

## *25 Jotun Hein MMSC Case Study Projects*

*I have earlier formulated large number of projects, that changed a lot when the students started, ie I decided to involved the students in the project formulation as well.*

*One can home in on an interesting project in 2-3 conversations or possibly even faster if the student(s) have already something they want to do. If so, I just have to be convinced, that I can supervise it responsibly.*

*There at least 3 steps in finding an ideal project:*

*1 – what are the technical expertise of the student(s): algorithms, data analysis, stochastic models, dynamical systems,...*

*2 – what is the domain of application?*

*3 – what is the idea/investigation that should be pursued?*

*On the next slides (mostly without voice), that I hope that you will browse, you should find:*

*Mini-Projects from the Course, Topics in Computational Biology [TiCB].*

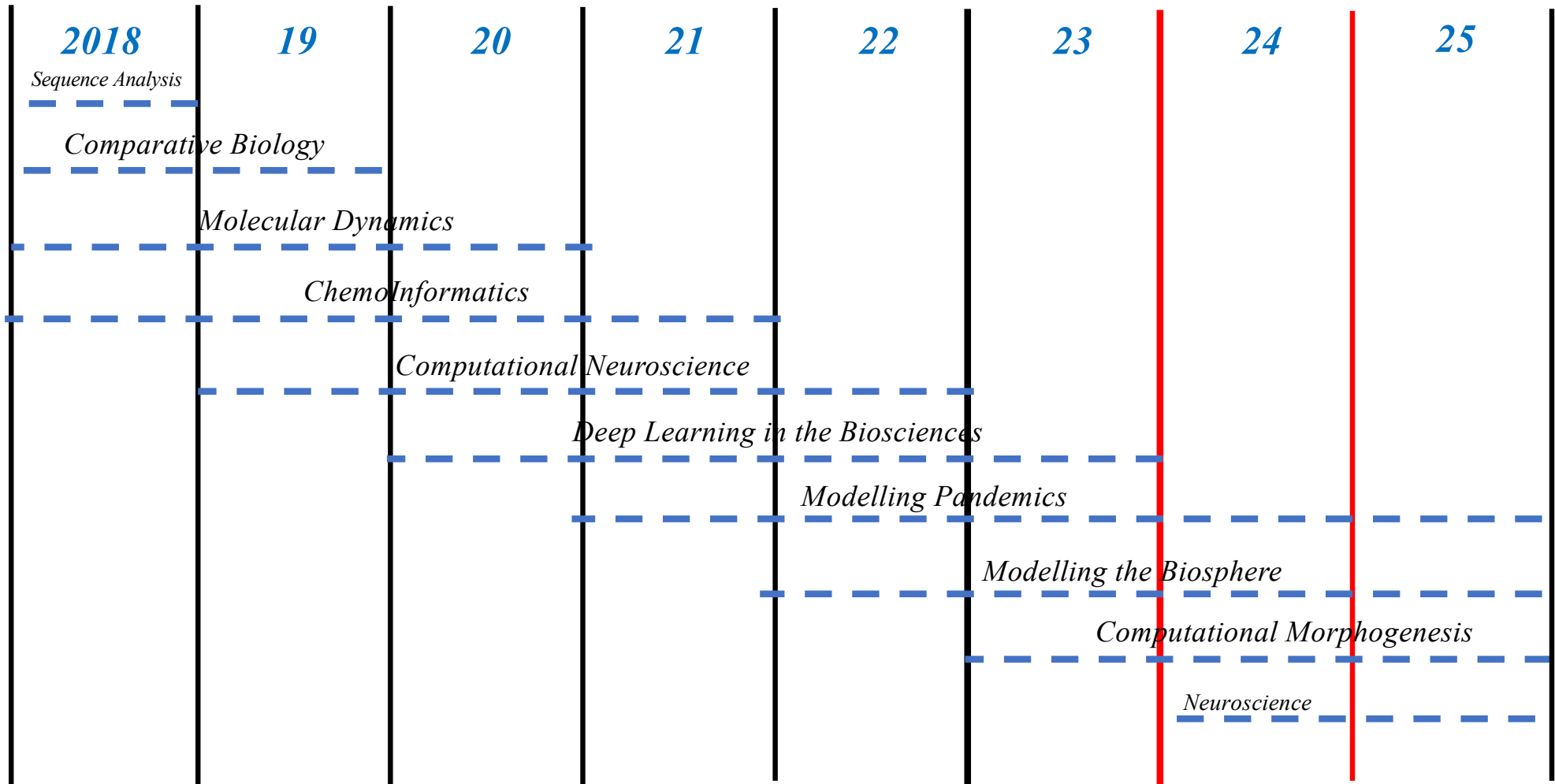
*Links to lectures mainly given in this course.*

*I hope this is interesting and defines what I could supervise in.*



# Topics in Computational Biology

*Each year a sub-topic is added [2 weeks] and a sub-topic is deleted*



*I spend 6-8 months reading and preparing slides. If I read 100 pages/day, this amounts to 20K pages. Sounds like much, but it is less than 10% of a real expert will have read.*

# *Some Earlier MiniProjects I*

Examination by Essay 2020.

1. [Stochastic Context Free Grammars in RNA Secondary Structure Prediction](#)
2. [Substitution Models with Rate Heterogeneity](#)
3. [Clustering Biological Networks](#)
4. [Quantum algorithms for Protein folding](#)
5. [Probability Theory of Networks](#)
6. [Inference of Gene Regulatory Networks using Differential Equations](#)
7. [Hidden Markov Models and their Applications to Bioinformatics](#)
8. [Computational Models for Gene prediction](#)
9. [Alignment-based De Novo Prediction of Non-coding RNA](#)
10. [Automated Annotation of Genes](#)
11. [Network flow and its applications to the analysis](#)
12. [Identification of Regulatory Elements – Phylogenetic Footprinting](#)
13. [Probable and Improbable Paths in Sequence Evolution](#)
14. [Simulative Approach to Alignment](#)
15. [Sequence RAFs](#)
16. [Birth and Death Processes](#)
17. [Haplotype Imputation](#)
18. [Prediction of drug targets](#)
19. [Computational Approaches to Nanopore Sequencing](#)
20. [Molecular Dynamics for Modelling Dimers](#)
21. [Persistent Homology](#)
22. [One Algorithm Fits All](#)
23. [Spiking Neural Networks](#)
24. [Statistical Approach against COVID-19 Pandemic](#)
25. [Multi-objective optimisation techniques applications to gene regulatory network inference](#)
26. [Deep Learning Methods for Modelling Quantitative Structure-Activity Relationships](#)
27. [Deep Learning and Gene expression](#)
28. [Protein Tertiary Structure Prediction](#)
29. [Trial and Error ThePowerof EvolutionandReinforcement Learning](#)
30. [Constructing Phylogenies TheBayesianWay](#)
31. [UnpackingAlphaFold A ReviewofAdvancesinProteinFolding](#)
32. [Deep LearningMethodsforDNA MethylationinCancerStudy](#)
33. [Computational Approaches to Molecular Representations Shallow and Deep Learning Methods](#)
34. [Statistical FineMapping in Genome Wide Association Studies](#)
35. [CRISPR-Cas9GeneEditingMechanism](#)
36. [Genetic AlgorithmsInDrugDesign](#)
37. [Dopamine and Reinforcement Learning](#)
38. [BayesianNeuralNetworksforManagingUncertainty in BiomedicalImaging](#)
39. [The Role of Deep Neural Networks in Neuroscience](#)
40. [Computational Approaches for RNA Secondary Structure Prediction From Nussinov's Algorithm to E2Efold](#)



# *Some Earlier MiniProjects II: 2021*

1. [Predictive Computational Modelling for Epidemiology](#)
2. [An Overview of Several Time Series Methods for COVID-19 Forecasting](#)
3. [Causal Inference Methods for Epidemiology of the COVID-19 Pandemic](#)
4. [Modelling the effects of non-pharmaceutical interventions on COVID-19 transmission](#)
5. [Computational Models for COVID-19 Prediction](#)
6. [Mathematical Models of Infectious Disease Transmission](#)
7. [LSTM and GRU for COVID-19 prediction and their limitations](#)
8. [Modelling spread of COVID-19 by incorporating infectious period as a time-delay](#)
9. [Statistical Approach against COVID-19 Pandemic](#)
10. [Deep Learning Methods for Non-Invasive Diagnosis of Neurodegenerative Diseases](#)
11. [ANALYSIS OF TRANSPARENCY AND EXPLAINABILITY IN APPLICATIONS OF MACHINE LEARNING TECHNIQUES IN CANCER DIAGNOSIS](#)
12. [Generative Adversarial Networks in De Novo Drug Design](#)
13. [NEUROSCIENCE AND REINFORCEMENT LEARNING- THE PAST, PRESENT AND FUTURE](#)
14. [Predictive Coding- Perspectives from Neuroscience and Artificial Intelligence](#)
15. [Deep Learning models for the prediction of gene expression ab initio from DNA sequence](#)
16. [BRAIN-COMPUTER INTERFACES AND THE USE OF DEEP LEARNING MODELS](#)
17. [Graph Representational Learning in Biomedicine](#)
18. [Computational Biology Essay- The Language of Life- Advances in Protein Language Modelling](#)
19. [MACHINE LEARNING AND LIQUID BIOPSY FOR EARLY DIAGNOSIS OF CANCER](#)
20. [Alzheimer's Disease & Machine Learning](#)
21. [Modelling functional connectivity networks](#)
22. [Unobserved Confounding Problem in Treatment Effect Estimation from Clinical Data](#)
23. [GRAPH NEURAL NETWORKS FOR LEARNING MOLECULAR REPRESENTATIONS IN DRUG DISCOVERY](#)
24. [Persistent Homology in Computational Neuroscience](#)
25. [Identity By Descent- bridging population genetics, genomics and evolutionary theory](#)
26. [Phylogenetic Insights into CRISPR-Cas](#)
27. [Convolutional Neural Networks Are the livelihoods of radiologists threatened by the improving segmentation capabilities of convolutional neural networks?](#)
28. [Machine Learning and Deep Learning in Binding Affinity Prediction](#)
29. [Protein contact prediction based on multiple sequence alignments](#)
30. [Predicting Aberrant Splicing with Deep Learning](#)
31. [Deep Learning for Protein Structure Prediction](#)
32. [Applying deep learning to coarse grained models of molecular dynamics](#)
33. [Generative Models of Mental Disorders](#)
34. [Applying Markov decision problem and optimal control theory methods to the stag hunt game to phenotype Autism Spectrum Condition](#)
35. [Markov State Models and Protein Folding](#)
36. [From Cognitive Maps to Basis Sets- Models of Cognition in Spatial and Relational Domains](#)

# *Some Earlier MiniProjects III: 2022*

1. [Evaluating Social Distancing Interventions in the Context of Covid-19 Using Social Network Approaches](#)
2. [The development of spatial models for HIV transmission between heterosexual populations](#)
3. [A Computational Overview of Oceanic Dimethylsulfide \(DMS\) and Its Impact on Climate Modelling](#)
4. [Genotype-Phenotype Mapping and Alzheimer's Disease](#)
5. [Assembly Calculus](#)
6. [Deep Learning in COVID-19 Diagnosis and Assessment](#)
7. [Applications of Neural Network based models in the COVID-19 pandemic](#)
8. [A Review on Protein Contact Prediction Methods and the Potential Future Topics in This Field](#)
9. [Neuroevolution of plastic spiking networks: replaying nature's creative process](#)
10. [A model for Alzheimer's disease based on amyloid cascade hypothesis](#)
11. [An assessment of Computational Models and Techniques applied to ADHD](#)
12. [Computational Models of Human Vision](#)
13. [Convolutional Neural Networks in Medical Diagnostics](#)
14. [Genomic Motif Representations in CNN First Layer Filters with Varying CNN Architectures](#)
15. [Comparing 2 Different Methods of Recombination Detection of SARS-CoV-2](#)
16. [AlphaFold, deep learning and protein structure prediction](#)
17. [Computational Neuroscience and Modern Deep Learning: It Takes Two to Tango](#)
18. [Application of Deep Learning in Diagnosis of Breast Cancer](#)
19. [Network Science in COVID-19 Modeling](#)
20. [Computational Network Biology: Predicting Parkinson's Disease Genes with Network Embedding](#)
21. [Application and comparison of compartmental and agent-based models in pandemics](#)
22. [Modeling and predicting climate extremes](#)
23. [Deep Generative Modelling for One-Shot Molecule Generation](#)
24. [Brain-Computer Interfaces and Neural Networks in BCIs](#)
25. [Assessing the effects of non-pharmaceutical interventions in the initial phase of the COVID-19 pandemic with an SIR type model](#)
26. [Graph Neural Network for Antibody Sequence-Structure Co-design](#)
27. [A Bayesian Approach to Statistical Genomics Applied to SARS-CoV-2](#)
28. [Biologically Plausible Neural Networks](#)
29. [Deep Learning Techniques for Drug-Target Binding Affinity Prediction](#)
30. [Deep learning generative methods and their impact on drug discovery](#)
31. [Biological Computing through On-Chip Neuronal Networks](#)
32. [Deep Learning in Anti-aging Drug Discovery](#)
33. [Is the brain a quantum computer?](#)
34. [From AlphaFold1 to AlphaFold2: Development of Deep Learning Methods in Protein Structure Prediction](#)
35. [Bayesian Inference for Infectious Disease Models](#)
36. [Graph Neural Networks for De Novo Drug Design](#)
37. [Modelling Aerosol-Cloud Interactions](#)
38. [Quantum Algorithms for Protein Folding](#)
39. [Self-supervised Protein Language modelling to generate de novo SARS-CoV-2 Mutational Landscape](#)
40. [Graph Representational Learning in Computational Neuroscience](#)

# *Mini-Projects IV: 2023*

1. [Study of Surrogate Model Approaches for Explaining the Importance of Blood Sample Features in the Prediction of COVID-19 Mortality](#)
2. [HIGH-LEVEL AGENT-BASED MODELLING OF THE COVID-19 PANDEMIC](#)
3. [Generalization of Prediction Models for COVID-19](#)
4. [Quantifying the Effects of Government Interventions on COVID-19 Spread](#)
5. [A Comparison of Two Mechanistic Models](#)
6. [An Overview of Mobility Network Models in Predicting Epidemics](#)
7. [Deterministic vs Stochastic Compartmental Models](#)
8. [Forecasting of the COVID-19 Pandemic](#)
9. [Evolution of disease modelling with SIR type models](#)
10. [Review - Modelling the transmission of Onchocerciasis](#)
11. [Infectious Disease Model and Model Fitting Techniques about A \(H1N1\) Influenza](#)
12. [Forecasting of the COVID-19 Pandemic - A Study of SIR, SEIR, and ARIMA Models](#)
13. [Forecasting of the COVID-19 Pandemic - Compartmental and Time-series models](#)
14. [\*\*Modelling HIV/AIDS pandemic to assess accuracy of treatment methods\*\*](#)
15. [An Overview of several compartmental epidemic models](#)
16. [Climate Modelling - A Brief History and the Evaluation of Uncertainty](#)
17. [A report on the Meta-Statistical Extreme Value Distribution as an alternative to the Generalised Extreme Value Distribution for modelling yearly rainfall maximums](#)
18. [A Comparison of Reinforcement Learning Models for Chemical Plume Tracing](#)
19. [The Evolution of El Niño Forecasting](#)
20. [A Review of the Carbon Cycle, and the Major Carbon Sinks](#)
21. [Representing Unresolved Processes into Climate Models through Machine-Learning-Based Parameterizations](#)
22. [MORTALITY ASSOCIATED WITH TEMPERATURE - CASE STUDY OF CHICAGO HEAT WAVE](#)
23. [Turing patterns in predator-prey modelling](#)
24. [Turing Patterns and Growth in Biology](#)
25. [Topological Data Analysis in Computational Biology](#)
26. [Multiscale modelling of environmental stimuli in plant tropism](#)
27. [A Monte Carlo Simulation of Protein Crystal Twinning](#)
28. [Hyperbolic Phylogenetics](#)
29. [The Emergence of Large Language Models in Healthcare](#)
30. [\*\*Deep Neural Networks for the Next Pandemic Diagnosis\*\*](#)
31. [Multi-task Learning with Deep Neural Networks for Discrete-time Survival Analysis](#)
32. [Deep learning in CRISPR gene editing via effective off-target predictions](#)
33. [Applications of Deep Learning and Genomics in Cancer Diagnosis](#)
34. [Deep Learning in Ovarian Cancer Diagnosis](#)
35. [Benchmarking Graph Neural Networks for Protein-Protein Interaction](#)
36. [Developments in Protein Structure Predictions](#)
37. [Convolutional Neural Networks in Drug Discovery](#)
38. [Metabolic Engineering and Machine Intelligence - a match made in heaven.](#)
39. [Diffusion Models on Generating Fluorescent Channel of the Cell Image](#)
40. [Towards the Design of Next-Generation Medical Chatbots - A Multimodal Deep Learning Approach](#)
41. [Geometric deep learning on non-Euclidean data for modelling cardiac anatomy.](#)



# Mini-Projects V: 2024

1. [Uncertainty in Climate Predictions](#)
2. [The East African Climate Paradox](#)
3. [A Review of Contrail Detection and Avoidance](#)
4. [Research Dissemination of Climate Change Science Through Interactive Climate Simulations](#)
5. [Analysing the Effect of Climate Change on Models for Tuberculosis](#)
6. [Attribution Studies in Climate Science and Policy](#)
7. [Measuring the effect of COVID-19 on Labour Supply and Demand](#)
8. [Predicting the Curtain Call - Computing the Probability of Outbreaks Ending](#)
9. [Models for Describing and Forecasting the COVID-19 Pandemic](#)
10. [Using Mathematical and Statistical Techniques to Investigate the Transmission of Dengue](#)
11. [Modelling Happiness on a Social Network with Multiple Modelling Paradigms from Epidemiology](#)
12. [Adaptations in epidemiological modelling for co-infection](#)
13. [The Mathematical Modelling of Zombie Outbreaks - Why it Went Viral](#)
14. [Forecasting of the COVID-19 Pandemic - Agent Based Models](#)
15. [Exploring Epidemic Seeding Effects Through Geometric Random Graph Models](#)
16. [The Evolution of Cancer and a Cancer's Evolution](#)
17. [Dating the Origins of Coronaviruses](#)
18. [Neural Learning in Brain Organoids](#)
19. [Chromatic Number of Intersection Graphs with Certain Girth in Different Dimensions](#)
20. [Hopfield Networks and their Applications](#)
21. [State of the Art Computational Models for Alzheimer's Disease Progression](#)
22. [Decoding Decision-Making - Integrating Reinforcement Learning with Neural Mechanisms](#)
23. [A Review on Hopfield Network - Past, Present and Future](#)
24. [Computational neuroscience models in ADHD](#)
25. [Computational Modelling of Seizure Dynamics - The Epileptor and Whole Brain Modelling of Epilepsy](#)
26. [In Silico Modelling of Adaptive Deep Brain Stimulation for Parkinson's Disease](#)
27. [Graph Machine Learning for Neuroimaging Analysis - A Wholistic Review](#)
28. [AI and Neuroscience - Answer to Unlocking AGI?](#)
29. [Decoding the Human Mind - The Capabilities and Challenges in Speech Brain-Computer Interfaces](#)
30. [Learning Mechanisms of Artificial Neural Networks and the Brain](#)
31. [Navigating the Brain - Charting the Intersection of Space, Relations, Memory and Artificial Intelligence](#)
32. [Mathematical modelling of cell proliferation and tumour growth using fluorescent cell-cycle indicators](#)
33. [Turing Patterns in Chemical Systems - an Investigation of the Chlorite-Iodide-Malonic Acid Reaction](#)
34. [Agent Based Modelling of Collective Movement in Animals](#)
35. [Cellular Automata versus Partial Differential Equations - Sea Shells and Traffic Flow](#)
36. [Towards Large-Scale Computational Morphogenesis](#)
37. [Active Matter in Embryo Morphogenesis](#)
38. [Kalman Filters in Computational Biology](#)
39. [Deep Learning in Biomedical Segmentation](#)
40. [Generative Diffusion Models for De Novo Drug Molecule and Protein Design](#)
41. [Natural Language Processing Techniques for De Novo Protein Design](#)
42. [Fairness in Healthcare Machine Learning Applications](#)
43. [Advances in de novo protein design by deep learning methods](#)
44. [Computational Models of Working Memory in Cognitive Neuroscience](#)
45. [Surveying Machine Learning Methods for Predicting the Outcomes of Prime Editing](#)
46. [Decoding Nature's Symphony - Advances in Computational Bioacoustics](#)
47. [Beyond Traditional SIR - Advanced Models for COVID-19 Forecasting](#)
48. [Meta-Reinforcement Learning and Neuroscience](#)
49. [A Comparative Study of Deep Learning Techniques for Prediction of CO2 Emission in Traffic Vehicles](#)
50. [the human ectoparasite hypothesis for plague transmission](#)

# *Why didn't somebody write on?:*

*There is a wonderful variety in the reports, but there also some clustering like a lot on Compartmental Models in Pandemics, so here are some topics I would have written on in case I was to hand in a set of reports to be graded.*

## **ChemoInformatics**

*How does one efficiently sample chemical space subject to different constraints?*

*Can one restrict the set of relevant molecular graphs by knowing they have to be 3D realizable?*

## **Pandemics**

*The Models are advanced, but the data noisy and biased. How does models deal with this?*

*How do you rank possible scenarios for the next Pandemic?*

*How do you measure the Cost of Covid? Of Lockdowns?*

## **Molecular Dynamics**

*What has MD done for Sars Cov-2?*

*How do you represent movements of a virion rationally?*

*Is comparative MD useful?*

## **Modelling the Biosphere**

*Why are Climate forecasts so noisy? Where does the error enter?*

*What is the cost of Climate Change? Social Cost of Carbon*

*Granger Causality – Causality between Processes*

## **Morphogenesis**

*Why did Catastrophe Theory go Extinct? Could it still have a role to play?*

## **Population Genetics, Recombination and Mapping**

*Formulate a Birth-Death Process with Recombination*

## **Machine Learning**

*How do you quantify memory in Artificial and in Biological Networks?*

*LLMs are a huge success? But how do you represent more abstract concepts (Space, mathematics,...) in an Artificial Structure?*

## **Comparative Biology**

*Is there a role for comparative modelling in morphogenesis?*

## **Computational Neuroscience**

*How do you quantify memory in Artificial and in Biological Networks?*

*What are the simplest Neural Systems and have they been modelled?*

*What function does a biological neuron implement?*

## **Sequence Analysis**

*Write up the recursions for Long Indel Model for  $n$  sequences in Statistical Alignment*

## **Modelling Cells, Organs, Individuals**

*What are the prospects for predictive modelling of a prokaryote?*



# *The Biosphere and Climate Change*

## *Why this topic?*

*It is the worlds most important problem!!!*

*It will only become more important in coming decades.*

*It will swamp all news ad nauseum*

*CC research needs statistics, modelling, computer science and more*

*I could have given: alignment, recombination, origin of life, ....- but this is more rewarding for all.*

## *The Biosphere*

*1 Climate Earth System*

*2 Models*

*3 Climate Statistics*

*4 Climate Change*

## *Preparation*

*[Modelling the Dynamics of the Biosphere](#)*

*[Demystifying global climate models for use in the life sciences](#)*

*[Extreme weather impacts of climate change: an attribution perspective](#)*

*Browse Stainforth!!*

## *Appendix Lectures*

*1 IPCC6 I*

*2 IPCC6 II-III*

*3 Climate Books*

*Extras Lectures: This Lecture + QA1 + QA2 + Expert (Stainforth)*

*Overall Assessment – A wonderful topic, with a great diversity of subtopics, ranging from difficult mathematics to applications of real consequence. And good reviews/textbooks on almost any subtopic.*

# *The Biosphere and Climate Change*

## *Overall Objectives:*

*Dynamic Modelling of Earth with Focus on Biosphere*

*Forecasting, Causality and Control*

*Consequences for Humanity*

## *Needed:*

*Formal Representation of Biosphere*

*Models*

*Quantitative Data*

*Model Selection, Parametrization and Quality Evaluation*

## *Challenges:*

*Size and Time-Scale of the System*

*Stochasticity*

*Unpredictability*

*Dependency on Human Behaviour*

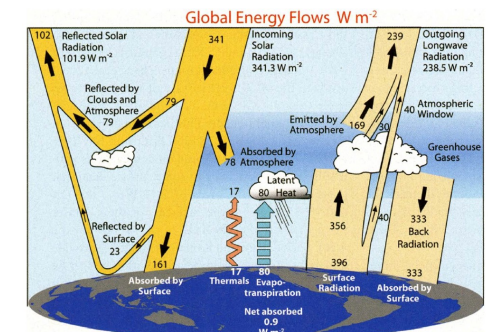
In even weeks you also you also have a class, where you present a paper of your choice in about 10 minutes and write about 1 page summary of your chosen paper. Below are 3 suggestions for papers, but students normally have no problems finding a paper themselves that they find interesting.

[A high-resolution record of early Paleozoic climate](#) [Trajectories of the Earth System in the Anthropocene](#)

[Estimating global mean sea-level rise and its uncertainties by 2100 and 2300 from an expert survey](#)



Wikipedia



Trenberth.

# *Computational Virology*

## *Why this topic?*

*Humanity Population Size is 8+ Billion, heading for 11+. This is a huge “Petri Dish” for pathogens, so more Pandemics is heading our way.*

*Like CC it will fill the media also post-COVID.*

*This research also needs statistics, modelling, computer science and more*

## *Computational Virology I : The Basics of COVID-19*

## *Computational Virology II : Epidemiological Models*

## *Computational Virology III: NPI, Forecasting, Cost, Epidemiological Parameters, Demography,..*

## *Computational Virology IV: Disease X, Clinical Trials, Variants, Outbreak Detection, ..*

***Overall Assessment** — Horrible to read about. A flood of un-reviewed papers and textbooks and it is impossible to sort it according to quality and conceptual clarity. For a period more than 700 papers and more than a book was published on this per day. After 4 years it is getting a lot better.*

# Computational Virology

## Overall Objectives:

*Map Key Epidemiological Parameters [infectiousness, mortality,..]*

*Functionally characterize the virus*

*Describe evolutionary processes and infer history of virus.*

*Describe virus-host interaction.*

*Forecast and evaluate control measures*

## Needed:

*Models*

*Data*

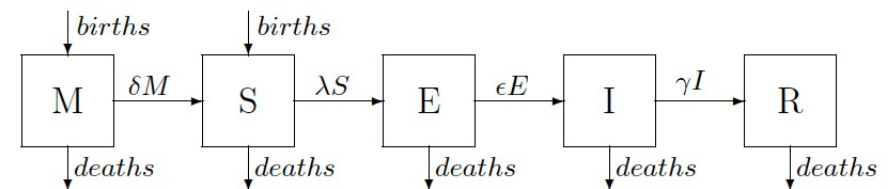
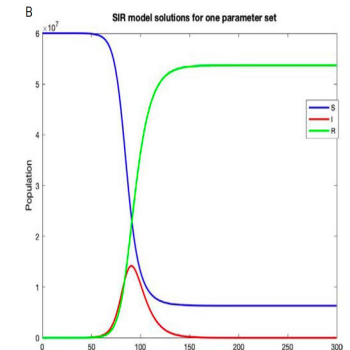
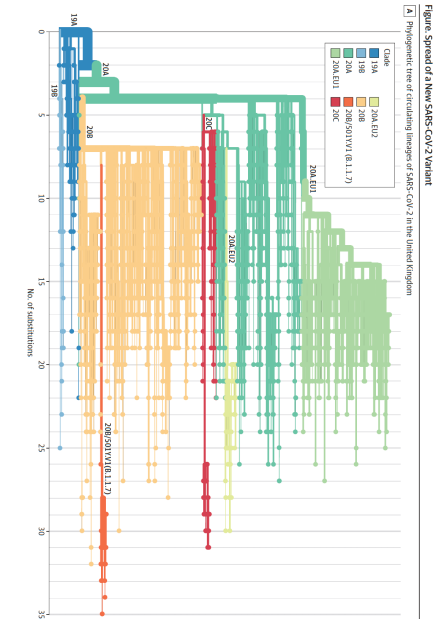
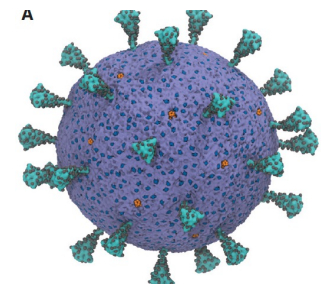
*Inference*

## Challenges:

*Computational Issues*

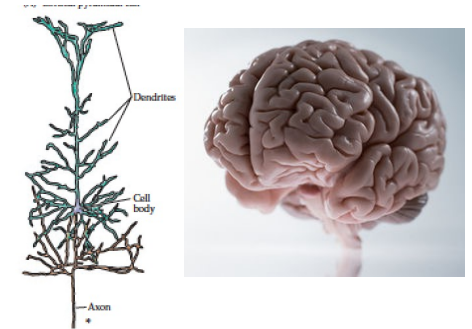
*Unbiased Data*

*Doing all in real-time*

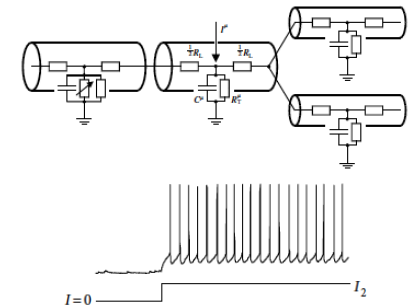


# Neuroscience

## The Biological and Artificial Neuron

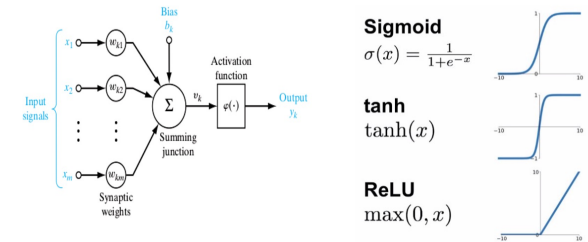


## Modelling Neurons



## Modelling Neural Networks

## Computational Modelling of the Biological Brain



**Overall Assessment** – *A massively challenging and fascinating topic that will dominate coming decades. The human brain is still “unknown” in terms of detailed wiring, despite massive progress. Even C. elegans [302 Neurons] cannot be simulated reliably. An interesting “parallel lives” between artificial and biological neurons/brains.*

# Computational Models of Shape and Form

## Why this topic?

*“Genotype  $\rightarrow$  Shape/Pattern” is one of the few Holy Grails of the Biosciences.*

*It will take decades or more to solve.*

*If you want to model or analyze Shape/Pattern, they must be represented.*

*The field is experiencing much progress presently.*

$$\begin{cases} u_t = D_u \nabla^2 u + \gamma(u + av - uv^2 - puv) \\ v_t = D_v \nabla^2 v + \gamma(bv + hu + uv^2 + puv) \end{cases}$$

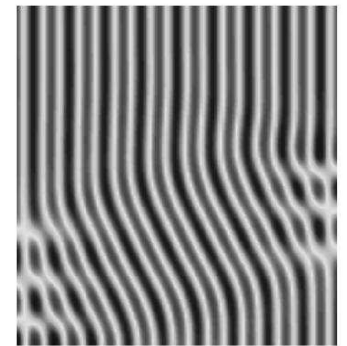
## Shape and Form in Multicellular Organisms

### The Classic Models

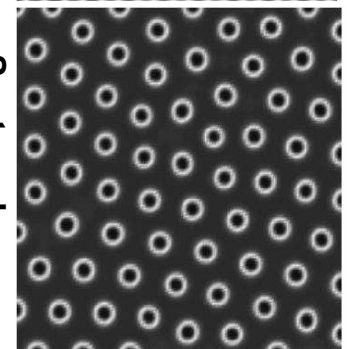
### Advanced Models

### Functions: Genomes $\rightarrow$ Shape/Form

Stripes:  $p$  small



Spots:  $p$  large





# *Computational/Mathematical Models of Shape and Form*

## *Overall Objectives:*

*Represent Shape/Pattern*

*Model/Predict Shapes/Patterns*

*Model Shape/Pattern Evolution*

*Compare Shapes/Patterns*



## *Needed:*

*Dynamic realistic models*

*Detailed Ontogenic [Life History] Shape/Pattern Data*

## *Challenges:*

*Organism are big computationally*

*Data not sufficiently detailed*

*The Genotype → Shape/Pattern Map is simply now known at present*



***Overall Assessment*** – *A wonderful topic, with a great readable literature. It is bound to be of increasing importance in coming decades.*

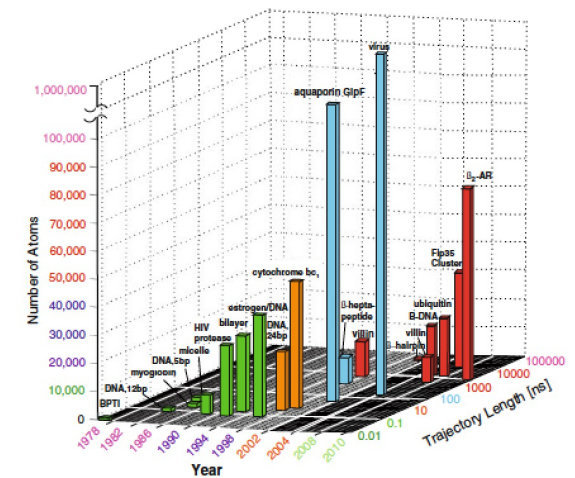
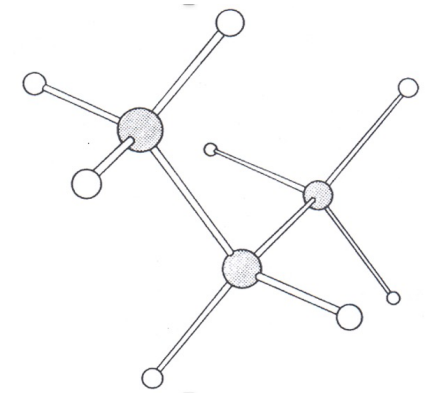
# Earlier Topic: Molecular Dynamics

## Physics of Molecules: QM and MM

## Integrators and Approximations

## MD of Protein Folding

## History and Applications of MD



Schlick, 2010

**Overall Assessment** – *At the bottom of the ladder of reductionism and is closer to explaining biology in terms of physics. However, the model are very heuristic and main exciting/progress is in more computation and thus applying them to larger and larger systems.*

# ***Earlier Topic: Deep Learning [DL] in the Biosciences [BS]***

## Goal of the 4 Lectures:

1. Basic introduction to ML+DL
2. Describing domains and achievements in 3 Bioscience Areas.

## ***Why this topic?***

*DL/ML (Machine Learning) has amazing achievements within the last decade within BS.*

*It will only become more important in coming decades.*

*Where will go? And what will be the role of traditional physical and evolutionary models?*

*There are many other application areas than the 3 chosen.*

## ***The Basics of Machine Learning [ML] and Deep Learning***

### ***Deep Learning and Structural Biology***

### ***Deep Learning and ChemoInformatics***

### ***Deep Learning and Genomics***

***Overall Assessment** – Amazingly successful in the last decade, but challenging to read about, since the field evolves very fast. There is no established terminology and results can be hard to understand since it can't be translated into model-parameters.*

# Deep Learning [DL] in the Biosciences

## Overall Objectives:

*Predict*

*Classify*

*Transfer Knowledge*

## Needed:

*Large Data Sets*

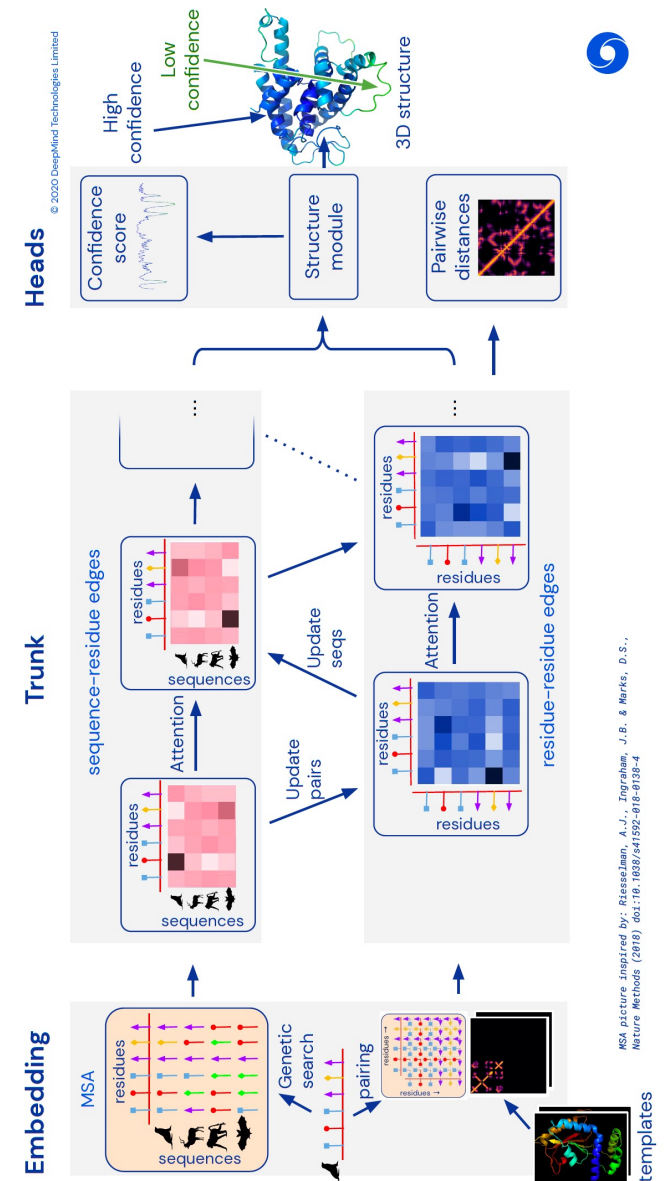
*Appropriate Deep Learning Model*

## Challenges:

*Interpretation*

*Relate it to traditional methods*

Nature 2021 July AlphaFold2



# Earlier Topic: ChemoInformatics

Mathematical Models of Origins of Life

Combinatorics of Small Molecules

Chemical Space

Reactions

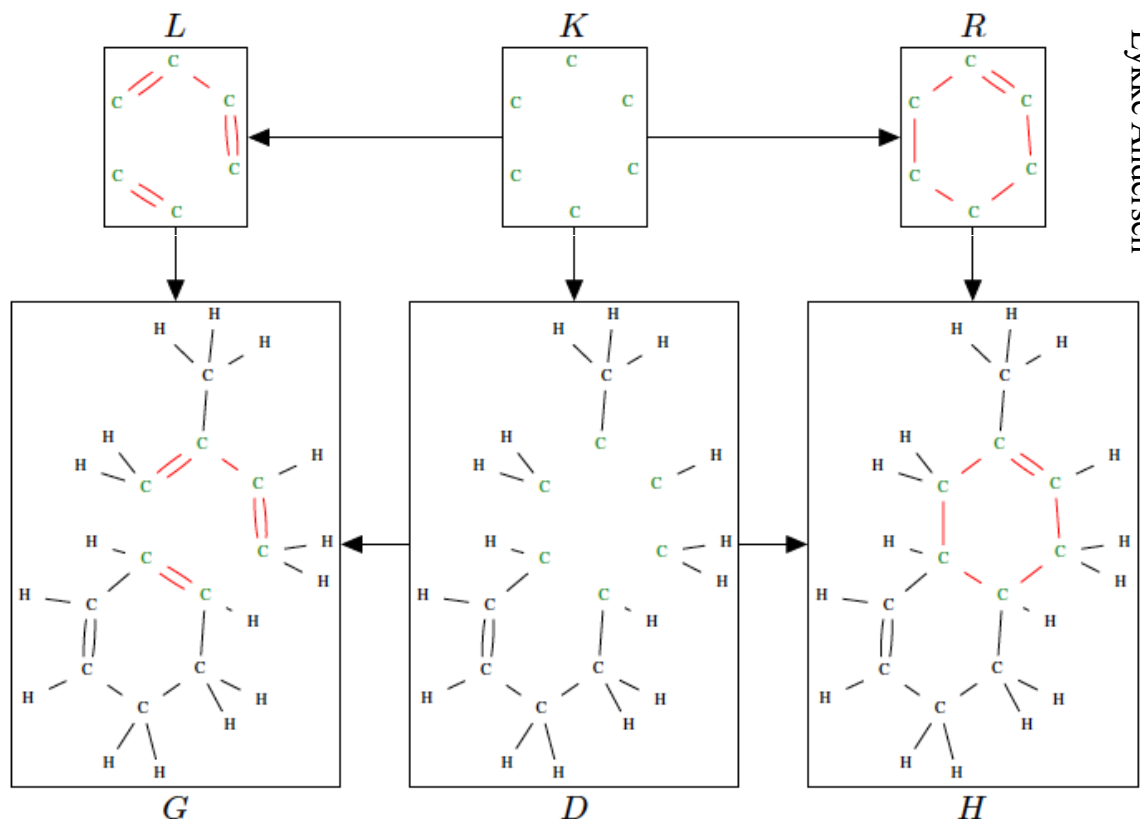


Table 2. Overview of the Structure Generation Process

nodes	graphs <sup>a</sup>	generated <sup>b</sup>	accepted <sup>c</sup>	unique tautomers (GDB) <sup>d</sup>	all tautomers	stereoisomers <sup>e</sup>
1	1	4	4	4	4	4
2	1	10	9	9	9	9
3	2	52	20	20	21	20
4	4	332	80	80	88	87
5	8	2294	357	352	397	469
6	20	18 066	1906	1850	2135	2911
7	57	154 542	10 953	10 568	12 438	19 904
8	194	1 445 073	69 563	66 706	79 899	153 601
9	705	14 213 741	464 402	444 313	540 002	1 258 963
10	2822	146 004 340	3 259 036	3 114 041	3 827 907	10 898 065
11	11 912	1 558 491 448	23 875 101	22 796 628	28 240 425	98 645 474
total	15 726	1 720 329 902	27 681 431	26 434 571	32 703 325	110 979 507

JP Reymond 2011

**Overall Assessment** – *Being close to drug-design, this is well funded and close to industry. It is good fun with lots of combinatorics, graph algorithms and DL has recently invaded it big time.*

# *Earlier Topic: Sequence Analysis*

1. Optimization Alignment - Pairwise

2. Optimization Alignment – Multiple: Exact Algorithms

3. Optimization Alignment – Multiple: Heuristics

4. Alignment Variants: HMM, Profile, Local, ....

5. Indel Processes and the Basic Model (TKF91)

6. Long Indel and Irreversible Model.

7. Multiple Statistical Alignment

8. Non-homologous Sequence Modeling

9. Applications of Alignment I: Ancestral Sequences, CoEvolution and Function

10. Applications of Alignment II: Annotation

11. Alignment beyond Sequences: Multigenes

12. Alignment beyond Sequences: Proteins

13. Alignment beyond Sequences: Networks

14. Alignment beyond Sequences: Various

15. Genome Comparison Algorithms

16. NGS Algorithms

17. Summary and Open Problems

18. Appendix Lecture : Tree Enumeration

**CTAGG**

**TT-GT**

**V-LSPADKTNVWGKVG AHAGEY**

**VHLTPEEKSVLWGKV--NVDEV**

**V-LSPADKTNVWGKVG AHAGEY**

**VHLTPEEKSVLWGKV--NVDEV**

***Overall Assessment – This is CLASSIC Bioinformatics with surprisingly much to do still.***



## *Earlier Topic: Population Genetics and Evolution*

## *Molecular Evolution*

## *Models in Molecular Evolution*

## *Rates in Molecular Evolution*

## *The Mutational Process*

## *Selection in Molecular Evolution*

## *The Evolution of Multigene Families*

# *Molecular Evolution and the Organism*

## *Population Genetics*

## *Models in Population Genetics*

## *The Coalescent*

## *The Coalescent and Recombination I*

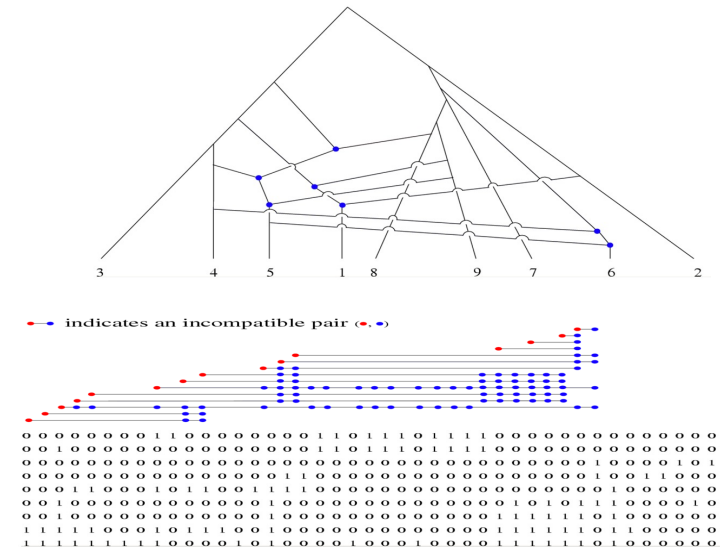
## *The Coalescent and Recombination II*

### *The Coalescent and Recombination III*

## *Data Analysis in Population Genetics*

# Human Population Genetics

## *Birth-Death Processes*



***Overall Assessment*** – Population Genetics and Evolution used to be fun little fringe topics, but have in the last 2 decades have become functionally important and thus exploded in funding. Evolution is the foundation of Comparative Genomics and PG of Genetic Mapping.