ELECTIVE I

19CHP105-CMOLECULAR MODELLING & DRUG DESIGN4H4CInstruction Hours/week: L:4 T:0 P:0Marks: Internal:40 External: 60 Total:100

Course Objectives

- 1. The students should be acquainted with theoretical and practical knowledge of molecular modeling tools and techniques for drug design and discovery.
- 2. The knowledge gained molecular modeling software will be useful for commercial projects related to drug discovery and developments.
- 3. The detailed knowledge and skill is given in the course and the students get acquired the same after studying the course.

Course Outcome

- 1. The students gained the knowledge on the molecular modeling and field effects as a part of drug discovery.
- 2. Students have understood on the various stages and various targets of drug discovery.
- 3. Learned the importance of the pharmacophores in drug discovery.
- 4. They have learned the importance of the role of computer aided drug design in drug discovery.

UNIT I

Introduction to Molecular Modelling:

Introduction-Useful concepts in molecular modelling: Coordinate systems. Potential energy surfaces. Molecular graphics. Surfaces. Computer hardware and software. The molecular modelling literature.

UNIT II

Force Fields:

Fields. Bond stretching. Angle bending. Introduction to nonbonded interactions. Electrostatic interactions. Van der Waals Interactions. Hydrogen bonding in molecular mechanics. Force field models for the simulation of liquid water.

UNIT III

Basics of molecular modelling, methods, steps involved in MM, selection of target and template, homology modelling, refinement and validation-SAVES server, the critical assessment of protein structure prediction (CASP), superposition of proteins using different tools, RMSD, presentation of protein conformations, hydrophobicity factor, shape complementary.

UNIT IV

Pharmacophore

Historical perspective and viewpoint of pharmacophore, functional groups considered as pharmacophores, Ehrlich's "Magic Bullet", Fischer's "Lock and Key", two-dimensional pharmacophores, three-dimensional approach of pharmacophores, criteria for pharmacophore model, pharmacophore model generation software tools, molecular alignments, handling flexibility, alignment techniques, scoring and optimization, pharmacophores, validation and usage, automated pharmacophore generation methods, GRID-based pharmacophore models, pharmacophores for hit identification, pharmacophores for human ADME/tox-related proteins.

UNIT V

Computer aided Chemistry: Structure Prediction and Drug Design:

Introduction to molecular docking, rigid docking, Flexible docking, manual docking, advantage and disadvantage of flex-X, flex-S, AUTODOCK and other docking software, scoring functions, simple interaction energies, GB/SA scoring (implicit solvation), CScore (consensus scoring algorithms).

SUGGESTED READINGS: Text Books:

- 1. Leach, A. R. (2001). *Molecular Modelling Principles and Application* (II Edition). Longman: Prentice Hall.
- 2. Haile, J. M. (1997). *Molecular Dynamics Simulation Elementary Methods* Edition). UK: John Wiley and Sons.

(I

Reference Books:

1. Gupta, S. P. (2008). *QSAR and Molecular Modeling* (I Edition). Springer- Netherlands: Anamaya Publishers.



(Deemed to be University) (Established Under Section 3 of UGC Act, 1956) Coimbatore – 641 021.

LECTURE PLAN DEPARTMENT OF CHEMISTRY

STAFF NAME: Dr. S. SAVITHIRI SUBJECT NAME: MOLECULAR MODELLING & DRUG DESIGN

SUB.CODE:19CHP105-C

| SEMESTER: I | | CLASS: I-M.Sc (CHEMISTRY) | | |
|--|------------|---|---------------------------------|--|
| S.No. Lecture Topics to be Covered Duration Period | | Topics to be Covered | Support Material/Page Nos | |
| | | UNIT-I (Introduction to Molecular Modelling) | | |
| 1 | 1 | Introduction | T1:1 | |
| 2 | 1 | Useful concepts in molecular modelling | T1:1 | |
| 3 | 1 | Coordinate systems | T1:2 | |
| 4 | 1 | Potential energy Surfaces. | T1:4 | |
| 5 | 1 | Molecular graphics | T1:5 | |
| 6 | 1 | Surfaces | T1:6 | |
| 7 | 1 | Computer hardware and software | T1:8 | |
| 8 | 1 | The molecular modelling literature | T1:9 | |
| 9 | 1 | Recapitulations and discussion of important questions | | |
| | Total No o | of Hours Planned For Unit 1=9 | | |
| | | UNIT-II (Force Fields) | | |
| 1 | 1 | Introduction: Force fields, Fields | T1: 165-168 | |
| 2 | 1 | Bond stretching, Angle bending | T1:170- 174 | |
| 3 | 1 | Introduction to nonbonded interactions | T1: 181 | |
| 4 | 1 | Electrostatic interactions | T1:181-204 | |
| 5 | 1 | Van der Waals Interactions | T1: 204-212 | |

| 6 | 1 | Hydrogen bonding in molecular mechanics | T1: 215 |
|---|------------|--|----------------------------------|
| 7 | 1 | Force field models for the simulation of liquid water | T1: 216-220 |
| 8 | 1 | Revision & Discussion of possible questions | |
| | Total No o | of Hours Planned For Unit II=8 | |
| | | UNIT-III | |
| | | | |
| 1 | 1 | Basics of molecular modeling, methods | T2: 109-111 |
| 2 | 1 | steps involved in MM, selection of target and template | T2: 45,98,134,136, 146,164 |
| 3 | 1 | homology modelling | T2:92, 96 |
| 4 | 1 | refinement and validation-SAVES server | T2:115,118,121 |
| 5 | 1 | the critical assessment of protein structure prediction (CASP) | T2:81,116 |
| 6 | 1 | superposition of proteins using different tools | T2:55,56 |
| 7 | 1 | RMSD, presentation of protein conformations | T2:81,82,113, 132 |
| 8 | 1 | hydrophobicity factor, shape complementary | T2:51,135,142 |
| 9 | 1 | Revision & Discussion of important questions | |
| | Total No o | of Hours Planned For Unit III=9 | |
| | | UNIT-IV (Pharmacophore) | |
| 1 | 1 | Historical perspective and viewpoint of pharmacophore | T3:572-584 |
| 2 | 1 | functional groups considered as pharmacophores | T3: 464-479 |
| 3 | 1 | Ehrlich's "Magic Bullet", Fischer's "Lock and Key" | T3: 519 |
| 4 | 1 | two-dimensional pharmacophores, three-dimensional approach of pharmacophores | T3: 573-575 |
| 5 | 1 | criteria for pharmacophore model, pharmacophore model generation software tools | T3:575-577 |
| 6 | 1 | molecular alignments, handling flexibility, alignment techniques, scoring and optimization | T3:592-594 |
| 7 | 1 | pharmacophores, validation and usage, automated pharmacophore generation methods. | T3:524-31 |
| 8 | 1 | GRID-based pharmacophore models, pharmacophores for hit identification | T3:259, 599 |

2019-2021

Batch

| 9 | 1 | pharmacophores for human ADME/tox-related proteins | T3:507 |
|---------|----------|--|------------|
| 10 | 1 | Revision & Discussion of important questions | |
| | Total No | of Hours Planned For Unit IV=10 | |
| | | UNIT-V (Computer aided Chemistry: | |
| | | Structure Prediction and Drug Design) | |
| 1 | 1 | Introduction to molecular docking, rigid docking | W1: |
| 2 | 1 | Flexible docking, manual docking, | W1: |
| 3 | | advantage and disadvantage of flex-X | W1: |
| 4 | 1 | flex-S, AUTODOCK | W1: |
| 5 | 1 | and other docking software | W1: |
| 6 | 1 | scoring functions, simple interaction energies | T2: 93 |
| 7 | 1 | GB/SA scoring (implicit solvation) | T2: 96,124 |
| 8 | 1 | CScore (consensus scoring algorithms). | T2: 149 |
| 9 | 1 | Revision & Discussion of important questions | |
| 10 | 1 | Discussion of important questions | |
| 11 | 1 | Discussion of important questions | |
| 12 | 1 | Discussion of important questions | |
| |] | Fotal No of Hours Planned for Unit V=12 | |
| Total | 48 | | |
| Planned | | | |
| Lours | | | |
| I HOUIS | 1 | | |

Text Book:

1. Leach, A. R. (2001). *Molecular Modelling Principles and Application* (II Edition). Longman: Prentice Hall.

2. Hans-Diels-Holtje and Gerd Folkers, copyright 1997: Molecular modeling. Basic principles and Applications. ISBN Latest edition 2008.

3. Camille George Wermuth (2008); The Practice of Medicinal Chemistry-III Edition Academic Press. W1: <u>https://sites.google.com/site/home/types</u> of docking



Class: I- M.Sc (Chemistry)Course Name: Molecular Modelling & Drug DesignCourse Code: 19CHP105-CUnit: I (Introduction to molecular modelling) Batch: 2019 -2021



Class: I- M.Sc (Chemistry)Course Name: Molecular Modelling & Drug DesignCourse Code: 19CHP105-CUnit: I (Introduction to molecular modelling) Batch: 2019 -2021

Prepared by Dr. S.SAVITHIRI, Asst. Prof., Department of Chemistry, KAHE Page 2/11



Class: I- M.Sc (Chemistry)Course Name: Molecular Modelling & Drug DesignCourse Code: 19CHP105-CUnit: I (Introduction to molecular modelling) Batch: 2019 -2021

UNIT I

Introduction to Molecular Modelling:

Introduction-Useful concepts in molecular modelling: Coordinate systems. Potential energy surfaces. Molecular graphics. Surfaces. Computer hardware and software. The molecular modelling literature.





Class: I- M.Sc (Chemistry)Course Name: Molecular Modelling & Drug DesignCourse Code: 19CHP105-CUnit: I (Introduction to molecular modelling) Batch: 2019 -2021

Introduction

Useful concepts in molecular modelling

What is molecular modelling? 'Molecular' clearly implies some connection with molecules. The *Oxford English Dictionary* defines 'model' as 'a simplified or idealised description of a system or process, often in mathematical terms, devised to facilitate calculations and predictions'. Molecular modelling would therefore appear to be concerned with ways to mimic the behaviour of molecules and molecular systems. Today, molecular modelling is invariably associated with computer modelling, but it is quite feasible to perform some simple molecular modelling studies using mechanical models or a pencil, paper and hand calculator. Nevertheless, computational techniques have revolutionised molecular modelling to the extent that most calculations could not be performed without the use of a computer. This is not to imply that a more sophisticated model is necessarily any better than a simple one, but computers have certainly extended the range of models that can be considered and the systems to which they can be applied.

The 'models' that most chemists first encounter are molecular models such as the 'stick' models devised by Dreiding or the 'space filling' models of Corey, Pauling and Koltun (commonly referred to as CPK models). These models enable three-dimensional representations of the structures of molecules to be constructed. An important advantage of these models is that they are interactive, enabling the user to pose 'what if ...' or 'is it possible to ...' questions. These structural models continue to play an important role both in teaching and in research, but molecular modelling is also concerned with more abstract models, many of which have a distinguished history. An obvious example is quantum mechanics, the foundations of which were laid many years before the first computers were constructed.

There is a lot of confusion over the meaning of the terms 'theoretical chemistry', 'computational chemistry' and 'molecular modelling'. Indeed, many practitioners use all three labels to describe aspects of their research, as the occasion demands! 'Theoretical chemistry' is often considered synonymous with quantum mechanics, whereas computational chemistry encompasses not only quantum mechanics but also molecular mechanics, minimisation, simulations, conformational analysis and other computer-based methods for understanding and predicting the behaviour of molecular systems. Molecular modellers use all of these methods and so we shall not concern ourselves with semantics but rather shall consider any theoretical or computational technique that provides insight into the behaviour of molecular systems to be an example of molecular modelling. If a distinction has to be made, it is in the emphasis that molecular modelling places on the representation and manipulation of the structures of molecules, and properties that are dependent upon those three-dimensional structures. The prominent part that computer graphics has played in molecular modelling has led some scientists to consider molecular modelling as little more than a method for generating 'pretty pictures', but the technique is now firmly established, widely used and accepted as a discipline in its own right.





Class: I- M.Sc (Chemistry) Course Code: 19CHP105-C Unit: I (Introduction to molecular modelling) Batch: 2019 -2021

Coordinate Systems

It is obviously important to be able to specify the positions of the atoms and/or molecules in the system to a modelling program^{*}. There are two common ways in which this can be done. The most straightforward approach is to specify the Cartesian (x, y, z) coordinates of all the atoms present. The alternative is to use *internal coordinates*, in which the position of each atom is described relative to other atoms in the system. Internal coordinates are usually written as a Z-matrix. The Z-matrix contains one line for each atom in the system. A sample Z-matrix for the staggered conformation of ethane (see Figure 1.1) is

as follows:

| 1 | С | | | | | | |
|---|---|------|---|-------|---|-------|---|
| 2 | С | 1.54 | 1 | | | | |
| 3 | Η | 1.0 | 1 | 109.5 | 2 | | |
| 4 | Η | 1.0 | 2 | 109.5 | 1 | 180.0 | 3 |
| 5 | Η | 1.0 | 1 | 109.5 | 2 | 60.0 | 4 |
| 6 | Η | 1.0 | 2 | 109.5 | 1 | -60.0 | 5 |
| 7 | Η | 1.0 | 1 | 109.5 | 2 | 180.0 | 6 |
| 8 | Н | 1.0 | 2 | 109.5 | 1 | 60.0 | 7 |

*For a system containing a large number of independent molecules it is common to use the term 'configuration' to refer to each arrangement; this use of the word 'configuration' is not to be confused with its standard chemical meaning as a different bonding arrangement of the atoms in a molecule



Fig. 11 The staggered conformation of ethane

Entry Instance I for Market I program I prog

KARPAGAM ACADEMY OF HIGHER EDUCATION

Class: I- M.Sc (Chemistry)Course Name: Molecular Modelling & Drug DesignCourse Code: 19CHP105-CUnit: I (Introduction to molecular modelling) Batch: 2019 -2021

In the first line of the Z-matrix we define atom 1, which is a carbon atom. Atom number 2 is also a carbon atom that is a distance of 1.54 Å from atom 1 (columns 3 and 4). Atom 3 is a hydrogen atom that is bonded to atom 1 with a bond length of 1.0 Å. The angle formed by atoms 2–1–3 is 109.5°, information that is specified in columns 5 and 6. The fourth atom is a hydrogen, a distance of 1.0 Å from atom 2, the angle 4–2–1 is 109.5°, and the torsion angle (defined in Figure 1.2) for atoms 4–2–1–3 is 180°. Thus for all except the first three atoms, each atom has three internal coordinates: the distance of the atom from one of the atoms previously defined, the angle formed by the atom and two of the previous atoms, and the torsion angle defined by the atom and three of the previous atoms. Fewer internal coordinates are required for the first three atoms because the first atom can be placed anywhere in space (and so it has no internal coordinates); for the second atom it is only necessary to specify its distance from the first atom and then for the third atom only a distance and an angle are required.

It is always possible to convert internal to Cartesian coordinates and vice versa. However, one coordinate system is usually preferred for a given application. Internal coordinates can usefully describe the relationship between the atoms in a single molecule, but Cartesian coordinates may be more appropriate when describing a collection of discrete molecules. Internal coordinates are commonly used as input to quantum mechanics programs, whereas calculations using molecular mechanics are usually done in Cartesian coordinates. The total number of coordinates that must be specified in the internal coordinate system is six fewer than the number of Cartesian coordinates for a non-linear molecule. This is because we are at liberty to arbitrarily translate and rotate the system within Cartesian space without changing

the relative positions of the atoms



Fig. 1.2 A torsion angle A-B-C-D is defined as the angle between the planes A, B, C and B, C, D A torsion angle can vary through 360° although the range -180° to $+180^{\circ}$ is most commonly used We shall adopt the IUPAC definition of a torsion angle in which an eclipsed conformation corresponds to a torsion angle of 0° and a trans or anti conformation to a torsion angle of 180° . The reader should note that this may not correspond to some of the definitions used in the literature, where the trans arrangement is defined as a torsion angle of 0° . If one looks along the bond B-C, then the torsion angle is the angle through which it is necessary to rotate the bond AB in a clockwise sense in order to superimpose the two planes, as shown

Potential Energy Surfaces



Class: I- M.Sc (Chemistry) Course Code: 19CHP105-C Unit: I (Introduction to molecular modelling) Batch: 2019 -2021

In molecular modelling the Born-Oppenheimer approximation is invariably assumed to operate. This enables the electronic and nuclear motions to be separated; the much smaller mass of the electrons means that they can rapidly adjust to any change in the nuclear positions. Consequently, the energy of a molecule in its ground electronic state can be considered a function of the nuclear coordinates only. If some or all of the nuclei move then the energy will usually change. The new nuclear positions could be the result of a simple process such as a single bond rotation or it could arise from the concerted movement of a large number of atoms. The magnitude of the accompanying rise or fall in the energy will depend upon the type of change involved. For example, about 3 kcal/molic .equired to change the covalent carbon-carbon bond length in ethane by 0.1 Å away from its equilibrium value, but only about 0.1 kcal/mol is required to increase the non-c ont separation between two argon atoms by 1 Å from their minimum energy separation. For mall isc and molecules, rotation about single bonds usually involves the smallest anges in e. . . For example, if we rotate the carbon-carbon bond in ethane, keeping all of ... Yond leng, ynd angles find in value, then the energy varies in an approximate'y sinusoio. ashion as wn in Figure 1.3, with in this ca. <u>an</u> e considered a minima at the three staggered conform ons. The ene. function of a single coordinate only (i.e. e torsior ane of the ca _n-carbon bond), e axis and the value of the and as such can be displayed graphically, w er .gy alons coordinate along the other.



Fig. 1.3 Variation energy with rotation of the carbon-carbon bond in ethane

Changes in the energy of a system can be considered as movements on a multidimensional 'surface' called the *energy surface*. We shall be particularly interested in stationary points on the energy surface, where the first derivative of the energy is zero with respect to the internal or Cartesian coordinates. At a stationary point the forces on all the atoms are zero. Minimum points are one type of stationary point; these correspond to stable structures. Methods for locating stationary points will be discussed in more detail in Chapter 5, together with a more detailed consideration of the concept of the energy surface.





Class: I- M.Sc (Chemistry)Course Name: Molecular Modelling & Drug DesignCourse Code: 19CHP105-CUnit: I (Introduction to molecular modelling) Batch: 2019 -2021

Molecular Graphics

Computer graphics has had a dramatic impact upon molecular modelling. It should always be remembered, however, that there is much more to molecular modelling than computer graphics. It is the interaction between molecular graphics and the underlying theoretical methods that has enhanced the accessibility of molecular modelling methods and assisted the analysis and interpretation of such calculations.

Molecular graphics systems have evolved from delicate and temperamental pieces of equipment that cost hundreds of thousands of pounds and occupied entire rooms, to today's inexpensive workstations that fit on or under a desk and yet are hundreds of times more powerful Over the years, two different types of molecular graphics display have been used in molecular modelling. First to be developed were vector devices, which construct pictures using an electron gun to draw lines (or dots) on the screen, in a manner similar to an oscilloscope. Vector devices were the mainstay of molecular modelling for almost two decades but have now been largely superseded by raster devices. These divide the screen into a large number of small 'dots', called pixels. Each pixel can be set to any of a large number of colours, and so by setting each pixel to the appropriate colour it is possible to generate the desired image.

Molecules are most commonly represented on a computer graphics screen using 'stick' or 'space-filling' representations, which are analogous to the Dreiding and Corey-Pauling-Koltun (CPK) mechanical models. Sophisticated variations on these two basic types have been developed, such as the ability to colour molecules by atomic number and the inclusion of shading and lighting effects, which give 'solid' models a more realistic appearance. Some of the commonly used molecular representations are shown in Figure 1.4 (colour plate section). Computer-generated models do have some advantages when compared with their mechanical counterparts. Of particular importance is the fact that a computer model can be very easily interrogated to provide quantitative information, from simple geometrical measures such as the distance between two atoms to more complex quantities such as the energy or surface area. Quantitative information such as this can be very difficult if not impossible to obtain from a mechanical model. Nevertheless, mechanical models may still be preferred in certain types of situation due to the ease with which they can be manipulated and viewed in three dimensions. A computer screen is inherently two-dimensional, whereas molecules are three-dimensional objects. Nevertheless, some impression of the three-dimensional nature of an object can be represented on a computer screen using techniques such as depth cueing (in which those parts of the object that are further away from the viewer are made less bright) and through the use of perspective. Specialised hardware enables more realistic three-dimensional stereo images to be viewed. In the future 'virtual reality' systems may enable a scientist to interact with a computer-generated molecular model in much the same way that a mechanical model can be manipulated.



Class: I- M.Sc (Chemistry)Course Name: Molecular Modelling & Drug DesignCourse Code: 19CHP105-CUnit: I (Introduction to molecular modelling) Batch: 2019 -2021

Even the most basic computer graphics program provides some standard facilities for the manipulation of models, including the ability to translate, rotate and 'zoom' the model towards and away from the viewer. More sophisticated packages can provide the scientist with quantitative feedback on the effect of altering the structure. For example, as a bond is rotated then the energy of each structure could be calculated and displayed interactively. For large molecular systems it may not always be desirable to include every single atom in the computer image; the sheer number of atoms can result in a very confusing and cluttered picture. A clearer picture may be achieved by omitting certain atoms (e.g. hydrogen atoms) or by representing groups of atoms as single 'pseudo-atoms' The techniques that have been developed for displaying protein structures nicely illustra', the range of computer graphics representation possible (the use of computational techniques to investigate the structures of proteins is considered in Chapter 10). Proteins are powers constructed from amino acids, and even a small protein may contain several thousand . ns. On ... v to produce a clearer picture is to dispense with the explicit represent tation of the come and the represent the protein using a 'ribbon'. Proteins are also comonly re, sented usin the cartoon drawings developed by J Richardson, ar example which is so wn in Figure 1.5 (colour plate section). The cylinders in this fig a represent a grangeme of a uno acids called an α -helix, and the flat arrows an alter. 've type α 'ar struct α called a β -strand. The regions between the cylinders and the ond nave no ch regalar structure and are represented as 'tubes'.

Surfaces

Many of the prob' ms that are studied using a subr modelling involve the non-covalent interaction between two or more more ales. The subry of such interactions is often facilitated



Fig 1.6. The van der Waals (vdw) surface of a molecule corresponds to the outward-facing surfaces of the van der Waals spheres of the atoms The molecular surface is generated by rolling a spherical probe (usually of radius 1.4 Å to represent a water molecule) on the van der Waals surface The molecular surface is constructed from contact and re-entrant surface elements The centre of the probe traces out the accessible surface



Class: I- M.Sc (Chemistry)Course Name: Molecular Modelling & Drug DesignCourse Code: 19CHP105-CUnit: I (Introduction to molecular modelling) Batch: 2019 -2021

by examining the van der Waals, molecular or accessible surfaces of the molecule. The van der Waals surface is simply constructed from the overlapping van der Waals spheres of the atoms, Figure 1.6. It corresponds to a CPK or space-filling model. Let us now consider the approach of a small 'probe' molecule, represented as a single van der Waals sphere, up to the van der Waals surface of a larger molecule. The finite size of the probe sphere means that there will be regions of 'dead space', crevices that are not accessible to the probe as it rolls about on the larger molecule. This is illustrated in Figure 1.6. The amount of dead space increases with the size of the probe; conversely, a probe of zero size would be able to access all of the crevices. The molecular surface [Richards 1977] is traced out http://www.crevices. The molecular surface [Richards 1977] is traced out http://www.crevices probe sphere as it rolls on the van der Waals surface of the molecule. The molecular surface contains two different types of surface element. The *ontact surface* corresponds to those regions where the probe is actually in contact with the der Waar surface of the 'target'. The re-entrant surface regions occur where ther are creases the are too narrow for the probe molecule to penetrate. The molecular sur is usually ned using a vater molecule as the probe, represented as a sphere of radius 1.4.

The accessible surface is also widely use <u>roriginally c</u>ed by Lee <u>d</u> achards [Lee and Richards 1971] this is the surface that is training the problem of t

Widely used algorithms for call dating in molecular and accessible surfaces were developed by Connolly [Connolly 19c3a, b] incluivers [e.g. Richmond 1984] have described formulae for the includation of exactor approximate values of the surface area. There are many ways to represent surfaces, so he of which are illustrated in Figure 1.7 (colour plate section). Shown it may also be polisible to endow a surface with a translucent quality, which enable the inside the inside to be displayed. Clipping can also be used to cut through the indecule inside the inside' to be viewed. In addition, properties such as the electrostatic point of the indecule formally extends to infinity. The 'hard sphere' representation is often velocitient and has certainly proved very valuable, but it may not be appropriate in all cases [Rouvray 1997, 1999, 2000].

Computer Hardware and Software



Class: I- M.Sc (Chemistry) Course Code: 19CHP105-C Unit: I (Introduction to molecular modelling) Batch: 2019 -2021

One cannot fail to be amazed at the pace of development in the computer industry, where the ratio of performance-to-price has increased by an order of magnitude every five years or so. The workstations that are commonplace in many laboratories now offer a real alternative to centrally maintained 'supercomputers' for molecular modelling calculations, especially as a workstation or even a personal computer can be dedicated to a single task, whereas the supercomputer has to be shared with many other users. Nevertheless, in the immediate future there will always be some calculations that require the power that only a supercomputer can offer. The speed of any computer system is ultimately constrained by the speed at which electrical signals can be transmitted. This means the 'there will come a time when no further enhancements can be made using machines' ith 'traditional' single-processor serial architectures, and parallel computers will play e^- ever more important role.

A parallel computer couples processors together in such vay the uncalculation is divided into small pieces with the results being combined at the end of one calculations are more amenable to parallel processing than others, and ensights a mount of fort is being spent converting existing algorithms to run efficiently on para. Tarchiter areas. In some cases completely new methods have for a developed take max. The advantage of the opportunities of parallel processing. The two cost of permitted computing means that large 'farms' of processors can be constructed to give significant computing power for relatively small outlay.

To perform molecular modell. L'culations one a requires appropriate programs (the software). The software used by me. Far modellers finges from simple programs that perform just a single task to highly complete the ages that integrate many different methods. There is also an fremely wide variation in the programs for software! Some programs have been so widely used an frested that they can be considered to have reached the status of a 'gold standard' gainst which similar programs are compared. One hesitates to specify such programs a forded the accolade These are the Gaussian series of programs for performing *ab initio* for the MOPAC/AMPAC programs for semi-empirical quantum mechanic. For the MM program for molecular mechanics.

Various pieces of softwa. Yere used to generate the data for the examples and illustrations throughout this book. Som of these were written specifically for the task; some were freely available programs; others were commercial packages. I have decided not to describe specific programs in any detail, as such descriptions rapidly become outdated. Nevertheless,

all items of software are accredited where appropriate. Please note that the use of any particular piece of software does not imply any recommendation!

Units of Length and Energy



Class: I- M.Sc (Chemistry) Course Code: 19CHP105-C Unit: I (Introduction to molecular modelling) Batch: 2019 -2021

It will be noted that our Z-matrix for ethane has been defined using the angstrom as the unit of length $(1 \text{ \AA} \equiv 10^{-10} \text{ m} \equiv 100 \text{ pm})$. The ångström is a non-SI unit but is a very convenient one to use, as most bond lengths are of the order of 1–2 Å. One other very common non-SI unit found in the molecular modelling literature is the kilocalorie (1 kcal \equiv 4.1840 kJ). Other systems of units are employed in other types of calculation, such as the atomic units used in quantum mechanics (discussed in Chapter 2). It is important to be aware of, and familiar with, these non-standard units as they are widely used in the literature and throughout this book.

The Molecular Modelling Literature

The number of scientific papers concerned with molecular modelling methods is rising rapidly, as is the number of journals in which such papers are published. This reflects the tremendous diversity of problems to which molecular modelling can be applied and the ever-increasing availability of molecular modelling methods. It does, however, mean that it can be very difficult to remain up to date with the field. A number of specialist journals are devoted to theoretical chemistry, computational chemistry and molecular modelling, each with their own particular emphasis. Relevant papers are also published in the more 'general' journals, and there are now a number of books covering aspects of molecular modelling, some aimed at the specialist reader, others at the beginner. Many scientists are now fortunate to have access to electronic catalogues of publications which can be searched to find relevant papers. As many journals are now available over the internet it is possible to perform a literature search and obtain copies of the relevant papers without even having to leave the office. Some of the journals which are devoted to short reviews of recent developments often include molecular modelling sections (such as the 'Current Opinion' series); in others, useful review articles appear on an occasional basis. One particularly valuable source of information on molecular modelling methods is the Reviews in Computational Chemistry, edited by Lipkowitz and Boyd, beginning in 1990 (see Further Reading), Each of these volumes contains chapters on a variety of subjects, each written by an appropriate expert A recent addition is the Encyclopaedia of Computational Chemistry by Schleyer et al. (1998) (see Further Reading), which contains many chapters that cover a wide range of topics.



Enable | Enlighten | Enrich (Deemed to be University) (Under Section 3 of UGC Act 1956)

KARPAGAM ACADEMY OF HIGHER EDUCATION (Deemed to be University Under Section 3 of UGC Act 1956) COIMBATORE-21

DEPARTMENT OF CHEMISTRY (For the candidates admitted from 2019 & onwards) 19CHP105-C MOLECULAR MODELLING & DRUG DESIGN Multiple Choice Questions for Unit I

| S. No | Question | Option 1 | Option 2 | Option 3 | Option 4 | Answer |
|-------|---|--------------------------------|----------------------------|----------------------------------|----------------------------------|----------------------------------|
| | Unit-I | | | | | |
| 1 | Which is considered synonymous with quantum mechanics | Theoretical chemistry | Computational chemistry | Molecular mechanics | Molecular dynamics | Theoretical chemistry |
| 2 | The torsion angle for the staggered conformation of ethane is | 90° | 180° | 270 | 360° | 180° |
| 3 | Change in energy of the system can be considered as a movement on a multi dimensional surface is called | Internal energy | Minimum energy | . Energy surface | . Kinetic energy | Energy surface |
| 4 | The software used by molecular modeling ranges from | Single to highly complex | Single to medium | Medium to highly complex | Medium to highly | Single to highly complex |
| 5 | Some programs in molecular modeling is widely used and tested so they are considered as | Gold standard | Silver standard | International standard | Molecular standard | Gold standard |
| 6 | Molecular modelling implies behaviour of molecules and | Atoms | Molecular system | Ions | Rotations | Molecular system |
| 7 | In molecular modelling CPK model is referred as | Cartesian Pole Kinetics | Core Push Key | Corey, Pauling and Kolthun | Critical Pressure Kinetics | Corey, Pauling and Kolthun |
| 8 | CPK model is also called | Remodelling | Space creating model | Rotating model | Space filling model | Space filling model |
| 9 | The number of torsional terms in benzene is | 6 | 18 | 24 | 30 | 24 |

| | | | | Morse | | |
|----|---|--|----------------------|-----------------------|------------------------|--|
| 10 | | Coulombs | F 1 1 1 | potential | TT 1 1 | Coulombs |
| 10 | The electrostatic interaction is calculated by using | law | Faraday's law | curve | Hookes law | law |
| 11 | The Contact surface is the region where the probe is in actually contact with the | Vander waal's surface of the target | Molecular surface | Water molecule | Accessible surface | Vander waal's surface of the target |
| 12 | Encyclopedia for computational chemistry by Schleyer et al was released in the year | 1990 | 1998 | 2002 | 2010 | 1998 |
| 13 | Change in energy of the system can be considered as a movement on a multi dimensional surface is called | Internal energy | Minimum energy | Energy surface | Kinetic energy | Energy surface |
| 14 | Computation Chemistry involves | Not only Quantum mechanics | Quantum mechanics | Theoretical chemistry | Computer graphics | Not only Quantum mechanics |
| 15 | The torsion angle for the staggered conformation of ethane is | 90° | 180° | 270° | 360° | 180° |
| 16 | In Z- matrix, for each atom it contains | Two lines | Four lines | One line | Three lines | One line |
| 17 | In which dimensional surface, there is a movement of change in energy of the system | Uni- dimensional | Bi- dimensional | Tri- dimensional | Multi- dimensional | Multi- dimensional |
| 18 | In molecular modeling, the first device used is | Vector | Scalar | Arithmetic | Logical | Vector |
| 19 | In vector devices, the pictures were constructed by using | Ball and stick | Pen | Electron gun | Scanning | Electron gun |
| 20 | In vector device now been largely superseded by | Scalar device | Raster device | Matrix device | Arithmetic calculation | Raster device |
| 21 | In raster device, the screen is divided into a large number of small dots calls | Pixels | Points | Arrays | Lines | Pixels |
| 22 | In raster device, each pixel can be set to any of a large number of | Points | Squares | Curves | Colours | Colours |
| 23 | The inclusion of which effect will give a solid model a | Shading | Lighting | Shading effect | Numbering | Shading and |

| | more realistic appearance | and lighting | effect | | effect | lighting |
|----|--|--------------------|----------------------|----------------------|---------------|----------------------------------|
| | | effect | | ~ | | effect |
| 24 | Which models have some more advantages when compared | Ball and | Computer- | Graphical | Space filling | Computer- |
| 25 | with mechanical counter parts? | Sticky | generated | T | Thurso | generated |
| 25 | A computer screen is inherently | One dimensional | Multi | 1W0 dimensional | limonoionol | 1W0 dimensional |
| 26 | | Quinensional | | | | Theres |
| 26 | Molecules are | One | Multi dimensional | 1W0 dimensional | Inree | Inree |
| 27 | Three dimensional nature of an abject can be remained | | dimensional | Commuter | Cusulting | aimensional Dan the area in a |
| 21 | on a computer screen, by using the technique called | cueing | Laser | image | Graphical | Depth cueing |
| 28 | Which enables more realistic three dimensional stereo | Specialized | Software tools | Specialized | Molecular | Specialized |
| | images to be viewed | Software | | hard ware | Graphics | hard ware |
| 29 | Which system in future may enable a scientist to interact | Virtual | Molecular | Space-filling | Molecular | Virtual |
| | with computer generated model in the same way as mechanical model? | reality | Graphics | 1 0 | Modelling | reality |
| 30 | In depth cueing technique, the parts of the object far away from the viewer are made | More bright | Less bright | Bright | Dark | Less bright |
| 31 | In which technique, the distance parts of the object are made less bright. | Depth cueing | Laser | Computer image | Graphical | Depth cueing |
| 32 | In molecular modeling, clear picture may be achieved by omitting which atom | Hydrogen | Carbon | Nitrogen | Oxygen | Hydrogen |
| 33 | Proteins are polymers constructed from | Amides | Acids | Amino acids | Esters | Amino acids |
| 34 | The number of atoms present in small proteins is in the order of | Lesser | Several hundreds | Several thousands | Fewer | Several thousands |
| 35 | The proteins are represented by using | Balls | Ribbons | Sticks | Circles | Ribbons |
| 36 | Proteins are commonly represented by cartoon drawing developed by | Richardson | Corey | Pauling | Kolthun | Richardson |
| 37 | The arrangement of amino acid in protein for α -helix is represented by | Tubes | Flat arrows | Squares | Cylinders | Cylinders |
| 38 | The arrangement of amino acid in protein for β -strand is represented by | Tubes | Flat arrows | Squares | Cylinders | Flat arrows |
| 39 | In the arrangement of amino acids in protein, the region between α -helix and β -strand is represented by | Tubes | Flat arrows | Squares | Cylinders | Tubes |
| 40 | The interaction between two or more molecules in | Ionic | Covalent | Non- | Dipole- | Non- |
| | molecular modelling is | | | Covalent | Dipole | Covalent |
| 41 | Which is constructed by overlapping Van der waal sphere | CPK model | Van der waal | Space filling | Molecular | Van der waal |
| | of atoms? | | surface | model | surface | surface |
| 42 | Which program is used for Molecular mechanics? | MM2 | MOPAC | AMPAC | MMH2 | MM2 |
| 43 | For eclipsed conformation, torsion angle is | 270° | 180° | 90° | 0° | 0° |
| 44 | For Anti conformation, torsion angle is | 270° | 180° | 90° | 0° | 180° |

| 45 | Z-Matrix of ethane has been defined using the angstrom as the unit of length, one angstrom is equivalent to | 60 pm | 80 pm | 100 pm | 120 pm | 100 pm |
|----|--|--------------------|-----------------------|---------------------|---------------------------------|-----------------------|
| 46 | One kilocalorie is equivalent to | 2.092kJ | 8.368 kJ | 6.276 kJ | 4.184 kJ | 4.184 kJ |
| 47 | Proteins are | Monomers | Dimers | Trimers | Polymers | Polymers |
| 48 | Proteins are polymers constructed from | Aminoacids | Aminoketones | Aminoesters | Alkylamides | Aminoacids |
| 49 | The number of atoms present in small protein is | 10 | 100 | 1000 | Several thousand | Several thousand |
| 50 | In molecular modeling proteins are represented by using which model? | Dots | Ribbon | Points | Lines | Ribbon |
| 51 | Which device were the main stay of molecular modeling for almost two decades? | Vector | Scalar | Electronic | Laser | Vector |
| 52 | Vector devices of molecular modeling have now been largely superseded by | Scalar devices | Laser devices | Raster devices | Electronic devices | Raster devices |
| 53 | Van der waals surface corresponds to | CPK model | Ball and stick model | Conceptual model | Computer simulation model | CPK model |
| 54 | Some programs have been so widely used and reached the status of | Silver standard | Gold standard | Bronze standard | ISO standard | Gold standard |
| 55 | The speed of any computer system is ultimately constrained by the speed of transmission of | Inputs | Datas | Electric signals | Arrays | Electric signals |
| 56 | Reviews in computational chemistry is edited by | Rouvray | Lipkowitz and Boyd | Connolly | Richmond | Lipkowitz and Boyd |
| 57 | Encyclopaedia of computational chemistry is edited by | Schleyer et al | Connolly | Richmond | Rouvray | Schleyer et al |
| 58 | Reviews in computational chemistry is edited in the year | 1998 | 1996 | 1994 | 1990 | 1990 |
| 59 | Encyclopedia of computational chemistry is edited in the year | 1998 | 1996 | 1994 | 1990 | 1998 |
| 60 | Wide range of topics about computational chemistry is present in | Review | Encyclopedia | Journal | Article | Encyclopedia |



Class: I- M.Sc (Chemistry) Course Code: 19CHP105-C Course Name: Molecular Modellingand Drug DesignUnit: II (Force Fields)Batch: 2019 - 2021

UNIT II

Force Fields:

Fields. Bond stretching. Angle bending. Introduction to nonbonded interactions. Electrostatic interactions. Van der Waals Interactions. Hydrogen bonding in molecular mechanics. Force field models for the simulation of liquid water.



Class: I- M.Sc (Chemistry) Course Name: Molecular Modelling and Drug Design Course Code: 19CHP105-C Unit: II (Force Fields) **Batch**: 2019 -2021



Class: I- M.Sc (Chemistry)Course Name: Molecular Modelling and Drug DesignCourse Code: 19CHP105-CUnit: II (Force Fields)Batch: 2019 - 2021

Force Fields: Introduction

Many of the problems that we would like to tackle in molecular modelling are unfortunately too large to be considered by quantum mechanics. Quantum mechanical methods deal with the electrons in a system, so that even if some of the electrons are ignored (as in the semiempirical schemes) a large number of particles must still be considered, and the calculations are time-consuming. Force field methods (also known as molecular mechanics) ignore the electronic motions and calculate the energy of a system as a function of the nuclear positions only. Molecular mechanics is thus invariably used to perform calculations on systems containing significant numbers of atoms. In some cases force fields can provide answers that are as accurate as even the highest-level quantum mechanical calculations, in a fraction of the computer time. However, molecular mechanics cannot of course provide properties that depend upon the electronic distribution in a molecule.

That molecular mechanics works at all is due to the validity of several assumptions. The first of these is the Born–Oppenheimer approximation, without which it would be impossible to contemplate writing the energy as a function of the nuclear coordinates at all. Molecular mechanics is based upon a rather simple model of the interactions within a system with contributions from processes such as the stretching of bonds, the opening and closing of angles and the rotations about single bonds. Even when simple functions (e.g. Hooke's law) are used to describe these contributions the force field can perform quite acceptably. Transferability is a key attribute of a force field, for it enables a set of parameters developed and tested on a relatively small number of cases to be applied to a much wider range of problems. Moreover, parameters developed from data on small molecules can be used to study much larger molecules such as polymers.

A Simple Molecular Mechanics Force field

Many of the molecular modelling force fields in use today for molecular systems can be interpreted in terms of a relatively simple four-component picture of the intra- and intermolecular forces within the system. Energetic penalties are associated with the deviation of bonds and angles away from their 'reference' or 'equilibrium' values, there is a function

that describes how the energy changes as bonds are rotated, and finally the force field contains terms that describe the interaction between non-bonded parts of the system. More sophisticated force fields may have additional terms, but they invariably contain these four components. An attractive feature of this representation is that the various terms can be ascribed to changes in specific internal coordinates such as bond lengths, angles, the rotation of bonds or movements of atoms relative to each other. This makes it easier to understand how changes in the force field parameters affect its performance, and also helps in the parametrisation process. One functional form for such a force field that can be used to model single molecules or assemblies of atoms and/or molecules is:





Class: I- M.Sc (Chemistry) **Course Code**: 19CHP105-C

$$\mathscr{V}(\mathbf{r}^{N}) = \sum_{\text{bonds}} \frac{k_{i}}{2} (l_{i} - l_{i,0})^{2} + \sum_{\text{angles}} \frac{k_{i}}{2} (\theta_{i} - \theta_{i,0})^{2} + \sum_{\text{torsions}} \frac{V_{n}}{2} (1 + \cos(n\omega - \gamma))$$
$$+ \sum_{i=1}^{N} \sum_{j=i+1}^{N} \left(4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \frac{q_{i}q_{j}}{4\pi\varepsilon_{0}r_{ij}} \right)$$
(4.1)

 $\mathscr{V}(\mathbf{r}^N)$ denotes the potential energy, which is a function of the positions (r) of N particles (usually atoms). The various contributions are schematically represented in Figure 4.1. The first term in Equation (4.1) models the interaction between pairs of bonded atoms, modelled here by a harmonic potential that gives the increase in energy as the bond length l_i deviates from the reference value $l_{i,0}$. The second term is a summation over all valence angles in the molecule, again modelled using a harmonic potential (a valence angle is the angle formed between three atoms A–B–C in which A and C are both bonded to B). The third term in Equation (4.1) is a torsional potential that models how the energy changes as a bond rotates. The fourth contribution is the non-bonded term. This is calculated between all pairs of atoms (*i* and *j*) that are in different molecules or that are in



Fig 4.1 Schematic representation of the four key contributions to a molecular mechanics force field bond stretching, angle bending and torsional terms and non-bonded interactions







Fig. 4.2. A typical force field model for propane contains ten bond-stretching terms, eighteen angle-bending terms, eighteen torsional terms and 27 non-bonded interactions

the same molecule but separated by at least three bonds (i.e. have a 1, *n* relationship where $n \ge 4$). In a simple force field the non-bonded term is usually modelled using a Coulomb potential term for electrostatic interactions and a Lennard-Jones potential for van der Waals interactions.

We shall discuss the nature of these different contributions in more detail in Sections 4.3-4.10, but here we consider how the simple force field of Equation (4.1) would be used to calculate the energy of a conformation of propane (Figure 4.2). Propane has ten bonds: two C–C bonds and eight C–H bonds. The C–C bonds are symmetrically equivalent but the C–H bonds fall into two classes, one group corresponding to the two hydrogens bonded to the central methylene (CH₂) carbon and one group corresponding to the six hydrogens bonded to the methyl carbons. In some sophisticated force fields different parameters would be used for these two different types of C-H bond, but in most force fields the same bonding parameters (i.e. k_i and $l_{i,0}$) would be used for each of the eight C-H bonds. This is an example of the way in which the same parameters can be used for a wide variety of molecules. There are 18 different valence angles in propane, comprising one C–C–C angle, ten C–C–H angles and seven H–C–H angles. Note that all angles are included in the force field model even though some of them may not be independent of the others. There are 18 torsional terms: twelve H–C–C–H torsions and six H–C–C–C torsions. Each of these is modelled with a cosine series expansion that has minima at the trans and gauche conformations. Finally, there are 27 non-bonded terms to calculate, comprising 21 H-H interactions and six H-C interactions. The electrostatic contribution would be calculated using Coulomb's law from partial atomic charges associated with



Class: I- M.Sc (Chemistry) Course Code: 19CHP105-C

Course Name: Molecular Modellingand Drug DesignUnit: II (Force Fields)Batch: 2019 - 2021

each atom and the van der Waals contribution as a Lennard-Jones potential with appropriate ε_{ij} and σ_{ij} parameters. A sizeable number of terms are thus included in the force field model, even for a molecule as simple as propane. Even so, the number of terms (73) is many fewer than the number of integrals that would be involved in an equivalent *ab initio* quantum mechanical calculation.

Some General Features of Molecular Mechanics Force Fields

To define a force field one must specify not only the functional form but also the parameters (i.e. the various constants such as k_i , V_n and σ_{ij} in Equation (4.1)); two force fields may use an identical functional form yet have very different parameters. Moreover, force fields with the same functional form but different parameters, and force fields with different functional forms, may give results of comparable accuracy. A force field should be considered as a single entity; it is not strictly correct to divide the energy into its individual components, let alone to take some of the parameters from one force field and mix them with parameters from another force field. Nevertheless, some of the terms in a force field are sufficiently independent of the others (particularly the bond and angle terms) to make this an acceptable approximation in certain cases.

The force fields used in molecular modelling are primarily designed to reproduce structural properties but they can also be used to predict other properties, such as molecular spectra. However, molecular mechanics force fields can rarely predict spectra with great accuracy (although the more recent molecular mechanics force fields are much better in this regard). A force field is generally designed to predict certain properties and will be parametrised accordingly. While it is useful to try to predict other quantities which have not been included in the parametrisation process it is not necessarily a failing if a force field is unable to do so.

Transferability of the functional form and parameters is an important feature of a force field. Transferability means that the same set of parameters can be used to model a series of related molecules, rather than having to define a new set of parameters for each individual molecule. For example, we would expect to be able to use the same set of parameters for all *n*-alkanes. Transferability is clearly important if we want to use the force field to make predictions. Only for some small systems, where particularly accurate work is required, may it be desirable to develop a model specific to that molecule.



| Class: I- M.Sc (Chemistry) | Course Name: Molecular Modelling | and Drug Design |
|----------------------------|----------------------------------|-------------------|
| Course Code: 19CHP105-C | Unit: II (Force Fields) | Batch: 2019 -2021 |

One important point that we should bear in mind as we undertake a deeper analysis of molecular mechanics is that force fields are *empirical*; there is no 'correct' form for a force field. Of course, if one functional form is shown to perform better than another it is likely that form will be favoured. Most of the force fields in common use do have a very similar form, and it is tempting to assume that this must therefore be the optimal functional form Certainly such models tend to conform to a useful picture of the interactions present in a system, but it should always be borne in mind that there may be better forms, particularly when developing a force field for new classes of molecule. The functional forms employed in molecular mechanics force fields are often a compromise between accuracy and computational efficiency; the most accurate functional form may often be unsatisfactory for efficient computation. As the performance of computers increases so it becomes possible to incorporate more sophisticated models. An additional consideration is that in order toreactechniques such as energy minimisation and molecular dynamics, it is usually desirable to be able to calculate the first and second derivatives of the energy with respect to the atomic coordinates.

A concept that is common to most force fields is that of an atom type. When preparing the input for a quantum mechanics calculation it is usually necessary to specify the atomic numbers of the nuclei present, together with the geometry of the system and the overall charge and spin multiplicity. For a force field the overall charge and spin multiplicity are not explicitly required, but it is usually necessary to assign an atom type to each atom in the system. The atom type is more than just the atomic number of an atom; it usually contains information about its hybridisation state and sometimes the local environment. For example, it is necessary in most force fields to distinguish between sp³-hybridised carbon atoms (which adopt a tetrahedral geometry), sp²-hybridised carbons (which are trigonal) and sp-hybridised carbons (which are linear). Each force field parameter is expressed in terms of these atom types, so that the reference angle θ_0 for a tetrahedral carbon atom would be near 109.5° and that for a trigonal carbon would be near 120°. The atom types in some force fields reflect the neighbouring environment as well as the hybridisation and can be quite extensive for some atoms. For example, the MM2, MM3 and MM4 force fields of Allinger and co-workers that are widely used for calculations on 'small' molecules [Allinger 1977; Allinger et al. 1989, 1990a, b, 1996a, b; Lii and Allinger 1989; Nevins et al. 1996a, b, c] distinguish the following types of carbon atom: sp³, sp², sp, carbonyl, cyclopropane, radical, cyclopropene and carbonium ion. In the AMBER force field of Kollman and co-workers [Weiner et al. 1984; Cornell et al. 1995] the carbon atom at the junction between a six- and a five-membered ring (e.g. in the amino acid tryptophan) is assigned





Fig. 4.3 AMBER atom types for the amino acids histidine, tryptophan and phenylalanine. There are three possible *protonation states of histidine.*

an atom type that is different from the carbon atom in an isolated five-membered ring such as histidine, which in turn is different from the atom type of a carbon atom in a benzene ring. Indeed, the AMBER force field uses different atom types for a histidine amino acid depending upon its protonation state (Figure 4.3). Other, more general, force fields would assign these atoms to the same generic 'sp² carbon' atom type. It is often found that force fields which are designed for modelling specific classes of molecule (such as proteins and nucleic acids, in the case of AMBER) use more specific atom types than force fields designed for general-purpose use.

We now discuss in some detail the individual contributions to a molecular mechanics force field, giving a selection of the various functional forms that are in common use. We shall then consider the important task of parametrisation, in which values for the many force constants are derived. Our discussion will be illuminated by examples chosen from contemporary force fields in widespread use and the MM2/MM3/MM4 and AMBER force fields in particular.

Bond Stretching

| Class: I- M.Sc (Chemistry) | Course Name: Molecular | Modelling and Drug Design |
|----------------------------|-------------------------|---------------------------|
| Course Code: 19CHP105-C | Unit: II (Force Fields) | Batch : 2019 -2021 |

The potential energy curve for a typical bond has the form shown in Figure 4.4. Of the many functional forms used to model this curve, that suggested by Morse is particularly useful The Morse potential has the form:

$$\nu(l) = D_{\rm e} \{1 - \exp[-a(l - l_0)]\}^2 \tag{4.2}$$

 $D_{\rm e}$ is the depth of the potential energy minimum and $a = \omega \sqrt{\mu/2D_{\rm e}}$, where μ is the reduced mass and ω is the frequency of the bond vibration. ω is related to the stretching constant of the bond, k, by $\omega = \sqrt{k/\mu}$. l_0 is the reference value of the bond. The Morse potential is not usually used in molecular mechanics force fields. In part this is because it is not particularly amenable to efficient computation but also because it requires three parameters to be specified for each bond. Moreover, it is rare in molecular mechanics calculations for



Fig 4 4: Variation in bond energy with interatomic separation.

bonds to deviate significantly from their equilibrium values; the Morse curve describes a wide range of behaviour from the strong equilibrium behaviour to dissociation. Consequently, simpler expressions are often used. The most elementary approach is to use a Hooke's law formula in which the energy varies with the square of the displacement from the reference bond length l_0 :

$$v(l) = \frac{k}{2}(l - l_0)^2 \tag{4.3}$$



| Class: I- M.Sc (Chemistry) | Course Name: Molecular M | lodelling and Drug Design |
|----------------------------|--------------------------|---------------------------|
| Course Code: 19CHP105-C | Unit: II (Force Fields) | Batch : 2019 -2021 |

The astute reader will have noticed our use of the term 'reference bond length' (sometimes called the 'natural bond length') for the parameter l_0 . This parameter is commonly called the 'equilibrium' bond length, but to do so can be misleading. The reference bond length is the value that the bond adopts when all other terms in the force field are set to zero. The equilibrium bond length, by contrast, is the value that is adopted in a minimum energy structure, when all other terms in the force field contribute. The complex interplay between the various components in the force field means that the bond may well deviate slightly from its reference value in order to compensate for other contributions to the energy. It is also important to recognise that 'real' molecules undergo vibrational motion (even at absolute zero, there is a zero-point energy due to vibrational motion). A true bond-stretching potential is not harmonic but has a shape similar to that in Figure 4.4, which means that the 'average' length of the bond in a vibrating molecule will deviate from the equilibrium value for the hypothetical motionless state. The effects are usually small, but they are significant if one wishes to predict bond lengths to thousandths of an angström. When comparing the results of calculations with experimental data, one must also remember that different experimental techniques measure different 'equilibrium' values, especially when the experiments are performed at different temperatures. The errors in experimentally determined bond lengths can be quite large; for example, libration of a molecule in a crystal means that the bond lengths determined by X-ray methods at room temperature may have errors as large as 0.015 Å. MM2 was parametrised to fit the values obtained by electron diffraction, which give the mean distances between atoms averaged over the vibrational motion at room temperature.

The forces between bonded atoms are very strong and considerable energy is required to cause a bond to deviate significantly from its equilibrium value. This is reflected in the magnitude of the force constants for bond stretching; some typical values from the MM2 force field are shown in Table 4.1, where it can be seen that those bonds one would

| Bo | ond | / ₀ (Å) | <i>k</i> (kcal mol ⁻¹ Å ⁻²) |
|----|----------------------------------|--------------------|--|
| Cs | p ³ -Csp ³ | 1.523 | 317 |
| Cs | p ³ -Csp ² | 1.497 | 317 |
| Cs | $p^2 = Csp^2$ | 1.337 | 690 |
| Cs | p ² =O | 1.208 | 777 |
| Cs | p ³ —Nsp ³ | 1.438 | 367 |
| C- | N (amide) | 1 345 | 719 |
| | | | |

Table 4 1 Force constants and reference bond lengths forselected bonds [Allinger 1977]



Class: I- M.Sc (Chemistry)Course Name: Molecular Modelling and Drug DesignCourse Code: 19CHP105-CUnit: II (Force Fields)Batch: 2019 -2021



Fig. 4.5: Comparison of the simple harmonic potential (Hooke's law) with the Morse curve

intuitively expect to be stronger have large force constants (contrast C–C with C=C and N=N). A deviation of just 0.2 Å from the reference value l_0 with a force constant of 300 kcal mol⁻¹ Å⁻² would cause the energy of the system to rise by 12 kcal/mol.

The Hooke's law functional form is a reasonable approximation to the shape of the potential energy curve at the bottom of the potential well, at distances that correspond to bonding in ground-state molecules. It is less accurate away from equilibrium (Figure 4.5). To model the Morse curve more accurately, cubic and higher terms can be included and the bond-stretching potential can be written as follows:

$$\nu(l) = \frac{k}{2}(l-l_0)^2 [1-k'(l-l_0)-k''(l-l_0)^2-k'''(l-l_0)^3\dots]$$
(4.4)



Class: I- M.Sc (Chemistry) Course Code: 19CHP105-C

Course Name: Molecular Modelling and Drug Design Unit: II (Force Fields) Batch: 2019 -2021



Fig 4 6. A cubic bond-stretching potential passes through a maximum but gives a better approximation to the Morse curve close to the equilibrium structure than the quadratic form

An undesirable side-effect of an expansion that includes just a quadratic and a cubic term (as is employed in MM2) is that, far from the reference value, the cubic function passes through a maximum. This can lead to a catastrophic lengthening of bonds (Figure 4.6). One way to accommodate this problem is to use the cubic contribution only when the structure is sufficiently close to its equilibrium geometry and is well inside the 'true' potential well. MM3 also includes a quartic term; this eliminates the inversion problem and leads to an even better description of the Morse curve.

Angle Bending

The deviation of angles from their reference values is also frequently described using a Hooke's law or harmonic potential:

$$\nu(\theta) = \frac{k}{2}(\theta - \theta_0)^2 \tag{4.5}$$

The contribution of each angle is characterised by a force constant and a reference value. Rather less energy is required to distort an angle away from equilibrium than to stretch or compress a bond, and the force constants are proportionately smaller, as can be observed in Table 4.2.

| Class: I- M.Sc (Chemistry) Course Code: 19CHP105-C | Course Nan Unit: II (Force F | ne: Molecular Modelling and Drug Desi Fields) Batch: 2019 -20 | ign 21 |
|---|---------------------------------|--|-----------|
| Angle | θο | <i>k</i> (kcal mol ⁻¹ deg ⁻¹) | |
| Csp ³ -Csp ³ -Csp ³ | 109.47 | 0.0099 | |
| Csp ³ –Csp ³ –H | 109.47 | 0.0079 | |
| H–Csp ³ –H | 109.47 | 0 0070 | |
| Csp ³ -Csp ² -Csp ³ | 117.2 | 0.0099 | |
| Csp ³ -Csp ² =Csp ² | 121.4 | 0.0121 | |
| Csp ³ -Csp ² =O | 122.5 | 0.0101 | |
| | | | |

| Table 4.2 | Force | constants | and | reference | angles | for | selected | angles |
|-----------|--------|-----------|-----|-----------|--------|-----|----------|--------|
| [Allinger | 1977]. | | | - | _ | - | | _ |

As with the bond-stretching terms, the accuracy of the force field can be improved by the incorporation of higher-order terms. MM2 contains a quartic term in addition to the quadratic term. Higher-order terms have also been included to treat certain pathological cases such as very highly strained molecules. The general form of the angle-bending term then becomes:

$$v(\theta) = \frac{k}{2} (\theta - \theta_0)^2 [1 - k'(\theta - \theta_0) - k''(\theta - \theta_0)^2 - k'''(\theta - \theta_0)^3 \dots]$$
(4.6)
rms

Torsional Terms

The bond-stretching and angle-bending terms are often regarded as 'hard' degrees of freedom, in that quite substantial energies are required to cause significant deformations from

their reference values. Most of the variation in structure and relative energies is due to the complex interplay between the torsional and non-bonded contributions.

The existence of barriers to rotation about chemical bonds is fundamental to understanding the structural properties of molecules and conformational analysis. The three minimumenergy staggered conformations and three maximum-energy eclipsed structures of ethane are a classic example of the way in which the energy changes with a bond rotation. Quantum mechanical calculations suggest that this barrier to rotation can be considered to arise from antibonding interactions between the hydrogen atoms on opposite ends of the molecule; the antibonding interactions are minimised when the conformation is staggered and are at a maximum when the conformation is eclipsed. Many force fields are used for modelling flexible molecules where the major changes in conformation are due to rotations about bonds; in order to simulate this it is essential that the force field properly represents the energy profiles of such changes.



| Class: I- M.Sc (Chemistry) | Course Name: Molecular | Modelling and Drug Design |
|----------------------------|-------------------------|---------------------------|
| Course Code: 19CHP105-C | Unit: II (Force Fields) | Batch : 2019 -2021 |

Not all molecular mechanics force fields use torsional potentials; it may be possible to rely upon non-bonded interactions between the atoms at the end of each torsion angle (the 1,4 atoms) to achieve the desired energy profile. However, most force fields for 'organic' molecules do use explicit torsional potentials with a contribution from each bonded quartet of atoms A-B-C-D in the system. Thus there would be nine individual torsional terms for ethane and 24 for benzene ($6 \times C-C-C-C$, $12 \times C-C-C-H$ and $6 \times H-C-C-H$). Torsional potentials are almost always expressed as a cosine series expansion. One functional form is:

$$v(\omega) = \sum_{n=0}^{N} \frac{V_n}{2} \left[1 + \cos(n\omega - \gamma) \right]$$
(4.7)

 ω is the torsion angle.

An alternative but equivalent expression is:

$$v(\omega) = \sum_{n=0}^{N} C_n \cos(\omega)^n$$
(4.8)

 V_n in Equation (4.7) is often referred to as the 'barrier' height, but to do so is misleading, obviously so when more than one term is present in the expansion. Moreover, other terms in the force field equation contribute to the barrier height as a bond is rotated, especially the non-bonded interactions between the 1,4 atoms. The value of V_n does, however, give a qualitative indication of the relative barriers to rotation; for example, V_n for an amide bond will be larger than for a bond between two sp³ carbon atoms. *n* in Equation (4.7) is the *multiplicity*; its value gives the number of minimum points in the function as the bond is rotated through 360°. γ (the phase factor) determines where the torsion angle passes through its minimum value. For example, the energy profile for rotation about the single bond between two sp³ carbon atoms could be represented by a single torsional term with n = 3 and $\gamma = 0^{\circ}$. This would give a threefold rotational profile with minima at torsion angles of $+60^{\circ}$, -60° and 180° and maxima at $\pm 120^{\circ}$ and 0° . A double bond between two sp² carbon atoms would have n = 2 and $\gamma = 180^{\circ}$, giving minima at 0° and 180° . The value of V_n would also be significantly larger for the double bond than for the single



Class: I- M.Sc (Chemistry) **Course Code:** 19CHP105-C



Fig 4.7. Torsional potential varies as shown for different values of V_n , n and γ .

bond. The effects of varying V_n , n and γ are illustrated in Figure 4.7 for commonly occurring torsional potentials.

Many of the torsional terms in the AMBER force field contain just one term from the cosine series expansion, but for some bonds it was found necessary to include more than one term. For example, to correctly model the tendency of O-C-C-O bonds to adopt a *gauche* conformation, a torsional potential with two terms was used for the O-C-C-O contribution:

$$\nu(\omega_{\rm C-O-O-C}) = 0.25(1 + \cos 3\omega) + 0.25(1 + \cos 2\omega)$$
(4.9)

The torsional energy for a OCH_2-CH_2O fragment (found in the sugars in DNA) varies with the torsion angle ω as shown in Figure 4.8. Another feature of the AMBER force field is its use of general torsional parameters. The energy profile for rotation about a bond that is described by a general torsional potential depends solely upon the atom types of the two




| Class: I- M.Sc (Chemistry) | Course Name: Molecular Modelling | and Drug Design |
|----------------------------|----------------------------------|-------------------|
| Course Code: 19CHP105-C | Unit: II (Force Fields) | Batch: 2019 -2021 |

Fig 4.8: Variation in torsional energy (AMBER force field) with O-C-C-O torsion angle (ω) for OCH_2-CH_2O fragment The minimum energy conformations arise for $\omega = 60^{\circ}$ and 300°

atoms that comprise the central bond and not upon the atom types of the terminal atoms. For example, all torsion angles in which the central bond is between two sp³-hybridised carbon atoms (e.g. H–C–C–H, C–C–C, H–C–C–C) are assigned the same torsional parameters, unless the torsion is a special case such as O-C-C-O. In its treatment of the torsional contribution, AMBER takes a position intermediate between those force fields which only ever use a single term in the torsional expansion and those which consistently use more terms for all torsions. MM2 falls into the latter category; it uses three terms in the expansion:

$$v(\omega) = \frac{V_1}{2} \left(1 + \cos\omega\right) + \frac{V_2}{2} \left(1 - \cos 2\omega\right) + \frac{V_3}{2} \left(1 + \cos 3\omega\right)$$
(4.10)

A physical interpretation has been ascribed to each of the three terms in the MM2 torsional expansion from an analysis of *ab initio* calculations on simple fluorinated hydrocarbons. The first, onefold term corresponds to interactions between bond dipoles, which are due to differences in electronegativity between bonded atoms. The twofold term is due to the effects of hyperconjugation (in alkanes) and conjugation effects (in alkenes), which provide 'double bond' character to the bond. The threefold term corresponds to steric interactions between the 1,4 atoms. It was found that the additional terms in the torsional potential were especially important for systems containing heteroatoms, such as the halogenated hydrocarbons and molecules containing CCOC and CCNC fragments.

With careful parametrisation a force field which uses more than one term in the torsional expansion will be more successful than a force field that uses only a single term (and this is borne out by the MM2 force field). The major drawback is that many parameters are required to model even a modest range of molecules.

Introduction to Nonbonded Interactions

Independent molecules and atoms interact through non-bonded forces, which also play an important role in determining the structure of individual molecular species. The non-bonded interactions do not depend upon a specific bonding relationship between atoms. They are 'through-space' interactions and are usually modelled as a function of some inverse power of the distance. The non-bonded terms in a force field are usually considered in two groups, one comprising electrostatic interactions and the other van der Waals interactions.

Electrostatic interactions

The Central Multipole Expansion

| Class: I- M.Sc (Chemistry) | Course Name: Molecular N | Modelling and Drug Design |
|----------------------------|--------------------------|---------------------------|
| Course Code: 19CHP105-C | Unit: II (Force Fields) | Batch : 2019 -2021 |

Electronegative elements attract electrons more than less electronegative elements, giving rise to an unequal distribution of charge in a molecule. This charge distribution can be represented in a number of ways, one common approach being an arrangement of fractional point charges throughout the molecule. These charges are designed to reproduce the electrostatic properties of the molecule. If the charges are restricted to the nuclear centres they are often referred to as *partial atomic charges* or *net atomic charges*. The electrostatic interaction between two molecules (or between different parts of the same molecule) is then calculated as a sum of interactions between pairs of point charges, using Coulomb's law:

$$\mathscr{V} = \sum_{i=1}^{N_{\rm A}} \sum_{j=1}^{N_{\rm B}} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \tag{4.19}$$

 $N_{\rm A}$ and $N_{\rm B}$ are the numbers of point charges in the two molecules. This approach to the representation and calculation of electrostatic interactions will be considered in more detail in Section 4.9.2. First, we shall consider an alternative approach to the calculation of electrostatic interactions which treats a molecule as a single entity and is (in principle at least) capable of providing a very efficient way to calculate electrostatic intermolecular interactions. This is the *central multipole expansion*, which is based upon the electric moments or multipoles: the charge, dipole, quadrupole, octopole, and so on introduced in Section 2.7.3. These moments are usually represented by the following symbols: q (charge), μ (dipole), Θ (quadrupole) and Φ (octopole). We are often interested in the lowest non-zero electric moment. Thus species such as Na^+ , Cl^- , NH_4^+ or $CH_3CO_2^-$ have the charge as their lowest non-zero moment. For many uncharged molecules the dipole is the lowest non-zero moment. Molecules such as N₂ and CO₂ have the quadrupole as their lowest non-zero moment. The lowest non-zero moment for methane and tetrafluoromethane is the octopole. Each of these multipole moments can be represented by an appropriate distribution of charges. Thus a dipole can be represented using two charges placed an appropriate distance apart. A quadrupole can be represented using four charges and an octopole by eight charges. A complete description of the charge distribution around a molecule requires all of the non-zero electric moments to the specified. For some molecules,







Fig 4 15: The electrostatic potential due to two point charges.

the lowest non-zero moment may not be the most significant and it may therefore be unwise to ignore the higher-order terms in the expansion without first checking their values. To illustrate how the multipolar expansion is related to a distribution of charges in a system, let us consider the simple case of a molecule with two charges q_1 and q_2 , positioned at $-z_1$ and z_2 , respectively (Figure 4.15). The electrostatic potential at point P (a distance *r* from the origin, r_1 from charge q_1 and r_2 from charge q_2) is then given by:

$$\phi(r) = \frac{1}{4\pi\varepsilon_0} \left(\frac{q_1}{r_1} + \frac{q_2}{r_2} \right)$$
(4.20)

By applying the cosine rule this can be written as follows (see Figure 4.15):

$$\phi(r) = \frac{1}{4\pi\varepsilon_0} \left(\frac{q_1}{\sqrt{r^2 + z_1^2 + 2rz_1\cos\theta}} + \frac{q_2}{\sqrt{r^2 + z_1^2 - 2rz_1\cos\theta}} \right)$$
(4.21)

If $r \gg z_1$ and $r \gg z_2$ then this expression can be expanded as follows:

$$\phi(r) = \frac{1}{4\pi\varepsilon_0} \left(\frac{q_1 + q_2}{r} + \frac{(q_2 z_2 - q_1 z_1)\cos\theta}{r^2} + \frac{(q_1 z_1^2 + q_2 z_2^2)(3\cos^2\theta - 1)}{2r^3} + \cdots \right)$$
(4.22)

We can now associate the appropriate terms in the expansion with the various electric moments:

$$\phi(r) = \frac{1}{4\pi\varepsilon_0} \left(\frac{q}{r} + \frac{\mu\cos\theta}{r^2} + \frac{\Theta(3\cos^2\theta - 1)}{2r^3} + \cdots \right)$$
(4.23)

| Class: I- M.Sc (Chemistry) | Course Name: Molecular | Modelling and Drug Design |
|----------------------------|-------------------------|---------------------------|
| Course Code: 19CHP105-C | Unit: II (Force Fields) | Batch : 2019 -2021 |

Thus $(q_1 + q_2)$ is the charge; $(q_2z_2 - q_1z_1)$ is the dipole; $(q_1z_1^2 + q_2z_2^2)$ is the quadrupole, and so on. One interesting feature about a charge distribution is that only the first non-zero moment is independent of the choice of origin. Thus, if a molecule is electrically neutral (i.e. $q_1 + q_2 = 0$) then its dipole moment is independent of the choice of origin. This can be demonstrated for our two-charge system as follows. If the position of the origin is now moved to a point -z', then the dipole moment relative to this new origin is given by:

$$\mu' = q_2(z_2 + z') - q_1(z_1 - z') = \mu + qz'$$
(4.24)

Only if the total charge on the system (q) equals zero will the dipole moment be unchanged. Similar arguments can be used to show that if both the charge and the dipole moment are zero then the quadrupole moment is independent of the choice of origin. For convenience, the origin is often taken to be the centre of mass of the charge distribution.

The electric moments are examples of *tensor properties*: the charge is a rank 0 tensor (which is the same as a scalar quantity); the dipole is a rank 1 tensor (which is the same as a vector, with three components along the x, y and z axes); the quadrupole is a rank 2 tensor with nine components, which can be represented as a 3×3 matrix. In general, a tensor of rank n has 3^n components.

For a distribution of charges (one not restricted to lie along one of the Cartesian axes), the dipole moment is given by:

$$\boldsymbol{\mu} = \sum q_i \mathbf{r}_i \tag{4.25}$$

The components of the dipole moment along the *x*, *y* and *z* axes are $\sum q_i x_i$, $\sum q_i y_i$ and $\sum q_i z_i$. The analogous way to define the quadrupole moment is as follows:

$$\Theta = \begin{pmatrix} \sum q_i x_i^2 & \sum q_i x_i y_i & \sum q_i x_i z_i \\ \sum q_i y_i x_i & \sum q_i y_i^2 & \sum q_i y_i z_i \\ \sum q_i z_i x_i & \sum q_i z_i y_i & \sum q_i z_i^2 \end{pmatrix}$$
(4.26)

This definition of the quadrupole is obviously dependent upon the orientation of the charge distribution within the coordinate frame. Transformation of the axes can lead to alternative definitions that may be more informative. Thus the quadrupole moment is commonly defined as follows:

$$\Theta = \frac{1}{2} \begin{pmatrix} \sum_{i} q_{i}(3x_{i}^{2} - r_{i}^{2}) & 3\sum_{i} q_{i}x_{i}y_{i} & 3\sum_{i} q_{i}x_{i}z_{i} \\ 3\sum_{i} q_{i}x_{i}z_{i} & \sum_{i} q_{i}(3y_{i}^{2} - r_{i}^{2}) & 3\sum_{i} q_{i}y_{i}z_{i} \\ 3\sum_{i} q_{i}x_{i}z_{i} & 3\sum_{i} q_{i}y_{i}z_{i} & \sum_{i} q_{i}(3z_{i}^{2} - r_{i}^{2}) \end{pmatrix}$$
(4 27)

In Equation (4.27) $r_i^2 = x_i^2 + y_i^2 + z_i^2$. This definition enables one to assess the deviation from spherical symmetry as a spherically symmetric charge distribution will have

$$\sum_{i} q_{i} x_{i}^{2} = \sum_{i} q_{i} y_{i}^{2} = \sum_{i} q_{i} z_{i}^{2} = \frac{1}{3} \sum_{i} q_{i} r_{i}^{2}$$
(4.28)

| Class: I- M.Sc (Chemistry) | Course Name: Molecular | Modelling and Drug Design |
|----------------------------|-------------------------|---------------------------|
| Course Code: 19CHP105-C | Unit: II (Force Fields) | Batch : 2019 -2021 |

and so the diagonal elements of the tensor will be zero. Quadrupoles are also reported in terms of the *principal axes*; these are three mutually perpendicular axes α , β and γ , which are linear combinations of *x*, *y* and *z* such that the quadrupole tensor is diagonal (i.e. off-diagonal elements are zero):

$$\Theta = \begin{pmatrix} \Theta_{\alpha\alpha} & 0 & 0\\ 0 & \Theta_{\beta\beta} & 0\\ 0 & 0 & \Theta_{\gamma\gamma} \end{pmatrix}$$
(4 29)

Let us now consider the effect of placing another molecule w^i a linear charge distribution (charges q'_1 and q'_2) with its centre of mass at the point P. T' e relative orientation of the two



Fig. 4.16: The relative on ... "on of two dipoles

molecules can be conscribed in terms of four parameters (the distance joining their centres of mass and mee ang is as shown in Figure 4.16). The electrostatic interaction between the two molecules in the local ed by multiplying end charge by the potential at that point and adding the result for the following expression is the result [Buckingham 1959]:



Class: I- M.Sc (Chemistry) Course Code: 19CHP105-C

$$\mathscr{V}(q,q') = \frac{1}{4\pi\varepsilon_0} \begin{cases} \frac{qq'}{r} \\ +\frac{1}{r^2} (q\mu'\cos\theta + q'\mu\cos\theta') \\ +\frac{\mu\mu'}{r^3} (2\cos\theta\cos\theta' + \sin\theta\sin\theta'\cos\zeta) \\ +\frac{1}{2r^3} [q\Theta'(3\cos^2\theta' - 1) + q'\Theta(3\cos^2\theta - 1)] \\ +\frac{3}{2r^4} [\mu\Theta'\{\cos\theta(3\cos^2\theta' - 1) + 2\epsilon \,\mu\theta\sin\theta'\cos\theta'\cos\zeta\} \\ +\mu'\Theta\{\cos\theta'(3\cos^2\theta - 1) + 2\sin\phi\cos\theta'\cos\zeta\}] \\ +\frac{3\Theta\Theta'}{4r^5} [1 - 5\cos^2\theta - 5\cos^2\phi - 17\cos^2\theta - \frac{2}{2}\theta' \\ +2\sin^2\theta\sin^2 - 3s^2\zeta + 16\sin\phi'\cos\theta - \frac{2}{3}\theta' \\ +\cdots \end{cases} \end{cases}$$
(4.30)

The energy of interaction between two charge stributions is as an infinite series that includes charge-charge, charge dipole, dipole-dipole charge-quadrupole, dipole-quadrupole interactions, quadrupole-q and q and q



Fig 4.17 The charge-charge energy decays much more slowly (αr^{-1}) than the dipole-dipole energy (αr^{-3})



between these groups then decays as r^{-3} rather than the r^{-1} dependence of each individual charge-charge interaction. This can be seen in Figure 4.17, in which the functions r^{-1} and r^{-3} have been plotted as a function of distance. Even when the dipole–dipole interaction energy has fallen off almost to zero the charge–charge interaction energy is still significant. In general, the interaction energy between two multipoles of order n and m decreases as $r^{-(n+m+1)}$. It should be emphasised again that these expressions are only valid when the separation of the two molecules, r, is much larger than the internal dimensions of the molecules. The favourable arrangements for the various multipoles are shown in Figure 4.18.

A central multipole expansion therefore provides a way to calculate the electrostatic interaction between two molecules. The multipole moments can be obtained from the wave-function and can therefore be calculated using quantum mechanics (see Section 2.7.3) or can be determined from experiment. One example of the use of a multipole expansion is



Fig. 4 18 The most favourable orientations of various multipoles (Figure adapted from Buckingham A D 1959 Molecular Quadrupole Moments. Quarterly Reviews of the Chemical Society **13**:183–214.)

the benzene model of Claessens, Ferrario and Ryckaert [Claessens *et al.* 1983]. Benzene has no charge and no dipole moment, but it does have a sizeable quadrupole. The inclusion of the quadrupole was found to give clearly superior results in molecular dynamics simulations of the liquid state over models that lacked any electronic contribution.

The main advantage of the multipolar description for calculating the electrostatic interactions between molecules is its efficiency. For example, the charge-charge interaction energy between two benzene molecules would require 144 individual charge-charge interactions with a partial atomic charge model rather than the single quadrupole-quadrupole term. Unfortunately, the multipole expansion is not applicable when the molecules are separated by distances comparable with the molecular dimensions. The formal condition



Class: I- M.Sc (Chemistry) Course Code: 19CHP105-C Course Name: Molecular Modellingand Drug DesignUnit: II (Force Fields)Batch: 2019 - 2021

for convergence of the multipolar interaction energy is that the distance between two interacting molecules should be larger than the sum of the distances from the centre of each molecule to the furthest part of its charge distribution. It a sphere is constructed around each molecule, positioned on its centre of mass, with a radius that encompasses all of the charge distribution, then the multipole expansion for the interaction between two molecules will converge if these spheres do not intersect. Even if one requires the sphere to encompass just the nuclei in a molecule (i.e. ignoring the fact that the charge distribution around a molecule extends to infinity) there may still be problems. For example, the convergence sphere for a molecule such as butane would extend beyond the van der Waals radiu in some directions, enabling other molecules to penetrate the convergence sphere, as illustrated in Figure 4.19. Another problem is that the multipolar expansion may be slow to converge. The multipolar expansion is often located at the centre of mass, but this may not be the best choice to achieve the most rapid convergence.



Fig. 4.19. The convergence sphere of the multipole expansion for a molecule such as butane may be penetrated by another molecule

There are other difficulties with the central multipole expansion. The multipole moments are properties of the entire molecule and so cannot be used to determine intramolecular interactions. The central multipole model thus tends to be restricted to calculations involving small molecules that are kept fixed in conformation during the calculation, and where the interactions between molecules act at their centres of mass. It can be a complicated procedure to calculate the forces acting on a molecule with a multipole model. The interaction between multipoles of zero order (i.e. charges) gives rise to a simple translational force. Multipoles of a higher order have directionality, and interactions between these produce a torque, or twisting force. Moreover, whereas the charge-charge forces are equal and opposite, the torque acting on molecule i due to another molecule j is not necessarily equal and opposite to the torque on molecule j due to molecule i.



| Class: I- M.Sc (Chemistry) | |
|----------------------------|--|
| Course Code: 19CHP105-C | |

Point-Charge Electrostatic models

We therefore return to the point-charge model for calculating electrostatic interactions. If sufficient point charges are used then all of the electric moments can be reproduced and the multipole interaction energy, Equation (4.30), is exactly equal to that calculated from the Coulomb summation, Equation (4.19).

An accurate representation of a molecule's electrostatic properties may require charges to be placed at locations other than at the atomic nuclei. A simple example of this is molecular nitrogen, which has a dipole moment of zero. The total charge on nitrogen is zero, and so an atomic partial charge model would put zero charge on each nucleus. However, nitrogen does have a quadrupole moment and this significantly affects its properties. The simplest way to model this is to place three partial charges along the bond: a charge of -q at each nucleus and +2q at the centre of mass. The quadrupole-quadrupole interaction between two nitrogen molecules can then be calculated by summing nine pairs of charge-charge interactions. The value of q can be calculated using the following relationship between the quadrupole moment and the partial charge:

$$\Theta = 2q(l/2)^2 \tag{4.31}$$

l is the bond length. The experimental quadrupole moment is consistent with a charge, *q*, of approximately 0.5*e*. In fact, a better representation of the electrostatic potential around the nitrogen molecule is obtained using the five-charge model shown in Figure 4.20.





Class: I- M.Sc (Chemistry) Course Code: 19CHP105-C Course Name: Molecular Modellingand Drug DesignUnit: II (Force Fields)Batch: 2019 -2021



Fig. 4 20: Two charge models for N_2 with the electrostatic potentials that they generate Also shown is the electrostatic potential calculated using ab initio quantum mechanics (6-31G^{*} basis set.) Negative contours are dashed and the zero contour is bold

Van der Waals Interactions

Electrostatic interactions cannot account for all of the non-bonded interactions in a system. The rare gas atoms are an obvious example; all of the multipole moments of a rare gas atom are zero and so there can be no dipole-dipole or dipole-induced dipole interactions. But there clearly must be interactions between the atoms, how else could rare gases have liquid and solid phases or show deviations from ideal gas behaviour? Deviations from ideal gas behaviour were famously quantitated by van der Waals, thus the forces that give rise to such deviations are often referred to as van der Waals forces.



| Class: I- M.Sc (Chemistry) | Course Name: Molecular Modelling | and Drug Design |
|----------------------------|----------------------------------|-------------------|
| Course Code: 19CHP105-C | Unit: II (Force Fields) | Batch: 2019 -2021 |

If we were to study the interaction between two isolated argon atoms using a molecular beam experiment then we would find that the interaction energy varies with the separation in a manner as shown in Figure 4.32. The other rare gases show a similar behaviour. The essential features of this curve are as follows. The interaction energy is zero at infinite distance (and indeed is negligible even at relatively short distances). As the separation is reduced, the energy decreases, passing through a minimum at a distance of approximately 3.8 Å for argon. The energy then rapidly increases as the separation decreases further. The force between the atoms, which equals minus the first derivative of the potential energy with respect to distance, is also shown in Figure 4.32. A variety of experiments have been used to provide evidence for the nature of the van der Waals interactions, including gas imperfections, molecular beams, spectroscopic studies and measurements of transport properties.

Modelling Van der Waals interactions

The dispersive and exchange-repulsive interactions between atoms and molecules can be calculated using quantum mechanics, though such calculations are far from trivial, requiring electron correlation and large basis sets. For a force field we require a means to model the interatomic potential curve accurately (Figure 4.32), using a simple empirical



Fig 4 34. The Lennard-Jones potential.

expression that can be rapidly calculated. The need for a function that can be rapidly evaluated is a consequence of the large number of van der Waals interactions that must be determined in many of the systems that we would like to model. The best known of the van der Waals potential functions is the *Lennard-Jones 12–6 function*, which takes the following form for the interaction between two atoms:



Class: I- M.Sc (Chemistry)Course Name: Molecular Modelling and Drug Design
Unit: II (Force Fields)Course Code: 19CHP105-CUnit: II (Force Fields)

$$v(r) = 4\varepsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^{6} \right]$$
(4.63)

The Lennard-Jones 12–6 potential contains just two adjustable parameters: the collision diameter σ (the separation for which the energy is zero) and the well depth ε . These parameters are graphically illustrated in Figure 4.34. The Lennard-Jones equation may also be expressed in terms of the separation at which the energy passes through a minimum, $r_{\rm m}$ (also written $r^{\tilde{s}}$). At this separation, the first derivative of the energy with respect to the internuclear distance is zero (i.e. $\partial v / \partial r = 0$), from which it can easily be shown that $r_{\rm m} = 2^{1/6} \sigma$. We can thus also write the Lennard-Jones 12–6 potential function as follows:



A is equal to εr_m^{12} (or $4\varepsilon \sigma^{12}$) and C is equal to $2\varepsilon r_m^6$ (or $4\varepsilon \sigma^6$).

The Lennard-Jones potential is characterised by an attractive part that varies as r^{-6} and a repulsive part that varies as r^{-12} . These two components are drawn in Figure 4.35. The r^{-6} variation is of course the same power-law relationship found for the leading term in theoretical treatments of the dispersion energy such as the Drude model. There are no strong theoretical arguments in favour of the repulsive r^{-12} , especially as quantum



Fig. 4 35 The Lennard-Jones potential is constructed from a repulsive component (αr^{-12}) and an attractive component (αr^{-6})

| Class: I- M.Sc (Chemistry) | Course Name: Molecular | Modelling and Drug Design |
|----------------------------|-------------------------|---------------------------|
| Course Code: 19CHP105-C | Unit: II (Force Fields) | Batch : 2019 -2021 |

mechanics calculations suggest an exponential form. The twelfth power term is found to be quite reasonable for rare gases but is rather too steep for other systems such as hydro-carbons. However, the 6–12 potential is widely used, particularly for calculations on large systems, as r^{-12} can be rapidly calculated by squaring the r^{-6} term. The r^{-6} term can also be calculated from the square of the distance without having to perform a computationally expensive square root calculation. Different powers have also been used for the repulsive part of the potential; values of 9 or 10 give a less steep curve and are used in some force fields. Lennard-Jones' original potential has been written in the following general form:

$$\nu(r) = k\varepsilon \left[\left(\frac{\sigma}{r}\right)^n - \left(\frac{\sigma}{r}\right)^m \right]; \qquad k = \frac{n}{n-m} \left(\frac{n}{m}\right)^{m/(n-m)}$$
(4.66)

Equation (4.66) returns the Lennard-Jones potential for n = 12 and m = 6.

Hydrogen Bonding in Molecular Mechanics

Some force fields replace the Lennard-Jones 6–12 term between hydrogen-bonding atoms by an explicit hydrogen-bonding term, which is often described using a 10–12 Lennard-Jones potential[.]

$$\nu(r) = \frac{A}{r^{12}} - \frac{C}{r^{10}} \tag{4.85}$$

This function is used to model the interaction between the donor hydrogen atom and the heteroatom acceptor atom. Its use is intended to improve the accuracy with which the geometry of hydrogen-bonding systems is predicted. Other force fields incorporate a more complicated hydrogen-bonding function that takes into account deviations from the geometry of the hydrogen bond and is thus dependent upon the coordinates of the donor and acceptor atoms as well as the hydrogen atom. For example, the YETI force field [Vedani 1988] uses the following form for its hydrogen bonding term:

$$\nu_{\rm HB} = \left(\frac{A}{r_{\rm H-Acc}^{12}} - \frac{C}{r_{\rm H-Acc}^{10}}\right) \cos^2\theta_{\rm Don-H-Acc} \cos^4\omega_{\rm H-Acc-LP}$$
(4.86)

The energy in Equation (4.86) depends upon the distance from the hydrogen to the acceptor, the angle subtended at the hydrogen by the bonds to the donor and the acceptor, and the deviation of the hydrogen bond from the closest lone-pair direction at the acceptor atom ($\omega_{\rm H-Acc-LP}$ in Equation (4.86), Figure 4.39).

The GRID program [Goodford 1985] that is used for finding energetically favourable regions in protein binding sites uses a direction-dependent 6–4 function:

$$v_{\rm HB} = \left(\frac{C}{d^6} - \frac{D}{d^4}\right) \cos^m \theta \tag{4.87}$$

 θ is the angle subtended at the hydrogen and *m* is usually set to 4.



Class: I- M.Sc (Chemistry)Course Name: Molecular Modelling and Drug Design
Unit: II (Force Fields)Course Code: 19CHP105-CBatch: 2019 -2021



Fig 4 39: Definition of hydrogen-bond geometry used in YETI force field

By no means do all force fields contain explicit hydrogen-bonding terms; most rely upon electrostatic and van der Waals interactions to reproduce hydrogen bonding.

Force Field Models for the Simulation of Liquid Water

Many of the concepts that we have considered so far can be illustrated by examining some of the empirical models that have been developed to study water. Despite its small size, water acts as a paradigm for the different force field models that we have discussed. Moreover, many of its properties can be easily determined using computer simulation methods and so readily compared with experiment. It is also one of the most challenging systems to model accurately. A wide range of water models have been proposed. The computational efficiency with which the energy can be calculated using a given model is often an important factor as there may be a very large number of water molecules present, together with a solute; most of the force fields used to simulate liquid water thus use effective pairwise potentials with no explicit three-body terms or polarisation effects.

Water models can be conveniently divided into three types. In the simple interaction-site models each water molecule is maintained in a rigid geometry and the interaction between molecules is described using pairwise Coulombic and Lennard-Jones expressions. Flexible models permit internal changes in conformation of the molecule. Finally, models have been developed that explicitly include the effects of polarisation and many-body effects.

Simple Water Models



| Class: I- M.Sc (Chemistry) | Course Name: Molecular Modelling | and Drug Design |
|----------------------------|----------------------------------|-------------------|
| Course Code: 19CHP105-C | Unit: II (Force Fields) | Batch: 2019 -2021 |

The 'simple' water models use between three and five interaction sites and a rigid water geometry. The TIP3P [Jorgensen *et al.* 1983] and SPC [Berendsen *et al.* 1981] models use a total of three sites for the electrostatic interactions; the partial positive charges on the hydrogen atoms are exactly balanced by an appropriate negative charge located on the oxygen atom. The van der Waals interaction between two water molecules is computed using a Lennard-Jones function with just a single interaction point per molecule centred on the oxygen atom; no van der Waals interactions involving the hydrogen atoms are calculated. The TIP3P and SPC models differ slightly in the geometry of each water molecule, in the

| | SPC | SPC/E | TIP3P | BF | TIP4P | ST2 |
|--|--------|---------|--------|-------|--------|-----------------|
| <i>г</i> (ОН), Å | 1.0 | 1.0 | 0.9572 | 0.96 | 0.9572 | 1.0 |
| HOH, deg | 109.47 | 109.47 | 104.52 | 105.7 | 104.52 | 109.47 |
| $A \times 10^{-3}$, kcal Å ¹² /mol | 629.4 | 629.4 | 582.0 | 560.4 | 600.0 | 238.7 |
| C, kcal Å ⁶ /mol | 625.5 | 625.5 | 595.0 | 837.0 | 610.0 | 268.9 |
| q(O) | 0.82 | -0.8472 | -0.834 | 0.0 | 0.0 | 0.0 |
| <i>q</i> (H) | 0.41 | 0.4238 | 0.417 | 0.49 | 0.52 | 0.2375 |
| q(M) | 0.0 | 0.0 | 0.0 | -0.98 | - 1.04 | -0 2 375 |
| r(OM), Å | 0.0 | 0.0 | 0.0 | 0.15 | 0.15 | 0.8 |

Table 4.3 A comparison of various water models [Jorgensen et al 1983]. For the ST2 potential, q(M) is the charge on the 'lone pairs', which are a distance 0.8 Å from the oxygen atom (see Figure 4.40)

hydrogen charges and in the Lennard-Jones parameters. These differences are indicated in Table 4.3, which also includes data for the SPC/E model [Berendsen *et al.* 1987], which is an updated version of the SPC model. The four-site models such as that of Bernal and Fowler [Bernal and Fowler 1933] (which is now relatively little used but is important for historical reasons as it dates from 1933) and Jorgensen's TIP4P model [Jorgensen *et al.* 1983] shift the negative charge from the oxygen atom to a point along the bisector of the HOH angle towards the hydrogens (Figure 4.40). The parameters for these two models are also given in the table. The most commonly used five-site model is the ST2 potential of Stillinger and Rahman [Stillinger and Rahman 1974]. Here, charges are placed on the hydrogen atoms and on two lone-pair sites on the oxygen. The electrostatic contribution is modulated so that for oxygen–oxygen distances below 2.016 Å it is zero and for distances greater than 3.1287 Å it takes its full value. Between these two distances the electrostatic contribution is modulated using a function that smoothly varies from 0.0 at the shorter distance to 1.0 at the longer distance (see Section 6.7.3).

The experimentally determined dipole moment of a water molecule in the gas phase is 1.85 D The dipole moment of an individual water molecule calculated with any of these simple models is significantly higher; for example, the SPC dipole moment is 2.27 D and that for TIP4P is 2.18 D. These values are much closer to the effective dipole moment of liquid water, which is approximately 2.6 D. These models are thus all effective pairwise models. The simple water models are usually parametrised by calculating various properties using molecular dynamics or Monte Carlo simulations and then modifying the



Fig 4.40. Some 'simple' water models (Table 4.3) [Jorgensen et al 1983].

parameters until the desired level of agreement between experiment and theory is achieved. Thermodynamic and structural properties are usually used in the parametrisation, such as the density, radial distribution function, enthalpy of vaporisation, heat capacity, diffusion coefficient and dielectric constant.^{*} It is found that some properties such as the density and the enthalpy of vaporisation are predicted rather well by all of the models, but there is significant variation in the values for other properties such as the dielectric constant [Jorgensen *et al.* 1983]. When comparing the different models, it is also important to take account of the computational effort each requires. Thus, nine site–site distances must be calculated for each water dimer using a three-site model; ten are required for a four-site model, and seventeen for the ST2 model.

The use of a rigid model for water is obviously an approximation, and it means that some properties cannot be determined at all. For example, only when internal flexibility is included can the vibrational spectrum be calculated and compared with experiment. Flexibility is most easily incorporated by 'grafting' bond-stretching and angle-bending terms onto the potential function for a rigid model. Such an approach needs to be done with care. For example, Ferguson has developed a flexible model for water that is based upon the SPC model [Ferguson 1995]. The partial charges and van der Waals parameters in this model were slightly different from those in the rigid model, and flexibility was achieved using cubic and harmonic bond-stretching terms and a harmonic angle-bending term. The calculated values compared well with experimental results for a wide range of thermodynamic and structural properties, including the dielectric constant and selfdiffusion coefficient.



Enable | Enlighten | Enrich (Deemed to be University) (Under Section 3 of UGC Act 1956)

KARPAGAM ACADEMY OF HIGHER EDUCATION (Deemed to be University Under Section 3 of UGC Act 1956) COIMBATORE-21

DEPARTMENT OF CHEMISTRY (For the candidates admitted from 2019 & onwards) 19CHP105-C MOLECULAR MODELLING & DRUG DESIGN Multiple Choice Questions for Unit II

| S. No | Question | Option 1 | Option 2 | Option 3 | Option 4 | Answer |
|----------|---|----------------------------|---------------------------|--------------------|-------------------------------|-------------------------|
| | Unit-II | | | | | |
| 1 | Force field method also known as | Molecular mechanics | Molecular graphics | Electronic motion | Molecular energy | Molecular mechanics |
| 2 | The number of difference of valence angle in propane is | 12 | 18 | 16 | 27 | 18 |
| 3 | The number of torsional terms in propane is | 6 | 12 | 18 | 27 | 18 |
| 4 | The bond stretching and angle bending terms are often regarded as | Hard degrees of freedom | Simple degrees of freedom | Degrees of freedom | Complex degrees of freedom | Hard degrees of freedom |

| 5 | The number of torsional terms in benzene is | 6 | 18 | 24 | 30 | 24 |
|----|--|---|---|---------------------------------------|---------------------------------------|---------------------------------------|
| 6 | The electrostatic interaction is calculated by using | Coulombs law | Faraday's law | Morse potential curve | Hookes law | . Coulombs law |
| 7 | For a distribution of charges the dipole moment is given by | $\mu = \pi q_i r_i$ | $\boldsymbol{\mu} = \sum \mathbf{q}_i \mathbf{r}_i$ | $\mu = 2 q_{\mu_1}^2$ | $\mu = \sum q_{i}/r_{i}$ | $\mu = \sum q_i r_i$ |
| 8 | The Lennard Jones potential is characterized by an attractive part that varies as | r ⁻² | r ⁻⁴ | r-6 | r ⁻¹² | r ⁻⁶ |
| 9 | Molecular mechanics will not work without | Born-oppenheimer approximation | Coulomb potential | Molecular geometry | Lenard Jones potential | Born- oppenheimer approximation |
| 10 | In propane, the number of H-C-C-H torsion is | 6 | 12 | 18 | 14 | 12 |
| 11 | In propane, the number of H-C-C-C torsion is | 6 | 12 | 18 | 14 | 6 |
| 12 | The Morse potential is | $v(l) = D_e \{ 1 + exp[-a(l-l_0)] \}^2$ | $v(l) = D_e \{ 1 - exp[-a(l+l_0)] \}^2$ | $v(l) = D_e \{1 - exp[-a(l-l_0)]\}^2$ | $v(l) = D_e \{ 1 - exp[-a(l-l_0)] \}$ | $v(l) = D_e \{1 - exp[-a(l-l_0)]\}^2$ |

| 13 | The Hookes law is given by | $v(l)=k(l-l_0)^2$ | $v(l)=k/2(l+l_0)^2$ | $v(l)=k/2(l-l_0)^2$ | v(l)=k/2(l-l ₀) | $v(l)=k/2(l-l_0)^2$ |
|----|--|--|------------------------------------|---|----------------------------------|---|
| 14 | The Hookes law for the deviation of angles from their reference value is given by | $\mathbf{v}(\theta) = \mathbf{k}(\theta - \theta_0)^2$ | $v(\theta)=k/2(\theta+\theta_0)^2$ | $\mathbf{v}(\mathbf{\theta}) = \mathbf{k}/2(\mathbf{\theta} \cdot \mathbf{\theta}_0)^2$ | $v(\theta)=k/2(\theta-\theta_0)$ | $\mathbf{v}(\boldsymbol{\theta}) = \mathbf{k}/2(\boldsymbol{\theta} \cdot \boldsymbol{\theta}_0)^2$ |
| 15 | The number of torsional term in benzene is | 6 | 12 | 18 | 24 | 24 |
| 16 | The torsional potential is always expressed in | Cosine series | Sin series | Tan series | Cot series | Cosine series |
| 17 | The number of bonds present in propane is | 3 | 6 | 10 | 12 | 10 |
| 18 | The number of C-C bonds present in propane is | 2 | 3 | 4 | 5 | 2 |
| 19 | The number of C-H bonds present in propane is | 2 | 4 | 6 | 8 | 8 |
| 20 | The C-C bonds in propane is | Unsymmetricaly equivalent | Symmetricaly equivalent | Symmetricaly unequivalent | UnSymmetricaly unequivalent | Symmetricaly equivalent |
| 21 | The C-H bond of propane falls in | One classes | Two classes | Three classes | Four classes | Two classes |
| 22 | The number of | 6 | 10 | 14 | 18 | 18 |

| | different | | | | | |
|----|------------------|-------------------|---------------|-------------------|------------------|---------------|
| | valence angle in | | | | | |
| | propane is | | | | | |
| 23 | The number of | 1 | 2 | 3 | 4 | 1 |
| | C-C-C angle in | | | | | |
| | propane is | | | | | |
| 24 | The number of | 3 | 4 | 7 | 10 | 10 |
| | C-C-H angle in | | | | | |
| | propane is | | | | | |
| 25 | The number of | 3 | 4 | 7 | 10 | 7 |
| | H-C-H angle in | | | | | |
| | propane is | | | | | |
| 26 | How many | 9 | 18 | 27 | 36 | 27 |
| | nonbonded | | | | | |
| | terms are | | | | | |
| | present in | | | | | |
| | propane? | | | | | |
| 27 | How many | 7 | 14 | 21 | 28 | 21 |
| | nonbonded | | | | | |
| | terms are | | | | | |
| | present in | | | | | |
| | propane as H-H | | | | | |
| | interactions? | | | | | |
| 28 | How many | 2 | 6 | 10 | 14 | 6 |
| | nonbonded | | | | | |
| | terms are | | | | | |
| | present in | | | | | |
| | propane as H-C | | | | | |
| | interactions? | | | | | |
| 29 | Force field | Single entity | Double entity | Triple entity | Multi entity | Single entity |
| | should be | | | | | |
| | considered as a | | | | | |
| 30 | The force field | Atomic properties | Structural | Dipole properties | Ionic properties | Structural |

| | used in molecular modeling are primarily designed to reproduce | | properties | | | properties |
|----|---|---------------------|------------------------|-------------------|-----------------|------------------------|
| 31 | The force field used in molecular modeling are also used to predict other properties such as | NMR spectra | IR spectra | Molecular spectra | EPR spectra | Molecular spectra |
| 32 | Transferability of functional form and parameter is an important feature of | Electrostatic field | Vanderwaals field | Nonbonded field | Force field | Force field |
| 33 | There is no correct form of force field in | Molecular dynamics | Molecular mechanics | Molecular energy | Surface energy | Molecular mechanics |
| 34 | Deep analysis of which one shows that force field is empirical? | Molecular dynamics | Molecular mechanics | Molecular energy | Surface energy | Molecular mechanics |
| 35 | Input for quantum mechanical calculation it is necessary to | Atomic number | Mass number | Neutron number | Positron number | Atomic number |

| | specify the nucleus | | | | | |
|----|--|----------------------|-------------|--------------------------|-----------------|-------------------------|
| 36 | Input for quantum mechanical calculation it is necessary to specify the system | Decay rate | Geometry | Binding energy | amu | Geometry |
| 37 | sp ³ hybridised carbon adopts a | Trigonal | Linear | Tetrahedral geometry | Square planar | Tetrahedral geometry |
| 38 | sp ² hybridised carbon adopts a | Trigonal | Linear | Tetrahedral geometry | Square planar | Trigonal |
| 39 | sp hybridised carbon adopts a | Trigonal | Linear | Tetrahedral geometry | Square planar | Linear |
| 40 | Reference angle (θ_0) for Tetrahedral carbon atom is near to | 90° | 120° | 109.5° | 180° | 109.5° |
| 41 | Reference angle (θ_0) for Trigonal carbon atom is near to | 90° | 120° | 109.5° | 180° | 120° |
| 42 | The carbon atom at the junction between 5 and 6-membered ring is assigned different from that of carbon atom in isolated | AMBER Force field | Force field | Ab initio calculation | MM2 calculation | AMBER Force field |

| | ring by | | | | | |
|----|--|-------------------------|----------------------|--------------------------------|----------------------|-------------------------|
| 43 | Frequency of | ω ² =kxμ | $\omega^2 = k/\mu$ | $\omega^2 = k/\mu^2$ | $\omega^2 = k^2/\mu$ | ω²=k/μ |
| | bond vibration | | | | | |
| | (ω) is related to | | | | | |
| | the stretching | | | | | |
| | constant of the | | | | | |
| | bond (k) by | | | | | |
| 44 | Even at absolute | Vibrational motion | Translational motion | Rotational motion | Linear motion | Vibrational |
| | zero, Real | | | | | motion |
| | molecules | | | | | |
| | undergoes | | | | | |
| 45 | A true bond | Harmonic | Not harmonic | Linear | Elliptical | Not harmonic |
| | stretching | | | | | |
| | potential is | | | | | |
| 46 | Real molecules | Vibrational motion | Translational motion | Rotational motion | Linear motion | Vibrational |
| | has Zero point | | | | | motion |
| | energy due to | | | | | |
| 47 | At absolute | No Zero point | Potential energy | Zero point energy | Kinetic energy | Zero point |
| | zero, due to | energy | | | | energy |
| | vibrational | | | | | |
| | motion real | | | | | |
| | molecule has | | | | | |
| 48 | A deviation of | Decrease by | Decrease | Rise by 12kcalmol ⁻ | Increase | Rise by |
| | just 0.2Å from | 12kcalmol ⁻¹ | exponentially | 1 | exponentially | 12kcalmol ⁻¹ |
| | the reference | | | | | |
| | value l_0 with a | | | | | |
| | force constant | | | | | |
| | of 300kcalmol ⁻¹ | | | | | |
| | Å ⁻² would cause | | | | | |
| | the energy of | | | | | |
| | the system to | | | | | |
| 49 | The θ_0 value for | 117.2 | 109.47 | 121.4 | 122.5 | 109.47 |
| | Csp ³ -Csp ³ -Csp ³ | | | | | |

| | is | | | | | |
|----|---|---|--|---|---|---|
| 50 | The θ_0 value for Csp^3 - Csp^3 -H is | 117.2 | 109.47 | 121.4 | 122.5 | 109.47 |
| 51 | The θ_0 value for H-Csp ³ -H is | 117.2 | 109.47 | 121.4 | 122.5 | 109.47 |
| 52 | The θ_0 value for Csp ³ -Csp ² -Csp ³ is | 117.2 | 109.47 | 121.4 | 122.5 | 117.2 |
| 53 | The θ_0 value for $Csp^3-Csp^2=Csp^2$ is | 117.2 | 109.47 | 121.4 | 122.5 | 121.4 |
| 54 | The θ_0 value for Csp^3 - $Csp^2=O$ is | 117.2 | 109.47 | 121.4 | 122.5 | 122.5 |
| 55 | Torsional potential expression is given by | $\mathbf{v}(\boldsymbol{\omega}) = \sum_{n=0}^{N} \mathbf{C}_{n} \mathbf{cos}(\boldsymbol{\omega})^{n}$ | $v(\omega) = \sum_{n=0}^{N} C_n \cos(2\omega)^n$ | $v(\omega) = \sum_{n=0}^{N} C_n \cos(\omega/2)^n$ | $v(\omega) = \sum_{n=0}^{N} C_n \cos(\omega)^{1/n}$ | $\mathbf{v}(\boldsymbol{\omega}) = \sum_{n=0}^{N} \mathbf{C}_n \mathbf{cos}(\boldsymbol{\omega})^n$ |
| 56 | The number of C-C-C-C torsional term in benzene is | 6 | 12 | 18 | 27 | 6 |
| 57 | The Lennard - Jones 12-6 function takes the form for the interaction between two atom is given by | $v(r) = 4\varepsilon [(\sigma/r)^6 - (\sigma/r)^{12}]$ | $v(\mathbf{r})=4\varepsilon[(\sigma/\mathbf{r})^{12}+(\sigma/\mathbf{r})^{6}]$ | ν(r)=4ε[(σ/r) ¹² - (σ/r) ⁶] | $v(r) = [(\sigma/r)^{12} - (\sigma/r)^6]$ | $v(\mathbf{r})=4\varepsilon[(\sigma/\mathbf{r})^{12}-(\sigma/\mathbf{r})^{6}]$ |
| 58 | The general form of Lennard-Jones original potential is | k=n/(n+m)(n/m) ^{(m/n-} m) | k=n/(n-m)(n/m) ^{(m/n-} ^{m)} | k=n/(n-m)(n/m) ^(m/n) | k=n/(m-n)(n/m) ^{(m/n-} ^{m)} | k=n/(n- m)(n/m) ^(m/n-m) |

| 59 | The 10-12 | $v(r) = A/r^{12} - C/r^{10}$ | $v(r) = A/r^{12} + C/r^{10}$ | $v(r) = A/r^{10} - C/r^{12}$ | $1/v(r) = A/r^{12} - C/r^{10}$ | $v(r) = A/r^{12} - C/r^{10}$ |
|----|------------------|--------------------------------|------------------------------------|------------------------------|------------------------------------|--|
| | Lennard-Jones | | | | | |
| | potential is | | | | | |
| | given by | | | | | |
| 60 | The GRID | $v_{(HB)} = (C/d^6 + D/d^4)co$ | $\nu_{(HB)} = (C/d^6 -$ | $v_{(HB)} = (C/d^4 -$ | $1/v_{(HB)} = (C/d^{6}-$ | ν _(HB) =(C/d ⁶ - |
| | program that is | s ^m θ | D/d^4)cos ^m θ | $D/d^6)\cos^m\theta$ | D/d^4)cos ^m θ | D/d ⁴)cos ^m θ |
| | used for finding | | , | , | , | |
| | energetically | | | | | |
| | favourable | | | | | |
| | regions in | | | | | |
| | protein binding | | | | | |
| | sites uses a | | | | | |
| | direction | | | | | |
| | dependent 6-4 | | | | | |
| | function, which | | | | | |
| | is given by | | | | | |



Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: III

Batch: 2019 -2021

UNIT III

Basics of molecular modelling, methods, steps involved in MM, selection of target and template, homology modelling, refinement and validation-SAVES server, the critical assessment of protein structure prediction (CASP), superposition of proteins using different tools, RMSD, presentation of protein conformations, hydrophobicity factor, shape complementary.





Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: III

Batch: 2019 -2021

Basics of molecular modeling

Molecular modeling is a powerful tool for drug design and molecular docking

Molecular modeling has become a valuable and essential tool to medicinal chemists in the drug design process. Molecular modeling describes the generation, manipulation or representation of three-dimensional structures of molecules and associated physico-chemical properties. It involves a range of computerized techniques based on theoretical chemistry methods and experimental data to predict molecular and biological properties. Depending on the context and the rigor, the subject is often referred to as 'molecular graphics', 'molecular visualizations', 'computational chemistry', or 'computational quantum chemistry'. The molecular modeling techniques are derived from the concepts of molecular orbitals of Huckel, Mullikan and classical mechanical programs' of Westheimer, Wiberg and Boyd.

What is Modeling and Molecular Modeling?

Modeling is a tool for doing chemistry. Models are central for understanding of chemistry. Molecular modeling allows us to do and teach chemistry better by providing better tools for investigating, interpreting, explaining and discovering new phenomena. Like experimental chemistry, it is a skill-demanding science and must be learnt by doing and not just reading. Molecular modeling is easy to perform with currently available software, but the difficulty lies in getting the right model and proper interpretation.

Computational chemistry is comprised of a theoretical (or structural) modeling part, known as *molecular modeling*, and a modeling of processes (or experimentations) known as *molecular simulation*.

Depending upon the level of theory that we observe in a computation, the following methods have been identified.

Ab Initio Calculations, Semiempirical Calculations, Modeling the Solid State, Molecular Mechanics, Molecular Simulation, Statistical Mechanics, Thermodynamics, Structure-Property Relationships, Symbolic Calculations, Artificial Intelligence, The Design of a Computational Research Program, Visualization

Steps involved in Molecular Modeling (MM).

1) **Topological properties:** description of the covalent connectivity of the molecules to be modeled





Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: III

Batch: 2019 -2021



2) **Structural properties:** the starting conformation of the molecule, provided by an X-ray structure, NMR data or a theoretical model

3) **Energetical properties:** a force field describing the force acting on each atom of the molecules.



4) Thermodynamical properties: a sampling algorithm that generates the thermodynamical ensemble that matchese experimental conditions for the system, e.g. *N*,*V*,*T*, *N*,*P*,*T*, ...



Selection of target and template

The initial step in comparative modeling is to assign the likely fold of the target sequence. Template identification can be achieved using any one of the many programs that scan sequence and structure databases, such as Protein Data Bank (PDB) structural classification of



Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: III

Batch: 2019 -2021

proteins (SCOP), distance-matrix alignment (DALI), and Class, Architecture, Topology, and Homology (CATH).Template search methods can be categorized into three different classes:

First, pairwise comparison methods, which include the popular programs Basic Local Alignment Search Tool (BLAST) and FASTA, align the target sequence with all the sequences in the database of known structures. The performance and efficiency of this class of methods has been studied extensively.

Second, sequenceprofile methods, such as position specific iterative (PSI)-BLAST and HMMER(http://hmmer.wustl.edu), rely on profiles derived from multiple sequence alignments to increase the sensitivity and accuracy of the template search. The profile enhances the sensitivity of the search. Profiles are also utilized by the intermediate sequence search algorithms that establish a homology between two remotely related sequences through an intermediary sequence.

Third, the so-called threading methods use a combination of sequence and structure considerations to detect similarities between sequences and structures.

In these methods, the target sequence is threaded through a library of 3-D profiles or folds, and each threading is assessed based on a certain scoring function. Commonly used methods and servers in this category include Superfamily and GenThreader. The threading methods are more effective in detecting homology at low sequence similarity than the methods relying on sequence information alone. The three different classes of methods are best suited for identifying templates in different regimes of the sequence-identity spectrum. The pairwise sequence comparison methods are the least sensitive and are best used to detect close homologs. The profile-based methods are usually capable of recognizing homologs sharing only approx 25% sequence identity. Threading methods can sometimes recognize common folds even in the absence of any statistically significant sequence similarity. Because most of the fold assignment methods involve sequence alignment, some of them are discussed in more detail in the following section about sequence-structure alignment. While a correct fold assignment can be used to build a useful model, an incorrect fold assignment renders the resulting model useless. Thus, when using a fold-recognition method, it is crucial to be aware of the accuracy of the method. In an assessment of different fold-recognition methods, the best method detected 75% of the closest structures correctly for a set of sequences related at the "family" level in the SCOP database. However, at the superfamily and fold levels, the accuracy dropped to 29 and 15%, respectively.

Once a list of all related protein structures is obtained, templates that are appropriate for the given modeling problem have to be selected. Usually, a higher overall sequence identity between the target and the template sequence yields a better template. Several other factors should also be considered in selecting templates.

Constructing a phylogenetic tree for the whole family can frequently help in selecting



Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: III

Batch: 2019 -2021

a template from the subfamily that is closest in structure to the target sequence.

Databases of structure-based phylogenies, such as the database of Phylogeny and Alignment (PALI), are useful in making a distinction between the sequence and structure similarity, which can be a key consideration for template identification.

Accuracy of the template structure is another important factor in template selection. The resolution and the R-factor of a crystallographic structure and the number of restraints per residue for an NMR structure are indicative of structure accuracy.

It is also crucial to compare the environment of the template to the required environment for the model. The term *environment* is used in a broad sense and includes all factors that determine protein structure, except its sequence (e.g., solvent, pH, ligands, and quaternary interactions). For example, if the objective of the model-building exercise is to dock ligands in the model, it is usually best to use a template that is itself bound to an identical or similar ligand. In general, prior biological information about the target sequence can be valuable in identifying an appropriate template.

Prioritization of the criteria for template selection depends on the purpose of the comparative model. For instance, if a protein–ligand model is to be constructed, the choice of the template that contains a similar ligand is probably more important than the resolution of the template. On the other hand, if a model is to be used to analyze the geometry of the active site of an enzyme, it is preferable to use a high-resolution template.

It is not necessary to select only one template. In fact, the use of several templates approximately equidistant from the target sequence generally increases the model accuracy.

Target–Template Alignment

After identifying the template(s), the next crucial step in comparative modeling is to accurately align the target sequence to the template(s). Although most template-recognition methods produce a target-template alignment, there is frequently a need to use a specialized alignment method to realign the sequences because the template-identification step is often optimized to identify distant relationships, sometimes at the expense of alignment accuracy. The sequence-structure alignment is a vital step in the model building process, and an erroneous alignment will almost certainly lead to the construction of an incorrect model.

An alignment between two sequences of residues is usually calculated by optimizing an alignment scoring function. The two common ingredients of the scoring function are a gap penalty function and a matrix of substitution scores for matching every residue in one sequence to every residue in the other sequence. The alignment score is usually a sum of the gap penalties, which depend linearly on the gap lengths, and the pairwise substitution scores, which depend on the matched residue types. The original and still widely used optimization method for sequence alignment is based on dynamic programming. Since its inception, the scoring function and its optimization by dynamic programming have been improved for alignment accuracy and speed, and applied to a variety of alignment problems.

In the next few paragraphs, we examine different methods to obtain substitution score matrices and gap penalties that optimize the accuracy of the output alignments.



Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: III

Batch: 2019 -2021

We examine the use of information from related multiple sequences and structures to enhance alignment accuracy and coverage, especially when target–template sequence identity decreases below 30%.

Homology Modelling

Homology modeling, also known as comparative modeling of protein is the technique which allows to construct an unknown *atomic-resolution model* of the "target" protein from:

1. its amino acid sequence and

2. an experimental 3D structure of a related homologous protein (the "template").

Homology modeling relies on the identification of one or more known protein structures likely to resemble the structure of the query sequence, and on the production of an alignment that maps residues in the query sequence to residues in the template sequence. It has been shown that protein structures are more conserved than protein sequences amongst homologues, but sequences falling below a 20% sequence identity can have very different structure. Evolutionarily related proteins have similar sequences and naturally occurring homologous proteins have similar protein structure. It has been shown that three dimensional protein structure is evolutionarily more conserved than expected due to sequence conservation. The sequence alignment and template structure are then used to produce a structural model of the

target. Because protein structures are more conserved than DNA sequences, detectable levels of sequence similarity usually imply significant structural similarity.



Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: III

Batch: 2019 -2021



Example

We will construct a 3D-structural model of the **Human** (Homo sapiens) variant of the **Rhodopsin** which has not been experimentally resolved, yet. Rhodopsin, also known as visual purple, is a biological pigment of the retina that is responsible for both the formation of the photoreceptor cells and the first events in the perception of light. Rhodopsins belong to the G-protein coupled receptor family and are extremely sensitive to light, enabling vision in low-light conditions. Exposed to light, the pigment immediately photobleaches, and it takes about 30 minutes to regenerate fully in humans.

Structurally, rhodopsin consists of the protein moiety *opsin* and a reversibly covalently bound cofactor, *retinal*. Opsin, a bundle of seven transmembrane helices connected to each other by protein loops, binds retinal (a photoreactive chromophore), which is located in a central pocket on the seventh helix at a lysine residue. Retinal lies horizontally with relation to the membrane. Each outer segment disc contains thousands of visual pigment molecules. About half the opsin is within the lipid bilayer. Retinal is produced in the retina from Vitamin A, from



Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: III

Batch: 2019 -2021

dietary betacarotene. Isomerization of 11-cis-retinal into all-trans-retinal by light induces a conformational change (bleaching) in opsin continuing with metarhodopsin II, which activates the associated G protein transducin and triggers a second messenger cascade.



1 - Getting the sequence

The first step in our procedure is to get the sequence of aminoacids for the human rhodopsin.

We use the famous UniProt database (UNIversal PROTein) reachable at the website: http://www.uniprot.org/

Insert in the "Query" field the string: Human rhodopsin

The field "Search in" can be left at the default: Protein Knowledgebase

Insert in the "Query" field the string: Human rhodopsin

The field "Search in" can be left at the default: Protein Knowledgebase (UniProtKB)

which identify the database to employ for the research.

The server will produce several output which has some relation with the content requested. However, only one entry (P08100) is the one we are interested in.



Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: III

Batch: 2019 -2021

It is always a good idea to read as more much information as possible from the database details about the file selected.

| UniProt | | | | | | | | Downloads · Contact · Documentation/H |
|--|--|---|----------------------------|--|--|--|---------|---------------------------------------|
| Search | Blast | Align | Retrieve | ID Ma | pping | | | |
| Search in | | Query | | | | | | |
| Protein Knowle | dgebase (UniProtKB) |) | | Search | Advanced Search » Clear | | | |
| ELCOME | | | | | NEWS | | | |
| ne mission of Ur omprehensive, h nd functional info /hat we provid | iProt is to provide the so igh-quality and freely acc rrmation. de | cientific community cessible resource of | with a protein sequence | | UniProt release 2011_12 - Between Charybdis and Cilia I (and DMDM | Dec 14, 2011 Cross-references to PATR | IIC | |
| UniProtKB Protein knowledgebase, consists of two sections: | | | | Statistics for UniProtKB: Swiss-Prot TrEMBL Forthcoming changes News archives Follow @uniprot 145 followers | | | | |
| JniRef | Sequence clusters, use searches. | ed to speed up sequ | ence similarity | | Management of Pro- | | | |
| JniParc | Sequence archive, use their identifiers. | d to keep track of s | equences and | | The second second | | | |
| Supporting data | Literature citations, taxe locations, cross-referen | onomy, keywords, s need databases and | ubcellular more. | | Control Information (Information Control Information Control Informatio Control Information Control I | 1967 | | |
| etting started | l | | | | As a second of the set of the second of the | | | |
| Text search Sequence Sequence | h similarity searches (BLA alionments | IST) | | | Learn how to make best use of | the tools and data on this | i site. | |
| Batch retri | eval | | | | PROTEIN SPOTLIC | ант | | |
| ick on t | the link of t | the select | ed entry | and d | ownload the co | rresponding | file i | n the |
| | | | | | | c a | | · c |
| sta forr | nat (using t | the last y | ellow but | ton at | the right top o | t the page) | which | is one of |

the most common format used by the scientific community for protein sequences.

If you open the file with an editor: edit

P08100.fasta we should see this text:

```
>sp|P08100|OPSD_HUMAN Rhodopsin OS=Homo sapiens GN=RHO PE=1 SV=1
MNGTEGPNFY VPFSNATGV VRSPFEYPQYYLAEPWQFSMLAAYMFLLIVLGFPINFLTL
YVTVQHKKLRTPLNYILLNLAVADLFMVLGGFTSTLYTSLHGYFVFGPTGCNLEGFFAT
LGGEIALWSLVVLAIERYVVVCKPMSNFRFGENHAIMGVAFTWVMALACAAPPLAGW
SRYIPEGLQCSCGIDYYTLKPEVNNESFVIYMFVVHFTIPMIIIFFCYGQLVFTVKEAAAQ
QQESATTQKAEKEVTRMVIIMVIAFLICWVPYASVAFYIFTHQGSNFGPIFMTIPAFFAKS
AAIYNPVIYIMMNKQFRNCMLTTICCGKNPLGDDEASATVSKTETSQVAPA
```

The first line after the symbol ">" is a comment which specify relevant information about the sequence. The rest of the lines are the aminoacidic sequence in the single letter code:

- G Glycine (Gly)
- P Proline (Pro)
- A Alanine (Ala)
- V Valine (Val)
- L Leucine (Leu)
- I Isoleucine (Ile)



Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: III

Batch: 2019 -2021

- M Methionine (Met)
- C Cysteine (Cys)
- F Phenylalanine (Phe)
- Y Tyrosine (Tyr)
- W Tryptophan (Trp)
- H Histidine (His)
- K Lysine (Lys)
- R Arginine (Arg)
- Q Glutamine (Gln)
- N Asparagine (Asn)
- E Glutamic Acid (Glu)
- D Aspartic Acid (Asp)
- S Serine (Ser)
- T Threonine (Thr)

2 – Template selection

The second step is to find a protein whose 3D-structure is known and that is as most similar as possible to our target regarding their aminoacid sequence: the procedure is called sequence alignment.

One of the most widely used algorithms for comparing primary biological sequences, such as aminoacid sequences, is **BLAST** (Basic Local Alignment Search Tool). A BLAST search enables a researcher to compare a query sequence with a library or database of sequences, and identify library sequences that resemble the query sequence above a certain threshold. We can do a BLAST search online by exploiting the webserver:

http://blast.ncbi.nlm.nih.gov/Blast.cgi

In particular click on the "protein blast" link to restrict the search on the protein database:





Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: III

Batch: 2019 -2021

| and a second sec | a side | Standard Protein BLAST | |
|--|--|--|---------------------|
| | | Statuard Protein BEAST | |
| to blastp blastk | Edwards Edwards | REASTP renorams asserth protain databases using a protain puery, more | Reset page Bookmark |
| Enter Query 5 | Sequence | and the first second | |
| inter accession | number(s), gl(s), or FASTA sequence(s) 🚱 🤇 🤤 | Query subrange 😧 | |
| | | From | |
| | | To | |
| | | | |
| r, upload file | Choose File) no file selected | | |
| ob Title | Participation of the second second second | Sequence in the fasta format | |
| Allan two or m | Enter a descriptive for your bLAS-I search 😼 | Sequence in the lasta for mat | |
| Augu tao or m | | | |
| Choose Searc | ch Set | | |
| atabase | Non-redundant protein sequences (nr) 🛟 🔒 | | |
| ptional | Ernar organism name or id-completions will be suggested] C Exclude | | |
| | Enter organism common name, binomial, or tax id. Only 20 top taxa will | be shown. 🔮 | |
| ixclude | □ Models (XM/XP) □ Uncultured/environmental sample sequences | Database selection | |
| intrez Query | | | |
| | Enter an Entrez query to limit search 🥹 | | |
| ptional | | | |
| Program Sele | ction | | |
| Program Sele Jgorithm | e blasto (protein-protein BLAST) | | |
| Program Sele Jgorithm | e blastp (protein-protein BLAST) O PSI-BLAST (Position-Specific Iterated BLAST) | | |
| Program Sele Jgorithm | e blastp (protein-protein BLAST) O PSI-BLAST (Position-Specific Iterated BLAST) O PH-BLAST (Pattern Hit Initiated BLAST) | | |
| Program Sele Igorithm | eblastp (protein-protein BLAST) ○ PSI-BLAST (Poelkon-Specific Iterated BLAST) ○ PHI-BLAST (Pattern Hit Initiated BLAST) Choose a BLAST algorithm € | | |
| Program Sele Igorithm | etion e blastp (protein-protein BLAST) ○ PSI-BLAST (Poetion-Specific Iterated BLAST) ○ PHI-BLAST (Pattern Hit Initiated BLAST) Choose a BLAST algorithm € | | |
| Program Sele Igorithm BLAST | Ction e blastp (protein-protein BLAST) p PSi-BLAST (Position-Specific Iterated BLAST) p PHI-BLAST (Pattern Hit Initiated BLAST) Choose a BLAST algorithm € Search database Non-redundant protein sequences (nr) using I | Blastp (protein-protein BLAST) | |

Button to search

- Upload your fasta file or copy and paste its content in the wide field at the top of the page.

- Choose the "Protein Data Bank proteins(pdb)" as database since it is the largest protein database which contains only experimentally resolved structures (in contrast to published models).

- Press the BLAST button to start the search.

After some seconds the server will output the result as a list of 3D protein structures ordered according to their "sequence identity percentage" with the target sequence. In particular, the sequence similarity of each line is summarized by the E value (Expected value): closer to zero higher level of sequence similarity.

The quality of the homology model is dependent on the quality of the sequence alignment and template structure. The approach can be complicated by the presence of alignment gaps (commonly called indels) that indicate a structural region present in the target but not in the template, and by structure gaps in the template that arise from poor resolution in the experimental procedure (usually X-ray crystallography) used to solve the structure. Model quality declines with decreasing sequence identity; a typical model has $\sim 1-2$ Å root mean square deviation between the matched C α atoms at 70% sequence identity but only 2–4 Å agreement at 25% sequence identity.

However, the errors are significantly higher in the loop regions, where the amino acid sequences of the target and template proteins may be completely different. *As a rule of thumb we should never use template with an E value larger than 1.*

In our case all the first founded structures have a sequence identity of 93% and an E


Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: III

Batch: 2019 -2021

value practically equal to zero, so anyone is a possible good candidate. However, since this step is crucial, several checks are mandatory before selecting the structure to be used as the template for the homology modeling:

Is an X-ray crystallography structure (NMR structure are usually much less resolved)?

• All the atoms are resolved in the selected structure?

• Is the chosen structure the best resolved one (typical good resolution is smaller than 2Å for membrane protein as rhodopsin and 1Å for the other proteins) among the structures with the same E value?

• ...

This implies a careful inspection of each candidate structure with a visualization program as VMD.

Note that inside the webpage associated to any BLAST result entry you can ask for the list of all the structures corresponding to the same sequence. This way for each structure you can identify the best resolved one, etc. To do so, click on the link "Identical Proteins". However, with this procedure also results from other databases are reported. Therefore, you should take into account only the experimentally resolved 3D structures as opposed to models.

Moreover, you should always read the articles associated to the structures (retrievable from the database) to understand all the conditions and the limitations related with them. Finally, we suggest also to try another search tool for cross-checking. A suggestion could be the HHsearch server: http://toolkit.tuebingen.mpg.de/hhpred

When the template structure is identified, download the corresponding pdb file in your folder or through the BLAST website or searching it on the PDB website.

3 – Sequence alignment

To create the model (with the procedure in the next chapter) we need a sequence alignment file (.aln file) between our target and the selected template sequence. To this aim, we will use another online server, the multiple sequence alignment ClustalW2: http://www.ebi.ac.uk/Tools/msa/clustalW2

which needs the two sequences in the fasta format: the sequence for the selected template can be obtained directly from the BLAST webpage or from the PDB website.

Insert the two sequences one after the other in the top box and click on the "Submit" button: all the default settings usually fulfill the needs of the most queries. You can also upload the sequences as text files.



Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: III

Batch: 2019 -2021



The result page will provide the sequence alignment.

When viewing your results, these are the consensus symbols used by ClustalW2: a. "*" means that the residues (or nucleotides in case of DNA sequence alignment) in that column are identical in all sequences in the alignment.

b. ":" means that conserved substitutions have been observed (different aminoacids in the sequences' position but belonging to the same type).

c. "." means that semi-conserved substitutions are observed (different aminoacids in the sequences' position which are somewhat similar).

If you would like to see your results in color, push the button that displays "Show Colors". Click Hide Colors to get rid of color. A table for the color code is shown

| below. | | |
|----------|---------|--|
| Residue | Color | Property |
| AVFPMILW | RED | Small (small+ hydrophobic (incl.aromatic -Y)) |
| DE | BLUE | Acidic |
| RK | MAGENTA | Basic - H |
| STYHCNGQ | GREEN | Hydroxyl + sulfhydryl + |





Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: III

Batch: 2019 -2021

| | | amine + G |
|--------|------|----------------------------------|
| Others | Grey | Unusual amino/imino acids etc |

From this webpage you can also download the Alignment File (aln file) needed for the next step.



4 – Building the model

Next step is to build the 3D structure of our target. To this aim we will use the SWISSMODEL website: http://swissmodel.expasy.org/

On the main page select the "Alignment Mode" link:

Unit: III



Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Batch: 2019 -2021

| SiB | BIOZENTRUM Universität Basel The Gester for Molecular (Jie Sciences | 1 | SWISS-M | ODEL | Works | pace | Repository | Documentation |
|-----------------------------|---|--------------------|------------------|-----------|-----------|-------------------|-------------|-----------------|
| [myWorkspa | ace] | | | | | | | [login] |
| SwissMo | del Alignment Moo | de 🥹 | | | Туре уои | email and | project tit | le |
| Email: Project Title | | + | | | Choose | e <u>ClustalW</u> | format | |
| Alignment Ir Cut & paste | nput Format: FASTA | Alignment: | | | _ | Upload or | paste the | <u>aln</u> file |
| Or upload an | n alignment file Choose | File) no file sele | cted | Submit th | e query | | | |
| nsert you | r email address | s and a pro | ject title on th | ne appro | priate bo | xes. | | |

- Choose the "ClustalW" format.
- Copy and paste or upload the .aln file from the ClustalW2 website.
- Submit the query

The server will analyze the file and will prompt you with the name of the two sequences and the request to insert the PDB code and the chain of the template one (PDB code:1U19; chain: A). Then, the server will show up the sequence alignment for further checking. If you press the "Submit alignment" button an email will send to your address with the instructions to access to the .pdb file of the final model.

SWISS-MODEL use a simple mapping of the coordinates of the template to create the model and very basic algorithms to define the positions of the atoms not present in the template structure. This means that the model should be refined with standard molecular dynamics techniques (inspection of the protonation states, energy minimization, water solvation and equilibration, etc) if we want to use the structure for dynamical studies.

The result webpage of SWISS-MODEL will provide several information to evaluate the model obtained:



Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: III

| 1 | 1 MA 2 / 12 | | 348 | |
|---|--|--|--|----------------------------------|
| | | | | Print/Save this page as |
| Nodel Summary | • | | | |
| 1 | | Model info Modelled r Quality in QMEAN Z- | rmation: esidue range: 1 to 348 formation: [details] Score: -6.27 | |
| gs: [Templates] • | Alignment]v [Modelling]v | Warning: proteins or reach such | Low QMEAN Z-scores! Or models of poor quality a n low scores. | inly membrane ine expected to |
| splay model: as wnload model: a: Global Model Qua | pdb]* - as [DeepView project]* - in [As s [pdb]± - as [Deepview project]± - as [lity Estimation : [+/-] | text/ewer]* text]± | | |
| | ores: 0 | | Local scores | |
| MEAN4 global so | ores. | | | |
| MEAN4 global so | Estimated absolute model quality @ | Score components | Coloring by residue error 😨 | Residue error plot @ |
| MEANscore4 @ | Estimated absolute model quality | Score components @ | Coloring by residue error | Residue error plot |

Unit: III



Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

R

Batch: 2019 -2021



```
Swiss Institute of Bioinformatics | About SWISS-MODEL | Help | Terms of ABack to the Top
```

SWISS-MODEL server has an automated procedure to get a model of a particular sequence. This procedure is effective in the case of high level of similarity (as in our case) between template and target but it can be not satisfactory otherwise. As a rule of thumb, automated sequence alignments are sufficiently reliable when target and template share more than 50% percent of sequence identity. This submission requires only the aminoacid sequence or the UniProt accession code of the target protein as input data. The pipeline will automatically



Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C U

Unit: III

Batch: 2019 -2021

select suitable templates based on a BLAST E-value limit (which can be adjusted upon submission), experimental quality, bound substrate molecules, or different conformational states of the template.

Go to the initial webpage and select "Automated Mode".

Insert the sequence of the human rhodopsin in fasta format and submit the request.

Compare the model obtained with the one got previously with "Alignment Mode".

3. Other online servers allow building models with some automated procedure.

One of this webserver is HHpred:

http://toolkit.tuebingen.mpg.de/hhpred

Use this website to get another structural prediction of the human rhodopsin and compare it with the model obtain with the manual procedure and the automated one in SWISS-MODEL.

5 – Model evaluation

Evaluation of model quality is a foundamental step in homology modeling. While the performance of the alignment and automated SWISS-MODEL pipeline has been evaluated extensively and updates are benchmarked carefully, the quality of individual models can vary significantly.

Therefore, SWISS-MODEL result webpage provides several tests to this aim. Graphical plots of

- Anolea mean force potential
- GROMOS empirical force field energy
- QMEAN

are provided to enable you to estimate the *local quality* of the predicted structure. In order to be able to rank alternative models of the same target protein, pseudo energies for the entire model as calculated by QMEAN and DFIRE are provided as well.

To facilitate the description of template and model structures, DSSP and Promotif can be invoked from the webpage to classify structural features.

Other tools available are Whatcheck and Procheck to analyze the stereo-chemistry of protein models and template structures.

Refer to SWISS-MODEL website (clicking on the red question marks on the result webpage) for a short description of each method and instructions about how to read them.

Moreover, a direct inspection with VMD is always of great worth. Inside VMD other basic checks can be employed, such as an analysis of the Ramachandran plot.

Finally, we suggest to use another online webserver, MolProbity, to have a summary statistics of your model:

http://molprobity.biochem.duke.edu/





Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: III

Batch: 2019 -2021

| PROBITY | Ма | in page | | | |
|---|---|---|--|--|--|
| Main page Evaluate X-ray Evaluate NMR Fix up structure Work with kins View & download files Lab notebook Feedback & bugs Site map | FILE UPLOAD/RETRIEVAL (MORE OPTIONS) Upload PDB/NDB code: Choose File no file selected | the .pdb file type: PDB coords type: PDB coords Upload > | | | |
| Save session Log out | Walk-thrus & tutorials: | Common questions: | | | |
| You are using 0% of your 200 Mb of disk | Evaluate X-ray structure: Typical steps for a published X-ray crystal structure or one still undergoing refinement. | Cite MolProbity: Chen et al. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallographica D66:12-21. | | | |
| | Evaluate NMR structure: Typical steps for a published NMR ensemble or one still undergoing refinement. Fix up structure: Rebuild the model to remove outliers as part of the refinement cycle. Work with kinemages: Create and view interactive 3-D graphics from your web browser. What's new in 3.19: jiffiLoop moved to main page, more options enabled. | and/or Davis et al. (2007) MolProbity: all-atom contacts and structure validation for prote and nucleic acids. Nucleic Acids Research 35:W375-W383. Cite KiNG: Chen et al. (2009) KiNG (Kinemage, Next Generation): A versatile interactive molecular and scientific visualization program. Protein Science 18:240 2409. Installing Java: how to make kinemage graphics work in your browser. Download MolProbity: how can I run a private MolProbity server, or m from the command line? NB: the back button doesn't work inside MolProbity | | | |

Just upload the .pdb file of the model in the main page of the server and run the evaluation.

Options available while running MolProbity are context-sensitive. Before loading a coordinate file, you had two panes:

"File Upload/Retrieval"

MolProbity information

after loading the .pdb file of the model, you also have

• a "Suggested Tools" pane to work on the indicated coordinate file

• a "Recently Generated Results" pane to manage the files in your work area above the original two.

The tools available in the "Suggested Tools" pane are also context sensitive

SUGGESTED TOOLS (ALL TOOLS)

Currently working on: Model.pdb





One could edit the .pdb file, if for instance, there were multiple identical chains in the asymmetric unit uploaded and you wanted to focus on just one.



Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C U:

Unit: III

Batch: 2019 -2021

Since the model obtained from the tutorial procedure has not the hydrogen atoms, we can use the "Add hydrogens" tool to do that:

- Choose the "Add hydrogens" function, and accept the defaults on the next dialog-page
- Click on "Start adding H >", to run the Reduce program allowing it to test Asn/Gln/His flips.
- All suggested flips are for Asn and Gln residues and seem like clear winnners

from the scores (Flip vs Flip-orig column).

• If you want to inspect the suggested flips choose "View in KiNG" for the ModelFH-flipnq.kin file. The KiNG "Views" pull down menu has an entry for each Asn and Gln, with * marking those flipped by Reduce; look at each * view.

Rotate the viewpoint to see the H-bond partner(s), and use the "a" key or the Animate arrows to compare the two flip states (Side chain is colored green in the preferred state).

• Close the KiNG window (button at bottom of page). You now have the choice of rejecting a flip if you don't agree with it.

Click the "Decements II..." button which reasons

• Click the "Regenerate H,..." button, which moves you on to a flip-report page. Note the information presented on this report and then "Continue >" to the MolProbity main page.

The Suggested Tools pane now includes the "Analyze all-atom contacts and geometry" tool as you are now working on a coordinate file with hydrogens. Select this tool, and then "Run..." with the default settings. Then, you will be redirected to the "Analyzed all-atom contacts and geometry for ModelFH.pdb" where you can see the **summary statistics** or choose to view any of the requested model quality assessments.



Viewing ModelFH-multi.table

When finished, you should close this window.

Hint: Use File | Save As... to save a copy of this page.

| All-Atom | Clashscore, all atoms: | 37.54 | 9 th percentile [*] (N=1784, all resolutions) | | |
|----------|---|--------|---|--|--|
| Contacts | Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms. | | | | |
| | Poor rotamers | 4.42% | Goal: <1% | | |
| | Ramachandran outliers | 4.34% | Goal: <0.2% | | |
| | Ramachandran favored | 86.71% | Goal: >98% | | |
| Protein | Cβ deviations >0.25Å | 0 | Goal: 0 | | |
| Geometry | MolProbity score^ | 3.17 | 17 th percentile [*] (N=27675, 0Å - 99Å) | | |
| | Residues with bad bonds: | 0.00% | Goal: 0% | | |
| | Residues with bad angles: | 1.15% | Goal: <0.1% | | |

* 100th percentile is the best among structures of comparable resolution; 0th percentile is the worst.
^ MolProbity score is defined as the following: 0.42574*log(1+clashscore) + 0.32996*log(1+max(0,pctRotOut-1)) + 0.24979*log(1+max(0,100-pctRamaFavored-2)) + 0.5

| # | Res | High B | Clash > 0.4Å | Ramachandran | Rotamer | C _β deviation | Bond lengths. | Bond angles. |
|---|-----|------------|-------------------------------|--|---|--------------------------|--------------------|--------------------|
| | | Avg: 50.57 | Clashscore: 37.54 | Outliers: 15 of 346 | Poor rotamers: 13 of 294 | Outliers: 0 of 326 | Outliers: 0 of 348 | Outliers: 4 of 348 |
| 1 | MET | 50 | 200 | 25 | 94.9% (mmm) chi angles: 303,299.3,289.5 | 0.059Å | | |
| 2 | ASN | 50 | - | Favored (69.18%) General case / -68.4,-27.4 | 77.5% (m-20) chi angles: 291.3,306.1 | 0.026Å | - | - |
| 3 | GLY | 50 | - | Favored (15.98%) Glycine / -126.0,179.2 | | - | | - |
| 4 | THR | 50 | 0.417Å CB with 20 VAL HG11 | Favored (3.56%) General case / -119.2,91.0 | 92.8% (m) chi angles: 300.3 | 0.002Å | | . |
| 5 | GLU | 50 | 121 | Favored (52.94%) General case / -67.1,137.3 | 58.2% (tt0) chi angles: 185.3,183.4,31.3 | 0.031Å | 4 | - |

The summary statistics will give you a simple-to-read evaluation set of parameters to judge the quality of the model.





Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: III

Batch: 2019 -2021

Validation

SAVES is a server for analyzing protein structures for validity and assessing how correct they use the following 5 programs: PROCHECK, WHAT_CHECK, ERRAT, VERIFY_3D, and PROVE.

Procheck

Through PROCHECK the stereochemical quality of a protein structure can be known. The aim of this program is to know that how normal, or how unusual, the geometry of the residues in a given protein structure is, as compared with stereochemical parameters derived from well-refined, high-resolution structures.

Errat

It is a method based on characteristic atomic interaction, use for differentiating between correctly and incorrectly determined regions of protein structures. There are different types of atoms distributed nonrandomly with respect to each other in proteins because of energetic and geometric effects. There will be more randomized distributions of the different atom types if there are errors in model building, which can be distinguished from correct distributions by statistical methods. This method identifies regions of error in protein crystal structures by examining the statistics of pairwise atomic interactions. This method provides a useful tool for model-building and structure verification.

Verify 3D

To test the correctness of a 3D protein model it is necessary to check the compatibility of the model to its own amino acid sequence which is measured by a 3D profile. To check the compatibility of the sequence with the model 3D profile score S is calculated which is the sum of 3D-1D scores of overall residue positions for the amino acid sequence of the protein. A graph is plotted for the compatibility of segments of the sequence with their 3D structures, having sequence number on the x-axis and the average 3D-1D score in a window of 21 windows on the y-axis. The 3D profile score S for amino acid sequence of the model is high for correct 3D protein models.

Model refinement

ProSA: The ProSA program (Protein Structure Analysis) tool has a large user base and is used in the refinement and validation of experimental protein structures, structure prediction and modeling. This program exploits the advantages of interactive web-based applications for the display of scores and energy plots that highlight potential problems spotted in protein structures. The ProSA-web service returns results instantaneously, i.e. the response time is in the order of seconds, even for large molecules.

ModLoop: ModLoop is a web server for automated modeling of loops in protein structures. The user has to provide the atomic coordinates of the protein structure in the Protein Data Bank format, and also the specification of the starting and ending residues of one or more segments to be modeled, containing not more than 20 residues in total. The result is the coordinates of the



Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: III

Batch: 2019 -2021

non-hydrogen atoms in the modeled segments. A user provides the input to the server via a simple web interface, and receives the output by e-mail. The server relies on the loop modeling routine in MODELLER that predicts the loop conformations by satisfaction of spatial restraints, without relying on a database of known protein structures.

Critical Assessment of protein Structure Prediction, or CASP,

It is a community-wide, worldwide experiment for protein structure prediction taking place every two years since 1994. CASP provides research groups with an opportunity to objectively test their structure prediction methods and delivers an independent assessment of the state of the art in protein structure modeling to the research community and software users. Even though the primary goal of CASP is to help advance the methods of identifying protein threedimensional structure from its amino acid sequence, many view the experiment more as a "world championship" in this field of science. More than 100 research groups from all over the world participate in CASP on a regular basis and it is not uncommon for entire groups to suspend their other research for months while they focus on getting their servers ready for the experiment and on performing the detailed predictions.

Selection of target proteins

In order to ensure that no predictor can have prior information about a protein's structure that would put him/her at an advantage, it is important that the experiment be conducted in a double-blind fashion: Neither predictors nor the organizers and assessors know the structures of the target proteins at the time when predictions are made. Targets for structure prediction are either structures soon-to-be solved by X-ray crystallography or NMR spectroscopy, or structures that have just been solved (mainly by one of the structural genomics centers) and are kept on hold by the Protein Data Bank. If the given sequence is found to be related by common descent to a protein sequence of known structure (called a template), comparative protein modeling may be used to predict the tertiary structure. Templates can be found using sequence alignmentmethods (e.g. BLAST or HHsearch) or protein threading methods, which are better in finding distantly related templates. Otherwise, de novo protein structure prediction must be applied (e.g. Rosetta), which is much less reliable but can sometimes yield models with the correct fold (usually, for proteins less than 100-150 amino acids). Truly new folds are becoming quite rare among the targets, making that category smaller than desirable.







Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: III

Batch: 2019 -2021



Cumulative plot of <u> α -carbon</u> accuracy, of all predicted models for target T0398 in CASP8, with the two best models labeled

The primary method of evaluation is a comparison of the predicted model <u> α -carbon</u> positions with those in the target structure. The comparison is shown visually by cumulative plots of distances between pairs of equivalents <u> α -carbon</u> in the alignment of the model and the structure, such as shown in the figure (a perfect model would stay at zero all the way across), and is assigned a numerical score <u>GDT-TS (Global Distance Test — Total Score)</u> describing percentage of well-modeled residues in the model with respect to the target. Free modeling (template-free, or *de novo*) is also evaluated visually by the assessors, since the numerical scores do not work as well for finding loose resemblances in the most difficult cases. High-accuracy template-based predictions were evaluated in CASP7 by whether they worked for molecularreplacement phasing of the target crystal structure with successes followed up later, and by fullmodel (not just <u> α -carbon</u>) model quality and full-model match to the target in CASP8.

Evaluation of the results is carried out in the following prediction categories:

- <u>tertiary structure</u> prediction (all CASPs)
- <u>secondary structure prediction</u> (dropped after CASP5)
- prediction of <u>structure complexes</u> (CASP2 only; a separate experiment <u>CAPRI</u> carries on this subject)
- residue-residue contact prediction (starting CASP4)
- disordered regions prediction (starting CASP5)
- <u>domain</u> boundary prediction (CASP6–CASP8)
- <u>function</u> prediction (starting CASP6)
- model quality assessment (starting CASP7)
- model refinement (starting CASP7)
- high-accuracy template-based prediction (starting CASP7)

Tertiary structure prediction category was further subdivided into





Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C Unit: III

Batch: 2019 -2021

- <u>homology modeling</u>
- fold recognition (also called <u>protein threading</u>; Note, this is incorrect as threading is a method)
- *de novo* structure prediction, now referred to as 'New Fold' as many methods apply evaluation, or scoring, functions that are biased by knowledge of native protein structures, such as an artificial neural network.

Starting with CASP7, categories have been redefined to reflect developments in methods. The 'Template based modeling' category includes all former comparative modeling, homologous fold based models and some analogous fold based models. The 'template free modeling (FM)' category includes models of proteins with previously unseen folds and hard analogous fold based models. Due limited number of template free targets (they are quite rare) in 2011 so called CASP ROLL was introduced. This continuous (rolling) CASP experiment aims at more rigorous evaluation of template free prediction methods through assessment of a larger number of targets outside of the regular CASP prediction season. Unlike LiveBench and EVA, this experiment is in the blind-prediction spirit of CASP, i.e. all the predictions are made on yet unknown structures.

The CASP results are published in special supplement issues of the scientific journal *Proteins*, all of which are accessible through the CASP website. A lead article in each of these supplements describes specifics of the experiment while a closing article evaluates progress in the field.

Superposition of proteins using different tools

Superposition (fitting) of protein structures

For models within structural ensembles, in order to reflect internal motions of proteins, it is necessary to have more or less the same orientation in space. This is most often achieved by fitting the models on each other or on a given model.

1. Least square fit

The superposition of two sets of identical atomic positions can be formulated in a mathematical way as the problem to minimize the RMSD or the weighted RMSD between them. In this case, for two sets of centered atomic positions \mathbf{r}_i and \mathbf{r}'_i such a rotation matrix \mathbf{R} is needed, that minimizes RMSD:

$$RMSD_{min} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} |\mathbf{r_i} - \mathbf{Rr'_i}|^2}$$

Any operation that results in an appropriate \mathbf{R} is generally called as least square fit.



Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C Unit: III

Batch: 2019 -2021

In the following, the not weighted RMSD minimization is discussed.

2. Kabsch algorithm

One of the many and one of the earliest algorithms for achieving a mathematically exact solution for minimum RMSD is described by Wolfgang Kabsch in 1976.

Two sets of centered atomic positions is given: **ri** and **r'i**. If the points are not centered, it is important before the operation to translate them in a way that their average coincides with the origin.

In the following, the 3x3 rotation matrix **R** is needed that rotates the points of **r'i** into **ri**.

As the first step, Nx3 matrixes **P** and **Q** are produced, that contain the coordinates of **ri** and **r'i** as row vectors, respectively.

$$\mathbf{P} = \begin{pmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ \vdots & \vdots & \vdots \\ x_n & y_n & z_n \end{pmatrix}$$

Then, cross-covariance matrix **A** is calculated as $\mathbf{A} = \mathbf{P}^{\mathrm{T}} \mathbf{Q}$

The rotation matrix is given as

$$\mathbf{R} = (\mathbf{A}^{\mathrm{T}} \mathbf{A})^{1/2} \mathbf{A}^{-1}$$

However, A is not guaranteed to have an inverse. To account for all special cases, singular value decomposition (SVD) of matrix A is carried out.

A=VSW^T

This way, the rotation matrix \mathbf{R} is given as

$$\mathbf{R} = \mathbf{W} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & d \end{pmatrix} \mathbf{V}^T$$

Where =sign(det (WV^{T}))

3. Implementations of least square fit



Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: III

Batch: 2019 -2021

Fitting two or more protein models on each other, needless to say, requires an exact definition of the subset of atoms used for the fitting.

In a structural ensemble, the models may be fitted to any of the models (e.g. the first one), or to the mean structure. The difference is significant, since least square fit guarantees minimal RMSD only if it is calculated the same way as the fitting has been carried out. That is, pairwise RMSD, for example, is not guaranteed to be minimal if all the models are fit to the first model. Similarly, if all the models are fit to the first model, the RMSD relative to the mean structure is not guaranteed to be minimal.

In some cases, a rotation matrix \mathbf{R} for minimizing weighted RMSD is searched for. The above described version of the Kabsch algorithm is not capable of yielding such rotation matrix. Instead, one should implement the algorithm described by McLachlan in 1979.

4. Iterative fitting

The mean structure of a protein ensemble clearly changes after fitting the ensemble. This also modifies the overall RMSD relative to the mean structure. To guarantee minimal RMSD, the fitting to the mean structure is carried out in an iterative way, and the mean structure is recalculated in each step of the iteration.

In some other cases, it is necessary to decide which regions of the proteins should be considered in fitting. Flexible regions for example would spoil the purpose of fitting and impede obtaining a low RMSD value. To avoid such ambiguity, local RMSD may be calculated after fitting, and after deciding, which residues have too high RMSD, the may be left out from the following step of the iterative fitting. Such automatic mechanism is not implemented in PDBFitter, since decision about rather flexible than static region requires thorough consideration.

How other programs do it

(i) Chimera

Chimera in general has a fancy graphical user interface, and is capable of performing a large scale of operations, both from command line and on the graphical surface. However, it is often not very customizable, and does not have the advantage of easily modifying the code to obtain a special value.

(ii) Model-model RMSD

Chimera easily yields the same fit pairwise model-model table as described in 1.1., algorithm 4. (Tools, Structure Comparison, Ensemble Match.) It also interactively rotates the models onto each other. Furthermore, it can draw an RMSD map, under the option MD movie.

3.1.2. Local RMSD

Chimera can calculate local RMSD for backbone, Ca and all atoms under Tools, Sequence, and then Structure, Associations, where all the structures have to be associated to the same sequence, and then Headers. The local RMSDs are calculated in a pairwise manner, and yield the same result as in Chapter 1.2, algorithm 1. Note that Chimera includes into the backbone the following atoms: N, C, CA, O.

3.1.3. Fitting



Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: III

Batch: 2019 -2021

Chimera has a sophisticated fitting tool, which is capable of performing sequence and 3D structure alignment. If identical sequences are superimposed, it is also possible to use an iterative method, in which atom pairs with a distance above a given threshold can be dropped. The fitting tool is available under Tools, Structure Comparison, MatchMaker.

The result of the non-iterative fitting can exactly be reproduced through the Kabsch algorithm that was implemented in PDBFitter.

3.2. MOLMOL

The MOLMOL 2.6 was released in 1998. It has a primitive graphical interface and is capable a number of operations and calculations on proteins. However, usage and algorithms are not well documented. Exact calculations sometimes can only be traced by looking at the source code, which, in the present report, is not discussed.

3.2.1. Model-model RMSD

MOLMOL yields the same model-model tables as described in Chapter 1.1., algorithms 1., 3., and 4., for backbone atoms, heavy atoms and the heavy atoms of the sidechains. The values can be reproduced exactly. It has to be noted that for MOLMOL the backbone consists of C, N and CA only. Also, it gives the variance, where PDBFitter the square of the variance.

3.2.2. Local RMSDs

MOLMOL computes the local RMSDs as the average of 3 consecutive RMSDs. The exact values cannot be reproduced, for it is unclear what algorithms the program uses. However, the tendencies are the same. Also, pairwise and relative-to-mean RMSDs can be calculated, which show the same relation as described in Chapter 1.2. MOLMOL also computes two other sets of data, which are called global and local displacements. It is unclear what these values represent.

3.2.3. Fitting

The results of least square fit in MOLMOL are the same as in other programs. The fit ensembles can exactly be reproduced with PDB Fitter.

Root mean square deviation (RMSD)

The root mean square deviation is usually used to compare two or more protein structures and quantify their deviation.

$$RMSD = \sqrt{\frac{1}{n} \sum_{i=1}^{n} |\mathbf{r_i} - \mathbf{r'_i}|^2}$$



Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: III

Batch: 2019 -2021

where

$$|\mathbf{r_{i}} - \mathbf{r'_{i}}|^{2} = (r_{i,x} - r'_{i,x})^{2} + (r_{i,y} - r'_{i,y})^{2} + (r_{i,z} - r'_{i,z})^{2}$$

is the square distance between two identical atomic positions.

Naturally, **ri** and **r'i** may refer to a subset of the entire molecule, such as $C\alpha$ atoms, backbone atoms or heavy atoms. Final RMSD value depends on the exact definition of the atom set **ri**.

It is possible to define the weighted RMSD, in which the atomic distances are multiplied by the weights (e.g. atomic masses) **wi**.

$$RMSD = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \mathbf{w_i} |\mathbf{r_i} - \mathbf{r'_i}|^2}$$

Structural ensembles of proteins usually contain more than two models of the same molecule. For such ensembles, RMSD can be calculated in three ways. Overall RMSD is usually given as the average and the variance of the single model-model RMSDs. (In the following formulae, assume that the ensemble consist of N models, and the RMSD between models k and l is denoted with RMSDk, l.)

1. RMSD can be calculated in a pairwise manner, and be given as an $N\mathbb{N}$ matrix. Such matrix contains as many independent values as the number of elements in the triangular matrix (diagonal not considered, since the RMSD of each model to itself is 0), which is N(N-1)/2. The overall average RMSD is therefore

$$RMSD = \frac{2}{N(N-1)} \sum_{k < l} RMSD_{k,l}$$

2. RMSD can be given for each model relative to any chosen model, for example the first one. In this case





Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: III

Batch: 2019 -2021

$$RMSD = \frac{1}{N} \sum_{k} RMSD_{k,1}$$

3. RMSD of each model can be calculated relative to the mean structure. It has to be kept in mind, that the mean structure no always represents a chemically realistic structure, especially if the ensemble is composed of more than one significantly different conformation. Let the atomic coordinates of the mean structure be $\mathbf{r}\mathbf{i}$ and the $\mathbf{r}\mathbf{i}$ atomic coordinate in model k be $\mathbf{r}\mathbf{i}k$. In this case the mean structure is

$$\bar{\mathbf{r}}_i = \frac{1}{N} \sum_{k=1}^{N} \mathbf{r}_i^k$$

and the overall average RMSD is given by

$$RMSD = \frac{1}{N} \sum_{k} RMSD_{k,mean}$$

assuming that $RMSD_{k,mean}$ is the RMSD between model k and the mean structure.

4. Finally, model-model RMSD can be calculated in a similar way as in method 1., but comparison of any two models is preceded by fitting of the same two models on each other. This results in a necessarily lower-or-equal average RMSD than that of method 1., since the pairwise RMSDs are minimized in each step.

It is of utmost importance to clearly define how the RMSD was calculated, i.e. based on which subset of the atomic positions (backbone, heavy atoms, etc.) and relative to which structure (pairwise manner, to the mean structure or to a given model). The reason for this is that fitting algorithms guarantee minimal RMSD only if the same subset and algorithms are used for both fitting and RMSD calculation.

Local RMSD

Internal motion of a specific region within the protein can be quantified with the local RMSD. Local RMSD is often calculated for a single atom or a group of atoms, e.g. a single residue.



Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C Unit: III

Batch: 2019 -2021

Similarly, it has to be accurately defined, which atoms of the residue are used for local RMSD calculation.

The RMSD of a single atom (e.g. the $C\alpha$ of a given residue) can be calculated in two different ways.

1. In a pairwise manner, i.e. square distances are computed between all the possible pairs of identical atoms in different models. This also results in N(N-1)/2 distances, that have to be averaged.

$$RMSD_{pairwise}(\mathbf{r_i}) = \sqrt{\frac{2}{N(N-1)}} \sum_{k < l} |\mathbf{r_i}^k - \mathbf{r_i}^l|^2$$

where the indexes k and l go over the models of the ensemble.

2. Relative to the mean position of the given atom:

$$RMSD_{to mean}(\mathbf{r_i}) = \sqrt{\frac{1}{N}\sum_{k} |\mathbf{r_i}^k - \overline{\mathbf{r_i}}|^2}$$

It can be shown, that for high N

$$RMSD_{pairwise}(\mathbf{r_i}) = \sqrt{2} RMSD_{to mean}(\mathbf{r_i})$$

Proof of the above statement includes complicated statistical calculations. Local RMSD is usually calculated for more than one atom in each residue. In this case, the local RMSD of a residue (denoted by **resi**) is given as

$RMSD(\mathbf{res}_i) = \overline{RMSD(\mathbf{r}_i)}$

which is the average of the RMSDs of the single atoms within the given residue.

Maximum distances

Instead of the local RMSD, the highest distance can be given between two identical atomic positions between two different models. If more than one atom is considered in a residue, than distances between different types of atoms will not be measured and compared, only between identical atoms in different models.

Presentation of protein conformations



Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: III

Batch: 2019 -2021

Importance of Proteins

- Muscle structure depends on protein-protein interactions
- Transport across membranes involves protein-solute interactions
- Nerve activity requires transmitter substance-protein interactions
- Immune protection requires antibody-antigen interactions

Overview of Protiens It has

- Primary Structure
- Secondary Structure
- Tertiary Structure
- Quaternary Structure

Primary Structure

- Polypeptide chains \rightarrow Amino Acids
- Largest polypeptide chain approx has 5000AA but most have less than 2000AA
- Amino Acid Basic Structure H2N-CH-COOH
- Arrangement of the 20 amino acids in the polypeptide is the amino acid sequence which composes the primary structure of the protein



20 Amino Acids [(i) Nonpolar, hydrophobic, (ii) Polar, uncharged, (iii) Polar, charged]



Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 18CHP105-C

Unit: III







Course Name: Molecular Modelling and Drug Design

Course Code: 18CHP105-C

Unit: III

Batch: 2018 -2020



A Venn diagram showing the relationship of the 20 naturally occurring amino acids to a selection of physio-chemical properties thought to be important in the determination of protein structure.

Stereochemistry

- Configuration of amino acids in proteins
- The CORN Law



Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 18CHP105-C

Unit: III





Value Information Network Information Control of Million Action Description Des

Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 18CHP105-C

Unit: III







Course Name: Molecular Modelling and Drug Design

Course Code: 18CHP105-C

Unit: III

Batch: 2018 -2020



Protein Secondary Structure

Regular local structures formed by single strands of peptide chain due to constraints on backbone conformation

Peptide Bond



- Resonance
- C-N bond length of the peptide is 10% shorter than that found in usual C-N amine bonds
- Peptide bond planer
- ω , angle around peptide bond, 0^0 for cis. 180⁰ for trans





Course Name: Molecular Modelling and Drug Design

Course Code: 18CHP105-C

Unit: III





Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 18CHP105-C

Unit: III





Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 18CHP105-C

Unit: III







Course Name: Molecular Modelling and Drug Design

Course Code: 18CHP105-C

Unit: III





Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 18CHP105-C

```
Unit: III
```

Batch: 2018 -2020

Beta Sheet



Beta Sheet Features

- Sheets can be made up of any number of strands
- Orientation and hydrogen bonding pattern of strands gives rise to flat or twisted sheets
- Parallel sheets buried inside, while Antiparallel sheets occurs on the surface

More Beta Structures

Beta Bulge chymotrypsin (1CHG.PDB) involving residues 33 and 41-42

Anti parallel



Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 18CHP105-C

Unit: III





Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 18CHP105-C

Unit: III







Course Name: Molecular Modelling and Drug Design

Course Code: 18CHP105-C

Unit: III

Batch: 2018 -2020



Relative probabilities that a given amino acid will occur in the three common types of secondary structure.

Interactions

- Covalent bonds Disulphide bond (2.2 ⁰A) between two Cys residues
- Non-covalent bonds
 Long range electrostatic interaction
 Short range (4 ⁰A) van der Waals interaction
 Hydrogen bond (3 ⁰A)





Course Name: Molecular Modelling and Drug Design

Course Code: 18CHP105-C

Unit: III

Batch: 2018 -2020

Tertiary Protein Structure

- Defines the three dimensional conformation of an entire peptide chain in space
- Determined by the primary structure
- Modular in nature Aspects which determine tertiary structure
- Covalent disulfide bonds from between closely aligned cysteine residues form the unique Amino Acid cystine.
- Nearly all of the polar, hydrophilic R groups are located in the surface, where they may interact with water
- The nonpolar, hydropobic R groups are usually located inside the molecule Motifs and Domains
- Motif a small structural domain that can be recognized in a variety of proteins
- Domain Portion of a protein that has a tertiary structure of its own. In larger proteins each domain is connected to other domains by short flexible regions of polypeptide.







Course Name: Molecular Modelling and Drug Design

Course Code: 18CHP105-C

Unit: III







Course Name: Molecular Modelling and Drug Design

Course Code: 18CHP105-C

Unit: III

Batch: 2018 -2020



Modular Nature of Protiens

- Epidermal Growth Factor (EGF) domain is a module present in several different proteins illustrated here in orange.
- Each color represents a different domain




Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 18CHP105-C

Unit: III

Batch: 2018 -2020



Protein Folding



Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 18CHP105-C

Unit: III

Batch: 2018 -2020

- Protein folding constitutes the process by which a poly-peptide chain reduces its free energy by taking a secondary, tertiary, and possibly a quaternary structure Thermodynamics
- Proteins follow spontaneous reactions to reach the conformation of lowest free energy
- Reaction spontaneity is modeled by the equation $\Delta G = \Delta H T \Delta S$



Hydrophobicity factor

It is generally accepted today that the hydrophobic force is the dominant energetic factor that leads to the folding of polypeptide chains into compact globular entities. This principle was first explicitly introduced to protein chemists in 1938 by Irving Langmuir, past master in the application of hydrophobicity to other problems, and was enthusiastically endorsed by J.D. Bemal.

Being a more tangible idea, directly expressed in structural terms, the cyclol hypothesis received more attention than the hydrophobic principle and the latter never actually entered the mainstream of protein science until 1959, when it was thrust into the limelight in a lucid review by W. Kauzmann.

A theoretical paper by H.S. Frank and M. Evans, not itself related to protein folding, probably played a major role in the acceptance of the hydrophobicity concept by protein



Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 18CHP105-C

Unit: III

Batch: 2018 -2020

chemists because it provided a crude but tangible picture of the origin of hydrophobicity per se in terms of water structure.

A "like to like" mechanism as an important factor in hydrocarbon segregation. In micelle or bilayer formation, where one is dealing with solute molecules that are homogeneous or nearly so, with very long aliphatic hydrocarbon chains, this concept (perhaps lining up in parallel) has some intrinsic appeal. In applying the idea to proteins, "like to like" lacks any such plausibility: the non-polar moieties of protein amino acid side chains are not only short in length, but some are aliphatic and some aromatic. Furthermore, they are all mixed up with polar groups along the polypeptide chain, rather than neatly segregated.

State of knowledge of protein structure

There was intense interest in protein chemistry because proteins were seen to control a huge variety of biological processes: enzymatic activity, antibody specificity, oxygen binding, and even genetics and inheritance. Most people still thought that proteins were the carriers of genetic information, proved it was DNA and many people were not finally convinced until the "Waring Blender" experiments of Hershey and Chase. It was established that proteins consist of long chains of aminoacids in peptide linkage and that free amino and carboxyl groups carry actual ionic charges at neutral pH (and that these groups in proteins behave normally in response to changes in pH, much as they do in amino acids and other small molecules). Molecular weights were known for many proteins-often (e.g., for hemoglobin) quite accurate in the light of modem definitive values. The distinction between "fibrous" and "globular" proteins was established and a considerable number of the latter were being obtained in crystallographic results that the globular proteins were folded tightly into small compact particles. The phenomenon of protein "denaturation" was known and fitted into this picture as an unfolding of the tight globular structure.

In the case of protein folding, the hydrophobic effect is important to understanding the structure of proteins that have hydrophobic amino acids (suchas glycine, alanine, valine, leucine,

isoleucine, phenylalanine, tryptophan and methionine) clustered together within the protein. Structures of water-soluble proteins have a hydrophobic core in which side chains are buried from water, which stabilizes the folded state. Charged and polar side chains are situated on the solvent-exposed surface where they interact with surrounding water molecules. Minimizing the number of hydrophobic side chains exposed to water is the principal driving force behind the folding process, although formation of hydrogen bonds within the protein also stabilizes protein structure.

The energetics of DNA tertiary structure assembly were determined to be driven by the hydrophobic effect, in addition to Watson-Crick base pairing, which is responsible for sequence selectivity, and stacking interactions between the aromatic bases.

In biochemistry, the hydrophobic effect can be used to separate mixtures of proteins based on their hydrophobicity. Column chromatography with a hydrophobic stationary phase



Class: I- M.Sc (Chemistry) Course

Course Name: Molecular Modelling and Drug Design

Course Code: 18CHP105-C

Unit: III

Batch: 2018 -2020

such as phenyl-sepharose will cause more hydrophobic proteins to travel more slowly, while less hydrophobic ones elute from the column sooner. To achieve better separation, a salt may be added (higher concentrations of salt increase the hydrophobic effect) and its concentration decreased as the separation progresses.

Shape complementary

To investigate the effects of shape complementarity on the protein-protein interactions. By monitoring different kinds of protein shape-complementarity modes, we gave a clear mechanism to reveal the role of the shape complementarity in the protein-protein interactions, i.e., when the two proteins with shape complementarity approach each other, the conformation of lipid chains between two proteins would be restricted significantly. The lipid molecules tend to leave the gap formed by two proteins to maximize the configuration entropy, and therefore yield an effective entropy-induced protein-protein attraction, which enhances the protein aggregation. Definitely, the shape complementarity is the third key factor affecting protein aggregation and complex, besides the electrostatic-complementarity and hydrophobic complementarity.

To investigate the role of protein shape complementarity in the protein aggregation dissipative particle dynamic (DPD) simulations is used. In this simulation, Two main protein models were designed. One is typically cylindrical protein, and the other is "bowknot" protein with a groove body. Through regulating the size of the groove, we design three kinds of "bowknot" proteins which are larger, equal and smaller than the cylindrical protein radius, as shown in <u>Figure 1</u>.

Figure 1: Models of different proteins and the correct binding modes of proteins.

Complementary Shape Matching

A pose, π , between the context shapes CS_X^R from protein P_R (the receptor) and CS_Y^L from protein P_L (the ligand), with $X, Y \in \{vol, ses, sis, core, inK, outK\}$ and $K \in [1, 4]$, was represented by a one-to-one mapping of the context rays between the two context shapes. The feasibility of the pose was assessed using the overlap volume, defined as the volume that is labeled as inside in both CSs.

For a given pose π , the overlap volume of protein P_L 's context shape CS_{vol}^L with respect to the context shape CS_X^R for layer X of protein P_R , is given as:

$$OV(CS_{vol}^L, CS_X^R, \pi) = \sum_{i=1}^{\kappa} V(CR_i^L \wedge CR_{i_\pi}^R)$$
(1)



Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 18CHP105-C

Unit: III

Batch: 2018 -2020

where $CR_i^L \in CS_{vol}^L$, and $CR_{i_{\pi}}^R \in CS_X^R$ is a context ray mapped to CR_i^L according to pose π . Here $CR_i^L \wedge CR_{i_{\pi}}^R$ denotes the bitwise AND operation between the two context rays. The overlap volume within a single thin cone-shaped segment of the sphere is given as $V(CR_i^L \wedge CR_{i_{\pi}}^R) = \sum_{j=1}^{\beta} \nu(j)V[j]$, where $\nu(j) = (CR_i^L[j] \wedge CR_{i_{\pi}}^R[j])$ and V[j] is the actual volume corresponding to the *j*-th segment of the context ray. Depending on the choice of the layer X above, we obtain different kinds of overlap volumes. The total overlap volume between two context shapes is a symmetric quantity, and is given as $OV(CS_{vol}^L, CS_{vol}^R, \pi)$. It represents the overlap of one protein's volume with the other, for a given pose. On the other hand the layered overlap volume is asymmetric, and is given as $OV(CS_{vol}^L, CS_{vol}^R, CS_{vol}^L, \pi)$, where X is one of the *inK* or *outK* layers. It represents the part of one protein's volume that overlaps a given layer X in the other protein, for a given pose.







Enable | Enlighten | Enrich (Deemed to be University) (Under Section 3 of UGC Act 1956)

KARPAGAM ACADEMY OF HIGHER EDUCATION (Deemed to be University Under Section 3 of UGC Act 1956) COIMBATORE-21

DEPARTMENT OF CHEMISTRY (For the candidates admitted from 2019 & onwards) 19CHP105-C MOLECULAR MODELLING & DRUG DESIGN

Multiple Choice Questions for Unit III

| S. No | Question | Option 1 | Option 2 | Option 3 | Option 4 | Answer |
|-------|---|--|---|--|---|---|
| | Unit III | | | | | |
| 1 | The term CASP means | Crystal Analysis Structure Prediction | Critical analysis of structural protein | Critical Assesment of protein Structure Prediction | Critical Assay Structural Protein | Critical Assessment of protein Structure Prediction |
| 2 | In arrangement of protein sequences which one will work well. | Homology modelling | Computation al modelling | Threading | Theoretical modelling | Homology modelling |
| 3 | Comparative modeling is | Homology modeling and Threading | Computation al modeling and Theoretical modelling | Threading and Theoretical modelling | Homology modeling and Computational modeling | Homology modeling and Threading |
| 4 | Threading is an | Wave function | Energy function | Interaction | Isolation | Energy function |
| 5 | In protein threading, the total | E _p +E _s | | E_p+E_g | | $\mathbf{E}_{\mathbf{p}}$ + $\mathbf{E}_{\mathbf{s}}$ + $\mathbf{E}_{\mathbf{g}}$ |

| | energy is given by | | $E_p+E_s+E_g$ | | E _s +E _g | |
|----|---|--|--|--|---|---|
| 6 | Homology modeling needs sequence identity of | >30% | < 30% | 10-20% | 1-10% | >30% |
| 7 | To determine the 3D structure of target protein which one is the best | X-ray crystallograph y | Homology modelling | NMR | IR | Homology modelling |
| 8 | In Homology modeling, which software is used as best. | SWISS MODEL | SWISS Prot | MOE | GPMAW | SWISS MODEL |
| 9 | Trick for fast threading is | Sequence structure- Sequence contact matrix- Sequence 1D profile | Sequence Contact matrix- Sequence structure- Sequence 1D profile | Sequence structure- Sequence1D profile- Sequence contact matrix | Sequence 1D profile- Sequence structure- Sequence contact matrix | Sequence structure- Sequence contact matrix- Sequence 1D profile |
| 10 | In protein structure prediction, if template is hard to identify it is often called | Comparative modelling | Homology modelling | Fold recognition | Computative modelling | Fold recognition |
| 11 | In protein structure prediction, threading needs | >80% sequence identity | 60 – 80% sequence identity | < 30% sequence identity | 50 – 60% sequence identity | < 30% sequence identity |
| 12 | In Homology modeling, the accuracy of the results depends on | Sequence similarity | Aminoacid | Protein | Folding | Sequence similarity |
| 13 | Homology modeling needs | Known structure | Similar sequence | Known structure with similar sequence | Any structure | Known structure with similar sequence |
| 14 | Protein folding is used to find | 3D Protein structure | 2D Protein structure | 1D Protein structure | Sequence similarity | 3D Protein structure |
| 15 | Real protein fold over a time scale of | 10,000 secs | 0.1-1000 secs | 10 ¹⁴ secs | 1000- 5000secs | 0.1-1000 secs |

| 16 | Which one is used as a | Known | Unknown | Known | UnKnown | Known protein |
|----|------------------------------|----------------|---------------|----------------|---------------|----------------------|
| | template to model target | protein with | protein with | protein with | protein with | with best |
| | structure? | best | best | any score | any score | similarity score |
| | | similarity | similarity | | | |
| | | score | score | | | |
| 17 | Comparative modeling is | Homology | Computation | Threading and | Homology | Homology |
| | | modeling and | al modeling | Theoretical | modeling and | modeling and |
| | | Threading | and | modelling | Computational | Threading |
| | | | Theoretical | | modeling | |
| | | | modeling | | | |
| 18 | Topological property of | Ionic | Covalent | Coordinate | Charge | Covalent |
| | molecular modeling describes | connectivity | connectivity | connectivity | transfer | connectivity |
| | the | | | | connectivity | |
| 19 | NMR data gives the | Topological | Structural | Energetic | Thermodynam | Structural |
| | | property | property | property | ical property | property |
| 20 | A force field describing the | Topological | Structural | Energetic | Thermodynam | Energetic |
| | force acting on each atom of | property | property | property | ical property | property |
| | the molecules is called | | | | | |
| 21 | A sampling algorithm that | Topological | Structural | Energetic | Thermodvna | Thermodynami |
| | generates the | property | property | property | mical | cal property |
| | thermodynamical ensemble | r · r · · J | | r · r · J | property | ···· F - · F - · · J |
| | that matches experimental | | | | | |
| | conditions for the system is | | | | | |
| | called | | | | | |
| 22 | PDB means | Protein Data | Polymer Data | Program Data | Palm Data | Protein Data |
| | | Bank | Bank | Base | Base | Bank |
| 23 | SCOP means | Structural | Structural | Scientific | Scientific | Structural |
| | | Classification | Classificatio | Classification | Code Of | Classification |
| | | Of Polymers | n Of | Of Polymers | Polymers | Of Proteins |
| | | | Proteins | | | |
| 24 | DALI means | Donor- | Donor | Distance- | Donor- | Distance- |

| | | Acceptor LIgand | Activity of LIgand | matrix ALIgnment | Acceptor LInear complex | matrix ALIgnment |
|----|---|--|---|--|---|--|
| 25 | CATH means | Critical Assesment of THyroid | Critical Analogue of THyroid | Classical Analysis of Thyroid Harmone | Class, Architecture, Topology, and Homology | Class, Architecture, Topology, and Homology |
| 26 | Template search methods can be categorized into how many classes? | 1 | 2 | 3 | 4 | 3 |
| 27 | Pairwise comparison methods, which include the popular program of | DALI | BLAST and FASTA | SCOP | PALI | BLAST and FASTA |
| 28 | Sequence profile method use the program of | FASTA | BLAST | PSI-BLAST | PALI | PSI-BLAST |
| 29 | Threading methods use a combination of | Sequence and Structure | BLAST and PALI | BLAST and FASTA | DALI and PALI | Sequence and Structure |
| 30 | In which method the target sequence is threaded through a library of 3-D profiles or folds, and each threading is assessed based on a certain scoring function | Pairwise comparison method | Sequence profile method | Threading methods | Comparative method | Threading methods |
| 31 | In threading method, each threading is assessed based on a certain | Criteria | Scoring function | Property | Identity | Scoring function |
| 32 | BLAST means | Basic Logarithmic And Sequential Tool BLAST and | Basic Logarithmic Alignment of Search Tool | Basic Local Alignment Search Tool | Basic Logarithmic And Scientific Tool | Basic Local Alignment Search Tool Superfamily |

| | Commonly used methods and servers is | FASTA | and GenThreade r | | PALI | and GenThreader |
|----|--|----------------------------|---|-------------------------------|-------------------------|---|
| 34 | The threading methods are more effective in detecting | Homology | Topology | Alignment | Tools | Homology |
| 35 | Which is the least sensitive and are best used to detect close homologs? | Threading methods | Pairwise sequence comparison methods | Sequence profile method | Sequence method | Pairwise sequence comparison methods |
| 36 | PALI means | Protein ALIgnment | Pairwise ALIgnment | Phylogeny and ALIgnment | Property ALIgnment | Phylogeny and ALIgnment |
| 37 | Which is useful in making a distinction between the sequence and structure similarity? | FASTA | PALI | DALI | SCOP | PALI |
| 38 | Homology modeling, also known as | Space filling modelling | Mechanical modelling | Comparative modeling | Automodellin g | Comparative modeling |
| 39 | Which modeling allows to construct an unknown <i>atomic-</i> <i>resolution model</i> of the "target" protein? | Space filling modelling | Mechanical modelling | Comparative modeling | Automodellin g | Comparative modeling |
| 40 | Comparative modeling also called as | Homology modeling | Space filling modelling | Automodellin g | Mechanical modelling | Homology modeling |
| 41 | Rhodopsin is a biological pigment of the | Cornea | Retina | Īris | Pupil | Retina |
| 42 | Rhodopsin also known as | Lens | Pupil | Visual purple | Iris | Visual purple |
| 43 | Which modeling allows to construct an unknown <i>atomic-</i> <i>resolution model</i> of the "target" protein? | Space filling modelling | Mechanical modelling | Homology modeling | Automodellin g | Homology modeling |
| 44 | Homo sapiens is the Zoological name for? | Birds | Human | Animals | Insects | Human |

| 45 | Rhodopsin belongs to | Aminoacid family | Alkaloid family | G-protein coupled receptor family | Enzyme family | G-protein coupled receptor family |
|----|---|------------------------|-----------------------------------|--|--------------------------|---|
| 46 | Rhodopsin consists of the protein moiety | Retinal | Opsin | Retinol | GLy-Ala | Opsin |
| 47 | Which server has an automated procedure to get a model of a particular sequence? | SWISS Prot | SWISS- MODEL | MOE | GPMAW | SWISS- MODEL |
| 48 | Which server is for analyzing protein structures for validity and assessing how correct? | MOE | GPMAW | SAVES | SWISS Prot | SAVES |
| 49 | How many programmes are used for analyzing protein structures for validity and assessing how correct? | 2 | 3 | 4 | 5 | 5 |
| 50 | The stereochemical quality of a protein structure can be known by | PROCHECK | WHAT_CHE CK | ERRAT | VERIFY_3D | PROCHECK |
| 51 | Which method is based on characteristic atomic interaction, use for differentiating between correctly and incorrectly determined regions of protein structures? | PROCHECK | WHAT_CHE CK | ERRAT | VERIFY_3D | ERRAT |
| 52 | ProSA means | Pro Server Analysis | Protein Structural Analysis | Prosa Structured Analysis | ProServer Association | Protein Structural Analysis |
| 53 | Which-web service returns results instantaneously? | ERRAT | GPMAW | ProSA | SAVES | ProSA |
| 54 | The response time for even | Seconds | Few minutes | Minutes | Hours | Seconds |

| | large molecules in ProSA server is in the order of | | | | | |
|----|---|------------------------------------|--------------------------------------|-------------------------------------|---------------------------------------|-------------------------------------|
| 55 | Which web server is used for automated modeling of loops in protein structures? | ProSA | GPMAW | ModLoop | SAVES | ModLoop |
| 56 | High-accuracy template-based predictions for protein were evaluated in | CASP4 | CASP5 | CASP7 | CASP8 | CASP7 |
| 57 | Prediction of structure complexes of protein is done by | CASP2 | CASP4 | CASP6 | CASP7 | CASP2 |
| 58 | Protein model refinement is done by | CASP4 | CASP5 | CASP7 | CASP8 | CASP7 |
| 59 | Structure of protein is classified into | 2 types | 3 types | 4 types | 5 types | 4 types |
| 60 | The three dimensional conformation of an entire peptide chain in space is given by | Primary Structure of protein | Secondary Structure of protein | Tertiary Structure of protein | Quaternary Structure of protein | Tertiary Structure of protein |



Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: IV (Pharmacophore)

Batch: 2019 -2021

UNIT IV Pharmacophore

Historical perspective and viewpoint of pharmacophore, functional groups considered as pharmacophores, Ehrlich's "Magic Bullet", Fischer's "Lock and Key", two-dimensional pharmacophores, three-dimensional approach of pharmacophores, criteria for pharmacophore model, pharmacophore model generation software tools, molecular alignments, handling flexibility, alignment techniques, scoring and optimization, pharmacophores, validation and usage, automated pharmacophore generation methods, GRID-based pharmacophore models, pharmacophores for hit identification, pharmacophores for human ADME/tox-related proteins.



Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: IV (Pharmacophore)

Batch: 2019 -2021

Historical Perspective

Early Considerations About Structure-Activity Relationships

In his interesting Edelstein award lecture, presented at the 224th American Chemical Society Meeting in Boston, MA, in August 2002 and entitled "To Bond or Not to Bond: Chemical Versus Physical Theories of Drug Action", John Parascandola [8] relates the early history of structure–activity relationships.

Regarding drug selectivity, he cites Earles, who states: "The fact that drugs may exert a selective action on specific organs of the body had long been recognized empirically and expressed vaguely in the traditional designation of certain remedies as cordials (acting on the heart), hepatics (acting on the liver), etc." [9].

One of the earliest to recognize structure–activity relationships was Robert Boyle in 1685, who tried to explain the specific effects of drugs in terms of mechanical philosophy by suggesting that since the different parts of the body have different textures, it is not implausible that when the corpuscles of a substance are carried by the body fluids throughout the organism, they may, according to their size, shape and motion, be more fit to be detained by one organ than another [10].

Later, at the turn of the 20th century, the German scientist Sigmund Fränkel argued that the selective action of drugs can only be understood by assuming that certain groups in the drug molecule enter into a chemical union with the cell substance of a particular tissue. Once fixed in the cell in this manner, the drug can exert its pharmacological action [11].

Despite this pioneering view, the understanding of the nature of chemical bonding and of cellular structure and function was still in its infancy at the beginning of the 20th century. Thus there was significant controversy over whether the physical or the chemical properties of a substance could best explain its pharmacological action and over the value of attempts to relate the physiological activity of a drug to its chemical structure. As an example, in 1903 Arthur Cushny, Professor of Materia Medica and Therapeutics at the University of Michigan, published a paper in the *Journal of the American Medical Association* entitled "The pharmacologic action of drugs: is it determined by chemical structure or by physical characters?" [12]. To a chemist today, such a question might seem odd. Finding convincing answers to it became possible only after the discovery of the existence and role of pharmacological receptors.



Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C Unit: IV (Pharmacophore)

Batch: 2019 -2021

Early Considerations About the Concept of Receptors

The idea that drugs act upon receptors began with Langley in 1878 [13], who introduced the term "receptive substance" [14]. However, the word "receptor" was introduced later, by Paul Ehrlich [15, 16]. During the first half of the 20th century, several observations highlighted the critical features associated with the concept of receptors [17].

"Three striking characteristics of the actions of drugs indicate very strongly that they are concentrated by cells on small, specific areas known as receptors. These three characteristics are (i) the high dilution (often 10⁻⁹ M) at which solutions of many drugs retain their potency, (ii) the high chemical specificity of drugs, so discriminating that even D- and L-isomers of a substance can have different pharmacological actions, and (iii) the high biological specificity of drugs, e.g. adrenaline has a powerful effect on cardiac muscle, but very little on striatal muscle." [17].

Functional Groups Considered as Pharmacophores: the Privileged Structure Concept

The retrospective analysis of the chemical structures of the various drugs used in medicine led medicinal chemists to identify some molecular motifs that are associated with high biological activity more frequently than other structures. Such molecular motifs were called privileged structures by Evans et al. [2], to represent substructures that confer activity to two or more different receptors. The implication was that the privileged structure provides the scaffold and that the substitutions on it provide the specificity for a particular receptor. Two monographs deal with the privileged structure concept [3, 4].

Among the most popular privileged structures, historical representatives are arylethylamines (including indolylethylamines), diphenylmethane derivatives, tricyclic psychotropics and sulfonamides. Dihydropyridines [5], benzodiazepines, [2, 5], *N*-arylpiperazines, biphenyls and pyridazines [6] are more recent contributions.

A statistical analysis of NMR-derived binding data on 11 protein targets indicates that the biphenyl motif is a preferred substructure for protein binding [7].

Paul Ehrlich Magic Bullet



Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: IV (Pharmacophore)

Batch: 2019 -2021

Paul Ehrlich (Nobel Laureate: 1908) developed the concept of Magic Bullet. For staining any bacteria, we use different dyes. Basic dyes bind the surface of the bacteria. Magic Bullet is referred to the dyes which bind selectively to the pathogen without harming the human cells.

Magic Bullet (Medicine)

The **magic bullet** was a scientific concept developed by a German Nobel laureate Paul Ehrlich in 1900.[[] While working at the Institute of Experimental Therapy (*Institut für experimentelle Therapie*), Ehrlich formed an idea that it could be possible to kill specific microbes (such as bacteria) that cause diseases without harming the body itself. He named the hypothetical agent as *Zauberkugel*, the magic bullet.[[] He envisioned that just like a bullet fired from a gun to hit a specific target, there could be a way to specifically target invading microbes. His continued research to discover the magic bullet resulted in further knowledge of the functions of the body's immune system, and in the development of Salvarsan, the first effective drug for syphilis, in 1909. His works were the foundation of munology, and for his contributions he shared the 1908 Nobel Prize in Physiology or Medicine with Élie Metchnikoff.

Ehrlich's discovery of Salvarsan in 1909 for the treatment of syphilis is termed as the first magic bullet. This led to the foundation of the concept of chemotherapy.

Ehrlich joined the Institute of Experimental Therapy (*Institut für experimentelle Therapie*) at Frankfurt am Main, Germany, in 1899, becoming the director of its research institute the Georg–Speyer Haus in 1906. Here his research focused on testing arsenical dyes for killing microbes. Arsenic was an infamous poison, and his attempt was criticised. He was publicly lampooned as an imaginary "Dr Phantasus".

But Ehrlich's rationale was that the chemical structure called side chain forms antibodies that bind to toxins (such as pathogens and their products); similarly, chemical dyes such as arsenic compounds could also produce such side chains to kill the same microbes. This led him to propose a new concept called "side-chain theory". (He later, in 1900, revised his concept as "receptor theory".) Based on his new theory, he postulated that in order to kill microbes, "*wir müssen chemisch zielen lernen*" ("we have to learn how to aim chemically"). His institute was convenient as it was adjacent to a dye factory.



Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: IV (Pharmacophore)

Batch: 2019 -2021

He began testing a number of compounds against different microbes. It was during his research that he coined the terms "chemotherapy" and "magic bullet". Although he used the German word *zauberkugel* in his earlier writings, the first time he introduced the English term "magic bullet" was at a Harben Lecture in London in 1908. By 1901, with the help of Japanese microbiologist Kiyoshi Shiga, Ehrlich experimented with hundreds of dyes on mice infected with trypanosome, a protozoan parasite that causes sleeping sickness. In 1904 they successfully prepared a red arsenic dye they called Trypan Red for the treatment of sleeping sickness.

Examples of Magic Bullet:

- 1. **Trypan Red:** It binds the Trypanosomes which cause African sleeping sickness. It was first discovered by Kiyoshi Shiga and Paul Ehrlich.
- 2. **Prontosil Red:** It is used for staining leather. However, when injected into the mice, it protected the mice completely against pathogenic staphylococci and Streptococci.

All the dyes are commonly having sulfur and are structurally similar. Sulfonamide/ sulfa drugs were discovered by Gerhard Domagk.

Discovery of the first magic bullet-Salvarsan

In 1906 Ehrlich developed a new derivative of arsenic compound, which he code-named Compound 606 (the number representing the series of all his tested compounds). The compound animals.[[] In was effective against malaria experimental 1905. Fritz infection in Schaudinn and Erich Hoffmann identified a spirochaete bacterium (Treponema pallidum) as the causative organism of syphilis. With this new knowledge, Ehrlich tested Compound 606 (chemically arsphenamine) on a syphilis-infected rabbit. He did not recognise its effectiveness. Sahachiro Hata went over Ehrlich's work and found on 31 August 1909 he found that the rabbit, which had been injected with Salvarsan 606 was cured using only a single dose, the rabbit showed no adverse effect. (The normal treatment procedure of syphilis at the time involved two to four years routine injection with mercury.) Ehrlich, after receiving this information, performed experiments on human patients with the same success. After convincing clinical trials, the



Class: I- M.Sc (Chemistry) Course Name: Molect

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: IV (Pharmacophore)

Batch: 2019 -2021

compound number 606 was given a trade name "Salvarsan", a portmanteau for "saving arsenic". Salvarsan was commercially introduced in 1910, and in 1913, a less toxic form "Neosalvarsan" (Compound 914) was released in the market. These drugs became the principal treatments of syphilis until the arrival of penicillin and other novel antibiotics towards the middle of the 20th century. Ehrlich's research on the magic bullet was the foundation of pharmaceutical research.

Fischer's "Lock and Key"

To quote Fischer (1868) himself, "There is a relation between the unknown structure of an active enzyme and that of substrate, they are complementary and the one may be said to fit the other as a key fits a lock."

Fischer's Lock and Key Model Previously, the interaction of substrate and enzyme was visualized in terms of a lock and key model (also known as template model), proposed by Emil Fischer in 1898. According to this model, the union between the substrate and the enzyme takes place at the active site more or less in a manner in which a key fits a lock and results in the formation of an enzyme substrate complex as shown below.



Formation of an enzyme-substrate complex according to Fischer's lock and key model

In fact, the enzyme-substrate union depends on a *reciprocal fit* between the molecular structure of the enzyme and the substrate. And as the two molecules (that of the substrate and the enzyme) are involved, this hypothesis is also known as the **concept of intermolecular fit**. The enzyme-substrate complex as shown below is highly unstable and almost immediately this complex decomposes to produce



Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: IV (Pharmacophore)

Batch: 2019 -2021



Structure of an enzyme-substrate complex

(i) The enzyme cytochrome P-450 is illustrated bound to its substrate camphor and

(ii) In the active site, the substrate is surrounded by residues from the enzyme. Note also the presence of a heme cofactor.

the end products of the reaction and to regenerate the free enzyme. The enzyme-substrate union results in the release of energy. It is this energy which, in fact, raises the energy level of the substrate molecule, thus inducing the *activated state*. In this activated state, certain bonds of the substrate molecule become more susceptible to cleavage.

Evidences Proving the Existence of an ES Complex:

The existence of an ES complex during enzymatically-catalyzed reaction has been shown in many ways :

1. The ES complexes have been directly observed by electron microscopy and x-ray crystallography.

2. The physical properties of enzymes (*esp.*, solubility, heat sensitivity) change frequently upon formation of an ES complex.

3. The spectroscopic characteristics of many enzymes and substrates change upon formation of an ES complex. It is a case parallel to the one in which the absorption spectrum of deoxyhemoglobin changes markedly, when it binds oxygen or when it is oxidized to ferric state.

4. Stereospecificity of highest order is exhibited in the formation of ES complexes. For example, D-serine is not a substrate of tryptophan synthetase. As a matter of fact, the D-isomer does not even bind to the enzyme.

5. The ES complexes can be isolated in pure form. This may happen if in the reaction, $A + B \Longrightarrow C$,



Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: IV (Pharmacophore)

Batch: 2019 -2021

the enzyme has a high affinity for the substrate A and also if the other reactant B is absent from the mixture.

6. A most general evidence for the existence of ES complexes is the fact that at a constant concentration of enzyme, the reaction rate increases with increase in the substrate concentration until a maximal velocity is reached.

Two-dimensional Pharmacophores

Sulfonamides and PABA

The recognition of the quantitatively almost unmatched ability of *p*-aminobenzoic acid (PABA) to oppose the bacteriostatic efficiency of the sulfonamides led Woods and Fildes [19, 20] to formulate the fundamentals of the theory of metabolite antagonism (Fig. 1.1).





Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: IV (Pharmacophore)

Batch: 2019 -2021



Fig. 1.1 PABA and *p*-aminobenzenesulfonamide show similar critical distances. The incorporation of the sulfonamide instead of PABA inhibits the biosynthesis of tetrahydrofolic acid.

Estrogens

Another early achievement (Fig. 1.2) was the synthesis and the pharmacological evaluation of *trans*-diethylstilbestrol as an estrogenic agent showing similarities with estradiol [21]. Here again the proposed model was two-dimensional [22], despite the fact that the non-planar conformation of estradiol was already known.





An Early Three-dimensional Approach: the Three-point Contact Model



Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: IV (Pharmacophore)

Batch: 2019 -2021

When an asymmetric center is present in a compound, it is thought that the substituents on the chiral carbon atom make a three-point contact with the receptor. Such a fit insures a very specific molecular orientation which can only be obtained for one of the two isomers (Fig. 1.3). A three-point fit of this type was first suggested by Easson and Stedman [23], and the corresponding model proposed by Beckett [24] in the case of (R)-(–)-adrenaline [= (R)-(–)-epinephrine]. The more active natural (R)-(–)-adrenaline establishes contacts with its receptor through the three interactions shown in Fig. 1.3.

In simply assuming that the natural (R)-(–)-epinephrine establishes a threepoint interaction with its receptor (A), the combination of the donor–acceptor interaction, the hydrogen bond and the ionic interaction will be able to generate energies of the order of 12–17 kcal mol⁻¹, which corresponds [25] to binding constants of 10^{-9} – 10^{-12} . The less active isomer, (*S*)-(+)-epinephrine, may establish only a two-point contact (B). The loss of the hydrogen bond interaction equals ~ 3 kcal mol⁻¹, hence this isomer should possess an ~ 100-fold lesser affinity. Experience confirms this estimate. If we consider less abstract models, it becomes apparent that the less potent enantiomer also is able to develop three intermolecular bonds to the receptor, provided that it approaches the receptor in a different manner. However, the probability of this alternate binding mode to trigger the same biological response is close to zero.



Fig. 1.3 Interaction capacities of the natural (R)-(-)-epinephrine and its (S)-(+)-antipode.



Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: IV (Pharmacophore)

Batch: 2019 -2021

Criteria for a Satisfactory Pharmacophore Model

To be recognized as a useful tool, a pharmacophore model has to provide valid information for the medicinal chemist exploring structure-activity relationships. 1. First, it has to highlight the functional groups involved in the interaction with the target, the nature of the non-covalent bonding and the different intercharge distances. This means that worthless images of ribbon and spaghetti models [33], without indication of the molecular features of the interacting partners, have to be avoided. This is true also for many unnecessary and opaque theoretical digressions. The model also has to show some *predictive power* and lead to the design of new, more potent compounds or, even better, of totally novel chemical structures, not evidently deriving from the translation of structural elements from one active series into the other. An interesting aspect of pharmacophore-based analogue design is referred to as scaffold hopping. It consists in the design of functional analogues by searching within large virtual compound libraries of isofunctional structures, but based on a

different scaffold. The objective is to escape from a patented chemical class in identifying molecules in which the central scaffold is changed but the essential function-determining points are preserved and form the basis of a relevant pharmacophore [34].



Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: IV (Pharmacophore)

Batch: 2019 -2021

- 2. The second criterion for a valid pharmacophore model is that it should discriminate stereoisomers. Stereospecificity is one of the principal attributes of pharmacological receptors and a perfect stereochemical complementarity between the ligand and the binding-site protein is an essential criterion for high affinity and selectivity. A convincing example of enantiomeric discrimination was observed for GABA-A receptor antagonists [35].
- 3. In a similar manner, the ideal model should distinguish between agonists and antagonists. This is relatively easy for the specific category of antagonists which, according to Ariëns et al. theory [36], derive from the agonists simply through the addition of some supplementary aromatic rings which play the role of additional binding sites (e.g. the passage from muscarinic agonists to muscarinic antagonists [37] or from GABA agonists to GABA antagonists [35]). The discrimination between the two categories becomes less evident when the passage from agonist to antagonist relies on relatively subtle changes such as one observes for glutamate, oxotremorine and benzodiazepine antagonists.
- 4. Sometimes a good pharmacophore model can *explain* apparently *paradoxical observations*, e.g. the unexpected affinity reversal found in *R* and *S*-enantiomers of the sulpiride series on changing *N*-ethyl to *N*-benzyl derivatives [38].
- 5. Finally, it has to account for the *lack of activity* of certain analogues of the active structures. The knowledge of structural or electronic parameters leading to poorly active or inactive compounds is a cost-lowering factor that allows the number of compounds to be synthesized to be reduced.

Pharmacophores: the Viewpoint of a Medicinal Chemist

Even before the advent of computer-aided drug design, simple pharmacophores were described in the literature and considered as tools for the design of new drug molecules. Initial structure–activity relationship considerations were accessible in the 1940s thanks to the knowledge of the bond lengths and the van der Waals sizes which allowed the construction of simple two-dimensional model structures. With the availability of X-ray analysis and conformational chemistry, access to three-dimensional models became possible in the 1960s.

Pharmacophore Model Generation Software Tools



Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: IV (Pharmacophore)

Batch: 2019 -2021

Introduction

Although the concept of pharmacophores constituting a simple representation of molecules and chemical groups in certain order was introduced nearly a century ago [1], there has been increasing interest and focus on pharmacophores in recent years following the advances in computational chemistry research. The historical development of the pharmacophore concept has recently been reviewed [2].

Often, all alignment-based methods and molecular field and potential calculations are classified as pharmacophore perception techniques. We will include most of these methods in this review; however, when using the term pharmacophore model, we will be referring mainly to one specific type of perception, namely three-dimensional feature-based pharmacophore models represented by geometry or location constraints, qualitative or quantitative. An extrapolation of the pharmacophore approach to a set of multi-dimensional descriptors (pharmacophore fingerprints) has been developed mostly for library design and focusing purposes [3–8].

At the beginning of this chapter we will look into the different automated alignment methods as correct alignment is the first and most important prerequisite for a successful pharmacophore identification process. Further, we will elaborate how essential issues of pharmacophore modeling such as conformational search, pharmacophore feature definitions, compounds structure storage and screening are handled by various available software packages.

Various ways of perceiving pharmacophores have been explored, known issues with pharmacophore modeling have been addressed in one way or another and several computer-based applications with a pharmacophore focus have been created since the 1980s. Many of these programs are not intensively used today, but we consider that they should be mentioned in this review: ALADDIN [9], DANTE [10–13], APOLLO [14], RAPID [15], SCREEN and its PMapper from ChemAxon [16] and ChemX fingerprints [3] from Chemical Design (now Accelrys).



Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: IV (Pharmacophore)

Batch: 2019 -2021

This review is based on literature data and on the personal experience of the authors and should not represent a direct comparison between packages but rather a snapshot of the current developments in pharmacophore perception technology from our perspective.

Molecular Alignments

Although the terms molecular alignment and superposition and pharmacophore elucidation are often used interchangeably, it is probably more accurate to differentiate alignment as providing a prerequisite to pharmacophore development. Conversely, some alignment methods require a pharmacophore as a starting point [17–19]. In this section, we briefly overview the molecular alignment methods available; extensive reviews and summaries of different superposition algorithms over the last 10 years are available elsewhere [20–22]. Of course, molecular alignment is not limited to just providing a basis for pharmacophore elucidation; it can also be used to derive 3D-QSAR models that potentially can estimate binding affinities, in addition to indirectly providing insight into the spatial and chemical nature of the receptor–ligand interaction of the putative receptor. Essentially, an alignment endeavors to produce a set of plausible relative superpositions of different ligands, hopefully approximating their putative binding geometry.

Many of the issues and concerns in the generation of pharmacophore models are inherent in different alignment methods. These issues can be used to differentiate or categorize the plethora of available algorithms.







Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: IV (Pharmacophore)

Batch: 2019 -2021

Primary among these issues is that of ligand flexibility, vital in the determination of the relevant binding conformation for each of the ligands concerned. Alignment methods can be considered rigid, semi-flexible or flexible. Rigid methods, while generally simpler and faster, require a presumption of the bioactive conformation of the ligands; this is often not possible and also removes the impartiality of the method. Semi-flexible methods are those that are fed with pre-generated conformers which are processed in either a sequentially, iterative or combinatorial manner. These methods lead to a further series of considerations such as whether the weighting, number and spread of conformers are determined by energy cut-offs or Boltzmann probability distributions and whether solvation models should be used. Flexible methods are considered to be those in which the conformational analysis is performed on-the-fly and these are generally the most time consuming as they require rigorous optimization.

Alignment Techniques



Property-based algorithms, often also termed field-based, make use of grid or field descriptors, the most popular of which are those obtained from the program GRID, developed by Goodford [25]. These are generated by defining a three-dimensional grid around a ligand and calculating the energy of interaction between the ligand and a given probe at each grid point. These diverse descriptors include various molecular properties such as molecular shape and volume, electron density, charge distribution such as molecular electrostatic potentials and even high-level quantum mechanical calculations.



Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: IV (Pharmacophore)

Batch: 2019 -2021

These algorithms are commonly broken down into three stages, which are subject to much variation. First, each ligand is represented by a set of spheres or Gaussian functions displaying the property or properties of interest. Usually the property is first calculated on a grid and subsequently transformed to the sphere or Gaussian representation. A number of random or systematically sampled starting configurations are then generated depending on the degrees of freedom considered, rotational, translational and conformational. Finally, local optimizations are performed with some variant of the classical similarity measure of the intermolecular overlap of the Gaussians as the objective function. While earlier property-based alignment methods were commonly gridbased, these have been surpassed by Gaussian molecular representation and Gaussian overlap optimization. These provide high information contents and avoid the dependence on additional parameters such as grid spacing while also providing a substantial increase in speed.

Variations on these algorithms have included the application of Fourier space methods to optimize the electron density overlap, similar to the molecular replacement technique in X-ray crystallography [26] and differentially weighted molecular interaction potentials or field terms [27, 28]. Another interesting alternative has been to apportion the conformational space of the ligands into fragments, compute the property field on pairs of fragments and determine the alignment by a pose clustering and incremental build-up procedure of retrieved fragment pairs [29].

Scoring and Optimization





Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: IV (Pharmacophore)

Batch: 2019 -2021

All alignment methods require some quantitative measure or fitness function, to assess the degree of overlap between the ligands being aligned and to monitor the progression of that optimization. This is most often manifested as a molecular similarity score or alignment index [22].

Typically in point-based algorithms, the optimization process endeavors to reduce the root-mean-square (RMS) deviation of the distances between the points or cliques by least-squares fitting. However, interesting variations have been developed including the use of distance matrices to represent any given conformation of a ligand [30]. Simulated annealing is used to optimize the fitness function, which is a quantification of the sum of the elements of the difference distance matrix created by calculating the magnitude of the difference for all corresponding elements of two matrices.

Another optimization method, related to the least-squares fitting used in point-based algorithms, is the directed tweak method [31]. This is a torsional space optimizer, in which the rotatable bonds of the ligands are adjusted at search time to produce a conformation which matches the 3D query as closely as possible. As directed tweak involves the use of analytical derivatives, it is very fast and allows for an RMS fit to consider ligand flexibility.

In property-based alignments where the molecular fields are represented by sets of Gaussian functions, the intermolecular overlap of the Gaussians is used as the fitness function or similarity index. The two most common optimization methods are Monte Carlo and simulated annealing [32, 33]. Other straightforward optimization algorithms include gradient-based methods and the simplex method, which seeks the vector of parameters corresponding to the global extreme (maximum or minimum) of any *n*-dimensional function, searching through the parameter space [34].

Further, more sophisticated, optimization algorithms include neural networks and genetic algorithms which mimic the process of evolution as they attempt to identify a global optimization [35]. In an alignment procedure chromosomes may encode the conformation of each ligand in addition to intramolecular feature correspondences, orientational degrees of freedom, torsional degrees of freedom or other information such as molecular electrostatic potential fields. During the optimization the chromosome undergo manipulation by genetic operators such as crossover and mutation.



Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: IV (Pharmacophore)

Batch: 2019 -2021

Alignment methods are also known to combine different optimization methods, such as a genetic algorithm and a directed tweak method [36].

Although this summary has highlighted the most common differentiators that can be used to categorize the plethora of available algorithms, further issues are significant to the alignment dilemma. Such issues include the degree of human intervention required, how to address the relative significance or weighting of some ligands over other ligands and how some algorithms generate multiple alignment solutions rather than an optimum superposition.

Pharmacophores, Validation and Usage



- 1. Statistical significance analysis, randomization tests.
- 2. Enrichment based methods. These focus on recovering active molecules from a test database in which a small number of known actives have been hidden in a large database of randomly selected compounds. Database mining and the utilization of receiver operating characteristic (ROC) curves [43] can be included in this category.
- 3. Biological testing of a selection of compounds.

The main utility of pharmacophores is their use as screening tools. Many examples in the literature show their successful usage in finding new scaffolds [44–51].

Automated Pharmacophore Generation Methods

In order to have an objective view of the different available pharmacophore perception software tools, we chose to analyze these using the following criteria whenever possible:



Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: IV (Pharmacophore)

Batch: 2019 -2021

- 1. Compound builder: is there a molecular builder? Which file formats are supported?
- 2. Stereochemistry: how is the stereochemistry of molecules handled in the program?
- 3.3D conformations: does the program contain conformer generation methods?
- 4. Pharmacophore generation engine: which type of pharmacophore perception engine is implemented in the program?
- 5. Fitness function: how are the pharmacophores evaluated?
- 6. Alignment method: does the program require pre-alignment of molecules? On what basis are the molecules aligned together?
- 7. Pharmacophore definition: description of the type of pharmacophore locations and associated functions. Can other descriptors be added to the pharmacophore alignment?
- 8. Database searching: is a database search engine implemented in the program? Which types of database searches are possible, e.g. substructure, pharmacophore, shape, exclusion?
- 9. Scoring of hits: Can database search hits be ranked?

The currently available pharmacophore perception methods are reviewed here in three major categories: geometry- and feature-based methods, field-based methods and pharmacophore fingerprints. Finally, the methods that do not fall into any of the above categories are described in an additional section.





Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: IV (Pharmacophore)

Batch: 2019 -2021

The pharmacophore concept is a widely accepted and useful approach in both early drug discovery stages of hit determination and lead optimization [1]. A vast number of slightly different methods to build such models exist and some of them are discussed in this book. Usually pharmacophore models are created by collecting the most relevant structural features of biologically active compounds. When no three-dimensional structure information on the target exists most cases of the time, chemical intuition is necessary for completing the ligand-based pharmacophore generation, in ambiguous cases possibly leading to erroneous models. One of the most advanced applications of pharmacophore models is to use them as virtual screening filters of large compound database sets against a wide variety of multiple macromolecular targets [2]. This emerging technique, now considered as a new source of novel drug leads [3], is attracting more and more the attention of industrial pharmaceutical research.

Among the computational methodologies widely adopted in drug design studies, Goodford's GRID [4] program is very well accepted and trusted in the scientific community. It works by mapping the three-dimensional space around molecular targets with probes mimicking the main chemical properties of most common atom types and small moieties that are found in ligands. GRID data can be used to identify the best probe locations as map display and also 3D information for chemometric analysis [5–8]. The large availability of crystal and NMR structures of macromolecular complexes deposited in the Protein Data Bank (PDB) [9] is an excellent source for studying interactions between molecules of different nature (proteins, nucleic acids, small organic ligands).

In this chapter, we describe the results of our studies we aimed at the development of a general computational procedure to generate automatically and unbiased objective pharmacophore models using the GRID approach and starting with PDB macromolecular complexes. Within the context of structure-based pharmacophore modeling, it represents an approach that is somehow complementary to that described in Chapter 6. We have used logically combined maps



Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: IV (Pharmacophore)

Batch: 2019 -2021

computed with the GRID force field in order to derive essential information on the interactions between occurring within the molecules of a PDB complexes for to generating chemical feature-based pharmacophore models. However, this approach can also be extended to cases where only one macromolecular complex partner is present, since the computation of GRID maps requires at least one binding partner. The versatility of the new computational approach has been tested benchmarked in several application examples using molecular complexes of different nature.

Theoretical Basis of the GBPM Method



The GRID-based pharmacophore model (GBPM) is created in a six-step procedure as depicted in Fig. 7.1.

The first step is dedicated to the PDB file pretreatment, which often contains water molecules and no hydrogen atoms. In the pretreatment, the user should fix typical problems such as missing residues, missing side-chains and wrong bond orders, especially for bound organic compounds. The GREAT and GRIN modules of the GRID software help contribute to this task and allow the preparation of the GRID mapping procedure. Assuming that the complex has two interacting molecules *a* and β , as in the case of protein–protein or protein–ligand complexes, the main goal of this first step is to obtain three interaction energy maps with from the *a*+ β , *a* and β subunits, keeping the atomic coordinates of the original PDB model (Fig. 7.1).

The *second step* performs the GRID calculation with a given probe on the three subunit models. In order to make the application of Boolean operations with the map files as easy as possible, the matrix dimension of the GRID box is exactly maintained as in the largest model, i.e. that with $a+\beta$ subunits, maintaining, for both subunits, the original complex atom coordinates. The three maps obtained are named **A**, **B** and **C**, respectively (Fig. 7.1).



Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: IV (Pharmacophore)

Batch: 2019 -2021

The *third step* is based on the GRAB procedure, implemented in GRID v. 21, performing a Boolean operation [10] between the maps **B** and **A**. The resulting map **D** has, by definition, the same matrix dimension of the original maps and reports, with negative energy values, the $a-\beta$ interaction areas. According to the GRAB algorithm [10], the *a* components are converted into positive or zero values comparing maps **D** and **C**. The resulting map **E** reports the acceptance degree of a certain probe into the $a-\beta$ binding site. Such an indication represents a first, interesting, advantage of the GBPM method, since actually no indication has been given in order to identify the right positioning and the extension of map **E**. Definitely, each point of the $a-\beta$ interaction area is automatically defined with unbiased influence of the user.

The *fourth step* is dedicated to the identification of the most important interaction areas of map \mathbf{E} . This task is carried out using the MINIM utility included into the GRID program. This program collects all points within a certain energy

threshold, allowing the interpolation of the closest ones. The choice of an energy threshold value is a biased task *per se* but, considering a pharmacophore model as a minimum interaction descriptor built by few features, we have generally found an energy threshold about 10% higher than the global minimum value to be appropriate. This means, in most cases, about 1 kcal mol⁻¹ above the global minimum energy determined. Actually, such a value allows at least one feature to be collected for each probe used. Often the above energy threshold yields too complicated pharmacophore models that can be reduced using the GRID energy as a cutting criterion.

In order to design a suitable pharmacophore model, all reported operations should be repeated using at least three different probes: the hydrophophic probe (DRY), a hydrogen bond acceptor (O) and a hydrogen bond donor (N1). This choice allows a basic characterization of most of the interaction areas; however, more sophisticated and selective models can be obtained by adding other GRID probes such as halogen or charged atoms. In the *fifth step*, the information originating from the different probe experiments are simply merged into a preliminary pharmacophore model (multiple probe features of Fig. 7.1).



Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: IV (Pharmacophore)

Batch: 2019 - 2021

The *sixth step* is dedicated to the validation of this the preliminary model and eventually its modulation in terms of number of features (i.e. its complexity). The quality of the pharmacophore model is tested as the capability to recognize selectively the original ligand present in the PDB file. Technically the evaluation step can be carried out by the Catalyst software [11], in particular using the Ci-Test fit module [12]. The preliminary model is imported converting the GBPM points into Catalyst features. The GRID energies are also included in the fit analysis as feature weight according to the following equation:

 $wF_{ij} = EF_i | AEF_j \tag{1}$

where wF_{ij} is the weight for the feature *i* into the hypothesis *j*, *EF*_i is the GRID energy for the features *i* and *AEF*_j is the average GRID energy value for the hypothesis *j*. This approach allows a maximum fit value (MFV) equal to the total number of features available for the hypothesis *j*. Taking into account the GRID energies, several preliminary models (hypotheses) can be designed reducing the number of features. Unfortunately, owing to the high variability, i.e. extension and interaction type, of the *a*- β subunit interface, the number of preliminary models can not be predefined. Therefore, in order to identify the best one, all possible models are submitted to a CiTest fit.

A fit index (FI), defined as ratio between the CiTest fit and MFV, is used for the evaluation of each hypothesis and as a choice criterion for the identification of the best GBPM.

Moreover, the FI descriptor, which makes possible comparisons among models with different numbers of features, can be used to extend the evaluation step including other molecules known to interact with the same β subunit binding site. Such an eventuality was found to improve strongly the quality of the final model.




Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: IV (Pharmacophore)

Batch: 2019 -2021





Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: IV (Pharmacophore)

Batch: 2019 -2021

Fig. 7.1 Flow chart of the GBPM starting from a PDB complex. The bottom figure represents a generic feature-based pharmacophore model.

Pharmacophores for Hit Identification

Many tools and protocols are now available for the scientist involved in the drug discovery process. When the structure of the macromolecular target (usually termed the receptor) is unknown, the ligand-based drug design approach can be applied for different purposes. As an example, given a set of compounds acting through the same mechanism of action (that is, able to bind to the same site of a receptor), one can investigate the chemical features responsible for the activity and summarize them in terms of pharmacophoric models. However, a problem that sometimes arises in pharmacophore-based approaches is the need to take into account possible adverse steric interactions between inactive compounds in a dataset and the target protein counterpart. The most common situation encountered in the literature is connected with the mining of large databases. In this case, the most likely outcome of gueries based on relatively simple pharmacophore hypotheses (that contain three or four features) would be very large hit lists of several hundreds of compounds, difficult to evaluate critically. Addition of excluded volume spheres to pharmacophores or ligand-forbidden zones to constrain the models is consequently expected to reduce the number of retrieved hits considerably.

Based on the above considerations, we report in this chapter two case studies where pharmacophore generation and handling plays a pivotal role in finding new hits. In the first example, a classical computational strategy consisting of pharmacophore building, pharmacophore validation, database mining, hit identification and hit optimization is described, aiming at the identification of potent antagonists of the a_1 adrenergic receptor. Additionally, we also report how this original pharmacophore model for a_1 adrenoceptor antagonists evolved towards a_{1d} subtype selectivity. In the second example, in contrast, the rationalization of the antifungal activities of azole compounds is exploited to discuss the importance and utility of adding excluded volume spheres (representing regions of the space forbidden to the ligands) to a pharmacophore.



Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: IV (Pharmacophore)

Batch: 2019 -2021

Pharmacophores for Human ADME/Tox-related Proteins

The last decade has witnessed an enormous increase in the number of compounds flowing through the drug discovery and development pipeline in the pharmaceutical industry, primarily owing to the advent of combinatorial chemistry and high-throughput screening (HTS) (Rodrigues 1997). These new technologies may have increased the chances of finding new lead compounds beyond traditional medicinal chemistry methods. However, expensive phase II and phase III clinical trial failures related to unsatisfactory ADME/Tox properties have also increased. In order to improve the rate of success in the more costly downstream stages of drug development, ADME/Tox evaluations have been shifted into the very early part of the discovery process. New technologies such as in vitro HTS and in silico approaches have been developed to meet the new challenges of large compound numbers and shortened cycle times that are characteristic of this phase of drug discovery. As the most cost-effective method, in silico screening has the additional advantage of being able to reduce significantly the experimental effort in the screening phase of drug discovery (Boobis et al. 2002). In silico approaches include three-dimensional quantitative structure-activity relationship (3D-QSAR) and pharmacophore modeling, which can be used directly as database searching methods. Recently, there have been several reviews focused on different aspects of computational ADME/Tox and more recently one of these (Ekins and Swaan 2004) has reviewed in silico approaches to modeling the specific proteins involved in determining ADME/Tox properties. The primary aim of this chapter is to review briefly some of the recent applications of pharmacophore technologies in drug discovery ADME/Tox studies. For other OSAR methods the reader is referred to several useful recent reviews (Barratt and Rodford 2001; Wessel and Mente 2001; Butina et al. 2002; Greene 2002; van de Waterbeemd and Gifford 2003).



Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: IV (Pharmacophore)

Batch: 2019 -2021

Absorption into the bloodstream is the first step for drugs to reach their targets followed by distribution to tissues. The drugs are then metabolized into more readily excreted forms. All of these aspects are significantly mediated or influenced by transporters, enzymes, ion channels and receptors. This complex interplay of different proteins coordinates to absorb nutrients and protect against the accumulation and toxic compounds. The potential overlap or competition of multiple compounds with affinity for the same protein raises the potential for possible drug-drug interactions (DDI), which could result in either reduced drug efficiency or increased drug toxicity owing to extended bioavailability. Methods to predict these types of interactions effectively are highly desirable and in recent years we have seen the focus of much research on computational methods. The interest in computational models based on in vitro data for predicting potential drug interactions via these multiple proteins (Ekins and Swaan 2004) follows the computational assessment of properties such as absorption (Palm et al. 1996, 1998; Wessel et al. 1998; Clark 1999; Kelder et al. 1999; Norinder et al. 1999; Oprea and Gottfries 1999; Stenberg et al. 1999; Egan et al. 2000; Ekins et al. 2001b; Raevsky et al. 2001; Stenberg et al. 2001; Zhao et al. 2001; Niwa 2003), which now occurs much earlier in drug discovery than perhaps a decade ago or less (Ekins et al. 2000c; Ekins and Rose 2002). These recently developed different computational approaches are undergoing validation yet they represent a means to improve the productivity of the drug discovery process. Owing to their highly parallel nature, computational methods are also the fastest and most cost-effective method for indication of possible toxic consequences caused by interfering with the above multiple proteins (Ekins et al. 2000b), providing molecular insight and suggesting new hypotheses for rapid testing in vitro. This ability to screen large numbers of molecules computationally parallels the increase in throughput of in vitro assays for drug discovery over the past decade. Both in vitro and computational approaches can be used in tandem in an iterative manner to improve the developed models.



Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: IV (Pharmacophore)

Batch: 2019 -2021

A pharmacophore is the representation of the spatial arrangement of structural features that are required for a certain biological activity. Pharmacophore development theory and applications have been explained in detail elsewhere (Guner 2000; Kurogi and Guner 2001; Guner 2002; Guner et al. 2004). Three widely used pharmacophore perception programs, Catalyst, GASP and DISCO, have been thoroughly described and compared by Patel et al. (2002) and the interested reader is referred to their paper for further details of the methods. The ultimate goal of in silico studies in ADME/Tox is to predict the disposition behavior of drugs in the whole body by incorporating all kinetic processes into one global model. However, currently only a very limited number of preliminary models at the protein level have been conducted (Yamashita and Hashida 2004). We are starting to observe a more "systems-based" approach to ADME/Tox as various databases on the interactions of small molecules with proteins are combined with multiple QSAR models and other ADME/Tox tools (Ekins et al. 2005 a, c, d). However, most published studies describe modeling of the individual protein targets related to a single ADME/Tox property, and these models will be the primary focus of this chapter.

The key proteins that have been modeled with such pharmacophore methods include the major cytochrome P450 (CYP) enzymes, UDP-glucuronosyltransfer-

ase (UGT), P-glycoprotein (P-gp), breast cancer resistance protein (BCRP), peptide transporter (PepT1), apical sodium-dependent bile acid transporter (ASBT), sodium taurocholate-transporting polypeptide (NTCP), nucleoside transporter, organic cation transporter (OCT), multiple nuclear hormone receptors including the pregnane X receptor (PXR) and human ether-a-go-go (hERG) potassium channel. We describe below some of the pharmacophore modeling efforts to date in more detail.



Enable | Enlighten | Enrich (Deemed to be University) (Under Section 3 of UGC Act 1956)

KARPAGAM ACADEMY OF HIGHER EDUCATION (Deemed to be University Under Section 3 of UGC Act 1956) COIMBATORE-21

DEPARTMENT OF CHEMISTRY (For the candidates admitted from 2019 & onwards) 19CHP105-C MOLECULAR MODELLING & DRUG DESIGN

Multiple Choice Questions for Unit IV

| S. No | Question | Option 1 | Option 2 | Option 3 | Option 4 | Answer |
|----------|---|-------------------|----------------------|-------------|------------------|----------------------|
| | Unit IV | | | | | |
| 1 | The substance possessing distinct functional group and biological activity is called | Pharmaco phore | Chromophore | Auxochrome | Chemiluminesence | Pharmacophore |
| 2 | The term pharmacophore was called as receptive substance by | . Ehrlich | Langley | Fischer | Marshall | Langley |
| 3 | An example for Pharmacophoric group is | Cyanamide | Toluamide | Sulfonamide | Acetamide | Sulfonamide |
| 4 | The term lock and key concept is designed by | Ehrlich | Fischer | Kier | Marshall | Fischer |
| 5 | Who have pioneered the development of Pharmacophore concept and the application in SAR? | Ehrlich | Kier and Marshall | Fischer | Langley | Kier and Marshall |

| 6 | Based on initial SAR consideration simple 2D moleules was introduced in the year | 2010 | 1990 | 1940 | 1920 | 1940 |
|----|---|--------------------------------|-------------------------|-------------------------------------|--------------------------|---------------------------------|
| 7 | Which one is considered as tools for the discovery of novel molecules. | Enzymes | Pharmacophores | Receptors | Nucleicacids | Pharmacophores |
| 8 | Which is a computational procedure for determining energetically favourable binding sites on molecules of known structure. | GRID | QSAR | QSPR | FlexX | GRID |
| 9 | Who introduced the concept Magic bullet? | Fischer | Ehrlich | Kier | Marshall | Ehrlich |
| 10 | The Lock and Key mechanism is for | Protein complex | Aminoacid | Enzyme- Substrate complex | Polypeptide | Enzyme- Substrate complex |
| 11 | A well developed pharmacophore model gives the information about the | Receptor- binding cavity | Acceptor-binding cavity | Acceptor- Receptor Antagonist | Enzyme-binding cavity | Receptor- binding cavity |
| 12 | The majority of automated pharmacophore generated program do not consider the activity of | Metals | Ligands | Metal- Ligand complex | Enzymes | Ligands |
| 13 | The term pharmacophore was called as receptive | Ehrlich | Langley | Fischer | Marshall | Langley |

| | substance by | | | | | |
|----|--|--------------------------------|------------------------------|-----------------------|--------------------|------------------------------|
| 14 | Who was awarded for his lecture entitled "To Bond or Not to Bond: Chemical versus Physical theories of drug action"? | Edelstein | Robert Boyle | Sigmund Frankel | Arthur Cushny | Edelstein |
| 15 | Drug acting on the heart is called as | Somatic | Cordials | Hepatics | Emetic | Cordials |
| 16 | Drug acting on the liver is | Somatic | Emetic | Hepatics | Cordials | Hepatics |
| 17 | Who introduce the term "receptive substance"? | Robert | Frankel | Langley | Edelstein | Langley |
| 18 | Who introduce the term "receptor"? | Robert | Paul Ehrlich | Cushny | Edelstein | Paul Ehrlich |
| 19 | Adrenaline has a powerful effect on which muscle? | Cardiac | Skeletal | Striatal | Smooth | Cardiac |
| 20 | Adrenaline has a very little effect on which muscle? | Cardiac | Skeletal | Striatal | Smooth | Striatal |
| 21 | If D and L isomers of a drug substance has different pharmacological action means they are called as | Low Chemical specificity | High Chemical specificity | Stereospecifi city | Stereo selectivity | High Chemical specificity |
| 22 | Which is a preferred sub structure for protein binding? | Hexyl motif | Tolyl motif | Biphenyl motif | Glycosyl motif | Biphenyl motif |
| 23 | Which is referred to the dyes which bind selectively to the pathogen without harming the human cells? | Photophore | Magic Bullet | Auxochrome | Chromophore | Magic Bullet |

| 24 | Who introduced this term "the hypothetical agent as <i>Zauberkugel</i> "? | Robert | Paul Ehrlich | Cushny | Edelstein | Paul Ehrlich |
|----|---|--------------------------|------------------------|-----------------------|------------------|------------------------|
| 25 | Who formed an idea that it could be possible to kill specific microbes (such as bacteria) that cause diseases without harming the body itself? | Robert | Paul Ehrlich | Cushny | Edelstein | Paul Ehrlich |
| 26 | The treatment of syphilis is termed as the | First magic bullet | Second Magic bullet | Antibiotics | Antiviral | First magic bullet |
| 27 | Which led to the foundation of the concept of chemotherapy? | First magic bullet | Second Magic bullet | Antibiotics | Antiviral | First magic bullet |
| 28 | In Ehrlich research focused on testing dyes for killing microbes, Which attempt was criticized? | Cobalt dye | Arsenic dye | Copper dye | Manganese dye | Arsenic dye |
| 29 | Who was publicly lampooned as an imaginary "Dr Phantasus"? | Paul Ehrlich | Cushny | Edelstein | Robert | Paul Ehrlich |
| 30 | Who propose a new concept called "side-chain theory? | Paul Ehrlich | Cushny | Edelstein | Robert | Paul Ehrlich |
| 31 | Ehrlich revised the concept of side chain | Acceptor theory | Receptor theory | Multi chain theory | Pharmacokinetics | Receptor theory |

| | theory as | | | | | |
|----|--|-------------------------------|---------------------------|---------------------------------|--------------------------|-----------------------------|
| 32 | Sulfonamide/ sulfa drugs were discovered by | Edelstein | Robert | Gerhard Domagk | Cushny | Gerhard Domagk |
| 33 | Examples of magic bullet is | Congo red | Trypan Red | Acid violet | Malachite green | Trypan Red |
| 34 | Ehrlich developed a new derivative of arsenic compound, which he code-named | Compund4 04 | Compound 505 | Compound 606 | Compound707 | Compound 606 |
| 35 | The Compound 606 was effective against | Viral infection | Malaria infection | Fungi infection | Algae infection | Malaria infection |
| 36 | Ehrlich tested Compound 606, but he did not recognise its effectiveness on | Malaria infected rabbit | Malaria infected mouse | Syphilis- infected rabbit | Malaria infected species | Syphilis-infected rabbit |
| 37 | Chemically Compound 606 is | Calciphena mine | Arsphenamine | Kaliphenami ne | Sodphenamine | Arsphenamine |
| 38 | Compound number 606 was given a trade name | Salvarsan | Methadone | Cocaine | Amphetamines | Salvarsan |
| 39 | Examples of magic bullet is | Congo red | Prontosil Red | Acid violet | Malachite green | Prontosil Red |
| 40 | Salvarsan was commercially introduced in the year | 1900 | 1906 | 1910 | 1918 | 1910 |
| 41 | Compound number 914 was given a trade name | NeoSalvar san | Methadone | Cocaine | Amphetamines | NeoSalvarsan |
| 42 | NeoSalvarsan was commercially introduced in the year | 1906 | 1913 | 1918 | 1920 | 1913 |
| 43 | Neosalvarsan is | Less toxic | Medium toxic | Highly toxic | Potentially toxic | Less toxic |
| 44 | Which was the foundation of Pharmaceutical research? | Pharmacok inetics | Magic bullet | Pharmacody anamics | Pharmacogenomic s | Magic bullet |

| 45 | There is a relation | Robert | Cushny | Fischer | Kiyoshi Shiga | Fischer |
|-----------|-----------------------------|------------|-----------------|------------|---------------|-----------------|
| | between the unknown | | | | | |
| | structure of an active | | | | | |
| | enzyme and that of | | | | | |
| | substrate, they are | | | | | |
| | complementary and the | | | | | |
| | one may be said to fit the | | | | | |
| | other as a key fits a lock. | | | | | |
| | This statement is given | | | | | |
| | by | | | | | |
| 46 | Trypan Red is a | Red Cobalt | Red Arsenic dye | Acid Red | Congo Red dye | Red Arsenic dye |
| | | dye | | dye | | |
| 47 | Trypan Red is used for | Cough | Nausea | Sleeping | Vomitting | Sleeping |
| | the treatment of | | | sickness | | sickness |
| 48 | Kiyoshi Shiga is a | Biochemist | Microbiologist | Therapist | Pharmacist | Microbiologist |
| | Japanese | | | | | |
| 49 | Ehrlich successfully | Robert | Fischer | Kiyoshi | Frankel | Kiyoshi Shiga |
| | prepared Trypan Red dye | | | Shiga | | |
| | with the help of | | | | | |
| 50 | Trypan Red dye and | Selenium | Sulphur | Manganese | Zinc | Sulphur |
| | Prontosil Red dye | | | | | |
| | commonly has | | | | | |
| 51 | Emil Fischer proposed | 1868 | 1878 | 1888 | 1898 | 1898 |
| | Lock and Key model in | | | | | |
| | the year | | | | | |
| 52 | Lock and Key model is | Template | Sequence model | Threading | Folding | Template model |
| | also called as | model | | | | |
| 53 | Enzyme-Substrate | Stable | Highly Unstable | Ionic | Polymer | Highly Unstable |
| | complex is | | | | | |
| 54 | The ES complexes have | ESR | NMR | Electron | IR | Electron |
| | been directly observed by | | | microscopy | | microscopy |
| 55 | Upon formation of ES | Change | Not change | Be same | Be identical | Change |
| | complex, the physical | frequently | | | | frequently |

| | - | | • | | | |
|----|----------------------------|-------------|-----------------|---------------|--------------------|-----------------|
| | property of enzyme will | | | | | |
| 56 | A most general evidence | Increases | Increases with | Decreases | Decreases | Increases with |
| | for the existence of ES | with | Decrease in the | with Increase | exponentially with | Increase in the |
| | complexes is the fact that | Increase in | substrate | in the | Increase in the | substrate |
| | at a constant | the | concentration | substrate | substrate | concentration |
| | concentration of enzyme, | substrate | | concentratio | concentration | |
| | the reaction rate | concentrat | | n | | |
| | | ion | | | | |
| 57 | Which is an estrogenic | trans- | trans-Dimethyl | trans- | trans-Diethyl | trans-Diethyl |
| | agent? | Diethyl | acetylene | Diethyl | cinnamate | stilbestrol |
| | | Maleate | dicarboxylate | stilbestrol | | |
| 58 | Sulfonamides and PABA | Polymers | 2-D- | 3-D- | Antioxidants | 2-D- |
| | are | | Pharmacophores | Pharmacoph | | Pharmacophores |
| | | | | ores | | |
| 59 | GBPM means | GRID | Group Based | Genetic | Grip Bio Polymer | GRID based |
| | | based | Pharmacophore | Based | model | Pharmcophore |
| | | Pharmcop | model | Protein | | model |
| | | hore | | model | | |
| | | model | | | | |
| 60 | DDI means | Drug | Drug-Drug | Drug | Drug Dose | Drug-Drug |
| | | Dosage | Interaction | Disease | Immune power | Interaction |
| | | Interaction | | Interaction | | |



Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: V (Computer aided Chemistry)

Batch: 2019 -2021

18CHP105-C MOLECULAR MODELLING & DRUG DESIGN

UNIT V

Computer aided Chemistry: Structure Prediction and Drug Design: Introduction to molecular docking, rigid docking, Flexible docking, manual docking, advantage and disadvantage of flex-X, flex-S, AUTODOCK and other docking software, scoring functions, simple interaction energies, GB/SA scoring (implicit solvation), CScore (consensus scoring algorithms).



Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: V (Computer aided Chemistry)

Batch: 2019 -2021

Computer aided Chemistry: Structure Prediction and Drug Design:

Molecular Docking

Molecular docking is the process that involves placing molecules in appropriate configurations to interact with a receptor. Molecular docking is a natural process which occurs within seconds in a cell. 行 In molecular modeling the term "molecular docking" refers to the study of how two or more molecular structures fit together



Molecular Docking Models

Over the years biochemists have developed numerous models to capture the key elements of the molecular recognition process. Although very simplified, these models have proven highly useful to the scientific community.



Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: V (Computer aided Chemistry)

Batch: 2019 -2021

| year | model | author(s) | | | |
|--|-----------------------|------------------|--|--|--|
| 1890 | lock-and-key | Emil Fischer | | | |
| 1958 | induced-fit | Daniel Koshland | | | |
| 2003 | conformation ensemble | Buyong Ma et al. | | | |
| " All models are wrong, some are useful " (George Box) | | | | | |

Rigid Docking Methods

If we assume that the molecules are rigid, we are then looking for a transformation in 3D space of one of the molecules which brings it into optimal fit with the other molecule in terms of a scoring function

In rigid-body docking, the search space is restricted to three rotational and three translational degrees of freedom



Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: V (Computer aided Chemistry)



Degrees of Freedom in Flexible Docking

The rigid-body docking approaches are often not sufficient to predict the structure of a protein complex from the separate unbound structures

The incorporation of molecular flexibility into docking algorithms requires to add conformational degrees of freedom to translations and rotations

Approximation algorithms need to be introduced to reduce the dimensionality of the problem and produce acceptable results within a reasonable computing time



Course Name: Molecular Modelling and Drug Design Class: I- M.Sc (Chemistry)

Course Code: 19CHP105-C

Unit: V (Computer aided Chemistry) Batch: 2019 - 2021 position orientation P = 3 + 3rigid docking flexible docking $P = 3 + 3 + N_r$ intractable problem number of parameters number of flexible bonds to optimize rigid docking flexible docking flexible docking 6 ~ 1500 ~ 3000 protein-ligand protein-protein Manual docking Molecular docking

Molecular docking can be divided into two separate sections.

1) Search algorithm - The algorithm should create an optimum number of configurations that include the experimentally determined binding modes. The following are the various algorithms used for docking analysis.

- Molecular dynamics
- Monte Carlo methods
- Genetic algorithms
- Fragment-based methods
- Point complementary methods



Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: V (Computer aided Chemistry) Batch:

Batch: 2019 -2021

Distance geometry methods

Systematic searches

2) Scoring Function –These are mathematical methods used to predict the strength of the noncovalent interaction called as binding affinity, between two molecules after they have been docked. Scoring functions have also been developed to predict the strength of other types of intermolecular interactions, for example between two proteins or between protein and DNA or

protein and drug. These configurations are evaluated using scoring functions to distinguish the experimental binding modes from all other modes explored through the searching algorithm.

• Empirical scoring function of Igendock

Fitness = vdW + Hbond + Elec

Binding Energy

 $\Delta G_{bind} = \Delta G_{vdw} + \Delta G_{hbond} + \Delta G_{elect} + \Delta G_{conform} + \Delta G_{tor} + \Delta G_{sol}$

General concept of the algorithm:

1) A 'negative' image of the binding site is made - a collection of spheres of varying radii, each of which touches the molecular surface at just 2 points.





Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: V (Computer aided Chemistry)

Batch: 2019 -2021



2) Ligand atoms are then matched to sphere centers where at least four distances between ligand atoms are matched to sphere center-sphere center distances.





Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: V (Computer aided Chemistry)



- 3) Proper orientation is achieved by a least squares fit of ligand atoms to the sphere centers.
- 4) Orientation is checked for any steric clashes between ligand and receptor.
- 5) If acceptable, then interaction energy is computed as a 'score' for that binding mode
- 6) New orientations are obtained by matching different sets of atoms and sphere centers
- 7) Top-scoring orientations are retained for subsequent analysis

Types of Docking -

The following are majorly used method for docking-



Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Class: I- M.Sc (Chemistry)

Unit: V (Computer aided Chemistry) Batch: 2019 -2021

- Lock and Key\Rigid Docking In rigid docking, both the internal geometry of the receptor and ligand is kept fixed and docking is performed.
- Induced fit\Flexible Docking An enumeration on the rotations of one of the molecules (usually smaller one) is performed. Every rotation the surface cell occupancy and energy is calculated; later the most optimum pose is selected.

Major steps in molecular docking:

Step I – Building the Receptor

This step the 3D structure of the receptor should be considered which can be downloaded from PDB; later the available structure should be processed. This should include removal of the water molecules from the cavity, stabilizing the charges, filling the missing residues, generation the side chains etc according to the parameters available. The receptor should be biological active and stable state.

Step II – Identification of the Active Site

After the receptor is built, the active site within the receptor should be identified. The receptor may have many active sites but the one of the interest should be selected. Most of the water molecules and heteroatom if present should be removed.

Step III – Ligand Preparation

Ligands can be obtained from various databases like ZINC, PubChem or can be sketched using tools Chemsketch. While selecting the ligand, the LIPINSKY'S RULE OF 5 should be applied. The rule is important for drug development where a pharmacologically active lead structure is optimized stepwise for increased activity and selectivity, as well as drug-like properties as described.

For selection of a ligand according to the LIPINSKY'S RULE:



Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: V (Computer aided Chemistry) Batch

Batch: 2019 -2021

Not more than 5 –H bond donors.

Molecular Weight NOT more than 500 Da.

Log P not over 5.

NOT more than 10 H bond acceptors.

Step IV- Docking

This is the last step, where the ligand is docked onto the receptor and the interactions are checked. The scoring function generates score depending on which the best fit ligand is selected.

Softwares available for Molecular Docking:

SANJEEVINI.

SCHRODINGER

DOCK

AUTOLOCK TOOLS.

DISCOVERY STUDIO.

iGemDock

Autodock

- Interaction between biomolecules lie at the core of all metabolic processes and life activities
- The number of solved protein structures available in the databases is expanding exponentially

Prepared by Dr. E. Karthikeyan, Asst. Prof., Department of Chemistry, KAHE Page 10/17



Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Class: I- M.Sc (Chemistry)

Unit: V (Computer aided Chemistry)

Batch: 2019 -2021

Page

- To understand their functions it is essential to elucidate the interaction mechanisms between the different molecules
- Primary importance lies in rational drug design
- Depending upon the success of the docked molecules the docking ligand may be redesigned or its structure further refined.
- Also important in the area of immunology to study antigen-antibody interaction.

The AutoDock Software

- Developed by AJ Olson's group in 1990.
- AutoDock uses free energy of the docking molecules using 3D potential-grids
- Uses heuristic search to minimize the energy. Search Algorithms used:
 - Simulated Annealing
 - Genetic Algorithm

Lamarckian GA (GA+LS hybrid)

Algorithms Overview

Simulated Annealing

- Based on temperature effects
- Start with high temperature and global search
- Lower temperature local search

Genetic Algorithm

- Charles Darwin's Theory of Evolution
- Genotype to Phenotype
- Lamarckian Algorithm (Jean –Baptiste de Lamarck)
- Phenotype to Genotype

A number of docking programs have been developed during the last two decades and made available to academic institutions at little or no charge. Table 1 summarizes current proteinligand docking tools. Basic characteristics such as supported platforms, license terms, as well as applied docking algorithms and scoring functions are presented.



Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: V (Computer aided Chemistry)

Batch: 2019 -2021

Table 1. Basic characteristics for current protein-ligand docking tools*.

| Entry | Program Ref ^{**} | Designer / Company | Licence terms | Supported platforms | Docking approach | Scoring function |
|-------|------------------------------|--|---|---|--|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 1 | AutoDock [5] | D. S. Goodsell and A. J. Olson The Scripps Research Institute | Free for academic use | Unix, Mac OSX, Linux, SGI | Genetic algorithm Lamarckian genetic algorithm Simulated annealing | AutoDock (force-field methods) |
| 2 | DOCK [6] | I. Kuntz University of California, San Francisco | Free for academic use | Unix, Linux, Sun, IBM AIX, Mac OSX, Windows | Shape fitting (sphere sets) | ChemScore, GB/SA solvation scoring, other |
| 3 | FlexX [7] | T. Lengauer and M. Rarey BioSolveIT | Commercial Free evaluation (6 weeks) | Unix, Linux, SGI, Sun Windows, | Incremental construction | FlexXScore, PLP, ScreenScore, DrugScore |
| 4 | FRED [8] | OpenEye Scientific Software | Free for academic use | Unix, Linux, SGI, Mac OSX, IBM AIX, Windows | Shape fitting (Gaussian) | ScreenScore, PLP, Gaussian shape score, user defined |
| 5 | Glide [9] | Schrödinger Inc. | Commercial | Unix, Linux, SGI, IBM AIX | Monte Carlo sampling | GlideScore, GlideComp |
| 6 | GOLD [10] | Cambridge Crystallographic Data Centre | Commercial Free evaluation (2 months) | Linux, SGI, Sun, IBM, Windows | Genetic algorithm | GoldScore, ChemScore user defined |
| 7 | LigandFit [11] | Accelrys Inc. | Commercial | Linux, SGI, IBM AIX | Monte Carlo sampling | LigScore, PLP, PMF |

*Other current docking tools are: ICM [12], ProDock [13], QXP [14], Slide [15], Surflex [16]. *Internet addresses of selected home pages are given [17].

Scoring Functions:

Concept and Application Scoring function is the most important component in structurebased drug design for evaluating the efficacy of ligands binding to their target proteins. In molecular docking experiments, protein-ligand complexes need to be rapidly and accurately assessed. As molecular docking experiments generate thousands of ligand binding

Prepared by Dr. E. Karthikeyan, Asst. Prof., Department of Chemistry, KAHE Page 12/17



Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: V (Computer aided Chemistry)

Batch: 2019 -2021

orientations/conformations, scoring functions are used to rank these complexes and differentiate the accurate binding mode predictions from inaccurate predictions. The goal of an ideal scoring function is to rank the complex as determined empirically. Additionally, the scoring function should be able to predict the absolute binding affinity of the complex to facilitate identification of the potential hits/lead candidates against any therapeutic target from a large library of compounds as used in virtual screening. Scoring functions are very helpful to screening libraries of compounds or individual compounds based on their binding mode and affinity. Over the years, various scoring functions that exhibit different accuracies and computational efficiencies have been developed. In this section, we briefly review the scoring functions in literature developed for protein-ligand interactions in molecular docking. Fig. 2 shows different scoring functions currently in use. Scoring functions have been categorized into four different types: 1. Force-field or molecular mechanics-based scoring functions. 2. Empirical scoring functions. 3. Knowledge-based scoring functions. 4. Consensus scoring functions.

Force-Field or Molecular Mechanics-Based Scoring Functions

Classic molecular mechanics are used by force-field scoring functions for energy calculations. These scoring functions use various physical features such as van der Waals (VDW) interactions, electrostatic interactions, and bond stretching/bending/torsional forces. Force-field or molecular mechanics-based scoring functions utilize parameters derived from both experimental and *ab initio* quantum mechanical calculations. These scoring functionsestimate the binding free energy of protein-ligand complexes by the sum of the van der Waals (VDW) interactions and electrostatic electrostatic interactions. Despite its various successful applications, a major challenge associated with force field scoring functions is their inability to treat solvent molecules in ligand binding. To overcome this shortcoming, variables from the empirical scoring functions are often taken into consideration along with force-field functions.

Empirical Scoring Functions

These scoring functions are based on counting the number of different types of interactions between two binding partners. These functions count the number of atoms within a ligand and receptor that are in contact with each other or calculate changes in the solvent accessible surface area (SASA) in the complex and the uncomplexed structure of the protein and ligand. These interaction terms of the function may include favorable contacts (hydrophobichydrophobic), unfavorable contacts (hydrophobic-hydrophilic), favorable contributions to affinity (especially if shielded from solvent), no contribution if solvent exposed (number of hydrogen bonds), and unfavorable conformational entropy contribution (number of rotatable bonds immobilized in complex formation).

Knowledge-Based Scoring

This scoring function attempts to capture knowledge about the receptor (target) - ligand binding available in the protein data bank (PDB) by statistical analysis of structural data alone. Frequency of occurrence of individual contacts is assumed to measure their energetic contribution to the binding. A specific contact that occurs more frequently than an average or random distribution indicates attractive interaction, whereas less frequent occurrence indicates repulsive interaction, e.g., PMF score (potentials of mean force).



Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: V (Computer aided Chemistry)

Batch: 2019 -2021

Consensus Scoring Function

Despite the availability of some good scoring functions, consensus scoring functions have been developed. Every scoring function currently in use has some limitations and advantages. The consensus scoring function was developed while considering the advantages of different scoring functions to achieve high accuracy. Consensus scoring functions, which are the most advanced scoring technique, improve the probability of finding the correct solution via a combination of different scoring functions. The best aspect of consensus scoring functions is their ability to score predicted binding poses using different scoring functions. Commonly used consensus scoring strategies include:

(1) Weighted combinations of scoring functions, vote by number strategy.

(2) Vote by number strategy in which a cutoff value is established for each scoring method used and the final decision is made based on the number of passes a molecule has

(3) Rank by number strategy in which each compound is ranked by its average normalized score.
(4) Rank by rank strategy in which the compounds are sorted on the basis of their average rank and predicted by individual scoring functions.

Simple interaction energies

In the course of a calculation the total energy is minimized with respect to the atomic coordinates, and it consists of a sum of different contributions that compute the deviations from equilibrium of bond lengths, angles, torsions and non-bonded interactions:

$E_{tot} = E_{str} + E_{bend} + E_{tors} + E_{vdw} + E_{elec} + \dots$

where E_{tot} is the total energy of the molecule, E_{str} is the bond-stretching energy term, E_{bend} is the angle-bending energy term, E_{tors} is the torsional energy term, E_{vdw} is the van der Waals energy term, and E_{elec} is the electrostatic energy term. The equilibrium values of bond lengths and bond angles are the corresponding force constants used in the potential energy function in the force field and it defines a set known as force field parameters. Each deviation from these equilibrium values will result in increasing total energy of the molecule. So, the total energy is a measure of intramolecular strain relative to a hypothetical molecule with an ideal geometry of equilibrium. By itself the total energy has no strict physical meaning, but differences in total energy between two different conformations of the same molecule can be compared.¹⁶⁻¹⁹

Energy-Minimizing Procedures

Prepared by Dr. E. Karthikeyan, Asst. Prof., Department of Chemistry, KAHE Page 14/17



Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: V (Computer aided Chemistry) Batch: 2019 -2021

Energy minimization methods can be divided into different classes depending on the order of the derivative used for locating a minimum on the energy surface. Zero order methods are those that only use the energy function to identify the regions of low energy through a grid search procedure. The most well-known method of this kind is the SIMPLEX method. Within first-derivative techniques, there are several procedures like the steepest descent method or the conjugate gradient method that make use of the gradient of the function. Second-derivative methods, like the Newton-Raphson algorithm make use of the hessian to locate minima.^{20,21}

<u>PBSA</u>

Predictions of solvation free energies using PBE models are often augmented with a term that describes the non-polar contribution to solvation. The additional non-polar term is usually estimated from the solvent-accessible surface area (SA), and PBE models used in conjunction with SA are referred to as PBSA models [40,41]. The non-polar solvation free energy $\Delta G_{sol}^{non-pol}$ that accounts for the cost of cavity formation and Van der Waals interactions with the solvent can be approximated using an equation of the following form:

 $\Delta G_{sol}^{non-pol} = \gamma A + b \tag{11}$

with A representing the solvent-accessible surface area, and γ and b being the adjustable parameters. The values of γ and b are usually calibrated using experimentally determined solvation energies of small molecules.

With recent improvements in calculating the polar contributions of solvation free energies using MD simulations in conjunction with PBE/GB models (see the next section), the limitations of this simple SA model (Eq.11) are becoming more obvious [42-44]. New efforts to develop an extended treatment of non-polar solvation free energies, especially by separating the costs of Van der Waals interactions with the solvent and costs of cavity formations, have been constructed [45] and will hopefully increase the level of accuracy of current continuum models.



Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: V (Computer aided Chemistry) Batch: 2019 -2021

<u>GB formalism</u>

Binding free energies of protein-ligand complexes can be evaluated using the PBE (or PBSA) models in minutes of computer time. However, the screening of thousands of potential drugs usually requires higher computational efficiency than the PBE model can deliver. Therefore, very fast analytical methods based on the Born equation (Eq.9), Generalized Born (GB) models, have been widely used in the calculation of free energies of ligand binding, in computer-aided drug design (CADD), and in conformational analysis of proteins.

The GB models for ligand binding are based on the assumption that the screening of electrostatic interactions between two charges can be estimated by the degree to which the charges interact with the solvent [46]. The solutes in the GB models are described by collection of atoms, where each atom is represented by a sphere of a radius a_i , referred to as "effective Born solvation radius". The point charge of each atom q_i is positioned at the sphere's centre. Like in the PBE models, the molecule in the GB model is described as a low-dielectric volume ε_p surrounded by a high-dielectric aqueous environment ε_w . Electrostatic interaction energies between point charges in GB models are calculated using Coulomb interactions *in vacuo* (Eq.2). The electrostatic solvation energy ("polarization energy") of transferring a molecule from the protein to the aqueous environment is given by [40]:

$$\Delta G_{pol} = -166kcal/mol(\frac{1}{\varepsilon_p} - \frac{1}{\varepsilon_w})\sum_{i}^{N}\sum_{j}^{N}\frac{q_i q_j}{f_{GB}(r_{ij}, a_i, a_j)}$$
(12)

where N is the number of atoms in the system, r_{ij} is the distance between atoms *i* and *j*, and f_{GB} is the function that involves several empirical parameters. The effective Born solvation radius a_i of the atom *i* can be thought as the distance between the charge and protein-solvent boundary. This parameter is adjustable and its value is usually determined using the solvation free energy from the PBE by:



Class: I- M.Sc (Chemistry)

Course Name: Molecular Modelling and Drug Design

Course Code: 19CHP105-C

Unit: V (Computer aided Chemistry) Batch

Batch: 2019 -2021

$$a_i = -166 \text{\AA}(\frac{1}{\varepsilon_p} - \frac{1}{\varepsilon_w}) \frac{1}{\Delta G_{pol}^i}$$
(13)

where ΔG_{pol}^{i} designates the solvation free energy of the unit charge positioned at the centre of the atom *i* in the uncharged solute, whereas ε_{p} and ε_{w} represent relative dielectric constants of the solute and the solution in the PBE.

GB-based models are unable to calculate electrostatic potential maps of proteins and their parameters are usually optimized using the results obtained with PBE solvers [47-51]. Optimized and calibrated GB models are usually augmented by the hydrophobic contribution to binding and referred to as GBSA models. With improvements in recent years, current GBSA models are capable of achieving the similar level of accuracy as numerical PBSA methods [42,52]. Their high computational efficiency makes them suitable for drug-design applications and different versions of GBSA models are implemented now in many ligand-docking programs [53-56].





Enable | Enlighten | Enrich (Deemed to be University) (Under Section 3 of UGC Act 1956)

KARPAGAM ACADEMY OF HIGHER EDUCATION (Deemed to be University Under Section 3 of UGC Act 1956) COIMBATORE-21

DEPARTMENT OF CHEMISTRY (For the candidates admitted from 2019 & onwards) 19CHP105-C MOLECULAR MODELLING & DRUG DESIGN

Multiple Choice Questions for Unit V

| S. No | Question | Option 1 | Option 2 | Option 3 | Option 4 | Answer |
|-------|--|-----------------------------------|---------------------------------------|---------------------------|--|-----------------------------------|
| | Unit V | | | | | |
| 1 | Which is designed to predict how small molecules such as substrates, bind to a receptor of known 3D structure. | AutoDock | GRID | FlexX | QSAR | AutoDock |
| 2 | Whch is a key tool in structural molecular biology and computer assisted drug design. | AutoDocking | Rigiddockin g | Molecular docking | Flexible docking | Molecular docking |
| 3 | In which docking molecules cannot change their spatial shape during the docking process. | Flexible | Rigid | Molecular | Autodockin g | Rigid |
| 4 | Computer-aided drug design, often called | Structure based drug design | Conformatio n based drug design | Atom based drug design | Reactivit y based drug design | Structure based drug design |

| 5 | Which predicts the binding orientation and affinity of a ligand to target. | Docking | AADS | ADAM | FlexS | Docking |
|----|---|----------------------------------|-----------------------------------|------------------------------------|------------------------------------|------------------------------------|
| 6 | Computer aided drug design involves the biochemical information of | Protein-Ligand interaction | Enzyme- Protein interaction | Protein- Protein interaction | Ligand- Receptor interaction | Ligand- Receptor interaction |
| 7 | Docking can be classified as Rigid and flexible based on | Ligand- protein flexibilty | Ligand flexibility | Protein flexibility | Structure of ligand | Ligand- protein flexibilty |
| 8 | The most common software packages used forligand based pharmacophore generation include | GRID | DISCO | GRIN | GRIP | DISCO |
| 9 | Which is used to describe the binding constant between probe atoms and the target. | SCORE | MOE | DISCO | GASP | SCORE |
| 10 | The process that involves placing molecules in appropriate configurations to interact with a receptor is called | Auto Docking | Molecular Docking | Rigid Docking | Flexible Docking | Molecular Docking |
| 11 | Which docking is a natural process which occurs within seconds in a cell? | Auto Docking | Molecular Docking | Rigid Docking | Flexible Docking | Molecular Docking |
| 12 | The author of Induced-Fit model is | Emil Fischer | Buyong Ma et al | Daniel Koshland | Robert | Daniel Koshland |
| 13 | The author of Conformation ensemble model is | Emil Fischer | Buyong Ma et al | Daniel Koshland | Robert | Buyong Ma et al |
| 14 | Lock and Key model was introduced in the year | 1928 | 1890 | 1958 | 1968 | 1890 |
| 15 | In molecular modeling which | Molecular | Rigid | Flexible | Auto | Molecular |

| | term refers to the study of how two or more molecular structures fit together? | Docking | Docking | Docking | Docking | Docking |
|----|---|--|---|--|---|--|
| 16 | The Induced-Fit model was introduced in the year | 1958 | 1998 | 1989 | 1969 | 1958 |
| 17 | The Conformation ensemble model was introduced in the year | 1998 | 2003 | 2008 | 2010 | 2003 |
| 18 | In rigid body docking, the search space is restricted to | Two rotational and Three translational degree of freedom | Two rotational and Two translational degree of freedom | Three rotational and Three translationa l degree of freedom | Three rotational and Two translational degree of freedom | Three rotational and Three translationa l degree of freedom |
| 19 | The approach of which docking is not sufficient to predict the structure of a protein complex. | Auto Docking | Molecular Docking | Rigid Docking | Flexible Docking | Rigid Docking |
| 20 | The incorporation of molecular flexibility into docking algorithm requires to add which degree of freedom? | Vibrational | Conformati onal | Translationa l | Rotational | Conformati onal |
| 21 | Flexible docking $P = 3+3+N_r$, P refers to | Number of Parameters optimized | Degree of freedom | Number of Parameters to optimize | Number of Parameters required | Number of Parameters to optimize |
| 22 | Flexible docking $P = 3+3+N_r$, N _r refers to | Number of rigid bonds | Number of flexible bonds | Number of reference bonds | Number of Proteins | Number of flexible bonds |
| 23 | Rigid docking $P = 3+3$, first 3 refers to | Position | Orientation | Number of bonds | Number of proteins | Position |
| 24 | Rigid docking $P = 3+3$, second 3 refers to | Position | Orientation | Number of bonds | Number of proteins | Orientation |

| 25 | Molecular docking can be divided into How many separate sections? | 2 | 3 | 4 | 5 | 2 |
|----|---|------------------------|------------------------|--------------------------------|-----------------------------|--------------------------------|
| 26 | Molecular dynamics, Mont carlo method, Genetic algorithm etc are the various algorithm used for | Scoring function | Docking analysis | Score interpretatio n | Building the receptor | Docking analysis |
| 27 | Fragment based method, Point complementary methods, Distance geometry method etc are the various algorithm used for | Scoring function | Docking analysis | Score interpretatio n | Building the receptor | Docking analysis |
| 28 | The Search algorithm and Scoring function are the parts of | Auto Docking | Molecular Docking | Rigid Docking | Flexible Docking | Molecular Docking |
| 29 | Scoring function is used to predict the strength of | Covalent interaction | Ionic interaction | Non covalent interaction | Dipole interaction | Non covalent interaction |
| 30 | Scoring function is a | Mathematical method | Computatio nal method | Physical method | Mechanical method | Mathematic al method |
| 31 | Scoring function is used to predict the strength of Non covalent interaction called as | Bond order | Binding affinity | Bond energy | Dissociation energy | Binding affinity |
| 32 | The intermolecular interaction between two proteins is studied by | Molecular dynamics | Monte carlo methods | Scoring function | Fragment based method | Scoring function |
| 33 | The intermolecular interaction between protein and DNA is studied by | Molecular dynamics | Monte carlo methods | Scoring function | Fragment based method | Scoring function |
| 34 | The intermolecular interaction between protein and drug is studied by | Molecular dynamics | Monte carlo methods | Scoring function | Fragment based method | Scoring function |
| 35 | Empirical Scoring function of | vdW+Hbond+ | vdW+Hbon | vdW-Hbond | vdW+Elec | vdW+Hbon |

| | Igemdock is given by Fitness= | Elec | d | +Elec | | d+Elec |
|----|---|--------------|----------------------|---------------------|---------------------|---------------------|
| 36 | Lock and Key is also called | Auto Docking | Molecular Docking | Rigid Docking | Flexible Docking | Rigid Docking |
| 37 | Induced fit is also | Auto Docking | Molecular Docking | Rigid Docking | Flexible Docking | Flexible Docking |
| 38 | Both the internal geometry of receptor and ligand is kept fixed and docking is performed in which docking? | Auto Docking | Molecular Docking | Rigid Docking | Flexible Docking | Rigid Docking |
| 39 | In which docking? An enumeration on the rotations of one of the molecules is performed. Every rotation the surface cell occupancy and energy is calculated and most optimum pose is selected. | Auto Docking | Molecular Docking | Rigid Docking | Flexible Docking | Flexible Docking |
| 40 | In receptor, How many active sites are present? | 1 | 2 | 3 | Many | Many |
| 41 | For ligand preparation, ligand can be obtained from the data base like | DB2 | ZINC | ORACLE | ADABAS | ZINC |
| 42 | To select the ligand, which rule should be applied? | Veber rule | MDDR rule | LIPINSKY' S Rule | BBB rule | LIPINSKY' S Rule |
| 43 | Which rule states that ligand should not have more than 5-H bond donors? | Veber rule | MDDR rule | LIPINSKY' S Rule | BBB rule | LIPINSKY' S Rule |
| 44 | Which rule states that ligand should not have more than 10- H bond acceptors? | Veber rule | MDDR rule | LIPINSKY' S Rule | BBB rule | LIPINSKY' S Rule |
| 45 | Which rule states that ligand should not have Molecular | Veber rule | MDDR rule | LIPINSKY' S Rule | BBB rule | LIPINSKY' S Rule |

| | weight more than 500. | | | | | |
|----|--|----------------------|----------------------|-------------------------------|---------------------|----------------------|
| 46 | For ligand preparation, ligand can be obtained from the data base like | PubChem | ADABAS | IBMDB2 | SQLite | PubChem |
| 47 | SCHRODINGER is the software for | Auto Docking | Molecular Docking | Rigid Docking | Flexible Docking | Molecular Docking |
| 48 | Interaction between biomolecules lie at the core of all metabolic processes and life activities. This called | Auto Docking | Molecular Docking | Rigid Docking | Flexible Docking | Auto Docking |
| 49 | The number of solved protein structures available in the databases is expanding exponentially in | Auto Docking | Molecular Docking | Rigid Docking | Flexible Docking | Auto Docking |
| 50 | Auto docking software is developed by | Evan Miller group | AJ Olson's group | Jo Erskine Hannay group | Donald group | AJ Olson's group |
| 51 | Simulated Annealing is the algorithm of | Auto Docking | Molecular Docking | Rigid Docking | Flexible Docking | Auto Docking |
| 52 | Protein-Ligand complexes are rapidly and accurately assessed in | Auto Docking | Molecular Docking | Rigid Docking | Flexible Docking | Molecular Docking |
| 53 | Which is used to rank these complexes and differentiate the accurate binding mode predictions from inaccurate predictions? | Auto Docking | Molecular Docking | Rigid Docking | Scoring function | Scoring function |
| 54 | Which is very helpful to screening libraries of | Auto Docking | Molecular Docking | Rigid Docking | Scoring function | Scoring function |

| | compounds or individual compounds based on their binding mode and affinity? | | | | | |
|----|--|---|----------------------------------|---|----------------------------------|--|
| 55 | The scoring functions in literature developed for protein-ligand interactions is in | Auto Docking | Molecular docking | Rigid Docking | Flexible Docking | Molecular docking |
| 56 | Scoring functions have been categorized into how many different types? | 1 | 2 | 3 | 4 | 4 |
| 57 | Which scoring functions utilize parameters derived from both experimental and <i>ab</i> <i>initio</i> quantum mechanical calculations? | Forcefield based scoring function | Empirical scoring function | Knowledge based scoring function | Consensus scoring function | Forcefield based scoring function |
| 58 | Which scoring functions are based on counting the number of different types of interactions between two binding partners? | Forcefield based scoring function | Empirical scoring function | Knowledge based scoring function | Consensus scoring function | Empirical scoring function |
| 59 | Which scoring function attempts to capture knowledge about the receptor (target) - ligand binding available in the protein data bank (PDB) by statistical analysis of structural data alone? | Forcefield based scoring function | Empirical scoring function | Knowledge based scoring function | Consensus scoring function | Knowledge based scoring function |
| 60 | Which scoring function was developed while considering the advantages of different scoring functions to achieve high accuracy? | Forcefield based scoring function | Empirical scoring function | Knowledge based scoring function | Consensus scoring function | Consensus scoring function |