



## Prediction Methods - Homology Modelling

### I.Fold Assignment

- Early Fold Assignment Projects - 3D Profiles
- Bowie, Luthy and Eisenberg (1991) Science 253,pg.164
  - Incorporates three structure environment terms
    - Secondary structure
    - Solvent accessibility
    - Degree of burial of apolar vs polar atoms
  - Problems
    - As sequences diverge internal environments can change beyond detectability

## Prediction Methods - Homology Modelling

### I.Fold Assignment

- Protein fold recognition using secondary structure
  - Rost, B (1997) JMB 287, pg 797
  - Matches secondary structure prediction strings with a large secondary structure string database
  - Dependent on correct secondary structure prediction

## Prediction Methods - Homology Modelling

### I.Fold Assignment

- GenTHREADER (1999) JMB 287, pg 797
  - Fast, First in CASP4 (Asilomar 2000)
  - Profile based alignments using PSI-BLAST
  - PSI-BLAST profiles evaluated by neural network
  - Gives a thread
  - <http://www.pspred.net/>



## Prediction Methods - Homology Modelling

### II. Threading

```

#|056491|CCR4_PANAN|NAHFRRLPLPTIYIIFLSTLVVQSGVLLMVGKQLRIMTD-----KYLRLQVADLL
#|028474|CCR4_MGCPA|NAHFRRLPLPTIYIIFLSTLVVQSGVLLMVGKQLRIMTD-----KYLRLQVADLL
#|062747|CCR4_DIBTO|NAHFRRLPLPTIYIIFLSTLVVQSGVLLMVGKQLRIMTD-----KYLRLQVADLL
#|025930|CCR4_BOVEN|NAHFRRLPLPTIYIIFLSTLVVQSGVLLMVGKQLRIMTD-----KYLRLQVADLL
#|14993|ISR2_BOVEN|SR---AVLDFIIPFYVYVGLGRLTIVYVLEFARMTTTH-----YIHLALADEL
#|10460|ISR2_RAT|SR---AVLDFIIPFYVYVGLGRLTIVYVLEFARMTTTH-----YIHLALADEL
#|13075|ISR2_MOUSE|SR---AVLDFIIPFYVYVGLGRLTIVYVLEFARMTTTH-----YIHLALADEL
#|15374|ISR2_MOUSE|DR--HLEALVPLVYVYVGLGRLTIVYVLEFARMTTTH-----YIHLALADEL
#|15381|ISR2_RAT|DR--HLEALVPLVYVYVGLGRLTIVYVLEFARMTTTH-----YIHLALADEL
#|15050|ISR2_HUMAN|DR--HLEALVPLVYVYVGLGRLTIVYVLEFARMTTTH-----YIHLALADEL
#|02926|YF88_CARLE|LWVQLDFPPTAYRHTLVLDVLDLNRKRVVLETTREYR--LAKLPLRGLAIIIPPTI
#|01200|YF88_YEAST|LWVQLDFPPTAYRHTLVLDVLDLNRKRVVLETTREYR--LAKLPLRGLAIIIPPTI
#|010479|YF87_SCDPO|LWVQLDFPPTAYRHTLVLDVLDLNRKRVVLETTREYR--LAKLPLRGLAIIIPPTI
#|04030|YF86_YEAST|LWVQLDFPPTAYRHTLVLDVLDLNRKRVVLETTREYR--LAKLPLRGLAIIIPPTI
#|10259|SHV_MCVB|VWGGSTDFLWVLTGVLGSGDFPFRPNSGAVFRSGD--YLDLPLSPFPPTG
#|01374|SHV_MCVG|VWGGSTDFLWVLTGVLGSGDFPFRPNSGAVFRSGD--YLDLPLSPFPPTG
#|05887|YF03_CARLE|YRHTLDFPPTAYRHTLVLDVLDLNRKRVVLETTREYR--LAKLPLRGLAIIIPPTI
    
```

- Methods
  - Alignment Methods
  - Structure Based
    - Machine Learning
    - Rule Based

*Difficult Problem That Is Critical To The Model Quality*

## Prediction Methods - Homology Modelling

### II. Threading

- Threading Introduction
  - Task: Align a sequence with a homologous structure
  - Corrections after this step are difficult
- Common Tools
  - Multiple alignment tools, Blast, PSI-BLAST, ClustalW, etc
  - Secondary structure prediction tools, etc
  - Empirical protein potentials

## Prediction Methods - Homology Modelling

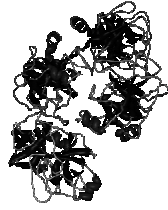
### II. Threading

- Empirical Protein Potential Methods, Sippl MJ (1990) JMB 213 pg 859
  - Idea:
    - Use residue-residue interaction statistics of protein structure databases to infer "energy-like" terms using statistical mechanics
    - Alignments that minimize energy are accepted
    - Computationally challenging
  - For residues  $R_1$  and  $R_2$  separated by a distance  $D$ ,  $P(R_1R_2|D)/P(R_1)P(R_2)$
  - Of interest, Thomas and Dill, JMB (1996) 257(2):457-69

### Prediction Methods - Homology Modelling

#### III. Model Building

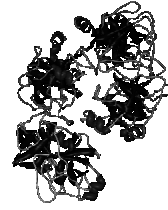
- **Methods**
  - All Atom
  - Backbone Based
- **Best solutions**
  - Modeller (guitar.rockefeller.edu)
  - SCWRL
  - Swiss Model
- **Problem**
  - The best solution is one that is closer to the actual sequence than any of the templates are.
  - Difficult because you can't sample many conformations



### Prediction Methods - SCWRL and General Methods

#### III. Model Building

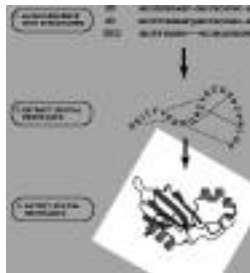
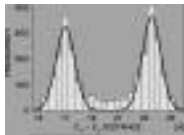
- **Generally,**
  - Threading by PSI-Blast
  - Template structure and aligned sequence are swapped.
  - Often, the backbone remains fixed and residues are swapped and repacked (SCWRL) Dunbrack, R Proteins 1999;Suppl 3:81-7



### Prediction Methods - Homology Modelling

#### III. Model Building

- **Satisfaction of spatial restraints** (JMB (1993) 234, 779-815)
  - Used by Modeller
  - For more info on spatial restraint satisfaction see:
    - Williams, et al. (2001) J Comp Bio 8(5):523-547



From <http://guitar.rockefeller.edu/modeller/manual/node12.html>

### Prediction Methods - Homology Modelling

#### IV. Model Evaluation

- **Methods**
  - Data Driven
  - Rule Based
  - Machine Learning Methods
- **Best solutions**
  - Come from x-ray crystallography and structure world
- **Refinement**
  - Molecular Simulation
  - Simulated Annealing
  - Rotamer Rule Based Methods

### Prediction - Structure Evaluation Methods

- **Biotech Validation Suite For Protein Structures**
  - (<http://biotech.ebi.ac.uk:8400/>)
  - Evaluates a model

<b>Procheck</b> <ul style="list-style-type: none"> <li>• Geometry</li> <li>• Planarity</li> <li>• Dihedral Angles</li> <li>• Chirality</li> <li>• Hydrogen Bonds</li> <li>• Disulphide Bonds</li> <li>• etc</li> </ul>	<b>Prove</b> <ul style="list-style-type: none"> <li>• Volume Analysis</li> </ul>	<b>Whatif</b> <ul style="list-style-type: none"> <li>• Bond Angles</li> <li>• Bond Lengths</li> <li>• Proline Ring Puckers</li> <li>• Quality Check</li> <li>• Rotamer Check</li> <li>• Water Clusters</li> <li>• Torsion Angles</li> <li>• Atomic Occupancy</li> <li>• and more</li> </ul>
--	--	---

### Prediction - Structure Evaluation Methods

- **Prosa II**
  - (<http://www.came.sbg.ac.at/Services/prosa.html>)
  - Evaluates a model based on knowledge of other structures
  - Sippl, M.J. Proteins 17:355-362 (1992)



## Prediction Methods - Homology Modelling

### IV. Model Evaluation

- Biotech Validation Suite For Protein Structures
  - (<http://biotech.ebi.ac.uk:8400/>)
  - Evaluates a model

#### Procheck

- Geometry
- Planarity
- Dihedral Angles
- Chirality
- Hydrogen Bonds
- Disulphide Bonds
- etc

#### Prove

- Volume Analysis

#### Whatif

- Bond Angles
- Bond Lengths
- Proline Ring Puckers
- Quality Check
- Rotamer Check
- Water Clusters
- Torsion Angles
- Atomic Occupancy
- and more

## Prediction Methods - Homology Modelling

### IV. Model Evaluation

- Prosa II
  - (<http://www.came.sbg.ac.at/Services/prosa.html>)
  - Evaluates a model based on knowledge of other structures
  - Sippl, M.J. Proteins 17:355-362 (1992)

## Prediction Methods - Homology Modelling

### IV. Model Evaluation

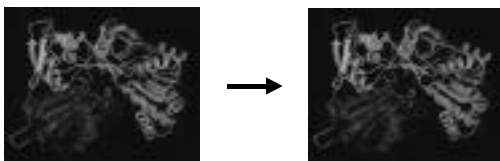
- Protein Secondary Structure Prediction
  - PredictProtein (<http://maple.bioc.columbia.edu/predictprotein/>)
  - Chou-Fasman (<http://fasta.biotech.virginia.edu/fasta/chofas.htm>)
  - GOR (<http://molbio.soton.ac.uk/computer/GOR.html>)
  - NN Predict (<http://www.cmpharm.ucsf.edu/~nomi/nnpredict.html>)
- Methods
  - Neural Network
  - Rule Based
  - Other Machine Learning
  - Homology Based

## The Fifth Step of Prediction - Refinement

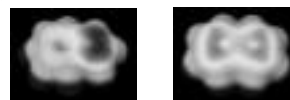
- Molecular Mechanics Methods
  - Popular Forcefields
    - AMBER ([www.amber.ucsf.edu](http://www.amber.ucsf.edu))
    - CHARMM
- Method
  - Parameters for atom types are determined
    - Usually developed from experiments or calculations on small molecules
  - Atoms are simulated by solving Newton's equations of force
- QM/MM methods integrate quantum calculations with molecular simulation

## Prediction - Refinement

- Molecular Mechanics Methods
  - Refinements don't always lead to better structures



## Structure - QM Methods



Credit: [www.gaussian.com](http://www.gaussian.com)

- Tools
  - Gaussian94/98 - ([www.gaussian.com](http://www.gaussian.com))
- Method
  - Calculates structures/energies of very small molecules very accurately
  - Takes a lot of time to learn, run and interpret

### Prediction - What do the results mean?

- The quality of the results vary on the difficulty of the problem and the context of the question being asked.
  - Never build a model without a question
  - Always make sure your model resolution can answer that question

*It is very easy to build an incorrect structure, you will build many of them. This is because it is much easier to find one of the 2<sup>N</sup> different structures than the ground state structure.*

### Scientific Projects

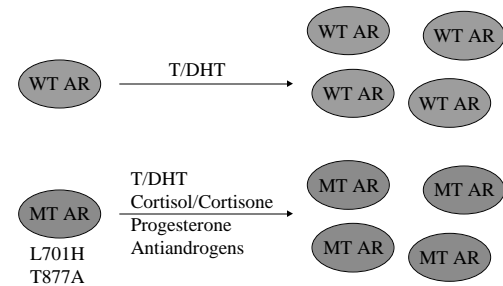
- The good: Androgen Receptor Ligand Binding Domain (Bruce Gottlieb, McGill University)
- The less good: Granzyme M (Sami Mahrus)
- The ugly: Modeling Nitric Oxide Synthase (Gigi Knudsen)

### Androgen Independent Prostate Cancer

- Prostate cancers often respond to endocrine therapy.
- Conversion of prostate cancers to an androgen independent state is often inevitable
  - AR is still expressed, sometimes overexpressed
  - Mutations in AR are associated with this conversion
  - Underlying mechanisms are unclear in most cases, but there are examples

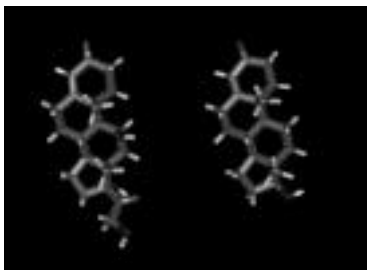
### Molecular Causes of Disease in AR

- Prostate Cancer



Nature Medicine 6:6, pg 628 (2000)

### Molecular Basis For Promiscuousness

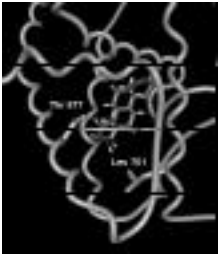



### Scientific Projects: The good: Androgen Receptor Ligand Binding Domain (Bruce Gottlieb, McGill University)

- Quantitatively model different ligands to show differences when disease associated mutations are present
- High degree of identity to other templates
  - Used progesterone (>40%) and estrogen (>30%) receptor LBD's as templates
  - Aligned with ClustalW
  - Built with modeler
  - Minimized and calculated in AMBER
- Less than 3 Angstrom RMSD with actual structure



### Molecular Basis For Promiscuosness

- Mutations:
  - L701H - Affinity for cortisone
  - L701H and T877A - Affinity for cortisone, testosterone and progesterone

### Experimental Protocol

- Simulations of proteins
  - Simulate wildtype protein with T, DHT Cortisol and Cortisone
  - Simulate mutant protein with T, DHT, Cortisol and Cortisone
  - Use MM/PBSA method to estimate binding free energy differences
  - Use results to correlate with observed trend and to build a convincing molecular model of the mutations

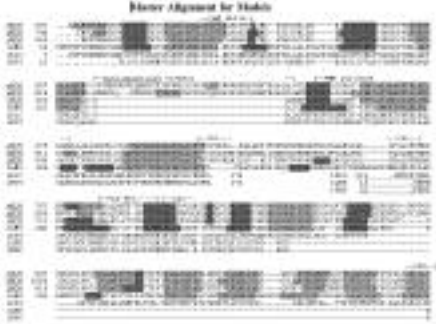
### Scientific Projects: The less good (but still pretty good): Granzyme M (Sami Mahrus)

- Question: Determinants of specificity of a serine protease
- High degree of identity to other templates
  - Two stage modeling process
    - Step One: Model structure w/ 4 templates
    - Step Two: Using better templates for active site remodel active site
  - Templates aligned using structure alignment software then sequence aligned with ClustalW, then refined by hand
  - Worked after lots of trial and error

### Scientific Projects: The ugly: Modeling Nitric Oxide Synthase (Gigi Knudsen)

- Question: Role of specific loops in protein function
- Low degree of identity to other templates, very large protein
  - Aligned with ClustalW, evaluated with secondary structure prediction
  - Templates aligned using structure alignment software then sequence aligned with ClustalW, then refined by hand
  - Loops are predicted to be near binding site, experiments confirmed.


### Scientific Projects: The ugly: Modeling Nitric Oxide Synthase (Gigi Knudsen)



Blower Alignment for N14616

Seq ID	Accession	Length	Identity	Positives	Negatives	Gaps
1	U00000	14616	100%	14616	0	0
2	U00001	14616	100%	14616	0	0
3	U00002	14616	100%	14616	0	0
4	U00003	14616	100%	14616	0	0
5	U00004	14616	100%	14616	0	0
6	U00005	14616	100%	14616	0	0
7	U00006	14616	100%	14616	0	0
8	U00007	14616	100%	14616	0	0
9	U00008	14616	100%	14616	0	0
10	U00009	14616	100%	14616	0	0
11	U00010	14616	100%	14616	0	0
12	U00011	14616	100%	14616	0	0
13	U00012	14616	100%	14616	0	0
14	U00013	14616	100%	14616	0	0
15	U00014	14616	100%	14616	0	0
16	U00015	14616	100%	14616	0	0
17	U00016	14616	100%	14616	0	0
18	U00017	14616	100%	14616	0	0
19	U00018	14616	100%	14616	0	0
20	U00019	14616	100%	14616	0	0
21	U00020	14616	100%	14616	0	0
22	U00021	14616	100%	14616	0	0
23	U00022	14616	100%	14616	0	0
24	U00023	14616	100%	14616	0	0
25	U00024	14616	100%	14616	0	0
26	U00025	14616	100%	14616	0	0
27	U00026	14616	100%	14616	0	0
28	U00027	14616	100%	14616	0	0
29	U00028	14616	100%	14616	0	0
30	U00029	14616	100%	14616	0	0
31	U00030	14616	100%	14616	0	0
32	U00031	14616	100%	14616	0	0
33	U00032	14616	100%	14616	0	0
34	U00033	14616	100%	14616	0	0
35	U00034	14616	100%	14616	0	0
36	U00035	14616	100%	14616	0	0
37	U00036	14616	100%	14616	0	0
38	U00037	14616	100%	14616	0	0
39	U00038	14616	100%	14616	0	0
40	U00039	14616	100%	14616	0	0
41	U00040	14616	100%	14616	0	0
42	U00041	14616	100%	14616	0	0
43	U00042	14616	100%	14616	0	0
44	U00043	14616	100%	14616	0	0
45	U00044	14616	100%	14616	0	0
46	U00045	14616	100%	14616	0	0
47	U00046	14616	100%	14616	0	0
48	U00047	14616	100%	14616	0	0
49	U00048	14616	100%	14616	0	0
50	U00049	14616	100%	14616	0	0
51	U00050	14616	100%	14616	0	0
52	U00051	14616	100%	14616	0	0
53	U00052	14616	100%	14616	0	0
54	U00053	14616	100%	14616	0	0
55	U00054	14616	100%	14616	0	0
56	U00055	14616	100%	14616	0	0
57	U00056	14616	100%	14616	0	0
58	U00057	14616	100%	14616	0	0
59	U00058	14616	100%	14616	0	0
60	U00059	14616	100%	14616	0	0
61	U00060	14616	100%	14616	0	0
62	U00061	14616	100%	14616	0	0
63	U00062	14616	100%	14616	0	0
64	U00063	14616	100%	14616	0	0
65	U00064	14616	100%	14616	0	0
66	U00065	14616	100%	14616	0	0
67	U00066	14616	100%	14616	0	0
68	U00067	14616	100%	14616	0	0
69	U00068	14616	100%	14616	0	0
70	U00069	14616	100%	14616	0	0
71	U00070	14616	100%	14616	0	0
72	U00071	14616	100%	14616	0	0
73	U00072	14616	100%	14616	0	0
74	U00073	14616	100%	14616	0	0
75	U00074	14616	100%	14616	0	0
76	U00075	14616	100%	14616	0	0
77	U00076	14616	100%	14616	0	0
78	U00077	14616	100%	14616	0	0
79	U00078	14616	100%	14616	0	0
80	U00079	14616	100%	14616	0	0
81	U00080	14616	100%	14616	0	0
82	U00081	14616	100%	14616	0	0
83	U00082	14616	100%	14616	0	0
84	U00083	14616	100%	14616	0	0
85	U00084	14616	100%	14616	0	0
86	U00085	14616	100%	14616	0	0
87	U00086	14616	100%	14616	0	0
88	U00087	14616	100%	14616	0	0
89	U00088	14616	100%	14616	0	0
90	U00089	14616	100%	14616	0	0
91	U00090	14616	100%	14616	0	0
92	U00091	14616	100%	14616	0	0
93	U00092	14616	100%	14616	0	0
94	U00093	14616	100%	14616	0	0
95	U00094	14616	100%	14616	0	0
96	U00095	14616	100%	14616	0	0
97	U00096	14616	100%	14616	0	0
98	U00097	14616	100%	14616	0	0
99	U00098	14616	100%	14616	0	0
100	U00099	14616	100%	14616	0	0
101	U00100	14616	100%	14616	0	0

### Scientific Projects: The ugly: Modeling Nitric Oxide Synthase (Gigi Knudsen)



## Post-Genomic Applications

- Useful Structure Databases
  - ModBase ([pipe.rockefeller.edu/modbase/](http://pipe.rockefeller.edu/modbase/))
  - Dali ([www.ebi.ac.uk/dali/domain/2.0/](http://www.ebi.ac.uk/dali/domain/2.0/))
  - FSSP ([www.ebi.ac.uk/dali/fssp/fssp.html](http://www.ebi.ac.uk/dali/fssp/fssp.html))
- Future
  - Improvement of ab initio methods
  - Determination of all naturally evolved scaffolds