Personalized Drug Therapy and the Benefits of Artificial Intelligence

Erdem Buyukbingol, Ph. D.

0026-895X/88/060852-11$02.00/0

MOLECULAR PHARMACOLOGY, 34:852-862

MOLECULAR PHARMACOLOGY

An Artificial Intelligence Approach to the Study of the Structural Moieties Relevant to Drug-Receptor Interactions in Aldose Reductase Inhibitors

GILLES KLOPMAN and ERDEM BUYUKBINGOL
Case Western Reserve University, Department of Chemistry, Cleveland, Ohio 44106
Received November 6, 1987; Accepted August 5, 1988.

1988

SUMMARY

The computer-automated structure evaluation program has been used to study 482 compounds relevant to the inhibition of the aldose reductase enzyme. Major activating/inactivating fragments were generated *automatically*. The significance of these molecular descriptors with respect to the activity of the compounds is discussed.

The aldose reductase enzyme, AR, involved in the sorbitol pathway (Fig. 1), which is an important mechanism in the regulation of mammalian glucose metabolism, has been found to play a physiologically significant role in the initiation of diabetic complications (1, 2). Therefore, over the past few years, considerable efforts have been made by several research groups to design inhibitors of this enzyme. The enzyme is primarily found in the corneal epithelium, retina, optic nerve, placenta, brain, kidney, muscle, and sperm (3–5).

Using NADPH as a cofactor, the AR enzyme catalyzes the transformation of glucose into sorbitol, which, in diabetes, is believed to accumulate in certain tissues such as nerve, kidney, pancreas, retina, and lens. Increased concentration of sorbitol can cause damage to these tissues, leading to diabetic complications such as microangiopathy, nephopathy, neuropathy, ret-

fatty acids (10) and tetramethylene glutaric acid (11), many natural as well as synthetic compounds have been studied, under both experimental and clinical conditions, for the treatment of the aforementioned complications. One class of inhibitory compounds found to be effective is the flavonoids, which are derived from many natural sources and exhibit a broad range of bioactivity (5). Numerous analogs have been prepared in the hope of improving their pharmacological profile with respect to the inhibition of the AR enzyme (8, 12, 13). However, few of the flavonoids have sufficient activity to warrant further investigation. Their low water solubility as well as their inability to penetrate biological membranes such as the blood-retina barrier are additional deterrents to further evaluation of these compounds.

Numerous other molecules have been screened for potential

JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

Cite This: J. Chem. Inf. Model. 2019, 59, 4654−4662

pubs.acs.org/jcim

2019

Three-Dimensional Analysis of Binding Sites for Predicting Binding Affinities in Drug Design

Ozlem Erdas-Cicek,*[,†] Ali Osman Atac,[‡] A. Selen Gurkan-Alp,[§] Erdem Buyukbingol,[§,⊥] and Ferda Nur Alpaslan[‡]

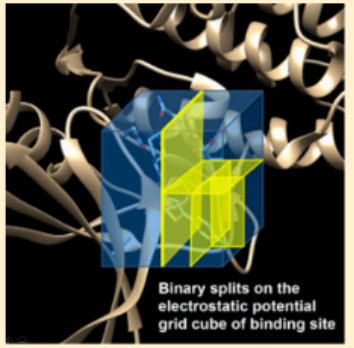
[†]Department of Computer Engineering, Faculty of Engineering, Alanya Alaaddin Keykubat University, Alanya, 07425 Antalya, Turkey

[‡]Department of Computer Engineering, Middle East Technical University, Cankaya, 06800 Ankara, Turkey

[§]Department of Pharmaceutical Chemistry, Faculty of Pharmacy, Ankara University, Yenimahalle, 06560 Ankara, Turkey

[⊥]Pharmaceutical Chemistry, Faculty of Pharmacy, Afyonkarahisar Health Sciences University, 03200 Afyonkarahisar, Turkey

ABSTRACT: Understanding the interaction between drug molecules and proteins is one of the main challenges in drug design. Several tools have been developed recently to decrease the complexity of the process. Artificial intelligence and machine learning methods offer promising results in predicting the binding affinities. It becomes possible to do accurate predictions by using the known protein−ligand interactions. In this study, the electrostatic potential values extracted from 3-dimensional grid cubes of the drug−protein binding sites are used for predicting binding affinities of related complexes. A new algorithm with a dynamic feature selection method was implemented, which is derived from Compressed Images For Affinity Prediction (CIFAP) study, to predict binding affinities of Checkpoint Kinase 1 and Caspase 3 inhibitors.

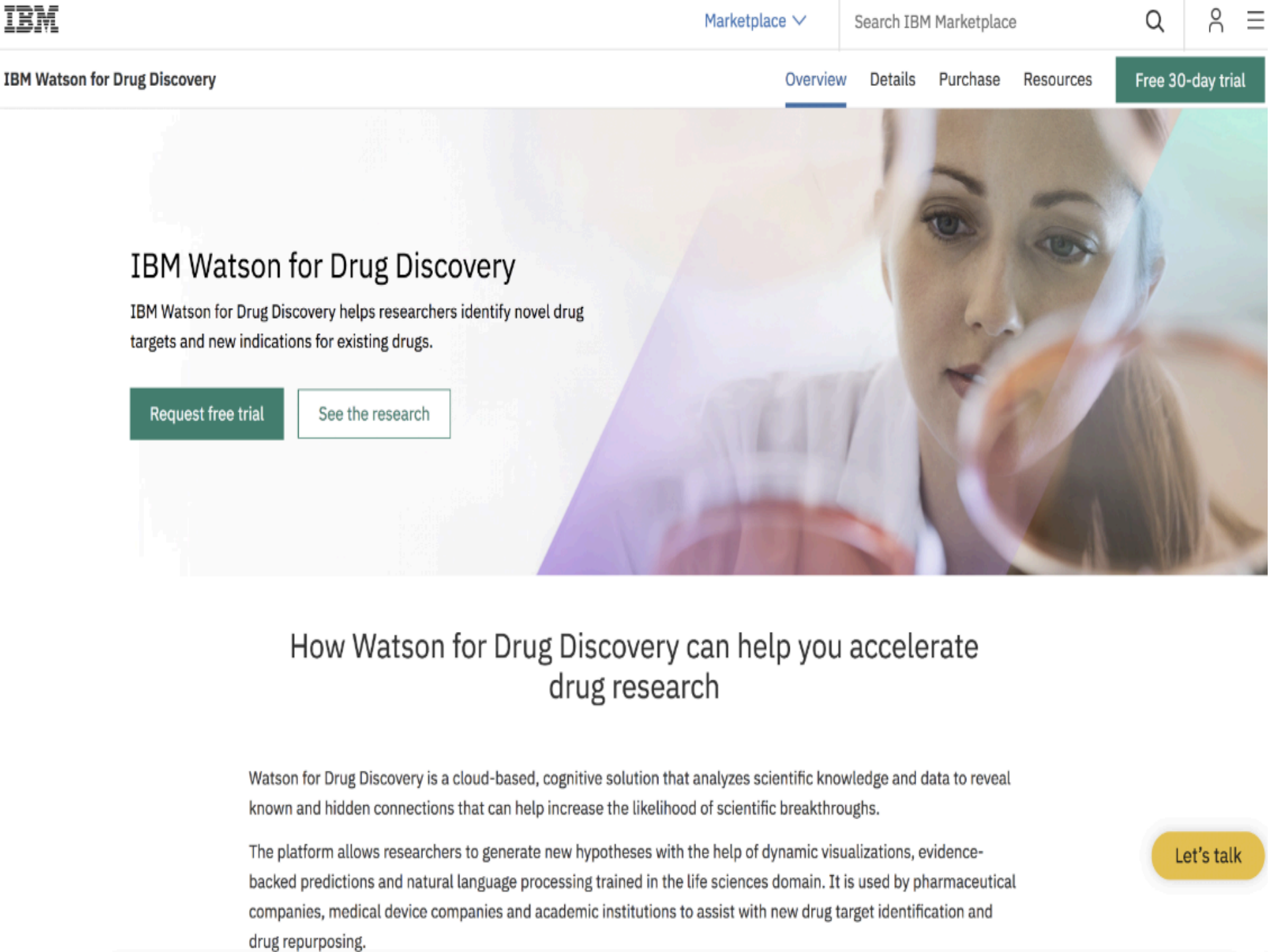


INTRODUCTION

Throughout history, mankind has struggled to cure diseases. In the conventional approach, studies are conducted in laboratories or by trial-and-error testing of drug candidates on animals.[1] Till now, there is a complementary approach called "in silico drug discovery" such that drug experiments can be performed in simulation environments with the help of high-performance computers. Although this has eventually reduced the cost of the process, the drug design is still a challenging task with many successes and failures. In the last decade, more successful approaches using machine learning techniques have become popular in drug discovery. Especially, pharmaceutical companies extensively use machine learning methods to model, analyze, and predict the biological results of a candidate drug in the discovery process.[2]

binding sites which gave the best results in predicting the binding affinity values. Even with binding site analysis, there is a large set of points that should be considered for electrostatic potential values. Thus, there is a need for a data model that is both efficient for time complexity and low on data-loss. In this study, we propose a new machine learning method along with a data model, which is used for predicting the binding affinities of Checkpoint Kinase 1 (CHK1) inhibitors[8] and Caspase3 (CASP3) inhibitors[9] effectively. The idea is inspired from the basic idea defined in the Compressed Images For Affinity Prediction (CIFAP) study.[10,11]

In the previous study of Erdas et al.,[10,11] electrostatic potential values of the three-dimensional (3D) structure of the protein−ligand interaction are used in the data modeling method called CIFAP. In this approach, a candidate ligand is docked into the

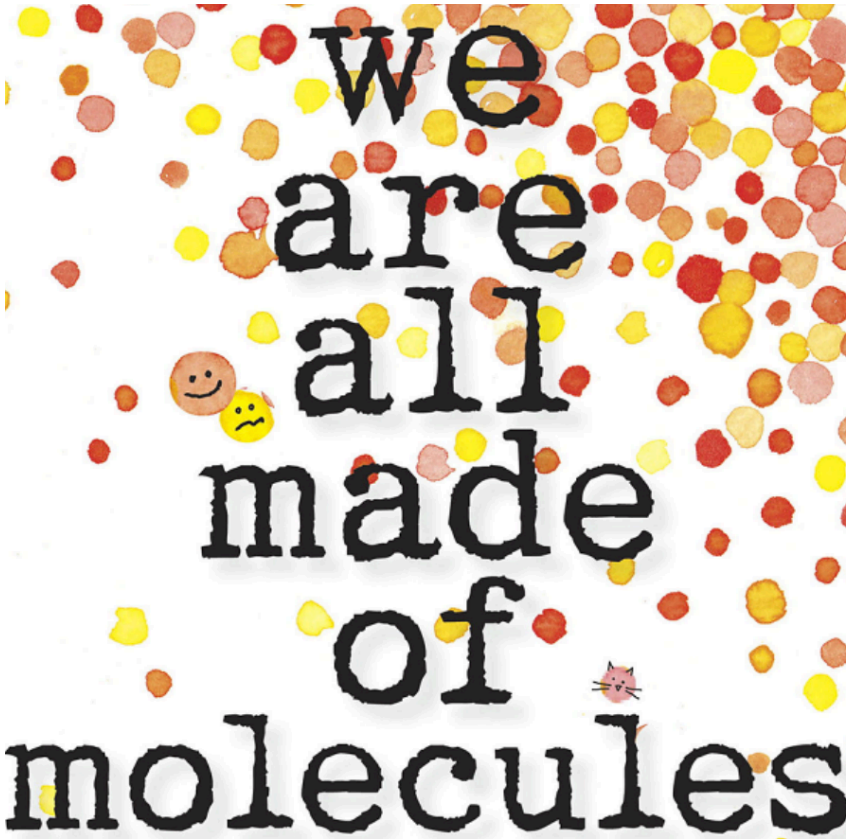


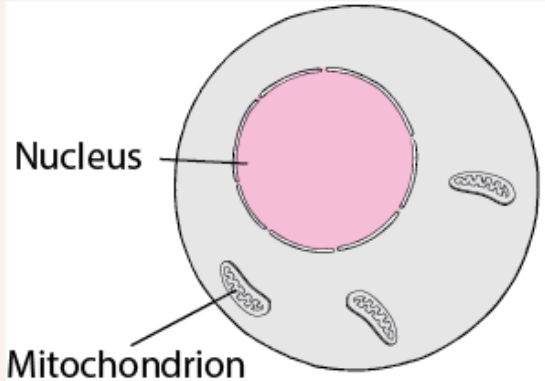
Pfizer Collaborating with IBM

Pharmaceutical giant Pfizer in late 2016 announced a collaboration that will utilize IBM Watson for Drug Discovery. Pfizer is using IBM's AI technology on its immuno-oncology research, a strategy of using a body's immune system to help fight cancer. Based on our research, this appears to be one of the first significant uses of Watson for drug discovery. The move was a big public announcement for both Pfizer and IBM.

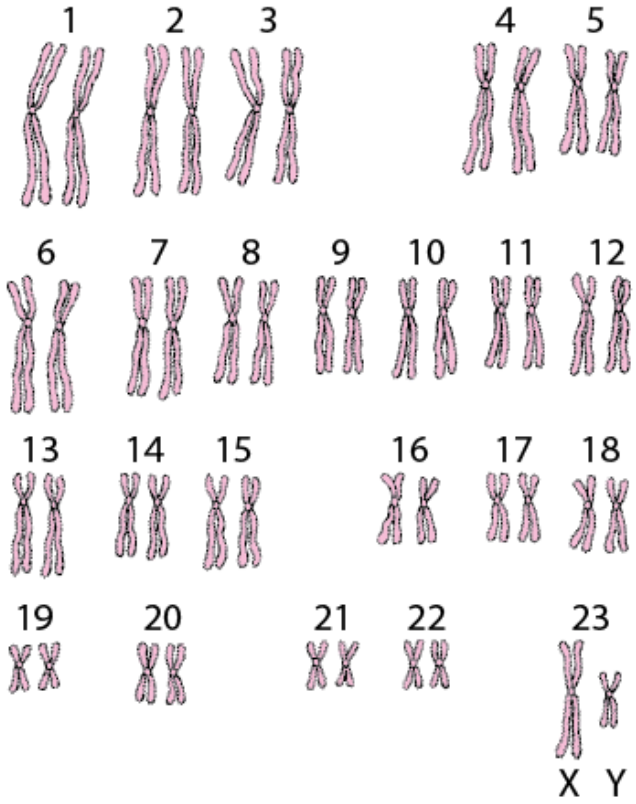
Why do people respond differently to the same drug?

Personalized Drug Therapy
Precision Medicine

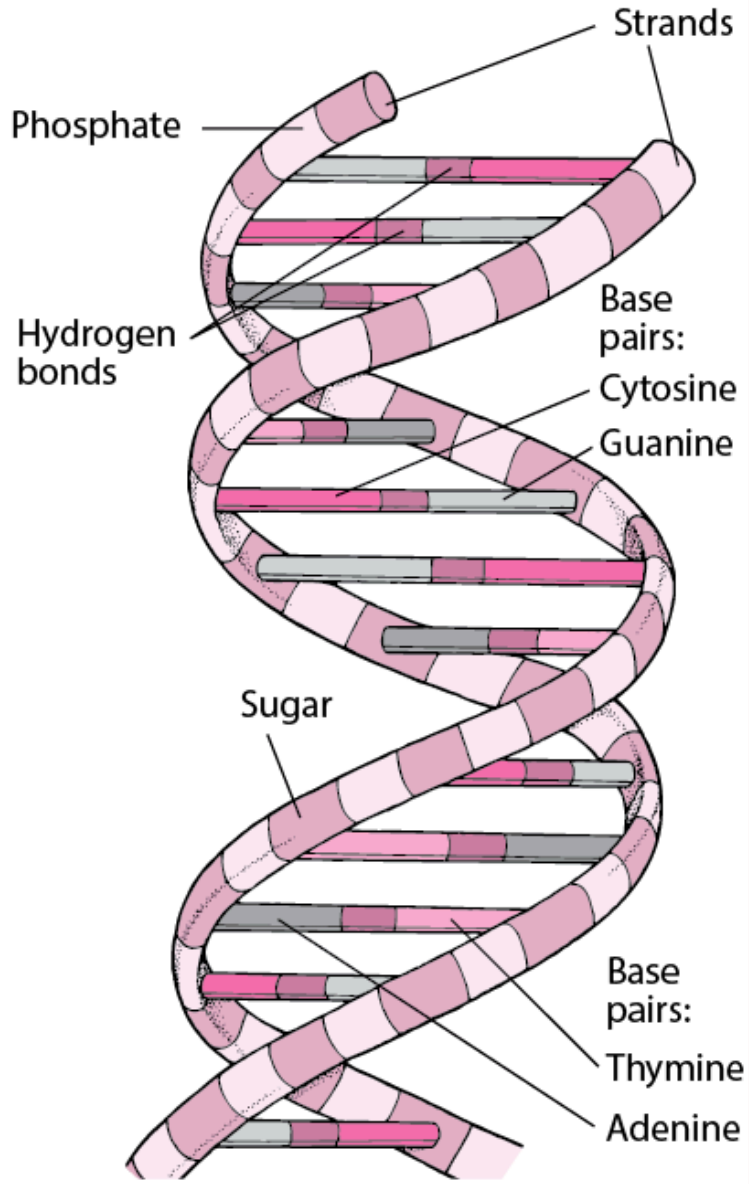




Cell



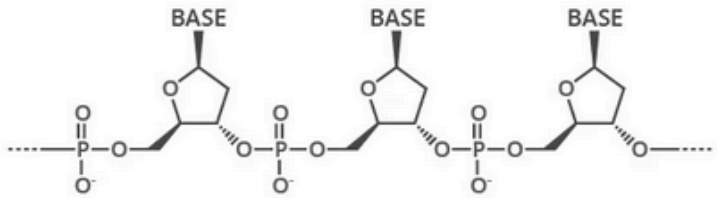
Pairs of Chromosomes in a Human Cell



DNA Double Helix

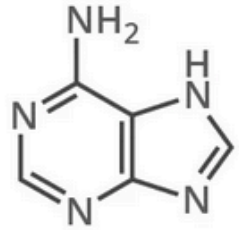
THE CHEMICAL STRUCTURE OF DNA

THE SUGAR PHOSPHATE 'BACKBONE'

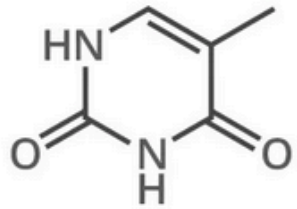


DNA is a polymer made up of units called nucleotides. The nucleotides are made of three different components: a sugar group, a phosphate group, and a base. There are four different bases: adenine, thymine, guanine and cytosine.

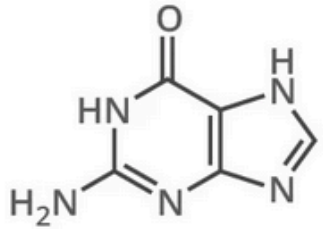
A ADENINE



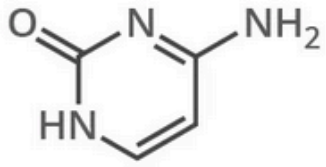
T THYMINE



G GUANINE

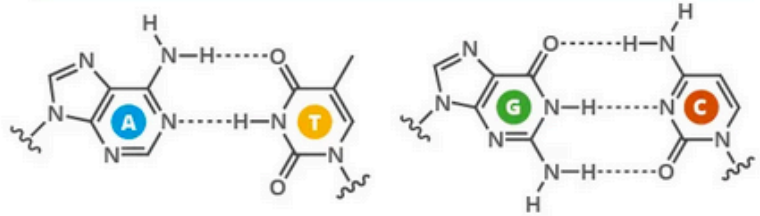


C CYTOSINE



WHAT HOLDS DNA STRANDS TOGETHER?

DNA strands are held together by hydrogen bonds between bases on adjacent strands. Adenine (A) always pairs with thymine (T), while guanine (G) always pairs with cytosine (C). Adenine pairs with uracil (U) in RNA.

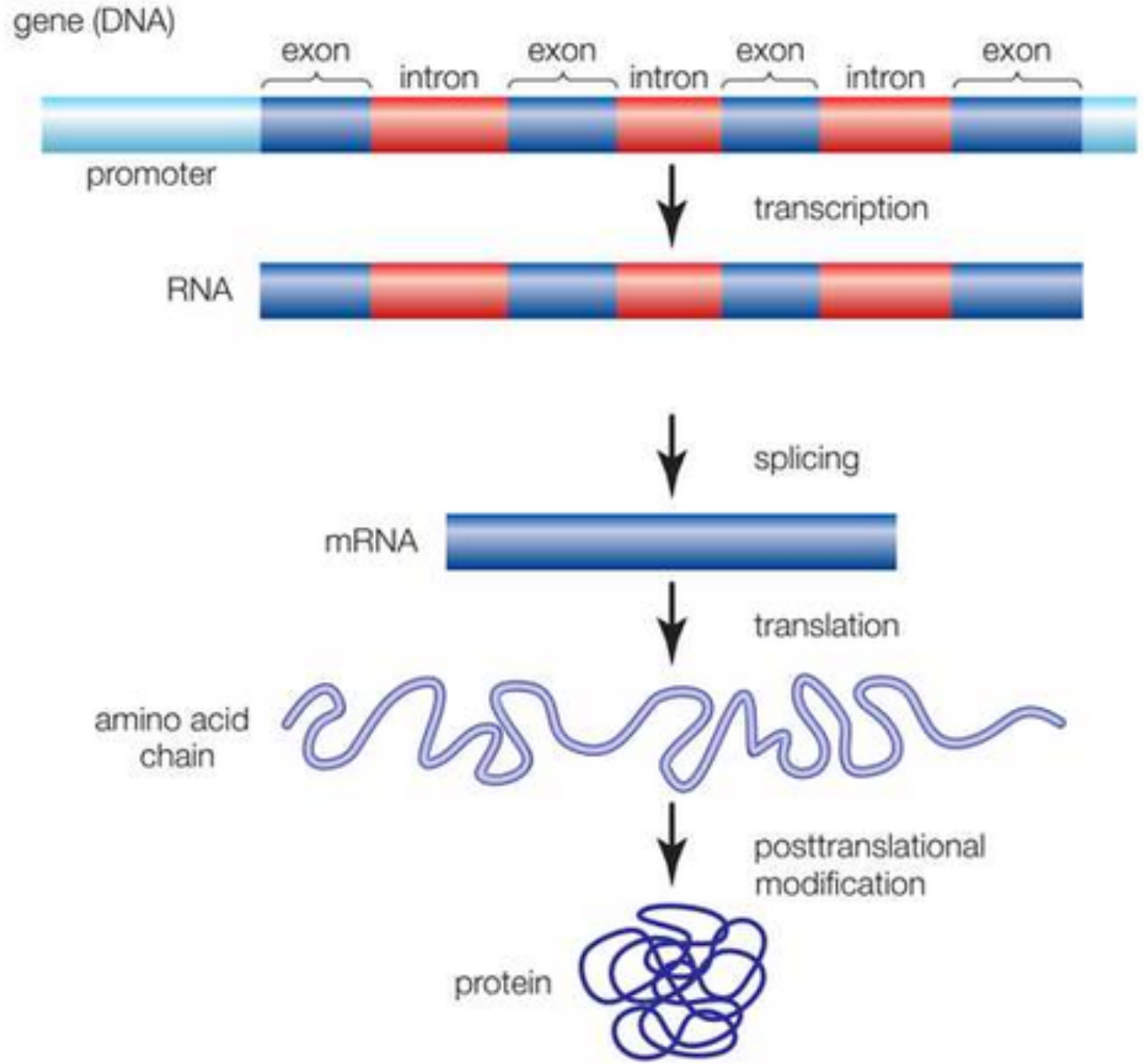


FROM DNA TO PROTEINS

The bases on a single strand of DNA act as a code. The letters form three letter codons, which code for amino acids - the building blocks of proteins.

An enzyme, RNA polymerase, transcribes DNA into mRNA (messenger ribonucleic acid). It splits apart the two strands that form the double helix, then reads a strand and copies the sequence of nucleotides. The only difference between the RNA and the original DNA is that in the place of thymine (T), another base with a similar structure is used: uracil (U).

In multicellular organisms, the mRNA carries genetic code out of the cell nucleus, to the cytoplasm. Here, protein synthesis takes place. 'Translation' is the process of turning the mRNA's 'code' into proteins. Molecules called ribosomes carry out this process, building up proteins from the amino acids coded for.


gene (DNA)
exon
intron
exon
intron
exon
intron
exon
promoter
transcription
RNA
splicing
mRNA
translation
amino acid chain
posttranslational modification
protein

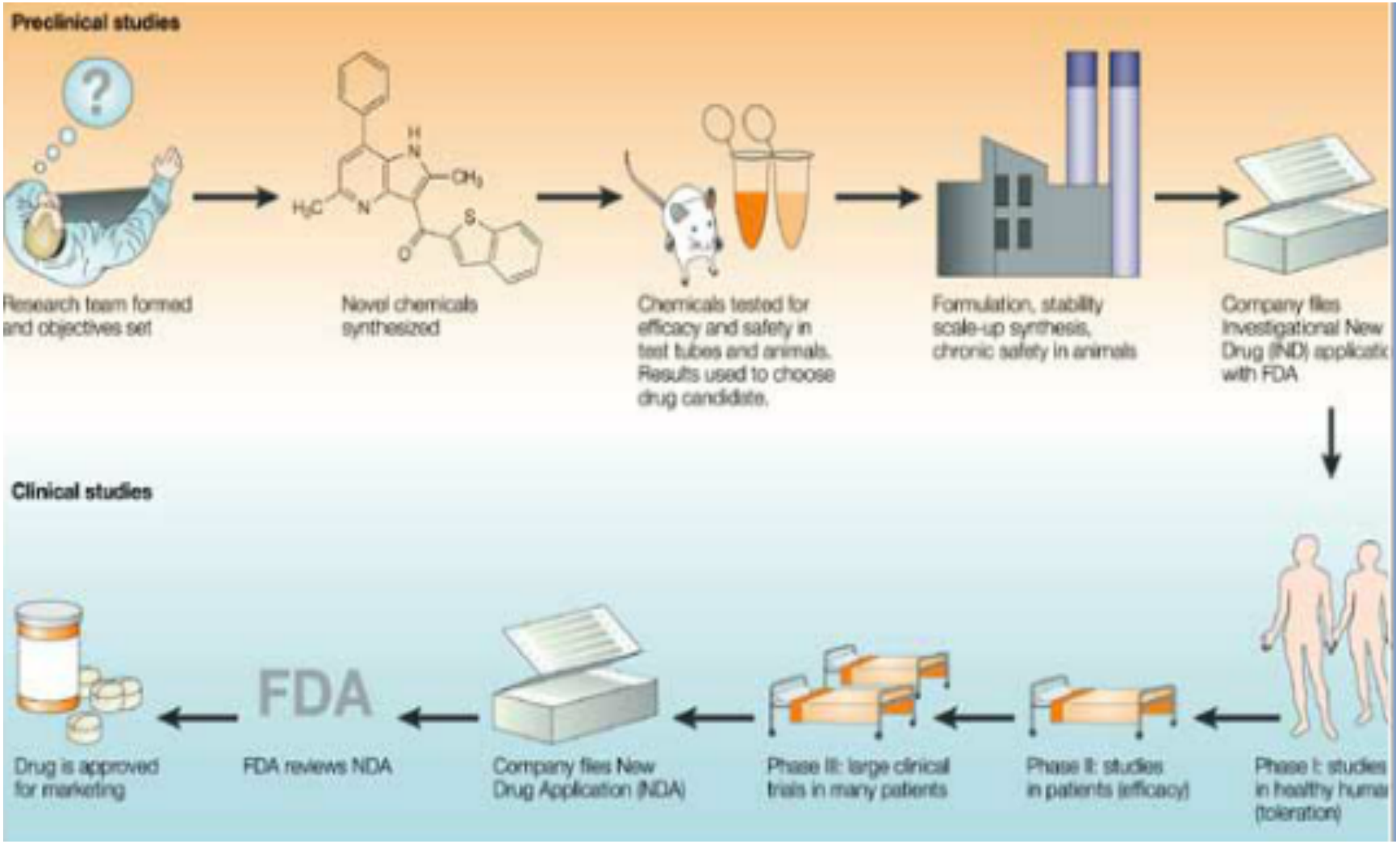
Applications of AI

11 COMMON APPLICATIONS OF ARTIFICIAL INTELLIGENCE IN HEALTHCARE

1. Managing Medical Records and Other Data
2. Analyzing tests, X-Rays, CT scans, data entry
3. Treatment Design
4. Digital Consultation
5. Virtual Nurses
6. Medication Management (Disease Identification/Diagnosis)
7. **Precision Medicine**
   (Personalized Treatment/Behavioral Modification)
8. Drug Design and Development
9. Health Monitoring
10. Healthcare System Analysis
11. **Clinical Trial Research**

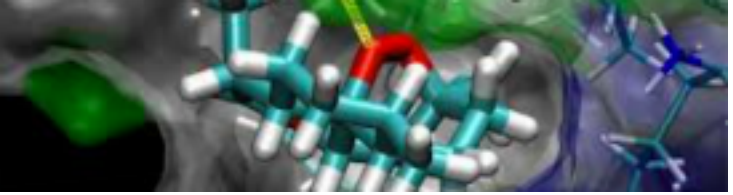
Drug Design and Development

Developing pharmaceuticals through clinical trials can take more than a decade and cost billions of dollars. Making this process faster and cheaper could change the world.



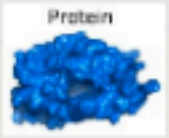
On an average, it costs USD 2.5 billion to bring a new drug to the market. For big pharmaceutical companies, this average is around USD 4 billion and has been shown to go as high as USD 11 billion…!

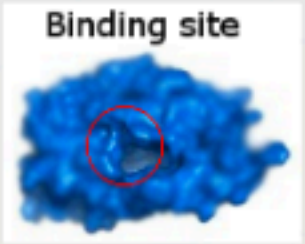
Designing a new drug that binds to any specific target requires a large amount of time, as well as computing power. In many ways, deep learning algorithms are being developed to accelerate this process. It is anticipated that digital solutions for drug discovery may save significant time and money...!

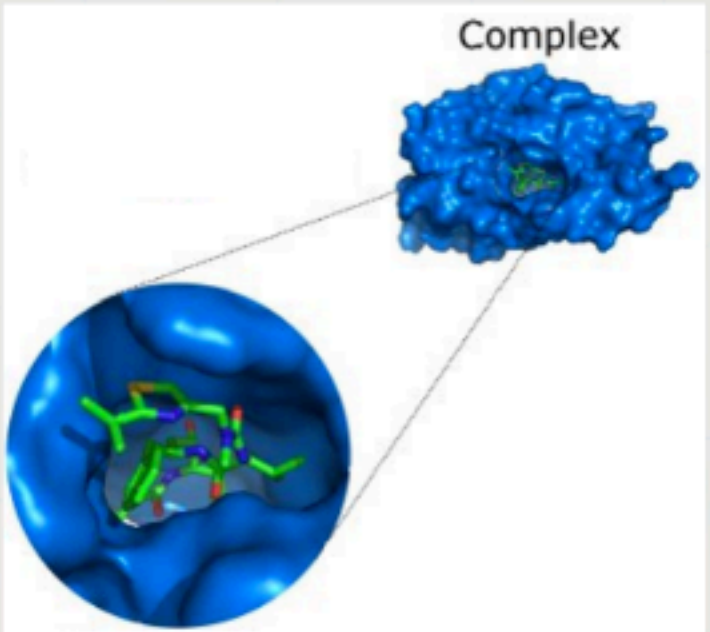


Drug Design and Development

Problem Statement: Protein-Ligand Interactions







- Propose a data model for predicting the binding affinity representing the strength of binding.
- Find patterns which define the interaction.
- Determine the type of relationship between the model and the binding site.

Cyberinfrastructure...

The problem is...

How a ligand recognize its binding site?

Will we be able to express protein-ligand interaction as a mathematical model?

OR

Facial Recognition

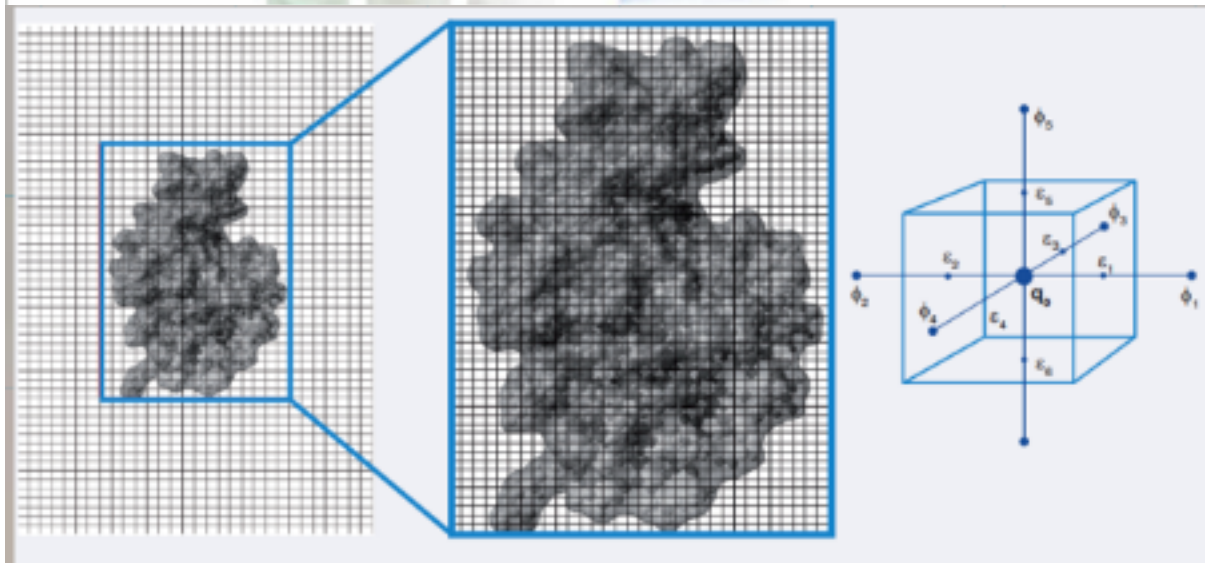
Fingerprint Recognition

Problem Statement: Protein-Ligand Interactions

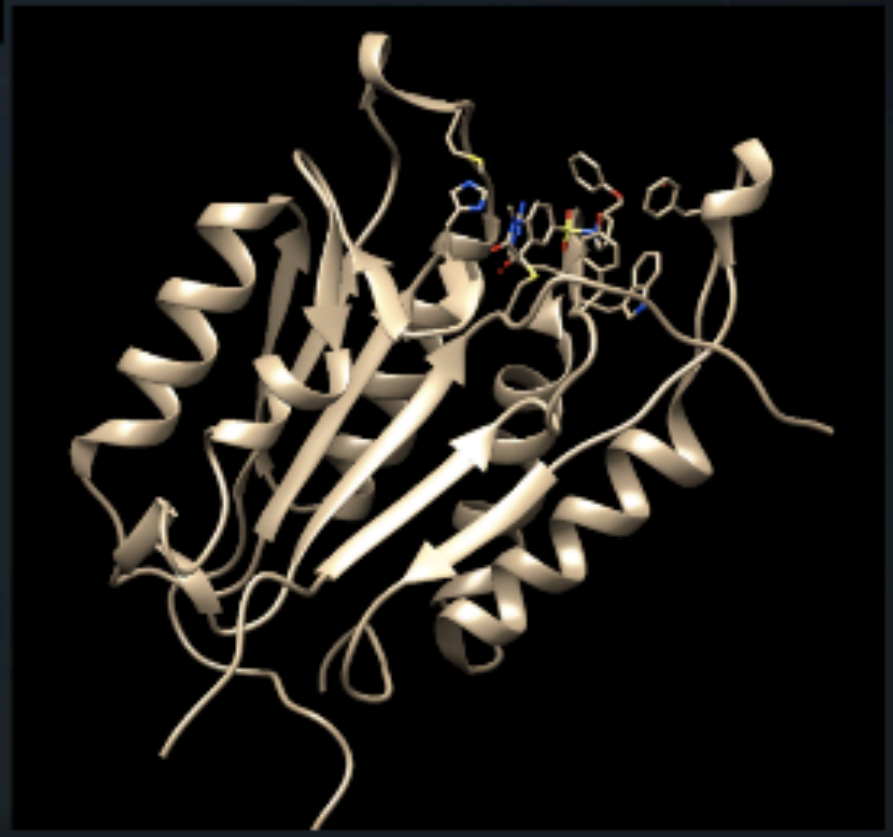
Facial Recognition System

A computer application for automatically identifying or verifying a person from a digital image. One of the ways to do this is by comparing selected facial features from the image and a facial database.

It is thought that electrostatic accommodation of diverse ligands by the receptor might be a prerequisite for the evolution of cognate recognition and may provide the mechanism for interference by noncognate molecules.



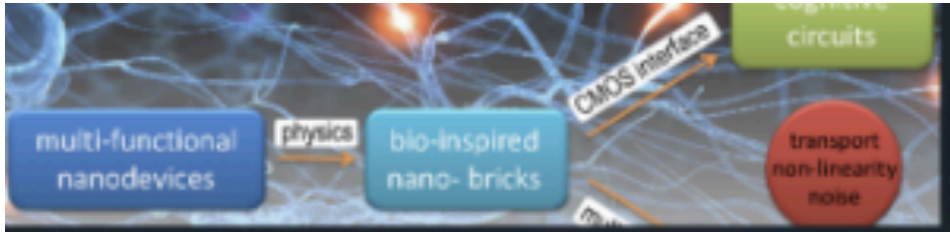
Problem Statement: Protein-Ligand Interactions



CASP3-Compound1
PDB ID: 1GFW

Caspases have been strongly implicated to play an essential role in apoptosis.

* The ligand fits to the target like "a hand in a glove".
* The strength of the protein-ligand interaction is measured by binding affinity.
* Measuring binding affinity is difficult.



DATA MODELLING METHODS

Ligand Preparation and Docking

Obtaining 3D Electrostatic Potential Grid Maps

Compressing 3D Cube into 2D Image

Feature Selection

Sequential Forward Selection

Algorithm 1 :Sequential Forward Selection
Input: P = - initial feature set
Q - the full set of features
J - criterion function to minimize
Output: P - final feature set
repeat
for all x Q do
set P P {x}
calculate J(P)
end for
set P P {x+} where x+ = argmin[J(P)]
set Q Q Ä {x+}
until no further improvement in J

Sequential Backward Elimination

Algorithm 2 :Sequential Backward Elimination
Input: P - the full set of feature set
J - criterion function to minimize
Output: P - final feature set
repeat
for all x P do
set P P Ä {x}
calculate J(P)
end for
set P P Ä {x−} where x− = argmax[J(P)]
until no further improvement in J

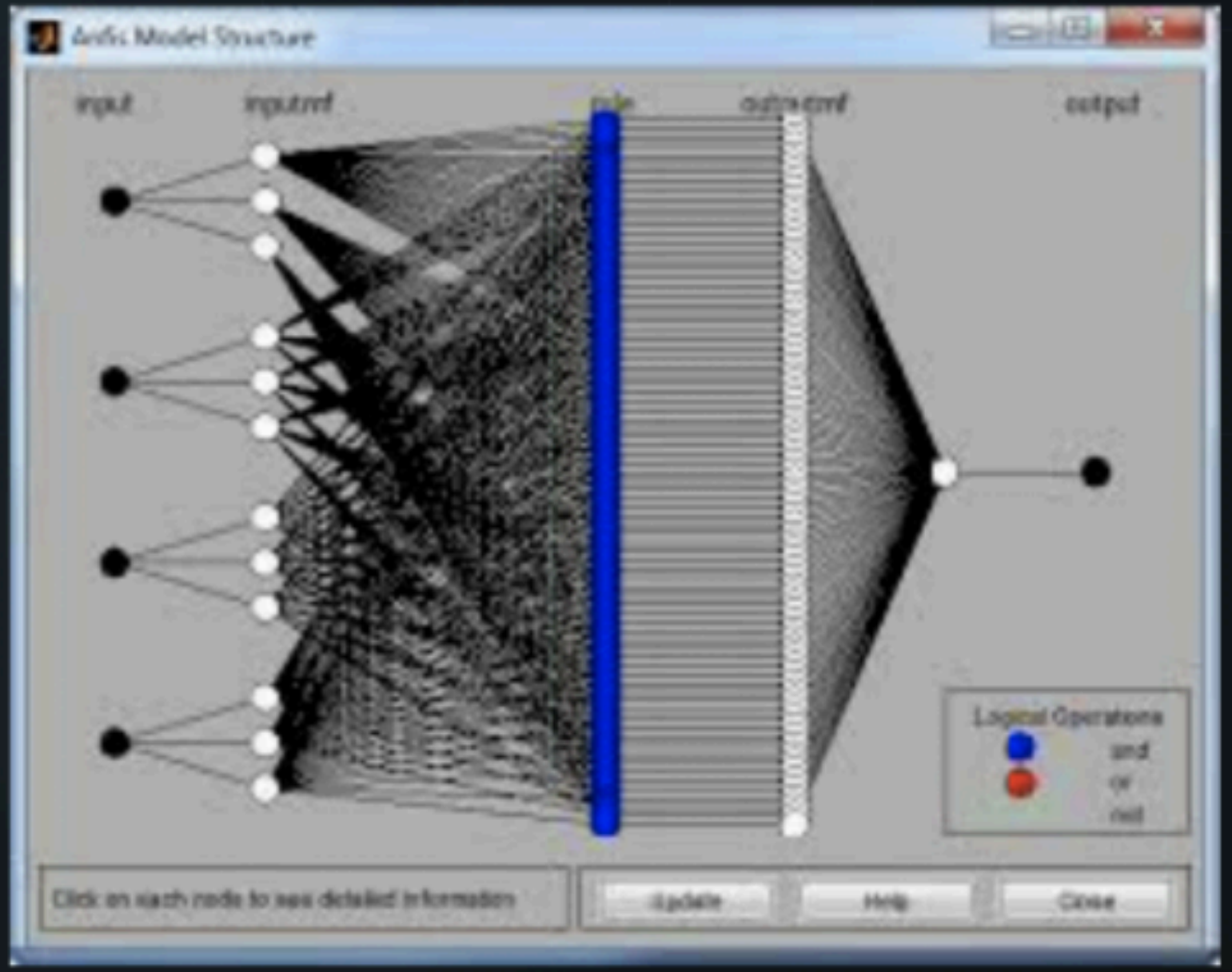
Sequential Forward Floating Selectio

Algorithm 3 :Sequential Forward Floating Selection
Input: P = - initial feature set
Q - the full set of features
J - criterion function to minimize
Output: P - final feature set
repeat
Step 1. Select the best feature x+ = argmin[J(P {x+})]
set P P {x+}
Step 2. Select the worst feature x− = argmax[J(P Ä {x−})]
if J(P Ä {x−}) < J(P) then
set P P Ä {x−}
go to Step 2
else
go to Step 1
end if
until no further improvement in J

Adaptive Neuro-Fuzzy Inference System

L1: input layer
L2: fuzzyfication layer,
computing the membership value.

L3: rule layer, each rule is a node using soft-min or product to
compute the rule matching factor ∏i
L4: normalization layer, each ∏i is scaled into Ni in the
normalization layer
L5: defuzzification layer, Each Ni weighs the result of its linear
regression fi = aix1 + bix2 in the function layer
L6: output layer, sum of outputs of L5.



3D ANALYSIS OF THE BINDING SITE IMAGES FOR PREDICTING BINDING AFFINITIES IN DRUG DESIGN

HARVARD
John A. Paulson
School of Engineering
and Applied Sciences

Site People

ABOUT SEAS | ACADEMICS | FACULTY & RESEARCH | NEWS & EVENTS | OFFICES & SERVICES | MAKE A GIFT

News & Events > Programming smart molecules

Programming smart molecules

Harvard machine-learning algorithms could make chemical reactions intelligent

December 12, 2013

Ryan P. Adams is an assistant professor of computer science at Harvard SEAS. (Photo by Eliza Grinnell, SEAS Communications)

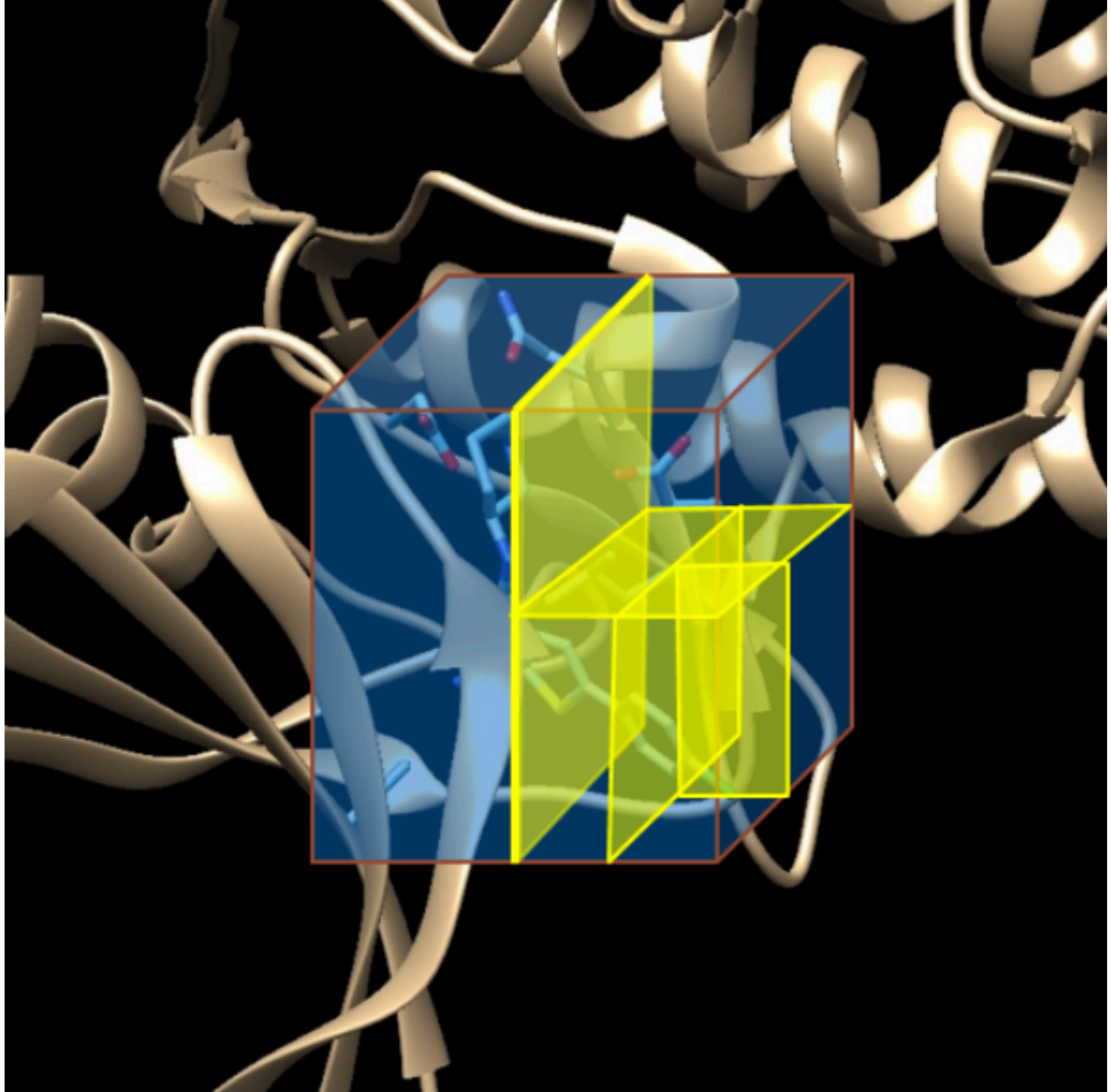
Cambridge, Mass. – December 12, 2013 – Computer scientists at the **Harvard School of Engineering and Applied Sciences** (SEAS) and the **Wyss Institute for Biologically Inspired Engineering at Harvard University** have joined forces to put powerful probabilistic reasoning algorithms in the hands of bioengineers.

PRESS CONTACT

Smart molecules...

If we agree the fact that the consciousness of the molecules acting on an relevant proteins and/or DNA/RNA zones,

Then we should be able to teach the molecules to find the specific target proteins to deal with the biological activities…!



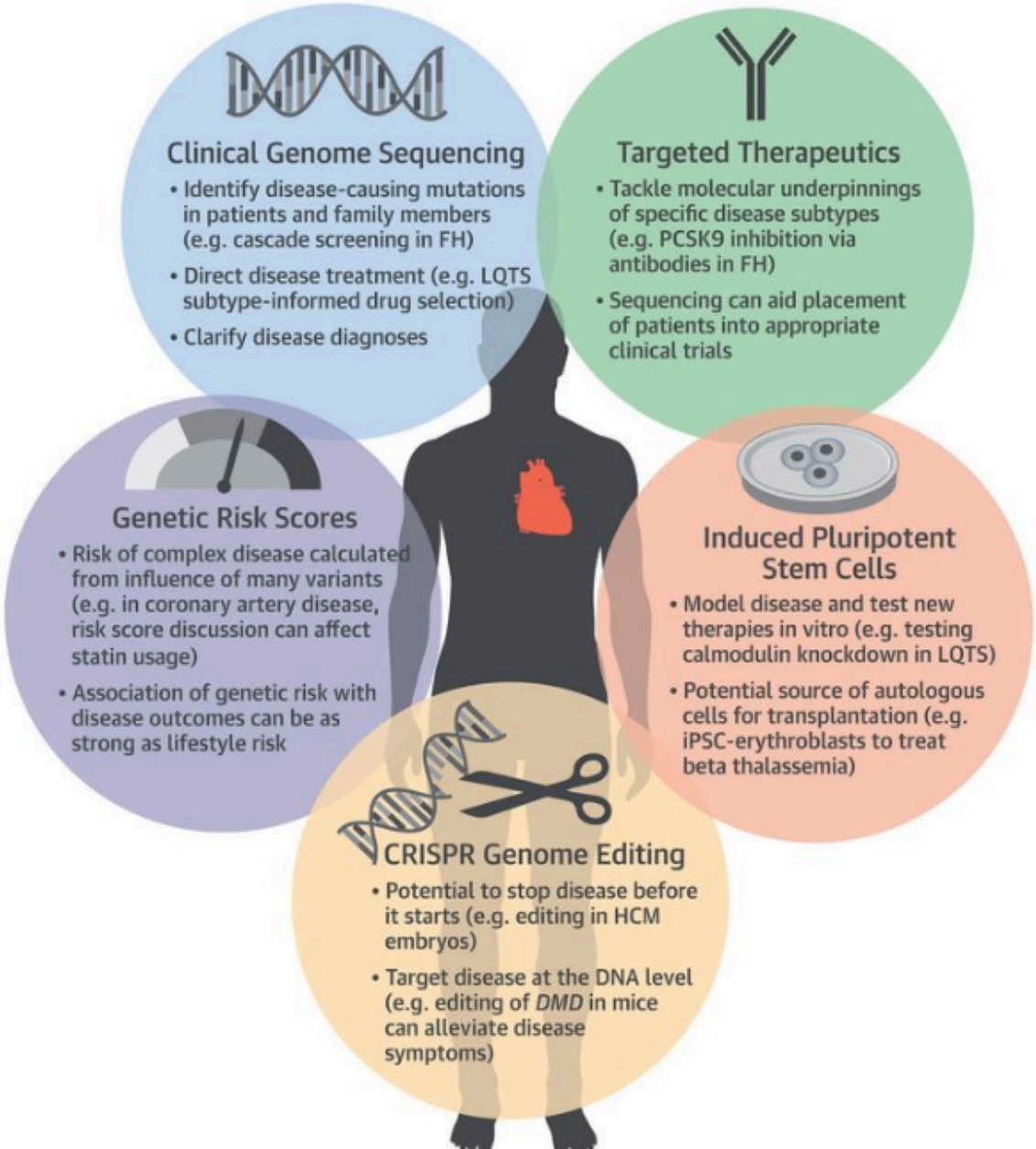
Individual sub-cube method (Exhaustive search)

How do we know how close we are to the solution?

At level 10: $3x10^{89}$

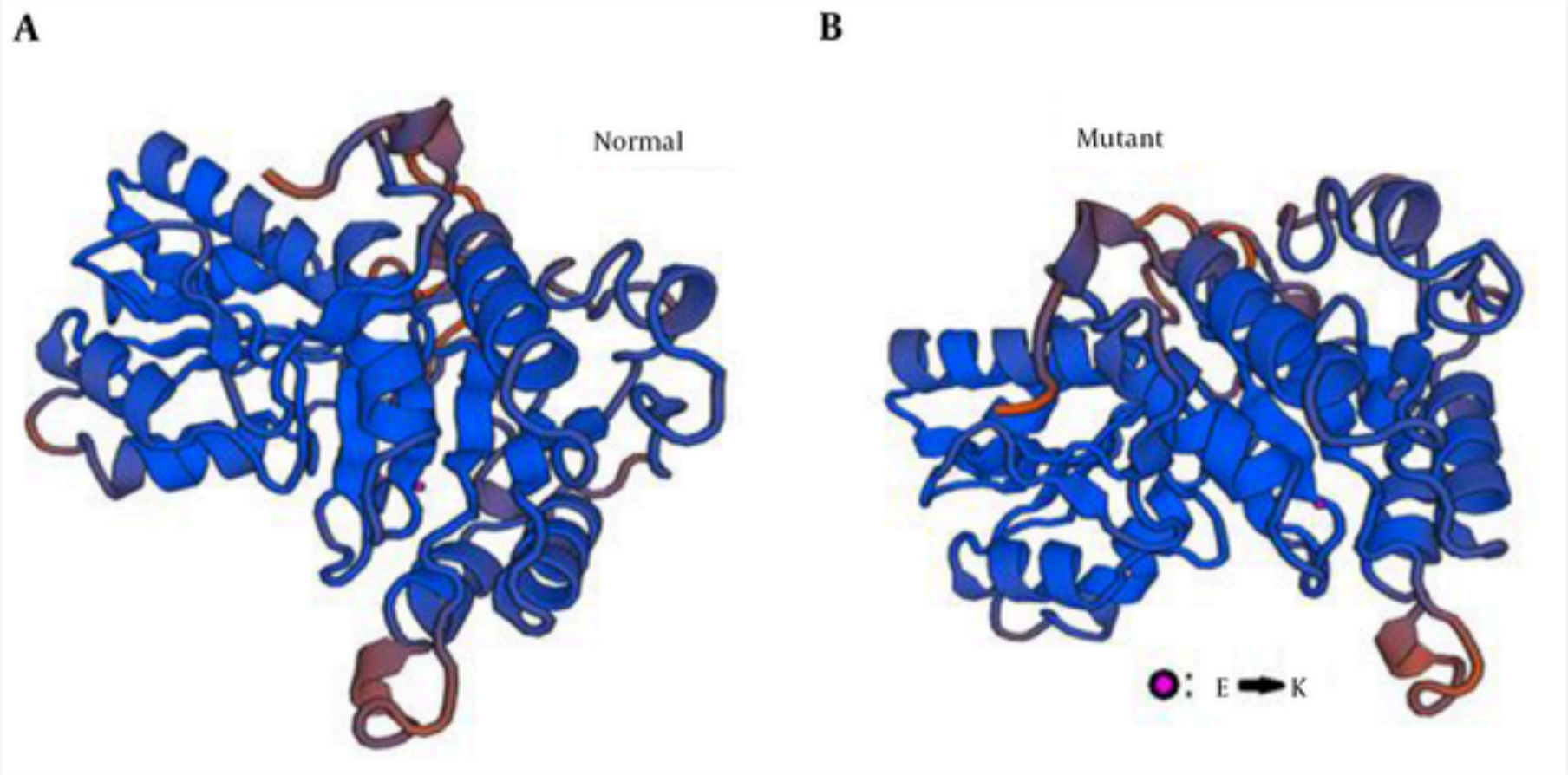
We need more deeper inside artificial intelligence methods…!

Genomics in personalized medicine



Genomics in personalized medicine

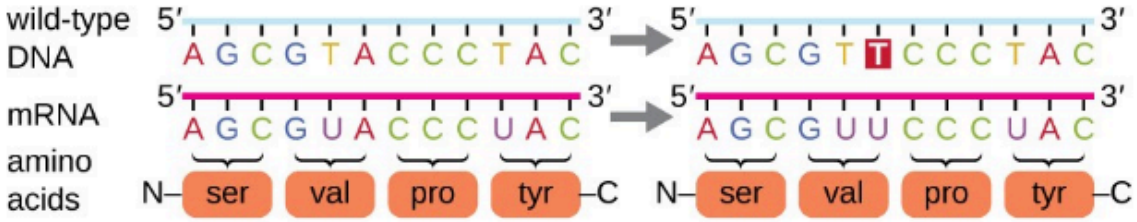
Effects of Mutations on Protein Structure and Function



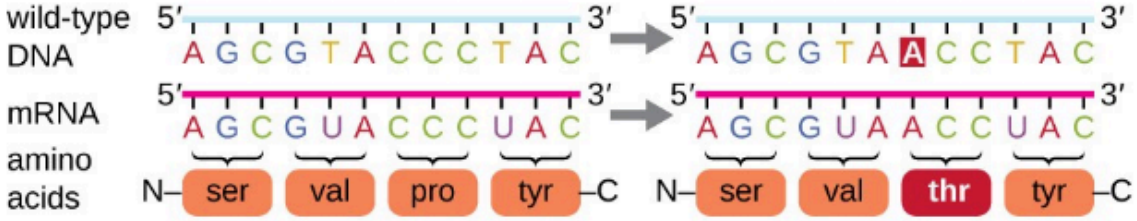
A, the predicted 3D structure of normal protein product of the BCKDHB gene (wild-type E1β protein); B, the E330K mutant E1β protein structure. (Pictures prepared by Swiss model software).

point mutation: substitution of a single base

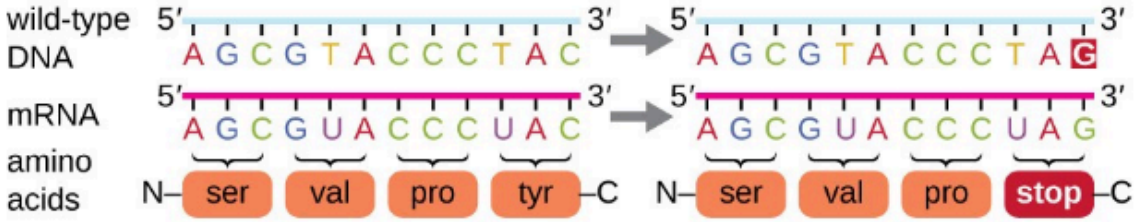
silent: has no effect on the protein sequence



missense: results in an amino acid substitution

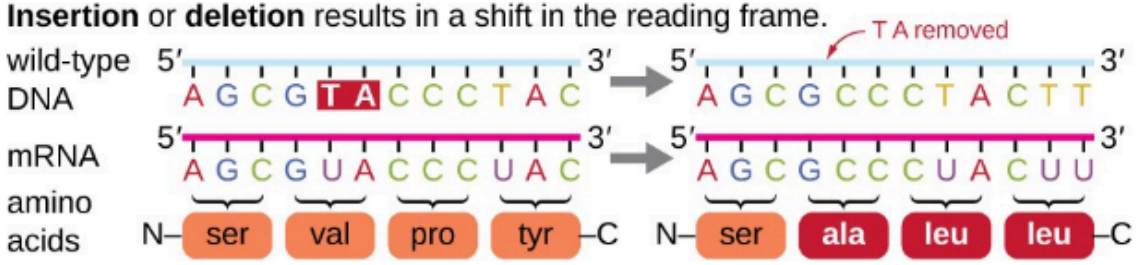


nonsense: substitutes a stop codon for an amino acid



frameshift mutation: insertion or deletion of one or more bases

Insertion or **deletion** results in a shift in the reading frame.



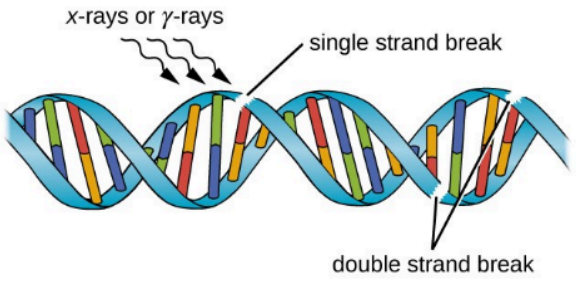
Causes of Mutations

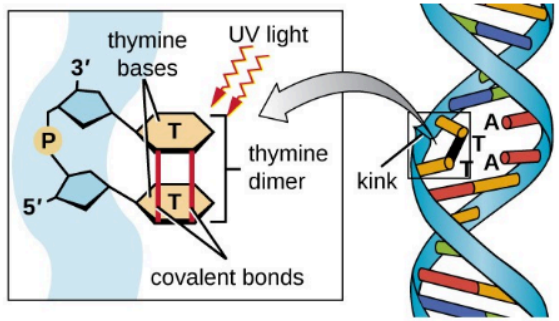
Mistakes in the process of DNA replication can cause **spontaneous mutations** to occur

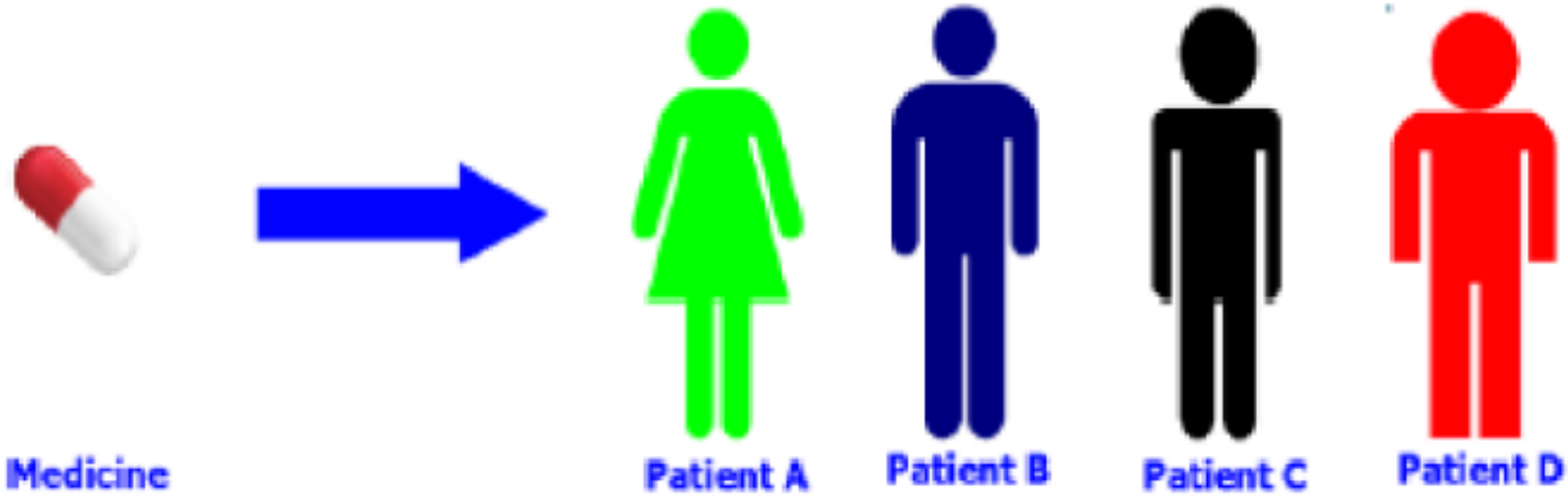
Various types of chemical mutagens interact directly with DNA either by acting as nucleoside analogs or by modifying nucleotide bases.

Strong **ionizing radiation** like **X-rays** and **gamma rays** can cause single- and double-stranded breaks in the DNA backbone through the formation of hydroxyl radicals on radiation exposure

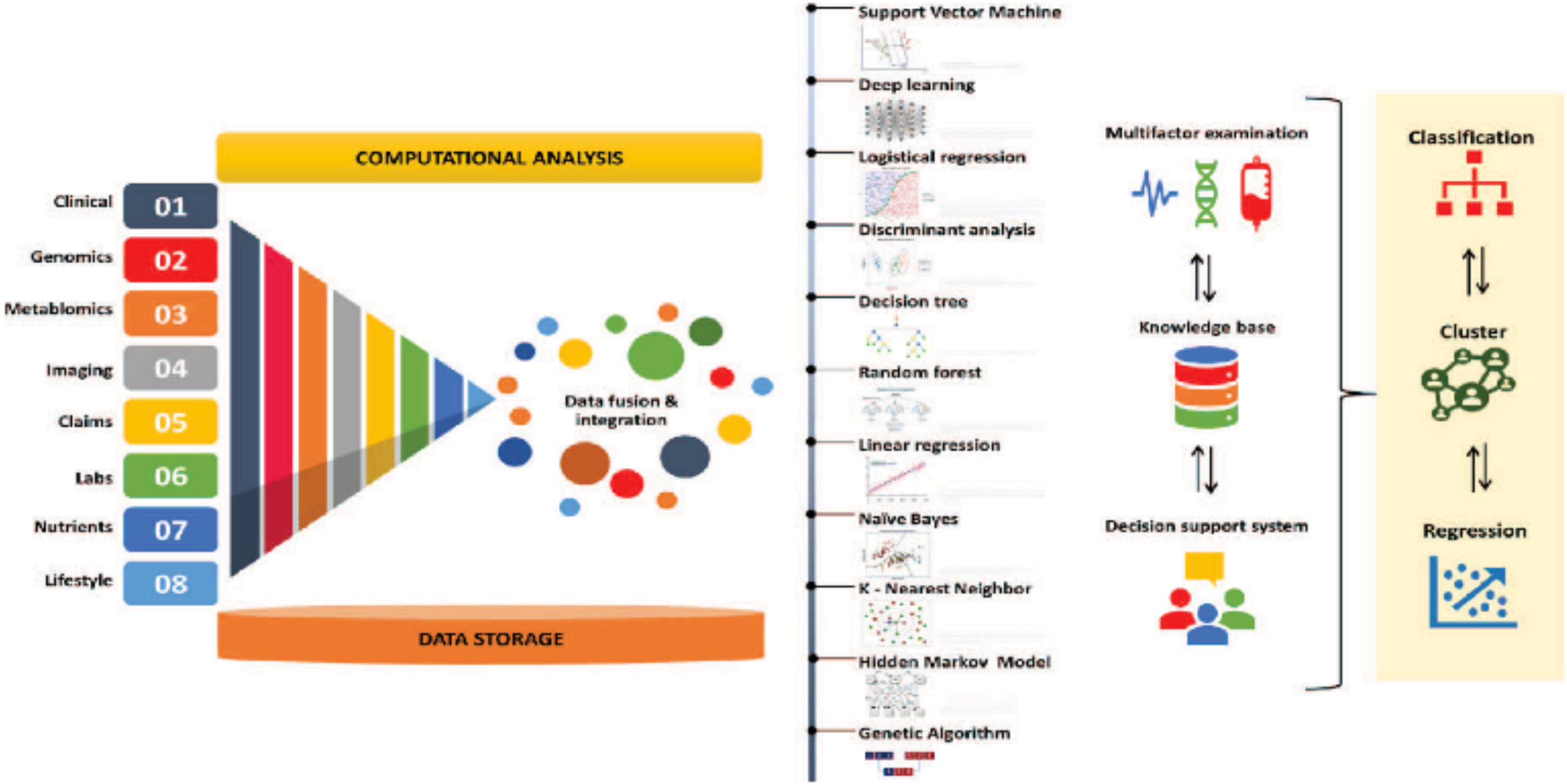
Nonionizing radiation, like ultraviolet light, is not energetic enough to initiate these types of chemical changes

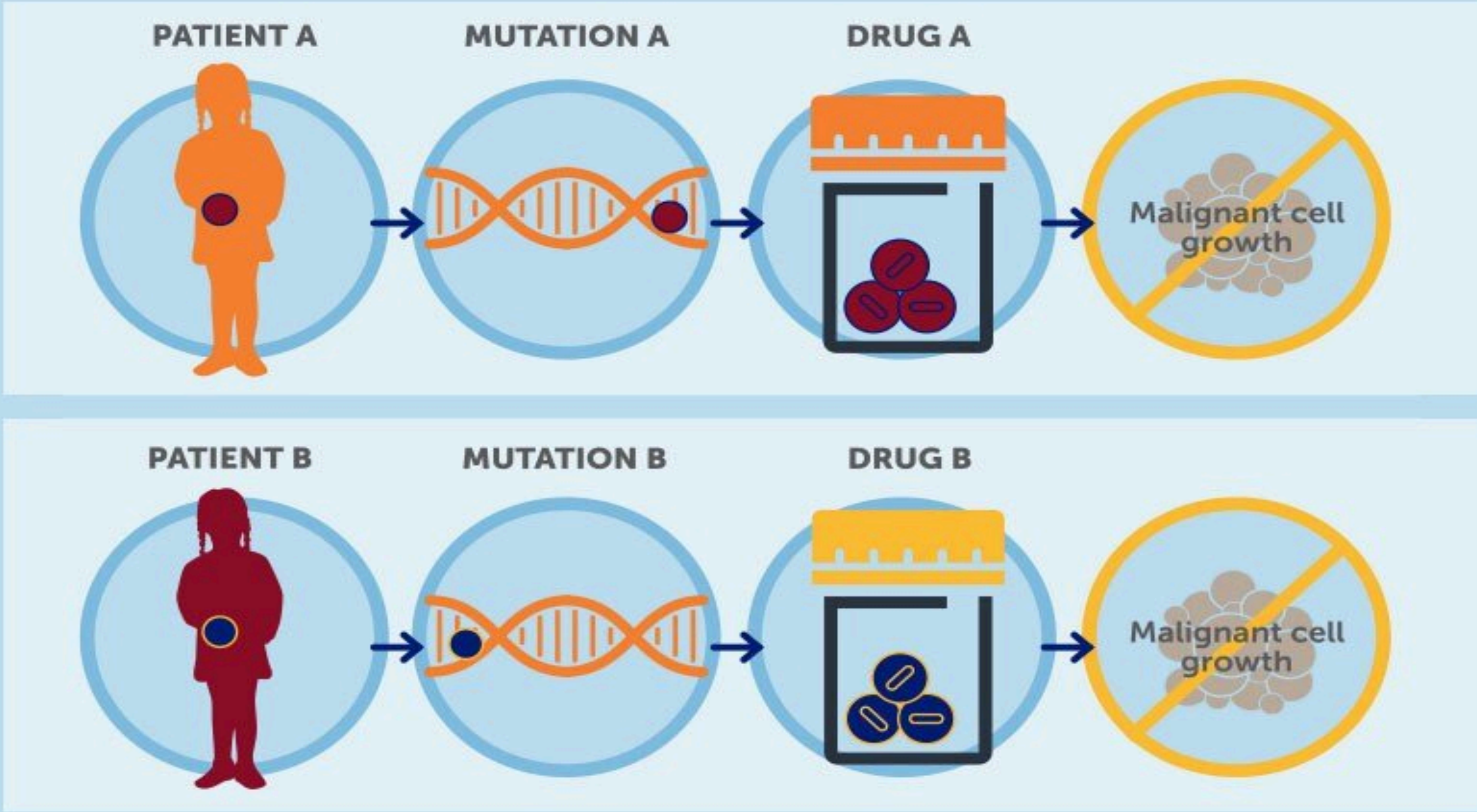






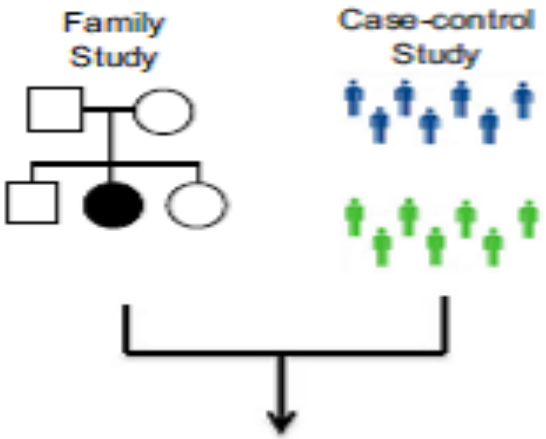
Traditional Medicine


COMPUTATIONAL ANALYSIS
Clinical
01
Genomics
02
Metablomics
03
Imaging
04
Claims
05
Labs
06
Nutrients
07
Lifestyle
08
Data fusion & integration
DATA STORAGE
Support Vector Machine
Deep learning
Logistical regression
Discriminant analysis
Decision tree
Random forest
Linear regression
Naïve Bayes
K - Nearest Neighbor
Hidden Markov Model
Genetic Algorithm
Multifactor examination
Knowledge base
Decision support system
Classification
Cluster
Regression


PATIENT A
MUTATION A
DRUG A
Malignant cell growth
PATIENT B
MUTATION B
DRUG B
Malignant cell growth

We have to learn from biodiversity...!

Identifying Causal Genes



Genome Sequencing

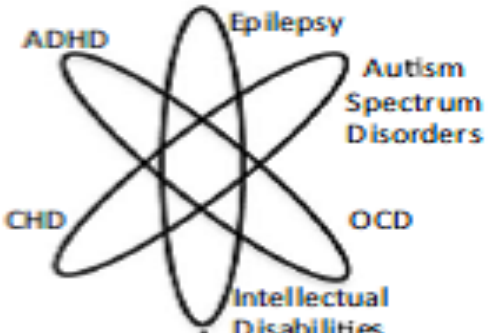
↓

Machine Intelligence Prediction or Classification Model

↓

Detection of Causal or Associated Genes

Phenotypic and Genetic Heterogeneity



↓

- Pleiotropy
- Incomplete penetrance
- Environmental factor
- Dosage Imbalance

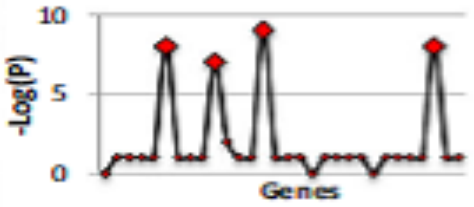
↓

Machine Intelligence Clustering or Classification Model

↓

Identifying Genotype-Phenotype Correlation

Polygenic Risk Score and Gene-Gene Interactions



↓

- Gene Expression
- Regulatory OMIC Databases
- Population Scale Mutational Databases

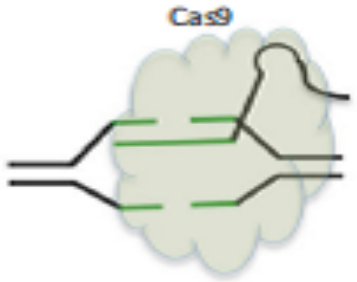
↓

Machine Intelligence Algorithm Inferring Polygenic Risk Score and Gene-Gene Interactions

↓

Quantifying Polygenic Risk Score and Gene-Gene Interactions

Drug Discovery



↓

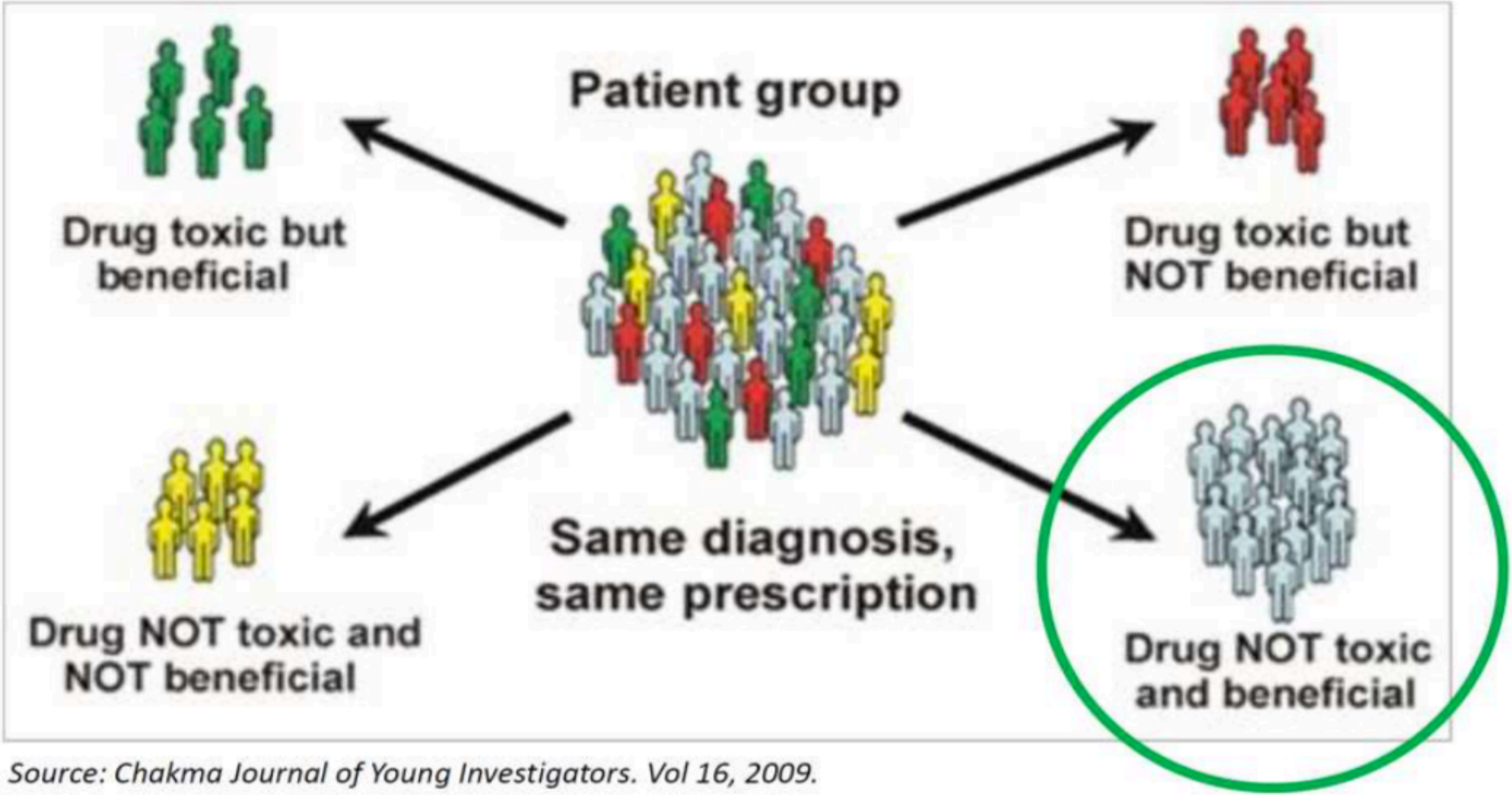
- *in vitro* (Cell Line, Brain Organoid) and *in vivo* (animal model) model to characterize genetic mutations

↓

Machine Intelligence Algorithm Inferring Biomarker

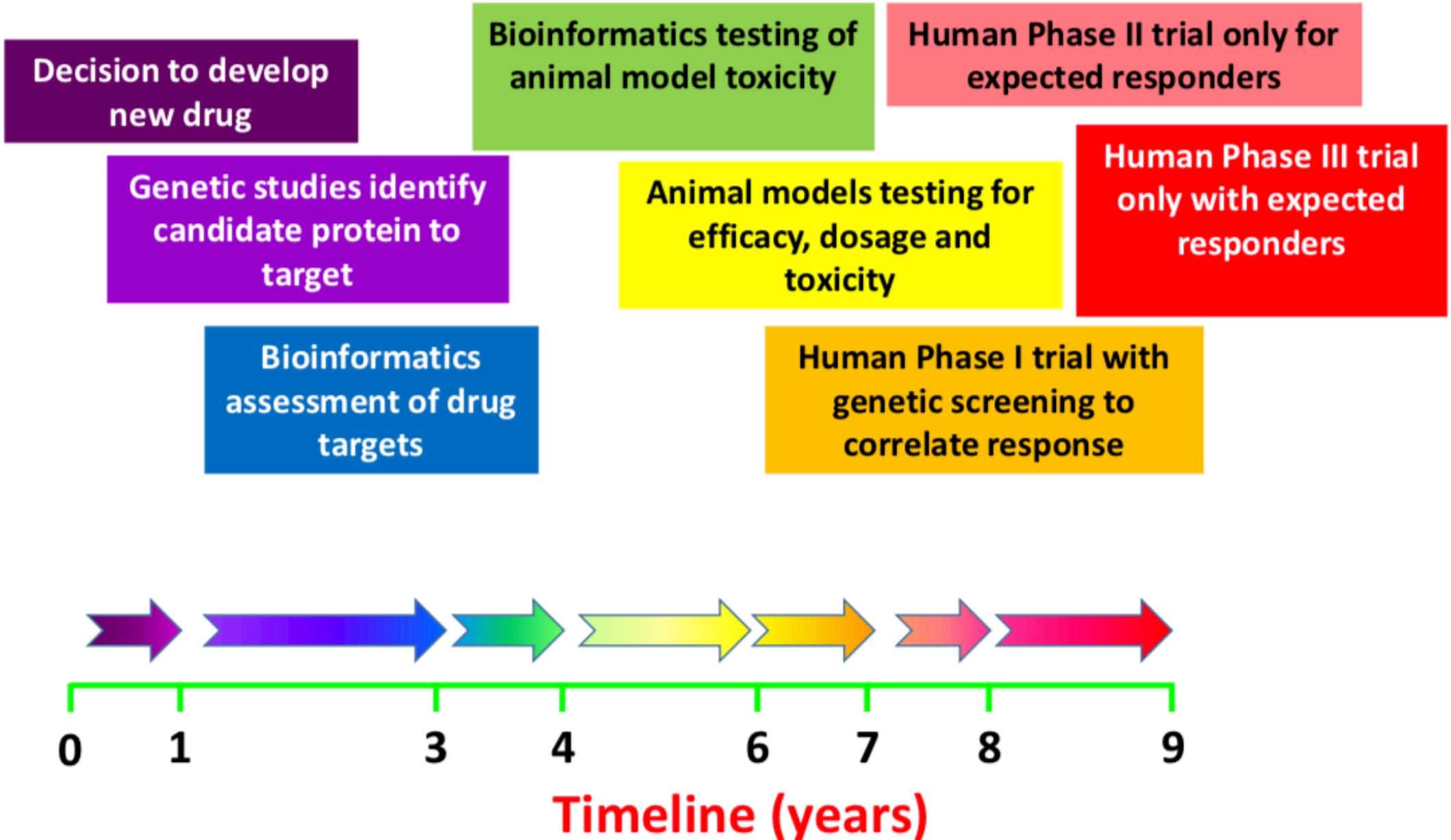
↓

Multi-stage Clinical Trials

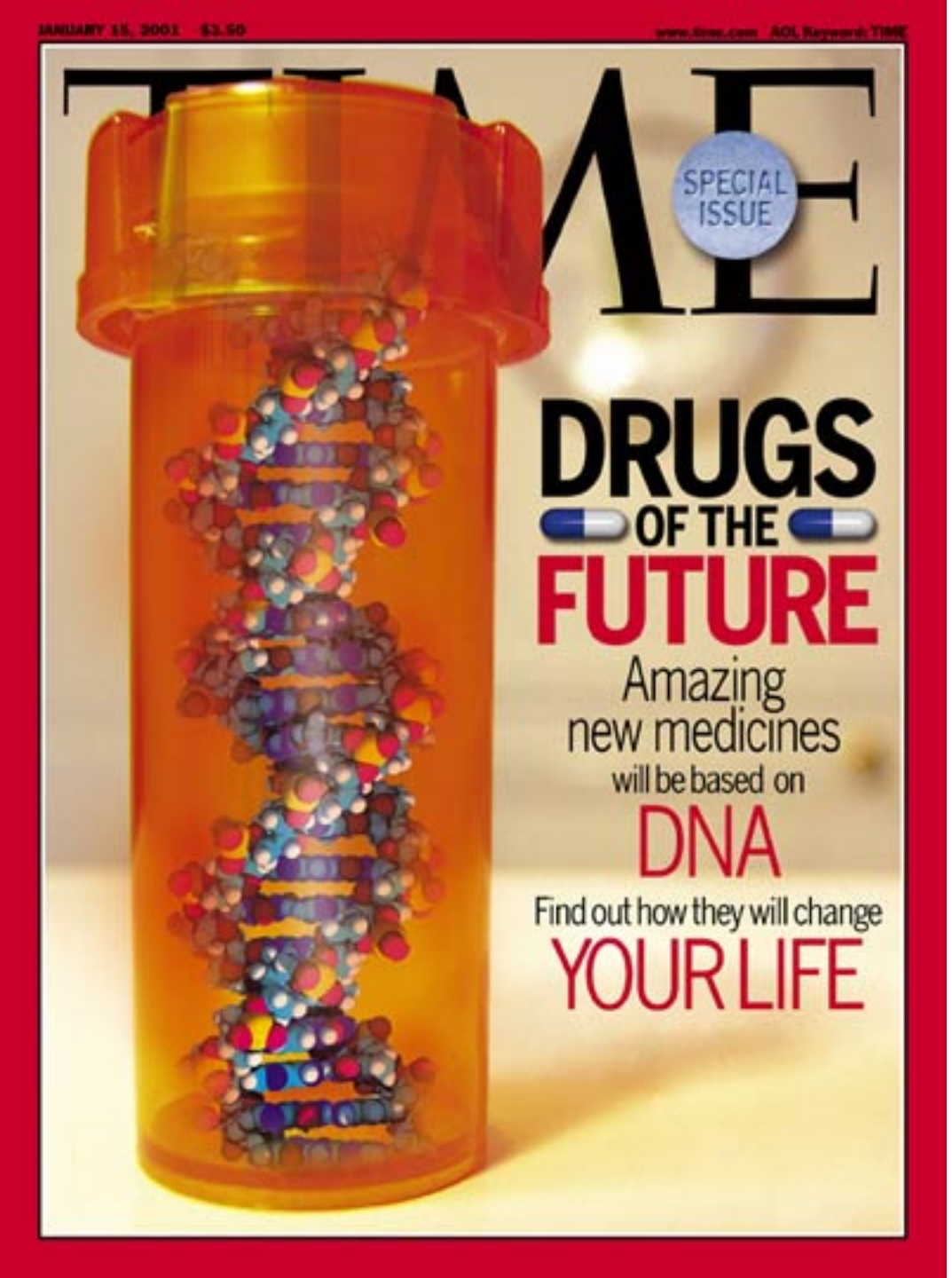


Source: Chakma Journal of Young Investigators. Vol 16, 2009.

Future (current) Drug Development



MinION: A complete DNA sequencer on a USB stick



Personalized Medicine Revolution

The Right Molecule

The Right Patient

For The Right Disease

At The Right Time

With The Right Dosage

Future Direction for Using Artificial Intelligence to Predict and Manage Gene Analysis and Editing

HOPE

or

Hype

We might not be there yet
But we are closer than we were yesteday...

Thank you...