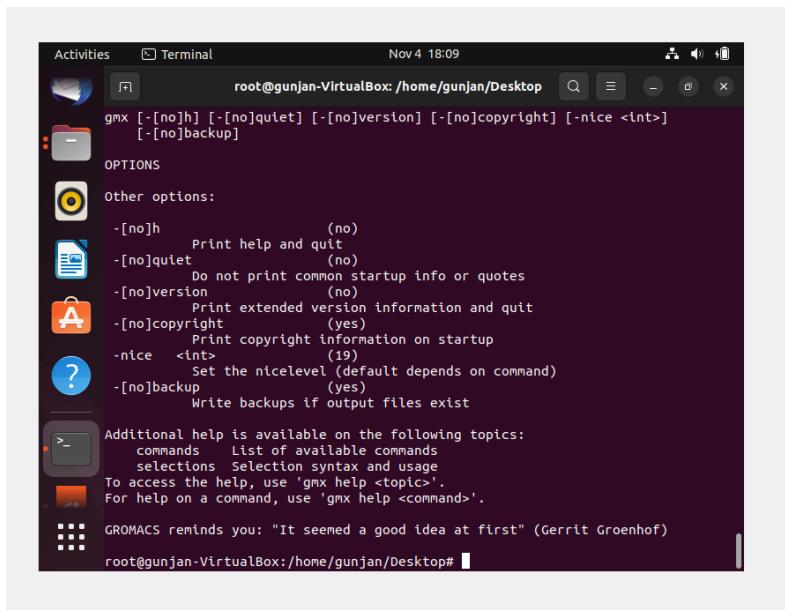


Antimicrobial peptide without water molecules in discovery studio visualizer.

Step 5: Install Gromacs in Ubuntu.

Step 6: Set the path of the peptide folder.

```
root@gunjan-VirtualBox:/home/gunjan/Desktop# gmx
:-) GROMACS - gmx, 2021.4-Ubuntu-2021.4-2 (-:
          GROMACS is written by:
Andrey Alekseenko          Emile Apol          Rossen Apostolov
          Paul Bauer           Herman J.C. Berendsen      Par Bjelkmar
          Christian Blau        Viacheslav Bolnykh       Kevin Boyd
          Aldert van Buuren     Rudi van Drunen        Anton Feenstra
          Gilles Gouaillardet   Alan Gray            Gerrit Groenhof
          Anca Hamuraru         Vincent Hindriksen    M. Eric Irrgang
          Aleksei Iupinov        Christoph Junghans   Joe Jordan
          Dimitrios Karkoulis   Peter Kasson         Jiri Kraus
          Carsten Kutzner       Per Larsson         Justin A. Lemkul
          Viveca Lindahl         Magnus Lundborg    Erik Marklund
          Pascal Merz           Pieter Meulenhoff   Teemu Murtola
          Szilard Pall           Sander Pronk        Roland Schulz
          Michael Shirts        Alexey Shvetsov    Alfons Sijbers
```



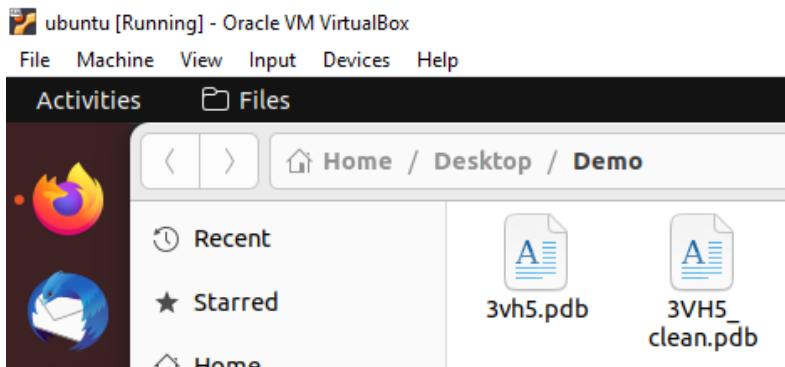
```
Activities Terminal Nov 4 18:09
root@gunjan-VirtualBox:/home/gunjan/Desktop
gmx [-[no]h] [-[no]quiet] [-[no]version] [-[no]copyright] [-nice <int>]
    [-[no]backup]

OPTIONS

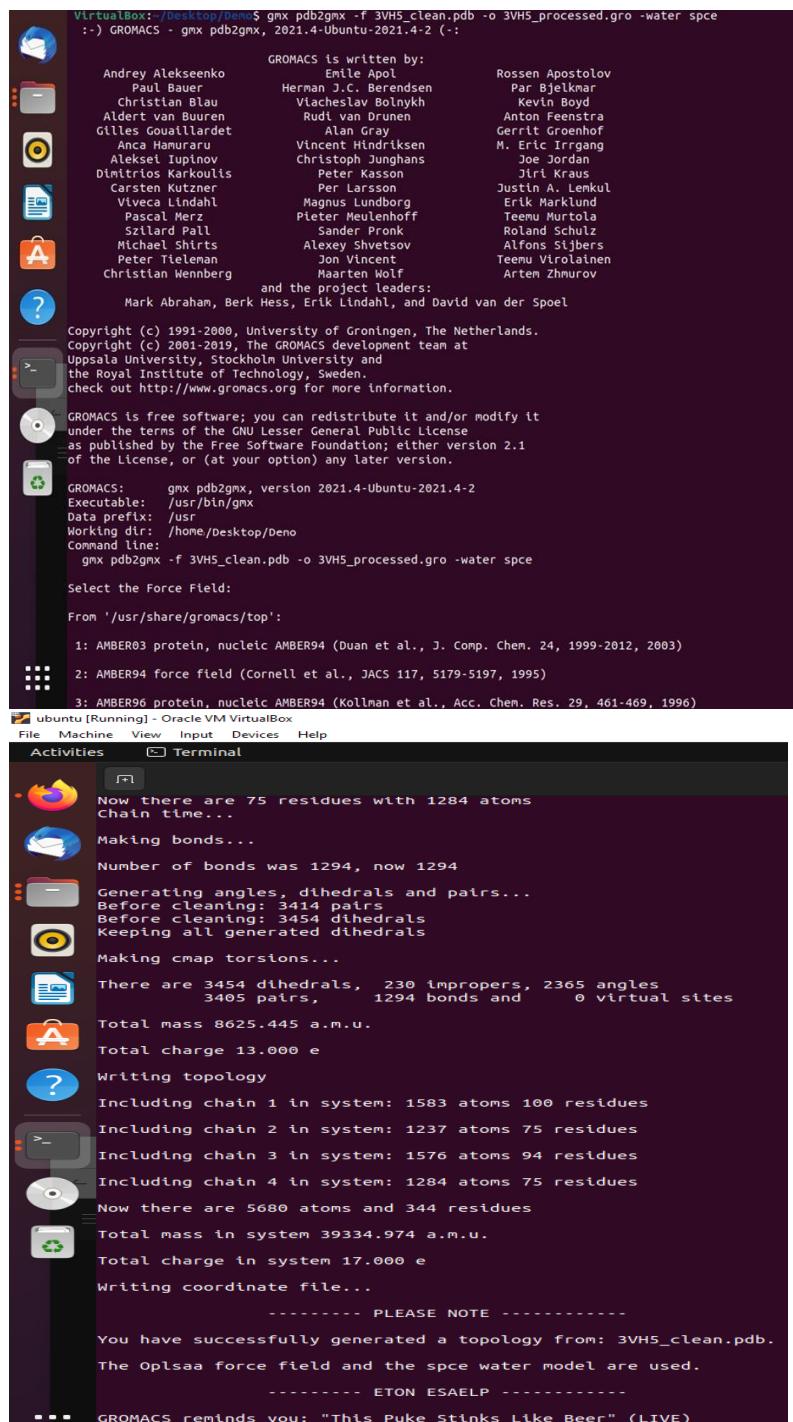
Other options:
-[no]h      (no)
-[no]quiet   (no)
-[no]version  (no)
-[no]copyright (yes)
-[no]copyright (yes)
-nice <int>  (19)
-set the nicelevel (default depends on command)
-[no]backup  (yes)
-[no]backup  (yes)

Additional help is available on the following topics:
  commands  List of available commands
  selections Selection syntax and usage
To access the help, use 'gmx help <topic>'.
For help on a command, use 'gmx help <command>'.

GROMACS reminds you: "It seemed a good idea at first" (Gerrit Groenhof)
root@gunjan-VirtualBox:/home/gunjan/Desktop#
```



Step 7: Than we will have to execute pdb2gmx by issuing a command also, the structure will be processed by pdb2gmx, and we will be prompted to choose a force field, where we select the force field as 15:



```

VirtualBox: /home/Desktop/ $ gmx pdb2gmx -f 3VH5_clean.pdb -o 3VH5_processed.gro -water spce
:-) GROMACS - gmx pdb2gmx, 2021.4-Ubuntu-2021.4-2 (-:
GROMACS is written by:
  Andrey Alekseenko      Emile Apol      Rossen Apostolov
  Paul Bauer             Herman J.C. Berendsen  Par Bjelkmar
  Christian Blau          Vlacheslav Bolyntkh  Kevin Boyd
  Albert van Buuren       Rudi van Drunen   Anton Feenstra
  Gilles Gouallardet     Alan Gray        Gerrit Groenhof
  Anca Hamararu          Vincent Hindriksen M. Eric Irrgang
  Aleksei Iupinov         Christoph Junghans  Joe Jordan
  Dimitrios Karkoulis    Peter Kasson     Jiri Kraus
  Carsten Kutzner        Per Larsson    Justin A. Lemkul
  Viveca Lindahl          Magnus Lundborg Erik Marklund
  Pascal Merz             Pieter Meulenhoff Teemu Murtola
  Szilard Pall            Sander Pronk   Roland Schulz
  Michael Shirts          Alexey Shvetsov Alfons Slijbers
  Peter Tieleman          Jon Vincent    Teemu Virolainen
  Christian Wennberg     Maarten Wolf   Artem Zmurov
                                and the project leaders:
                                Mark Abraham, Berk Hess, Erik Lindahl, and David van der Spoel

Copyright (c) 1991-2000, University of Groningen, The Netherlands.
Copyright (c) 2001-2019, The GROMACS development team at
Uppsala University, Stockholm University and
the Royal Institute of Technology, Sweden.
check out http://www.gromacs.org for more information.

GROMACS is free software; you can redistribute it and/or modify it
under the terms of the GNU Lesser General Public License
as published by the Free Software Foundation; either version 2.1
of the License, or (at your option) any later version.

GROMACS:  gmx pdb2gmx, version 2021.4-Ubuntu-2021.4-2
Executable: /usr/bin/gmx
Data prefix: /usr
Working dir: /home/Desktop/Deno
Command line:
  gmx pdb2gmx -f 3VH5_clean.pdb -o 3VH5_processed.gro -water spce

Select the Force Field:

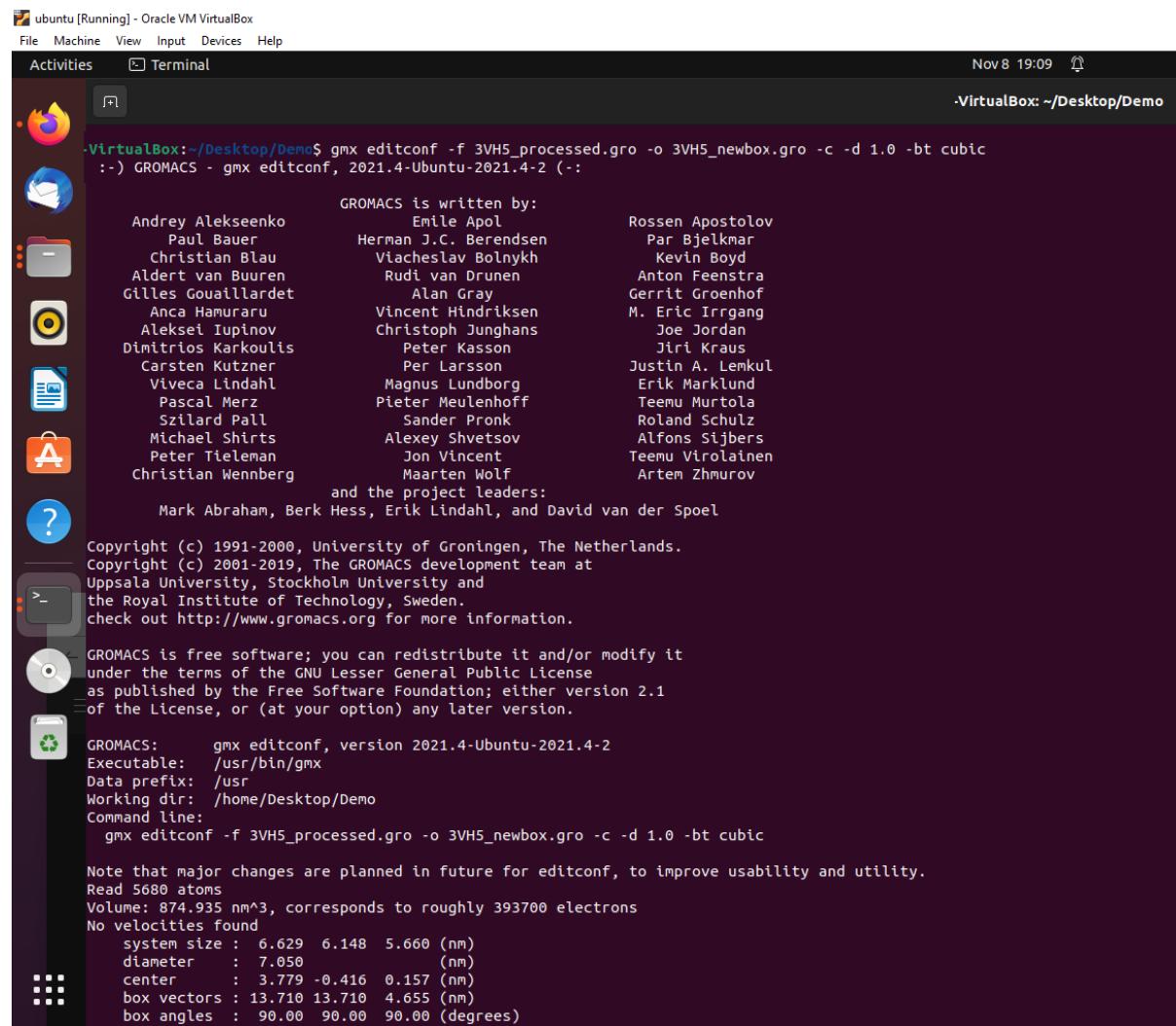
From '/usr/share/gromacs/top':
  1: AMBER03 protein, nucleic AMBER94 (Duan et al., J. Comp. Chem. 24, 1999-2012, 2003)
  2: AMBER94 force field (Cornell et al., JACS 117, 5179-5197, 1995)
  3: AMBER96 protein, nucleic AMBER94 (Kollman et al., Acc. Chem. Res. 29, 461-469, 1996)

ubuntu [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Activities Terminal
Now there are 75 residues with 1284 atoms
Chain time...
Making bonds...
Number of bonds was 1294, now 1294
Generating angles, dihedrals and pairs...
Before cleaning: 3414 pairs
Before cleaning: 3454 dihedrals
Keeping all generated dihedrals
Making cmap torsions...
There are 3454 dihedrals, 230 impropers, 2365 angles
  3405 pairs, 1294 bonds and 0 virtual sites
Total mass 8625.445 a.m.u.
Total charge 13.000 e
Writing topology
Including chain 1 in system: 1583 atoms 100 residues
Including chain 2 in system: 1237 atoms 75 residues
Including chain 3 in system: 1576 atoms 94 residues
Including chain 4 in system: 1284 atoms 75 residues
Now there are 5680 atoms and 344 residues
Total mass in system 39334.974 a.m.u.
Total charge in system 17.000 e
Writing coordinate file...
----- PLEASE NOTE -----
You have successfully generated a topology from: 3VH5_clean.pdb.
The Oplsaa force field and the spce water model are used.
----- ETON ESAELP -----
GROMACS reminds you: "This Puke Stinks Like Beer" (LIVE)
```

Step 8: So, after that the command which we execute centers the protein in the box (-c), and places it at least 1.0 nm from the box edge (-d 1.0).

Also, the box type is defined as a cube (-bt cubic) and we know that the distance to the edge of the box is an important parameter. Since we will be using periodic boundary conditions, we must satisfy the minimum image convention.

Now, that is, a protein should never see its periodic image, otherwise the forces calculated will be spurious. Specifying a solute-box distance of 1.0 nm will mean that there are at least 2.0 nm between any two periodic images of a protein. This distance will be sufficient for just about any cutoff scheme commonly used in simulation



The screenshot shows a Linux desktop environment with a dark theme. A terminal window is open in the center, displaying the output of a GROMACS command. The command is:

```
VirtualBox:~/Desktop/Demo$ gmx editconf -f 3VH5_processed.gro -o 3VH5_newbox.gro -c -d 1.0 -bt cubic
```

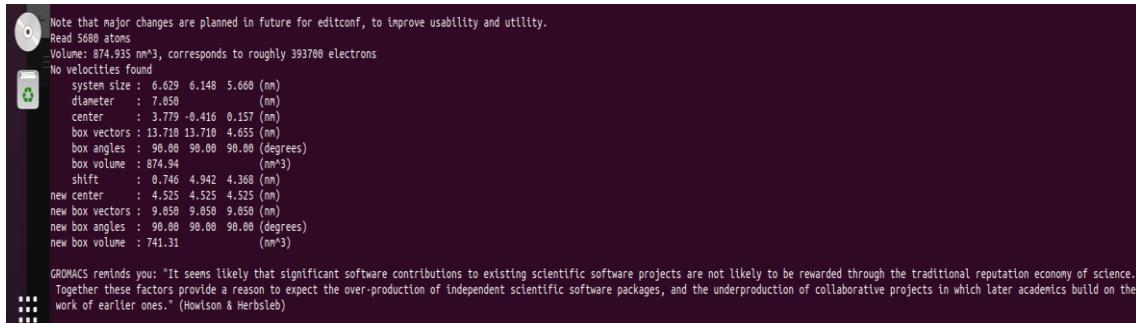
The terminal output includes the GROMACS copyright notice, a list of contributors, and the command-line parameters used. The command-line parameters are:

```
GROMACS:      gmx editconf, version 2021.4-Ubuntu-2021.4-2
Executable:   /usr/bin/gmx
Data prefix:  /usr
Working dir:  /home/Desktop/Demo
Command line:
  gmx editconf -f 3VH5_processed.gro -o 3VH5_newbox.gro -c -d 1.0 -bt cubic
```

Note that major changes are planned in future for editconf, to improve usability and utility.

Read 5680 atoms  
Volume: 874.935 nm<sup>3</sup>, corresponds to roughly 393700 electrons  
No velocities found

system size :	6.629	6.148	5.660	(nm)
diameter :	7.050			(nm)
center :	3.779	-0.416	0.157	(nm)
box vectors :	13.710	13.710	4.655	(nm)
box angles :	90.00	90.00	90.00	(degrees)



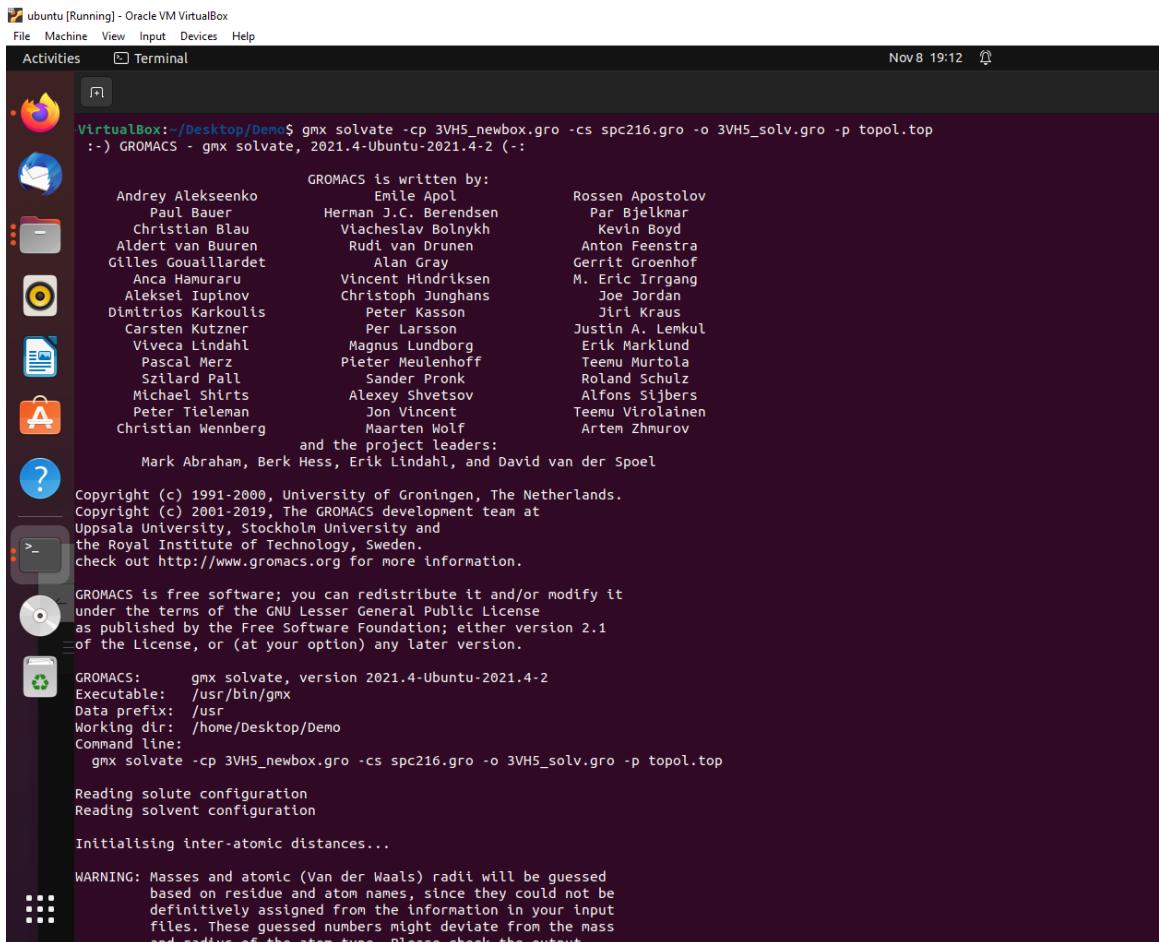
```

Note that major changes are planned in future for editconf, to improve usability and utility.
Read 5680 atoms
Volume: 874.935 nm3, corresponds to roughly 393700 electrons
No velocities found
  system size : 6.629 6.148 5.660 (nm)
  diameter   : 7.058          (nm)
  center     : 3.779 -0.416 0.157 (nm)
  box vectors : 13.710 13.710 4.655 (nm)
  box angles  : 90.00 90.00 90.00 (degrees)
  box volume  : 874.94          (nm3)
  shift       : 0.746 4.942 4.368 (nm)
  new center  : 4.525 4.525 4.525 (nm)
  new box vectors : 9.058 9.058 9.058 (nm)
  new box angles : 90.00 90.00 90.00 (degrees)
  new box volume : 741.31          (nm3)

GROMACS reminds you: "It seems likely that significant software contributions to existing scientific software projects are not likely to be rewarded through the traditional reputation economy of science. Together these factors provide a reason to expect the over-production of independent scientific software packages, and the underproduction of collaborative projects in which later academics build on the work of earlier ones." (Howison & Herbsleb)

```

Step 9: So, as the configuration of the protein (-cp) is contained in the output of the previous editconf step, and the configuration of the solvent (-cs) is part of the standard GROMACS installation. Also, we know that there are using spc216.gro, which is a generic equilibrated 3-point solvent model. Here we can use spc216.gro as the solvent configuration for SPC, SPC/E, or TIP3P water, since they are all three-point water models, and the output is called 1AKI\_solv.gro, and we tell solvate the name of the topology file (topol.top) so it can be modified. Note the changes to the [ molecules ] directive of topol.top.



```

ubuntu [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Activities Terminal Nov 8 19:12 ⓘ
VirtualBox:~/Desktop/Demo$ gmx solvate -cp 3VH5_newbox.gro -cs spc216.gro -o 3VH5_solv.gro -p topol.top
::) GROMACS - gmx solvate, 2021.4-Ubuntu-2021.4-2 (-:

GROMACS is written by:
  Andrey Alekseenko      Emile Apol      Rossen Apostolov
  Paul Bauer             Herman J.C. Berendsen  Par Bjelkmar
  Christian Blau         Viacheslav Bolnykh  Kevin Boyd
  Aldert van Buuren      Rudi van Drunen  Anton Feenstra
  Gilles Gouillardet    Alan Gray        Gerrit Groenhof
  Anca Hanuraru         Vincent Hindriksen M. Eric Irrgang
  Aleksei Iupinov        Christoph Junghans Joe Jordan
  Dimitrios Karkoulis    Peter Kasson    Jiri Kraus
  Carsten Kutzner        Per Larsson    Justin A. Lemkul
  Viveca Lindahl          Magnus Lundborg Erik Marklund
  Pascal Merz            Pieter Meulenhoff Teemu Murtola
  Szilard Pall           Sander Pronk  Roland Schulz
  Michael Shirts         Alexey Shvetsov Alfons Sijbers
  Peter Tielemans        Jon Vincent   Teemu Virolainen
  Christian Wennberg    Maarten Wolf   Artem Zhmurov
  and the project leaders:
  Mark Abraham, Berk Hess, Erik Lindahl, and David van der Spoel

Copyright (c) 1991-2000, University of Groningen, The Netherlands.
Copyright (c) 2001-2019, The GROMACS development team at
Uppsala University, Stockholm University and
the Royal Institute of Technology, Sweden.
check out http://www.gromacs.org for more information.

GROMACS is free software; you can redistribute it and/or modify it
under the terms of the GNU Lesser General Public License
as published by the Free Software Foundation; either version 2.1
of the License, or (at your option) any later version.

GROMACS:      gmx solvate, version 2021.4-Ubuntu-2021.4-2
Executable:   /usr/bin/gmx
Data prefix:  /usr
Working dir:  /home/Desktop/Demo
Command line:
  gmx solvate -cp 3VH5_newbox.gro -cs spc216.gro -o 3VH5_solv.gro -p topol.top

Reading solute configuration
Reading solvent configuration

Initialising inter-atomic distances...

WARNING: Masses and atomic (Van der Waals) radii will be guessed
based on residue and atom names, since they could not be
definitively assigned from the information in your input
files. These guessed numbers might deviate from the mass
and radius of the atom type. Please check the output.

```

```

Reading solute configuration
Reading solvent configuration
Initialising inter-atomic distances...
WARNING: Masses and atomic (Van der Waals) radii will be guessed
based on residue and atom names, since they could not be
definitively assigned from the information in your input
files. These guessed numbers might deviate from the mass
and radius of the atom type. Please check the output
files if necessary.

NOTE: From version 5.0 gmx solvate uses the Van der Waals radii
from the source below. This means the results may be different
compared to previous GROMACS versions.

++++ PLEASE READ AND CITE THE FOLLOWING REFERENCE +++
A. Bondi
van der Waals Volumes and Radii
J. Phys. Chem. 68 (1964) pp. 441-451
----- --- Thank You ----- -----

Generating solvent configuration
Will generate new solvent configuration of 5x5x5 boxes
Solvent box contains 81000 atoms in 27000 residues
Removed 8355 solvent atoms due to solvent-solvent overlap
Removed 530 solvent atoms due to solute-solvent overlap
Sorting configurations
Found 1 molecule type:
    SOL ( 3 atoms): 22446 residues
Generated solvent containing 67338 atoms in 22446 residues
Writing generated configuration to 3VHS_solv.gro

Output configuration contains 73018 atoms in 22790 residues
Volume           :    741.313 (nm3)
Density          :     1000.39 (g/l)
Number of solvent molecules: 22446

Processing topology
Adding line for 22446 solvent molecules with resname (SOL) to topology file (topol.top)

Back Off! I just backed up topol.top to ./#topol.top.1#
GROMACS reminds you: "Jesus Not Only Saves, He Also Frequently Makes Backups." (Myron Bradshaw)

```

Step 10: Below we here I have assemble our .tpr file with the following:

```

ubuntu [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Activities Terminal Nov 8 19:14
VirtualBox:~/Desktop/Demos$ gmx grompp -f ionsmdp -c 3VHS_solv.gro -p topol.top -o ions.tpr
:-) GROMACS - gmx grompp, 2021.4-Ubuntu-2021.4-2 (-:
GROMACS is written by:
Andrey Alekseenko      Emil Apol            Rossen Apostolov
Paul Bauer              Herman J.C. Berendsen  Par Bjelkmar
Christian Blau          Viacheslav Bolnykh   Kevin Boyd
Albert van Buuren        Rudi van Drunen       Anton Feenstra
Gilles Gouallardet      Alan Gray           Gerrit Groenhof
Anca Hammaru            Vincent Hindriksen  M. Eric Irrgang
Aleksel Iupinov          Christoph Jungnанс  Joe Jordan
Dimitrios Karkoulis      Peter Kasson         Jiri Kraus
Carsten Kutzner          Per Larsson          Justin A. Lenkul
Viveca Lindahl          Magnus Lundborg    Erik Marklund
Pascal Merz              Pieter Meulenhoff  Teemu Murtola
Szilard Pall             Sander Pronk         Roland Schulz
Michael Shirts           Alexey Shvetsov   Alfonso Sijbers
Peter Tieleman           Jon Vincent          Teemu Virolainen
Christian Wennberg      Maarten Wolf         Artem Zhmurov
                                         and the project leaders:
                                         Mark Abraham, Berk Hess, Erik Lindahl, and David van der Spoel
Copyright (c) 1991-2000, University of Groningen, The Netherlands.
Copyright (c) 2001-2019, The GROMACS development team at
Uppsala University, Stockholm University and
the Royal Institute of Technology, Sweden.
check out http://www.gromacs.org for more information.

GROMACS is free software; you can redistribute it and/or modify it
under the terms of the GNU Lesser General Public License
as published by the Free Software Foundation; either version 2.1
of the License, or (at your option) any later version.

GROMACS:      gmx grompp, version 2021.4-Ubuntu-2021.4-2
Executable:   /usr/bin/gmx
Data prefix:  /usr
Working dir:  /home/Desktop/Demo
Command line:
  gmx grompp -f ionsmdp -c 3VHS_solv.gro -p topol.top -o ions.tpr

Ignoring obsolete mdp entry 'ns_type'

NOTE 1 [file ionsmdp]:
With Verlet lists the optimal nstlist is >= 10, with GPUs >= 20. Note
that with the Verlet scheme, nstlist has no effect on the accuracy of
your simulation.

Setting the LD random seed to 2111301631

```

```

NOTE 1 [file constтоп]:
With Verlet lists the optimal nstlist is >= 10, with GPUs >= 20. Note
that with the Verlet scheme, nstlist has no effect on the accuracy of
your simulation.

Setting the LD random seed to 2111301631

Generated 330891 of the 330891 non-bonded parameter combinations
Generating 1-4 interactions: fudge = 0.5

Generated 330891 of the 330891 1-4 parameter combinations

Excluding 3 bonded neighbours molecule type 'Protein_chain_A'

Excluding 3 bonded neighbours molecule type 'Protein_chain_D'

Excluding 3 bonded neighbours molecule type 'Protein_chain_T'

Excluding 3 bonded neighbours molecule type 'Protein_chain_W'

Excluding 2 bonded neighbours molecule type 'SOL'

NOTE 2 [file topol.top, line 52]:
System has non-zero total charge: 17.000000
Total charge should normally be an integer. See
http://www.gromacs.org/Documentation/Floating_Point_Arithmetic
for discussion on how close it should be to an integer.

Analysing residue names:
There are: 344 Protein residues
There are: 22446 Water residues
Analysing Protein...
Number of degrees of freedom in T-Coupling group rest is 151713.00

NOTE 3 [file ions.mdp]:
You are using a plain Coulomb cut-off, which might produce artifacts.
You might want to consider using PME electrostatics.

This run will generate roughly 6 Mb of data
There were 3 notes

GROMACS reminds you: "This simulation is not as the former." (Malvolio, Act II, scene V of Shakespeare's Twelfth Night)

```

Step 11: So, after that now we have an atomic-level description of our system in the binary file ions.tpr. We will pass this file to genion:

```

ubuntu [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Activities Terminal Nov 8 19:17 ⓘ
VirtualBox:~/Desktop/Demo$ gmx genion -s ions.tpr -o 3VH5_solv_ions.gro -p topol.top -pname NA -nname CL -neutral
:-) GROMACS - gmx genion, 2021.4-Ubuntu-2021.4-2 (-:
GROMACS is written by:
Andrey Alekseenko Emile Apol Rossen Apostolov
Paul Bauer Herman J.C. Berendsen Par Bjelkmar
Christian Blau Viacheslav Bolynkh Kevin Boyd
Aldert van Buuren Rudt van Drunen Anton Feenstra
Gilles Gouillaudet Alan Gray Gerrit Groenhof
Anca Hamararu Vincent Hindsight M. Eric Irrgang
Aleksel Juphov Christoff Jungmans Joe Jordan
Dimitrios Karkoulis Peter Kasson Jirí Kraus
Carsten Kärner Per Larsson Justin A. Lemkul
Uteca Lindahl Magnus Lundberg Erik Marklund
Pascal Marzilier Pieter Neulendorff Teemu Niemala
Sjåstrand Päll Sandra Pronk Roland Schulz
Michael Shirts Alexey Sivakov Almas Sühers
Peter Tieleman Jon Vlent Teemu Virolainen
Christian Wennberg Maarten Wolf Arten Zhusarov
and the project leaders:
Mark Abraham, Berk Hess, Erik Lindahl, and David van der Spoel

Copyright (c) 1991-2000, University of Groningen, The Netherlands.
Copyright (c) 2001-2019, The GROMACS development team at
Uppsala University, Stockholm University and
the Royal Institute of Technology, Sweden.
check out http://www.gromacs.org for more information.

GROMACS is free software; you can redistribute it and/or modify it
under the terms of the GNU Lesser General Public License
as published by the Free Software Foundation; either version 2.1
of the License, or (at your option) any later version.

GROMACS:      gmx genion, version 2021.4-Ubuntu-2021.4-2
Executable:   /usr/bin/gmx
Data prefix:  /usr
Working dir:  /home/Desktop/Demo
Command line:
  gmx genion -s ions.tpr -o 3VH5_solv_ions.gro -p topol.top -pname NA -nname CL -neutral

Reading file ions.tpr, VERSION 2021.4-Ubuntu-2021.4-2 (single precision)
Reading file ions.tpr, VERSION 2021.4-Ubuntu-2021.4-2 (single precision)
Will try to add 0 NA ions and 17 CL ions.
Select a continuous group of solvent molecules
Group  0 (      System) has 73018 elements
Group  1 (    Protein) has 5680 elements
Group  2 (  Protein-H) has 2771 elements
Group  3 (    C-alpha) has 344 elements
Group  4 ( Backbone) has 1032 elements

```

```

Reading file ions.tpr, VERSION 2021.4-Ubuntu-2021.4-2 (single precision)
Reading file ions.tpr, VERSION 2021.4-Ubuntu-2021.4-2 (single precision)
Will try to add 0 NA ions and 17 CL ions.
Select a continuous group of solvent molecules
Group  0 (      System) has 73018 elements
Group  1 (      Protein) has 5680 elements
Group  2 (      Protein-H) has 2771 elements
Group  3 (      C-alpha) has 344 elements
Group  4 (      Backbone) has 1032 elements
Group  5 (      MainChain) has 1380 elements
Group  6 (      MainChain+Cb) has 1712 elements
Group  7 (      MainChain+H) has 1724 elements
Group  8 (      SideChain) has 3956 elements
Group  9 (      SideChain-H) has 1391 elements
Group 10 (      Prot-Masses) has 5680 elements
Group 11 (      non-Protein) has 67338 elements
Group 12 (      Water) has 67338 elements
Group 13 (      SOL) has 67338 elements
Group 14 (      non-Water) has 5680 elements
Select a group: 13
Selected 13: 'SOL'
Number of (3-atomic) solvent molecules: 22446

Processing topology
Replacing 17 solute molecules in topology file (topol.top) by 0 NA and 17 CL ions.

Back Off! I just backed up topol.top to ./#topol.top.2#
Using random seed -8814785.
Replacing solvent molecule 242 (atom 6406) with CL
Replacing solvent molecule 6493 (atom 25159) with CL
Replacing solvent molecule 6300 (atom 24580) with CL
Replacing solvent molecule 2131 (atom 12073) with CL
Replacing solvent molecule 17145 (atom 57115) with CL
Replacing solvent molecule 13176 (atom 45208) with CL
Replacing solvent molecule 8679 (atom 31717) with CL
Replacing solvent molecule 13207 (atom 45301) with CL
Replacing solvent molecule 13869 (atom 47287) with CL
Replacing solvent molecule 21210 (atom 69310) with CL
Replacing solvent molecule 3952 (atom 17536) with CL
Replacing solvent molecule 42 (atom 5806) with CL
Replacing solvent molecule 18548 (atom 61324) with CL
Replacing solvent molecule 7003 (atom 26689) with CL
Replacing solvent molecule 1225 (atom 9355) with CL
Replacing solvent molecule 17325 (atom 57655) with CL
Replacing solvent molecule 19773 (atom 64999) with CL

GROMACS reminds you: "Product of optimism and knowledge is a constant." (Lev Landau)

```

Step 12: Now we assemble the binary input using grompp using this input parameter file:

```

ubuntu [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Activities Terminal Nov 8 19:21 ⓘ
VirtualBox:-/Desktop/Demo$ gmx grompp -f minim.mdp -c 3VHS_solv_ions.gro -p topol.top -o em.tpr
:-) GROMACS - gmx grompp, 2021.4-Ubuntu-2021.4-2 (-:)

GROMACS is written by:
Andrey Alekseenko          Emile Apol           Rossen Apostolov
Paul Bauer                  Herman J.C. Berendsen   Par Bjelkmar
Christian Blau              Viacheslav Bolnykh     Kevin Boyd
Albert van Buuren            Rudi van Drunen       Anton Feenstra
Gilles Gouillaudet          Alan Gray           Gerrit Groenhof
Anca Hamararu              Vincent Hindrikxen   M. Eric Irrgang
Aleskel Iupinov             Christoph Junghans  Joe Jordan
Dimitrios Karkoulis          Peter Kasson         Jiri Kraus
Carsten Kutzner              Per Larsson        Justin A. Lenkul
Viveca Lindahl              Magnus Lundborg    Erik Marklund
Pascal Merz                  Pieter Meulenhoff   Teemu Murtola
Szilard Pall                 Sander Pronk       Roland Schulz
Michael Shirts               Alexey Shvetsov    Alfons Sijbers
Peter Tieleman               Jon Vincent        Teemu Virolainen
Christian Wennberg          Maarten Wolf       Artem Zhmurov
and the project leaders:
Mark Abraham, Berk Hess, Erik Lindahl, and David van der Spoel

Copyright (c) 1991-2000, University of Groningen, The Netherlands.
Copyright (c) 2001-2019, The GROMACS development team at
Uppsala University, Stockholm University and
the Royal Institute of Technology, Sweden.
check out http://www.gromacs.org for more information.

GROMACS is free software; you can redistribute it and/or modify it
under the terms of the GNU Lesser General Public License
as published by the Free Software Foundation; either version 2.1
of the License, or (at your option) any later version.

GROMACS:   gmx grompp, version 2021.4-Ubuntu-2021.4-2
Executable: /usr/bin/gmx
Data prefix: /usr
Working dir: /home/Desktop/Demo
Command line:
  gmx grompp -f minim.mdp -c 3VHS_solv_ions.gro -p topol.top -o em.tpr

Ignoring obsolete mdp entry 'ns_type'

NOTE 1 [file minim.mdp]:
With Verlet lists the optimal nstlist is >= 10, with GPUs >= 20. Note
that with the Verlet scheme, nstlist has no effect on the accuracy of
your simulation.

Setting the LD random seed to 2130698239
Generated 330891 of the 330891 non-bonded parameter combinations

```

```

Ignoring obsolete mdp entry 'ns_type'

NOTE 1 [file minimmdp]:
With Verlet lists the optimal nstlist is >= 10, with GPUs >= 20. Note
that with the Verlet scheme, nstlist has no effect on the accuracy of
your simulation.

Setting the LD random seed to 2130698239

Generated 330891 of the 330891 non-bonded parameter combinations
Generating 1-4 interactions: fudge = 0.5

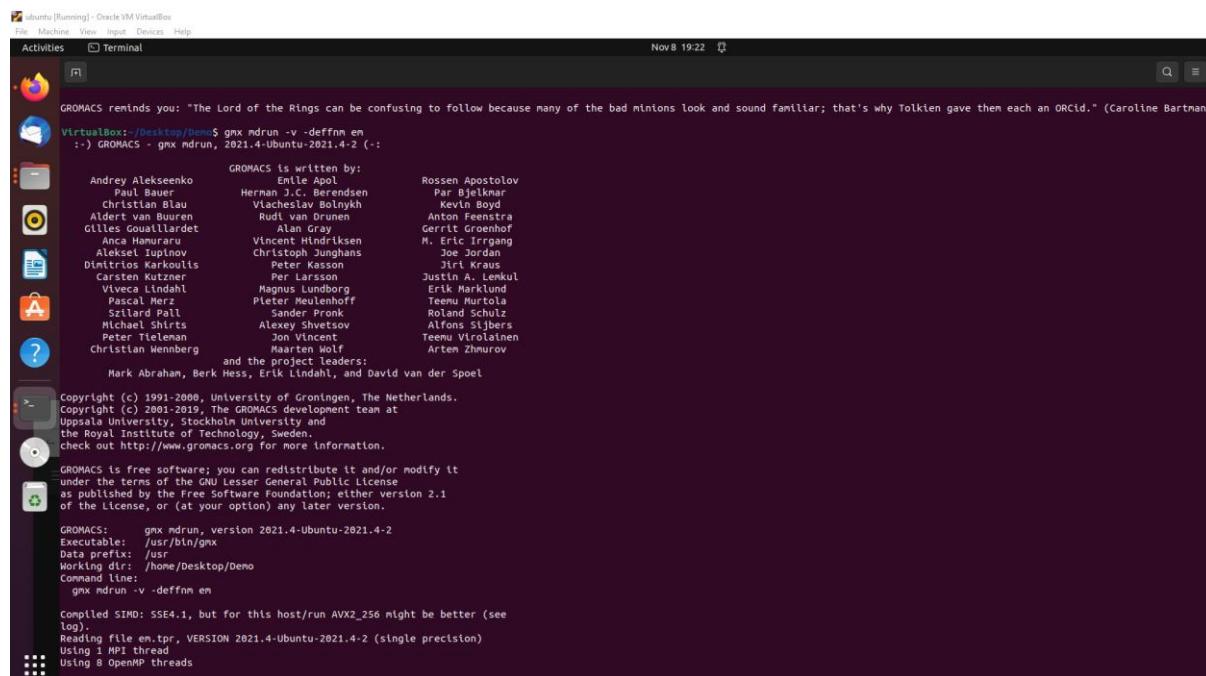
Generated 330891 of the 330891 1-4 parameter combinations

Excluding 3 bonded neighbours molecule type 'Protein_chain_A'
Excluding 3 bonded neighbours molecule type 'Protein_chain_D'
Excluding 3 bonded neighbours molecule type 'Protein_chain_T'
Excluding 3 bonded neighbours molecule type 'Protein_chain_W'
Excluding 2 bonded neighbours molecule type 'SOL'
Excluding 1 bonded neighbours molecule type 'CL'
Analysing residue names:
There are: 344 Protein residues
There are: 22429 Water residues
There are: 17 Ion residues
Analysing Protein...
Analysing residues not classified as Protein/DNA/RNA/Water and splitting into groups...
Number of degrees of freedom in T-Coupling group rest is 151662.00
Calculating Fourier grid dimensions for X Y Z
Using a Fourier grid of 80x80x80, spacing 0.113 0.113 0.113
Estimate for the relative computational load of the PME mesh part: 0.19
This run will generate roughly 6 Mb of data
There was 1 note

GROMACS reminds you: "The Lord of the Rings can be confusing to follow because many of the bad minions look and sound familiar; that's why Tolkien gave them each an ORCID." (Caroline Bartman)

```

## Step 13: Also, now we are ready to invoke mdrun to carry out the EM:



```

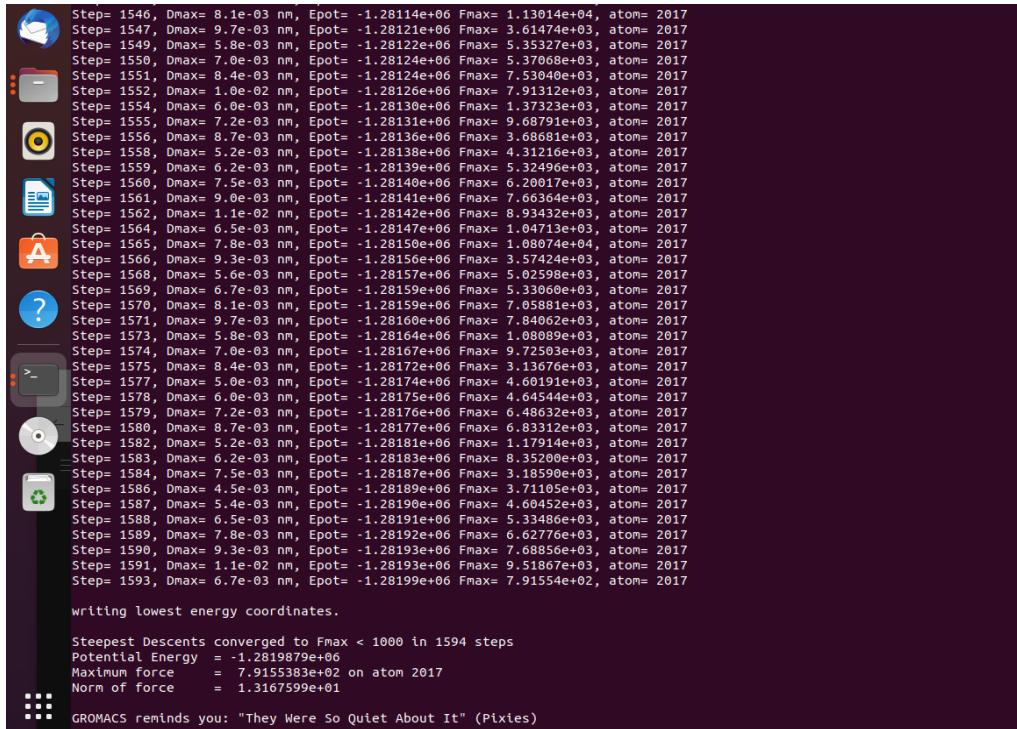
Ubuntu [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Activities Terminal Nov 8 19:22
GROMACS reminds you: "The Lord of the Rings can be confusing to follow because many of the bad minions look and sound familiar; that's why Tolkien gave them each an ORCID." (Caroline Bartman)
VirtualBox:~/Desktop/Demo$ gmx mdrun -v -deffnm em
:-) GROMACS - gmx mdrun, 2021.4-Ubuntu-2021.4-2 (-:
GROMACS is written by:
Andrey Alekseenko Emile Apol Rossen Apostolov
Paul Bauer Herman J.C. Berendsen Par Bjelkmar
Christian Blau Vlacheslav Bolnykh Kevin Boyd
Albert van Buuren Rudi van Drunen Anton Feenstra
Gilles Cauquilardet Alan Gray Gerrit Groenhof
Andrea Giuffrida Vincent Jonckheere M. Erik Jorgenson
Aleksei Iupinov Christopher Junghans Joe Jordan
Dimitrios Karkoulis Peter Kasson Jiri Kraus
Carsten Kutzman Per Larsson Justin A. Lemkul
Viveca Lindahl Magnus Lundborg Erik Marklund
Pascal Merz Pieter Meulenhoff Teemu Murtola
Szilard Pali Sander Pronk Roland Schulz
Michael Shirts Alexey Savchenko Alfonso Silbers
Peter Tieleman Jon Vinkent Teemu Virolainen
Christian Wenneberg Marten Wolf Artem Zhusarov
and the project leaders:
Mark Abraham, Berk Hess, Erik Lindahl, and David van der Spoel
Copyright (c) 1991-2006 University of Groningen, The Netherlands.
Copyright (c) 2007-2021 GROMACS development team at
Uppsala University, Stockholm University and
the Royal Institute of Technology, Sweden.
check out http://www.gromacs.org for more information.

GROMACS is free software; you can redistribute it and/or modify it
under the terms of the GNU Lesser General Public License
as published by the Free Software Foundation; either version 2.1
of the License, or (at your option) any later version.

GROMACS:   gmx mdrun, version 2021.4-Ubuntu-2021.4-2
Executable: /usr/bin/gmx
Data prefix: /usr
Working dir: /home/Desktop/Demo
Command line:
gmx mdrun -v -deffnm em

Compiled SIMD: SSE4.1, but for this host/run AVX2_256 might be better (see
log).
Reading file em.tpr, VERSION 2021.4-Ubuntu-2021.4-2 (single precision)
Using 1 MPI thread
Using 8 OpenMP threads

```



```

Step= 1546, Dmax= 8.1e-03 nm, Epot= -1.28114e+06 Fmax= 1.13014e+04, atom= 2017
Step= 1547, Dmax= 9.7e-03 nm, Epot= -1.28121e+06 Fmax= 3.61474e+03, atom= 2017
Step= 1549, Dmax= 5.8e-03 nm, Epot= -1.28122e+06 Fmax= 5.35327e+03, atom= 2017
Step= 1550, Dmax= 7.0e-03 nm, Epot= -1.28124e+06 Fmax= 5.37068e+03, atom= 2017
Step= 1551, Dmax= 8.4e-03 nm, Epot= -1.28124e+06 Fmax= 7.53040e+03, atom= 2017
Step= 1552, Dmax= 1.0e-02 nm, Epot= -1.28126e+06 Fmax= 7.91312e+03, atom= 2017
Step= 1554, Dmax= 6.0e-03 nm, Epot= -1.28130e+06 Fmax= 1.37323e+05, atom= 2017
Step= 1555, Dmax= 7.2e-03 nm, Epot= -1.28131e+06 Fmax= 9.68791e+03, atom= 2017
Step= 1556, Dmax= 8.7e-03 nm, Epot= -1.28136e+06 Fmax= 3.68681e+03, atom= 2017
Step= 1558, Dmax= 5.2e-03 nm, Epot= -1.28138e+06 Fmax= 4.31216e+03, atom= 2017
Step= 1559, Dmax= 6.2e-03 nm, Epot= -1.28139e+06 Fmax= 5.32496e+03, atom= 2017
Step= 1560, Dmax= 7.5e-03 nm, Epot= -1.28140e+06 Fmax= 6.20017e+03, atom= 2017
Step= 1561, Dmax= 9.0e-03 nm, Epot= -1.28141e+06 Fmax= 7.66364e+03, atom= 2017
Step= 1562, Dmax= 1.1e-02 nm, Epot= -1.28142e+06 Fmax= 8.93432e+03, atom= 2017
Step= 1564, Dmax= 6.5e-03 nm, Epot= -1.28147e+06 Fmax= 1.04713e+03, atom= 2017
Step= 1565, Dmax= 7.8e-03 nm, Epot= -1.28150e+06 Fmax= 1.08074e+04, atom= 2017
Step= 1566, Dmax= 9.3e-03 nm, Epot= -1.28156e+06 Fmax= 3.57424e+03, atom= 2017
Step= 1568, Dmax= 5.0e-03 nm, Epot= -1.28157e+06 Fmax= 5.02598e+03, atom= 2017
Step= 1569, Dmax= 6.7e-03 nm, Epot= -1.28159e+06 Fmax= 5.33060e+03, atom= 2017
Step= 1570, Dmax= 8.1e-03 nm, Epot= -1.28159e+06 Fmax= 7.05881e+03, atom= 2017
Step= 1571, Dmax= 9.7e-03 nm, Epot= -1.28160e+06 Fmax= 7.84062e+03, atom= 2017
Step= 1573, Dmax= 5.8e-03 nm, Epot= -1.28164e+06 Fmax= 1.08089e+03, atom= 2017
Step= 1574, Dmax= 7.0e-03 nm, Epot= -1.28167e+06 Fmax= 9.72503e+03, atom= 2017
Step= 1575, Dmax= 8.4e-03 nm, Epot= -1.28172e+06 Fmax= 3.13676e+03, atom= 2017
Step= 1577, Dmax= 5.0e-03 nm, Epot= -1.28174e+06 Fmax= 4.60191e+03, atom= 2017
Step= 1578, Dmax= 6.0e-03 nm, Epot= -1.28175e+06 Fmax= 4.64544e+03, atom= 2017
Step= 1579, Dmax= 7.2e-03 nm, Epot= -1.28176e+06 Fmax= 6.48632e+03, atom= 2017
Step= 1580, Dmax= 8.7e-03 nm, Epot= -1.28177e+06 Fmax= 6.83312e+03, atom= 2017
Step= 1582, Dmax= 5.2e-03 nm, Epot= -1.28181e+06 Fmax= 1.17914e+03, atom= 2017
Step= 1583, Dmax= 6.2e-03 nm, Epot= -1.28183e+06 Fmax= 8.35208e+03, atom= 2017
Step= 1584, Dmax= 7.5e-03 nm, Epot= -1.28187e+06 Fmax= 3.18596e+03, atom= 2017
Step= 1586, Dmax= 4.5e-03 nm, Epot= -1.28189e+06 Fmax= 3.71105e+03, atom= 2017
Step= 1587, Dmax= 5.4e-03 nm, Epot= -1.28190e+06 Fmax= 4.60452e+03, atom= 2017
Step= 1588, Dmax= 6.5e-03 nm, Epot= -1.28191e+06 Fmax= 5.33486e+03, atom= 2017
Step= 1589, Dmax= 7.8e-03 nm, Epot= -1.28192e+06 Fmax= 6.62776e+03, atom= 2017
Step= 1590, Dmax= 9.3e-03 nm, Epot= -1.28193e+06 Fmax= 7.68856e+03, atom= 2017
Step= 1591, Dmax= 1.1e-02 nm, Epot= -1.28193e+06 Fmax= 9.51867e+03, atom= 2017
Step= 1593, Dmax= 6.7e-03 nm, Epot= -1.28199e+06 Fmax= 7.91554e+02, atom= 2017

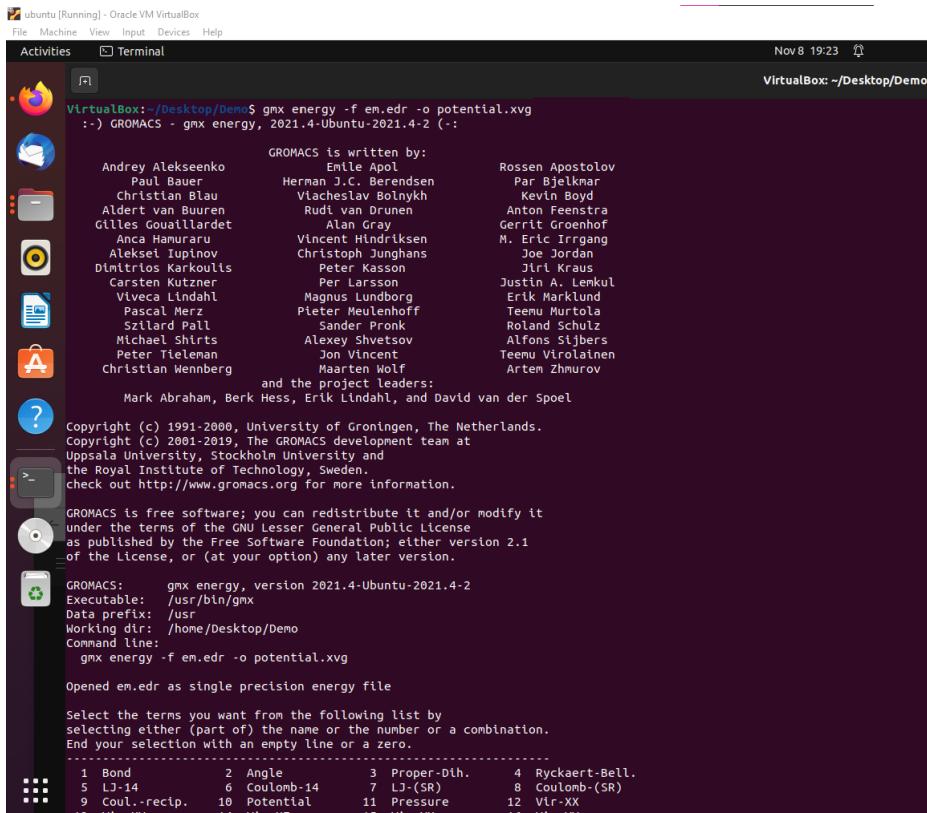
writing lowest energy coordinates.

Steepest Descents converged to Fmax < 1000 in 1594 steps
Potential Energy = -1.2819879e+06
Maximum Force = 7.9155383e+02 on atom 2017
Norm of Force = 1.3167599e+01

GROMACS reminds you: "They Were So Quiet About It" (Pixies)

```

Step 14: The em.edr file contains all of the energy terms that GROMACS collects during EM. So now we can analyze any .edr file using the GROMACS energy module.



```

ubuntu [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
Activities Terminal Nov 8 19:23
VirtualBox:~/Desktop/Demo$ gmx energy -f em.edr -o potential.xvg
:-) GROMACS - gmx energy, 2021.4-Ubuntu-2021.4-2 (::

GROMACS is written by:
Andrey Alekseenko Emile Apol Rossen Apostolov
Paul Bauer Herman J.C. Berendsen Par Bjelkmar
Christian Blau Vlacheslav Bolnykh Kevin Boyd
Albert van Buuren Rudi van Drunen Anton Feenstra
Gilles Gouallardet Alan Gray Gerrit Groenhof
Anca Hamuraru Vincent Hindriksen M. Eric Irrgang
Aleksei Iupinov Christoph Junghans Joe Jordan
Dmitrios Karkoulis Peter Kasson Jiri Kraus
Carsten Kutzner Per Larsson Justin A. Lemkul
Viveca Lindahl Magnus Lundborg Erik Marklund
Pascal Merz Pieter Meulenhoff Teemu Murtola
Szilard Pall Sander Pronk Roland Schulz
Michael Shirts Alexey Shvetsov Alfons Stijbers
Peter Tieleman Jon Vincent Teemu Virolainen
Christian Wennberg Maarten Wolf Artem Zhuarov
and the project leaders:
Mark Abraham, Berk Hess, Erik Lindahl, and David van der Spoel

Copyright (c) 1991-2000, University of Groningen, The Netherlands.
Copyright (c) 2001-2019, The GROMACS development team at
Uppsala University, Stockholm University and
the Royal Institute of Technology, Sweden.
check out http://www.gromacs.org for more information.

GROMACS is free software; you can redistribute it and/or modify it
under the terms of the GNU Lesser General Public License
as published by the Free Software Foundation; either version 2.1
of the License, or (at your option) any later version.

GROMACS: gmx energy, version 2021.4-Ubuntu-2021.4-2
Executable: /usr/bin/gmx
Data prefix: /usr
Working dir: /home/Desktop/Demo
Command line:
gmx energy -f em.edr -o potential.xvg

Opened em.edr as single precision energy file

Select the terms you want from the following list by
selecting either (part of) the name or the number or a combination.
End your selection with an empty line or a zero.

1 Bond 2 Angle 3 Proper-Dih. 4 Ryckaert-Bell.
5 LJ-14 6 Coulomb-14 7 LJ-(SR) 8 Coulomb-(SR)
9 Coul.-recip. 10 Potential 11 Pressure 12 Vlr-XX
13 Vir-XY 14 Vir-XZ 15 Vir-YX 16 Vir-YY

```

```

? Select the terms you want from the following list by
? selecting either (part of) the name or the number or a combination.
? End your selection with an empty line or a zero.

1 Bond      2 Angle      3 Proper-Dih.   4 Ryckaert-Bell.
5 LJ-14     6 Coulomb-14 7 LJ-(SR)      8 Coulomb-(SR)
9 Coul.-recip. 10 Potential 11 Pressure 12 Vir-XX
13 Vir-XY    14 Vir-XZ    15 Vir-YX    16 Vir-YY
17 Vir-YZ    18 Vir-ZX    19 Vir-ZY    20 Vir-ZZ
21 Pres-XX   22 Pres-XY   23 Pres-XZ   24 Pres-YX
25 Pres-YY   26 Pres-YZ   27 Pres-ZX   28 Pres-ZY
29 Pres-ZZ   30 #Surf*SurfTen 31 T-rest

10
0
Last energy frame read 1261 time 1593.000

Statistics over 1594 steps [ 0.0000 through 1593.0000 ps ], 1 data sets
All statistics are over 1262 points (frames)

Energy          Average   Err.Est.      RMSD  Tot-Drift
-----
Potential      -1.24445e+06    19000    50092.5   -126860 (kJ/mol)

GROMACS reminds you: "I always think there is something foreign about jolly phrases at breakfast." (Mr. Carson in Downtown Abbey)

```

Step 15: Also, once we are done with the above step, we run the following code to install grace so that we can analyze the Energy Minimization Graph.

```

VirtualBox:/Desktop/Demo$ sudo apt install grace
[sudo] password for harsh:
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
The following package was automatically installed and is no longer required:
  systemd-hwe-hwdb
Use 'sudo apt autoremove' to remove it.
The following additional packages will be installed:
  gconf-service gconf-service-backend gconf2 gconf2-common gsfonts libaerc0
  libgconf-2-4 libhdfs5-103.1 libhdfs5-hl-100 libmotif-common libnetcdf19 libs2z
  libutempter0 libxbae4m libxm4 libxhtml1.1 xfonts-100dpi xterm
Suggested packages:
  gconf-defaults-service texlive-extra-utils xfonts-cyrillic
The following NEW packages will be installed:
  gconf-service gconf-service-backend gconf2 gconf2-common grace gsfonts
  libaerc0 libgconf-2-4 libhdfs5-103.1 libhdfs5-hl-100 libmotif-common
  libnetcdf19 libs2z libutempter0 libxbae4m libxm4 libxhtml1.1 xfonts-100dpi
xterm
0 upgraded, 19 newly installed, 0 to remove and 171 not upgraded.
Need to get 13.3 MB of archives.
After this operation, 34.1 MB of additional disk space will be used.
Do you want to continue? [Y/n] y
Get:1 http://in.archive.ubuntu.com/ubuntu jammy/universe amd64 libmotif-common all 2.3.8-3 [11.0 kB]
Get:2 http://in.archive.ubuntu.com/ubuntu jammy/universe amd64 libxm4 amd64 2.3.8-3 [1,001 kB]
Get:3 http://in.archive.ubuntu.com/ubuntu jammy/universe amd64 gconf2-common all 3.2.6-7ubuntu2 [698 kB]
Get:4 http://in.archive.ubuntu.com/ubuntu jammy/universe amd64 libgconf-2-4 amd64 3.2.6-7ubuntu2 [86.0 kB]
Get:5 http://in.archive.ubuntu.com/ubuntu jammy/universe amd64 gconf-service-backend amd64 3.2.6-7ubuntu2 [59.3 kB]
Get:6 http://in.archive.ubuntu.com/ubuntu jammy/universe amd64 gconf-service amd64 3.2.6-7ubuntu2 [17.4 kB]
Get:7 http://in.archive.ubuntu.com/ubuntu jammy/universe amd64 gconf2 amd64 3.2.6-7ubuntu2 [83.9 kB]
Get:8 http://in.archive.ubuntu.com/ubuntu jammy/universe amd64 gsfonts all 1:8.11+urwcyr1.0.7-pre44-4.5 [3,120 kB]
Get:9 http://in.archive.ubuntu.com/ubuntu jammy/main amd64 libutempter0 amd64 1.2.1-2build2 [8,848 B]
Get:10 http://in.archive.ubuntu.com/ubuntu jammy/universe amd64 xterm amd64 372-1ubuntu1 [857 kB]
Get:11 http://in.archive.ubuntu.com/ubuntu jammy/universe amd64 libaerc0 amd64 1.0.6-1 [20.1 kB]
Get:12 http://in.archive.ubuntu.com/ubuntu jammy/universe amd64 libs2z amd64 1.0.6-1 [5,354 B]
Get:13 http://in.archive.ubuntu.com/ubuntu jammy/universe amd64 libhdfs5-103.1 amd64 1.10.7+repack-4ubuntu2 [1,295 kB]
Get:14 http://in.archive.ubuntu.com/ubuntu jammy/universe amd64 libhdfs5-hl-100 amd64 1.10.7+repack-4ubuntu2 [59.1 kB]
Get:15 http://in.archive.ubuntu.com/ubuntu jammy/universe amd64 libnetcdf19 amd64 1:4.8.1-1 [456 kB]
Get:16 http://in.archive.ubuntu.com/ubuntu jammy/universe amd64 libxbae4m amd64 4.60.4-9 [106 kB]
Get:17 http://in.archive.ubuntu.com/ubuntu jammy/universe amd64 libxhtml1.1 amd64 1.1.10-4 [212 kB]
Get:18 http://in.archive.ubuntu.com/ubuntu jammy/universe amd64 grace amd64 1:5.1.25-12build1 [1,391 kB]
Get:19 http://in.archive.ubuntu.com/ubuntu jammy/universe amd64 xfonts-100dpi all 1:1.0.4+nmu1.1 [3,818 kB]

```

```

Selecting previously unselected package libmetacity1:amd64.
Preparing to unpack .../libmetacity1:amd64_1:4.8.1+1~.1.10-4_amd64.deb ...
Selecting previously unselected package libbaize1:amd64 ...
Selecting previously unselected package libbaize0:amd64 ...
Unpacking libbaize4:amd64 (4.09.4.9) ...
Preparing to unpack .../libbaize1:amd64_1:4.8.1+1~.1.10-4_amd64 ...
Unpacking libbaize0:amd64 (4.09.4.9) ...
Preparing to unpack .../libbaizetl1:amd64_1:4.8.1+1~.1.10-4_amd64 ...
Unpacking libbaizetl1:amd64 (1.1.10-4) ...
Preparing to unpack .../libfontconfig1:amd64_2.13.1-2ubuntu1 ...
Preparing to unpack .../libfontconfig1:amd64_2.13.1-2ubuntu1 ...
Unpacking fontconfig (2.13.1-25.12build1) ...
Preparing to unpack .../libgrace1:amd64_1:5.1.25-12build1 ...
Unpacking grace (1:5.1.25-12build1) ...
Selecting previously unselected package fonts-100dpi ...
Preparing to unpack .../fonts-100dpi_1:9.0.4+mmui.1 ...
Unpacking xfonts-100dpi_1:9.0.4+mmui.1 ...
Setting up libmetacity-common (3.2.6-3) ...
Setting up libbaize0:amd64 (3.2.6-7ubuntu2) ...
Setting up libbaize1:amd64 (3.2.6-7ubuntu2) ...
Creating config file /etc/gconf/2/path with new version
Setting up gfonts (1:8.1+urxvtc1.0.7-pre44-4.5) ...
Setting up libutempter0:amd64 (1.2.1-2ubuntu1) ...
Setting up libbz2:amd64 (1.0.6.1) ...
Setting up libxml2:amd64 (2.9.3+dfsg-1) ...
Setting up libxslt1.1:amd64 (1.1.29-1) ...
Setting up libxslt1.1:amd64 (1.1.10-4) ...
Setting up libxmlhtml1:1:amd64 (1:1.10-4) ...
Setting up libxmlhtml1:1:amd64 (1:1.10-7) ...
Setting up libxmlhtml1:1:amd64 (1:1.10-7+repack-kubuntu2) ...
Setting up libmetacity1:amd64 (1:4.8.1+1) ...
Processing triggers for libshared-mime-info (2.1-2) ...
Processing triggers for sgml-base (1.39) ...
Processing triggers for libfontconfig1:amd64 (2.13.1-2ubuntu1) ...
Processing triggers for fontconfig (2.13.1-2ubuntu1) ...
Processing triggers for desktop-file-utils (0.26-1ubuntu3) ...
Setting up libgconf-2-4:amd64 (3.2.6-7ubuntu2) ...
Processing triggers for gnome-menus (3.36.0-1ubuntu3) ...
Processing triggers for libgl1-mesa-glx (2.6.0-1ubuntu2) ...
Setting up gconf-service-backend (3.2.6-7ubuntu2) ...
Setting up powerdevil (1.2.6-7ubuntu2) ...
Setting up libgudev-1.0-0:amd64 (1:3.16.6-1ubuntu2) ...
Setting up grace (1:5.1.25-12build1) ...

```

Step 16: So, after that we have to just run xmgrace the file for the graph has been generated in the folder.

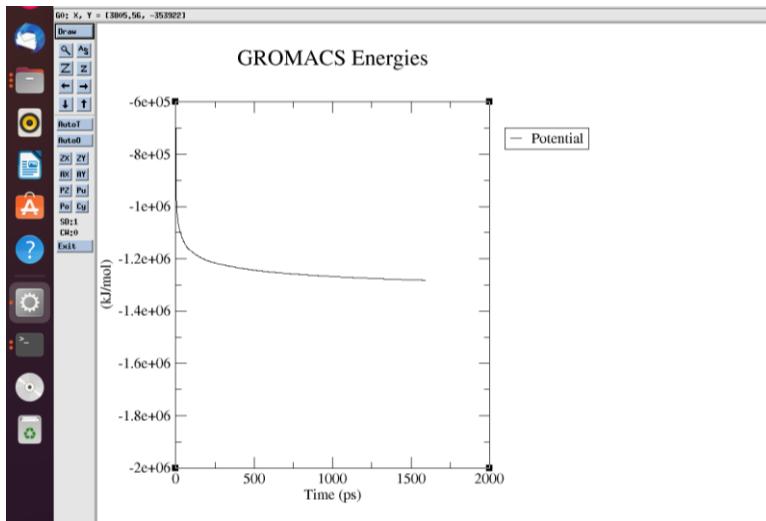
```

Ubuntu:~$ cd /Desktop/Demo
Ubuntu:~/Desktop/Demo$ ./xmgrace
Nov 8 19:25 VirtualBox:~/Desktop/Demo

Actions Terminal Nov 8 19:25 VirtualBox:~/Desktop/Demo

Selecting previously unselected package libmetacity1:amd64.
Preparing to unpack .../libmetacity1:amd64_1:4.8.1+1~.1.10-4_amd64 ...
Unpacking libmetacity1:amd64_1:4.8.1+1~.1.10-4_amd64 ...
Selecting previously unselected package libbaize1:amd64 ...
Selecting previously unselected package libbaize0:amd64 ...
Unpacking libbaize4:amd64 (4.09.4.9) ...
Preparing to unpack .../libbaize1:amd64_1:4.8.1+1~.1.10-4_amd64 ...
Unpacking libbaize0:amd64 (4.09.4.9) ...
Preparing to unpack .../libbaizetl1:amd64_1:4.8.1+1~.1.10-4_amd64 ...
Unpacking libbaizetl1:amd64 (1.1.10-4) ...
Preparing to unpack .../libfontconfig1:amd64_2.13.1-2ubuntu1 ...
Preparing to unpack .../libfontconfig1:amd64_2.13.1-2ubuntu1 ...
Unpacking fontconfig (2.13.1-25.12build1) ...
Preparing to unpack .../libgrace1:amd64_1:5.1.25-12build1 ...
Unpacking grace (1:5.1.25-12build1) ...
Selecting previously unselected package fonts-100dpi ...
Preparing to unpack .../fonts-100dpi_1:9.0.4+mmui.1 ...
Unpacking xfonts-100dpi_1:9.0.4+mmui.1 ...
Setting up libmetacity-common (3.2.6-3) ...
Setting up libbaize0:amd64 (3.2.6-7ubuntu2) ...
Creating config file /etc/gconf/2/path with new version
Setting up gfonts (1:8.1+urxvtc1.0.7-pre44-4.5) ...
Setting up libutempter0:amd64 (1.2.1-2ubuntu1) ...
Setting up libbz2:amd64 (1.0.6.1) ...
Setting up libxml2:amd64 (2.9.3+dfsg-1) ...
Setting up libxslt1.1:amd64 (1.1.29-1) ...
Setting up libxslt1.1:amd64 (1.1.10-4) ...
Setting up libxmlhtml1:1:amd64 (1:1.10-4) ...
Setting up libxmlhtml1:1:amd64 (1:1.10-7) ...
Setting up libxmlhtml1:1:amd64 (1:1.10-7+repack-kubuntu2) ...
Setting up libmetacity1:amd64 (1:4.8.1+1) ...
Processing triggers for libshared-mime-info (2.1-2) ...
Processing triggers for sgml-base (1.39) ...
Processing triggers for libfontconfig1:amd64 (2.13.1-2ubuntu1) ...
Processing triggers for fontconfig (2.13.1-2ubuntu1) ...
Processing triggers for desktop-file-utils (0.26-1ubuntu3) ...
Setting up libgconf-2-4:amd64 (3.2.6-7ubuntu2) ...
Processing triggers for gnome-menus (3.36.0-1ubuntu3) ...
Processing triggers for libgl1-mesa-glx (2.6.0-1ubuntu2) ...
Setting up gconf-service-backend (3.2.6-7ubuntu2) ...
Setting up powerdevil (1.2.6-7ubuntu2) ...
Setting up libgudev-1.0-0:amd64 (1:3.16.6-1ubuntu2) ...
Setting up grace (1:5.1.25-12build1) ...
VirtualBox:~/Desktop/Demo$ xmgrace
xmgrace: cannot convert file "xmgrace.helvetica-medium-r-normal--12--*-**--**" to type fontStruct
VirtualBox:~/Desktop/Demo$ ^C

```



## Practical 6

### Title: - Gene annotation studies in Ensembl genome browser

#### Introduction:

Ensembl is a genome browser for vertebrate genomes that supports research in comparative genomics, evolution, sequence variation and transcriptional regulation. Ensembl annotate genes, computes multiple alignments, predicts regulatory function and collects disease data. Ensembl tools include BLAST, BLAT, BioMart and the Variant Effect Predictor (VEP) for all supported species.

Step 1: Go to ensembl genome browser and type BRCA1

Step 2: Click on a link.

The screenshot shows the Ensembl genome browser homepage at [asia.ensembl.org/index.html](http://asia.ensembl.org/index.html). The top navigation bar includes links for BLAST/BLAT, VEP, Tools, BioMart, Downloads, Help & Docs, and Blog. A search bar is present, with the placeholder "Search all species...". Below the header, there are four main tool links: Tools, BioMart, BLAST/BLAT, and Variant Effect Predictor. The BLAST/BLAT link has a sub-link for "All tools". The Variant Effect Predictor link has a sub-link for "All variants". The main content area features a search form with "Human" selected in the dropdown and "BRCA1" entered in the search field. Below the search form, there is a "Direct Links" section listing various BRCA1-related entries. To the right of the search form, there is a summary of Ensembl's mission and a "Ensembl Release 108 (Oct 2022)" section with a bulleted list of changes. At the bottom, there is a "Ensembl Rapid Release" section, a cookie consent banner, and a system status bar showing the date and time.

## BID 19006

Step 3: Now we have to locate the gene on chromosome with position and then click on location.

BID 19006

Step 4: Region in detail allows you to browse genes, variants, sequence conservation, and other annotation along the genome.

Structural Variant...

- splice donor 5th base variant
- frameshift variant
- start lost
- inframe deletion
- protein altering variant
- splice region variant
- splice donor region variant
- synonymous variant
- 5 prime UTR variant
- non coding transcript exon variant

Regulation Legend

- CNV
- CTCF
- open chromatin
- promoter flank
- enhancer
- promoter
- transcription factor binding

Age of Base Legend

- Human-specific base
- Appeared in mammals (paler = older)

There are currently 871 tracks turned off.

Ensembl Home sapiens version 108.38 (GRCh38.p13) Chromosome 17: 43,044,295 - 43,170,245

Add/remove tracks | Custom tracks | Share | Resize image | Export image | Reset configuration | Reset track order | Export this image | Permanent link • View in archive site

About Us

Get help

Our sister sites

Follow us

About us

Using this website

Ensembl Bacteria

Blog

https://asia.ensembl.org/Homo\_sapiens/ImageExport/ImageFormats?compo...

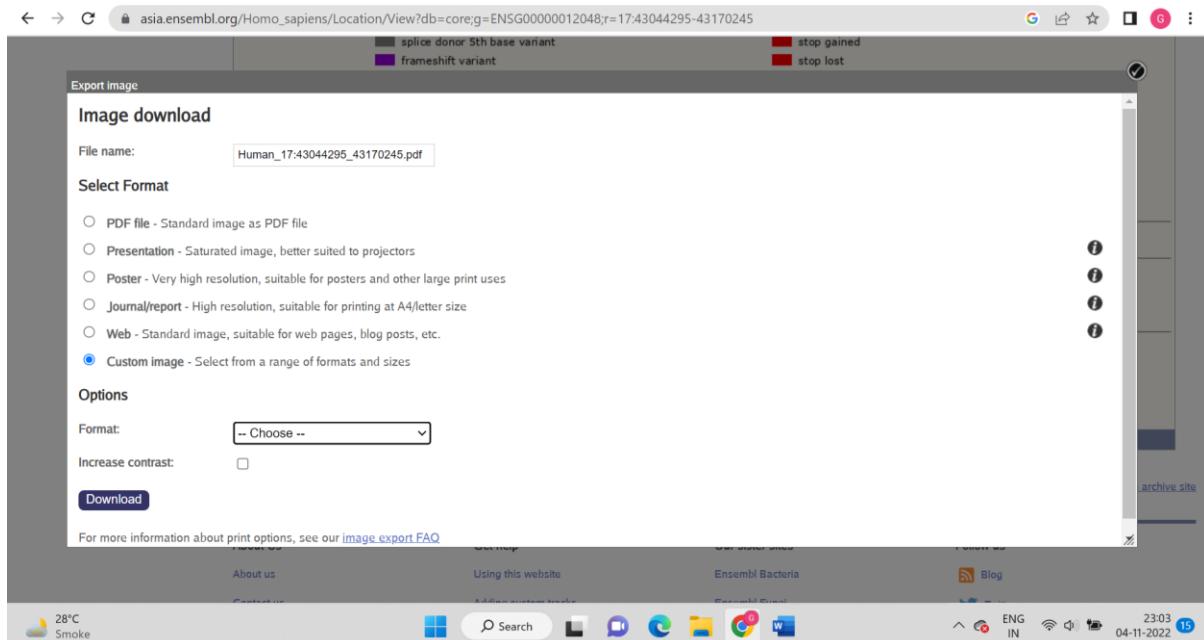
28°C  
Smoke

ENG IN

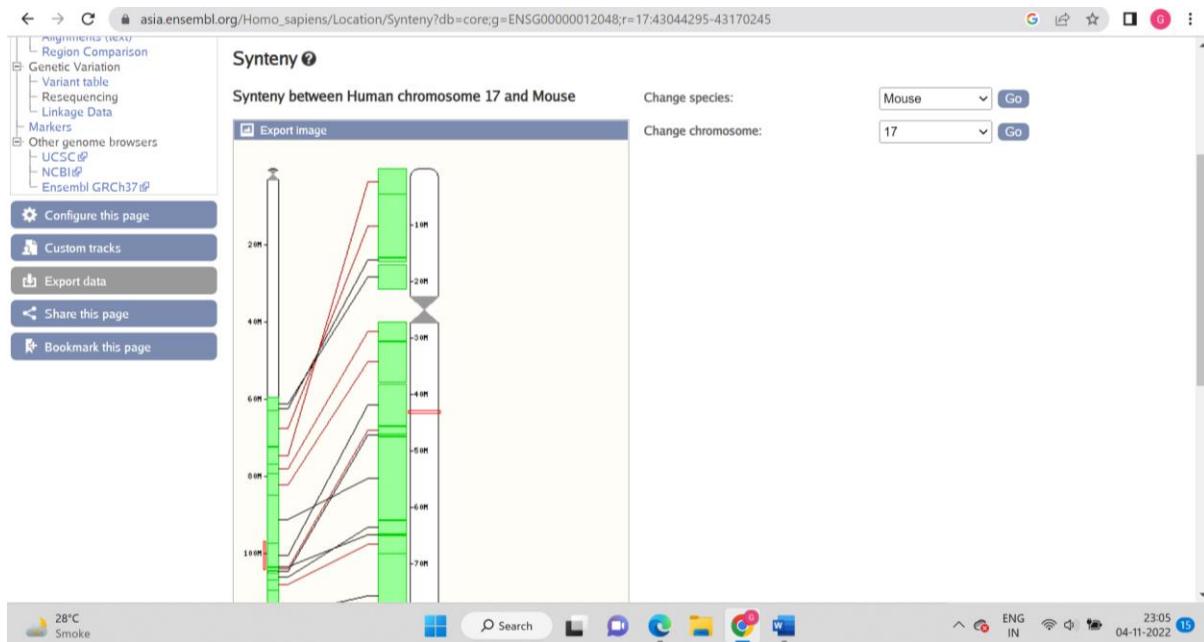
23:03 04-11-2022

## BID 19006

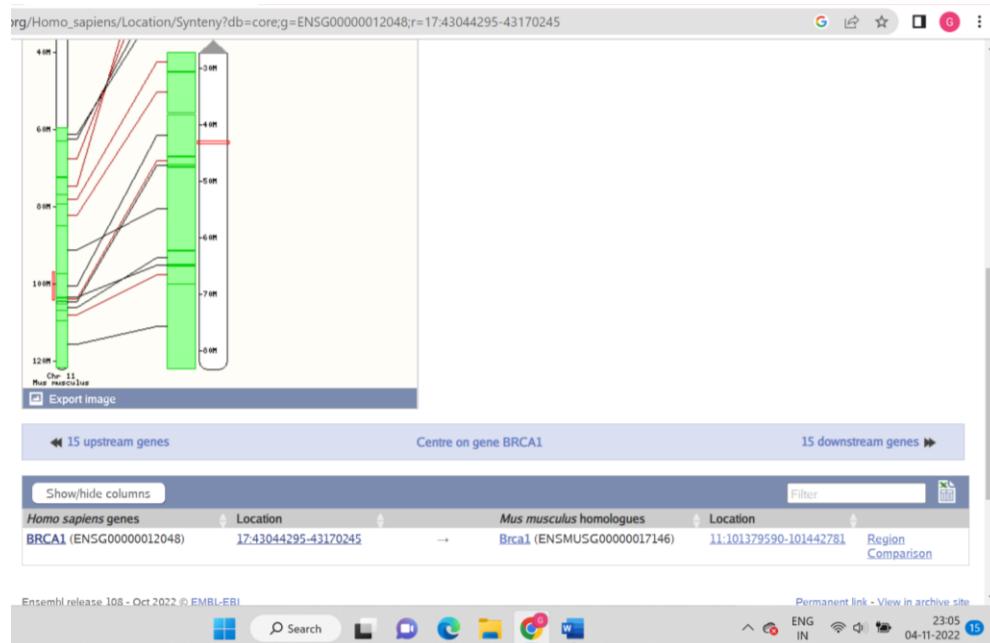
Step 5: After that just scroll down and click on the export image icon after that download in custom image format.



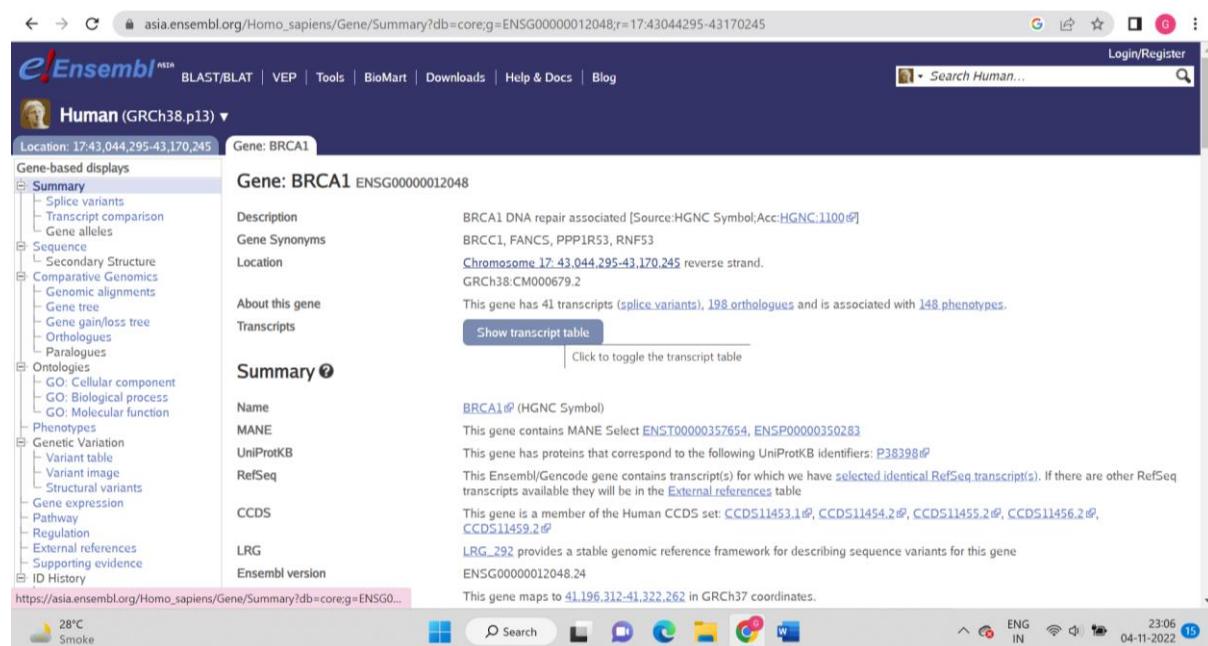
Step 6: Synteny – So, on the left side panel go and click on synteny.



## BID 19006



Step 7: So see how many transcripts are available for the gene of interest and show any one in detail also , now click on gene than at the top to get the number of transcripts.



# BID 19006

**Transcript ID** ENST00000472490.1 **Name** BRCA1-211 **bp** 561 **Protein** No protein **Biotype** Retained intron **CCDS** **UniProt Match** - **RefSeq Match** - **Fl**

**Summary**

**Same** [BRCA1](#) (HGNC Symbol)

**MANE** This gene contains MANE Select: [ENST00000357654](#), [ENSP00000350283](#)

**UniProtKB** This gene has proteins that correspond to the following UniProtKB identifiers: [P38398](#)

**RefSeq** This Ensembl/Gencode gene contains transcript(s) for which we have selected identical RefSeq transcript(s). If there are other RefSeq transcripts available they will be in the [External references](#) table

**CCDS** This gene is a member of the Human CCDS set: [CCDS11453.1](#), [CCDS11454.2](#), [CCDS11455.2](#), [CCDS11456.2](#), [CCDS11459.2](#)

**RG** LRG\_292 provides a stable genomic reference framework for describing sequence variants for this gene

**Ensembl version** ENSG00000012048.24

**Other assemblies** This gene maps to [41.196.312-41.322.262](#) in GRCh37 coordinates. View this locus in the GRCh37 archive: [ENSG00000012048](#)

**Gene type** Protein coding

**Annotation method** Annotation for this gene includes both automatic annotation from Ensembl and Havana manual curation, see [article](#).

**Annotation Attributes** overlapping locus [[Definitions](#)]

Go to [Region in Detail](#) for more tracks and navigation options (e.g. zooming)

Add/remove tracks | Custom tracks | Share | Resize image | Export image | Reset configuration | Reset track order | 145.95 kb Forward strand

43.04Mb 43.06Mb 43.08Mb 43.10Mb 43.12Mb 43.14Mb 43.16Mb 43.18

Genes (Comprehensive set from GENCODE 42) RPL21P4-201 - ENST00000497954 > NBR2-204 - ENST00000657841 >

Windows Search Task View Internet Explorer File 18:49 ENG IN 09-11-2022

org/Homo\_sapiens/Gene/Summary?db=core:g=ENSG00000012048;r=17:43044295-43170245

Go to [Region in Detail](#) for more tracks and navigation options (e.g. zooming)

Add/remove tracks | Custom tracks | Share | Resize image | Export image | Reset configuration | Reset track order | 145.95 kb Forward strand

43.04Mb 43.06Mb 43.08Mb 43.10Mb 43.12Mb 43.14Mb 43.16Mb 43.18

Genes (Comprehensive set from GENCODE 42) RPL21P4-201 - ENST00000497954 > processed pseudogene

NBR2-204 - ENST00000657841 > IncRNA

NBR2-202 - ENST00000460115 > IncRNA

NBR2-201 - ENST00000356906 > IncRNA

NBR2-203 - ENST00000467245 > IncRNA

ENST00000587322 > transcribed unprocessed pseudogene

ENST00000464237 > transcribed processed pseudogene

NBR1-201 - protein codin

NBR1-204 - protein codin

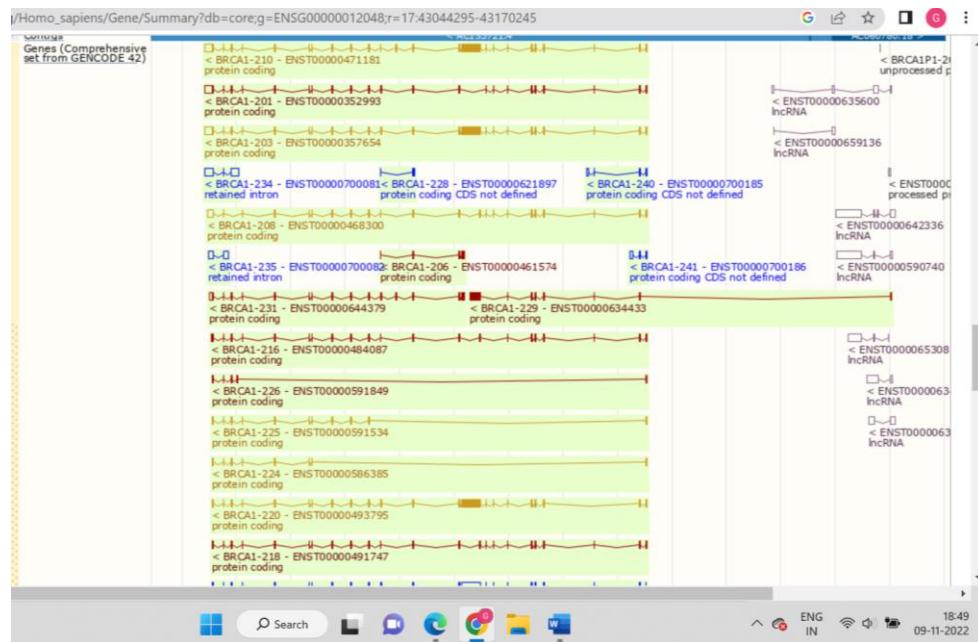
NBR1-202 - protein codin

NBR1-206 - protein codin

NBR1-207 - protein codin

Windows Search Task View Internet Explorer File 18:49 ENG IN 09-11-2022

## BID 19006



## BID 19006

### Step 8: Transcript of 1<sup>st</sup> id :

Transcript ID	Name	bp	Protein	Biotype	CCDS	UniProt Match	RefSeq Match	MANE Select	Ensembl Canonical
ENST00000357654.9	BRCA1-203	7088	1863aa	Protein coding	CCDS11453	P38398	NM_007294.4	-	
ENST00000471181.7	BRCA1-210	7270	1884aa	Protein coding	CCDS11456	P38398-7	-	-	GENCODE basic
ENST00000493795.5	BRCA1-220	5732	1816aa	Protein coding	CCDS11459	P38398-8	-	-	GENCODE
ENST00000354071.7	BRCA1-202	4497	1399aa	Protein coding		Q5YLB2	-	-	GENCODE
ENST00000352993.7	BRCA1-201	3668	721aa	Protein coding		P38398-5	-	-	GENCODE
ENST00000468300.5	BRCA1-208	3273	699aa	Protein coding	CCDS11455	P38398-6	-	-	GENCODE
ENST00000644379.1	BRCA1-231	2571	659aa	Protein coding		A0A2R8Y7V5	-	-	CDS 5
ENST0000064433.1	BRCA1-229	2534	798aa	Protein coding		A0A0U1RRA9	-	-	TSL:5 CI

### Step 9: Region in detail of the 1st id:

## BID 19006

[https://www.ensembl.org/Homo\\_sapiens/Transcript/Summary?db=core;g=ENSG0000012048;r=17:43044295-43170245;t=ENST00000357654](https://www.ensembl.org/Homo_sapiens/Transcript/Summary?db=core;g=ENSG0000012048;r=17:43044295-43170245;t=ENST00000357654)

Transcript ID	Name	bp	Protein	Biotype	CCDS	UniProt Match	RefSeq Match
ENST00000472490.1	BRCA1-211	561	No protein	Retained intron			

**Summary**

**Export image** |

< BRCA1-203 - ENST00000357654  
protein coding

Reverse strand ————— 81.07 kb —————

**Statistics**  
Exons: 23, Coding exons: 22, Transcript length: 7,088 bps, Translation length: 1,863 residues

**MANE**  
This MANE Select transcript contains [ENSP00000350283](#) and matches to [NM\\_007294.4](#) and [NP\\_009225.1](#).

**Uniprot**  
This transcript corresponds to the following Uniprot identifiers: [P38398](#).

**CCDS**  
This transcript is a member of the Human CCDS set: [CCDS11453](#).

**Transcript Support Level (TSL)**  
TSL:1  
ENST00000357654.9

**Type**  
Protein coding

**Annotation Method**  
Transcript where the Ensembl genebuild transcript and the Havana manual annotation have the same sequence, for every base pair. See [article](#).

**GENCODE basic gene**  
This transcript is a member of the [Gencode basic](#) gene set.

Ensembl release 108 - Oct 2022 © EMBL-EBI Permanent link · View in archive site

---

About Us Get help Our sister sites Follow us

Search ENG IN 18:50 09-11-2022

[https://www.ensembl.org/Homo\\_sapiens/Location/View?db=core;g=ENSG0000012048;r=17:43044295-43125364;t=ENST00000357654](https://www.ensembl.org/Homo_sapiens/Location/View?db=core;g=ENSG0000012048;r=17:43044295-43125364;t=ENST00000357654)

Add/remove tracks | Custom tracks | Share | Resize image | Export image | Reset configuration | Reset track order | Drag>Select:

Chromosome bands

Constrained elements for 91 eutherian mammals EPO-Extended

91 way GERP elements

Genes (Comprehensive set from GENCODE 42)

Contigs

Genes (Comprehensive set from GENCODE 42)

43.05Mb 43.06Mb 43.07Mb 43.08Mb 43.09Mb 43.10Mb 43.11Mb 43.12Mb

81.07 kb q21.31 RPL21P4-201 - ENST00000497954 > processed pseudogene

AC135721.4

BRCA1-210 - ENST00000471181 protein coding

BRCA1-201 - ENST00000352993 protein coding

BRCA1-203 - ENST00000357654 protein coding

BRCA1-234 - ENST00000700081 retained intron

BRCA1-208 - ENST00000468300 protein coding

BRCA1-235 - ENST00000700082 retained intron

BRCA1-231 - ENST00000493919 protein coding

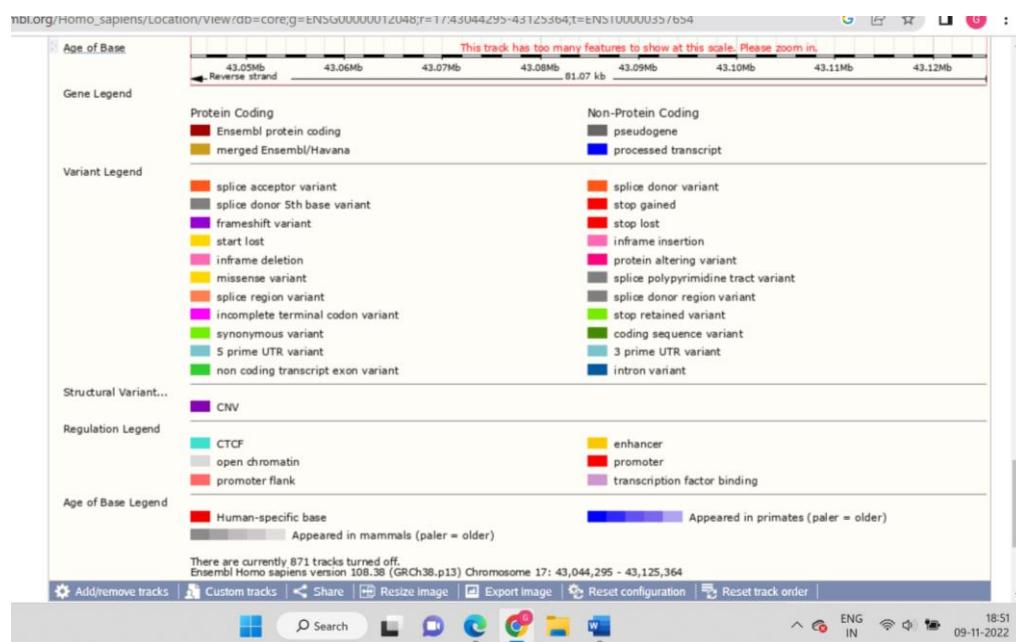
BRCA1-216 - ENST00000484087 protein coding

BRCA1-226 - ENST00000591849 protein coding

BRCA1-225 - ENST00000591534

18:51 09-11-2022

## Step 10: Variant legend , Gene legend , the age of base etc:



Step 11: Go to Gene tree:

**Number of genes**

207

**Number of speciation nodes**

176

**Number of duplication**

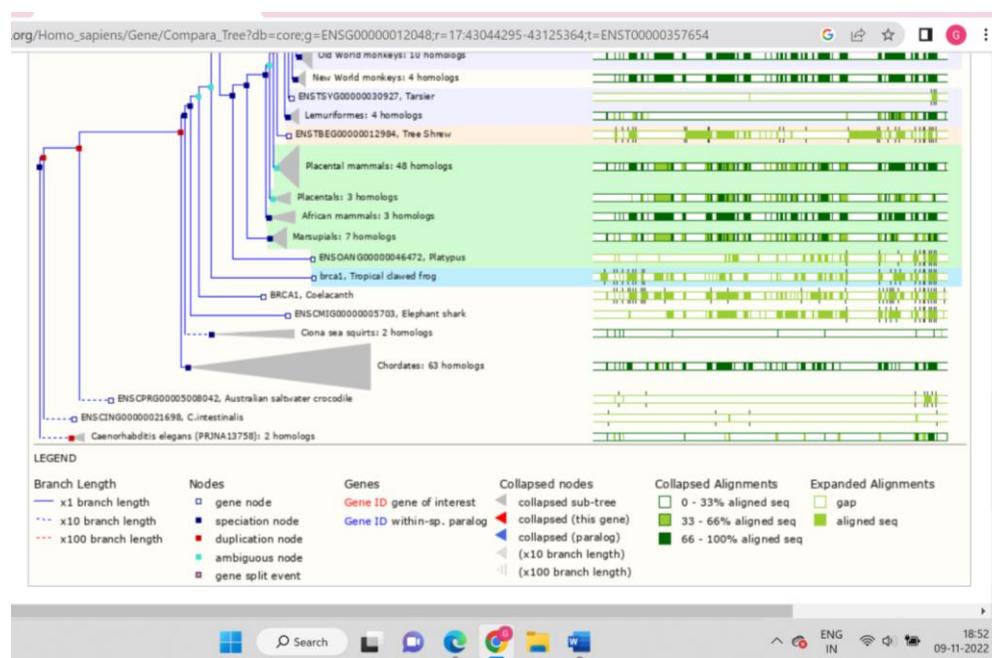
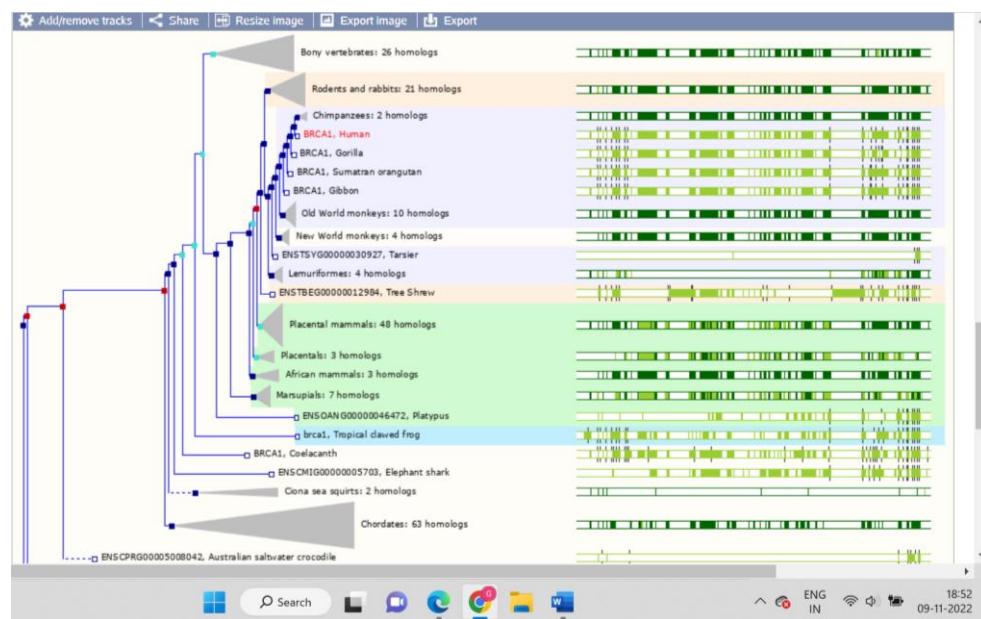
11

**Number of ambiguous**

10

**Number of gene split events**

9



Homo\_sapiens/Gene/Compara\_tree/db=core;g=ENSG0000012048;r=1/43044295-43125364;t=ENST00000357654

<a href="#">ENST00000700083.1</a>	BRCA1-236	1309	No protein	Retained intron
<a href="#">ENST00000700184.1</a>	BRCA1-239	919	No protein	Retained intron
<a href="#">ENST00000472490.1</a>	BRCA1-211	561	No protein	Retained intron

**Gene tree ?**

GeneTree [ENSGT00440000034289](#)

Number of genes 207

Number of speciation nodes 176

Number of duplication 11

Number of ambiguous 10

Number of gene split events 9

Step 12: Go to Ontologies and click on molecular functions:

Binding to a specific sequence of DNA that is part of a regulatory region that controls transcription of that section of the DNA. The transcribed region might be described as a gene, cistron, or operon.

l.org/Homo\_sapiens/Ontologies/molecular\_function/db=core;g=ENSG0000012048;r=1/43044295-43125364;t=ENST00000357654

GO: Molecular function ?

Show	All	entries	Show/hide columns (1 hidden)				Filter
Accession	Term	Evidence	Annotation source	Transcript IDs			
GO:0000976	transcription cis-regulatory region binding	IEA	Ensembl	ENST00000461221 ENST00000491747 ENST00000493795 ENST00000471181 ENST00000352993 ENST00000461798 ENST00000468300	• Search BioMart	• View on karyotype	
GO:0002039	p53 binding	IDA	DisProt	ENST00000352993 ENST00000357654 ENST00000493795 ENST00000491747 ENST00000461221 ENST00000471181 ENST00000461798 ENST00000468300	• Search BioMart	• View on karyotype	
GO:0003677	DNA binding	IEA	InterPro	ENST00000586385 ENST00000700182 ENST00000468300 ENST00000357654 ENST00000461221 ENST00000491747 ENST00000644379 ENST00000354071 ENST00000461798 ENST00000634433 ENST00000352993 ENST00000484087 ENST00000637707	• Search BioMart	• View on karyotype	

## BID 19006

Step 13: Go to Onlologies and click on cellular component:  
A protein complex that includes a ubiquitin-protein ligase and enables ubiquitin protein ligase activity. The complex also contains other proteins that may confer substrate specificity on the complex. PMID:9529603

GO: Cellular component <a href="#">?</a>					
Show All <a href="#">entries</a>		Show/hide columns (1 hidden)		Filter	
Accession	Term	Evidence	Annotation source	Transcript IDs	
GO:0000151 <a href="#">?</a>	ubiquitin ligase complex	NAS	UniProt	ENST00000468300 ENST00000352993 ENST00000491747 ENST00000461221 ENST00000471181 ENST00000357654 ENST00000493795 ENST00000461798	<ul style="list-style-type: none"><li>Search BioMart</li><li>View on karyotype</li></ul>
GO:0000152 <a href="#">?</a>	nuclear ubiquitin ligase complex	IDA	ComplexPortal	ENST00000357654 ENST00000471181 ENST00000461221 ENST00000491747 ENST00000352993 ENST00000468300 ENST00000461798 ENST00000493795	<ul style="list-style-type: none"><li>Search BioMart</li><li>View on karyotype</li></ul>
GO:0000793 <a href="#">?</a>	condensed chromosome	IEA	Ensembl	ENST00000357654	<ul style="list-style-type: none"><li>Search BioMart</li><li>View on karyotype</li></ul>
GO:0000794 <a href="#">?</a>	condensed nuclear chromosome	IEA	Ensembl	ENST00000357654	<ul style="list-style-type: none"><li>Search BioMart</li><li>View on karyotype</li></ul>
GO:0000800 <a href="#">?</a>	lateral element	IDA	MGI	ENST00000493795 ENST00000461798 ENST00000352993 ENST00000468300 ENST00000471181 ENST00000357654 ENST00000491747	<ul style="list-style-type: none"><li>Search BioMart</li><li>View on karyotype</li></ul>

Step 14: Go to Onlologies and click on biological process:

O: Biological process <a href="#">?</a>					
Show All <a href="#">entries</a>		Show/hide columns (1 hidden)		Filter	
Accession	Term	Evidence	Annotation source	Transcript IDs	
O:0000209 <a href="#">?</a>	protein polyubiquitination	IDA	ComplexPortal	ENST00000493795 ENST00000461798 ENST00000352993 ENST00000468300 ENST00000461221 ENST00000491747 ENST00000471181 ENST00000357654	<ul style="list-style-type: none"><li>Search BioMart</li><li>View on karyotype</li></ul>
O:0000724 <a href="#">?</a>	double-strand break repair via homologous recombination	IDA	HGNC-UCL	ENST00000468300 ENST00000352993 ENST00000471181 ENST00000357654 ENST00000491747 ENST00000461221 ENST00000493795 ENST00000461798	<ul style="list-style-type: none"><li>Search BioMart</li><li>View on karyotype</li></ul>
O:0006281 <a href="#">?</a>	DNA repair	IEA	InterPro	ENST00000461221 ENST00000491747 ENST00000477152 ENST00000352993 ENST0000044555 ENST00000354071 ENST00000586385 ENST00000493919 ENST00000642945 ENST00000492859 ENST00000478531 ENST00000461797	<ul style="list-style-type: none"><li>Search BioMart</li><li>View on karyotype</li></ul>

## BID 19006

Step 15: Go to CCDS :

Transcript ID	Name	bp	Protein	Biotype	CCDS	UniProt Match	RefSeq Match	MANE Select
ENST00000493795.5	BRCA1-220	5732	1816aa	Protein coding	CCDS11459	P38398-8	-	
ENST00000471181.7	BRCA1-210	7270	1884aa	Protein coding	CCDS11456	P38398-7	-	
ENST00000468300.5	BRCA1-208	3273	699aa	Protein coding	CCDS11455	P38398-6	-	
ENST00000491747.6	BRCA1-218	2379	759aa	Protein coding	CCDS11454	P38398-3	-	
ENST00000357654.9	BRCA1-203	7088	1863aa	Protein coding	CCDS11453	P38398	NM_007294.4	MANE Select
ENST00000354071.7	BRCA1-202	4497	1399aa	Protein coding		QSYLB2	-	
ENST00000352993.7	BRCA1-201	3668	721aa	Protein coding		P38398-5	-	
ENST00000644379.1	BRCA1-231	2571	659aa	Protein coding		A0A2R8Y7V5	-	
ENST00000634433.1	BRCA1-229	2534	798aa	Protein coding		A0A0U1RRA9	-	
ENST00000484087.6	BRCA1-216	2490	717aa	Protein coding		H0Y8B8	-	
ENST00000652672.1	BRCA1-233	2291	601aa	Protein coding		A0A494C182	-	
ENST00000700182.1	BRCA1-237	2269	544aa	Protein coding		-	-	
ENST00000470026.5	BRCA1-209	2108	649aa	Protein coding		E7EWNS	-	
ENST00000477152.5	BRCA1-214	1980	622aa	Protein coding		F9PH68	-	

Original	Current	Source	Nucleotide ID	Protein ID	Status in CCDS	Seq. Status	Links
✓	✓	EBI	ENST00000493795.5	ENSP00000418775.1	Accepted	alive	N P N P
✓	✓	NCBI	NM_007297.4	NP_009228.2	Accepted	alive	N P N P R

#### **Step 16: Go to Protein Sequence:**

## Step 17: Exons:

This transcript has 23 exons, is annotated with 121 domains and features, is associated with 31842 variant alleles and maps to 1304 oligo probes.

BID 19006

#### Step 18: Go to cDNA sequence:

[Step 16: Go to cDNA sequence](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=core&list_uids=143044295-43125364&dopt=fasta)

## cDNA sequence ?

 Download sequence

 BLAST this sequence

### Codons    Alternating codons    Alternating codons

Exons An exon Another exon

### Variants    3 prime UTR    5 prime U'

Protein altering

Other UTR

```

YYKW BKYMM SVHN*RRVNYR Y KR DKK SYY RR RS RRYBS*Y D
1 CCTGAGACTTCTCCTACGGGGGACACGGCTGTGCGGTTCTCGAGATAACTGGGCCCTCCG
.....  

.....  

Y S R SB B Y YB YS KR B RWY*RWY*MMSSMRMR*WVRNNNDMYW
61 CTCAGGACGCCCTTGACCTCTGCTCTGGTATAGTCATTGGAACAGAAGAAATGGATT
.....  

.....-M-D--  

.....  

BDYYWS**YW**YRWNSMMRMES***RRMW***VYWMHKSMNEYRRMMVR*****  

21 ATCTGCTCTTCGCTTGAGAGACTACAAAATGTCATTATGCTATGCGAAAAATCTTAG  

8 TATCTGCTCTTCGGCTTGAGAGACTACAAAATGTCATTATGCTATGCGAAAAATCTTAG  

3 L---S---A---L---R---V---E---E---V---Q---N---V---I---N---A---M---Q---K---I---L---  


```

## BID 19006



This transcript has 23 exons, is annotated with 121 domains and features, is associated with 31842 variant alleles and maps to 1304 oligo probes.

## Practical 7

### Title:- Gene prediction using FGENESH program

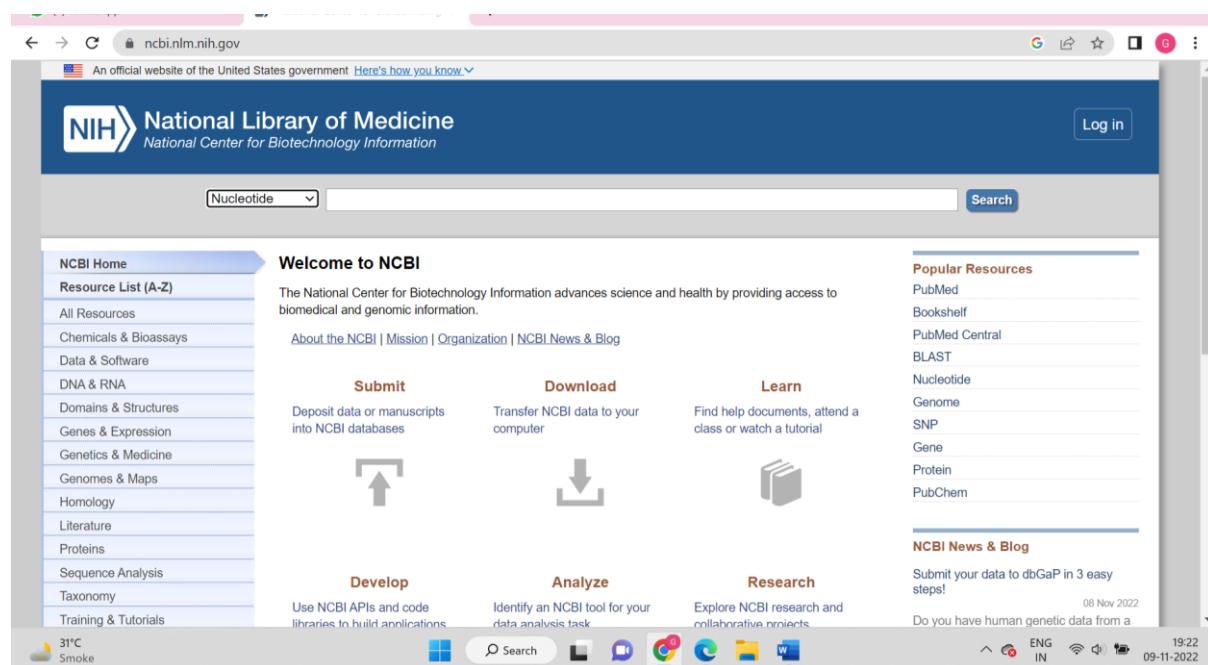
#### **Introduction:-**

FGENESH is the fastest and most accurate ab initio gene prediction program available. The Fgenesh gene-finder was selected as the most accurate program for plant gene identification. Plant Molecular Biology (2005), 57, 3, 445-460: "Five ab initio programs (FGENESH, GeneMark.hmm, GENSCAN, GlimmerR and Grail) were evaluated for their accuracy in predicting maize genes. FGENESH yielded the most accurate and GeneMark.hmm the second most accurate predictions" (FGENESH identified 11% more correct gene models than GeneMark on a set of 1353 test genes).

Web version of FGENESH can be used with parameters for the following genomes: human, mouse, Drosophila, nematode, dicot plants, monocot plants, yeast (*S.pombe*) and *Neurospora*.

Check appropriate genome/organism and and FGENESH program. Paste your sequence to the window or load your file with sequence in FASTA format and click Perform Search button.

**Step 1: Go to NCBI and click on all databases and select nucleotide**



**Step 2: In the search bar type Homo sapiens clone TMEM27-ACE2\_T5-A2 mRNA sequence**

The screenshot shows a web browser displaying the NCBI nucleotide database. The URL in the address bar is [ncbi.nlm.nih.gov/nucleotide/991820413](https://ncbi.nlm.nih.gov/nucleotide/991820413). The main content area is titled "Homo sapiens clone TMEM27-ACE2\_T5-A2 mRNA sequence". It provides detailed information about the sequence, including its definition, accession number (KM576724), version (KM576724.1), keywords, source (Homo sapiens (human)), and organism (Homo sapiens). The sequence itself is 864 bp long and is linear. The page also includes sections for "Analyze this sequence" (with options like Run BLAST, Pick Primers, Highlight Sequence Features, and Find in this Sequence), "Related information" (PubMed, Taxonomy, Full text in PMC), and "LinkOut to external resources" (Order TMEM27 cDNA clone/Protein/Antibody/RNAi [OriGene] and Order Ace2 cDNA clone/Protein/Antibody/RNAi [OriGene]). The bottom right corner shows the system status bar with "31°C Smoke", "ENG IN", and the date "09-11-2022" at 19:22.

**Step 3: Go to softberry homepage and click on Gene finding in Eukaryota:**

The screenshot shows the Softberry homepage. The top navigation bar includes links for "Run Programs Online" and "Not secure | softberry.com/berry.shtml". Below the header, there are several tool categories: "Cloud computing services" (described as data analysis using Softberry's public or clients' own pipelines in AWS cloud), "Next generation" (genome and transcript assembly, reads mapping, alternative transcripts, transomics pipeline, SNP discovery, and evaluation), "Annotation of Bacterial Genomes" (bacterial gene, promoters, terminators, operons identification, metagenomics, Fgenesb pipeline, microbiome sequence analysis and annotation), "Annotation of Animal Genomes" (gene identification, HMM Fgenesh gene finder and Fgenesh++ genome annotation pipeline, building gene models accounting RNASeq data, finding alternatively spliced isoforms, 329 custom made gene-finding parameters), "Annotation of Plant Genomes" (gene identification, Fgenesh gene finder and Fgenesh++ genome annotation pipeline, 42 custom made gene-finding parameters, largest Plant Gene Regulatory Elements database RegSite (~ 3000 entries)), "Alignment and Genome comparison" (alignment of sequences and genomes, multiple alignments, computing synteny between genomes, alignment of RNA and proteins onto genomic sequence), "Protein structure and functions" (comparison of 3D protein structures, finding functional sites and protein sub-cellular location, secondary structure prediction, protein structure, visualization, fold recognition, homology modeling, molecular docking, molecular mechanics and dynamics computations), and "RNA structure and functions" (RNA folding, finding palindrome, finding miRNA and their targets, search for bacterial terminators). The bottom right corner shows the system status bar with "31°C Smoke", "ENG IN", and the date "09-11-2022" at 19:23.

## BID 19006

The screenshot shows the Softberry homepage with several menu options: HOME, PRODUCTS, NEW PRODUCTS, SERVICES, MANAGEMENT TEAM, CORPORATE PROFILE, LINKS, and CONTACT. The main content area is divided into several sections:

- Gene Finding in Eukaryota:** Includes links to FGENESH (HMM-based gene structure prediction), FOENES (Pattern-based Human Gene structure), FOENES-M (multiple variants of Gene structure), FOENES-HM (multiple variants of Gene structure), variants of potential genes in genomic DNA), FGENESH\_GC (with possible donor GC), HMM based Human EST-ORF (Finding coding fragment EST-mRNA), FEX (Finding potential 5' internal and 3'-coding exons), SLM (Search for non-human splice sites using weight matrices), RNA-SPL (Search for exon-exon junction positions in cDNA), and FSPICE (Find splice sites in genomic DNA).
- Gene finding with similarity:** Includes links to FGENESH++ (HMM plus similar protein-based gene prediction), Speed & Accuracy of Fgenesh++, PROT MAP (mapping of a set of proteins on genome), FGENESH\_C (HMM plus similar cDNA-based gene structure), and FGENESH\_H2 (HMM gene prediction with two sequences of close organisms).
- Operon and Gene Finding in Bacteria:** Includes links to FGENESH\_B (Operon and Gene finding in Bacteria), AbSoLut (Separating archaea and bacterial genome fragments), FindTerm (Finding Terminators in bacterial genomes), Annotations (all bacteria), and Gene finding in Viral Genomes.
- Gene finding in Viral Genomes:** Includes links to FGENESV (Gene finding in Viral Genomes (Traced)), FGENESV (Gene finding in Viral Genomes (Markov chain-based viral gene prediction)), and FGENESV\_0 (Gene finding in Viral Genomes (Generic parameters Markov chain-based viral gene prediction)).
- GenomeSequence Explorer Infogene:** Includes a link to Human Genome Sequence Explorer (Visualization of Human Genome Annotations).
- Search for motifs /promoters/functional motifs/** Includes links to Regsite (List of Plant Regsite database factors used in TSSP and NSITE-PL programs), PERO (Human promoter prediction), PATTERNS (Search for significant patterns in the set of sequences), TSSP (Plants PLI - Promoter region and start of transcription), TSSW (Plants PLW - Promoter region and start of transcription), TSSW\_Human (Human Promoter region and start of transcription (ONLY for academic usage)), NSITE-PL (Recognition of PLN regulatory motifs), NSITE (Nearest neighbor with local alignments), NSITE\_HUMAN (Nearest neighbor with local alignments), NSITE\_HUMAN\_HUMAN (Nearest neighbor with local alignments), POLYAH (Recognition of polyadenylation region), BPROM (Promoter finding in Bacteria), PROMH (Q) (Find promoter with orthologs), PROMH (A) (Find promoter with orthologs (academic usage)), CpG Finder (Find GC-islands), ScanWMC (Search for weight matrix patterns of plant regulatory elements), Motif Explorer (Motif & promoter visualization sequences), PlantProm (Experimentally verified plant promoters database), and PlantProm - experimentally verified plant promoters database.
- Analysis of expression data:** Includes a link to SELTAG (Analysis of expression data).
- Alignment /Sequences&genomes/** Includes links to FMAP (mapping DNA/protein sequence on genome), SCAN2 (Comparison of 2 genomic sequences (with Java viewer)), SCAN2e (Comparison of 2 aminoacid sequences (with Java viewer)), DBSCAN (Comparing your sequence with Database (with Java viewer)), EST-mao (Mapping your mRNA/EST to Chromosome sequences of Human genome), PRO-Match (Comparison of a set of proteins on genome), Genomes Match (Comparison of 2 genomes or chromosomes), Genomes Match 2 (new version of comparison of 2 genomes or chromosomes), and Genome Match (Java Alignment Browser).
- Protein Location /pattern:** Includes links to ProtComp (Predict the subcellular localization for Animal/Unk), ProtComp (Predict the subcellular localization for Plant proteins), ProtComp (Localization of bacterial proteins), PSITE (Search for Prosite patterns with statistics), CTL-exope-Finder (Program for prediction of putative cytotoxic T-lymphotoxin (CTL) epitopes), and Protein Location /pattern.
- Protein structure:** Includes links to PSPPFinder (Prediction of protein secondary structure using Markov chains), SSPAL (Nearest-neighbors with local alignments SS prediction), NNSP (Nearest-neighbors SS prediction), SSP (Nearest-neighbors SS prediction), and SSENSE (Protein secondary structure and environment assignment from atomic coordinates).
- PDB-SEARCH (Protein-DNA 3D-Visual Works):** Includes links to GETATOMS (Atom coordinates using homologous protein), 3D-comp (Structure Alignment by Superposition), 3D-comp (Structure Alignment by Superposition), ABInitio (Ab initio folding), MM3SB (Multiple Structure Superposition), MM3SB (designed to perform multiple tasks with protein structure), HMod3DMM (Energy minimization program by molecular mechanics), CYS\_REC (Prediction of SS-bonding States of Cysteines in Protein Sequences), and 3DmodelFit (Program for comparison protein 3D model and its original structure).
- SeqMan:** Includes links to SeqMan (Manipulations with sequences), BestPal empirical (Calculates best palindrome for given DNA sequence, and also a set suboptimal palindromes (sorted by score)), and SeqMan (Manipulations with sequences).

Step 4: Now go to FGENESH: HMM-based gene structure prediction (multiple genes, both chains)

The screenshot shows the "Services Test Online" section for FGENESH. The left sidebar includes links to Home, Gene finding in Eukaryota (selected), Gene finding with similarity, Operon and Gene Finding in Bacteria, Gene Finding in Viral Genomes, Next Generation, Alignment (sequences and genomes), Genome visualization tools, Search for promoters/functional motifs, and Deep learning recognition.

The main content area has the following sections:

- FGENESH:** Used in more than 2800 publications. Reference: Solovyev V, Kosarev P, Seledsov I, Vorobev D. Automatic annotation of eukaryotic genes, pseudogenes and promoters. *Genome Biol.* 2006, 7, Suppl 1: P.10.1-10.12.
- Paste nucleotide sequence here:** A large text input field.
- Alternatively, load a local file with sequence in Fasta format:** A link to choose a file.
- Local file name:** A dropdown menu showing "Choose File" and "No file chosen".

Step 5: Paste the FASTA sequence from NCBI in the box show below :

The screenshot shows the NCBI Nucleotide search results for GenBank entry KM576724.1. The sequence is for the Homo sapiens clone TMEM27-ACE2\_T5-A2 mRNA. The sequence itself is a long string of nucleotides starting with >KM576724.1 Homo sapiens clone TMEM27-ACE2\_T5-A2 mRNA sequence, sequence. To the right of the sequence, there are several sections: 'Analyze this sequence' (Run BLAST, Pick Primers, Highlight Sequence Features), 'Related information' (PubMed, Taxonomy, Full text in PMC), and 'LinkOut to external resources' (Order TMEM27 cDNA clone/Protein/Antibody/RNAi [OriGene], Order Ace2 cDNA clone/Protein/Antibody/RNAi [OriGene]). The bottom of the screen shows standard Windows taskbar icons and system status.

The screenshot shows the Softberry gene finding tool interface. On the left, a sidebar lists various genomic analysis tools: Next Generation, Alignment (sequences and genomes), Genome visualization tools, Search for promoters/functional motifs, Deep learning recognition, Protein Location, RNA structures, Protein structure, Pathway prediction, Protein/DNA 3D-Visual Works, Manipulations with sequences, and Multiple alignments. The main panel is titled "Run Programs Online" and contains a text input field for nucleotide sequences. The sequence input field contains the FASTA sequence provided in the question. Below the input field, there are options for "Local file name:" (Choose File) and "Select organism specific gene-finding parameters:". A dropdown menu is set to "Human (Homo sapiens)". There are also links for "[Help]" and "[Show advanced options]". At the bottom, there are links for "[Example: Homo sapiens genomic beta globin region (HBB@) on chromosome 11]" and "[Example: Search in -chain]". The bottom of the screen shows the Windows taskbar and system status.

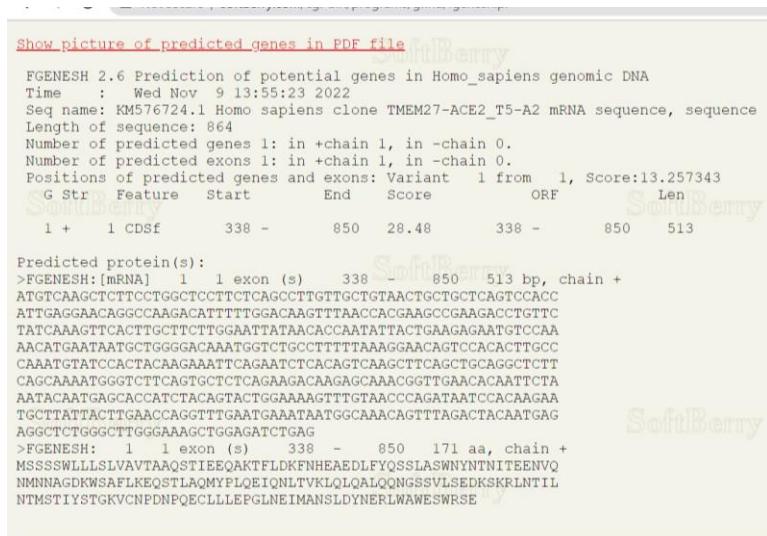
## BID 19006

Step 6: Click on search:

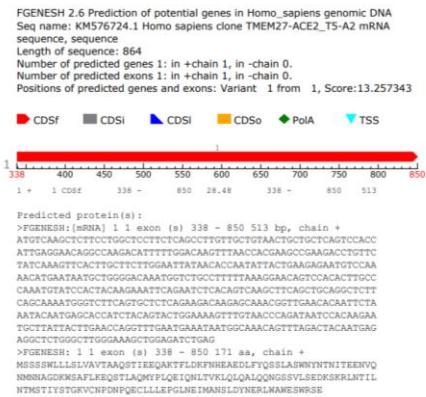
Length of sequence: 864

Number of predicted exons 1: in +chain 1, in -chain 0.

Positions of predicted genes and exons: Variant 1 from 1, Score:13.257343



Step 8: Predicted protein(s): >FGENESH:[mRNA] 1 1 exon (s) 338 - 850 513 bp, chain +



Interpretation:-

So, as shown that the length of the sequence is 864 bp. The number of genes predicted is 1. The start codon is predicted to be from the 338th position and stop codon is predicted to be from 850th position. There is only one exon region throughout the sequence. The score of the sequence is 28.48

## Practical 8

### Title:- Gene expression data analysis using GEO database

#### **Gene expression data analysis :**

Gene expression analysis is most simply described as the study of the way genes are transcribed to synthesize functional gene products — functional RNA species or protein products. The study of gene regulation provides insights into normal cellular processes, such as differentiation, and abnormal or pathological processes.

#### **Introduction:**

Gene Expression Omnibus (GEO) is a database for gene expression profiling and RNA methylation profiling managed by the National Center for Biotechnology Information (NCBI). These high-throughput screening genomics data are derived from microarray or RNA-Seq experimental data. These data need to conform to the minimum information about a microarray experiment (MIAME) format.

Step 1: Go to NCBI and go to GEO database homepage:

## BID 19006

Step 2: In the search box type the GEO accession number:

The screenshot shows the NCBI GEO homepage. At the top, there's a navigation bar with links for 'GEO Home', 'Documentation', 'Query & Browse', and 'Email GEO'. On the right, there's a 'Sign in to NCBI' link. Below the navigation, the 'Gene Expression Omnibus' logo is displayed. A search bar contains the accession number 'GSE66957' with a 'Search' button next to it. The main content area has three columns: 'Getting Started' (with links like Overview, FAQ, About GEO DataSets, etc.), 'Tools' (with links like Search for Studies at GEO DataSets, Search for Gene Expression at GEO Profiles, etc.), and 'Browse Content' (listing Repository Browser, DataSets: 4348, Series: 187089, Platforms: 24489, Samples: 5352935). Below these columns is a section titled 'Information for Submitters' with links for Login to Submit, Submission Guidelines, and MIAME Standards. The bottom of the page shows a Windows taskbar with various icons and system status information.

Step 3: In order to evaluate pathways that drive the development of ovarian cancer, we compared Affymetrix expression profiles from normal ovarian surface epithelium to those from primary ovarian carcinomas.

This screenshot shows the 'Accession Display' page for GSE66957. The top navigation bar includes 'HOME', 'SEARCH', 'SITE MAP', 'GEO Publications', 'FAQ', 'MIAME', and 'Email GEO'. It also indicates 'Not logged in | Login'. The main content area starts with a summary: 'Series GSE66957' was Public on Jun 22, 2015, with the title 'Pathways of ovarian cancer development'. The 'Organism' is listed as 'Homo sapiens'. The 'Experiment type' is 'Expression profiling by array'. The 'Summary' provides a brief description of the study目的: 'In order to evaluate pathways that drive the development of ovarian cancer, we compared Affymetrix expression profiles from normal ovarian surface epithelium to those from primary ovarian carcinomas.' Below this, the 'Overall design' is described as mRNA expression profiles for 57 ovarian carcinomas and 12 ovarian normal samples. The 'Contributor(s)' is Marchion DC. There's a note about citation missing: 'Has this study been published? Please [login](#) to update or notify GEO.' The 'Submission date' is Mar 16, 2015, and the 'Last update date' is Oct 19, 2022. The 'Contact name' is Chia-Ho Cheng, with the email address chiaho1@gmail.com. The 'Organization name' is Moffitt Cancer Center, located at 12902 USF Magnolia Drive, Tampa, ZIP/Postal code 33647. The bottom of the page features a Windows taskbar.

## BID 19006

Step 4: Now click on analyze with GEO2R:

Last update date: Oct 19, 2022  
Contact name: Chia-Ho Cheng  
E-mail(s): chiah01@gmail.com  
Organization name: Moffitt Cancer Center  
Street address: 12902 USF Magnolia Drive  
City: Tampa  
ZIP/Postal code: 33647  
Country: USA

Platforms (1): GPL15048 Rosetta/Merck Human RSTA Custom Affymetrix 2.0 microarray [HURSTA\_2a520709.CDF]

Samples (69):  
GSM1634925 Normal-ovarian sample1  
GSM1634926 Normal-ovarian sample2  
GSM1634927 Normal-ovarian sample3

Relations:  
BioProject: PRJNA278466

Analyze with GEO2R

Download family Format  
SOFT formatted family file(s) SOFT  
MINIMI formatted family file(s) MINIMI  
Series Matrix File(s) TXT

Supplementary file	Size	Download	File type/resource
GSE66957_RAW.tar	360.4 Mb	(http://custom)	TAR (of CEL)

Raw data provided as supplementary file  
Processed data included within Sample table

Step 5: Group the following data into normal and primary ovary:

ncbi.nlm.nih.gov/geo/geo2r/?acc=GSE66957

NCBI GEO Gene Expression Omnibus

GEO Publications FAQ MIAME Email GEO Login

NCBI > GEO > GEO2R > GSE66957

Use GEO2R to compare two or more groups of Samples in order to identify genes that are differentially expressed across experimental conditions. Results are presented as a table of genes ordered by significance. Full Instructions YouTube

GEO accession: GSE66957 Set Pathways of ovarian cancer development

Selected 0 out of 69 samples

Group	Accession	Title	Source name	Tissue	Cell type	Gender
-	GSM1634925	Normal-ovarian sample1	Normal ovarian tissue	normal ovary	surface epithelium	female
-	GSM1634926	Normal-ovarian sample2	Normal ovarian tissue	normal ovary	surface epithelium	female
-	GSM1634927	Normal-ovarian sample3	Normal ovarian tissue	normal ovary	surface epithelium	female
-	GSM1634928	Normal-ovarian sample4	Normal ovarian tissue	normal ovary	surface epithelium	female
-	GSM1634929	Normal-ovarian sample5	Normal ovarian tissue	normal ovary	surface epithelium	female
-	GSM1634930	Normal-ovarian sample6	Normal ovarian tissue	normal ovary	surface epithelium	female
-	GSM1634931	Normal-ovarian sample7	Normal ovarian tissue	normal ovary	surface epithelium	female
-	GSM1634932	Normal-ovarian sample8	Normal ovarian tissue	normal ovary	surface epithelium	female
-	GSM1634933	Normal-ovarian sample9	Normal ovarian tissue	normal ovary	surface epithelium	female

30°C Smoke ENG IN 19:31 09-11-2022

## BID 19006

Step 6: Now the green are primary ovary and purple are normal ovary:

Group	Accession	Source name	Tissue	Cell type	Gender
normal ovary	GSM1634925	Normal ovarian tissue	normal ovary	surface epithelium	female
normal ovary	GSM1634926	Normal ovarian tissue	normal ovary	surface epithelium	female
normal ovary	GSM1634927	Normal ovarian tissue	normal ovary	surface epithelium	female
normal ovary	GSM1634928	Normal ovarian sample4	Normal ovarian tissue	surface epithelium	female
normal ovary	GSM1634929	Normal ovarian sample5	Normal ovarian tissue	surface epithelium	female
normal ovary	GSM1634930	Normal ovarian sample6	Normal ovarian tissue	surface epithelium	female
normal ovary	GSM1634931	Normal ovarian sample7	Normal ovarian tissue	surface epithelium	female
normal ovary	GSM1634932	Normal ovarian sample8	Normal ovarian tissue	surface epithelium	female
normal ovary	GSM1634933	Normal ovarian sample9	Normal ovarian tissue	surface epithelium	female
normal ovary	GSM1634934	Normal ovarian sample10	Normal ovarian tissue	surface epithelium	female
normal ovary	GSM1634935	Normal ovarian sample11	Normal ovarian tissue	surface epithelium	female
normal ovary	GSM1634936	Normal ovarian sample12	Normal ovarian tissue	surface epithelium	female
primary carcinoma	GSM1634937	Ovarian sample1	Ovarian carcinoma tissue	primary ovarian carcinoma	surface epithelium
primary carcinoma	GSM1634938	Ovarian sample2	Ovarian carcinoma tissue	primary ovarian carcinoma	surface epithelium

Step 7: After grouping click on analyze:

If you edit Options after performing an analysis, click Reanalyze to apply the edits:

## BID 19006

ncbi.nlm.nih.gov/geo/geo2r/?acc=GSE66957

The screenshot shows the GEO2R interface for GSE66957. On the left, a sidebar lists samples grouped by 'normal ovary' (GSM1634925, GSM1634926, GSM1634927, GSM1634928, GSM1634929, GSM1634930, GSM1634931, GSM1634932, GSM1634933) and 'primary carcinoma' (GSM1634937, GSM1634938, GSM1634939). A 'Define groups' dialog is open, showing a list of samples under 'primary carcinoma' (57 samples) and 'normal ovary' (12 samples). The main panel displays sample metadata (Cell type, Gender) for 69 selected samples, all categorized as 'surface epithelium' female. Below this, a table shows sample details (Accession, Sample Name) for each group. A 'Quick start' section provides instructions for using the tool, and an 'Analyze' button is highlighted.

**Sample Groups**

Group	Accession
normal ovary	GSM1634925
normal ovary	GSM1634926
normal ovary	GSM1634927
normal ovary	GSM1634928
normal ovary	GSM1634929
normal ovary	GSM1634930
normal ovary	GSM1634931
normal ovary	GSM1634932
normal ovary	GSM1634933
primary carcinoma	GSM1634937
primary carcinoma	GSM1634938
primary carcinoma	GSM1634939

**Selected 69 out of 69 samples**

Cell type	Gender
surface epithelium	female

**Sample Groups**

Group	Accession
primary carcinoma	GSM1634937, GSM1634938, GSM1634939
normal ovary	GSM1634925, GSM1634926, GSM1634927, GSM1634928, GSM1634929, GSM1634930, GSM1634931, GSM1634932, GSM1634933

**Sample Details**

Accession	Sample Name
GSM1634937	Ovarian sample1
GSM1634938	Ovarian sample2
GSM1634939	Ovarian sample3

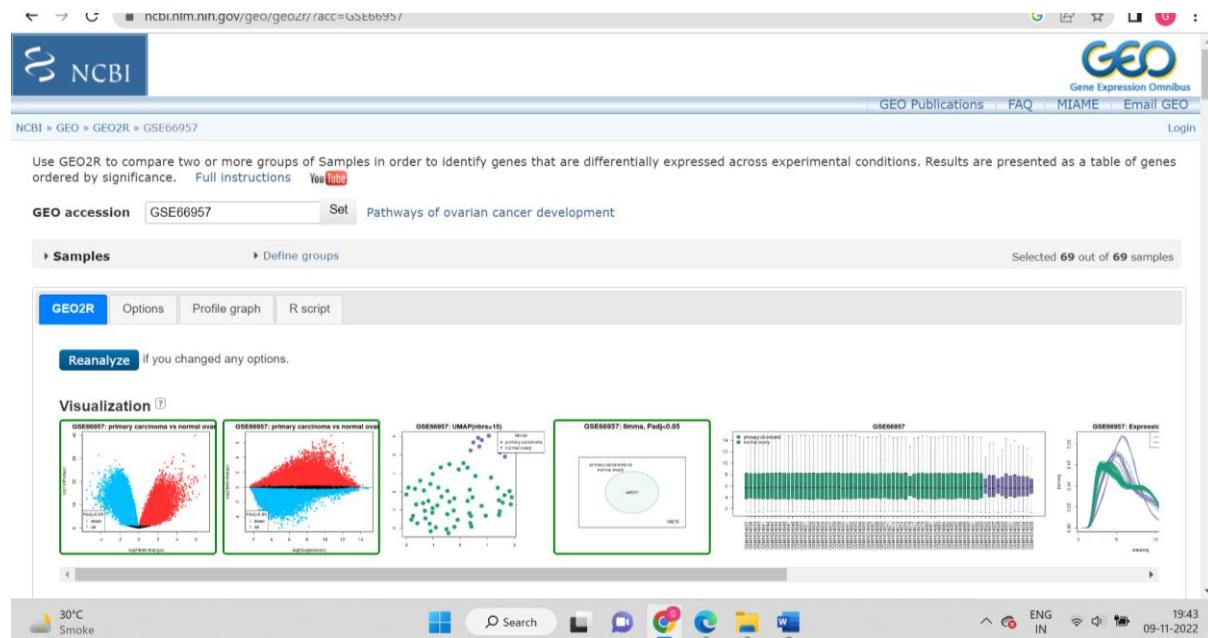
**Quick start**

- Specify a GEO Series accession and a Platform if prompted.
- Click 'Define groups' and enter names for the groups of Samples you plan to compare, e.g., test and control.
- Assign Samples to each group. Highlight Sample rows then click the group name to assign those Samples to the group. Use the Sample metadata (title, source and characteristics) columns to help determine which Samples belong to which group.
- Click 'Analyze' to perform the calculation with default settings.
- You may change settings in the Options tab.

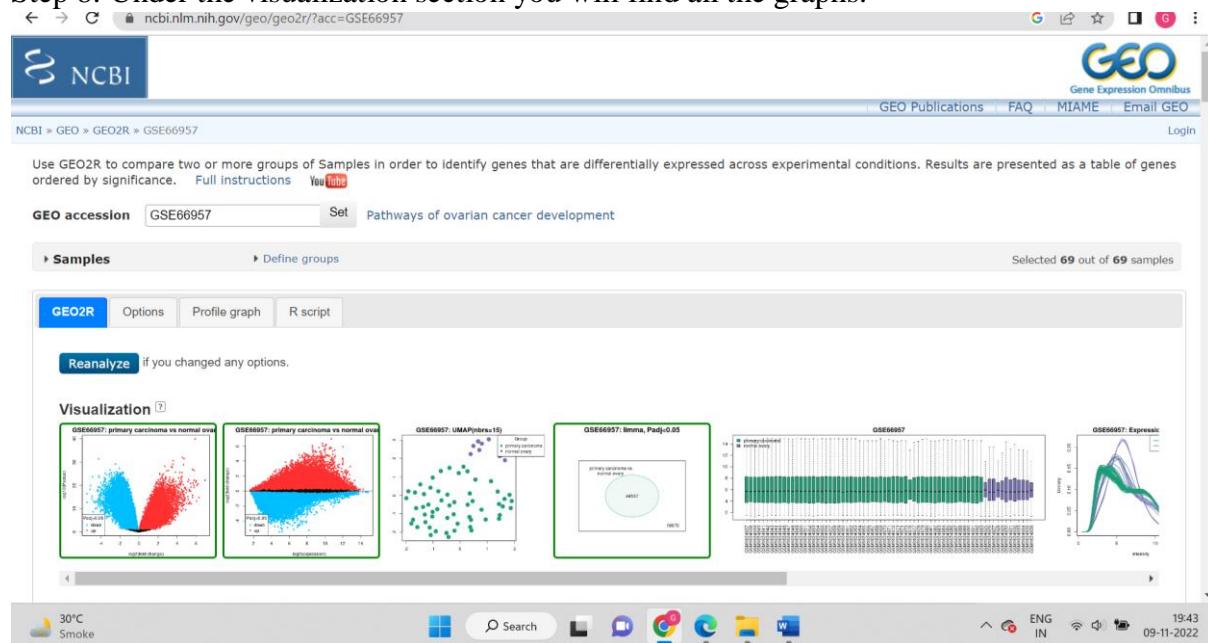
**Analyze**

Run the differential expression analysis. Distribution plots can be viewed without specifying Sample groups.

## BID 19006



Step 8: Under the visualization section you will find all the graphs:



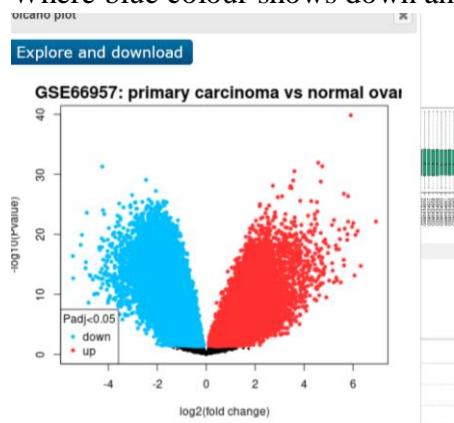
## BID 19006

Top differentially expressed genes <small>[?]</small>							
ID	adj.P.Val	P.Value	t	B	logFC	GB_LIST	SPOT_ID
merck-NM_006103_at	8.71e-36	1.44e-40	28.8	81.3	5.91	NM_006103	
merck-NM_139204_s_at	3.62e-28	1.19e-32	21.6	63.8	4.58	NM_133180 NM_017729	
merck2-AK025905_at	7.80e-28	4.63e-32	21.1	62.4	4.74	NM_022454	
merck-AF078528_x_at	7.80e-28	5.15e-32	-21.1	62.3	-4.24	AF078528	
merck2-NM_022572_a...	3.78e-27	3.12e-31	20.4	60.6	3.61	NM_015488 NM_022572	
merck-DR002919_at	8.51e-26	8.43e-30	-19.3	57.4	-2.45	NM_173552	
merck2-BM975589_a_at	9.34e-26	1.08e-29	19.3	57.1	3.56	NM_019590 NM_00109...	
merck-NM_022454_at	1.27e-25	1.67e-29	19.1	56.7	4.69	NM_022454	
merck2-BC009489_a_at	5.45e-25	8.10e-29	18.6	55.2	2.73	NM_138396	
merck-NM_015488_s_at	5.78e-25	9.54e-29	18.6	55	3.45	NM_015488 NM_022572	
merck2-BM975589_at	9.66e-25	1.75e-28	18.4	54.4	3.44	NM_019590	
merck2-NM_015488_a...	1.03e-24	2.03e-28	18.3	54.3	3.49	NM_015488 NM_022572	
merck-XM_928605_at	2.79e-24	5.99e-28	-18	53.2	-2.05	NG_006139	
merck2-NM_001038_at	7.63e-24	1.76e-27	17.6	52.1	5.63	NM_001038 NM_00115...	
merck-AK023831_at	1.20e-23	2.97e-27	-17.5	51.6	-3.02	AK023831	
merck-NM_001038_at	1.65e-23	4.49e-27	17.4	51.2	5.8	NM_001038 NM_00115...	
merck2-BC018764_at	1.65e-23	4.62e-27	17.4	51.2	3.1	NM_019590	
merck-NM_015202_at	2.21e-23	6.56e-27	17.2	50.8	2.92	NM_015202	
merck2-ENST0000037...	5.08e-23	1.59e-26	-17	50	-2.35	merck2-ENST00000378...	
merck-NM_015925_at	8.05e-23	2.66e-26	16.8	49.5	3.85	NM_015925 NM_20583...	
merck-NM_017650_at	8.25e-23	2.90e-26	16.8	49.4	4.29	NM_001166160 NM_00...	

Step 9: Following is Volcano plot for primary ovary vs normal ovary where the x axis shows log2 fold change and y axis shows -log10 p value:

Padj <0.05

Where blue colour shows down and red colour shows up:

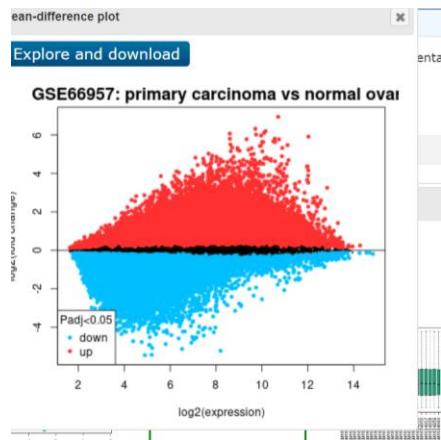


Step 10: Following is mean difference plot for primary ovary vs normal ovary where the x axis shows log2 expression and y axis shows log2 fold change:

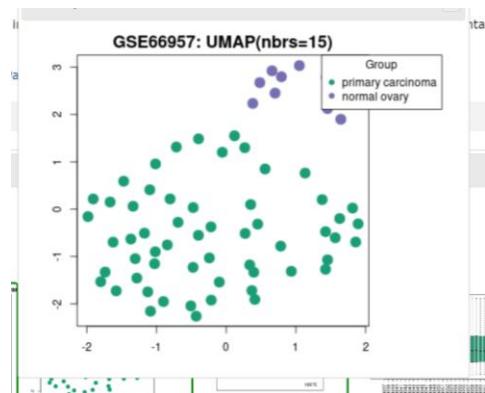
Padj <0.05 :-

Where blue colour shows down and red colour shows up:

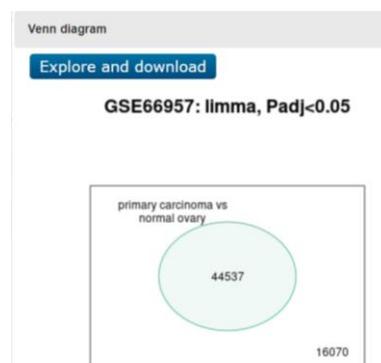
## BID 19006



Step 11: Following is UMAP plot for primary ovary vs normal ovary where nbrs = 15 and green dots indicate primary ovary and purple shows normal ovary:

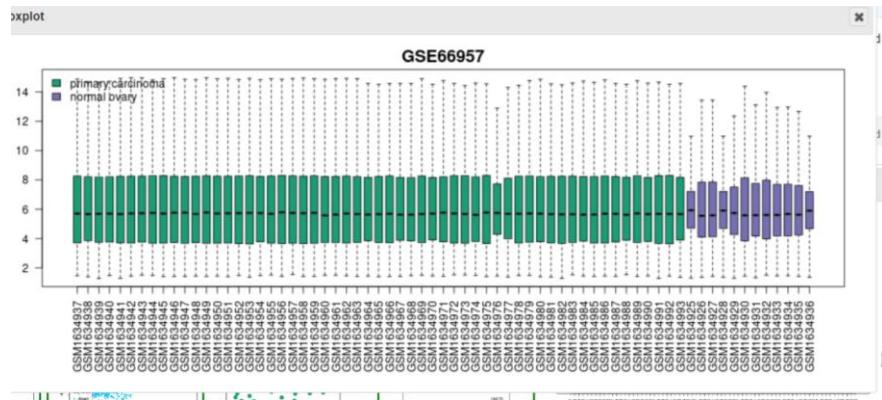


Step 12: Following is Venn diagram for primary ovary vs normal ovary where IIImma, Padj < 0.05:

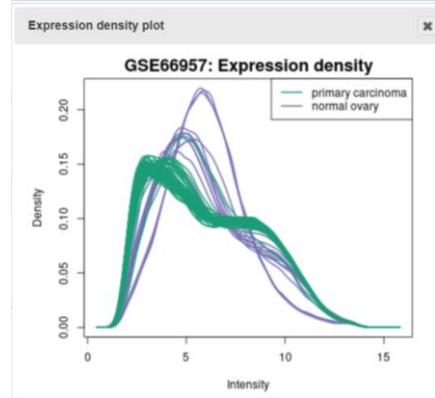


Step 13: Following is Boxplot for primary ovary vs normal ovary where green color bar are indicating primary ovary and purple colour bar shows normal ovary:

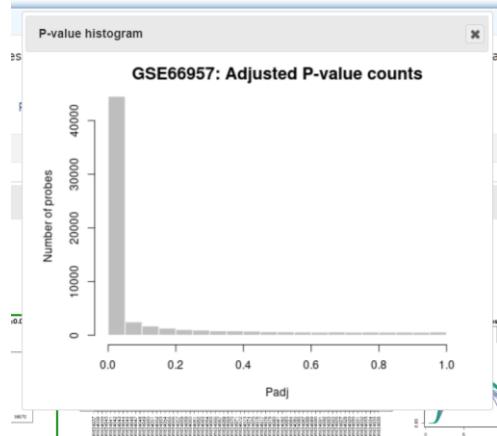
## BID 19006



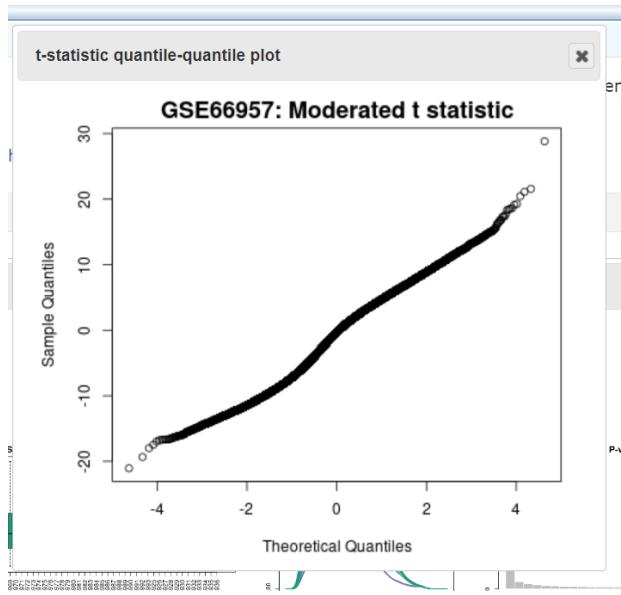
Step 14: Following is expression density plot for primary ovary vs normal ovary where green color is primary ovary and purple colour is normal ovary:



Step 15: Following is P-value histogram for primary ovary vs normal ovary where x axis shows Padj and Y axis shows Number of probes:



Step 16: Following is t-statistic quantile-quantile plot for primary ovary vs normal ovary where x axis shows Theoretical quantiles and y axis shows sample quantiles:



Step 17: Following is Mean-variance trend plot for primary ovary vs normal ovary where x axis shows average log expression and y axis shows log2 sigma:

60607 probes

