endogena

ENDOGENOUS REGENERATIVE MEDICINE

FOR OPHTHALMOLOGY AND BEYOND

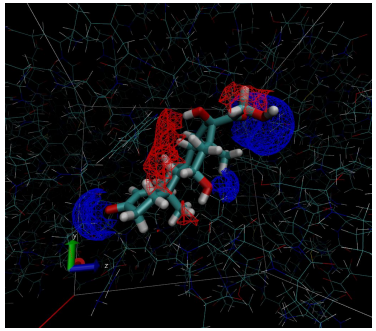
How Deep Learning has revolutionized 3D modelling in drug discovery

WHY DO WE MODEL THE WORLD IN A 3D SPACE?

FROM COMPLEXITY TO VISUAL SIMPLICITY

$$H\Psi = E\Psi$$

- Revolutionary equation
- Take extreme complex information
- Ψ (x,y,z) , H interactions of atoms, E total energy of a molecule



- Visualize this complex information
- Represent the 3D structure of a molecule
- Predict how molecules interact in a biological sense

time

Erwin Schrödinger
year 1925

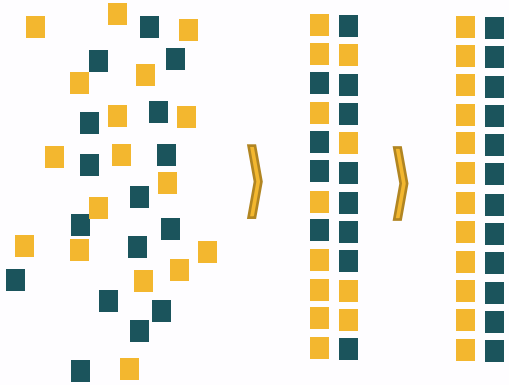
Us
year 2022

"PATTERN RECOGNITION" IN DRUG DISCOVERY

HUMAN APPROACH TO STRUCTURE DATA

Human approach

A



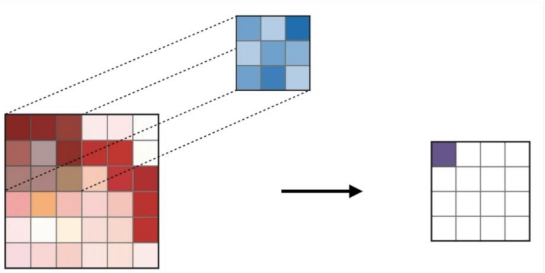
Non-structured information

Structured information

Automation

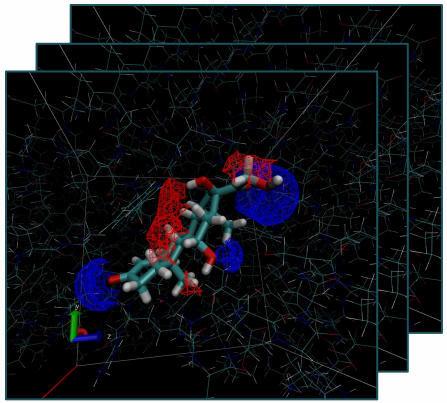
3D Conv Nets in general

B



Big data analysis

C

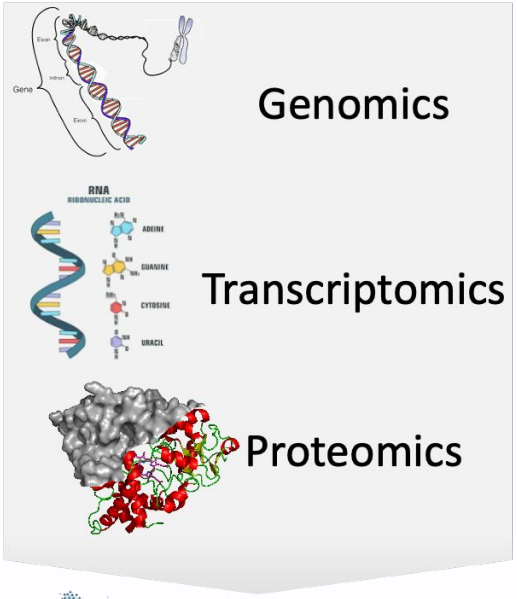


Human engineering

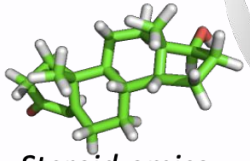
Deep learning is a fast way to analyse massive amount of data and to find patterns

RETENTION TIME PREDICTION VIA 3D CONVNET

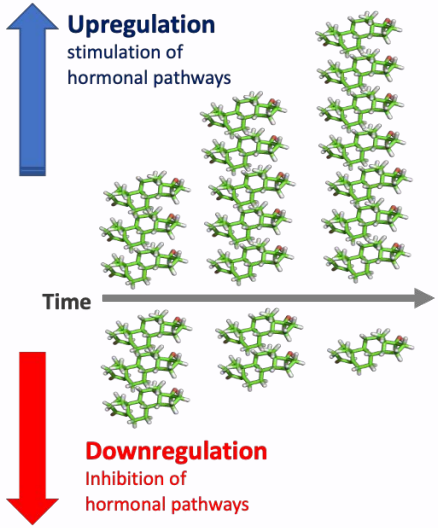
STEROIDOMIC CASE STUDY



Endocrine disrupting chemicals



Steroid-omics



Metabolic diseases

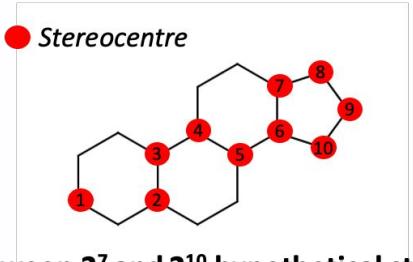
Journal of Chromatography A
Volume 1612, 8 February 2020, 460661

Steroid identification via deep learning retention time predictions and two-dimensional gas chromatography-high resolution mass spectrometry

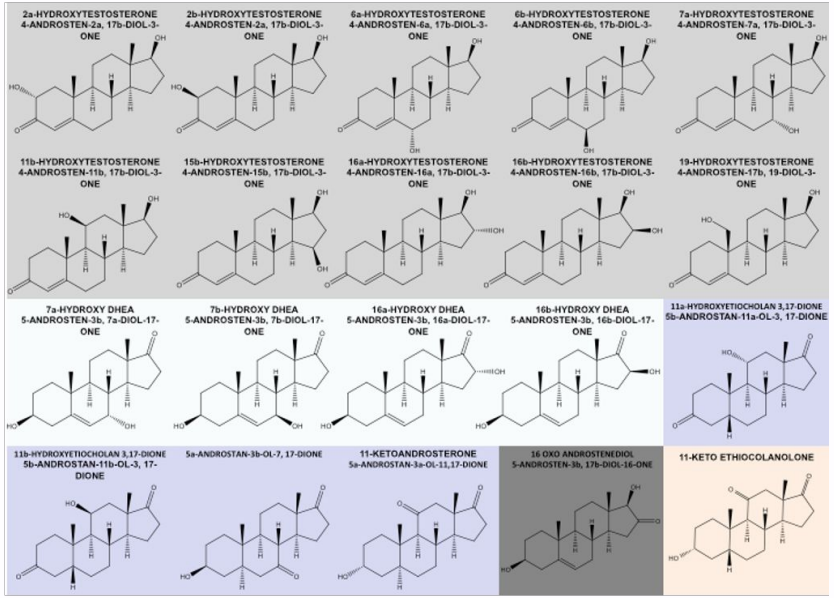
Giuseppe Marco Randazzo [a], Andrea Bileck [b], Andrea Danani [a], Bruno Vogt [b], Michael Groessl [b]

STEROIDS

THE IDENTIFICATION PROBLEM



… between 2^7 and 2^{10} hypothetical structures …



20 isomers for single exact mass!

Randazzo et al. 2016 doi:10.1016/j.aca.2016.02.014

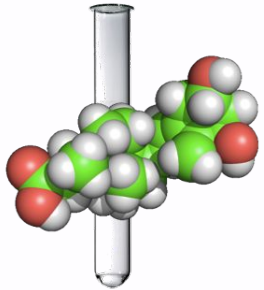
Unambiguous identification is an analytical challenge!

WETLAB 2D GAS CHROMATOGRAPHY ANALYSIS

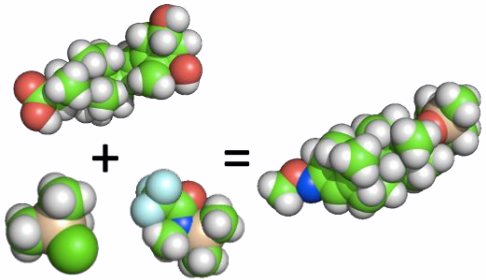
CAN YOU IDENTIFY MOLECULES?

Analytical Steps

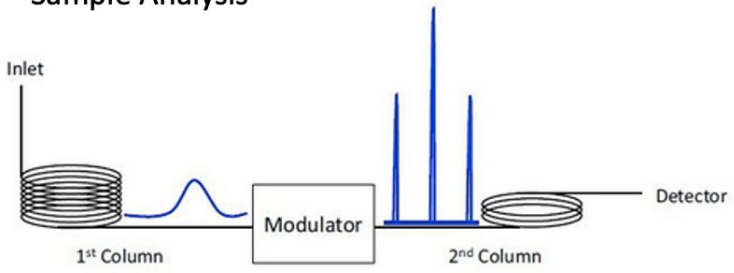
Sample preparation



Derivatization

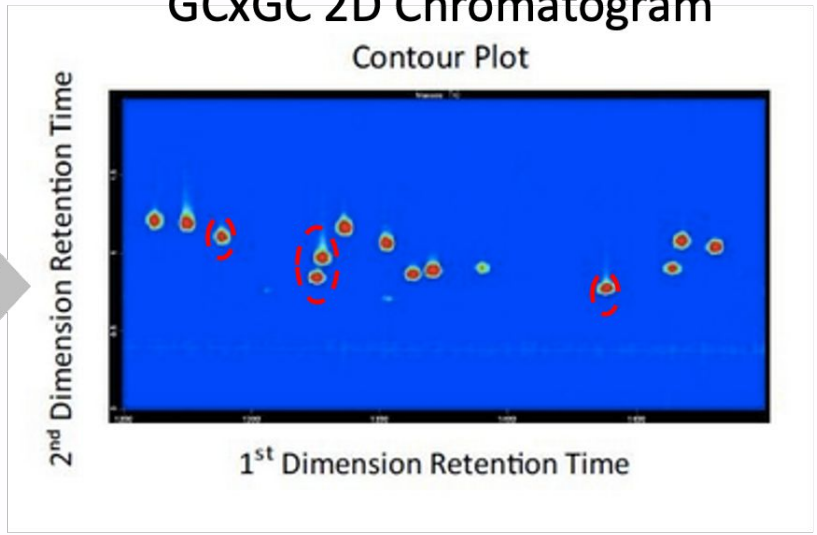


Sample Analysis



Output

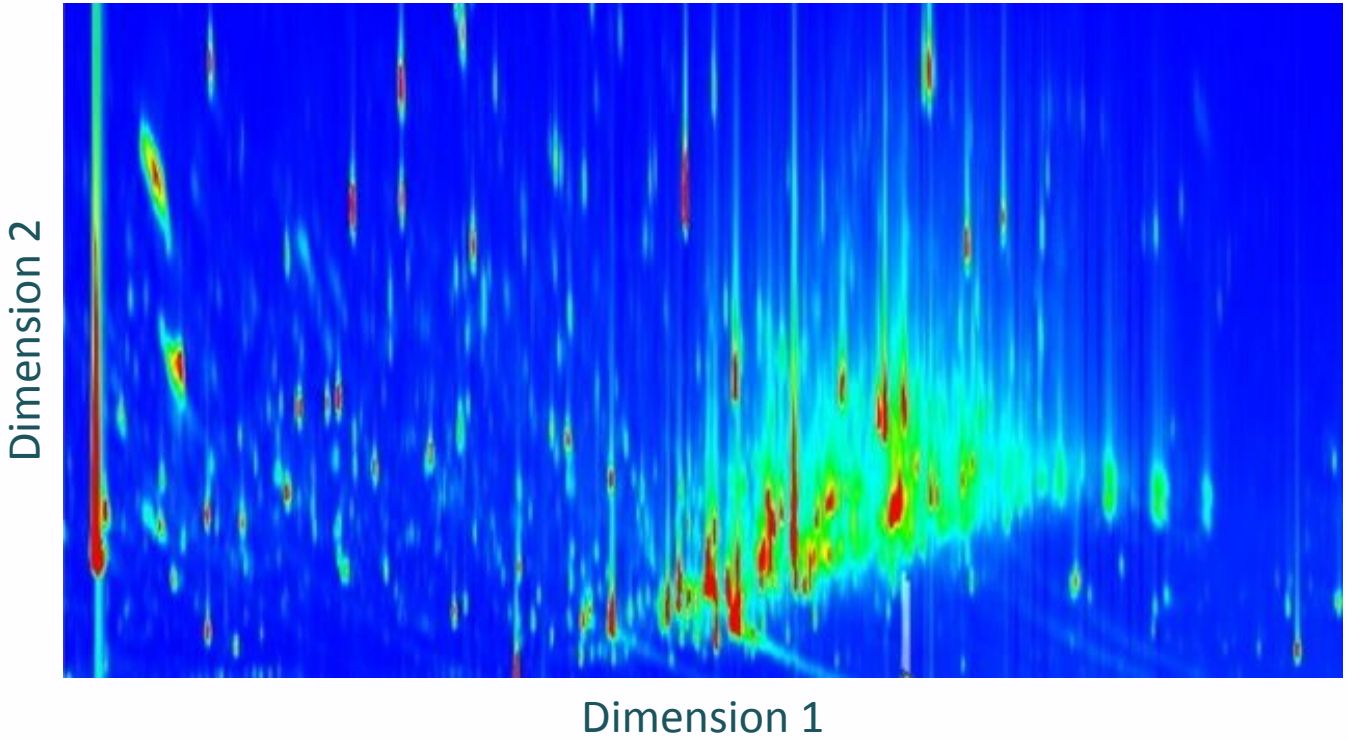
GCxGC 2D Chromatogram



Standard compounds are easy identified in this 2D plot

REAL SAMPLE

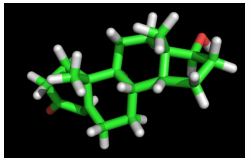
CAN YOU IDENTIFY MOLECULES?



WHAT DEEP LEARNING BRINGS US TODAY?

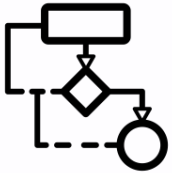
CLASSICAL ML METHODS VS NEW ONES: FAIR COMPARISON

Classical machine learning



Input

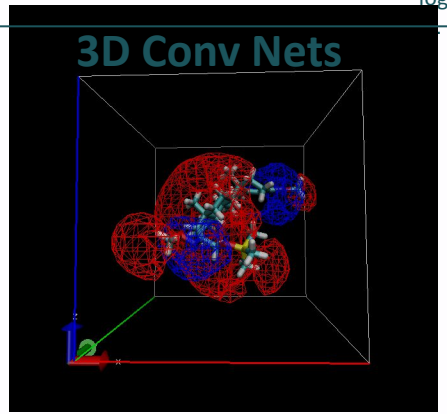
Handcrafted descriptors
C atoms
rotable bonds
log $P_{n\text{-}oct}$



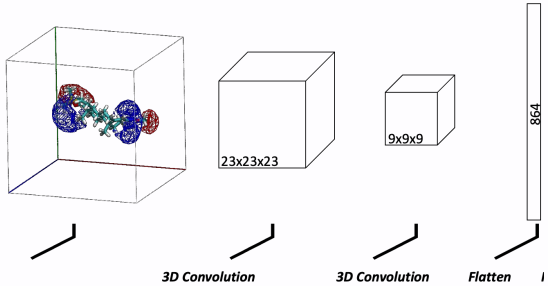
Regression

V.S.

3D Conv Nets



Input

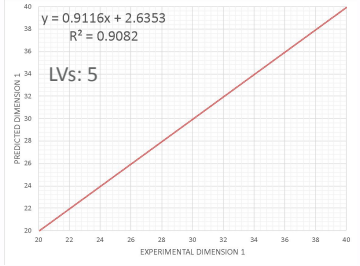
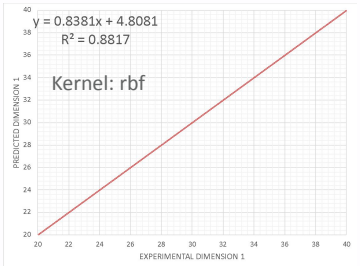
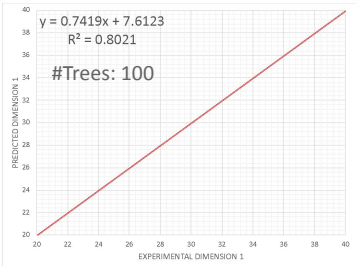
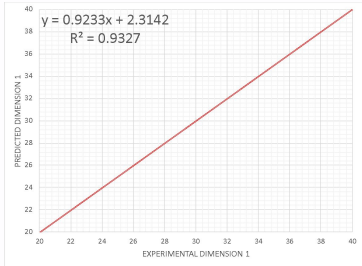
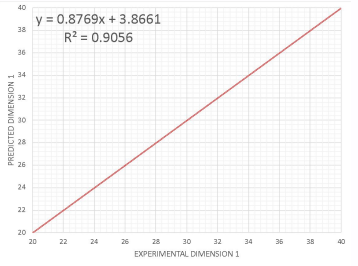
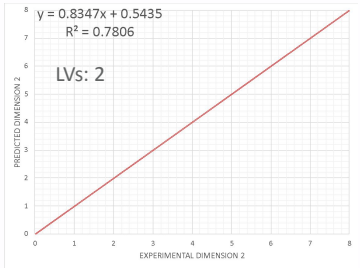
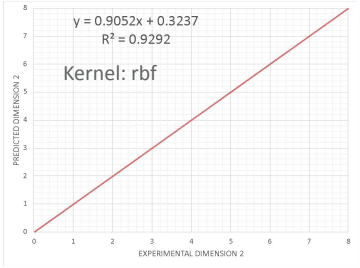
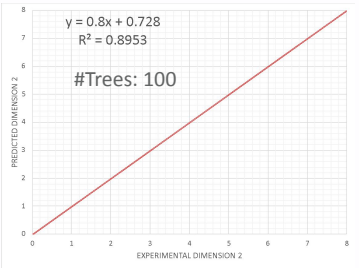
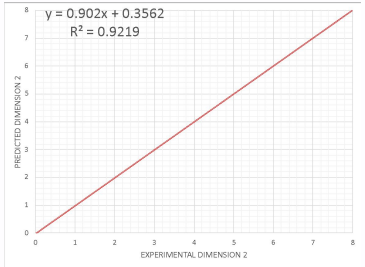
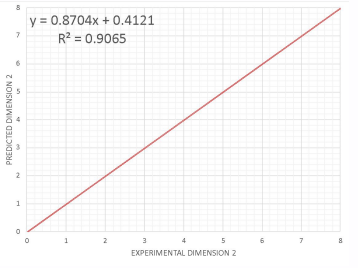


Neural Network

9

CROSS VALIDATION RESULTS

CLASSICAL ML METHODS VS 3D CONV NETS

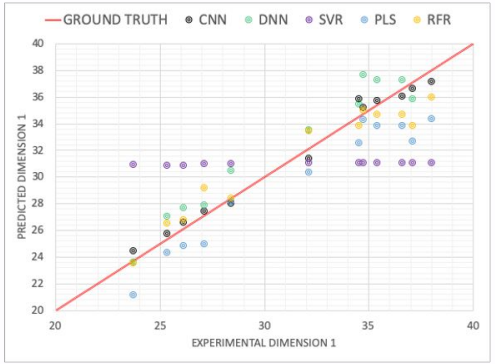
	PLS	SVM	Random forest	DNN	3D Conv Net
Dimension 1					
Dimension 2					

REAL IDENTIFICATION CASE STUDY

ALGORITHM COMPARISONS

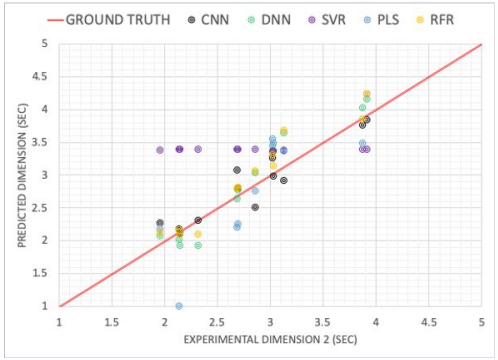
Target t_R dimension 1

Model	Q2	RMSE	MSE	MAE	% Avg. Rel. Err
3D CNN	0.98	0.66	0.44	0.60	1.92%
DNN	0.95	1.75	3.06	1.53	4.77%
PLS	0.94	2.32	5.38	1.99	6.21%
SVR	0.61	4.92	24.22	4.61	15.04%
RFR	0.93	1.53	2.33	1.21	3.70%



Target t_R dimension 2

Model	Q2	RMSE	MSE	MAE	% Avg. Rel. Err
3D CNN	0.89	0.21	0.04	0.16	5.97%
DNN	0.93	0.27	0.07	0.23	8.36%
PLS	0.74	0.71	0.50	0.58	23.04%
SVR	0.02	0.83	0.69	0.74	30.06%
RFR	0.94	0.23	0.05	0.18	6.35%



These models allow identification of steroids in a blood/urine sample

REVISITING A 1988 METHODOLOGY

WHAT'S THE BENEFIT OF DEEP LEARNING

J. Am. Chem. Soc. **1988**, *110*, 5959–5967

Comparative Molecular Field Analysis (CoMFA). 1. Effect of Shape on Binding of Steroids to Carrier Proteins

Richard D. Cramer, III,* David E. Patterson, and Jeffrey D. Bunce

Contribution from Tripos Associates, 1699 South Hanley Road, St. Louis, Missouri 63144. *Received January 5, 1988*

Abstract: Comparative molecular field analysis (CoMFA) is a promising new approach to structure/activity correlation. Its characteristic features are (1) representation of ligand molecules by their steric and electrostatic fields, sampled at the intersections of a three-dimensional lattice, (2) a new "field fit" technique, allowing optimal mutual alignment within a series, by minimizing the RMS field differences between molecules, (3) data analysis by partial least squares (PLS), using cross-validation to maximize the likelihood that the results have predictive validity, and (4) graphic representation of results, as contoured three-dimensional coefficient plots. CoMFA is exemplified by analyses of the affinities of 21 varied steroids to corticosteroid- and testosterone-binding globulins. Also described are the sensitivities of results to the nature of the field and the definition of the lattice and, for comparison, analyses of the same data using various combinations of other parameters. From these results, a set of ten steroid-binding affinity values unknown to us during the CoMFA analysis were well predicted.

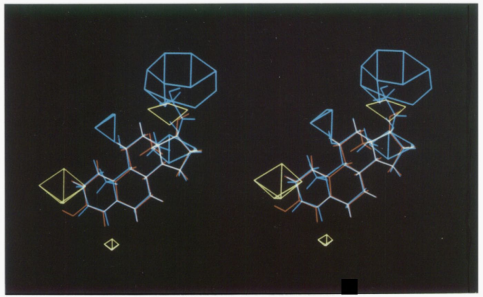
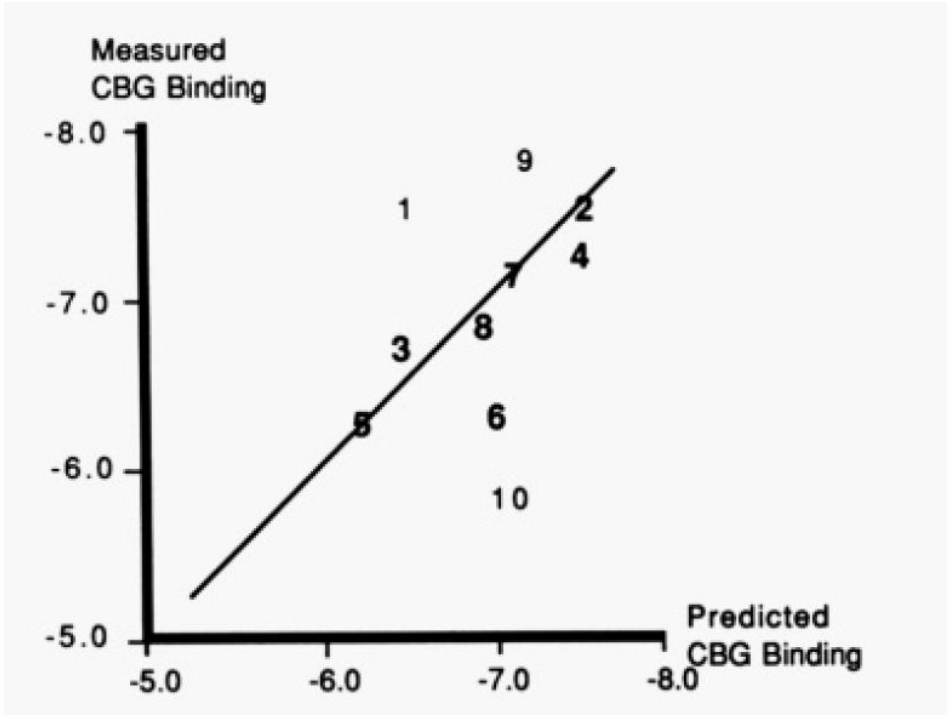
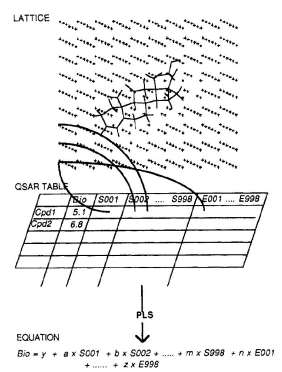


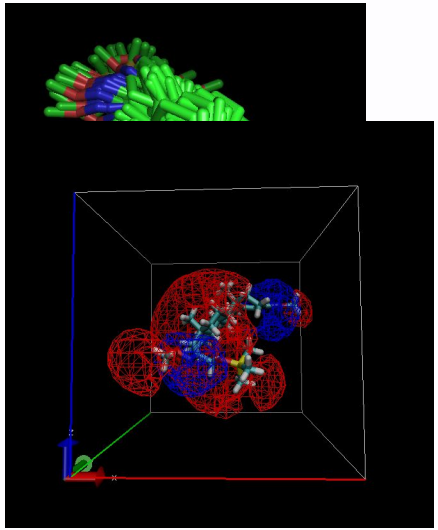
Figure 5. Stereoscopic views of the major steric features of the QSAR for steroid binding to corticosteroid-binding globulin (CBG). Yellow and red contours surround regions where a lower steric interaction would increase binding (the QSAR coefficient times the standard deviation of the corresponding column greater than +0.01 and +0.1, respectively). Blue and cyan contours surround regions where a higher steric interaction would increase binding (less than −0.01 and −0.1, respectively). The red molecule (estradiol) is poorly bound to CBG and the blue molecule (cortisol) strongly bound.



REVISITING A 1988 METHODOLOGY

RESULTS IN 2022





time

1988

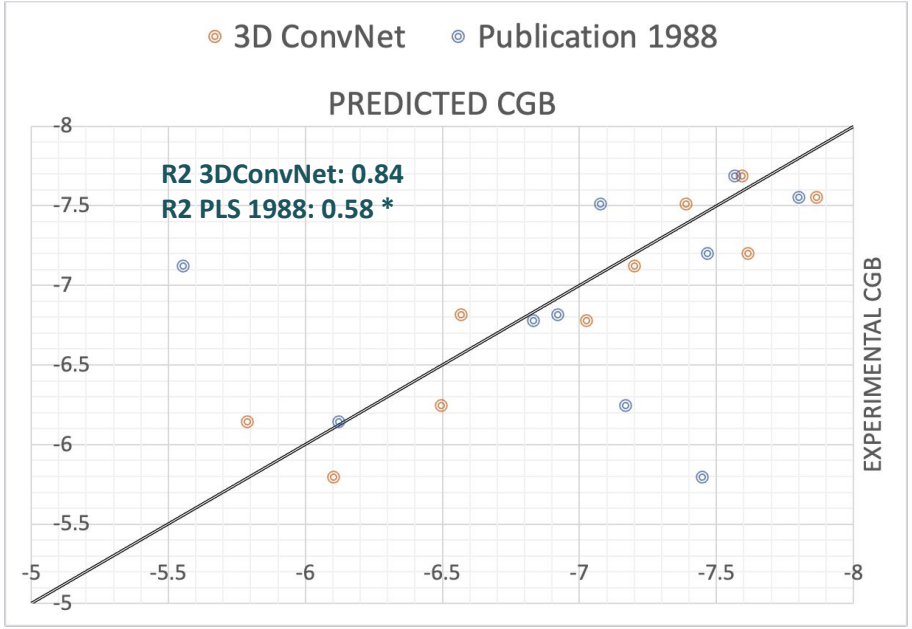
CoMFA

- Align Molecules
- Convert the 3D Voxel representation into a table

2022

3D ConvNet

- No need to align molecules
- No need to transform the 3D representation into a table



* Not considering outliers

IMPACT OF AI IN DRUG DISCOVERY

WHAT ARE THE BENEFITS OF THIS APPROACH TO A START-UP LIKE ENDOGENA

Deep Learning offers better numbers for confident decision making.

HOW?

- Access to a broad compound library selection despite limited resources
- Time efficient and simultaneous look at different areas (solubility, 3D structure, etc.)
- Reduction of experimental time and costs

CURING HIGH UNMET MEDICAL NEEDS
BY ACTIVATING ENDOGENOUS TISSUE REPAIR
AND REGENERATION