DEPTH OF BIOLOGY

Unit-5

QSAR Statistical methods - regression analysis, partial least

square analysis (PLS) and other multivariate statistical methods.

3D-QSAR approaches like COMFA and COMSIA
Prodrug design-Basic concept, Prodrugs to improve patient acceptability, Drug solubility, Drug absorption and distribution, site specific drug delivery and sustained drug action.

Rationale of
prodrug design and practical consideration of prodrug design

Made with Goodnotes

DEPTH OF BIOLOGY

QSAR means Quantitative Structure–Activity Relationship.

It is a mathematical/statistical method to find a relationship between the chemical structure of compounds and their biological activity.

Purpose: To predict activity of new compounds before synthesis and guide drug design.

Why Statistics in QSAR?

Biological activity (like IC50, MIC, binding affinity) is influenced by many structural factors (hydrophobicity, electronic properties, steric effects).
To connect these structural descriptors with biological data, statistical methods are used.
They help in:
Building models
Testing accuracy
Predicting unknown activities

DEPTH OF BIOLOGY

General Steps in QSAR Statistical Analysis-

-Data Collection

Collect activity data (e.g., IC50 values) of a series of compounds.

Convert them into log form (like pIC50).

-Descriptor Calculation

Calculate molecular descriptors (physicochemical properties, electronic, hydrophobic, steric).

-Statistical Modeling

Use statistical methods to find a correlation between descriptors and activity.

Made with Goodnotes

DEPTH OF BIOLOGY

-Model Validation

Test the reliability of the model (training/test sets, cross-validation).

-Prediction

Apply the model to predict activity of new molecules.

-Main QSAR Statistical Methods-

A. Regression Methods

These are the most common in QSAR.

Simple Linear Regression (SLR)

Relates one descriptor to activity.

Made with Goodnotes

DEPTH OF BIOLOGY

Equation:

$Y = a + bX$

$Y = a + bX$ where
Y = biological activity
X = molecular descriptor
a = intercept, b = slope

Example: activity depends only on lipophilicity (logP).

Made with Goodnotes

Multiple Linear Regression (MLR)-

Relates multiple descriptors to activity.

Equation:

$$Y = a + b1\ X1 + b2\ X2 + b3\ X3 + \ldots + bn\ Xn$$

$$Y = a + b1\ X1 + b2\ X2 + b3\ X3 + \ldots + bn\ Xn$$

Y = activity

X1, X2... = descriptors

b1, b2... = regression coefficients

Very popular in QSAR.

Made with Goodnotes

DEPTH OF BIOLOGY

Regression Analysis-

It is a method to connect structure with activity.

We check:
-If we change something in the chemical structure → how does the biological activity change?

So, regression gives a formula (equation) to explain this.

Why Use in QSAR?

Drugs have different properties (like size, shape, charge, solubility).
These properties affect how strongly they work.
Regression tells us which property is important and how much it affects activity.

Made with GoodNotes

DEPTH OF BIOLOGY

Types of Regression-

(A) Simple Linear Regression

Use only one property (descriptor) at a time.

Example: "Only lipophilicity (oil–water solubility) controls activity."

Result: A straight-line relation (when property ↑, activity ↑ or ↓).

(B) Multiple Linear Regression

Use two or more properties together.

Example: "Activity depends on lipophilicity + size + electronic effect."

More realistic because drugs usually depend on many factors.

Made with Goodnotes

DEPTH OF BIOLOGY

(C) Non-Linear Regression-
When the relationship is curved, not straight.

Example:
Little lipophilicity → low activity.

Medium lipophilicity → highest activity.

Too much lipophilicity → again low activity.

This gives a curve (U-shape or parabola), not a straight line.

Made with Goodnotes

DEPTH OF BIOLOGY

How It Works-

Collect activity values of drugs (like IC_{50}).

Collect their properties (size, charge, solubility, etc.).

Use regression to make an equation connecting activity with properties.

Test if equation is correct (validation).

Use that equation to predict activity of new drugs.

Made with Goodnotes

DEPTH OF BIOLOGY

Example-

Suppose we have 5 molecules.

We measure their lipophilicity (X) and activity (Y).

Regression says:

"If lipophilicity increases by 1 unit → activity increases by 2 units."

This becomes an equation → easy to use for predictions.

Why It Is Important

Saves time and money in drug design.

Helps to know which chemical feature is useful.

Gives a mathematical model to design better drugs.

Made with Goodnotes

Partial Least Squares (PLS) Analysis in QSAR

PLS is a statistical method used in QSAR when we have:
Many descriptors (X variables).
These descriptors are often correlated (related to each other).
Normal regression (MLR) fails in this situation because too many descriptors cause errors.
PLS solves this problem by reducing data and then building a model.

In simple words:
PLS = A method that compresses many descriptors into fewer "new variables" (components) and then relates them to activity.

Made with Goodnotes

DEPTH OF BIOLOGY

Why We Use PLS in QSAR-

Drug molecules can have hundreds of descriptors (hydrophobicity, steric, electronic, topological, etc.).

Many of them overlap or give similar information.

PLS removes redundancy and keeps only the most important parts.
Works better than MLR when data is large and complicated.

Made with Goodnotes

How PLS Works -

Start with data: descriptors (X) and activity values (Y).
Reduce X → PLS combines many descriptors into fewer latent variables (components).

Example: 100 descriptors → compressed into 3–5 components.

Make a regression model between these new components and activity (Y).
Validate the model → check prediction ability.

Predict new activities using this model.

Made with Goodnotes

DEPTH OF BIOLOGY

Example-

Suppose we have 50 molecules and 100 descriptors:

MLR fails because too many variables.

PLS takes those 100 descriptors and compresses them into, say, 4 components.

Then it makes an equation:

$$Y=a+b1\ C1\ +b2\ C2\ +b3\ C3\ +b4\ C4$$

where C_1, C_2, C_3, C_4 = new compressed variables.

-This model is much easier, stable, and accurate.

Made with Goodnotes

DEPTH OF BIOLOGY

Importance of PLS in QSAR-

Very popular in 3D-QSAR methods like CoMFA (Comparative Molecular Field Analysis).

Useful when descriptors >> molecules (common in modern QSAR).

Improves predictive power of QSAR models.

Made with Goodnotes

Multivariate Statistical Methods in QSAR

Multivariate" means dealing with many variables at the same time.

In QSAR, we often have many descriptors (X variables) for each molecule, and we want to relate them to biological activity (Y variable).

Since multiple variables interact together, multivariate methods are used.

1. Multiple Linear Regression (MLR)-

Simplest multivariate method.

Relates activity (Y) to several descriptors (X_1, X_2, X_3...).

Equation-

$Y=a+b1\ X1\ +b2\ X2\ +b3\ X3\ +...+bn\ Xn$

Problem: Works only if descriptors are not highly correlated.

Made with Goodnotes

2. Principal Component Analysis (PCA)-

A data reduction technique.

Converts many correlated descriptors into few new uncorrelated variables (principal components).

These new variables capture most of the information.

Used to visualize patterns, remove redundancy, and simplify data before regression.
Think of PCA as "summarizing" the data.

Made with Goodnotes

DEPTH OF BIOLOGY

3. Partial Least Squares (PLS)-

Similar to PCA but more powerful.

Compresses descriptors into components → but while doing so, it also considers the biological activity (Y).

Very useful when:
Number of descriptors >> Number of molecules.
Descriptors are correlated.

Commonly used in 3D-QSAR methods (like CoMFA).
PLS = "Dimension reduction + regression together."

Made with Goodnotes

4. Principal Component Regression (PCR)-

Combination of PCA + regression.
First, descriptors are reduced into principal components (via PCA).
Then, regression is applied between these components and activity.
Helps avoid problems of multicollinearity (overlapping variables).

5. Discriminant Analysis (DA)-

A classification method.
Instead of predicting numerical values, it separates compounds into groups:
Example: "Active" vs. "Inactive" drugs.
Finds descriptors that best discriminate between categories.

Made with Goodnotes

DEPTH OF BIOLOGY

3D-QSAR

3D-QSAR = A method that relates the three-dimensional properties of molecules (like shape, size, electrostatics, steric effects) to their biological activity.

Unlike traditional QSAR (1D or 2D descriptors), here 3D spatial information is considered.

-In simple words: 3D-QSAR studies how the 3D structure of a molecule controls its activity.

Made with Goodnotes

DEPTH OF BIOLOGY

Basic Idea of 3D-QSAR-

Align molecules in 3D space.

Calculate fields around molecules (steric, electrostatic, hydrophobic, etc.).

Correlate these 3D descriptors with biological activity using statistical methods (mostly PLS).

Generate contour maps → show regions where adding or removing groups will increase activity.

Made with Goodnotes

DEPTH OF BIOLOGY

Steps in 3D-QSAR-

Molecular Alignment → all molecules must be placed in a common 3D orientation.

Grid Generation → a 3D grid is placed around molecules.

Field Calculation → values (steric, electrostatic, etc.) are calculated at each grid point.

Statistical Modeling → usually Partial Least Squares (PLS) is used.

Contour Maps → visualization of important regions for activity.

Made with Goodnotes

DEPTH OF BIOLOGY

-Advantages of 3D-QSAR

Visual maps show where to modify molecules.
More realistic than 2D-QSAR (since molecules work in 3D).
Can guide lead optimization in drug design.

-Limitations

Requires correct alignment of molecules (difficult step).
Sensitive to structural flexibility (different conformations).
Needs large dataset for reliable results.

Made with Goodnotes

CoMFA (Comparative Molecular Field Analysis)

CoMFA = Comparative Molecular Field Analysis.

It is the most popular 3D-QSAR method.

It studies how the 3D fields (steric & electrostatic) around a molecule affect its biological activity.

In short: CoMFA checks how size and charge around a molecule control its activity.

Molecules interact with receptors through shape (steric effects) and charge distribution (electrostatic effects).
CoMFA measures these two properties in 3D space and builds a statistical model to connect them with activity.

Made with Goodnotes

DEPTH OF BIOLOGY

Steps in CoMFA-

1. Molecular Alignment

All molecules are placed in the same 3D orientation.

Important because results depend on correct alignment.

2. 3D Grid Generation

A 3D grid (like a box made of small cubes) is placed around the aligned molecules.

3. Field Calculation

At each grid point, two things are calculated:

Steric field → effect of size/bulk of molecule.

Electrostatic field → effect of charges (positive/negative).

Made with Goodnotes

DEPTH OF BIOLOGY

4. Statistical Analysis-
These field values are correlated with biological activity using PLS (Partial Least Squares).

5. Contour Map Generation-
Final output is a 3D contour map showing regions important for activity.

Made with Goodnotes

Interpretation of CoMFA Contour Maps-

Steric (size) contours:

Green region → bulky group increases activity.

Yellow region → bulky group decreases activity.

Electrostatic (charge) contours:

Blue region → positive charge increases activity.

Red region → negative charge increases activity.

These maps help chemists know where to add or remove groups in the molecule.

Made with Goodnotes

DEPTH OF BIOLOGY

5. Advantages of CoMFA-

Gives clear visual maps for drug design.

Relates 3D structure directly to biological activity.

Widely used in lead optimization.

6. Limitations-

Alignment-dependent → wrong alignment = wrong model.

Sensitive to conformational flexibility (different shapes of same molecule).

Needs good dataset of molecules.

Made with Goodnotes

CoMSIA (Comparative Molecular Similarity Indices Analysis)-

1. Meaning

CoMSIA = Comparative Molecular Similarity Indices Analysis.
It is an advanced 3D-QSAR method, developed as an improvement of CoMFA.

Like CoMFA, it relates 3D molecular properties to biological activity, but it uses a smoother and more reliable approach.

In short: CoMSIA is a refined version of CoMFA that uses more fields and gives smoother contour maps.

Made with Goodnotes

DEPTH OF BIOLOGY

Why CoMSIA was Developed?

CoMFA only uses steric and electrostatic fields.

CoMFA results can be unstable because of sudden energy changes near atoms.

CoMSIA solves this by using a Gaussian function, which makes field values smoother and easier to interpret

Made with Goodnotes

DEPTH OF BIOLOGY

Steps in CoMSIA (Similar to CoMFA)

1. Molecular Alignment → molecules are superimposed in 3D.
2. Grid Placement → a 3D lattice/grid surrounds the molecules.
3. Field Calculation → instead of just steric & electrostatic, CoMSIA calculates five similarity indices:

Steric (size)
Electrostatic (charge)
Hydrophobic
Hydrogen bond donor
Hydrogen bond acceptor

Statistical Analysis → uses PLS regression to relate these fields with activity.

Contour Map Generation → shows regions important for activity.

Made with Goodnotes

DEPTH OF BIOLOGY

Interpretation of CoMSIA Contour Maps-

Steric fields: bulky groups helpful or harmful.

Electrostatic fields: positive or negative charges favorable.

Hydrophobic fields: regions where hydrophobic/lipophilic groups increase activity.

H-bond donor fields: areas where adding donor groups improves activity.

H-bond acceptor fields: areas where acceptor groups improve activity.

Gives more detailed maps compared to CoMFA.

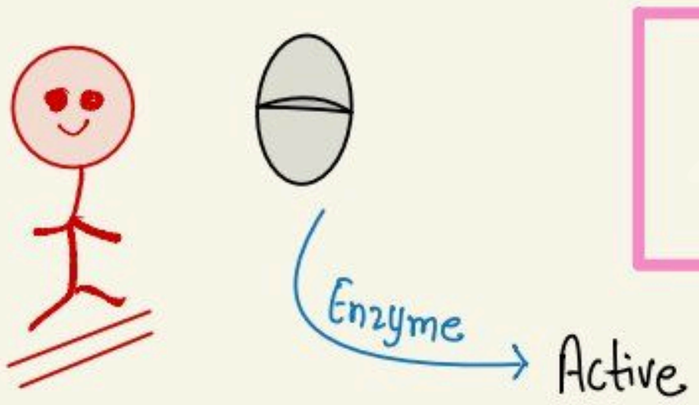
Made with Goodnotes

DEPTH OF BIOLOGY

Advantages of CoMSIA-
Considers five fields (not just 2 like CoMFA).
Smoother and more stable results (Gaussian function).
Less sensitive to alignment errors.
Provides better interpretation for medicinal chemists.

Limitations-
Still requires proper molecular alignment.
Needs a large and reliable dataset.
More computationally demanding than CoMFA.

Made with Goodnotes



PRODRUGS

DEPTH OF BIOLOGY

-Prodrugs are medications that turn into an active form once they enter the body. Prodrugs help improve a medication's effectiveness

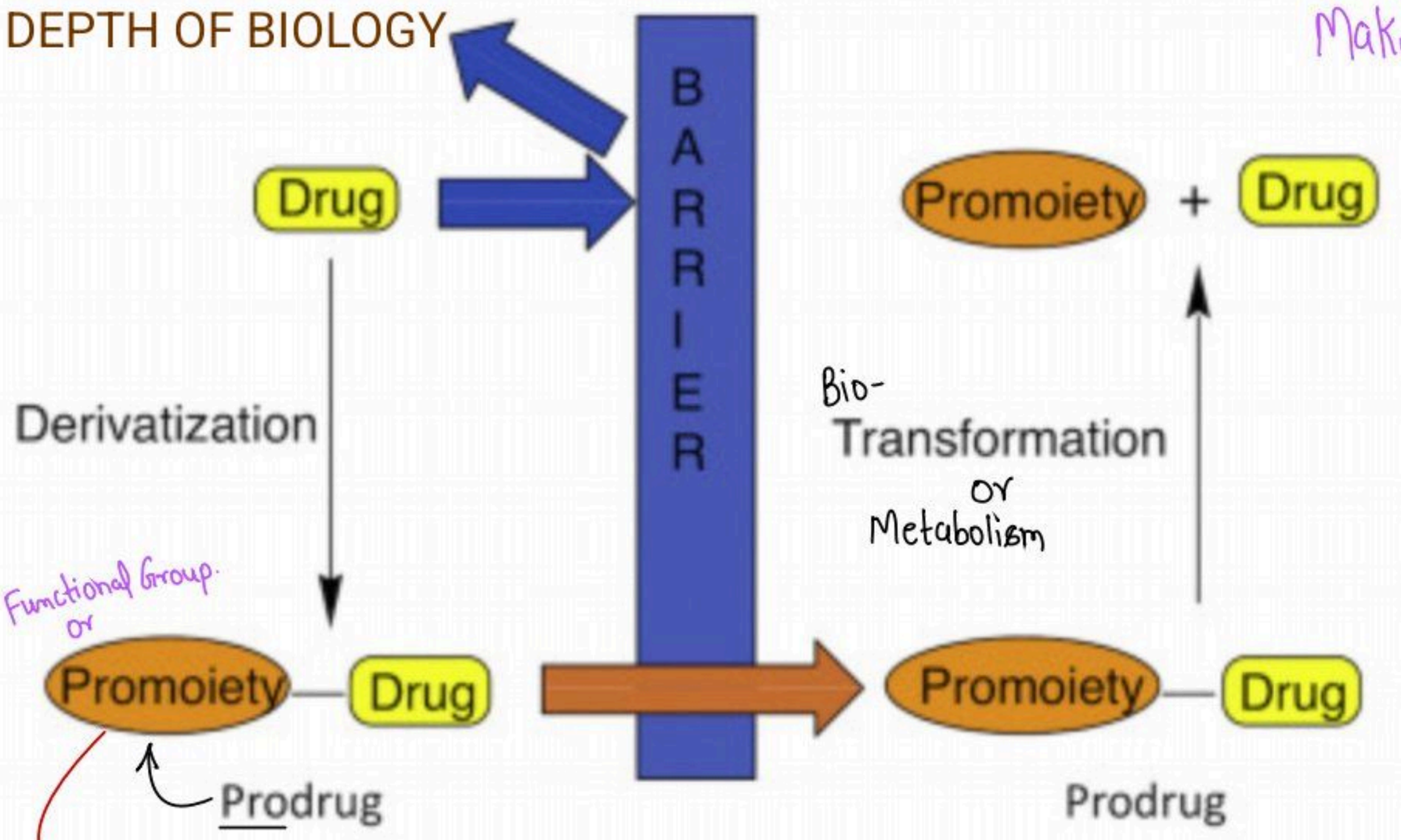
-Prodrug is a compound with little or no pharmacological activity that metabolizes inside the body and converts into a pharmacologically active drug compound.

-Prodrugs can be used to improve drug delivery or pharmacokinetics, to decrease toxicity, or to target the drug to specific cells or tissues , Physicochemical properties. , Acid Resistance

⊖ → Solub↑ Lipophilic↓

Make Prodrug.

DEPTH OF BIOLOGY



Barriers
- Limited solubility
- Poor permeability
- Toxicity/site irritation
- Poor physico-chemical properties
- Presystemic metabolism
- Controlled release desired
- Need for drug targeting

→ Toxic X , Safe ✓ , Stable ✓ ; Excrete from body ↑ .

DEPTH OF BIOLOGY

Made with Goodnotes

Made with Goodnotes

Scanned with CamScanner

A

Drug

DEPTH OF BIOLOGY

This drug is easily metabolized and results in a motabolite that is unable to activate its receptor

Enzymes

Receptor

Reduced/limited therapeutic effect

B

Prodrug/parent compound

-Promoiety

Enzymes

Site-specific enzymes

DEPTH OF BIOLOGY

Receptor

Prodrug's are parent compounds that contain regions/additional chemical groups that mask the drug's active zone and/or alters drug solubility (promoiety), but can be enzymatically or chemically cleaved/altered by site-specific enzymes, exposing the drug's active zone and allowing the drug to activate it's receptor resulting in a therapeutic effect

Therapeutic effect

DEPTH OF BIOLOGY

Made with Goodnotes

Made with Goodnotes

Scanned with CamScanner

Examples of prodrugs, their activation process and therapeutic effects:

Pro-drug	Active drug	Activation process	Key therapeutic use of the active drug
Azathioprine	6-mercaptopurine	Non-enzymatic conversion	antineoplastic
Codeine	Morphine	CYP2D6-mediated metabolism	analgesic
Clopidogrel	Thiol-derivative	CYP2C19-mediated oxidation	anticoagulant
Enalapril	enalaprilat	hydrolysis	anti-hypertensive
L-dopa	dopamine	decarboxylation	antiparkinsonian
Prednisone	prednisolone	reduction	anti-inflammatory
Simvastatin	beta-hydroxy acid metabolite	hydrolysis	Lipid-lowering drug
Sulfasalazine	5-aminosalicylic acid	Bacteria-mediated metabolism	anti-inflammatory

For a more comprehensive list of prodrugs, please refer to (Najjar & Karaman, 2019).

DEPTH OF BIOLOGY

-The prodrug approach allows for the development of more stable compounds in the gastrointestinal tract or more soluble than the parent compound, improving its bioavailability

For example, pivampicillin, a prodrug of ampicillin, carries an ester group which increases its lipophilicity. DEPTH OF BIOLOGY

Contain Ester Group.
↓
Lipophilicity.

-This chemical modification increases ampicillin's bioavailability from 32-55% to 87-94%. DEPTH OF BIOLOGY

-The prodrug approach can also improve the delivery of drugs to the central nervous system (e.g., L-dopa) and minimize side effects.

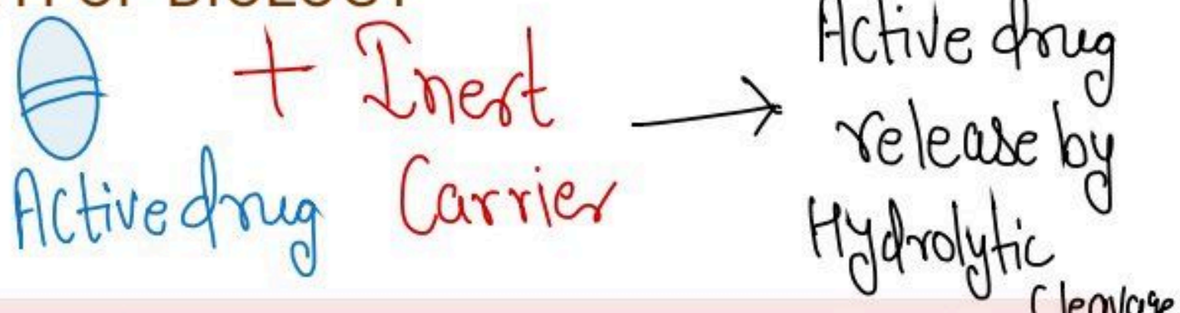
Made with Goodnotes

Made with Goodnotes

Scanned with CamScanner

CLASSIFICATION OF PRODRUGS:

DEPTH OF BIOLOGY



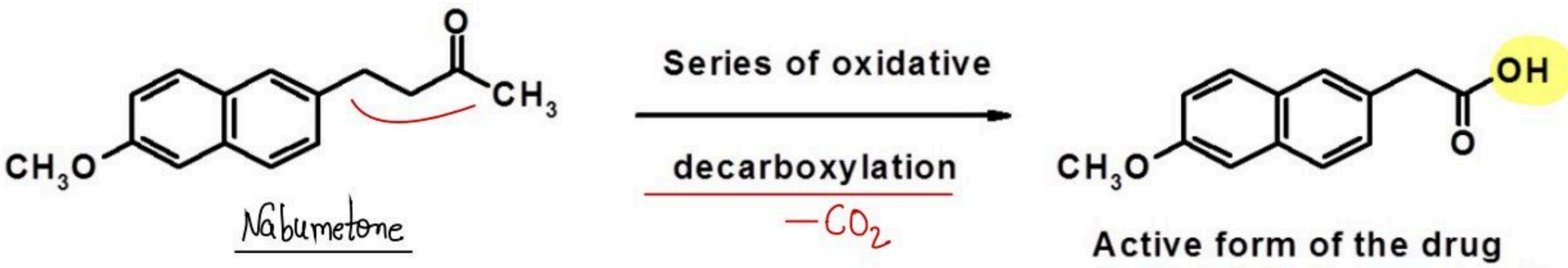
1.CARRIER LINKED PRODRUGS:

Active drug is **covalently linked** to an inert carrier or transporter moiety. They have **enhanced lipophilicity** due to attached carrier. The **active drug is released by hydrolytic cleavage, either chemically or enzymatically.**

2.BIOPRECURSORS:

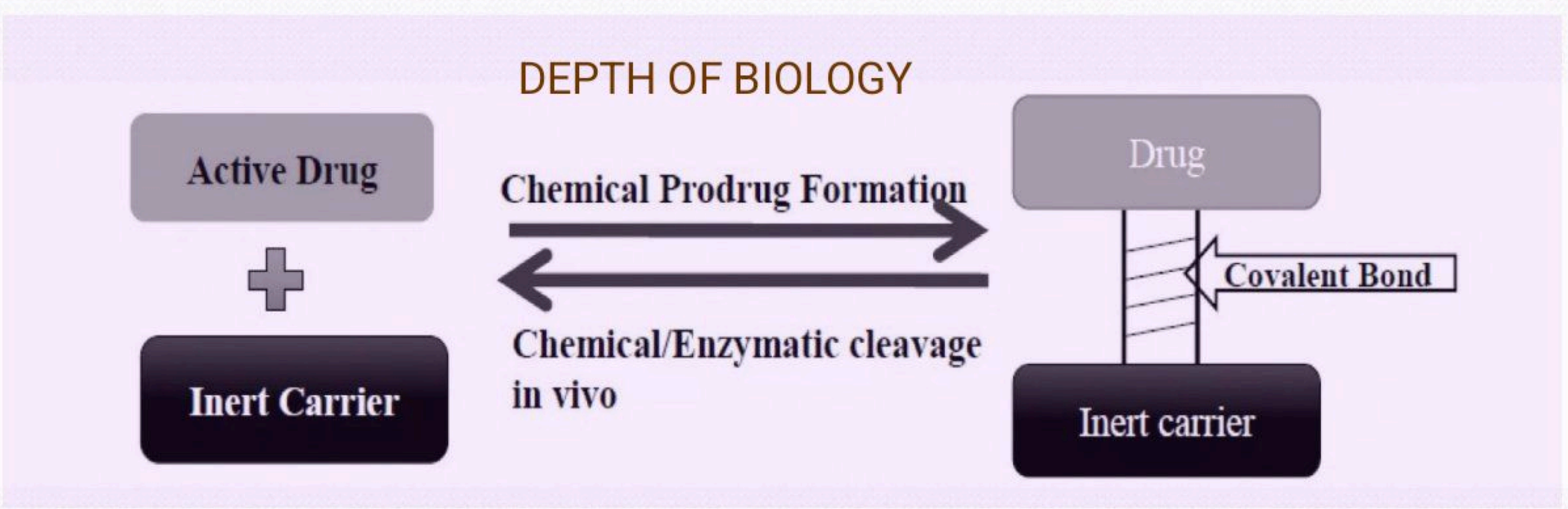


These are **inert molecules** obtained by chemical modification of the drug but **do not contain a carrier.** Such a molecule has the **same lipophilicity** as the parent drug and is bioactivated generally by redox biotransformation, only enzymatically. Ex: NSAID – nabumetone (relafen) - arthritis.



DEPTH OF BIOLOGY

- Carrier linked prodrug consists of the attachment of a carrier group to the active drug to alter its physicochemical properties.

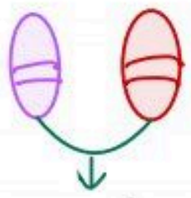


Made with Goodnotes

Made with Goodnotes

Scanned with CamScanner

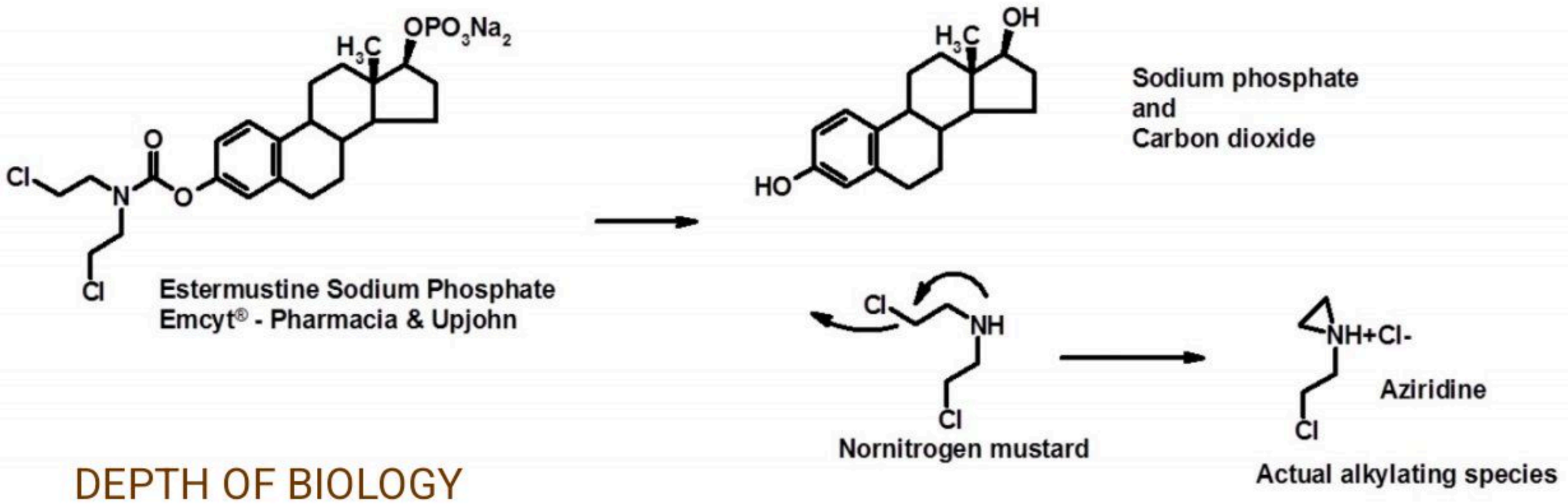
3. MUTUAL PRODRUGS:



DEPTH OF BIOLOGY

The prodrug comprises of two pharmacologically active agents coupled together to form a single molecule such that each acts as the carrier for the other.

Ex: Benorylate is mutual prodrug of NSAIDs aspirin and paracetamol.
Ex: Emcyt is a mutual prodrug containing estramustine and nornitrogen mustard linked to each other.



DEPTH OF BIOLOGY

CONSIDERATIONS IN THE DESIGN OF PRODRUGS:

1. Purpose
2. Expected properties
3. Site of transformation
4. Mechanism of transformation
5. Chemical half life

DEPTH OF BIOLOGY

Made with Goodnotes

Made with Goodnotes

Scanned with CamScanner

STRATEGIES FOR THE DESIGN OF PRODRUGS:

1. CARRIERS:

DEPTH OF BIOLOGY

- Carrier is an inert molecule or the promoiety attached to the active drug moiety through a metabolically labile linkage.
- The carrier imparts some desirable property to the drug such as increased lipid or water solubility.
- Carriers that help in directing the active moiety to the target site is called as specifier.

SPECIFIERS:

Targeting unit part of the prodrug which directs the active moiety to the target site.

DEPTH OF BIOLOGY

- Antibody Directed Enzyme Prodrug Therapy
- Gene Directed Enzyme Prodrug Therapy
- Polymer Directed Enzyme Prodrug Therapy/ Macromolecule Directed Enzyme Prodrug Therapy

LINKERS:

A releasable linker or spacer is incorporated between the specifier/carrier and the parent drug.

Reasons for the application of linkers:

1. Incorporation of appropriate linkage between the promoiety and the active drug.
2. Facilitation of enzymatic action on carrier linked prodrugs.

Classification of linkers:

DEPTH OF BIOLOGY

- Electronic cascade linkers
- Cyclization linkers

Doxorubicin attaches β-glucuronide was inert towards clevage by β-glucuronidase. When they were linked by a 1,6-elimination linker, the substrate showed susceptability to the enzyme.

Made with Goodnotes

Made with Goodnotes

Scanned with CamScanner

PRODRUG LINKAGE AND ENZYMES INVOLVED IN THE HYDROLYSIS OF LINKAGE: DEPTH OF BIOLOGY

Prodrug linkage	Hydrolyzing enzymes
Ester	Carboxylesterases
Short and medium chain	Cholinesterases
Aliphatic	Ester hydrolase , Lipase , Cholesterol Esterase, Acetylcholinesterase , Aldehyde oxidase, Carboxypeptidase
Long chain aliphatic carbonate	Pancreatic lipase,Pancreatin,Lipase, Carboxypeptidase, Cholinesterase
Phosphate, Organic	Acid phosphatase, Alkaline Phosphatase
Sulfate, organic	Steroid sulfatase

DEPTH OF BIOLOGY

PRODRUG

Advantages

- 1. It reduces adverse effects of drug.
- 2. Drug can be targeted to the desired sites.
- 3. Synergistic effects can be obtained without side effects.
- 4. Give additional biological action as that of parent drug.

DEPTH OF BIOLOGY

Disadvantages

- 1. Formation of toxic metabolites
- 2. **The active doses of two mutual prodrugs of the same** parent drugs may appear to be same in rats but may be quite different in clinical investigations.
- 3. The prodrug might consume a vital cell constituent such as glutathione during its activation stage which causes depletion of prodrug.

Made with Goodnotes

Made with Goodnotes

Scanned with CamScanner

Applications of Prodrugs

Pharmaceutical Applications

DEPTH OF BIOLOGY

- Masking Taste or Odour
- Reduction of gastric irritation
- Reduction in Pain at Site of Injection
- Enhancement of drug solubility and dissolution rate
- Enhancement of chemical stability
- Increase in patient compliance

DEPTH OF BIOLOGY

Pharmacokinetic Applications

- Enhancement of Bioavailablity
- Prevention of Presystemic metabolism
- Prolongation of duration of action
- Reduction Local and Systemic Toxicity of Drugs
- Site Specific Drug Delivery

DEPTH OF BIOLOGY

Made with Goodnotes

Made with Goodnotes

Scanned with CamScanner

DEPTH OF BIOLOGY

Rationale (Why we design prodrugs?)-

Drugs in their normal form often have problems:

Poor solubility
Poor absorption from the stomach/intestine
Do not reach the required organ or site
Act too fast or for too short a time
Cause irritation or bad taste

So, scientists modify the drug into a prodrug to overcome these problems

Made with Goodnotes

Prodrug Design for Specific Purposes-

(A) Improve Drug Solubility

Some drugs are not soluble in water → difficult for injections or absorption.

Prodrug is made more water-soluble (e.g., by making phosphate esters).

Example: Prednisolone → Prednisolone sodium phosphate (water soluble).

(B) Improve Drug Absorption and Distribution

Some drugs cannot easily cross cell membranes or the blood-brain barrier.

By making them more lipid-soluble, they can cross membranes.

Example: L-DOPA → enters brain → converts to dopamine.

Made with Goodnotes

DEPTH OF BIOLOGY

(C) Site-Specific Drug Delivery-

Some drugs cause side effects because they act everywhere.

Prodrug can be designed to activate only at target site.

Example: Sulfasalazine → broken in colon → releases 5-ASA (acts locally in colon).

(D) Sustained or Prolonged Action-

Some drugs act for a short time → frequent dosing is needed.

Prodrugs can be designed as depot (long-acting) forms → slow release.

Example: Fluphenazine decanoate (antipsychotic, given once in weeks).

Made with Goodnotes

DEPTH OF BIOLOGY

Practical Considerations of Prodrug Design-

When designing prodrugs, the following points must be kept in mind:

Conversion must be predictable – It should always change into the active drug.

Non-toxic by-products – The group that leaves after conversion should not be harmful.

Selectivity – It should activate only in the right place.

Stability – Must be stable in storage but break down inside the body.

Simple synthesis – Should be easy and cost-effective to make.

Made with Goodnotes