IJPSR (2010), Vol. 1, Issue 10 **(Review Article)**

INTERNATIONAL JOURNAL OF PHARMACEUTICAL SCIENCES AND RESEARCH

Received on 15 June, 2010; received in revised form 11 August, 2010; accepted 17 September, 2010

NOVEL NUCLEAR RECEPTOR SUPER FAMILY DATABASE

Akshay A. Bhat [1], Vinay Sridhar [1], Hemanth Kumar K. [1], Abhilash M.*[2]

Department of Biotechnology, the Oxford College of Engineering [1], Bangalore, Karnataka, India
UST Global, Technopark [2], Trivandrum, Kerala, India

ABSTRACT

The objective of the project is to develop a comprehensive database of Nuclear Receptor (NR) Superfamily-Ligand. Nuclear Receptor Superfamily represents one of the most important families of drug targets in pharmaceutical development. Nuclear Receptor) Superfamily-Ligand is a novel public NR-related chemical genomic database that is primarily focused on the correlation of information between NRs and their ligands. It provides correlation data between NRs and their ligands, along with chemical information on ligands, as well as access information to the various web databases regarding NRs. These data are connected with each other in a relational database, allowing users in the field of NR-related drug discovery to easily retrieve such information from either biological or chemical starting points. NR Superfamily-ligand database includes structure similarity search functions for the NRs and for their ligands. Thus database can provide correlation maps linking the searched homologous NRs (or ligands) with their ligands; we can gain more detailed knowledge about their interactions and improve drug design efforts by focusing on inferred candidates for NR-specific drugs.

Keywords:
NR,
Ligands,
Docking,
Database

Correspondence to Author:

Abhilash M.

UST Global Technopark,
Trivandrum , Kerala, India

INTRODUCTION: In the field of molecular biology, nuclear receptors are a class of proteins found within the interior of cells that are responsible for sensing the presence of steroid and thyroid hormones and certain other molecules. In response, these receptors work in concert with other proteins to regulate the expression of specific genes thereby controlling the development, homeostasis, and metabolism of the organism. Nuclear receptors (Table 1) are one of the most abundant classes of transcriptional regulators in animals (metazoans). They regulate diverse functions, such as homeostasis, reproduction, development and metabolism nuclear receptors have the ability to directly bind to DNA and regulate the expression of adjacent genes, hence these receptors are classified as transcription factors. The regulation of gene expression by nuclear receptors only happens when a ligand- a molecule which affects the receptor's behavior is present. More specifically, ligand binding to a nuclear receptor results in a conformational change in the receptor which in turn activates the receptor resulting in up-regulation of gene expression.

An orphan receptor is an apparent receptor that has a similar structure to other identified receptors but whose endogenous ligand has not yet been identified. If a ligand for an orphan receptor is later discovered, the receptor is referred to as an "adopted orphan". Examples of orphan receptors are found in the G protein-coupled receptor (GPCR) and nuclear receptor families. GPCR orphan receptors are usually given the name "GPR" followed by a number, for example GPR1. Adopted orphan receptors in the nuclear receptor group include the farnesoid X receptor (FXR), liver X receptor (LXR), and peroxisome proliferator-activated receptor (PPAR).

An assessment of the number of molecular targets that represent an opportunity for therapeutic intervention is crucial to the development of post-genomic research strategies within the pharmaceutical industry. Now that we know the size of the human genome, it is interesting to consider just how many molecular targets this opportunity represents. We start from the position that we understand the properties that are required for a good drug, and therefore must be able to understand what makes a good drug target [1].

The nuclear receptor superfamily describes a related but diverse array of transcription factors, which include nuclear hormone receptors (NHRs)[2] and orphan nuclear receptors. NHRs are receptors for which hormonal ligands have been identified, whereas orphan receptors are so named because their ligands are unknown, at least at the time the receptor is identified. Unlike hormones for cell surface receptors, lipophilic hormones can traverse the plasma membrane to the cell interiors where NHRs transduce signals from glucocorticoids, mineralocorticoids, the sex steroids (estrogen, progesterone, and androgen), thyroid hormones, and vitamin D_3. All of the nuclear receptors have common structural features

Given the widespread relevance of the superfamily of nuclear receptors to almost all aspects of normal human physiology, the role of these receptors in the etiology of many human diseases, and their importance as therapeutic targets for pharmaceuticals, it is obvious that a detailed understanding of these systems has major implications, not only for human biology but also for the understanding and development of new drug treatments [3].

Predicting interactions between small molecules and proteins is a key element in the drug discovery process. In particular, several classes of proteins such as G-protein-coupled receptors (GPCR), enzymes and ion channels represent a large fraction of current drug targets and important targets for new drug development. Understanding and predicting the interactions between small molecules and such proteins could therefore help in the discovery of

new lead compounds. Various approaches have already been developed and have proved very useful to address this *in silico* prediction issue [4]. The classical paradigm is to predict the modulators of a given target, considering each target independently from other proteins. Usual methods are classified into *ligand-based* and *structurebased* or *docking* approaches. Ligand-based approaches compare a candidate ligand to the known ligands of the target to make their prediction, typically using machine learning algorithms [5] whereas structure-based approaches use the 3D-structure of the target to determine how well each candidate binds the target.

Ligand-based approaches require the knowledge of sufficient ligands of a given target with respect to the complexity of the ligand/non-ligand separation to produce accurate predictors. If few or no ligands are known for a target, one is compelled to use docking approaches, which in turn require the 3D structure of the target and are very time consuming [6]. If for a given target with unavailable 3D structure no ligand is known, none of the classical approaches can be applied. This is the case for many GPCR as very few structures have been crystallized so far and many of these receptors, referred to as *orphan* GPCR, have no known ligand.

An interesting idea to overcome this issue is to stop considering each protein target independently from other proteins, and rather take the point of view of *chemogenomics*. Roughly speaking, chemogenomics aims at mining the entire *chemical space*, which corresponds to the set of all small molecules, for interactions with the *biological space*, i.e. the set of all proteins or at least protein families, in particular drug targets [7, 8]. A salient motivation of the chemogenomics approach is the realization that some classes of molecules can bind 'similar' proteins, suggesting that the knowledge of some ligands for a target can be helpful to determine ligands for similar targets. Besides, this type of method allows for a more rational approach to design drugs since controlling a whole ligand's selectivity profile is crucial to make sure that no side effect occurs and that the compound is compatible with therapeutical usage [9, 10].

TABLE 1: TYPES OF NUCLEAR RECEPTORS AND ITS THERAPEUTIC RELEVANCE

NAME	SUBTYPES	THERAPEUTIC RELEVANCE
Estrogen receptors	ERR-alpha, beta, gamma	muscle fatty acid metabolism
RAR-Related Receptor	ROR-alpha, beta, gamma	Cerebellum Development, bone maintenance, circadian rhythm.
Human nuclear factor 4	HNF4-alpha,gamma	Role in diabetes
Reverse erbA	Rev-erbA-alpha, beta	circadian rhythm
Testis receptor	TR2,TR4	Unknown
Tailless-like photoreceptor	TLX,PNR	Neural development, Photoreceptor differentiation.
Chicken Ovalbumin	COUP-TFI, TFII.	Neural development, vascular Development
Upstream promoter transcription factor	COUP-TFIII	Neural development, vascular Development
NGF induced factor B	NGFIB alpha	Role in thymocyte apoptosis
Nur-Related Factor 1	NGFIB beta	Role in dopaminergic neuron development
Neuron derived orphan receptor 1	NGFIB gamma	unknown
Steroidogenic factor 1	SF1	Role in mammalian sexual development
Liver receptor homologous protein 1	LRH1	Role in lipid homeostasis, cell cycle control
Germ cell nuclear factor	GCNF	Role in vertebrate embryogenesis

MATERIALS:

Pubchem: PubChem is a database of chemical molecules. The system is maintained by the national center for Biotechnology Information (NCBI), a component of the national library of medicine, which is part of the United States National Institutes of Health (NIH). PubChem contains substance descriptions and small molecules with fewer than thousand atoms and thousand bonds. After the initial literature survey, this database was used to obtain closely related compounds for the specific receptors viz., Ligands.

Marvin Sketch: Marvin Sketch is the molecule drawing tool of Marvin, a chemical structure drawing and visualizing package, including an integrated chemical file format converter. Marvin consists of Marvin Sketch, Marvin View (single and multiple structure viewer), Marvin Space (high performance 3D molecule visualizer) and Mol Converter (file format batch-conversion tool). We mainly used it for energy minimization and chemical file conversion of the formats. The potential energy calculated by summing the energies of various interactions is a numerical value for a single conformation.

This number can be used to evaluate a particular conformation, but it may not be a useful measure of a conformation because it can be dominated by a few bad interactions. For instance, a large molecule with an excellent conformation for nearly all atoms can have a large overall energy because of a single bad interaction, for instance two atoms too near each other in space and having a huge van der Waals repulsion energy. It is often preferable to carry out energy minimization on a conformation to find the best nearby conformation. Energy minimization is usually performed by gradient optimization: atoms are moved so as to reduce the net forces on them. The minimized structure has small forces on each atom and therefore serves as an excellent starting point for molecular dynamics simulations.

PDB: The Protein Data Bank (PDB) is repository for the 3-D structural data of large biological molecules, such as proteins and nucleic acids. The data, typically obtained by X-ray crystallography or NMR spectroscopy and submitted by biologists and biochemists from around the world, can be accessed at no charge on the internet. We used PDB to obtain the IDs of the receptors and we also used it for modeling receptors using Spdb viewer.

HEX: *Hex* is an interactive molecular graphics program for calculating and displaying feasible docking modes of pairs of protein and DNA molecules. *Hex* can also calculate protein-ligand docking, assuming the ligand is rigid, and it can superpose pairs of molecules using only knowledge of their 3D shapes. It is the only docking and superposition program to use *spherical polar Fourier* (SPF) correlations to accelerate the calculations of docking.

In *Hex*'s docking calculations, each molecule is modelled using 3D expansions of real orthogonal spherical polar basis functions to encode both surface shape and electrostatic charge and potential distributions. Essentially, this allows each property to be represented by a vector of coefficients (which are the components of the basis functions). *Hex* represents the surface shapes of proteins using a two-term surface skin plus van der Waals steric density model, whereas the electrostatic model is derived from classical electrostatic theory. By writing expressions for the *overlap* of pairs of parametric functions, one can obtain an overall docking score as a function of the six degrees of freedom in a rigid body docking search. With suitable scaling factors, this docking score can be interpreted as an interaction energy, which we seek to minimize.

SPDBV (Swiss PDB Viewer): Swiss-Pdb Viewer (aka Deep View) is an application that provides a

user friendly interface allowing analyzing several proteins at the same time. The proteins can be superimposed in order to deduce structural alignments and compare their active sites or any other relevant parts. Amino Acid mutations, H-bonds, angles and distances between atoms are easy to obtain thanks to intuitive graphic and menu interface. Working with these two programs greatly reduces the amount of work necessary to generate models, as it is possible to thread a protein primary sequence onto a 3D template and get an immediate feedback of how well the threaded protein will be accepted by the reference structure before submitting a request to build missing loops and refine side chain packing. We used this program to model receptor proteins which were not available in PDB.

QUANTUM: QUANTUM docking/library screening software docks a small-molecule on the active site of a protein, calculates IC50 (Kd, Ki, pKd) value for a protein-ligand complex, and performs screening of a compound library against a target protein. The QUANTUM drug discovery software was developed using a new paradigm in molecular modeling-applying fast quantum and molecular physics instead of the statistical approaches scoring-function-like and QSAR-like methods. In QUANTUM binding affinity of a protein-ligand complex is estimated on the basis of free binding energy calculations.

METHODS:

Collection of Molecules: Literature review was carried out to obtain a list of natural chemo preventive agents targeting the receptors. The literature review was carried out from journals like:

a. Pubmed central
b. Oxford journals
c. AACR Journals
d. Nature
e. Springerlink
f. Biomed Central
g. Science Direct etc.

We worked with fourteen orphan nuclear receptors, having twenty three subtypes in total. Chemical information regarding the nuclear receptors and ligands had to be found. Swiss Prot was used to obtain the PDB ID of the receptors and the correct crystal structure of this receptor was stored (Example: 1A6Y.pdb). An excel sheet was created to store all the chemical information regarding the ligands and receptors. Drug similarity and information related to 'Ligand chemical component' were assorted and stored in this excel sheet for further use.

Modeling the Protein: The receptors which did not have their crystal structure stored in the protein data bank had to be modeled using software called Swiss PDB viewer. This was done by obtaining the fasta sequence of the receptor from NCBI and saving it. The blast operation was performed for choosing the most similar sequence which could be used as the 'template' (score greater than 200). The structure of the template was downloaded from PDB and saved. The raw sequence (Fasta) and the template were loaded into the software program. Operations like 'Fit raw sequence', 'Magic fit' were done and this was submitted. The modeled structure which was obtained was saved for further use.

Search for Potential Ligands for the Orphan Nuclear Receptors: This was mainly done by the following two methods;

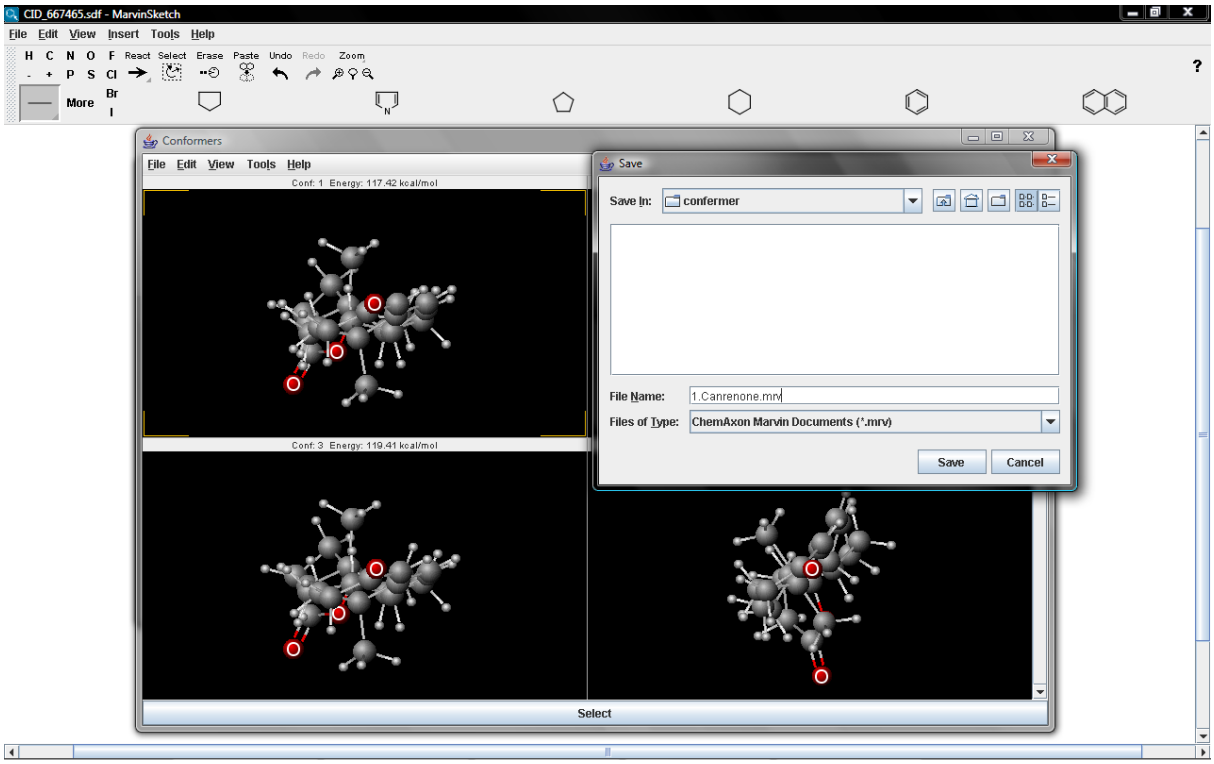
- Literature review
- Databases

Extensive searching of research papers was done to find any chemical compounds which would act as ligands to the specific receptors. This involved going through current research papers to check if any new compounds were present that could be potential ligands. Databases like Pubchem and Pubmed were used to search for

compounds which were similar to those obtained by the literature survey. This yielded closed to Two Hundred potential ligands for each receptor. All this information was assorted and saved in the above mentioned excel sheet.

Software Programs:

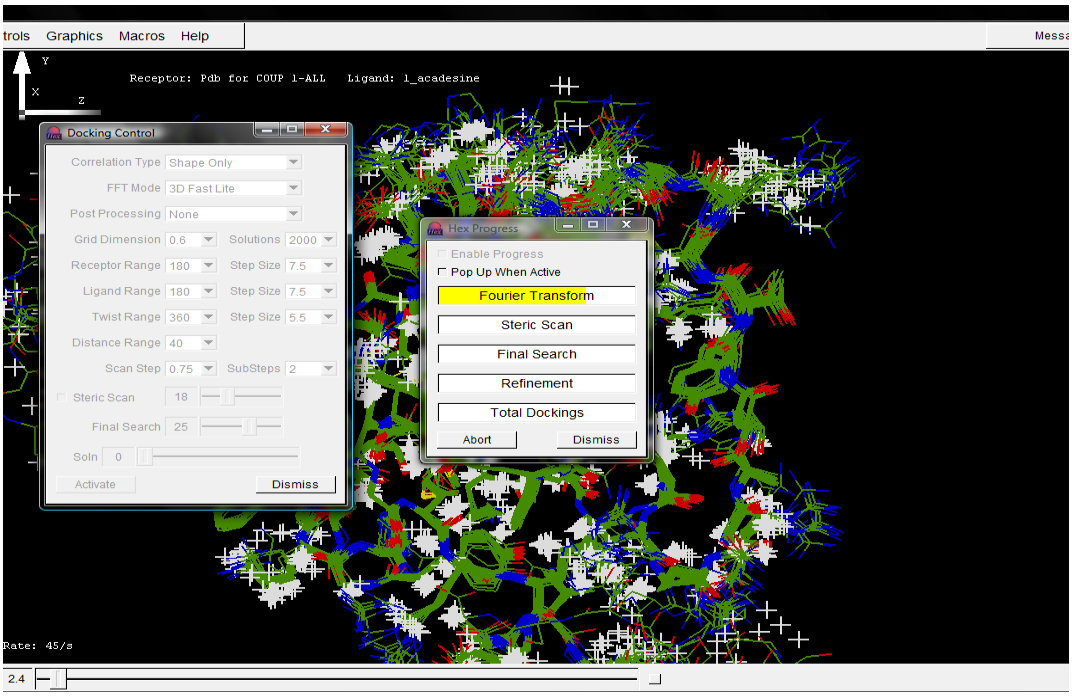
Marvin Sketch: Energy minimization was done on all the ligands using software called 'Marvin sketch'. The potential energy calculated by summing the energies of various interactions is a numerical value for a single conformation. This number can be used to evaluate a particular conformation, but it may not be a useful measure of a conformation because it can be dominated by a few bad interactions. For instance, a large molecule with an excellent conformation for nearly all atoms can have a large overall energy because of a single bad interaction, for instance two atoms too near each other in space and having a huge van der Waals repulsion energy. It is often preferable to carry out energy minimization on a conformation to find the best nearby conformation. Energy minimization is usually performed by gradient optimization: atoms are moved so as to reduce the net forces on them. The minimized structure has small forces on each atom and therefore serves as an excellent starting point for molecular dynamics simulations. Marvin Sketch is also used to change the format of the file so that they can be used by different programs (From .mrv to .sdf and .pdb).



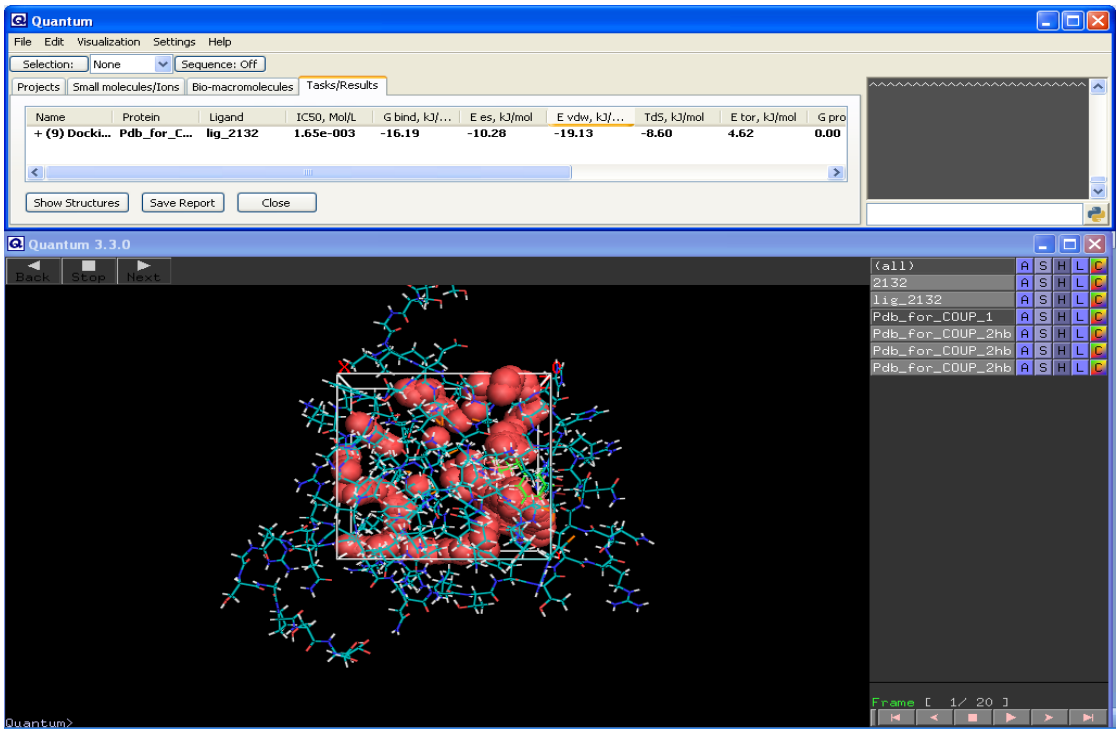
The main docking programs used were;

- Hex
- Quantum

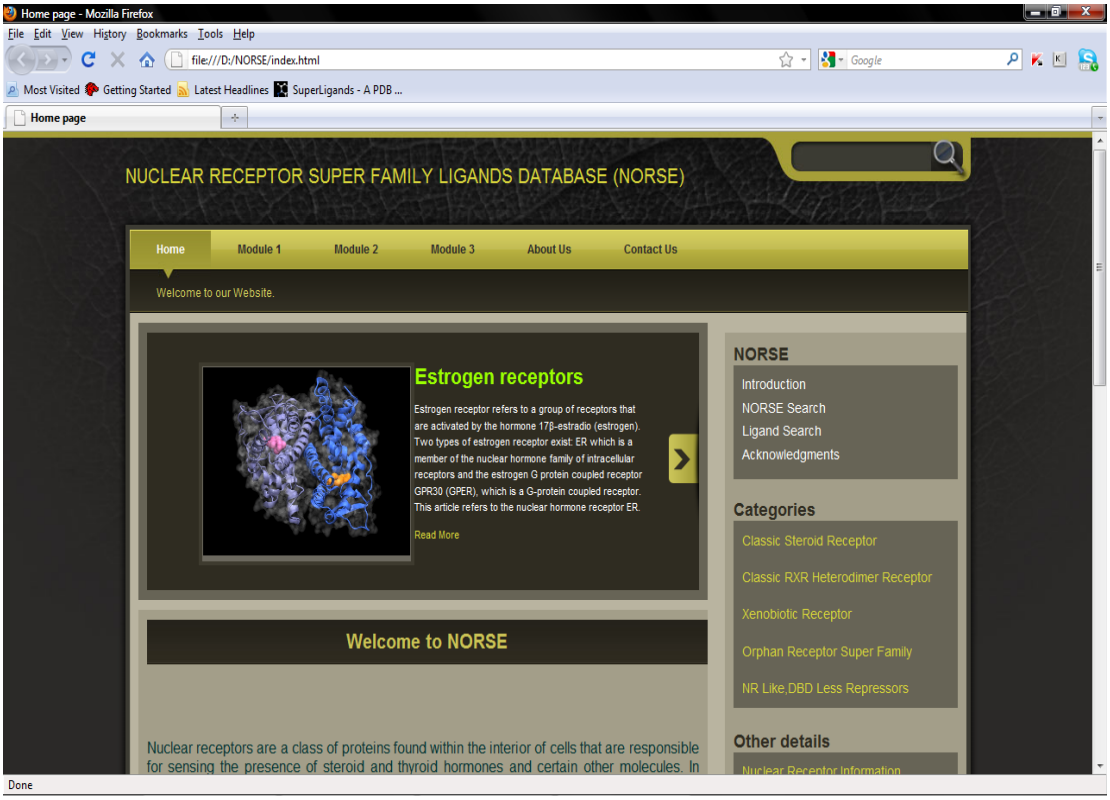
Hex is an interactive protein docking and molecular superposition program. Hex understands protein and DNA structures in PDB format, and it can also read small-molecule SDF files. It is mainly used to perform ligand-receptor docking studies to obtain energy values of the docking.



Similarly, another software called Quantum was used which is more specific. Basically what these do are study the 3D structure of the ligand and receptor and perform appropriate actions such as removal of water molecules and perform docking to get the best fit possible with the highest efficiency with the lowest energy input for the docking. Quantum selects upto ten regions on the compound and docs them at these locations to give you the best possible fit. Results are interpreted in terms of G bind energy values. All this information including the docking scores was saved in excel sheet and specific folders.



Database Creation: Adobe Dreamweaver (formerly Macromedia Dreamweaver) is a web development application originally created by Macromedia. Cascading Style Sheets (CSS) is a style sheet language used to describe the presentation semantics (that is, the look and formatting) of a document written in a markup language. It's most common application is to style web pages written in HTML and XHTML, but the language can also be applied to any kind of XML document, including SVG and XUL. CSS is designed primarily to enable the separation of document content (written in HTML or a similar markup language) from document presentation, including elements such as the layout, colors, and fonts. This separation can improve content accessibility, provide more flexibility and control in the specification of presentation characteristics, enable multiple pages to share formatting, and reduce complexity and repetition in the structural content (such as by allowing for table less web design). HTML, which stands for HyperText Markup Language, is the predominant markup language for web pages. It provides a means to create structured documents by denoting structural semantics for text such as headings, paragraphs, lists, links, quotes and other items. It allows images and objects to be embedded and can be used to create interactive forms. It is written in the form of HTML elements consisting of "tags" surrounded by angle brackets within the web page content. It can embed scripts in languages such as JavaScript which affect the behavior of HTML web pages. HTML can also be used to include Cascading Style Sheets (CSS) to define the appearance and layout of text and other material. The W3C, maintainer of both HTML and CSS standards, encourages the use of CSS over explicit presentational markup. Using HTML coding, CSS templates and Dream Weaver a relational database was created giving all the chemical information about the receptors, ligands, their interactions, other chemical information about the compounds and links to other websites.



RESULTS AND DISSCUSSION: A database comprising of the orphan nuclear receptors and their specific ligands were filtered by using two main docking programs viz., Hex and Quantum. Energy minimization was done using Marvin Sketch. All the information gathered through the three softwares, online databases and literature surveys were incorporated into the database that we created, NORSE. This is a relational database which has information leading to various databases for the user. It consists of receptor sequence information, in depth information about the ligands and other associated chemical properties. It is organized in a easy and receptor specific manner to make it simpler for the user to understand and use this information by downloading the necessary files et cetera from this database. Using the software Dream Weaver, codes can be changed to alter the database and its contents. This is done using HTML Coding. The result obtained was done by coding these tags into the program.

We found around two hundred ligands for each receptor through the various methods mentioned. Energy minimization was done on all these ligands (200* No. of receptors), Marvin sketch was used for this to calculate the base energy for all the ligands. These were filtered for interaction specificity using Hex. The filtered ligands were then put through Quantum which calculates interactions without water molecules so as to calculate the exact binding energy during the interaction. Quantum is more specific than Hex and is not available freely.

CONCLUSION: Nuclear receptors are a large superfamily of transcription factors involved in important physiological functions such as control of embryonic development, organ physiology, cell differentiation and homeostasis. Apart from the normal physiology, nuclear receptors have been identified to play a role in many pathological processes, such as cancer, diabetes, rheumatoid arthritis, asthma or hormone resistance syndromes. Therefore, despite their already long history, these transcriptional regulators are still of great interest in modern biomedical research and drug discovery. It is therefore important to create a database which can manage all this information regarding the Ligands and Receptors, into a package that can offer, say a scientist or a student, relevant information quickly and precisely. This is greatly achieved by creating a database of nuclear receptors and their Ligands, with information regarding docking scores, structures, formulae, et cetera so that correlations between them for further studies can be made, especially in the fields of Drug discovery, drug manufacture, Disease prevention and cure. Nuclear orphan receptor Databases will become the preferred method of storage for large multiuser applications, where coordination between many users is needed. Even individual users find them convenient, and many electronic mail programs and personal organizers are based on standard database technology.

REFERENCES:

1. Sluder AE, Maina CV, Nuclear receptors in nematodes: themes and variations. *Trends in Genetics,* (April 2001), *TIG* 17 (4): 206–13.
2. Pascual G, Glass CK Nuclear receptors versus inflammation: mechanisms of transrepression, 2006, *Trends Endocrinol Metab* 17 (8): 321–7.
3. Klinge CM, "Estrogen receptor interaction with co-activators and co-repressors, 2000, *Steroids* 65 (5): 227–51.
4. Benoit G, Cooney A, Giguere V, Ingraham H, Lazar M, Muscat G, Perlmann T, Renaud JP, Schwabe J, Sladek F, Tsai MJ, Laudet V, International Union of Pharmacology. LXVI. Orphan nuclear receptors, 2000, *Pharmacol. Rev.* 58 (4): 798–836.
5. Manly,C. *et al*, The impact of informatics and computational chemistry on synthesis and screening. *Drug Discov. Today*, 2000, 6, 1101–1110.
6. Butina,D. *et al*. Predicting ADME properties in silico: methods and models. *Drug Discov. Today*, 2002,7, S83–S88.
7. Byvatov,E. *et al*. Comparison of support vector machine and artificial neural network systems for drug/nondrug classification. *J. Chem. Inf. Comput. Sci.*, 2003, 43, 1882–1889.
8. Halperin, I. *et al*. Principles of docking: an overview of search algorithms and a guide to scoring functions. *Proteins*, 2002, 47, 409–443.
9. Laurent Jacob and Jean-Philippe Vert; Protein–ligand interaction prediction: an improved chemogenomics approach *Vol. 24 no. 19 2008, 2149–2156.*
10. Glass CK, Rosenfeld MG, The co-regulator exchange in transcriptional functions of nuclear receptors, 2008, *Genes Dev* 14 (2): 121–41.