

SCHEDA DESCRITTIVA DELL'UNITÀ DIDATTICA

Titolo del Modulo Codice	Analisi Correlative Struttura Attività 85730
Module Title Code	Structure activity relationships analysis 85730

Settore Scientifico di riferimento Reference Sector	CHIM08 (Chimica Farmaceutica) CHIM08 (Medicinal Chemistry)
--	---

Corso di Laurea Degree Course	Corso di laurea magistrale in Chimica e Tecnologia Farmaceutiche Degree Course in Pharmaceutical Chemistry and Technology
--	--

Anno di corso Year of the degree course	Terzo Third
--	----------------

Semestre Semester	Secondo Second
------------------------------------	-------------------

Numero di Crediti	6 CFU
Carico di lavoro globale (ore)	150 h
Lezioni frontali	42 h
Studio individuale	108 h
Credit Rating	6 CFU
Total workload (hours)	150 h
Class lectures	42 h
Individual work	108 h

Docente responsabile	Prof. Giuseppe Romeo Dipartimento di Scienze del Farmaco Edificio 2, Viale A. Doria, 6 – 95125 Catania. Tel: 095 738 4024; E-mail: gromeo@unict.it Orario di ricevimento: martedì ore 10-12.
Course Instructor	Prof. Giuseppe Romeo Department of Drug Sciences Building 2, Viale A. Doria, 6 – 95125 Catania. Phone: 095 738 4024; E-mail: gromeo@unict.it Office Hours: Tuesday, 10-12 AM.

Organizzazione generale del corso General organization of the course

Le attività del corso sono costituite da lezioni frontali e studi di casi in aula. È obbligatoria

la frequenza ad almeno il 70% delle lezioni. La frequenza alle lezioni è attestata dalla firma dello studente.

Activities of the course are constituted by class lectures and case studies. Attendance (at least 70%) at classes is mandatory. Attendance is attested by student's own signature.

Obiettivi formativi

Module aims

Il corso intende fornire allo studente conoscenze sulla correlazione tra struttura molecolare e attività biologica in composti di interesse farmaceutico. In particolare, il corso è indirizzato allo studio dei principali descrittori molecolari, alle tecniche di analisi regressionale per la costruzione di modelli matematici e ai principali parametri statistici utili per la loro validazione.

Knowledge about the relationships between molecular structure and biological activity in compounds of pharmaceutical interest. In particular, the course is directed to the study of the principal molecular descriptors, of the regression techniques for the construction of mathematical models, and to the comprehension of the main statistical parameters useful in model validation.

Contenuti (dettaglio)

Syllabus (detailed)

GENERALITÀ SUL CORSO.

RAPPRESENTAZIONE DELLE MOLECOLE.

van der Waals Surface Area e Solvent-Accessible Surface Area (SASA). Grafo di una molecola. Tabelle di connessione. Matrici di adiacenza. Matrici di distanza topologica. Matrici di distanza 3D. Structure data format (Molfile). Esempi di indici topologici: Topological autocorrelation vector. Notazione SMILES (Simplified Molecular Input Line Entry): regole principali ed esempi di notazione. Bit strings e Indici di similarità: Tanimoto coefficient. Motivi strutturali in una molecola secondo Bemis and Murcko: Ring systems, Linkers, Side-chains, Frameworks o Scaffolds. Motivi strutturali privilegiati e motivi strutturali indesiderati in drug-design.

INTERAZIONI FARMACO-RECETTORE.

Relazione tra variazione di energia libera e costante di equilibrio del complesso farmaco-recettore. Relazione tra variazione di energia libera standard e K_i . Competition binding assays: protocollo sperimentale. Curva di spiazzamento del ligando marcato e definizione di IC_{50} . Calcolo del valore di K_i : equazione di Cheng e Prusoff con esempi di calcolo. Processi che favoriscono o sfavoriscono l'interazione farmaco-recettore. Cenni ai diversi tipi di legami e interazioni tra farmaco e recettore: legame a idrogeno, interazioni elettrostatiche, effetto idrofobico, interazioni di van der Waals, interazioni tra sistemi aromatici. Studio di un Caso 1: Interazioni tra Carazololo e recettore β_2 -adrenergico (D. M. Rosenbaum et al. *Science*, 318, 1266-1273, 2007). Studio di un Caso 2: Meccanismo di binding tra Alprenololo e recettore β_2 -adrenergico (R. O. Dror et al., *PNAS*, 108, 13118-13123, 2011).

PROPRIETÀ DELLE MOLECOLE.

Parametri chimico-fisici e descrittori molecolari. Parametri di lipofilia: $\log P$; costante idrofobica del sostituente di Hansch-Fujita, π . Metodi per la misura sperimentale del $\log P$. Metodo dello Shake-flask, protocollo sperimentale secondo la OECD Guideline n° 107, vantaggi e svantaggi del metodo. Metodo in RP-Thin Layer Chromatography, definizione di R_M e relazione tra R_M e $\log P$. Metodo in RP-HPLC, definizione di fattore di capacità e sua relazione con $\log P$. Metodi di calcolo del $\log P$: Metodo per sostituzione. Metodi fragment-based (Nis and Rekker; Hansch and Leo) ed esempi di calcolo con CLOGP e ACD/Log P softwares. Metodi atom-based. Definizione di $\log D$. Formule per il calcolo di $\log D$ per composti acidi o basici. Grafici di $\log D$ in funzione del pH per diverse tipologie di principi attivi. Esempi generati con ACD/Log D software. Importanza del $\log D_{7.4}$ sulle drug-like properties. $\log P$ come descrittore molecolare in QSAR: esempi di modelli lineari e non lineari.

Parametri elettronici. Costanti di Hammett, σ , e loro definizione. Equazione di Hammett e definizione di ρ . Componenti \mathfrak{S} e \mathfrak{R} di Swain e Lupton.

Parametri sterici. Costante di Taft, E_s , sua definizione e correlazione con il raggio di van der Waals. Costante di Taft corretta, E_s^c . Rifrazione molare, MR, sua definizione e relazione con il volume molare. Volume di van der Waals, V_w . Parametri STERIMOL (L , B_1 , B_4 e L , B_1 , B_5) di Verloop e loro definizione.

Esempi di descrittori topologici. Total Adjacency Index con esempio di calcolo. Kier and Hall Connectivity Indices, χ , con esempi di calcolo per $^1\chi$ e $^2\chi$. Numero di gruppi accettori (HBA) e donatori (HBD) di legami a idrogeno in una molecola. Polar Surface Area (PSA) e Topological Polar Surface Area (TPSA), definizioni e metodi di calcolo.

LINEE GUIDA PER LA PROGETTAZIONE DI DRUG-LIKE COMPOUNDS.

Regole di Lipinski (Regola del 5): origine, definizione, utilizzo. Definizione di Rotatable bond. Esempi di identificazione dei Rotatable bonds in strutture di interesse farmaceutico. Regole di Veber. Regola del 3 e sua applicazione in Fragment-based Screening. Ligand Efficiency (LE) e Binding Efficiency Index (BEI), loro definizione e calcolo. Influenza delle dimensioni molecolari su LE.

ANALISI DI REGRESSIONE IN QSAR.

Analisi di regressione lineare semplice. Definizione e requisiti necessari per poter eseguire un'analisi di regressione lineare. Metodo dei minimi quadrati. Formule per il calcolo del coefficiente angolare e dell'intercetta. Coefficiente di correlazione, r , definizione e formula per il calcolo. Esempi di calcolo su un set di dati. Definizione di Total Sum of Squares (TSS), Explained Sum of Squares (ESS) e Residual Sum of Squares (RSS). Coefficiente di determinazione, r^2 , e formule per il calcolo. Errore standard di y , del coefficiente angolare e dell'intercetta. Calcolo degli intervalli di confidenza per il coefficiente angolare e l'intercetta. Analisi della Varianza (ANOVA) e Statistical Hypothesis Testing. Espressione dell'ipotesi nulla e dell'ipotesi alternativa nel contesto dell'analisi di regressione lineare semplice. Calcolo delle Mean Squares (MSE, MSR, MST). F -test come metodo statistico per l'accettazione o il rigetto dell'ipotesi nulla ed esempio di calcolo.

Analisi di regressione lineare multipla. Descrizione del modello e R^2 . F -test in analisi di regressione lineare multipla. Strategie per la creazione di un modello di regressione multipla: Backward Elimination e calcolo di F-to-remove; Forward Inclusion e calcolo di F-to-enter.

Procedure di cross-validazione di un modello regressionale. Leaving One sample Out at time (LOO) e calcolo di Q^2 . Definizione di Training set, Evaluation set e Test set. Utilizzo

del coefficiente di correlazione r tra attività predette e attività sperimentalmente determinate nella validazione di un modello. Analisi delle correlazioni casuali tra variabili in un modello regressionale: metodo dello Y scrambling.

APPROCCI DI HANSCH E DI FREE-WILSON IN QSAR

Generalità sull'approccio di Hansch. Equazioni di Hansch. Scelta delle variabili indipendenti e termini quadratici nell'equazione. Interpretazione dell'equazione in base al valore e al segno dei coefficienti. Esempi di equazioni di Hansch su sets di molecole ad attività α -adrenolitica e ad attività antimalarica. Grafico di Craig per la scelta dei sostituenti. Approccio di Free-Wilson: generalità e definizione di variabile indicatrice.

ANALISI DELLE COMPONENTI PRINCIPALI (PCA) E PARTIAL LEAST SQUARES (PLS)

Matrice di correlazione tra variabili. Collinearità. Pre-trattamento dei dati: autoscaling. Generalità e principi dell'analisi delle componenti principali. Interpretazione geometrica delle componenti principali. Scores e Loadings: matrici, plots e loro interpretazione. Eigenvalues. PCA come tecnica di classificazione. Studio di un Caso 3: Profili di binding di farmaci antipsicotici (J.H.L. Lange et al., *J. Med. Chem.* 50, 5103-5108, 2007). Regressione sulle componenti principali (PCR). Generalità e principi della tecnica PLS. Caratteristiche delle variabili latenti. Definizione e utilizzo della PRESS. Interpretazione geometrica delle variabili latenti.

3-D QSAR: COMPARATIVE MOLECULAR FIELD ANALYSIS (CoMFA)

Generalità e principi del metodo. Definizione IUPAC di farmacoforo. Campi sterici e elettrostatici. Interpretazione delle contour maps. Studio di un Caso 4: CoMFA study of piperidine analogues of cocaine at the dopamine transporter (H. Yuan et al. *J. Med. Chem.* 47, 6137-6143, 2004).

GENERALITIES ABOUT THE COURSE.

REPRESENTATION OF MOLECULES.

van der Waals Surface area and Solvent-Accessible Surface Area (SASA). Molecular graphs. Connection tables. Adjacency matrix. Topological distance matrix. 3D Distance matrix. Structure data format (Molfile). Topological indices examples: Topological autocorrelation vector. SMILES Notation (Simplified Molecular Input Line Entry): main rules and examples of SMILES notation. Similarity indices: Tanimoto coefficient. Molecular frameworks after Bemis and Murcko: Ring systems, Linkers, Side-chains, Frameworks or Scaffolds. Privileged Motifs and Unwanted Motifs in Drug Design.

DRUG-RECEPTOR INTERACTIONS.

Relationship between variation of free energy and equilibrium constant in ligand-receptor complex. Relationship between standard free energy and K_i . Competition binding assays: experimental protocol. Displacement curve and IC_{50} definition. Calculation of K_i : Cheng and Prusoff equation with examples. Favourable and unfavourable forces in the ligand-receptor interaction. Brief description of the different kinds of interactions in the ligand-receptor complex: hydrogen bond, electrostatic interactions, hydrophobic effect, van der Waals interactions, aromatic interactions. Case study 1: Interactions between Carazolol and the β_2 -adrenergic receptor binding site (D. M. Rosenbaum et al. *Science*, 318, 1266-1273, 2007). Case study 2: Mechanism of binding of Alprenolol to β_2 -adrenergic receptor (R. O. Dror et al., *PNAS*, 108, 13118-13123, 2011).

MOLECULAR PROPERTIES.

Physico-chemical parameters and molecular descriptors. Lipophilic parameters: $\log P$; Hansch-Fujita hydrophobic substituent constant, π . Experimental measure of $\log P$. Shake-flask method, experimental

protocol following the OECD Guideline n° 107, advantages of the method. Measure of logP by means of RP-Thin Layer Chromatography, R_M and relationship between R_M and logP. Measure of logP by means of RP-HPLC, capacity factor, k' and relationship with logP. Calculation methods of logP: Substitution method. Fragmental methods (Nis and Rekker; Hansch and Leo) with examples using CLOGP and ACD/Log P softwares. Atom-based methods. Definition of logD. Calculation of logD for acids and bases. logD Plots for different kinds of drugs. Examples with ACD/LogD software. Importance of the logD_{7.4} value for the drug-like properties. LogP as molecular descriptor in QSAR: examples of linear and non-linear models.

Electronic parameters. Hammett constants, σ , and their descriptions. Hammett equation and definition of ρ . Swain and Lupton Σ and \mathfrak{R} components.

Steric parameters. Taft E_s constant, definition and correlation with the van der Waals radius. Taft E_s^c corrected constant. Molar Refractivity, MR, definition and correlation with the Molar volume. van der Waals Volume, V_w . STERIMOL (Verloop) parameters (L , B_1 - B_4 e L , B_1 , B_5).

Examples of topologic descriptors: Total Adjacency Index and examples of calculation. Kier and Hall Connectivity Indices, χ , and examples of calculation of $^1\chi$ e $^2\chi$. Number of hydrogen bond acceptors (HBA) and donors (HBD) in a molecule. Polar Surface Area (PSA) and Topological Polar Surface Area (TPSA), definitions and calculation methods.

GUIDELINES FOR STRUCTURAL PROPERTIES OF DRUG-LIKE COMPOUNDS.

Lipinski Rules (Rule of 5): origin, definition, applications. Definition of Rotatable bond. Examples of identification of rotatable bonds in structures of pharmaceutical interest. Veber rules. Rule of 3 and application in Fragment-based Screening. Ligand Efficiency (LE) e Binding Efficiency Index (BEI), definitions and applications. Influence of the ligand molecular size on LE.

REGRESSION ANALYSIS IN QSAR.

Simple linear regression analysis. Assumptions for simple linear regression. Least squares method. Formulas for the slope, intercept and coefficient of correlation, r . Their calculation for a data set. Total Sum of Squares (TSS), Explained Sum of Squares (ESS), and Residual Sum of Squares (RSS). Coefficient of determination, r^2 . Standard errors of y , slope and intercept. Confidence intervals for the slope and intercept. Analysis of variance (ANOVA) and Statistical Hypothesis Testing. Null hypothesis and Alternative hypothesis in simple linear regression analysis. Calculation of the Mean Squares (MSE, MSR, MST). F -test as statistical test for the acceptance or rejection of the null hypothesis.

Multiple linear regression analysis. Description of the model and R^2 . F -test in multiple linear regression analysis. Strategies for creating a multiple regression model: Backward Elimination and F-to-remove calculation; Forward Inclusion and F-to-enter calculation.

Cross-validation of a regression model. Leaving One sample Out at time (LOO) and Q^2 calculation. Definitions of Training set, Evaluation set and Test set. Use of the coefficient of correlation r between actual and predicted activities for the model validation. Chance effects in a regression model: Y scrambling method.

HANSCH AND FREE-WILSON APPROACHES IN QSAR

Introduction to Hansch approach. Hansch equations. Choice of independent variables and quadratic terms in Hansch equations. Interpretation of the Hansch equations. Examples of Hansch equations for sets of α -adrenergic antagonists and antimalaric drugs. Craig plots.

Free-Wilson approach: brief description and definition of indicator variables.

PRINCIPAL COMPONENT ANALYSIS (PCA) E PARTIAL LEAST SQUARES (PLS)

Correlation matrix among variables. Collinearity. Data pre-treatment: autoscaling. Basic features of the Principal Component Analysis. Geometric interpretation of the principal components. Scores e Loadings: matrices, plots and their interpretations. Eigenvalues. PCA as classification technique. Case study 3: PCA and binding profile of antipsychotic drugs (J.H.L. Lange et al., *J. Med. Chem.* 50, 5103-5108, 2007). Regression on principal components (PCR).

Principles of PLS technique. Features of the latent variables. Definition of PRESS and its application in PLS. Geometric interpretation of the latent variables.

3-D QSAR: COMPARATIVE MOLECULAR FIELD ANALYSIS (CoMFA)

Principles of the method. IUPAC definition of "pharmacophore". Steric and electrostatic fields. Contour maps interpretation. Case study 4: CoMFA study of piperidine analogues of cocaine at the dopamine transporter (H. Yuan et al. *J. Med. Chem.* 47, 6137-6143, 2004).

Conoscenze, competenze e capacità acquisite

Learning outcomes

A completamento con successo del corso, lo studente avrà conoscenze e competenze su:

1. i diversi tipi di rappresentazione della struttura molecolare (1D, 2D, 3D). Sarà in grado di applicare la notazione SMILES e di costruire le tabelle di connessione e le matrici di adiacenza e di distanza topologica. Sarà in grado di calcolare il Tanimoto coefficient considerando le bitstrings di due diverse molecole;
2. le interazioni farmaco-recettore. Sarà anche in grado di descrivere il Radioligand Binding Assay e di calcolare il valore di K_i dalla curva di spiazzamento;
3. i principali descrittori molecolari di tipo sterico, elettronico, o di lipofilia e sui più comuni descrittori topologici per la parametrizzazione delle molecole di interesse; sarà capace di descrivere i metodi di calcolo, utilizzando software opportuno, o di misura sperimentale di alcuni descrittori (ad es.: $\log P$).
4. le principali linee guida nella progettazione di molecole drug-like. Sarà in grado di riconoscere i rotatable bonds in una struttura e di calcolare la ligand efficiency (LE) di un ligando;
5. le tecniche di analisi di regressione lineare semplice e multipla e sulla scelta del tipo e del numero di variabili indipendenti da inserire in un modello. Avrà conoscenze sull'approccio di Hansch e sull'utilizzo delle tecniche Principal Component Analysis (PCA) e Partial Least Square (PLS) in QSAR; sarà in grado di descrivere i principali parametri statistici per la validazione di un modello QSAR;
6. la tecnica ComFA in 3D-QSAR.

On successful completion of this course students will have knowledge and competence about:

1. different representations of molecular structure (1D, 2D, 3D). He or she will be able to apply SMILES notation and to construct connection tables and adjacency and topological distance matrices. He or she will be able to calculate the Tanimoto coefficient taking into account bitstrings of two different molecules;
2. the drug-receptor interactions. He or she will be also able to describe the Radioligand Binding Assay and to calculate the K_i value from the displacing curve;
3. the principal steric, electronic and lipophilic molecular descriptors and the main topological descriptors useful in the parameterization of molecules; he or she will be able to describe the calculation methods using suitable software or the experimental methods for the measure of some descriptors (e.g.: $\log P$);
4. the acknowledged guidelines for the design of drug-like molecules. He or she will be able to identify the rotatable bonds in a structure and to calculate the ligand efficiency (LE) of a ligand;
5. techniques of simple and multiple linear regression analysis and about the choice of the number and type of independent variables to be inserted in a model. He or she will be able to describe the Hansch approach and the use of Principal Component Analysis (PCA) and Partial Least Square (PLS) methods in QSAR; he or she will be able to describe the main statistical parameters for the validation of a QSAR model;
6. the ComFA method in 3D-QSAR.

Modalità d'insegnamento e delle verifiche d'apprendimento

Teaching and learning procedures

Il corso prevede delle lezioni frontali in aula durante le quali vengono presentati dettagliatamente tutti gli argomenti del programma. Gli studenti sono attivamente

chiamati alla discussione in aula e allo studio di casi, illustrativi di alcuni importanti aspetti del programma.

During class lectures, all the subjects are presented in details. Students are asked to actively discuss about them and about a number of case studies, descriptive of some important arguments of the course.

Descrizione delle modalità d'esame **Description of Assessment Items**

L'esame finale è costituito da una prova scritta (comprendente domande a risposta multipla e domande a risposta aperta).

Per sostenere l'esame è necessario prenotarsi on line accedendo con le proprie credenziali al Portale studenti sul sito dell'Università di Catania (www.unict.it). La lista di prenotazione è accessibile circa quindici giorni prima della data d'appello prevista sul calendario.

The final exam is a written test (containing a mix of multiple-choice questions and open questions). To take the exam, students need to book on line at Portale studenti of Università di Catania webpage (www.unict.it). Booking list is opened about fifteen days before the exam date.

Materiale didattico **Course materials**

Copia del materiale presentato durante le lezioni è disponibile presso l'aula studio *Pappalardo* della Biblioteca di Facoltà (Ed. 2).

A hard copy of slides presented during classes is available at the *Pappalardo* Reading room of the Faculty Library (Building 2).

Testi consigliati e bibliografia **Textbooks and bibliography**

1. G.L. Patrick, *Introduzione alla Chimica Farmaceutica*, II Edizione, Edises, 2010.
2. G. Schneider, K.-H. Baringhaus, *Molecular Design*, Wiley-VCH, 2008.
3. D. Livingstone, *A practical guide to scientific data analysis*, Wiley, 2009.
4. D.J. Abraham (Edited by), *Burger's Medicinal Chemistry & Drug Discovery*, Vol.1, Sixth edition, Wiley, 2003.
5. C. Hansch, *Comprehensive Medicinal Chemistry*, Vol. 4, Pergamon Press, 1990.