Cheminformatics and QSAR Modeling Training Course

Professional Development Training Course Outline

Category	Biotechnology and Pharmaceutical Development	**Duration**	10 Days
Base Fee	$4,000 USD	**Delivery Mode**	Online / On-site Hybrid

Course Introduction

The convergence of chemistry, computer science, and data analytics has given rise to Cheminformatics, a pivotal discipline in modern Drug Discovery and Materials Science. Cheminformatics and QSAR Modeling Training Course is meticulously designed to equip professionals and researchers with the cutting-edge computational skills to manage, analyze, and interpret vast chemical datasets. By mastering chemical information systems, molecular representation, and large-scale data curation, participants will learn to transform raw molecular structures into actionable knowledge. The core focus is on practical proficiency with industry-standard, Open-Source Toolkits like RDKit, KNIME, and scikit-learn, ensuring immediate applicability of learned concepts to real-world challenges in the Biopharma and chemical industries.

This program places a strong emphasis on Quantitative Structure-Activity Relationship (QSAR) Modeling, the fundamental technique for developing predictive models that link a molecule's structure to its properties. Leveraging the latest advancements in Artificial Intelligence (AI) and Machine Learning (ML), the course covers the entire QSAR Workflow, from generating robust Molecular Descriptors and Fingerprints to building, validating, and interpreting complex models. By learning to predict critical properties like ADMET (Absorption, Distribution, Metabolism, Ex Excretion, and Toxicity) in silico, participants will gain the ability to significantly accelerate the R&D Pipeline, reduce costly experimental work, and implement Rational Drug Design strategies for identifying and optimizing next-generation drug candidates.

Programme Curriculum

Cheminformatics and QSAR Modeling Training Course

Introduction

The convergence of chemistry, computer science, and data analytics has given rise to Cheminformatics, a pivotal discipline in modern Drug Discovery and Materials Science. Cheminformatics and QSAR Modeling Training Course is meticulously designed to equip professionals and researchers with the cutting-edge computational skills to manage, analyze, and interpret vast chemical datasets. By mastering chemical information systems, molecular representation, and large-scale data curation, participants will learn to transform raw molecular structures into actionable knowledge. The core focus is on practical proficiency with industry-standard, Open-Source Toolkits like RDKit, KNIME, and scikit-learn, ensuring immediate applicability of learned concepts to real-world challenges in the Biopharma and chemical industries.

This program places a strong emphasis on Quantitative Structure-Activity Relationship (QSAR) Modeling, the fundamental technique for developing predictive models that link a molecule's structure to its properties. Leveraging the latest advancements in Artificial Intelligence (AI) and Machine Learning (ML), the course covers the entire QSAR Workflow, from generating robust Molecular Descriptors and Fingerprints to building, validating, and interpreting complex models. By learning to predict critical properties like ADMET (Absorption, Distribution, Metabolism, Ex Excretion, and Toxicity) in silico, participants will gain the ability to significantly accelerate the R&D Pipeline, reduce costly experimental work, and implement Rational Drug Design strategies for identifying and optimizing next-generation drug candidates.

Course Duration

10 days

Course Objectives

1. Master the fundamentals of Cheminformatics Data Curation and management of large-scale chemical libraries.
2. Achieve proficiency in calculating and interpreting diverse Molecular Descriptors (0D-3D) and Molecular Fingerprints.
3. Develop robust QSAR Modeling and QSPR (Quantitative Structure-Property Relationship) models using both linear and non-linear regression.
4. Apply Machine Learning (ML) algorithms for predictive chemical tasks.
5. Implement Virtual Screening (VS) and High-Throughput Screening (HTS) techniques for Hit Identification.
6. Understand and predict ADMET properties and Computational Toxicology using validated models.
7. Gain hands-on expertise with key Open-Source Cheminformatics Tools
8. Master the principles and application of 3D-QSAR methods like CoMFA and CoMSIA for structural optimization.
9. Apply Deep Learning for Molecules in advanced property prediction.
10. Systematically validate and assess the Applicability Domain (AD) of QSAR models to ensure reliability.
11. Perform effective analysis of Structure-Activity Relationship (SAR) data for lead optimization.
12. Integrate Cheminformatics Workflows with Data Science practices for reproducible research.
13. Utilize Generative Chemistry models for de novo design of molecules with desired properties.

Target Audience

1. Medicinal Chemists and Synthetic Chemists.
2. Computational Chemists and Cheminformaticians.
3. Data Scientists and Bioinformaticians.
4. Toxicologists and Regulatory Scientists focused on In Silico Toxicology and risk assessment (REACH/Tox21).

5. Graduate Students (M.Sc./Ph.D.) and Postdoctoral Researchers in Chemistry, Biology, and Pharmacy.
6. R&D Professionals.
7. Software Developers and Bio-IT Specialists.
8. Academics and Educators.

Course Modules

Module 1: Foundations of Cheminformatics and Chemical Data

- Introduction to Cheminformatics and its role in Drug Discovery.
- Chemical structure representation
- Handling and standardizing chemical data from databases
- Chemical file formats and data exchange
- Case Study: Curating a quality-controlled dataset for a new therapeutic target from public databases.

Module 2: Molecular Descriptors and Fingerprints

- Classification of molecular descriptors
- Calculation of physicochemical properties
- Introduction to Molecular Fingerprints
- Similarity searching and chemical space analysis using Tanimoto coefficient.
- Case Study: Comparing the effectiveness of ECFP4 and RDKit topological fingerprints for compound clustering.

Module 3: Introduction to QSAR/QSPR Modeling

- Historical context and fundamental principles of QSAR
- The QSAR Workflow.
- Understanding the Structure-Activity Relationship (SAR) and its graphical representation.
- Activity standardization and outlier removal.
- Case Study: Developing a simple QSAR model for the inhibition of an enzyme using classical regression.

Module 4: Machine Learning Fundamentals for QSAR

- Supervised and Unsupervised Learning in Cheminformatics.
- Regression and Classification models
- Model selection, training, and test set division
- Evaluation metrics: R2, RMSE, AUC-ROC, Accuracy.
- Case Study: Implementing a Random Forest classifier to predict compound 'activity' or 'inactivity' against a cancer cell line.

Module 5: Advanced QSAR Algorithms

- Partial Least Squares (PLS) and its application in high-dimensional data.
- Support Vector Machines (SVM) for non-linear QSAR modeling.
- Feature selection techniques
- Ensemble methods for improved performance.
- Case Study: Using PLS to develop a robust model for predicting the aqueous solubility of small molecules.

Module 6: Predictive ADMET Modeling

- The importance of ADMET in the Pre-clinical phase.
- In silico models for Absorption and Distribution
- Predicting Metabolism and Excretion.
- Case Study: Building a QSAR model for predicting the blood-brain barrier penetration of CNS-active compounds.

Module 7: Computational Toxicology and Safety Assessment

- Fundamentals of In Silico Toxicology and the 3Rs principle.
- Predicting endpoints like Mutagenicity and Hepatotoxicity.
- Introduction to the OECD Principles for QSAR model validation.
- Model interpretation for mechanistic insights
- Case Study: Using an existing QSAR model to flag potential genotoxic risk in a new chemical library.

Module 8: Model Validation and Applicability Domain

- Internal validation techniques
- External validation using an independent test set.
- Defining and calculating the Applicability Domain (AD) of a QSAR model.
- The critical role of Y-Randomization in assessing model chance correlation.
- Case Study: Critically evaluating the predictive power and AD of a published QSAR model using a new external dataset.

Module 9: Working with RDKit in Python

- Introduction to the RDKit library for molecular manipulation.
- Generating 2D/3D molecular structures and visualization.
- Implementing descriptor and fingerprint calculations in Python scripts.
- Substructure searching and reaction modeling using RDKit.
- Case Study: Automating a workflow in Python/RDKit to normalize structures and calculate 50+ molecular descriptors for a batch of compounds.

Module 10: Visual and Unsupervised Cheminformatics

- Molecular visualization techniques
- Clustering chemical space
- Dimensionality reduction for visualization
- Analyzing Diversity and Coverage of chemical libraries.
- Case Study: Using PCA and t-SNE to map and visualize the chemical space of an internal compound library compared to a commercial one.

Module 11: 3D-QSAR and Conformation

- Generating and managing molecular 3D conformations.
- Alignment and Superposition methods for congeneric sets.
- Introduction to 3D-QSAR fields
- Basic concepts of Comparative Molecular Field Analysis methodology.
- Case Study: Performing molecular minimization and conformational analysis on a set of active ligands to identify the bioactive conformer.

Module 12: Virtual Screening and Library Design

- Overview of Ligand-Based Virtual Screening (LBVS) methods.
- Pharmacophore Modeling for hit identification.
- Similarity searching and its application in Scaffold Hopping.
- Applying Drug-Likeness filters to screen libraries.
- Case Study: Implementing a Pharmacophore search to screen a large database for novel hits with a specific activity profile.

Module 13: Introduction to Deep Learning for Molecules

- Review of Neural Networks and their architecture.
- Application of Deep Learning for complex property prediction.
- Introduction to Graph Neural Networks (GNNs) for molecular data.
- Transfer learning concepts in Cheminformatics.
- Case Study: Training a simple Deep Neural Network to predict an ADMET endpoint using Morgan fingerprints as input.

Module 14: Integrating Workflows with KNIME

- Introduction to KNIME for visual, node-based workflow design.
- Building end-to-end Cheminformatics pipelines without extensive coding.
- Data manipulation, filtering, and visualization within KNIME.
- Integrating RDKit and Machine Learning nodes in a KNIME workflow.
- Case Study: Designing a KNIME workflow that imports a chemical file, calculates descriptors, trains a QSAR model, and generates a prediction report.

Module 15: De Novo Design and Generative Models

- Principles of de novo design and lead optimization strategies.
- Introduction to Generative Adversarial Networks and VAEs for molecule generation.
- Controlling the generation process for molecules with desired properties.
- Retrosynthesis prediction using computational tools.
- Case Study: Exploring an open-source Generative Model to propose novel chemical structures with high predicted activity and low toxicity.

Training Methodology

This course employs a **blended, highly practical methodology** designed to maximize retention and application of skills:

- Interactive Lectures
- Hands-on Workshops
- Real-World Case Studies.
- Project-Based Learning.
- Live Q&A and Troubleshooting.

Register as a group from 3 participants for a Discount

Send us an email: info@fineskilltrainingcenter.org or call +254769199797

Certification

Upon successful completion of this training, participants will be issued with a globally- recognized certificate.

Tailor-Made Course

We also offer tailor-made courses based on your needs.

Key Notes

a. The participant must be conversant with English.

b. Upon completion of training the participant will be issued with an Authorized Training Certificate

c. Course duration is flexible and the contents can be modified to fit any number of days.

d. The course fee includes facilitation training materials, 2 coffee breaks, buffet lunch and A Certificate upon successful completion of Training.

e. One-year post-training support Consultation and Coaching provided after the course.

f. Payment should be done at least a week before commence of the training, to Fineskill Training Center account, as indicated in the invoice so as to enable us prepare better for you.

Upcoming Scheduled Sessions

Start Date	End Date	Location	Price
23 Nov 2026	04 Dec 2026	Online	$2,000
23 Nov 2026	04 Dec 2026	Physical	$3,000
23 Nov 2026	04 Dec 2026	Physical	$6,000
23 Nov 2026	04 Dec 2026	Physical	$6,000
23 Nov 2026	04 Dec 2026	Physical	$12,000
23 Nov 2026	04 Dec 2026	Physical	$9,000
23 Nov 2026	04 Dec 2026	Physical	$12,000
23 Nov 2026	04 Dec 2026	Physical	$12,000

Ready to enroll or request a customized program? Book online or contact us:

Register Online Now

Email: info@fineskilltrainingcenter.com | Website: www.fineskilltrainingcenter.com