

BMC WorkShop

# Protein Structure Prediction

## Introduction

Marc A. Marti-Renom & Damien Devos

Department of Biopharmaceutical Sciences, UCSF

June 17th and 18th, 2004

# Objective

**TO LEARN HOW-TO MODEL A  
3D-STRUCTURE FROM SEQUENCE**

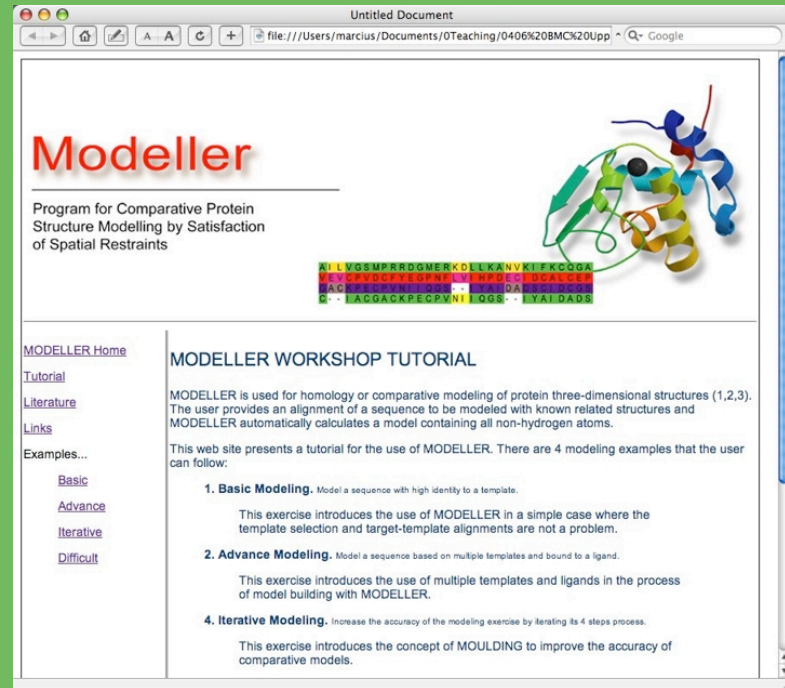
# Program

Template Search

Target – Template  
Alignment

Model Building


Model Evaluation



Untitled Document  
file:///Users/marcus/Documents/OTeaching/0406%20MCK%20Upp - Google

## Modeller

Program for Comparative Protein  
Structure Modelling by Satisfaction  
of Spatial Restraints



```
A I L V G S M P R R D G M E R K D L L K A N V K I F K G G A  
K Y C P V D C F Y E G P H E L Y I H P E C I D C A L C E  
R A C R E P R P R P R P R P R P R P R P R P R P R P R P R P R  
C - I A C G A C K P E C P V N I I Q G S - L V A I D A D S
```

[MODELLER Home](#)  
[Tutorial](#)  
[Literature](#)  
[Links](#)  
Examples...

### MODELLER WORKSHOP TUTORIAL

MODELLER is used for homology or comparative modeling of protein three-dimensional structures (1,2,3). The user provides an alignment of a sequence to be modeled with known related structures and MODELLER automatically calculates a model containing all non-hydrogen atoms.

This web site presents a tutorial for the use of MODELLER. There are 4 modeling examples that the user can follow:

- 1. Basic Modeling.** Model a sequence with high identity to a template.  
This exercise introduces the use of MODELLER in a simple case where the template selection and target-template alignments are not a problem.
- 2. Advance Modeling.** Model a sequence based on multiple templates and bound to a ligand.  
This exercise introduces the use of multiple templates and ligands in the process of model building with MODELLER.
- 4. Iterative Modeling.** Increase the accuracy of the modeling exercise by iterating its 4 steps process.  
This exercise introduces the concept of MOULDING to improve the accuracy of comparative models.

<http://www.salilab.org/modeller/workshop/>

# What are we going to do?

- Ask!
- Each day...
  - Basic introduction
  - Theory (representation-scoring-optimization)
  - Available programs
  - Application

# Nomenclature

**Homology:** Sharing a common ancestor, may have similar or dissimilar functions

**Similarity:** Score that quantifies the degree of relationship between two sequences.

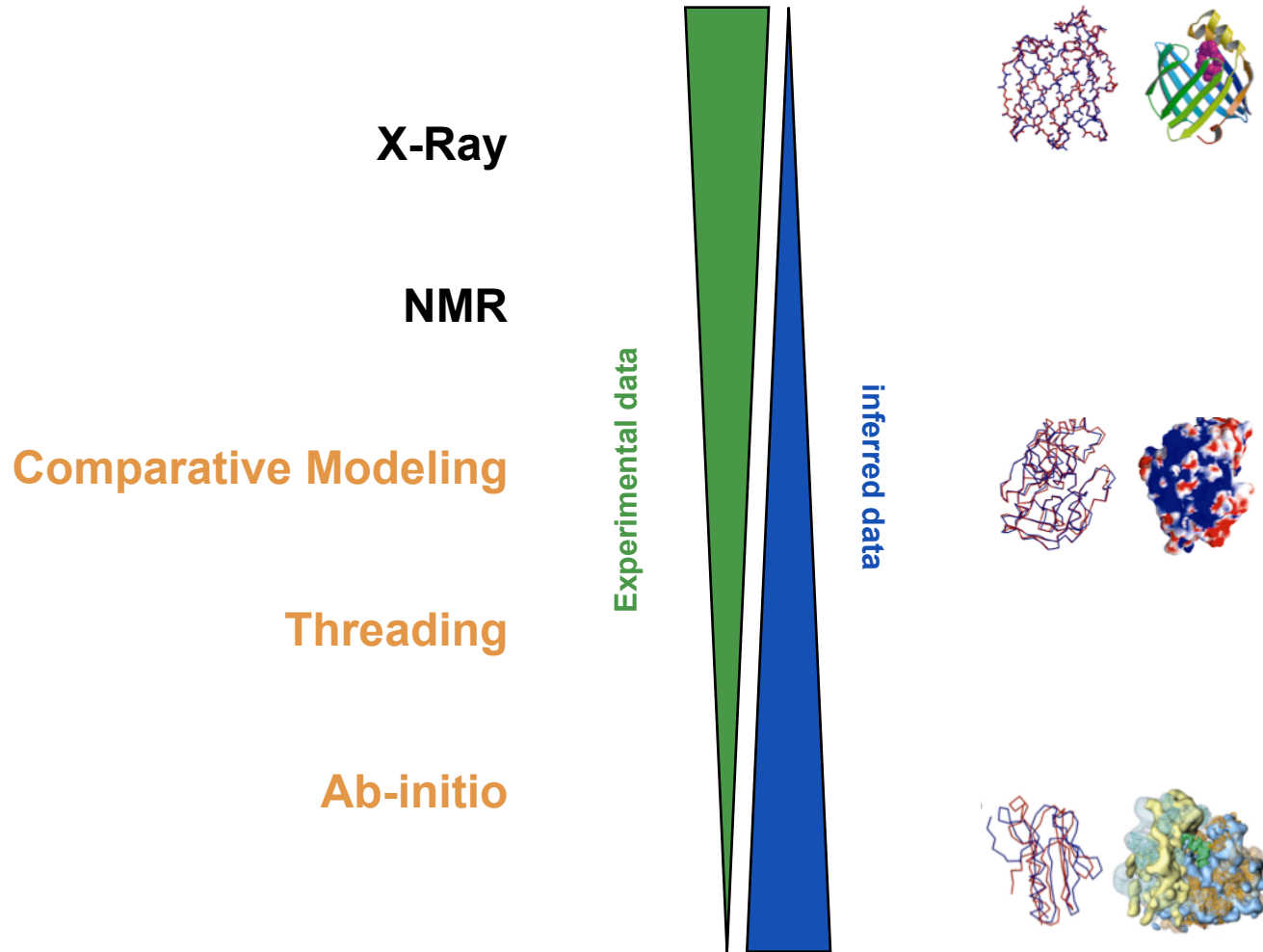
**Identity:** Fraction of identical aminoacids between two aligned sequences (case of similarity).

**Target:** Sequence corresponding to the protein to be modeled.

**Template:** 3D structure/s to be used during protein structure prediction.

**Model:** Predicted 3D structure of the target sequence.

# protein prediction .vs. protein determination



## Why protein structure prediction?

	Y 2004	Y 2006
Sequences	1,500,000	millions
Structures	28,000	50,000

## Why protein structure prediction?

	Y 2004
Sequences	1,500,000
Structures	400,000

<http://salilab.org/modbase/>

Theory



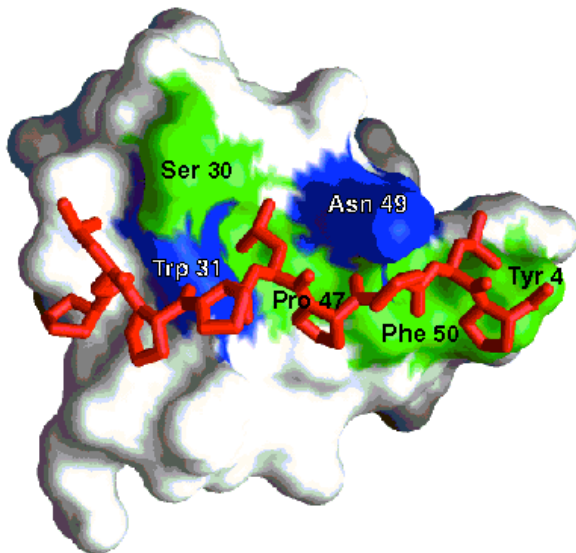
Experiment



## Why is it useful to know the **structure** of a protein, not only its sequence?

- The biochemical function (activity) of a protein is defined by its interactions with other molecules.
- 
- The biological function is in large part a consequence of these interactions.
- 
- The 3D structure is more informative than sequence because interactions are determined by residues that are close in space but are frequently distant in sequence.

YDL117W  
(15-64)      10      20      30      40      50  
K A R Y G W S G Q T K G D L G F L E G D I M E V T R I A G S W F Y G K L L R N K K C S G Y F P H N F

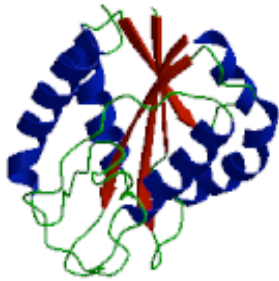
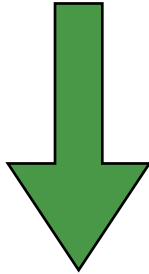


In addition, since evolution tends to conserve function and function depends more directly on structure than on sequence, **structure is more conserved in evolution than sequence.**

The net result is that **patterns in space are frequently more recognizable than patterns in sequence.**

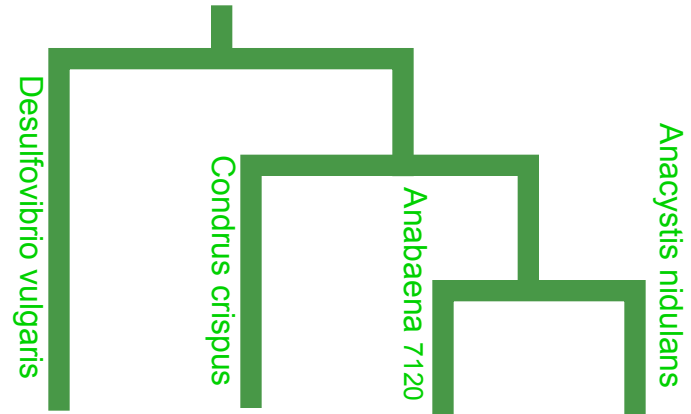
# Principles of Protein Structure

GFCHIKAYTRLIMVG...

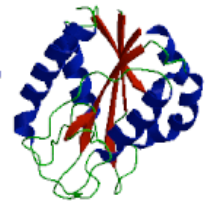


## Folding

Ab initio prediction



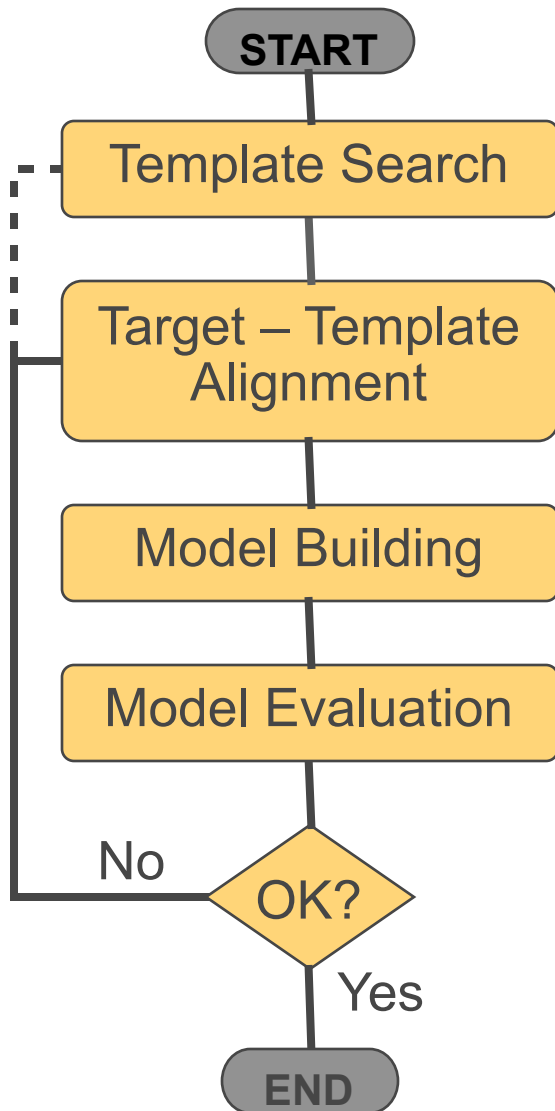
GFCHIK  
AYTRLI  
MVG...



## Evolution

Threading  
Comparative Modeling

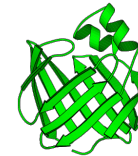
# Steps in Comparative Protein Structure Modeling



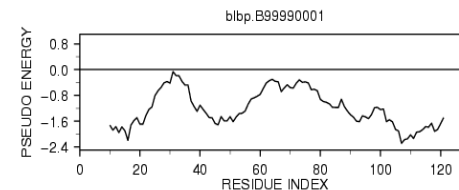
**TARGET**

ASILPKRFLFGNCEQTSDEG  
LKIERTPLVPHISAQNVCLKI  
DDVPERLIPERASFQWMN  
DK

**TEMPLATE**

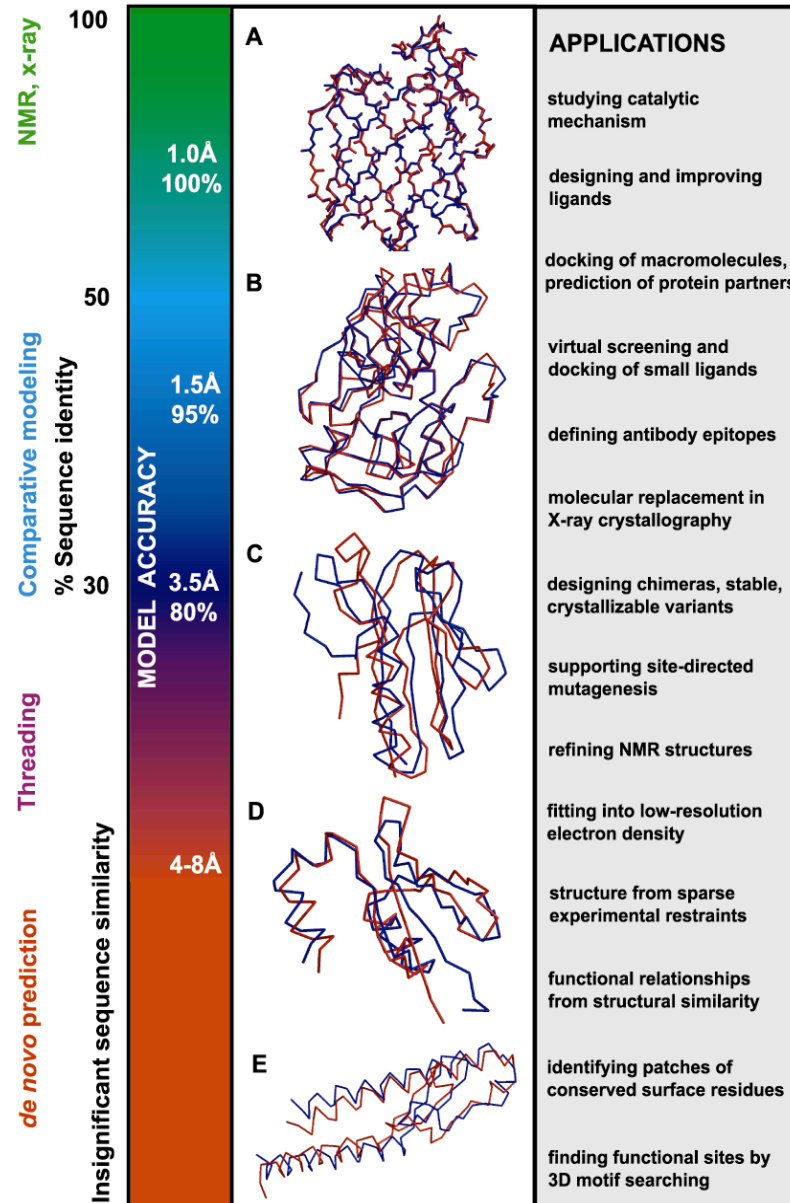


ASILPKRFLFGNCEQTSDEGLKIERTPLVPHISAQNVCLKIDDVPERLIPE  
MSVIPKRLYGNCEQTSEEAIRIEDSPIV---TADLVCLKIDEIPERLVGE



A. Šali, *Curr. Opin. Biotech.* 6, 437, 1995.  
R. Sánchez & A. Šali, *Curr. Opin. Str. Biol.* 7, 206, 1997.  
M. Marti et al. *Ann. Rev. Biophys. Biomolec. Struct.*, 29, 291, 2000.  
<http://salilab.org/>

# Utility of protein structure models, despite errors



D. Baker & A. Sali.  
*Science* **294**, 93, 2001.

# General References

## Protein Structure Prediction:

Marti-Renom et al. *Annu. Rev. Biophys. Biomol. Struct.* 29, 291-325, 2000.  
Baker & Sali. *Science* 294, 93-96, 2001.

## Comparative Modeling:

Marti-Renom et al. *Annu. Rev. Biophys. Biomol. Struct.* 29, 291-325, 2000.  
Marti-Renom et al. *Current Protocols in Protein Science* 1, 2.9.1-2.9.22, 2002.

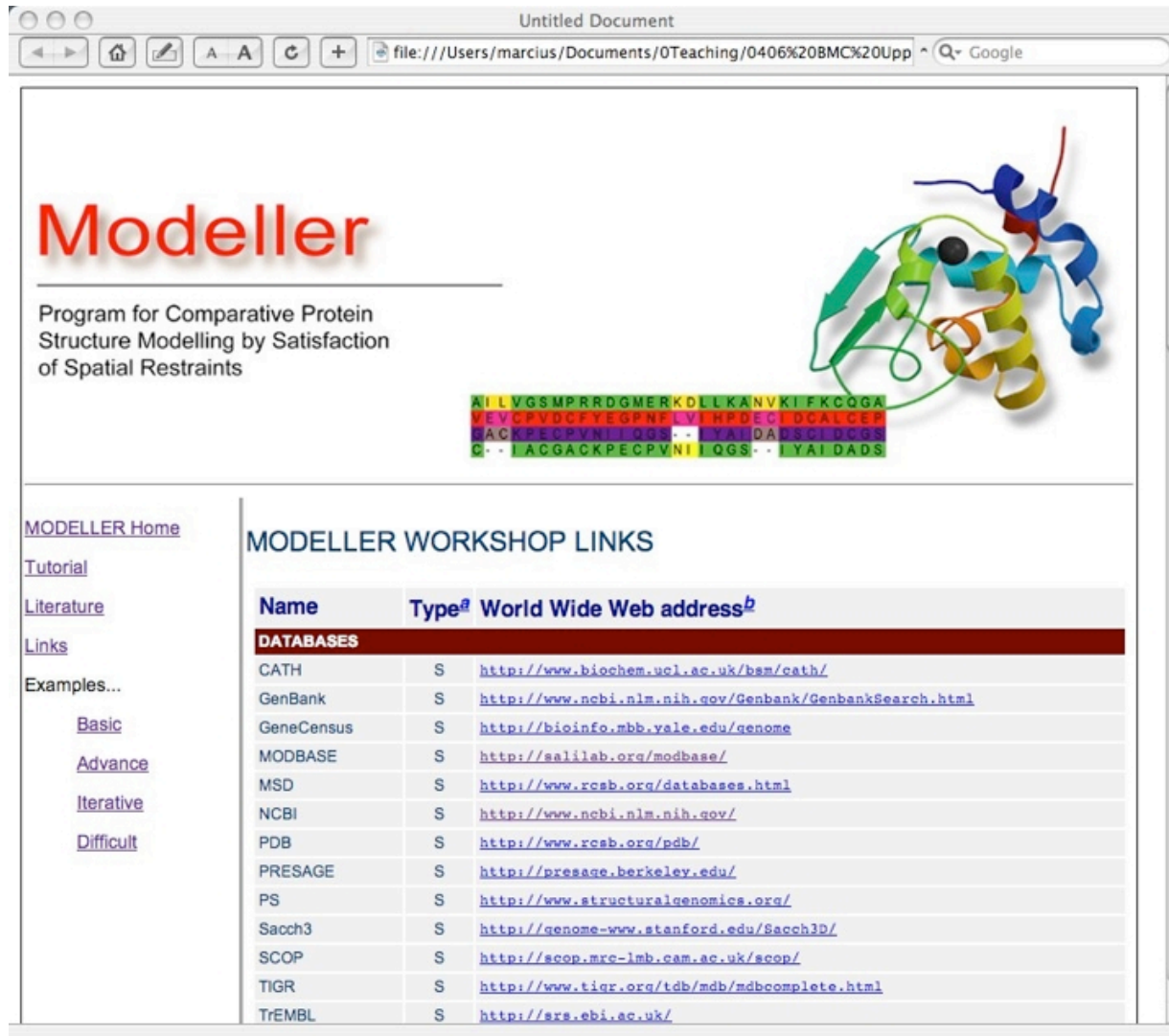
## MODELLER:

Sali & Blundell. *J. Mol. Biol.* 234, 779-815, 1993.

## Structural Genomics:


Sali. *Nat. Struct. Biol.* 5, 1029, 1998.  
Burley et al. *Nat. Genet.* 23, 151, 1999.  
Sali & Kuriyan. *TIBS* 22, M20, 1999.  
Sanchez et al. *Nat. Str. Biol.* 7, 986, 2000.  
Baker & Sali. *Science* 294, 93-96, 2001.  
Vitkup et al. *Nat. Struct. Biol.* 8, 559, 2001.

<http://www.salilab.org/modeller/workshop/links/>



# Modeller

Program for Comparative Protein Structure Modelling by Satisfaction of Spatial Restraints



```
A I L V G S M P R R D G M E R K D L L K A N V K I F K C O G A
V E V C P V D C F Y E G P N F L V I H P D E C L D C A L C E P
M A C K P E C P V N I I Q G S - - Y A I D A D S G I D C G S
C - - I A C G A C K P E C P V N I I Q G S - - Y A I D A D S
```

[MODELLER Home](#)  
[Tutorial](#)  
[Literature](#)  
[Links](#)  
[Examples...](#)

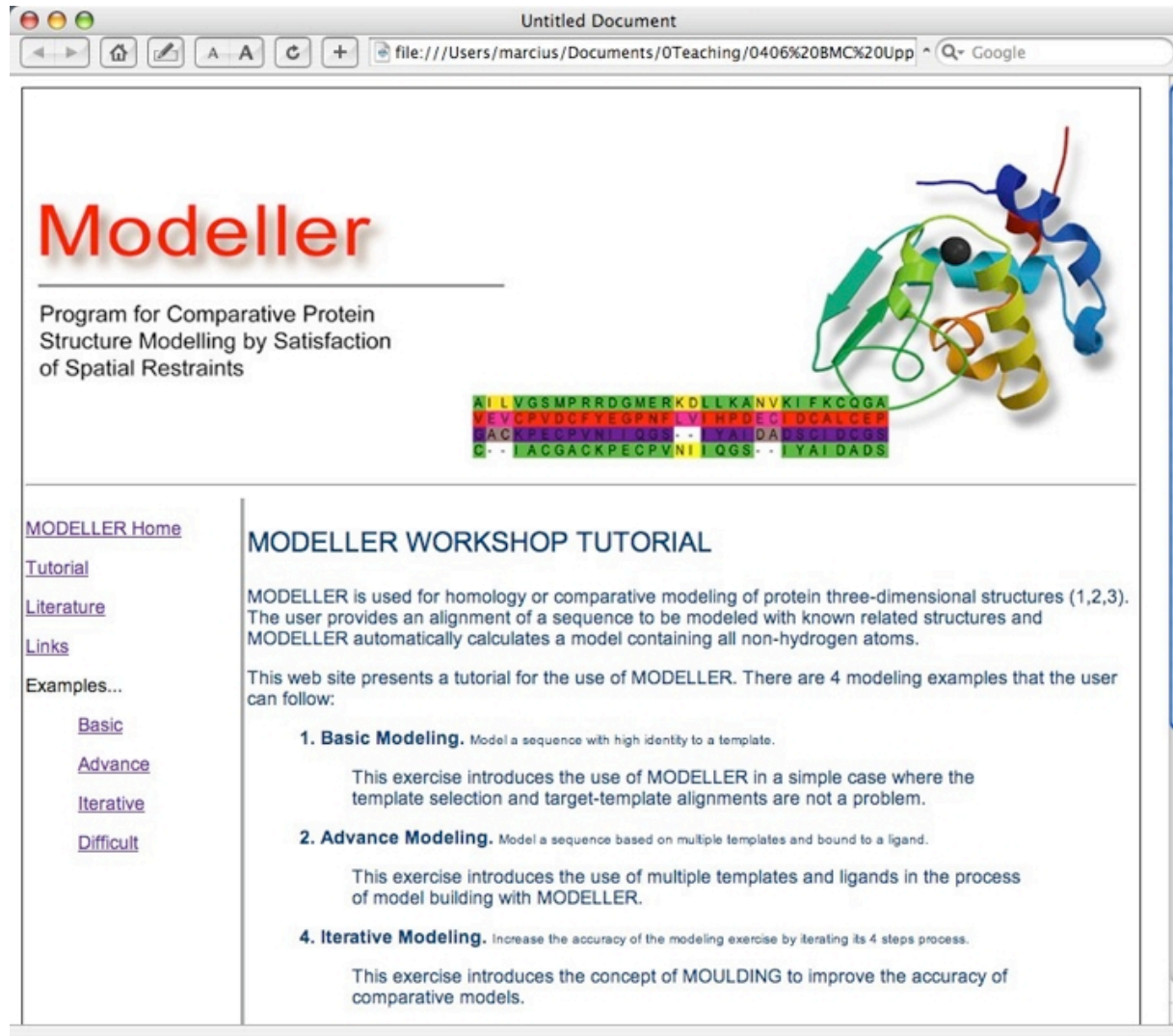
- [Basic](#)
- [Advance](#)
- [Iterative](#)
- [Difficult](#)

## MODELLER WORKSHOP LINKS

Name	Type <sup>a</sup>	World Wide Web address <sup>b</sup>
<b>DATABASES</b>		
CATH	S	<a href="http://www.biochem.ucl.ac.uk/bsm/cath/">http://www.biochem.ucl.ac.uk/bsm/cath/</a>
GenBank	S	<a href="http://www.ncbi.nlm.nih.gov/Genbank/GenbankSearch.html">http://www.ncbi.nlm.nih.gov/Genbank/GenbankSearch.html</a>
GeneCensus	S	<a href="http://bioinfo.mbb.yale.edu/genome">http://bioinfo.mbb.yale.edu/genome</a>
MODBASE	S	<a href="http://salilab.org/modbase/">http://salilab.org/modbase/</a>
MSD	S	<a href="http://www.rcsb.org/databases.html">http://www.rcsb.org/databases.html</a>
NCBI	S	<a href="http://www.ncbi.nlm.nih.gov/">http://www.ncbi.nlm.nih.gov/</a>
PDB	S	<a href="http://www.rcsb.org/pdb/">http://www.rcsb.org/pdb/</a>
PRESAGE	S	<a href="http://presage.berkeley.edu/">http://presage.berkeley.edu/</a>
PS	S	<a href="http://www.structuralgenomics.org/">http://www.structuralgenomics.org/</a>
Sacch3	S	<a href="http://genome-www.stanford.edu/Sacch3D/">http://genome-www.stanford.edu/Sacch3D/</a>
SCOP	S	<a href="http://scop.mrc-lmb.cam.ac.uk/scop/">http://scop.mrc-lmb.cam.ac.uk/scop/</a>
TIGR	S	<a href="http://www.tigr.org/tdb/mdb/mdbcomplete.html">http://www.tigr.org/tdb/mdb/mdbcomplete.html</a>
TrEMBL	S	<a href="http://srs.ebi.ac.uk/">http://srs.ebi.ac.uk/</a>

# Web site...

<http://www.salilab.org/modeller/workshop/>




Untitled Document

file:///Users/marcus/Documents/0Teaching/0406%20BMC%20Upp - Google

# Modeller

Program for Comparative Protein Structure Modelling by Satisfaction of Spatial Restraints



```
A I L V G S M P R R D G M E R K D L L K A N V K I F K C Q G A
V E Y C P V D C F T E G P N F L V H P D E C D C A L G E R
M A C K P E C P V N I Q G S - - Y A I D A D S C D G S
G - - I A C G A C K P E C P V N I Q G S - - Y A I D A D S
```

[MODELLER Home](#)

[Tutorial](#)

[Literature](#)

[Links](#)

Examples...

- [Basic](#)
- [Advance](#)
- [Iterative](#)
- [Difficult](#)

## MODELLER WORKSHOP TUTORIAL

MODELLER is used for homology or comparative modeling of protein three-dimensional structures (1,2,3). The user provides an alignment of a sequence to be modeled with known related structures and MODELLER automatically calculates a model containing all non-hydrogen atoms.

This web site presents a tutorial for the use of MODELLER. There are 4 modeling examples that the user can follow:

- 1. Basic Modeling.** Model a sequence with high identity to a template.

This exercise introduces the use of MODELLER in a simple case where the template selection and target-template alignments are not a problem.
- 2. Advance Modeling.** Model a sequence based on multiple templates and bound to a ligand.

This exercise introduces the use of multiple templates and ligands in the process of model building with MODELLER.
- 4. Iterative Modeling.** Increase the accuracy of the modeling exercise by iterating its 4 steps process.

This exercise introduces the concept of MOULDING to improve the accuracy of comparative models.

# Acknowledgments

## Protein Structure Modeling

**Andrej Sali**

**Bino John**

**Narayanan Eswar**

**Ursula Pieper**

Roberto Sánchez (MSSM)

András Fiser (AECOM)

Francisco Melo (CU, Chile)

Azat Badretdinov (Accelrys)

M. S. Madhusudhan

Ash Stuart

Nebojša Mirkovic

Valentin Ilyin (NE)

Eric Feyfant (GI)

Min-Yi Shen

Ben Webb

Rachel Karchin

Mark Peterson

<http://salilab.org>

## 1D to 3D for biologists

**David Huassler (UCSC)**

**Jim Kent (UCSC)**

**Daryl Thomas (UCSC)**

**Mark (UCSC)**

**Rolf Apweiler (EBI)**

## Yeast NPC

Tari Suprpto (RU)

Julia Kipper (RU)

Wenzhu Zhang (RU)

Liesbeth Veenhoff (RU)

Sveta Dokudovskaya (RU)

**J. Zhou (USC)**

**Mike Rout (RU)**

**Brian Chait (RU)**

## Chimera

**P. Babbitt**

**T. Ferrin**

## Ribosomes

**J. Frank**

## Brain Lipid Binding Protein

Liang Zhu (RU)

**Nat Heintz (RU)**

## Structural Genomics

**Stephen Burley (SGX)**

**John Kuriyan (UCB)**

NY-SGXRC

## BRCA1

**A. Monteiro (Cornel)**

## Fly p53

Shengkan Jin (RU)

**Arnie Levine (RU)**

## Mast Cell Proteases

**Rick Stevens (BWH)**

NIH

NSF

Sinsheimer Foundation

A. P. Sloan Foundation

Burroughs-Wellcome Fund

Merck Genome Res. Inst.

Mathers Foundation

I.T. Hirschl Foundation

The Sandler Family Foundation

Human Frontiers Science Program

SUN

IBM

Intel

Structural Genomix