

PharmacoLearn

MPL203T Principles of Drug Discovery, M.Pharm Semester 2

Unit 1: An Overview of the Modern Drug Discovery Process and Target Discovery

Free notes at pharmacolearn.fyi

Principles of Drug Discovery, subject code MPL203T, is a core Semester 2 subject of M.Pharm Pharmacology. It carries 4 credits and 60 hours of theory, taught at 4 hours per week, for 100 marks. This subject teaches how a medicine begins: not in the clinic and not in the factory, but in the discovery laboratory, where a biological target is chosen, tested for its role in disease, and then used to find and refine the chemical matter that will one day become a drug. Unit 1 maps the whole modern process (target identification, target validation, lead identification and lead optimization), sets out its economics, and then studies target discovery itself: genomics, proteomics and bioinformatics, the nucleic acid and protein microarrays, the antisense and siRNA tools, zinc finger proteins, and transgenic animals as validation instruments. Unit 2 is lead identification: combinatorial chemistry, high throughput screening, in silico lead discovery and assay development, followed by the protein structure knowledge that design consumes: levels of structure, domains, motifs and folds, homology modelling, threading, NMR and X-ray crystallography. Unit 3 contrasts traditional with rational drug design and develops the two rational families, structure based and pharmacophore based, ending in virtual screening with drug likeness screening and pharmacophore mapping. Unit 4 works through molecular docking in its rigid, flexible and manual forms, docking based screening and de novo design, and then builds quantitative structure activity relationships from their history: physicochemical parameters, Hansch analysis and Free Wilson analysis. Unit 5 completes the quantitative story with the statistics of QSAR (regression, partial least squares and the other multivariate methods), the 3D QSAR methods CoMFA and CoMSIA, and prodrug design, the deliberate engineering of a molecule that the body itself converts into the drug.

1.1 The four stages of modern drug discovery

Modern drug discovery is an ordered funnel. A programme begins with a disease and a hypothesis about a molecule whose modulation would treat it, and it ends, years later and after several rounds of elimination, with one optimized compound worth handing to development. The PCI syllabus divides the discovery half of that journey into four stages, and the whole subject is organized around them:

THE MODERN DRUG DISCOVERY FUNNEL

TARGET IDENTIFICATION

A disease is linked to a molecule that a drug could act on

|

v

TARGET VALIDATION

Evidence that modulating this molecule changes the disease

|

v

LEAD IDENTIFICATION

Screening and design find chemical matter that acts on the target

|

v

LEAD OPTIMIZATION

The lead is refined for potency, selectivity and drug like properties

|

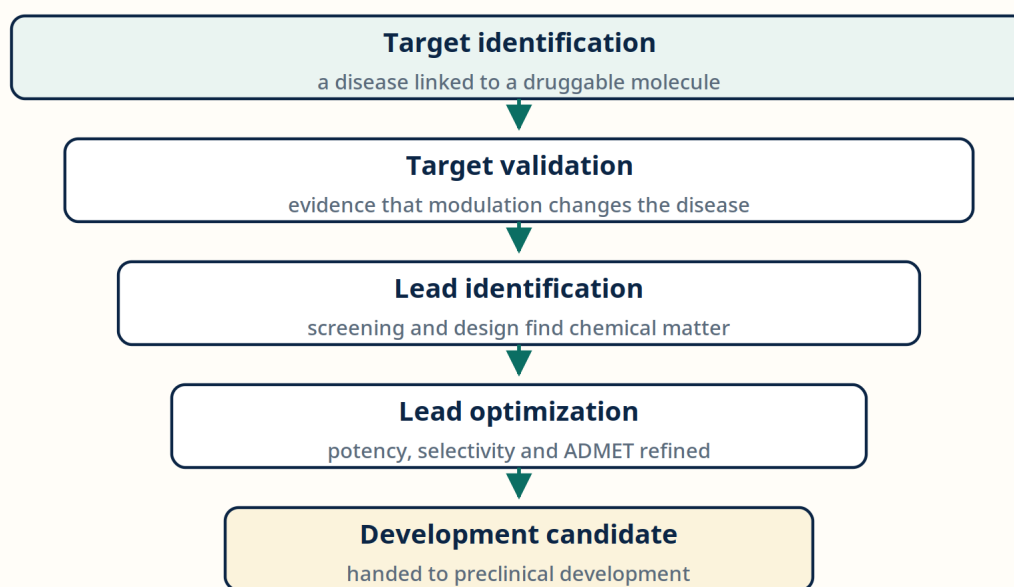
v

CANDIDATE -> preclinical development -> clinical trials -> medicine

Target identification asks which molecule, usually a protein, sits in a disease pathway at a point where a drug can usefully intervene. **Target validation** asks the harder question: is the evidence strong enough that changing this target's activity will change the disease in a patient, at an acceptable cost in safety? **Lead identification** finds the first chemical structures that act on the validated target with measurable, reproducible activity; these early actives are called hits, and the confirmed, chemically tractable ones become leads. **Lead optimization** is the long medicinal chemistry campaign that turns a lead into a candidate: potency is raised, selectivity against related targets is engineered, and the absorption, distribution, metabolism, excretion and toxicity (ADMET) properties are repaired one defect at a time, because a candidate must be a medicine, not merely a potent molecule.

Two disciplines run across all four stages. The first is elimination: at every stage most candidates are discarded, and the economics of the subject (section 1.4) flow from that attrition. The second is information: each stage buys knowledge the next stage spends. A target that cannot be assayed cannot be screened; a lead whose structure cannot be varied cannot be optimized. The stages are a logic, not a ritual: prove the biology, then find the chemistry, then perfect the chemistry.

The Modern Drug Discovery Funnel



Most programmes are eliminated at each stage: the funnel is a risk filter.

© 2026 PharmacoLearn, pharmacolearn.fyi

Figure: The modern drug discovery funnel. Programmes narrow stage by stage, and each stage exists to buy the information the next stage spends.

1.2 Target identification: what a drug target is

A **drug target** is a molecular structure, almost always a protein, whose activity a drug modifies to produce a therapeutic effect. The target classes that dominate marketed medicines are few, and the syllabus's "various targets for drug discovery" objective is met by knowing them:

Target class	How drugs act	Standard examples of the class at work
Enzymes	Usually inhibition, often of a disease driving step	HMG CoA reductase (statins), HIV protease (protease inhibitors), angiotensin converting enzyme (ACE inhibitors)
G protein coupled receptors	Agonism or antagonism of a signalling receptor	Beta adrenoreceptors, histamine H2 receptors, dopamine D2 receptors

Target class	How drugs act	Standard examples of the class at work
Ion channels	Block or modulation of ion flow	Voltage gated sodium channels (local anaesthetics), calcium channels (amlodipine)
Nuclear receptors	Ligand activated transcription factors	Steroid and thyroid hormone receptors, PPAR gamma (glitazones)
Transporters	Block of uptake or transport	Serotonin transporter (SSRIs), proton pump (an enzyme transporter, blocked by PPIs)
Structural and other proteins, nucleic acids	Binding that disables a function	Tubulin (vinca alkaloids, taxanes), bacterial ribosome (many antibiotics), DNA itself (alkylating agents)

A good target needs more than disease association. The working criteria taught with this unit are: **efficacy logic** (modulating it should change the disease, not merely accompany it); **druggability** (the protein must possess a site a small molecule can bind with useful affinity; enzymes and GPCRs score well because evolution built them to bind small molecules, while flat protein protein interaction surfaces score badly); **assayability** (its activity must be measurable in a robust laboratory test, or no screen can be run); **selectivity room** (it should differ enough from related proteins that a drug can hit it without hitting its cousins everywhere in the body); and **a safety margin in principle** (a target essential in every tissue is a toxicity risk from the first day). Target identification assembles candidates from disease biology: human genetics, pathway analysis, expression studies in diseased tissue, and the genomics and proteomics methods of sections 1.5 and 1.6.

1.3 Target validation: proving the target deserves a programme

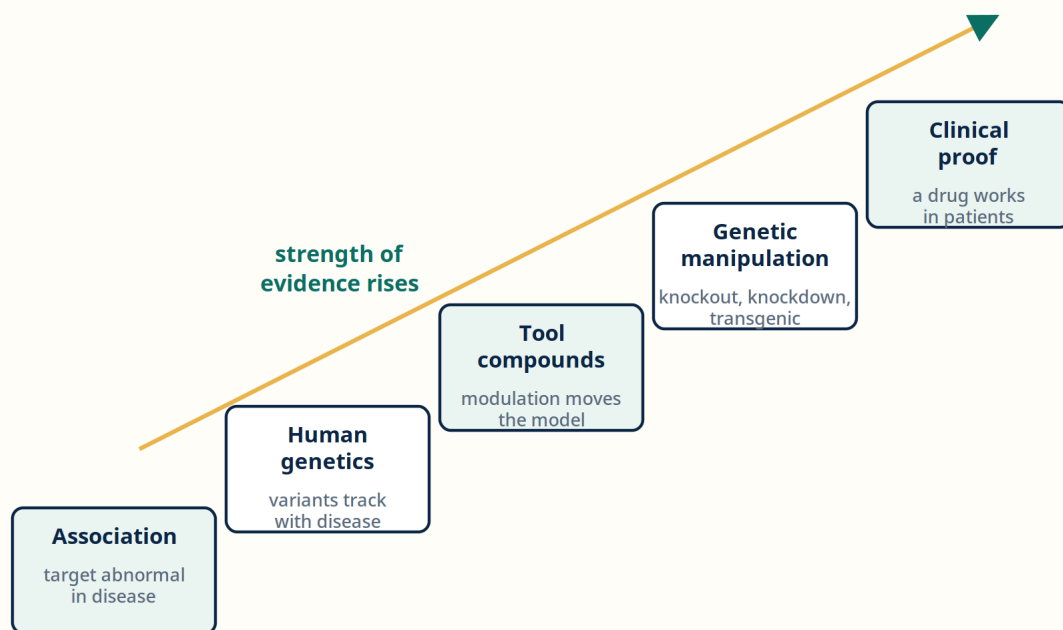
Identification proposes; validation disposes. **Target validation** is the accumulation of evidence that modulating the target, in the right direction, by a drug like intervention, will produce the desired clinical effect. It is deliberately distinguished from identification because the commonest expensive failure in discovery is a beautifully screened programme against a target that never mattered in patients.

Validation evidence forms a ladder, weakest at the bottom:

- **Association:** the target is overexpressed, mutated or otherwise abnormal in the disease. Suggestive, never sufficient: the abnormality may be a consequence of the disease rather than a cause.
- **Genetic evidence in humans:** people carrying loss or gain of function variants in the target gene show the disease, or are protected from it. This is among the strongest early evidence, because nature has already run the experiment of modulating that target for a lifetime.
- **Pharmacological evidence:** a tool compound, an antibody or a nucleic acid that modulates the target changes the disease process in cells and in animal models, in a dose related and reversible way.
- **Genetic manipulation in model systems:** switching the target off (knockout, antisense, siRNA) or on, in cells and in transgenic animals (sections 1.9 to 1.11), reproduces or reverses the disease phenotype.
- **Clinical proof,** the only complete validation: a drug against the target works in patients. It arrives too late to guide the original decision, which is why the earlier rungs must be made as strong as possible.

Two judgements inside validation deserve emphasis. First, direction: the evidence must say whether the target should be activated or inhibited, and programmes have failed by pushing the wrong way. Second, the therapeutic window in principle: a target whose modulation is lethal or intolerable in normal physiology fails validation however elegant the disease logic. Validation is therefore as much a safety discipline as an efficacy one, and it continues during lead optimization, because every toxic lead is also a question about the target.

The Target Validation Evidence Ladder



Validation is a ladder: each rung is stronger, and clinical proof is the only complete one.

© 2026 PharmacolLearn, pharmacollearn.fyi

Figure: The target validation evidence ladder. Each rung is stronger than the one below, and only clinical proof completes the validation.

1.4 The economics of drug discovery

The syllabus names the economics of drug discovery as a topic in its own right, and the honest treatment is structural rather than numerical, because published cost estimates differ widely with method and era. The structural facts are stable. Bringing one new medicine to market takes well over a decade from target selection. The overwhelming majority of programmes started are terminated, and the most expensive failures are the late ones, above all in clinical development, where a single large trial costs more than years of discovery work. Discovery itself is therefore run as risk management: money is spent in small, information buying steps early (validation, screening) precisely so that the large sums later are spent on candidates already stripped of every discoverable defect.

Three economic consequences shape the science in this subject. First, **fail early and cheap**: killing a weak target at validation costs a fraction of killing a weak candidate in Phase II, which is why validation rigour is an economic virtue, not academic caution. Second, **the funnel must be wide at the top**: because attrition is so high, lead identification screens very large compound collections (Unit 2), and anything that raises the quality of what enters the funnel pays for itself many times over. Third, **patent life runs from filing**, so years in discovery are years subtracted from protected sales, one pressure behind the computational methods of Units 3 to 5 that promise to shorten the search.

1.5 Genomics in target discovery and validation

Genomics studies the genome as a whole: its sequence, its variation between people, and which genes are switched on or off in a given tissue or disease. Its contributions to target discovery are threefold. First, **disease gene discovery**: comparing the genomes of patients and controls, through family linkage studies and genome wide association studies, points to genes whose variants track with disease, and each such gene is a candidate target or a signpost to one. Second, **target triage**: sequence comparison reveals

whether a candidate target belongs to a druggable family, whether close relatives exist that threaten selectivity, and whether the same gene is essential in normal tissues. Third, **expression genomics**: measuring which genes are transcribed in diseased versus healthy tissue (the measurement technology is the nucleic acid microarray of section 1.7) highlights the pathways the disease has disturbed.

Genomics also supplies validation of the strongest kind. A human population carrying a natural loss of function variant in a target gene is a lifelong, whole body experiment in inhibiting that target: if such people are healthy and protected from the disease, the target is validated for both efficacy direction and basic safety in a way no animal model can match. The discipline's limit is equally clear: genes are not proteins, association is not mechanism, and a genomic signal begins, rather than completes, the validation argument.

1.6 Proteomics and bioinformatics

Proteomics applies the same whole system ambition to proteins: identifying and quantifying the full protein content of a cell or tissue, mapping their interactions, and detecting their modifications, typically by separating complex mixtures and identifying the components by mass spectrometry. Where genomics lists what could be made, proteomics reports what is actually present, in what amount, in what modified state, and in whose company, which is closer to the drug target itself, since nearly all targets are proteins. In discovery it serves target identification (proteins abnormally expressed or modified in disease), biomarker discovery (proteins that report disease state or drug response, used to stratify experiments and trials), and mechanism studies (which proteins a candidate drug actually touches, the basis of selectivity screening).

Bioinformatics is the computational discipline that makes both omics usable: storage, alignment and comparison of sequences, prediction of protein families and domains, mining of expression databases, and modelling of pathways. It answers the first questions asked of any candidate target: What family is it? What does its sequence predict about structure and function? Where is it expressed? What happens to organisms that lack it? A target that survives this *in silico* interrogation earns the expensive wet laboratory work of validation; one that fails it is discarded at the cost of computer time, the cheapest elimination in the whole funnel.

1.7 Nucleic acid microarrays

A **nucleic acid microarray** (DNA chip) measures the expression of thousands of genes in one experiment. Short DNA probes, each specific to one gene's messenger RNA, are fixed at known positions in a grid on a solid surface. Messenger RNA from the sample is labelled, usually with fluorescence, and allowed to hybridize to the probes; the brightness at each spot reports how much of that gene's message the sample contained. Comparing a diseased sample with a healthy one, spot by spot, yields an **expression signature**: the pattern of genes turned up or down by the disease.

In target discovery the signature is mined for genes whose change is large, consistent and pathway coherent; such genes, or the proteins they encode, become candidate targets. In validation, the same technology tests whether modulating a candidate target moves the disease signature back toward health. Microarrays also serve pharmacology directly: the expression response of cells to a drug reveals its pathway effects and foreshadows toxicities. Their limitations are those of any messenger RNA measurement: message levels only approximate protein levels, and correlation with disease still has to be converted into causal evidence by the validation ladder of section 1.3.

1.8 Protein microarrays

A **protein microarray** fixes proteins, rather than nucleic acids, at known grid positions, and asks questions that only proteins can answer. In its common discovery forms, a labelled probe (a candidate drug, an antibody, another protein) is washed over the array, and the spots it binds identify its protein partners. Three uses matter for this subject. First, **interaction mapping**: which proteins bind a given protein helps place a candidate target in its pathway. Second, **drug selectivity profiling**: running a lead compound over an array of many proteins exposes off target binding early, when chemistry can still fix it. Third, **antibody and biomarker work**: arrays of antigens read out the antibody content of patient samples, reporting disease associated immune signatures.

Protein arrays are technically harder than DNA arrays, because proteins denature and lose activity when fixed to surfaces, so each array compromises between coverage and function. They are the protein level counterpart of the nucleic acid array: together the two technologies move discovery from one gene at a time to systems at a time.

1.9 Antisense technologies: antisense oligonucleotides and siRNA

Antisense technologies attack a target at its message rather than at the protein. An **antisense oligonucleotide** is a short, synthetic, single stranded nucleic acid whose sequence is complementary to the target's messenger RNA. Bound to that message, it blocks protein production, either by recruiting the enzyme RNase H to cut the RNA strand of the hybrid, or by physically blocking splicing or translation. **Small interfering RNA (siRNA)** reaches the same end by a cellular pathway: a short double stranded RNA is loaded into the RISC complex, one strand is discarded, and the retained strand guides RISC to cut messenger RNAs of matching sequence. In both cases the result is **knockdown**: a specific, dose adjustable, reversible reduction of one chosen protein, achievable within days of knowing the gene sequence, long before any small molecule inhibitor could exist.

That speed is why the syllabus places these tools inside target validation. If knocking down the target in cells or animals reverses the disease phenotype, validation strengthens; if nothing happens, the target is questioned before a screening campaign is funded. The same mechanisms are also medicines in a small number of diseases where the target tissue can be reached, but their primary discovery role, and the one examined here, is as validation instruments. Their limitations define their use: delivery into cells is imperfect, effects are temporary, and partial knockdown is not identical to full pharmacological blockade of the active site.

1.10 Zinc finger proteins

A **zinc finger** is a small protein domain, a motif in the language of Unit 2, in which a zinc ion held by cysteine and histidine residues stabilizes a fold that fits into the major groove of DNA. Natural transcription factors string several such fingers together, each finger reading a few base pairs, so that the assembly recognizes one chosen DNA sequence. Because the fingers are modular, they can be engineered: **zinc finger proteins** built to order can be aimed at a chosen gene and fitted with effector domains that switch the gene off or on, and **zinc finger nucleases** carry a cutting domain that breaks DNA at the chosen site, allowing a gene to be disrupted or edited in cells and animals.

For drug discovery the value is again instrumental. A zinc finger repressor aimed at a candidate target gene is a validation tool of great precision: it suppresses exactly one gene's expression without chemistry against the protein, complementing antisense and siRNA with a DNA level intervention. Zinc finger nucleases historically provided one route to the knockout animals of the next section. The syllabus includes them as part of the molecular toolkit that modern target validation assumes: genomics finds the gene, and zinc finger reagents let the investigator turn that gene down, off or out, on demand.

1.11 Transgenic animals in target validation

A **transgenic animal** carries an engineered change in its genome, and in validation work the engineering serves one purpose: to model, in a whole organism, the effect of modulating the target. The standard designs are:

Design	What is changed	Validation question it answers
Knockout	The target gene is disrupted so no functional protein is made	What happens when the target is fully inhibited, for life? Mimics a perfect antagonist, and exposes both efficacy logic and on target toxicity
Knockin	A modified version of the gene replaces the normal one (a disease mutation, or a drug resistant form)	Does the disease mutation itself cause the phenotype? Does the drug stop working when its binding site is altered, proving on target action?
Conditional or inducible knockout	The gene can be switched off in one tissue or at one chosen time	Separates the target's adult pharmacology from its developmental roles, and localises the effect to the tissue that matters
Humanized transgenic	The animal gene is replaced by the human version	Tests drugs and biologics that recognise only the human protein

The interpretive discipline matters as much as the technology. A knockout that produces no disease protection weakens the target, but developmental compensation can hide a real role, since the animal has had a lifetime to rewire around the missing protein; acute blockade or an inducible knockout may reveal what the germline knockout concealed. Conversely, a knockout may exaggerate a drug's effect, because a drug rarely abolishes a protein completely in every tissue. Transgenic evidence sits high on the validation ladder, below human genetics and clinical proof: strong, but not final.

*More free M.Pharm notes and unit wise PDFs at pharmacolearn.fyi
Free video lectures on the PharmacoLearn YouTube channel*