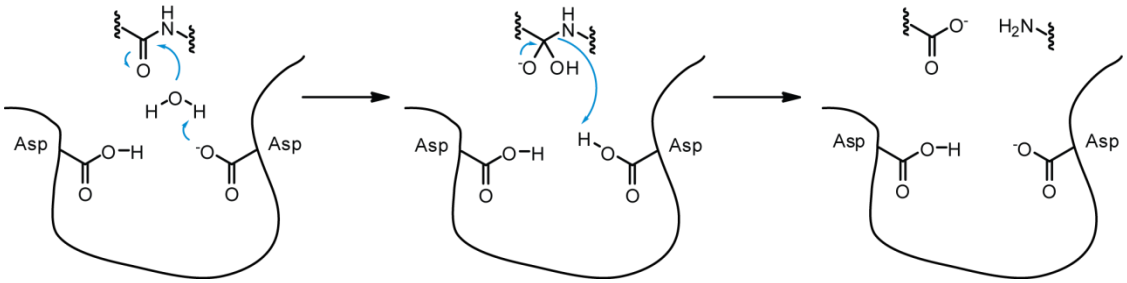
Analysing the Structure of an HIV-Protease using PDBe Tools

Jawahar Swaminathan, Ph.D.

Introduction

HIV-1 protease is a member of the aspartyl protease family and is essential for the life-cycle of HIV. Inhibition of this protease prevents maturation of HIV particles and has therefore, been the focus for the design and development of many drugs that inhibit this enzyme. There are over 100 structures of HIV proteases determined in complex with various drug candidates and peptidomimetic inhibitors. Proteases are a large family of enzymes that undertake proteolysis or the hydrolysis of the peptide bonds that link amino acids together in the polypeptide chain. Aspartyl proteases are so known because they utilize an aspartic acid residue in the active site for catalysis. These enzyme typically have two highly-conserved aspartates in the active site and function optimally at acidic pH. A rough schematic of the mechanism of action of an aspartyl protease enzyme is given below.



Purpose

This tutorial will analyse the structure of an HIV protease (EC Number: 3.4.23.16) using a variety of PDBe tools and services that are all available on the internet from the PDBe Portal at http://www.ebi.ac.uk/pdbe. It is hoped that at the end of this exercise, the user will have learned how to use the PDBE services and tools for searching and retrieving information from the PDBe using selected search criteria, exploring ligand/inhibitor binding sites, understanding and evaluating structures on the basis of quality, as well as appreciate the concepts of protein folds and quaternary structure assemblies.

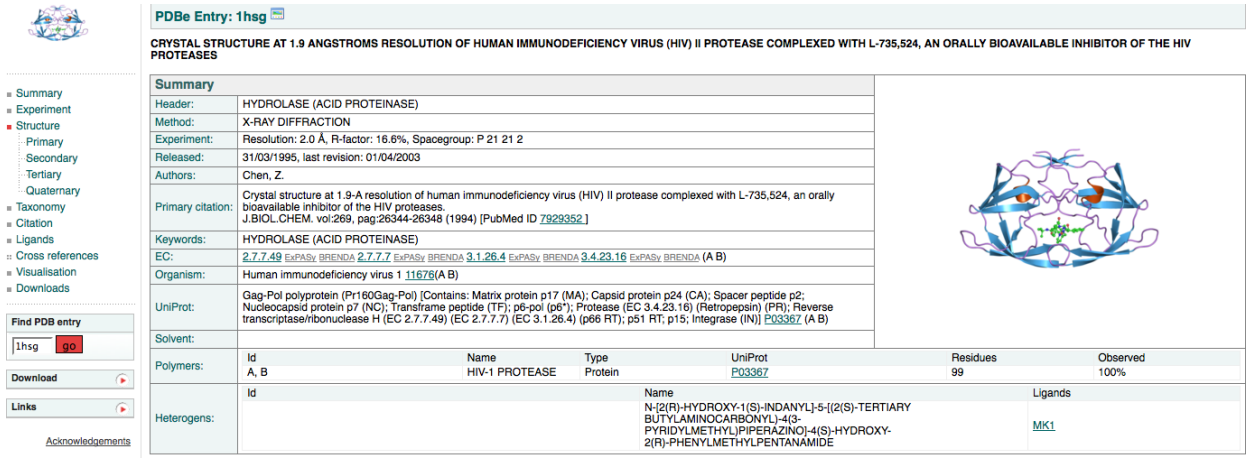
Requirements

a) A computer running any operating system connected to the internet.
b) Any modern web browser such as Internet Explorer/Firefox/Mozilla
c) Java Run-Time environment 1.5 or higher.
d) Rasmol/Raswin with mime-type chemical/x-pdb set in the browser to start rasmol/raswin when requested.
e) If running firefox, then the installation of Biobar (http://biobar.mozdev.org/) may help aid searching PDBe and other database.

Structure of HIV Protease (PDB entry 1HSG)

Let us start by loading the summary page for PDB entry 1HSG (XRAY Structure of HIV-1 PROTEASE at 2 A resolution). Go to the PDBE web page at http://www.ebi.ac.uk/pdbe and in the search box on the left side of the page titled "Get PDB by ID", type in the code 1HSG and click Go. This will load up a "Atlas" summary page for PDB entry 1HSG as stored in the PDBe database. The atlas page contains brief summary information regarding some important features of structure divided into tabs which are available on the left side of the page. Every underlined item on the pages are links to help text items which give more information about the item.

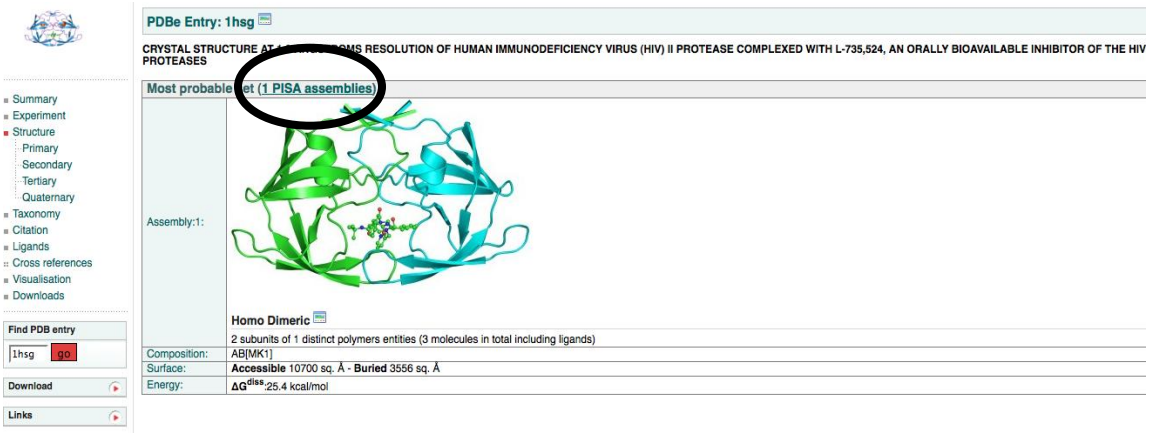
The Summary Page



The summary page contains at-a-glance information about the structure including sequence information as well as any bound molecules present in the structure. Let us look at some of the other information contained in the "Atlas" pages for this entry. Later on in this tutorial we will be coming back to this page for more information.

The Quaternary Structure Page

The quaternary structure link contains information about the probably quaternary structure of the protein as determined by the PDBePISA service (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html). PISA is an interactive tool for the exploration of macromolecular (protein, DNA/RNA and ligand) interfaces, prediction of probable quaternary structures (assemblies), database searches of structurally similar interfaces and assemblies, as well as searches on various assembly and PDB entry parameters. PDBepisa can also be used to upload one's own PDB file for the sort of analysis described below.



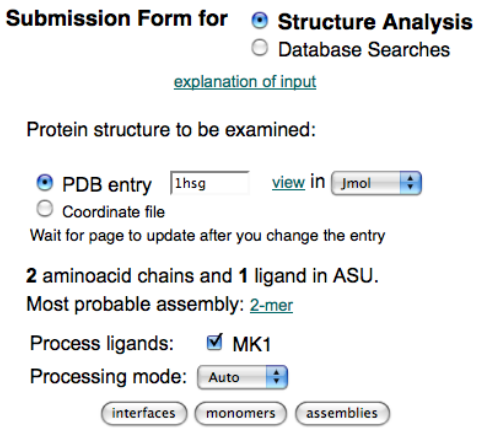
As you can see from the picture in the assembly page, there are two molecules of HIV-1 protease (colored green and cyan) with a single drug molecule bound in the center (in ball-and-stick). It would therefore, appear that the drug molecule binds in the interface of the two molecules of the enzyme with a 2:1 protein to inhibitor ratio. More information about this complex can be obtained from clicking on the link circled above. This should take you to the PDBePISA page for this entry. However, let us start PDBePISA independently as below.

In your browser go to http://www.ebi.ac.uk/pdbe and choose the PDBePISA link from the Services section on the right side of the page.

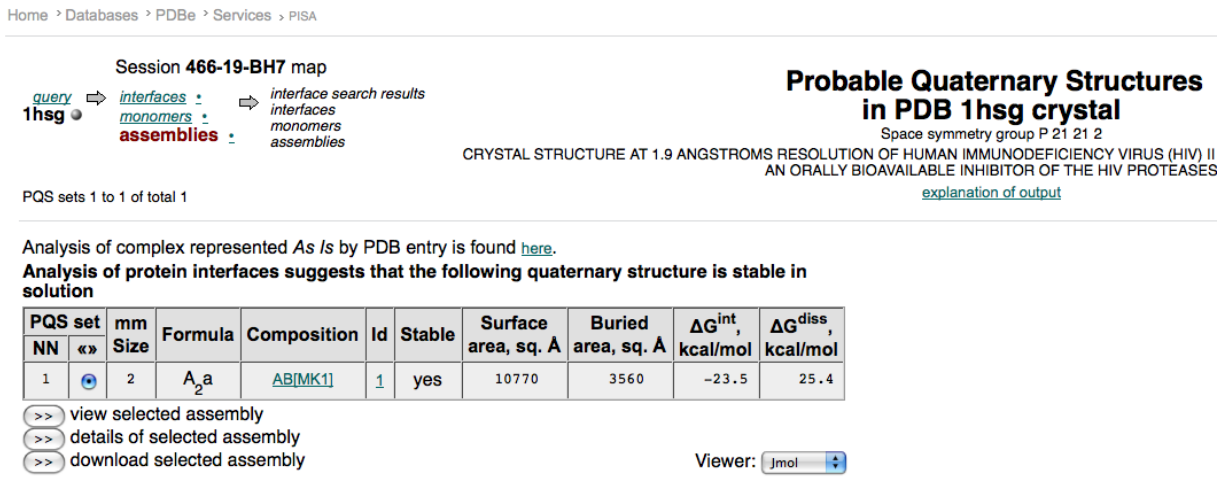
PDBePisa: Protein Interfaces, Surfaces and Assemblies

Click on "Start PISA" button and on the next page, type in 1HSG on the submission form. Wait for the page to update and ensure that the 'Process ligands' is checked and displays MK1. Click

on the assemblies button to get more information about the quaternary assembly predicted by PISA.



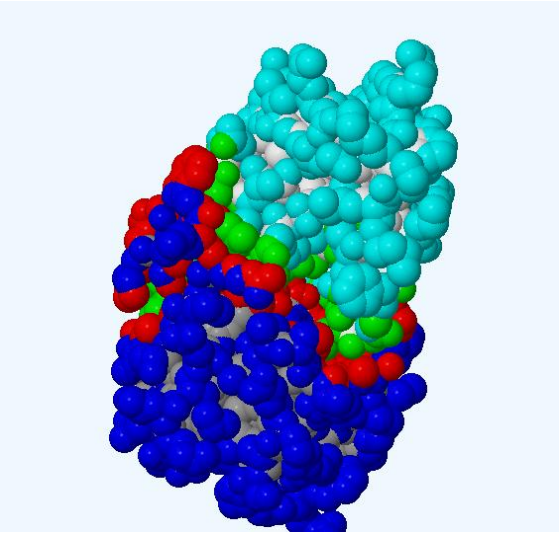
This page gives more information on the predicted quaternary structure this protein is likely to have in solution. PISA indicates that the assembly structure is dimeric (AB[MK1]) and is likely to be stable in solution, and that the energy required to break this complex in about 25.4 kcal/M. You can also view the assembly by clicking on the "View Selected Assembly" button.



Now choose the interfaces link on the top of this page to get more information about the various interfaces present in this structure.

Each interface between protein and ligand MK1 is analyzed and shown on the page. For each interface, the symmetry operation, and various statistics are provided. Let us look at the first interface on this page.

Click on the check box next to the 1st interface and choose view selected interface to pop-up the Jmol viewer in a separate window. The interface residues from each protein molecule are colored in red and green respectively. Rotate this and you should see a tunnel inside the complex which is the site at which the drug molecule MK1 binds to the protease.



Session **466-19-BH7** map

query 1hsg ⇨ interfaces • ⇨ interface search results
monomers • interfaces
assemblies • monomers
assemblies

Interfaces in PDB 1hsg crystal

Space symmetry group P 21 21 2

CRYSTAL STRUCTURE AT 1.9 ANGSTROMS RESOLUTION OF HUMAN IMMUNODEFICIENCY VIRUS (HIV) II PROTEASE COMPLEXED WITH L- 735,524 AN ORALLY BIOAVAILABLE INHIBITOR OF THE HIV PROTEASES

explanation of output

Found interfaces

##		Structure 1			×	Structure 2					Interface area, Å^2	Δ^iG kcal/mol	Δ^iG P-value	N_{HB}	N_{SB}	N_{DS}	CSS
NN	«»	Range	$^iN_{at}$	$^iN_{res}$		Range	Symmetry op-n	Sym.ID	$^iN_{at}$	$^iN_{res}$							
1	◉	B	174	42	◊	A	x,y,z	1_555	173	42	1746.9	-22.5	0.235	30	7	0	0.974
2	○	[MK1]-:902	37	1	◊	B	x,y,z	1_555	44	15	396.7	0.8	0.392	4	0	0	0.000
3	○	[MK1]-:902	36	1	◊	A	x,y,z	1_555	53	17	385.5	3.2	0.549	2	0	0	0.000
4	○	A	31	8	◊	B	x-1/2,-y+1/2,-z	4_455	29	7	254.5	-2.7	0.488	0	0	0	0.000
5	○	B	20	4	◊	A	-x+1/2,y-1/2,-z	3_545	24	6	215.7	1.0	0.832	4	0	0	0.000
6	○	B	19	8	◊	A	x,y,z-1	1_554	23	7	197.4	-2.0	0.527	1	0	0	0.000
7	○	B	18	4	◊	A	x-1/2,-y+1/2,-z	4_455	16	3	140.2	-1.8	0.493	2	0	0	0.000
8	○	B	12	4	x	B	-x+1/2,y-1/2,-z	3_545	9	1	92.0	-2.2	0.280	0	0	0	0.000
9	○	A	9	5	x	A	x-1/2,-y+1/2,-z	4_455	4	1	56.3	-1.5	0.216	0	0	0	0.000
10	○	[MK1]-:902	4	1	f	A	x-1/2,-y+1/2,-z	4_455	4	1	31.1	-1.0	0.263	0	0	0	0.100

(>>) view selected interface
(>>) details of selected interface
(>>) download selected interface
(>>) search PDB for interfaces between structures similar to those making the selected interface

Viewer: Jmol

Go back to the PDBePisa page for this entry and choose “Details of selected interface” to get a page of all interactions between the two protein molecules including hydrogen bonds, salt bridges etc.

Number of atoms				
interface	174	(23.0%)	173	(22.9%)
surface	513	(67.8%)	497	(65.7%)
total	757	(100.0%)	757	(100.0%)
Number of residues				
interface	42	(42.4%)	42	(42.4%)
surface	94	(94.9%)	92	(92.9%)
total	99	(100.0%)	99	(100.0%)
Solvent-accessible area, Å^2				
interface	1756.1	(26.0%)	1737.6	(26.1%)
total	6764.4	(100.0%)	6656.2	(100.0%)
Solvation energy, kcal/M				
isolated structure	-90.0	(100.0%)	-94.9	(100.0%)
gain at complexation	-12.3	(13.7%)	-10.2	(10.8%)
average gain	-9.4	(10.5%)	-7.7	(8.1%)
P-value	0.224		0.246	

This interface scored
0.974
in complexation significance score (CSS).
CSS ranges from 0 to 1 as interface relevance to complexation increases.

Achieved CSS implies that the interface plays an essential role in complexation

Hydrogen bonds

##	Structure 1	Dist. [Å]	Structure 2
1	B:ILE 3[N]	2.87	A:LEU 97[O]
2	B:ARG 8[NE]	3.27	A:ASP 29[OD1]
3	B:ARG 8[NH2]	2.82	A:ASP 29[OD2]
4	B:THR 26[N]	3.12	A:THR 26[OG1]
5	B:THR 26[OG1]	3.47	A:THR 26[OG1]
6	B:THR 26[OG1]	2.66	A:LEU 24[O]
7	B:GLY 52[N]	3.80	A:GLY 51[O]
8	B:ARG 87[NH1]	3.24	A:TRP 6[O]

Salt bridges

##	Structure 1	Dist. [Å]	Structure 2
1	B:ARG 8[NE]	3.27	A:ASP 29[OD1]
2	B:ARG 8[NE]	3.50	A:ASP 29[OD2]
3	B:ARG 8[NH2]	3.91	A:ASP 29[OD1]
4	B:ARG 8[NH2]	2.82	A:ASP 29[OD2]
5	B:ASP 29[OD1]	3.22	A:ARG 8[NE]
6	B:ASP 29[OD2]	2.82	A:ARG 8[NH2]
7	B:ASP 29[OD2]	3.41	A:ARG 8[NE]

Also look at other interfaces in this structure, particularly the ones between MK1 and chains A and B of HIV-1 protease to get an idea of the interactions between the protein and the drug molecule. You can also view every interface as above.

Another interesting feature on the interfaces page is the “search PDB for interfaces between structures similar to those making the selected interface” which will search the PDB for all interfaces similar to the one seen on the page. Clicking on this button will perform an exhaustive search of the PDB for list any other interfaces in the PDB that are similar. There are 298 structures in the PDB which have interfaces similar to the one seen in 1HSG. If you scroll through the list you can see from the titles that all the interfaces which are 70% or more similar to the interface in 1HSG are all protease enzymes. This suggests that this kind of interface is highly specific to aspartyl proteases.

Session **466-19-BH7** map

query 1hsg ⇨ interfaces • monomers • assemblies • ➡ interface search results: interfaces, monomers, assemblies

Search PDB for similar-structure interfaces

explanation of input

Search PDB for interfaces between:

Monomer 1: at least 70% similar to PDB 1hsg:A (view), **and**

Monomer 2: at least 70% similar to PDB 1hsg:B (view)

Return matches, where:

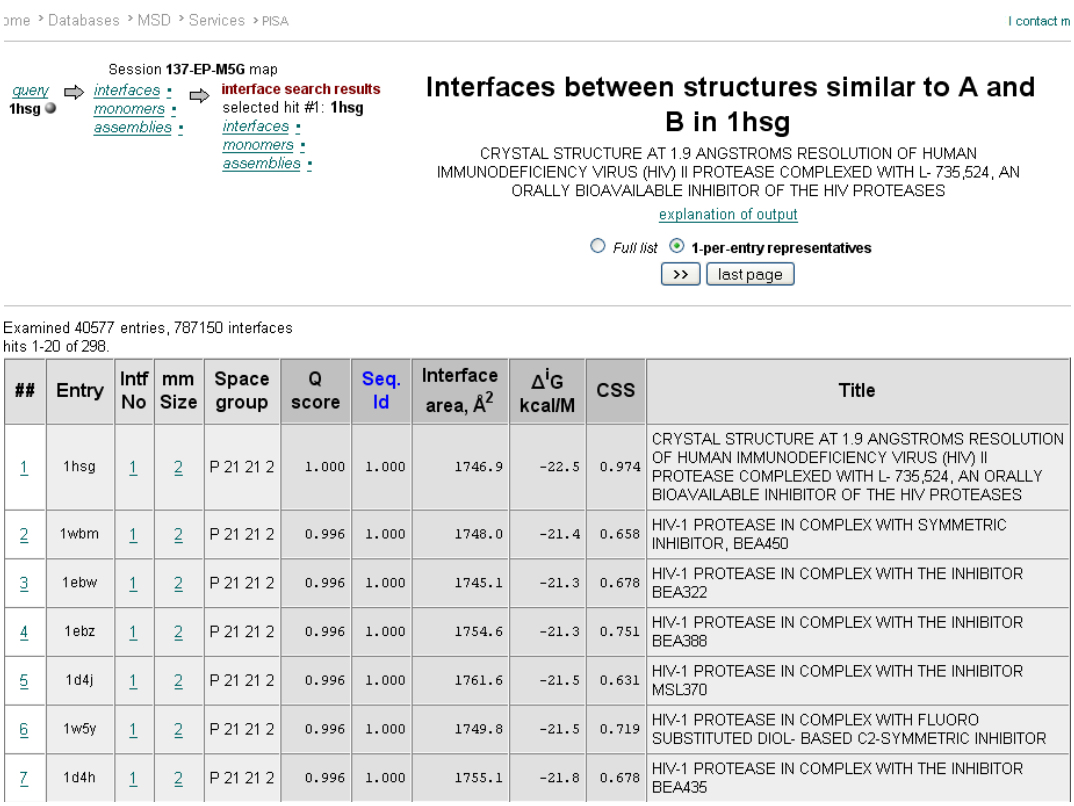
a multimeric assembly is or is not found

and interface 1hsg[1]B:A is or is not found

and any other interface from 1hsg is or is not found

in order to change the search monomer(s) or the query interface, make another selection in the interface list)

Submit for search Viewer: Jmol



ome › Databases › MSD › Services › PISA I contact m

Session **137-EP-M5G** map

query **1hsg** ⇨ *interfaces* ▪ *monomers* ▪ *assemblies* ▪ ⇨ **interface search results** selected hit #1: **1hsg** *interfaces* ▪ *monomers* ▪ *assemblies* ▪

Interfaces between structures similar to A and B in 1hsg

CRYSTAL STRUCTURE AT 1.9 ANGSTROMS RESOLUTION OF HUMAN IMMUNODEFICIENCY VIRUS (HIV) II PROTEASE COMPLEXED WITH L- 735,524, AN ORALLY BIOAVAILABLE INHIBITOR OF THE HIV PROTEASES

explanation of output

Full list **1-per-entry representatives**

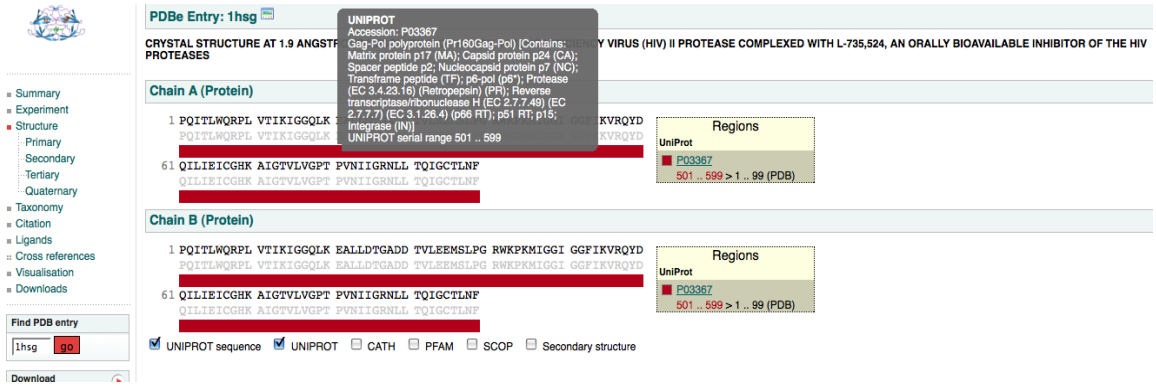
>> last page

Examined 40577 entries, 787150 interfaces
hits 1-20 of 298.

##	Entry	Intf No	mm Size	Space group	Q score	Seq. Id	Interface area, Å^2	Δ^iG kcal/M	CSS	Title
1	1hsg	1	2	P 21 21 2	1.000	1.000	1746.9	-22.5	0.974	CRYSTAL STRUCTURE AT 1.9 ANGSTROMS RESOLUTION OF HUMAN IMMUNODEFICIENCY VIRUS (HIV) II PROTEASE COMPLEXED WITH L- 735,524, AN ORALLY BIOAVAILABLE INHIBITOR OF THE HIV PROTEASES
2	1wbm	1	2	P 21 21 2	0.996	1.000	1748.0	-21.4	0.658	HIV-1 PROTEASE IN COMPLEX WITH SYMMETRIC INHIBITOR, BEA450
3	1ebw	1	2	P 21 21 2	0.996	1.000	1745.1	-21.3	0.678	HIV-1 PROTEASE IN COMPLEX WITH THE INHIBITOR BEA322
4	1ebz	1	2	P 21 21 2	0.996	1.000	1754.6	-21.3	0.751	HIV-1 PROTEASE IN COMPLEX WITH THE INHIBITOR BEA388
5	1d4j	1	2	P 21 21 2	0.996	1.000	1761.6	-21.5	0.631	HIV-1 PROTEASE IN COMPLEX WITH THE INHIBITOR MSL370
6	1w5y	1	2	P 21 21 2	0.996	1.000	1749.8	-21.5	0.719	HIV-1 PROTEASE IN COMPLEX WITH FLUORO SUBSTITUTED DIOL- BASED C2-SYMMETRIC INHIBITOR
7	1d4h	1	2	P 21 21 2	0.996	1.000	1755.1	-21.8	0.678	HIV-1 PROTEASE IN COMPLEX WITH THE INHIBITOR BEA435

Feel free to explore PDBePisa in your leisure time with your favorite structures at any time.

Let us now go back to our Summary Page for PDB entry 1HSG. http://www.ebi.ac.uk/pdbe-srv/view/entry/1hsg/. Go to the Primary Structure link on the left side of the page to load the sequence information. Make a note of the UniProt ID: P03367 for future reference below.



The Uniprot link on the right side will take you the Uniprot page for this sequence. A mouse over on the sequence will provide more information pertaining to the sequence. Similarly, you can also explore the Tertiary and Secondary structure links to get more information about the protein. There are links to Cath, Scop and Pfam database links based on the UniProt sequence database entry for the protein. Also provided are sequence alignments between the sequence from the protein structure and the UniProt sequence database.

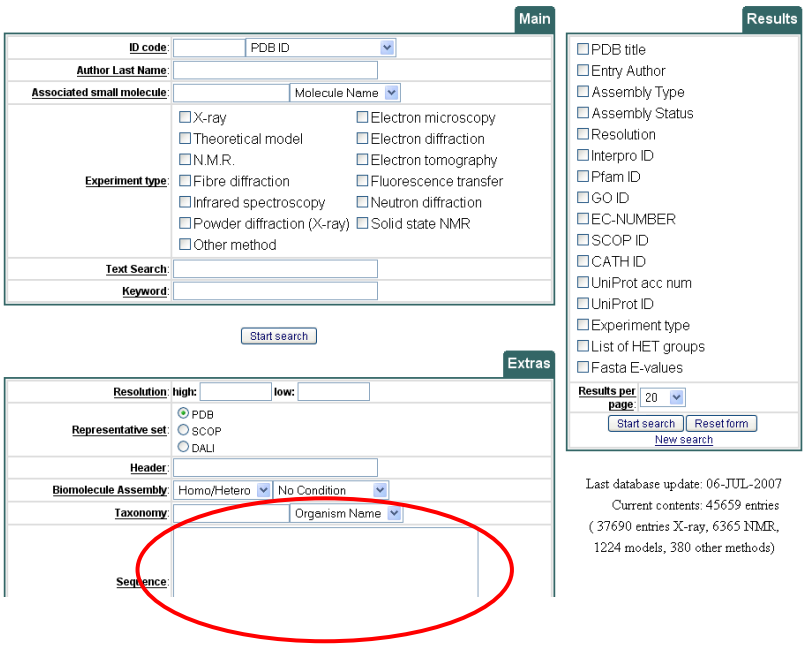
Copy the sequence of the protein with your mouse for the next step of this exercise. You can do this by choosing the Downloads -> mmCif and in the file that opens, take the sequence from:

```
_entity_poly.pdbx_seq_one_letter_code_can

;PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGIGGFIKVRQYDQIL
IEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNF
```

Searching for information in the PDBe using the PDBelite service

The PDBelite service provides a form-based search functionality to the PDBe database. This service allows the user to choose one or many criteria to search the database as well as to control what to see in the search results. Start the PDBelite service from the PDBe home page (http://www.ebi.ac.uk/pdbe/) by choosing the appropriate link from the Services section.



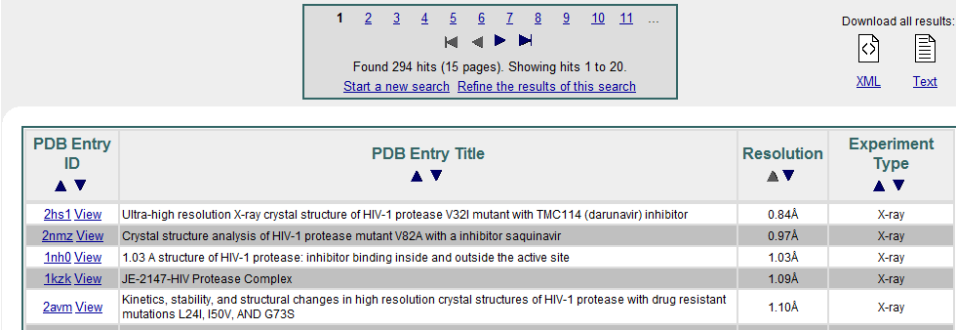
Paste the sequence of 1HSG that you copied over in the previous section inside the sequence box and click "Start Search". This will perform a Fasta search of the entire PDBe database and return only those results that match the sequence provided within a Fasta E-value of 1E-2. The results page will look like that shown below.

PDB Entry ID	PDB Entry Title	Resolution	Experiment Type
1nh0 View	1.03 A structure of HIV-1 protease: inhibitor binding inside and outside the active site	1.03Å	X-ray
2bb9 View	Structure of HIV1 protease and AKC4p_133a complex.	1.35Å	X-ray
2pwr View	HIV-1 PROTEASE IN COMPLEX WITH A CARBAMOYL DECORATED PYRROLIDINE-BASED INHIBITOR	1.50Å	X-ray
2pqz View	HIV-1 PROTEASE IN COMPLEX WITH A PYRROLIDINE-BASED INHIBITOR	1.55Å	X-ray
1xl5 View	HIV-1 Protease in complex with amidhyroxysulfone	1.73Å	X-ray
2uxz View	TWO-CARBON-ELONGATED HIV-1 PROTEASE INHIBITORS WITH A TERTIARY-ALCOHOL-CONTAINING TRANSITION-STATE MIMIC	1.75Å	X-ray
2uy0 View	TWO-CARBON-ELONGATED HIV-1 PROTEASE INHIBITORS WITH A TERTIARY-ALCOHOL-CONTAINING TRANSITION-STATE MIMIC	1.76Å	X-ray
1ec0 View	HIV-1 protease in complex with the inhibitor bea403	1.79Å	X-ray
1hvk View	INFLUENCE OF STEREOCHEMISTRY ON ACTIVITY AND BINDING MODES FOR C2 SYMMETRY-BASED DIOL INHIBITORS OF HIV-1 PROTEASE	1.80Å	X-ray
2cem View	P1' EXTENDED HIV-1 PROTEASE INHIBITORS ENCOMPASSING A TERTIARY ALCOHOL IN THE TRANSITION-STATE MIMICKING SCAFFOLD	1.80Å	X-ray
1d4h View	HIV-1 Protease in complex with the inhibitor BEA435	1.81Å	X-ray
1d4i View	HIV-1 protease in complex with the inhibitor BEA425	1.81Å	X-ray
1d4j View	HIV-1 protease in complex with the inhibitor MSL370	1.81Å	X-ray
1iiq View	CRYSTAL STRUCTURE OF HIV-1 PROTEASE COMPLEXED WITH A HYDROXYETHYLAMINE PEPTIDOMIMETIC INHIBITOR	1.83Å	X-ray
2bpv View	HIV-1 PROTEASE-INHIBITOR COMPLEX	1.90Å	X-ray
3bqb View	HIV-1 PROTEASE IN COMPLEX WITH A ISOBUTYL DECORATED OLIGOAMINE	1.90Å	X-ray
1ajv View	HIV-1 PROTEASE IN COMPLEX WITH THE CYCLIC SULFAMIDE INHIBITOR AHA006	2.00Å	X-ray
1ajx View	HIV-1 PROTEASE IN COMPLEX WITH THE CYCLIC UREA INHIBITOR AHA001	2.00Å	X-ray
1ec2 View	HIV-1 protease in complex with the inhibitor BEA428	2.00Å	X-ray
1npa View	crystal structure of HIV-1 protease-hup	2.00Å	X-ray

There appear to be many structures that match the sequence we provided in the PDBelite form. There are two links for each search result. The idcode link will open up the summary page for that entry, whereas the "view" link will open up AstexViewer that may be used to view and analyze the structure in further detail.

You can refine the results and whittle them down if necessary by choosing the "Refine the results of this search" link and adding more constraints.

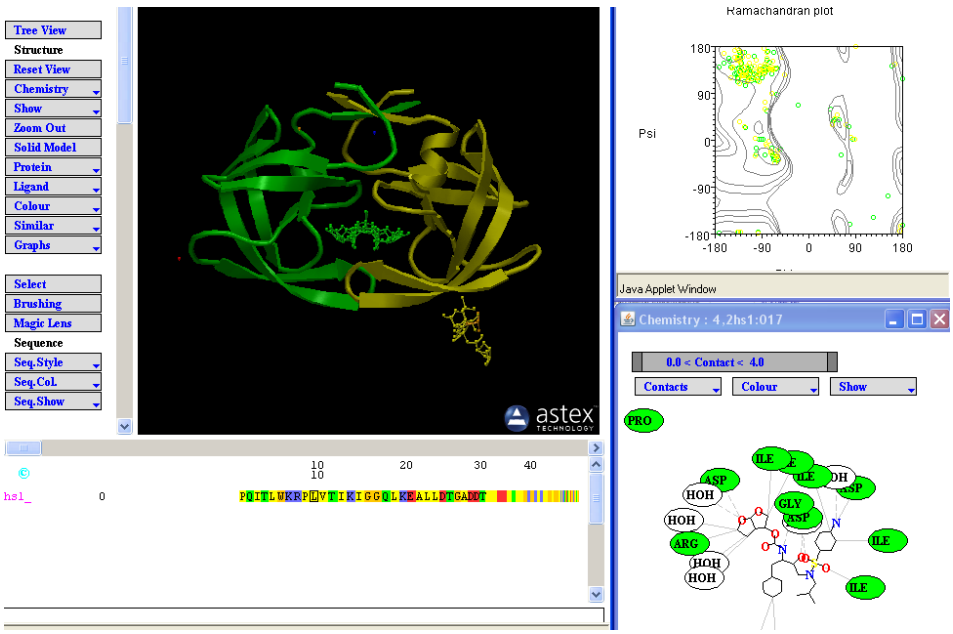
However, now let us start a new PDBelite session by going to http://www.ebi.ac.uk/msd-srv/msdlite. In the search page, entry HIV Protease in Text Search and then start a new search. From the result list, sort the entries by resolution by clicking on the triangles on the resolution column.



PDB Entry ID	PDB Entry Title	Resolution	Experiment Type
2hs1 View	Ultra-high resolution X-ray crystal structure of HIV-1 protease V32I mutant with TMC114 (darunavir) inhibitor	0.84Å	X-ray
2nmz View	Crystal structure analysis of HIV-1 protease mutant V82A with a inhibitor saquinavir	0.97Å	X-ray
1nh0 View	1.03 A structure of HIV-1 protease: inhibitor binding inside and outside the active site	1.03Å	X-ray
1kzk View	JE-2147-HIV Protease Complex	1.09Å	X-ray
2avm View	Kinetics, stability, and structural changes in high resolution crystal structures of HIV-1 protease with drug resistant mutations L24I, I50V, AND G73S	1.10Å	X-ray

The highest resolution structure of a HIV-1 protease is PDB entry 2HS1 which is at 0.84A resolution. Make a note of this entry for the future. Feel free to look over the summary pages for this entry. This entry is in complex with another drug molecule assigned the code 017 and is very similar in configuration and structure to drug MK1 seen in 1HSG.

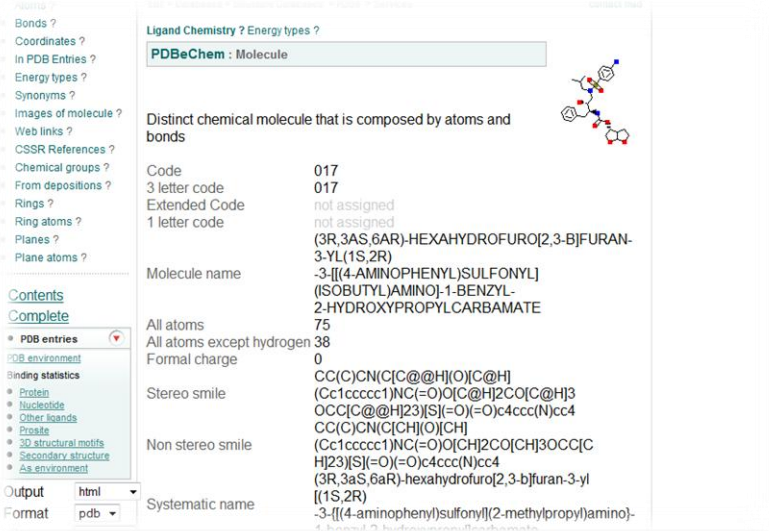
Let us view the structure of 2HS1 interactively. From the search list click on the "view" link or from the summary pages for 2HS1, choose the "Visualisation" link from the left side and click on "the PDB entry using AstexViewer™@MSD-EBI." This will open up a java window with the PDBe visualization tool.



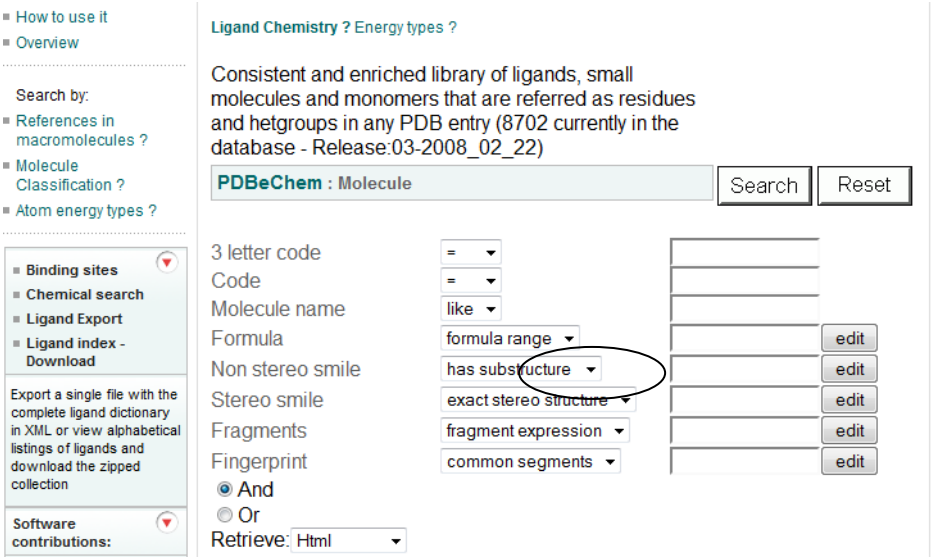
The viewer provides a powerful analysis interface with interactive views of sequence, structure quality graphs, chemistry information as well as excellent graphics quality. A beginners tutorial for AstexViewer is available separately at http://www.ebi.ac.uk/msd-srv/docs/Tutorials/Viewerframe.html. Have a play around with the structure and explore the chemistry views for ligand 017 and the Ramachandran graphs. Try clicking on a circle inside the Ramachandran Graph pane to see what happens in the viewer and sequence window. Then try clicking on one of the residues in the chemistry pane to see the effect on the viewer and sequence panes. Use the Zoom out button to reset your view.

Finding molecules similar to the one seen in 1HSG

It is possible to use the PDBe services to interactively search the PDBe for molecules based on their chemical structures. For this we use the PDBeChem service. The PDBeChem service is a consistent and enriched library of ligands, small molecules and monomers that are referred as residues and hetgroups in any PDB entry. For each unique small molecule observed in the PDB, there is a PDBeChem database entry. For example, you can view the PDBeChem entry for ligand MK1 we saw in PDB entry 1HSG directly from the summary pages for PDB entry 1HSG. Just go to the Ligand section at http://www.ebi.ac.uk/pdbe-srv/view/entry/2hs1/ligands and click on the picture of the ligand MK1 to load information about this ligand.

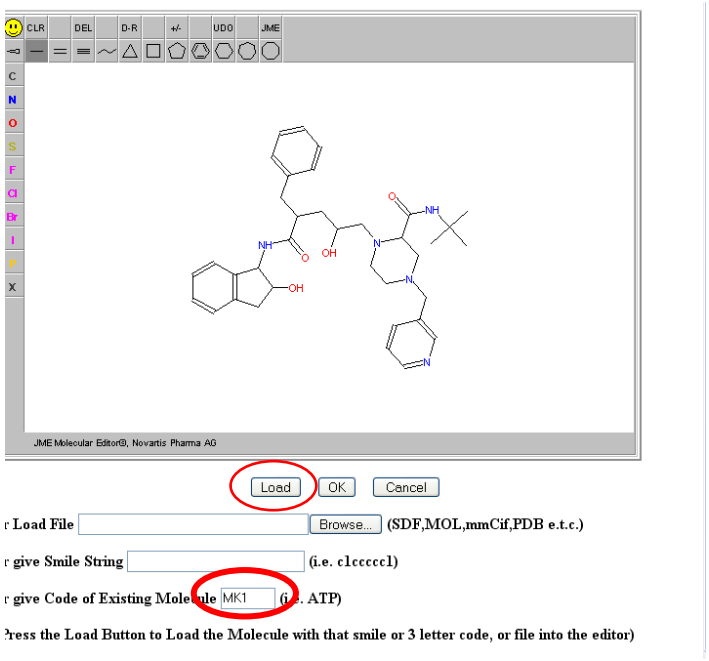


This page can be used to download model coordinates (from a representative structure), or ideal coordinates (based on standard chemical values) for docking experiments or other analysis as well as get information about the names, charge, molecular weight etc for small molecules.



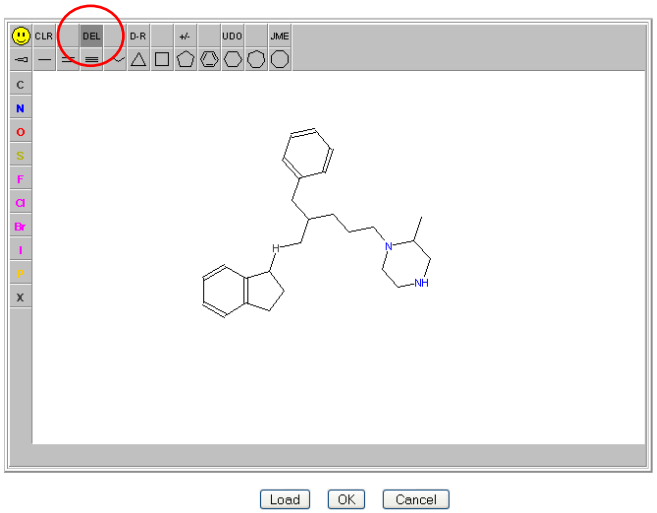
Now go to the main PDBeChem service, either from the PDBe home page (http://www.ebi.ac.uk/pdbe/) and choose PBDeChem from the Services section, or click on the "Ligand Chemistry" link on the page shown above.

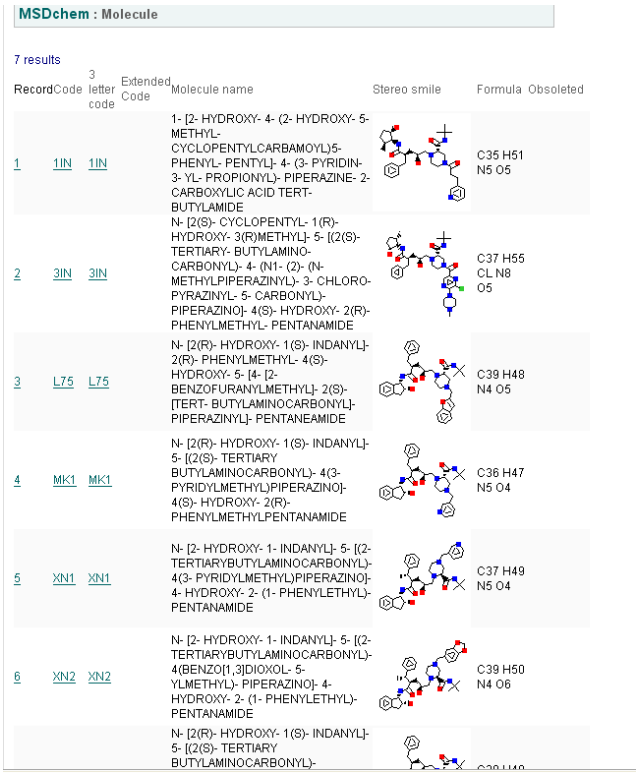
There are various search options on this page. One could search the PDBeChem database on formula, unique PDB 3-letter code etc. For our purposes of looking at molecules similar to MK1, choose the edit button on "Non-stereo smile" to open up new window. In the box provided for "or give Code of Existing Molecule" type in MK1 and click on the Load button to load the molecule into the java editor.



Using the "DEL" button on the editor, chop off a few atoms from the molecule and substitute the N atom near the indole ring with X (any atom), such as shown below. Then click on the "OK" button to close this window. On the main PDBeChem a smile string will have been pasted in the search box. Now click the "Search" button on PDBeChem.

This will search the database for other molecules that have a similar substructure. This search could take a few minutes since this is a database intensive process. If the search takes more than a few minutes, leave the process running and come back later on.





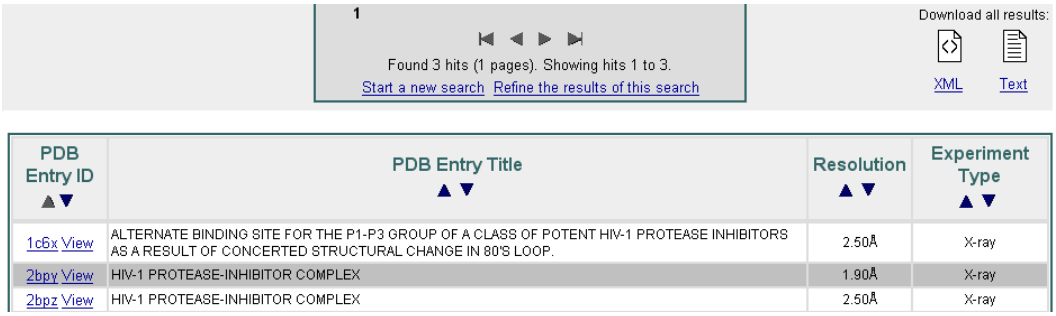
MSDchem : Molecule

7 results

Record	Code	3 letter code	Extended Code	Molecule name	Stereo smile	Formula	Obsoleted
1	1IN	1IN		1- [2- HYDROXY- 4- (2- HYDROXY- 5- METHYL- CYCLOPENTYLCARBAMOYL)5- PHENYL- PENTYL]- 4- (3- PYRIDIN- 3- YL- PROPIONYL)- PIPERAZINE- 2- CARBOXYLIC ACID TERT- BUTYLAMIDE		C35 H51 N5 O5	
2	3IN	3IN		N- [2(S)- CYCLOPENTYL- 1(R)- HYDROXY- 3(R)METHYL]- 5- [(2(S)- TERTIARY- BUTYLAMINO- CARBONYL)- 4- (N1- (2)- (N- METHYLPIPERAZINYL)- 3- CHLORO- PYRAZINYL- 5- CARBONYL)- PIPERAZINO]- 4(S)- HYDROXY- 2(R)- PHENYLMETHYL- PENTANAMIDE		C37 H55 CL N8 O5	
3	L75	L75		N- [2(R)- HYDROXY- 1(S)- INDANYL]- 2(R)- PHENYLMETHYL- 4(S)- HYDROXY- 5- [4- [2- BENZOFURANYLMETHYL]- 2(S)- [TERT- BUTYLAMINOCARBONYL]- PIPERAZINYL]- PENTANEAMIDE		C39 H48 N4 O5	
4	MK1	MK1		N- [2(R)- HYDROXY- 1(S)- INDANYL]- 5- [(2(S)- TERTIARY BUTYLAMINOCARBONYL)- 4(3- PYRIDYLMETHYL)PIPERAZINO]- 4(S)- HYDROXY- 2(R)- PHENYLMETHYLPENTANAMIDE		C36 H47 N5 O4	
5	XN1	XN1		N- [2- HYDROXY- 1- INDANYL]- 5- [(2- TERTIARYBUTYLAMINOCARBONYL)- 4(3- PYRIDYLMETHYL)PIPERAZINO]- 4- HYDROXY- 2- (1- PHENYLETHYL)- PENTANAMIDE		C37 H49 N5 O4	
6	XN2	XN2		N- [2- HYDROXY- 1- INDANYL]- 5- [(2- TERTIARYBUTYLAMINOCARBONYL)- 4(BENZO[1,3]DIOXOL- 5- YLMETHYL)- PIPERAZINO]- 4- HYDROXY- 2- (1- PHENYLETHYL)- PENTANAMIDE		C39 H50 N4 O6	
				N- [2(R)- HYDROXY- 1(S)- INDANYL]- 5- [(2(S)- TERTIARY BUTYLAMINOCARBONYL)-			

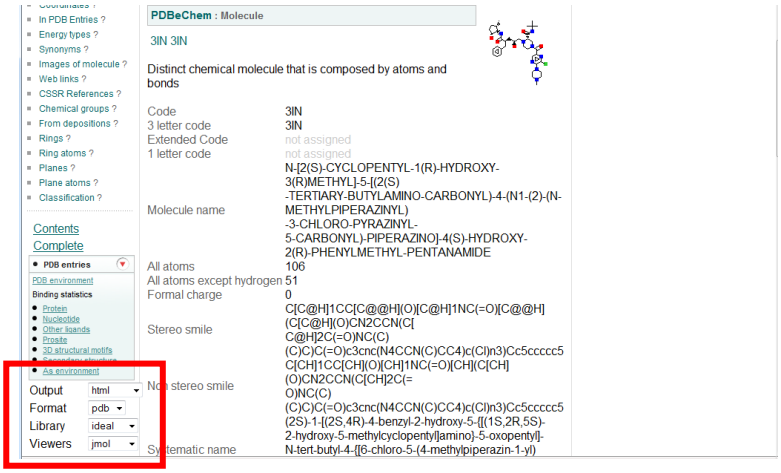
The results will show a variety of compounds that have a similar backbone structure to the one seen in 1HSG (Ligand MK1). These compounds are all various designed inhibitors of the HIV-1 protease. Click on any one of the results on this page to get more information regarding the compound. For example, click on the compounds 3IN on the results page to go to summary information regarding this compound. If the search was aborted in the previous step or takes too long, go back to the PDBeChem starting page and type in 3IN in the "3-letter code" box and click "Search". Click on the 3IN link on the next page to load the PDBeChem info page for this ligand. On the bottom left side there are various links to other resources.

Click on the "PDB entries" item on the sidebar menu to see which PDB entries have this molecule bound to them. You should see a few entries show up and the titles should tell you they are HIV-1 protease enzyme structures.



PDB Entry ID	PDB Entry Title	Resolution	Experiment Type
1c6x View	ALTERNATE BINDING SITE FOR THE P1-P3 GROUP OF A CLASS OF POTENT HIV-1 PROTEASE INHIBITORS AS A RESULT OF CONCERTED STRUCTURAL CHANGE IN 80'S LOOP.	2.50Å	X-ray
2bpy View	HIV-1 PROTEASE-INHIBITOR COMPLEX	1.90Å	X-ray
2bpz View	HIV-1 PROTEASE-INHIBITOR COMPLEX	2.50Å	X-ray

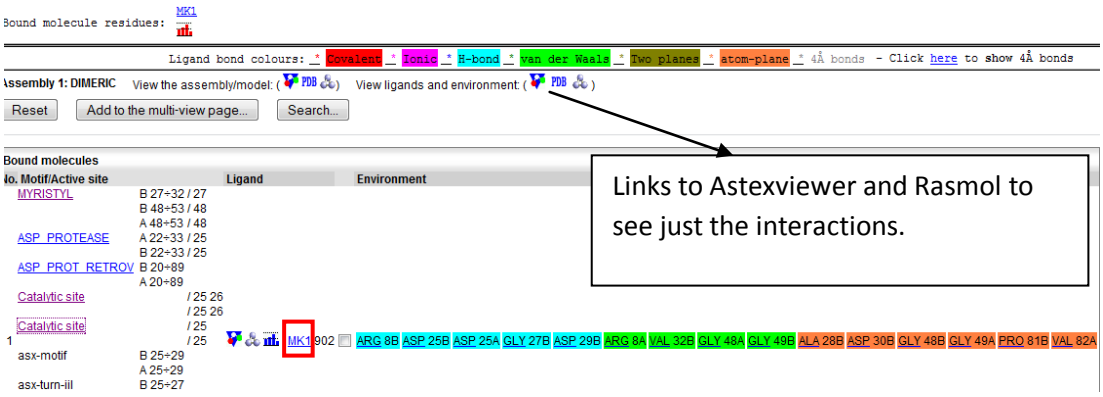
To view or download the ideal/model coordinates for this compound, choose the boxes which are highlighted on the PDBeChem page in red. Other items on the sidebar which relate to the PDBeSite database being described in the next section. The most pertinent among these is the Binding statistics link which will show you all the residues in the PDB which interact with 3IN (PDB entry 2BPZ for example).



Analyzing the Binding Site of HIV-1 protease

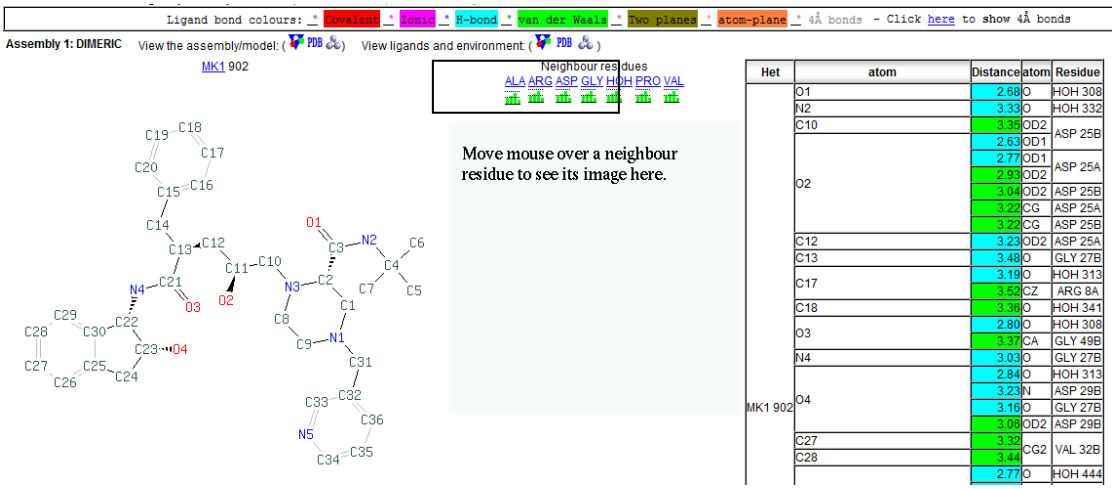
The summary pages for entries that contain any bound molecules have a 'Ligands' item in the sidebar of the atlas pages. Go to the Ligands Page for PDB entry 1HSG (http://www.ebi.ac.uk/pdbe-srv/view/entry/1hsg/summary).

Click on the "interaction statistics " link to take you to the PDEeMotif database pages for this entry. The PDBeMotif database is an integral part of the PDBe structural database and contains information about small molecule binding sites, the environment surrounding such molecules, as well as detailed information regarding interactions between small molecules.

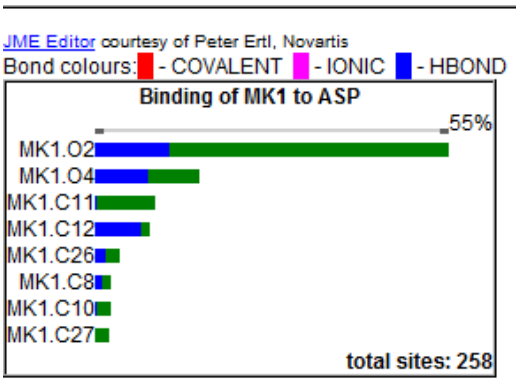


This page has information about ligand MK1 and the protein residues that interact with MK1. The interacting residues are color codes according to the nature of the interaction (cyan = H-bond, green = van der Waals, olive = two planes etc). As expected, there are two aspartic acid residues from each protease molecule that interact with the ligand.

Click on the ligand highlighted above to see all interactions between MK1 and PDB entry 1HSG.

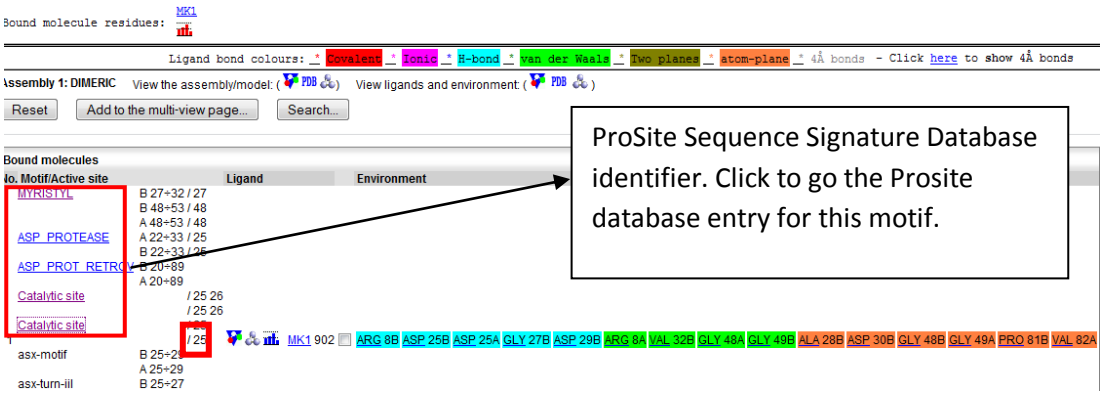


On the page as shown above, clicking on any of the green graph signs will show further detail of the nature of interactions between MK1 and the corresponding residue from the protein. Lets click on the graph symbol beneath ASP.

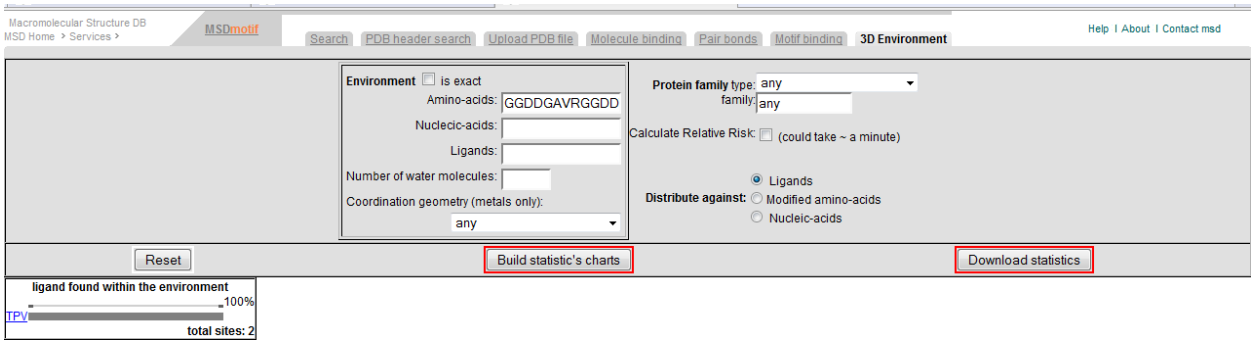


The information shown above contains a summary of all MK1-Aspartic acid residue interactions in the whole PDB. The predominant interaction appears to be interactions between atom O2 of MK1 and atoms of aspartic acid. Each bar on the graph is a link to PDB entries that contain that specific interaction only.

Lets now go back a few pages to the page which contains a summary of all interactions between MK1 and protein.

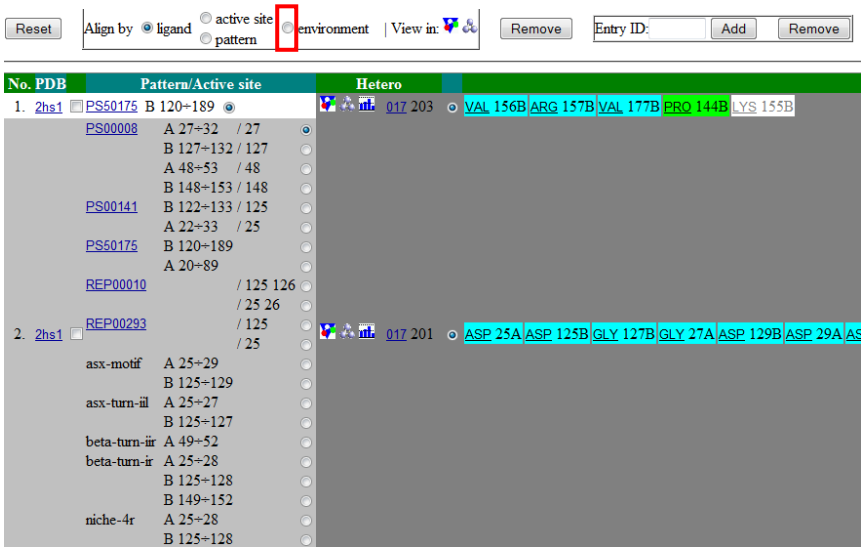


Click on the little blue graph symbol next to the MK1 link you chose before. This will show you all liagnds in the PDB that have the same environment to which MK1 binds in 1HSG.

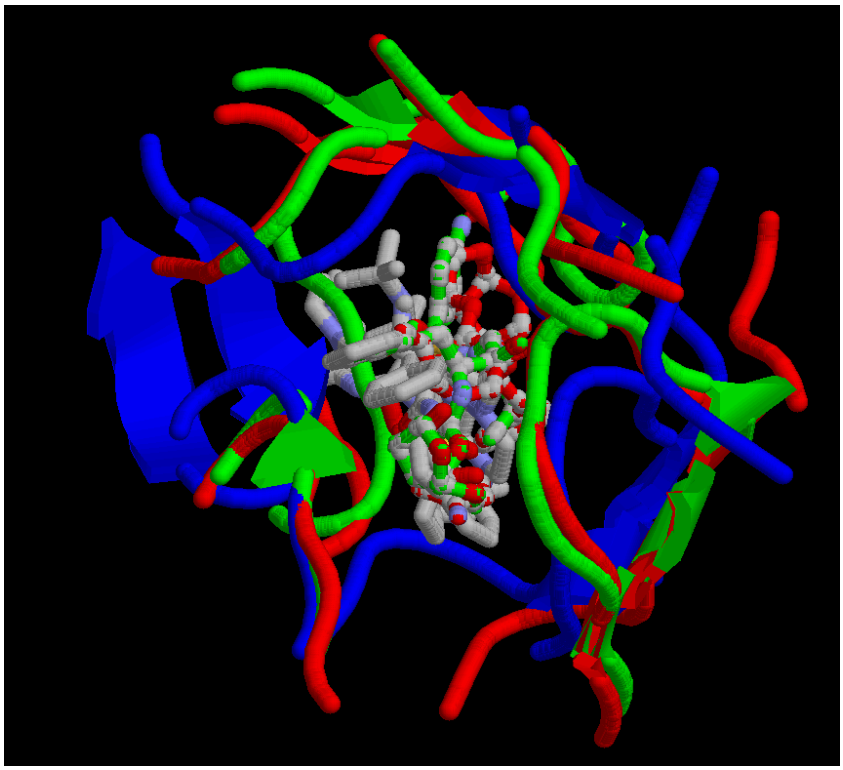


As can be seen, there is only one other ligand in the PDB that contain the same environment as in 1HSG MK1 binding site. You may modify the amino acids in the table above to see other related environments. Clicking on any of the ligands will show you just the entries where this ligand is present in the environment similar to that of MK1 in 1HSG.

Go back to the PDBeMotif summary page as before. Choose the “Add to the Multiview page” to add this structure for comparison. On the new page, lets add a few more HIV-1 protease structures we have recently encountered. Input 2BPZ in the box at the top and click add. Repeat the same step with 2HS1. This will show up a page as below. Check the boxes of 2HS1 and 2BPZ and 1HSG at places where they are with ligand 017, 3IN or MK1.



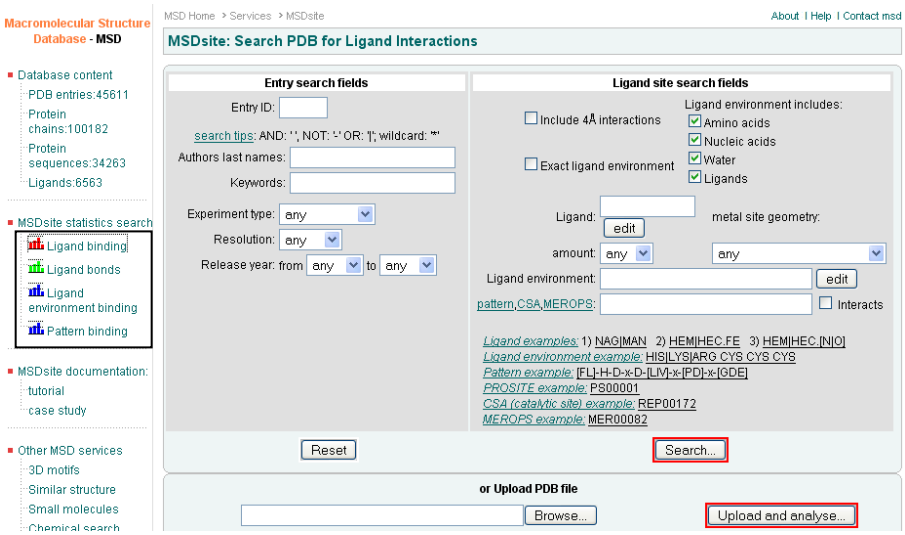
Now ensure that the box at the top says “Align by environment” and click on one of the icons to fire up a graphical viewer.



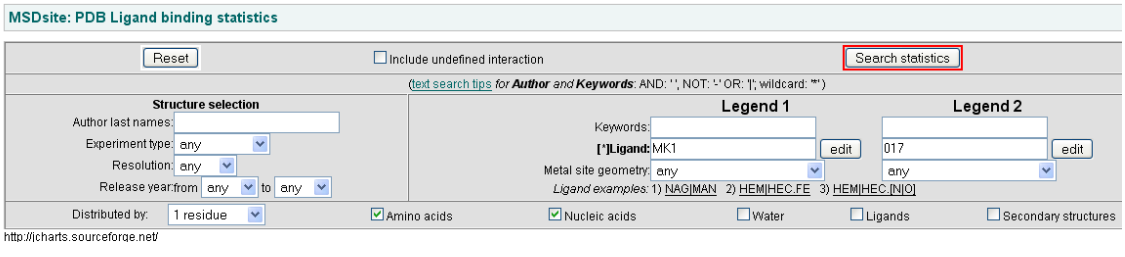
This will show the ligands with their surrounding environments overlaid on one another. It should be obvious that all the three compounds occupy roughly similar positions within the HIV-1 protease binding pocket.

Comparing binding sites of multiple compounds

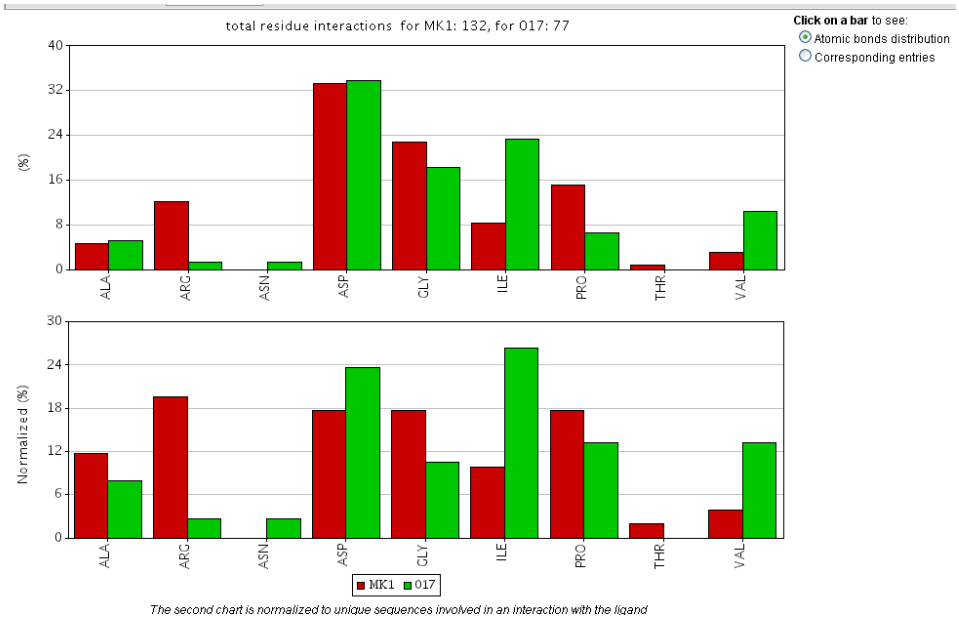
PDBeSite (http://www.ebi.ac.uk/msd-srv/msdsite/) is a service containing binding site statistics and ligand statistics in the PDB. This service has been superceded by PDBeMotif and is no longer updated but offers an easy and intuitive way to quickly analyze the differences in binding sites of similar compounds. Start with the PDBeSite main page (http://www.ebi.ac.uk/msd-srv/msdsite/). Choose the "Ligand Binding" on the left sidebar of the main page.



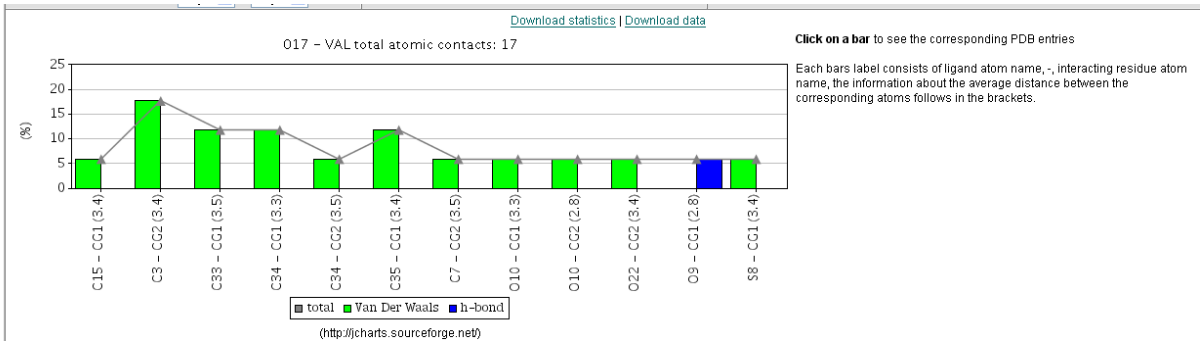
Input MK1 (PDB entry 1HSG) and 017 (PDB entry 2HS1) in the boxes as shown below. Click "Search Statistics".



The result is a graph that shows the environment preferences of MK1 and 017. The second graph is normalized. It would appear from these results that MK1 has a higher tendency to interact with ARG than 017 does. On the other hand, compound 017 tends to interact more with ILE and VAL residues in a binding pocket.



Clicking on any of the bars will show further details of ligand-residue interaction. Lets look at the interaction of VAL with 017.



As before, clicking on any of the bars on this page will show you the entries where such an interaction has been observed.

Try a similar analysis of the binding sites of GLC (alpha-D-glucose) and BGC (beta-D-glucose). You will discover that the two sugars have very different binding site residues even though they only differ in stereochemistry. PDBeSite has many other options which you are encouraged to

try out at your leisure and explore the world of small molecule-protein interactions. The other search options in PDBeSite are:

Ligand Bonds: Analyse the interaction between one small molecule and a protein residue across the whole PDB. A typical question might be: Are there any interactions between Methionine (MET) and NADP (NAP) in the whole PDB and if so, (1) what kind of interactions are these, and (2) which entries show such interaction ?

Ligand Environment Binding: What is likely to bind in a pocket comprised of residues HIS, LYS, ASP, TYR and GLY ? Go to increased complexity and specify the distance range and secondary structure constraints on the environment.

Pattern Binding: Look at standard sequence signatures and see if there are any compounds in the PDB that bind such a pattern. For example. What would be likely to bind to a sequence pattern [FL]-H-D-x-D-[LIV]-x-[PD]-x-[GDE] ?

Looking for other structures that have a fold similar to HIV-1 proteases.

You can also use PDBe services (PDBefold: http://www.ebi.ac.uk/msd-srv/ssm/) to look for other structures in the PDB that have similar topology and fold to our query structure 1HSG. This may be done from within the PDBe summary page for this entry (http://www.ebi.ac.uk/pdbe-srv/view/entry/1hsg/summary). Expand the Links section on the bottom left of the page and choose PDBeSSM.

This will fire off a PDBeFold computation job on the PDBe computer farms which will compare the overall fold of PDB entry 1HSG against the whole PDB archive. The results at the end of the computation will look similar to the one below.

Structure Alignment Results

explanation of output

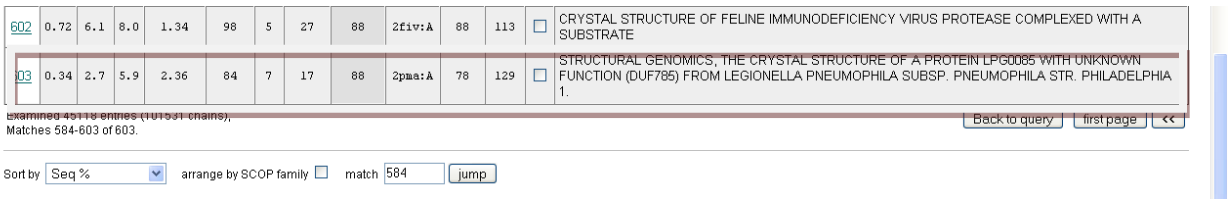
Query: pdb entry **1hsg**, chain A : 99 residues.
CRYSTAL STRUCTURE AT 1.9 ANGSTROMS RESOLUTION OF HUMAN IMMUNODEFICIENCY VIRUS (HIV) II PROTEASE COMPLEXED WITH L- 735,524, AN ORALLY BIOAVAILABLE INHIBITOR OF THE HIV PROTEASES

Examined 45118 entries (101531 chains),
Matches 1-20 of 603.

Back to query | >> | last page
resort results

##	Scoring			Rmsd	N_{algn}	N_g	$\%_{seq}$	Query	Target (PDB entry)				
	Q	P	Z					$\%_{sse}$	Match	$\%_{sse}$	N_{res}	×	Title
1	1.00	14.9	11.4	0.00	99	0	100	100	1hsg:A	100	99	☐	HIV-1 PROTEASE; 1HSG 4
2	0.99	10.1	9.6	0.25	99	0	100	75	1ebw:A	100	99	☐	HIV-1 PROTEASE IN COMPLEX WITH THE INHIBITOR BEA322
3	0.99	12.5	10.6	0.26	99	0	100	88	1wbm:A	100	99	☐	HIV-1 PROTEASE IN COMPLEX WITH SYMMETRIC INHIBITOR, BEA450
4	0.99	7.9	9.6	0.26	99	0	100	75	1ec2:A	86	99	☐	HIV-1 PROTEASE IN COMPLEX WITH THE INHIBITOR BEA428
5	0.99	10.6	10.6	0.26	99	0	100	88	2bpv:A	88	99	☐	HIV-1 PROTEASE-INHIBITOR COMPLEX
6	0.99	12.5	10.6	0.26	99	0	100	88	1ec3:A	100	99	☐	HIV-1 PROTEASE IN COMPLEX WITH THE INHIBITOR MSA367
7	0.99	10.1	9.6	0.27	99	0	100	75	1ebz:A	100	99	☐	HIV-1 PROTEASE IN COMPLEX WITH THE INHIBITOR BEA388
8	0.99	12.5	10.6	0.27	99	0	99	88	4phv:A	100	99	☐	HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 (HIV-1) PROTEASE COMPLEX 4PHV 3 WITH N,N-BIS(2(R)-HYDROXY-1(S)-INDANYL)-2,6-(R,R)- 4PHV 4 DIPHENYLMETHYL-4-HYDROXY-1,7-HEPTANDIAMIDE 4PHV 5
9	0.99	10.1	9.6	0.27	99	0	100	75	1d4i:A	100	99	☐	HIV-1 PROTEASE IN COMPLEX WITH THE INHIBITOR BEA425
10	0.99	10.0	9.6	0.27	99	0	100	75	1ec1:A	100	99	☐	HIV-1 PROTEASE IN COMPLEX WITH THE INHIBITOR BEA409
11	0.99	10.1	9.6	0.28	99	0	100	75	1w5v:A	100	99	☐	HIV-1 PROTEASE IN COMPLEX WITH FLUORO SUBSTITUTED DIOL- BASED C2-SYMMETRIC INHIBITOR
12	0.99	9.6	10.1	0.29	99	0	100	88	2bpw:A	88	99	☐	HIV-1 PROTEASE-INHIBITOR COMPLEX
13	0.99	10.5	10.6	0.29	99	0	100	88	2bpy:A	88	99	☐	HIV-1 PROTEASE-INHIBITOR COMPLEX
14	0.99	10.1	9.6	0.29	99	0	100	75	1w5y:A	100	99	☐	HIV-1 PROTEASE IN COMPLEX WITH FLUORO SUBSTITUTED DIOL- BASED C2-SYMMETRIC INHIBITOR
15	0.99	10.1	9.6	0.29	99	0	100	75	2hvw:A	100	99	☐	HIV-1 PROTEASE IN COMPLEX WITH INHIBITOR AHA455

There are over 700+ structures where 70% of the matched structure has similar topology with 70% of the query structure. Scroll to the bottom of the page and choose to sort the results by "Seq %". This will update the page and rank entries according to sequence identity. Now click on the button at the top of the page that says "last page". The last hit on this page appears interesting.



602	0.72	6.1	8.0	1.34	98	5	27	88	2fiv:A	88	113	☐	CRYSTAL STRUCTURE OF FELINE IMMUNODEFICIENCY VIRUS PROTEASE COMPLEXED WITH A SUBSTRATE
603	0.34	2.7	5.9	2.36	84	7	17	88	2pma:A	78	129	☐	STRUCTURAL GENOMICS, THE CRYSTAL STRUCTURE OF A PROTEIN LPG0085 WITH UNKNOWN FUNCTION (DUF785) FROM LEGIONELLA PNEUMOPHILA SUBSP. PNEUMOPHILA STR. PHILADELPHIA 1.

Examined 45118 entries (101531 chains),
Matches 584-603 of 603.

Back to query | first page | <<

Sort by Seq % | arrange by SCOP family ☐ | match 584 | jump

Match 721 of 721

Back to match list | first match | << | Back to query

Query PDB 1hsg:A				Alignment				Target PDB 2pma:A			
N_{res}	%res	N_{SSE}	%SSE	Q	P	RMSD	N_{align}	N_{res}	%res	N_{SSE}	%SSE
99	85	8	88	0.342	2.73	2.357	84	129	65	9	78
CRYSTAL STRUCTURE AT 1.9 ANGSTROMS RESOLUTION OF HUMAN IMMUNODEFICIENCY VIRUS (HIV) II PROTEASE COMPLEXED WITH L- 735,524, AN ORALLY BIOAVAILABLE INHIBITOR OF THE HIV PROTEASES				%seq	Z	N_{SSE}	N_{gaps}	STRUCTURAL GENOMICS, THE CRYSTAL STRUCTURE OF A PROTEIN LPG0085 WITH UNKNOWN FUNCTION (DUF785) FROM LEGIONELLA PNEUMOPHILA SUBSP. PNEUMOPHILA STR. PHILADELPHIA 1.				
				16.7	5.85	7	7				

view | download | view superposed | view | download

superpose whole entries Viewer: Jmol

Secondary Structure Alignment

```
1hsg:A   -SSSsSS-HS
2pma:A   sSSS-SSsHS
```

Query PDB 1hsg:A / **Target PDB 2pma:A**

```
1|SD    6|A|LEU  10 |ILE  15 |    <->    2|SD    7|A|VAL  30 |LEU  35 |
2|SD    7|A|GLN  18 |LEU  24 |    <->    3|SD    7|A|LEU  40 |LEU  46 |
3|SD    3|A|VAL  32 |GLU  34 |    <->    4|SD    5|A|ALA  53 |HIS  56 |
5|SD   15|A|GLY  52 |ILE  66 |    <->    5|SD   12|A|LYS 112 |LEU 122 |
6|SD   10|A|HIS  69 |GLY  78 |    <->    6|SD   14|A|LYS 125 |THR 134 |
7|H1    6|A|GLY  86 |THR  91 |    <->    8|H1    8|A|GLY 146 |PHE 153 |
8|SD    3|A|THR  96 |ASN  98 |    <->    9|SD    4|A|GLY 155 |ASP 158 |
```

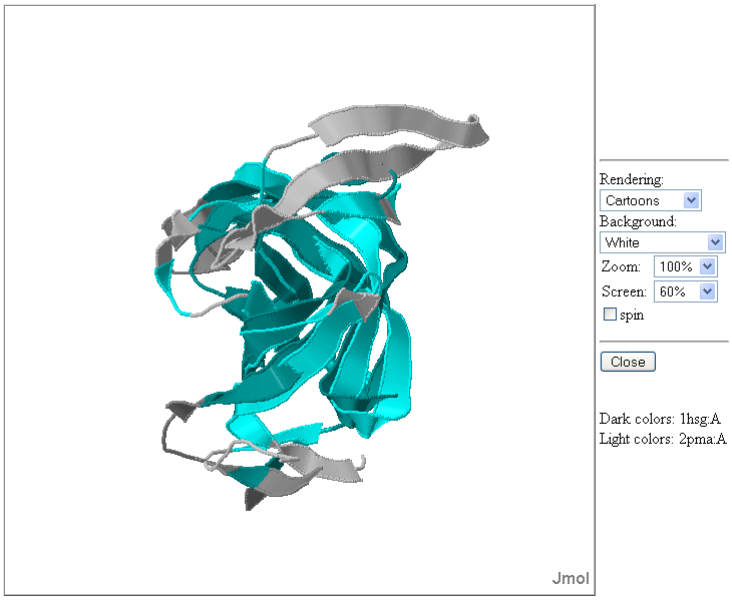
SCOP: domain 26541 , family b.50.1.1

PDBe Atlas | PDBe Motif | OCA | GeneCensus | FSSP | 3Dee | CATH | PDBsum

PDBe Atlas | PDBe Motif | OCA | GeneCensus | FSSP | 3Dee | CATH | PDBsum

view | download sequence | view superposed | view | download sequence

The overall RMSD between the two structures (1HSG and 2PMA) is 2.35A over 84 residues that could be aligned. The overall sequence identity is only 16.7%. Click on the "View superposed" button to open this structure up in Jmol.

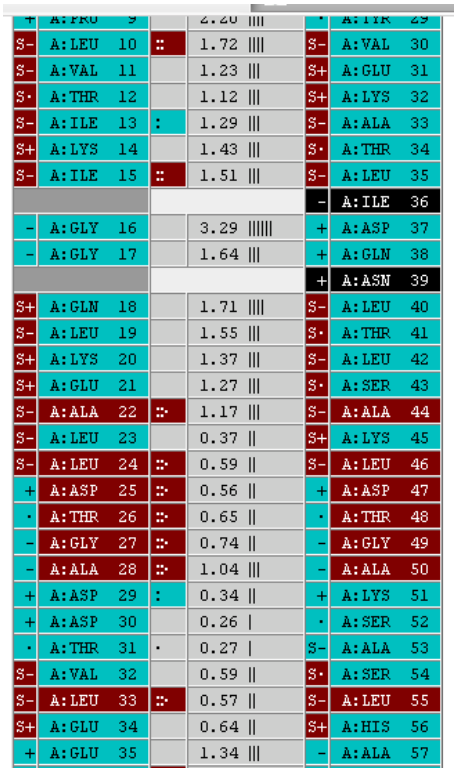


The query structure 1HSG is in cyan and the matched structure (2PMA) is in grey. The two structures have almost identical secondary structure core structures. The matched structure is a structural genomics consortium and has no assigned function.

In 1HSG the residues that were important for binding the inhibitor were ARG 8 ASP 25 ASP 29, GLY 27 ASP 30 VAL 32 GLY 48 GLY 49 PRO 81.

Go back to the match page between 2PMA and 1HSG and scroll down to the residue-by-residue listing and see if 2PMA has any similar residues in structurally corresponding positions.

Residues 24 to 28 in our query structure are Leu, Asp, Thr, Gly and Ala. Intriguingly enough, this stretch is exactly identical to residues 46 to 50 in 2PMA. However, the other aspartic acid residues at position 29 and 30 are not conserved in 2PMA, but are replaced with Lys and Ser at the corresponding positions. The narrow bars in the the picture with numbers between the two protein sequences indicate the distance between the corresponding positions in a 3D alignment. It would appear from these results that residues 19-35 in 1HSG and 41-57 align very closely with each other in 3d-space. However, there are large sections of 2PMA for which there are no equivalent residues in 1HSG.



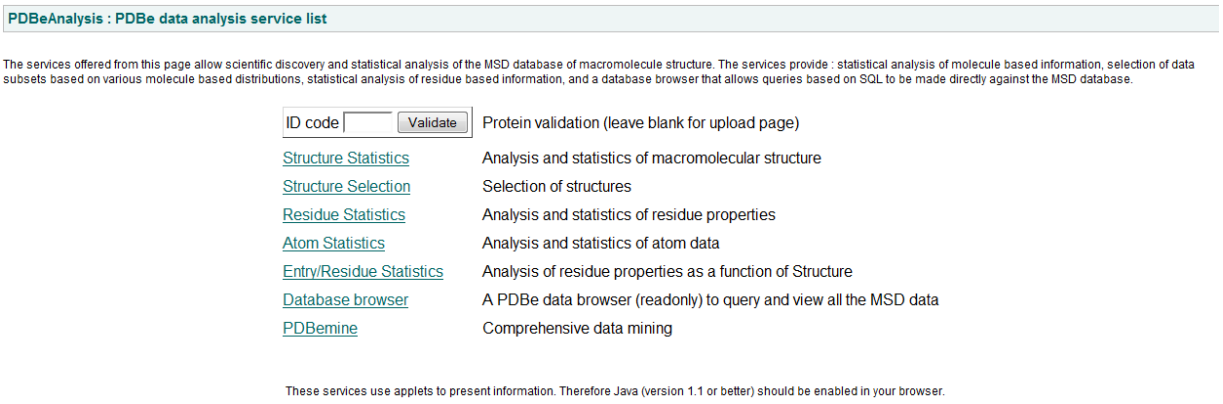
S-	A:LEU	10	::	1.72	S-	A:VAL	30
S-	A:VAL	11		1.23	S+	A:GLU	31
S·	A:THR	12		1.12	S+	A:LYS	32
S-	A:ILE	13	:	1.29	S-	A:ALA	33
S+	A:LYS	14		1.43	S·	A:THR	34
S-	A:ILE	15	::	1.51	S-	A:LEU	35
					-	A:ILE	36
-	A:GLY	16		3.29	+	A:ASP	37
-	A:GLY	17		1.64	+	A:GLN	38
					+	A:ASN	39
S+	A:GLN	18		1.71	S-	A:LEU	40
S-	A:LEU	19		1.55	S·	A:THR	41
S+	A:LYS	20		1.37	S-	A:LEU	42
S+	A:GLU	21		1.27	S·	A:SER	43
S-	A:ALA	22	:·	1.17	S-	A:ALA	44
S-	A:LEU	23		0.37	S+	A:LYS	45
S-	A:LEU	24	:·	0.59	S-	A:LEU	46
+	A:ASP	25	:·	0.56	+	A:ASP	47
·	A:THR	26	:·	0.65	·	A:THR	48
-	A:GLY	27	:·	0.74	-	A:GLY	49
-	A:ALA	28	:·	1.04	-	A:ALA	50
+	A:ASP	29	:	0.34	+	A:LYS	51
+	A:ASP	30		0.26	·	A:SER	52
·	A:THR	31	·	0.27	S-	A:ALA	53
S-	A:VAL	32		0.59	S·	A:SER	54
S-	A:LEU	33	:·	0.57	S-	A:LEU	55
S+	A:GLU	34		0.64	S+	A:HIS	56
+	A:GLU	35		1.34	-	A:ALA	57

In order to find even more distant matches one can lower the acceptable match criteria to 60%. In order to do this, click on the "Back to Query Button" at the top of the page, and change the value of 70% to 60% in both the boxes. Resubmit the job in the same session. Once this match is finished, sort the results according to Seq % and scroll to the last page. This now throws up 630+ results. Look at any of the last few results. These should indicate that there are other structures (for example Cathepsin D) which have some folds similar to our query structure. This is understandable as Cathepsin D also belongs to the protease family and are therefore, likely to share some evolutionary relationships. Look over the residue by residue listing to see if there are any important catalytic residues conserved between the two proteins.

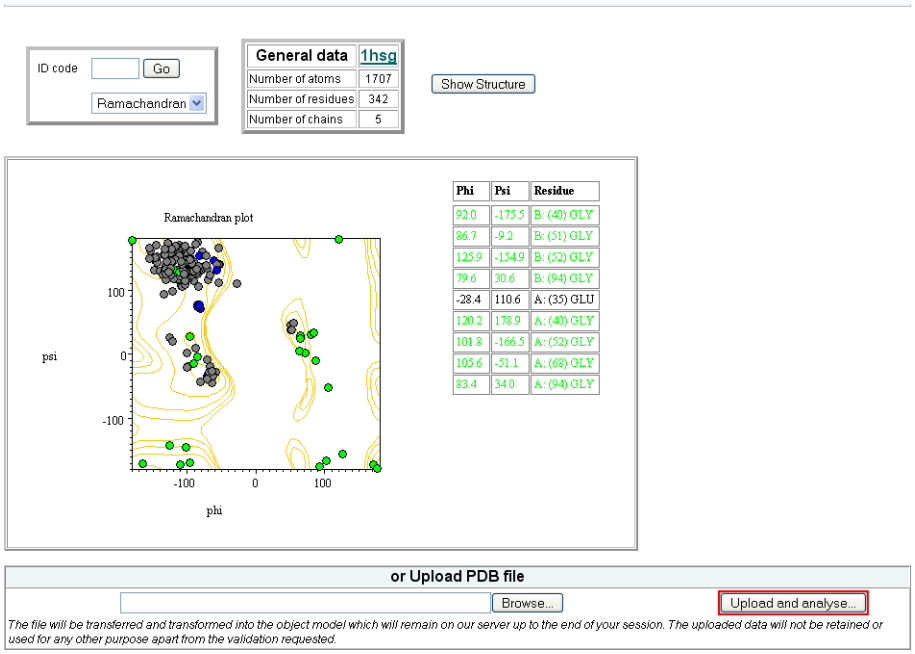
You can also upload your own PDB file, or sets of PDB files into PDBeFold for pairwise alignment between themselves or against the whole PDB archive. You can also align your structures against SCOP representative sets. Explore PDBeFold at your leisure at any time.

Structural Quality Assessment using PDBeAnalysis

The PDBeAnalysis service provides a quick way of assessing the quality of structures in the PDB or for uploaded files in PDB format. Start PDBeAnalysis from the PDBe home page (http://www.ebi.ac.uk/pdbe) and choosing the relevant service from the list provided.



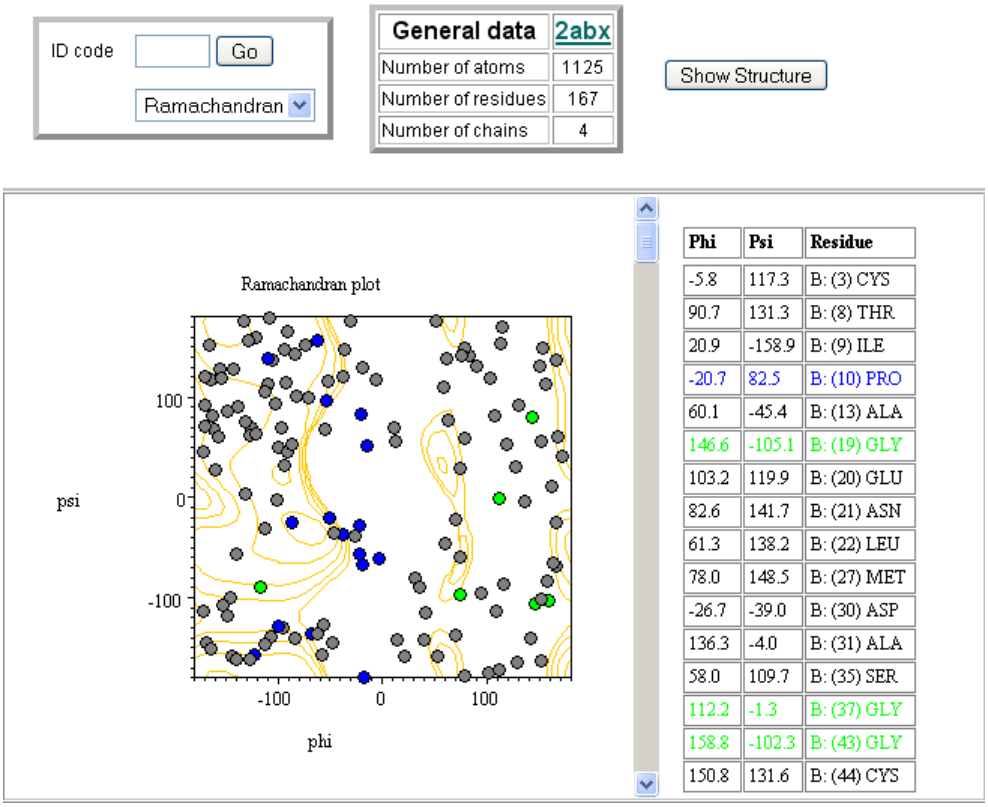
Type in 1HSG and click on the validate button. This will open up a page with the Ramachandran Graph of the structure showing any stereochemical outliers present in the structure.



The residues highlighted in green are glycine residues which are flexible. However, there is one outlier (residue A35 GLU). In an ideal structure there should be no outliers but it is not uncommon to see structures with a couple of outliers from standard graphs. Clicking on any outlier in the table will highlight the point in the graph. Use the dropdown on the top right to see other quality graphs for this structure.

Compare this graph with that of another entry. In the IDcode box at the top of the page, type in 2ABX and click “Go”.

A large number of residues of this protein are outside the allowed torsion angle area. This indicates that this structure is of poor quality and should only be used with extreme caution for any analysis. Any conclusions made from this structure could have serious errors. Also look at other “dodgy” entries like 1CYC and 1PYP. In contrast, it should be clear that 1HSG is a structure which has no real quality concerns as far as geometric parameters go.



Conclusions

We hope this tutorial has introduced you to the various PDBe services and tools that are available for the search, analysis and retrieval of protein structures. Detailed help is available for every service on the top right side of the page. There are other tutorials which delve in greater detail into our services. These tutorials can be accessed from http://www.ebi.ac.uk/pdbe/docs/education/Tutorial.html. In case you wish need to contact the PDBe, please write to msdhelp@ebi.ac.uk with your query and we will be happy to assist you in every way possible.