

Protein structure prediction



INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY **DELHI**

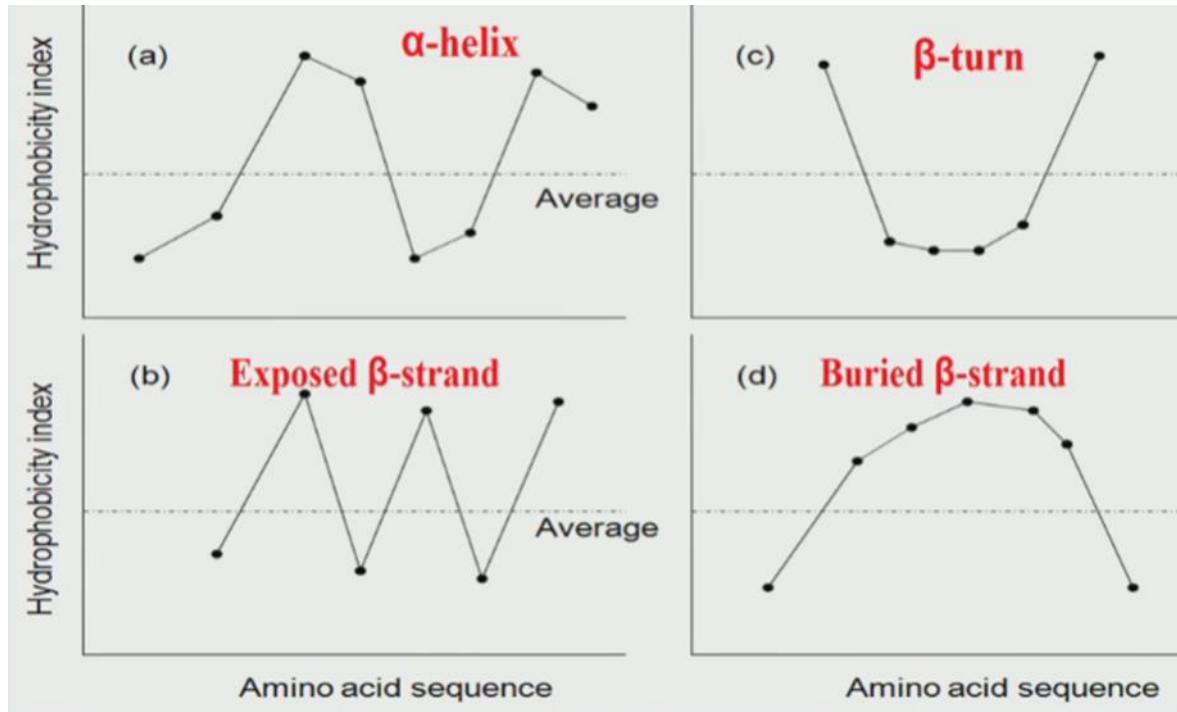
Dr. Jaspreet Kaur Dhanjal

Assistant Professor, Department of Computational Biology

Email ID: jaspreet@iiitd.ac.in

October 07, 2025

Hydrophobicity Profile



Multiple sequence alignment

- Average of helix, strand, coil and turn parameters (GOR) for all the aligned residues at each position
- Insertion is given a value of zero
- Confidence of the prediction is related to the conservation score (0-1) at the position
- Zpred – First SS predictor using MSA

Jpred 4
Incorporating Jnet

A Protein Secondary Structure Prediction Server

[Home](#) [BEST API](#) [About](#) [News](#) [FAQ](#) [Help & Tutorial](#) [Monitoring](#) [Contact](#) [Publications](#)

IMPORTANT MESSAGE: Do you use JPred and/or our other resources?

We are applying to renew funding for the next 5 years, so please help keep the services available for your use by writing us a support letter.

THANK YOU to those who have already written, but it is not too late for others to help us! Please write us a support letter by **Friday Night (22nd September)**

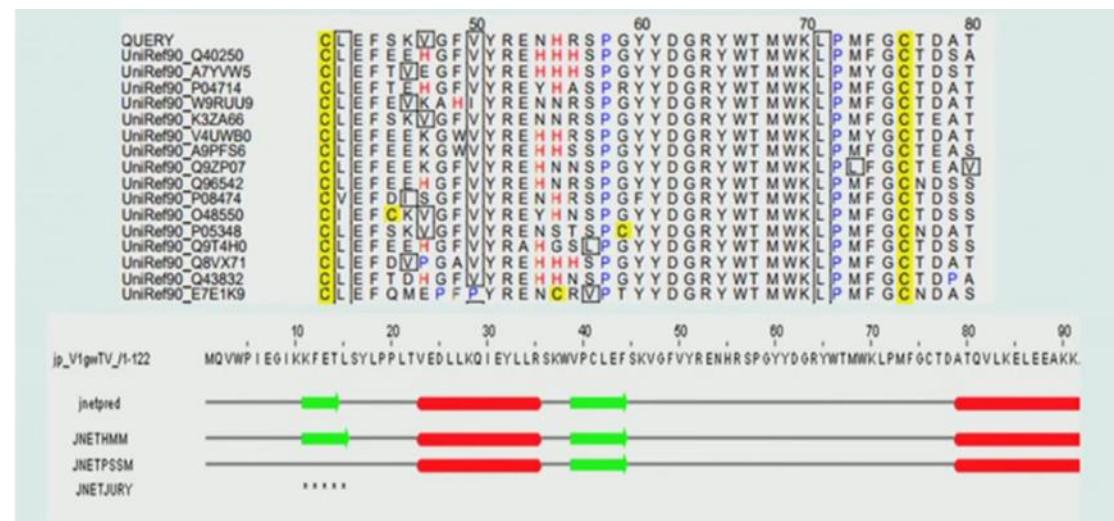
In your letter please say how you use our services and how important they are to you in your research/teaching. Please send the letter as a PDF, ideally on headed paper to: support_jpred@univengroup.org.

Thank you in advance for your help
Geoff Barton

Input sequence: IGVWPGKAFETLIVLPPVYDQGLAKEYLURSIVKPCDFKIGPVYRENHRSPGYYDGRYWTMVKLPMFGCTDATQLKELEAKK.
ADGKLEKELEAKKAPAVIPIQDFKIGPVYRENHRSPGYYDGRYWTMVKLPMFGCTDATQLKELEAKK

[Advanced options \(click to show/hide\)](#)

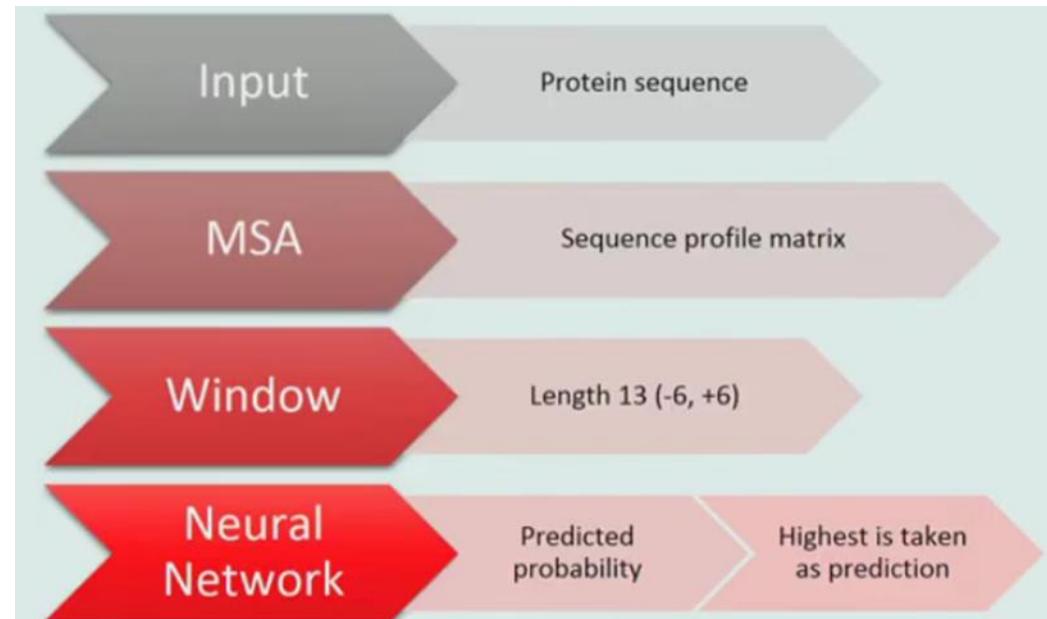
[Make Prediction](#) [Reset Form](#)



Machine learning techniques

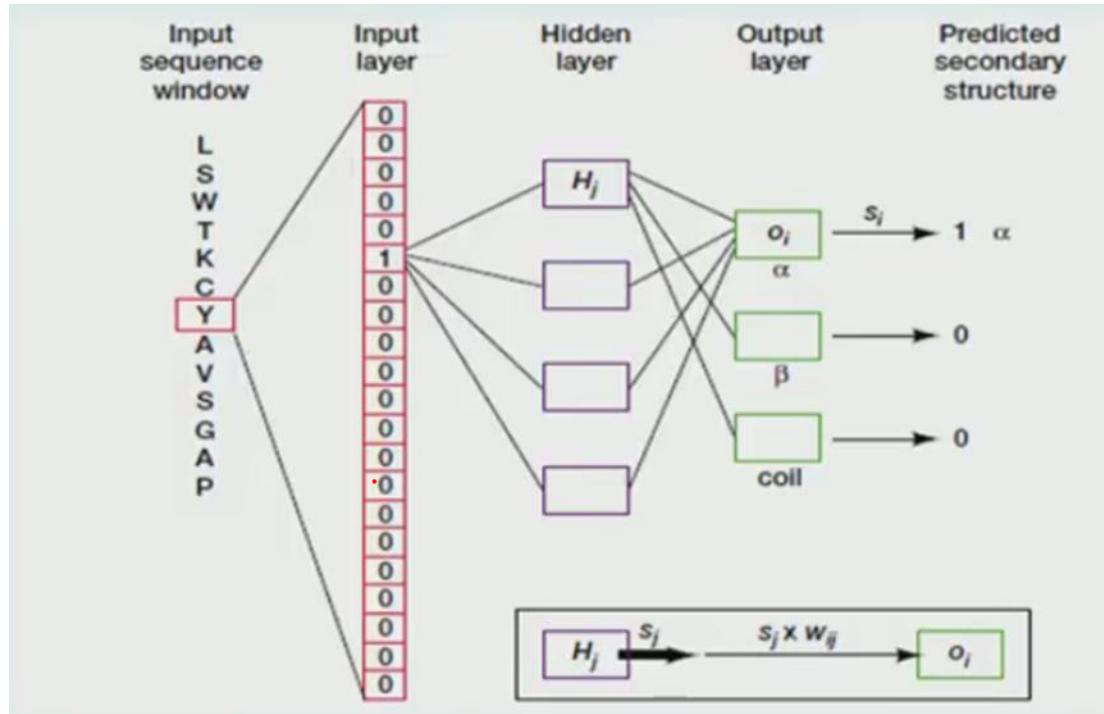
Profile neural network systems from Heidelberg (PHD)

- Aspects:
 - Multiple sequence alignment
→ Evolutionary information
 - Neural networks →
Learning secondary structure
 - amino acid residue pattern/information
 - Window approximation →
Captures effect of neighboring residues



Machine learning techniques

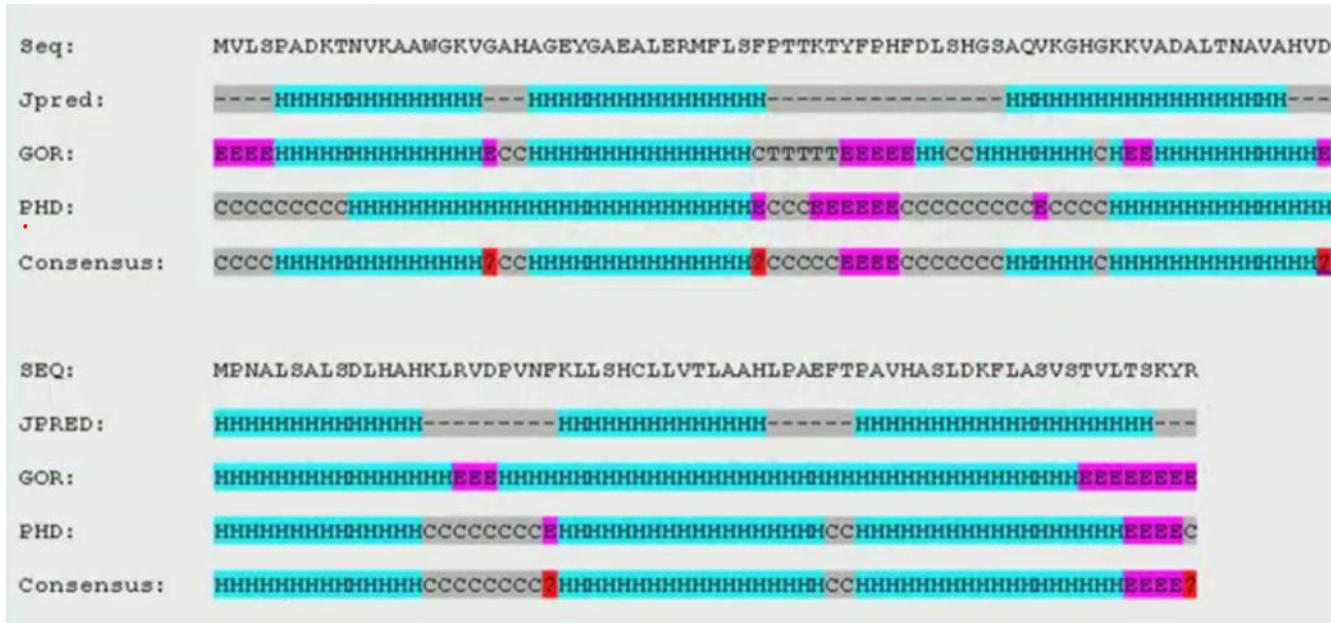
Profile neural network systems from Heidelberg (PHD)



Consensus (joint) prediction methods

1. Consensus

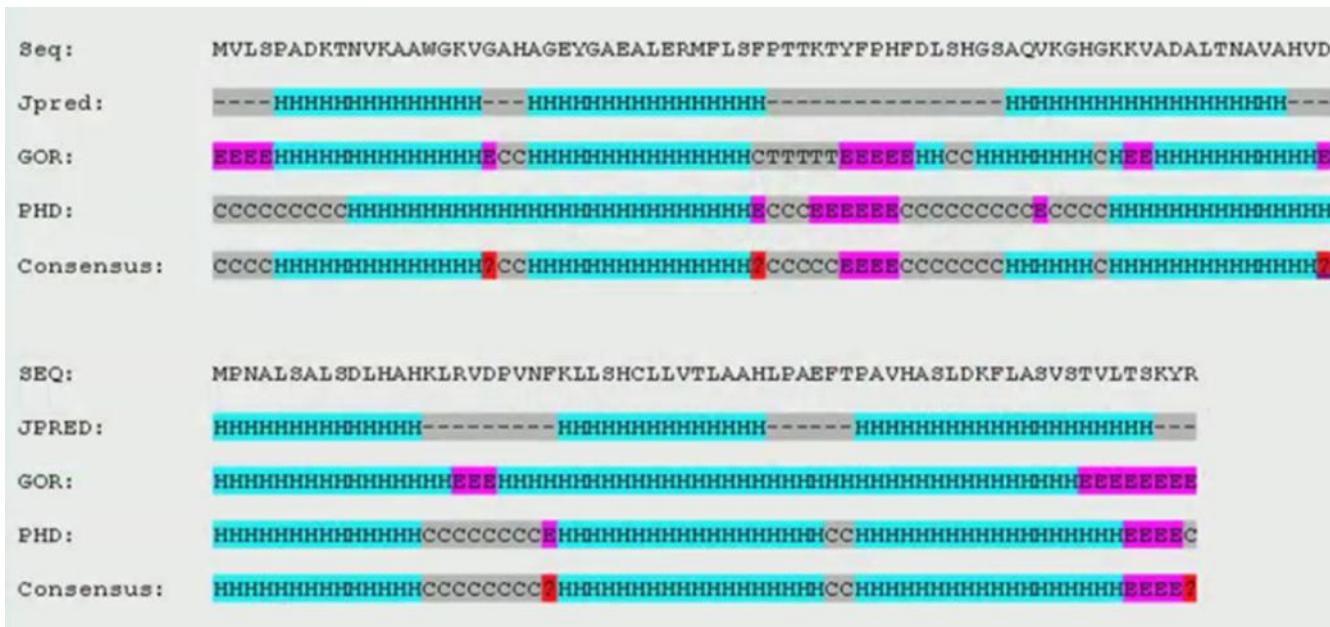
- Input: Prediction results from other methods
 - Strategy: Majority of voting
 - Output: Based on high vote



Consensus (joint) prediction methods

1. Meta prediction

- Input: Prediction results from other methods
 - Strategy: Train on machine learning methods
 - Output: Improved outcome



Summary of secondary structure prediction methods

Methods:

1. Statistical analysis

(Preference of residues, by Chou and Fasman in 1974)

2. Information theory (GOR method, by Garnier, Osguthorpe, and Robson in 1978)

3. Hydrophobicity Profile

4. Multiple sequence alignment

5. Machine learning techniques

(Neural networks, support vector machines, etc.)

6. Consensus (Joint)

Knowledge based prediction of protein tertiary structure

Homology Modeling

- Need homologues of known protein structure
- Backbone modelling
- Side chain modelling
- Fails in absence of similarity with other proteins

Threading Based Methods

- New way of fold recognition
- Sequence is tried to fit in known structures
- Motif recognition
- Loop & Side chain modelling
- Fails in absence of known example

Homology modeling of proteins

- Structure of a protein is uniquely determined by its amino acid sequence. Therefore, knowing the sequence should, at least in theory, suffice to obtain the structure.
- During evolution, a structure is more stable and changes at a slower rate as compared to its sequence, so that similar sequences adopt practically identical structures, and distantly related sequences still fold into similar structures.
- If the percentage of similarity or homology is 40%, two sequences are practically guaranteed to adopt a similar structure.

Step 1- Finding the best template

<i>Homo sapiens</i>	GWTPGTLWGVGDSAGNSSELGVFQGPDLCCREHDRCPQNISPLQNYGI	51	
<i>Bos taurus</i>	GTVPGTLWGVGDSAGNSSELGVFQGPDLCCREHDRCPHNVSPFQYNYGI	51	
<i>Mus musculus</i>	GWTIPGTLWGVGNSAENASELGVFHGPDLCCREHDQCPQTISPLQNYGI	51	
<i>Heloderma suspectum</i>	AFIMPGLWGAGNAASDYSQLGTEKDTDMCCRDHGENWISALEYKHGM	51	
<i>Apis mellifera</i>	-IIYPGTLWGHGNKSSGPNELGRFKHTDACCRTHDMDPDVMSAGESKHGL	50	
<i>Homo sapiens</i>	RNYRFHTISHCDCDTRFQQCLQNQHDSIS-DIVGVAFFNVLEIPCFVLEE	100	
<i>Bos taurus</i>	RNYRFHTISHNCDAFRFQQCLQDQRDSVS-DIMGVAFFNVLAIPCFVLEE	100	
<i>Mus musculus</i>	RNFRFHTISHCDCDVRFQQCLRSQGDSIS-DIMGVAFFNVLEIPCFVLEE	100	
<i>Heloderma suspectum</i>	RNYYPSTISHCDCDNQFRSCLMKLKDGTA-DYVGQTYFNVLKIPCFELEE	100	
<i>Apis mellifera</i>	TNTASHTRLSCDCDKFYDCLKNSADTISSYFVGKMYFNLIDTKCYKLEH	100	
	Identity	Acc. No.	
<i>Homo sapiens</i>	QEACVAWYWUWGGCRMYGTVPLARLQPRTFYNASUSSRATSPT-	142 -- -	Q9NZ20
<i>Bos taurus</i>	QEACVEWYWUWGGCRRYGSVPFARLQPRTFYNASUSSPATSLT-	142 88%	Q1JPB9
<i>Mus musculus</i>	QEACVAUNUWGGCRAYGSTPLAHLRPRIIYNASUWKAEATSYT-	142 81%	Q6AXHO
<i>Heloderma suspectum</i>	GEGCVDUNFULECTESKIMPVAKLVSAAQYQAQAEQSGEGR.	141 50%	P16354
<i>Apis mellifera</i>	PTVGCGERTEGRCLHYTVDKSKPKVYQWFDLRKY.....	135 44%	P00630

Step 1- Finding the best template

BLAST® » blastp suite » results for RID-5V06HD4V013

Job Title	O15393:RecName: Full=Transmembrane protease...	Filter Results																																																																																																																																										
RID	Descriptions	Graphic Summary	Alignments	Taxonomy																																																																																																																																								
Sequences producing significant alignments											Download	New Select columns	Show 100	?																																																																																																																														
<input checked="" type="checkbox"/> select all 100 sequences selected											GenPept	Graphics	Distance tree of results	Multiple alignment	New MSA Viewer																																																																																																																													
<table><thead><tr><th></th><th>Description</th><th>Scientific Name</th><th>Max Score</th><th>Total Score</th><th>Query Cover</th><th>E value</th><th>Per. Ident</th><th>Acc. Len</th><th>Accession</th></tr></thead><tbody><tr><td><input checked="" type="checkbox"/></td><td>human plasmakallikrein protease domain in complex with active site directed inhibitor [Homo sapiens]</td><td>Homo sapiens</td><td>213</td><td>213</td><td>47%</td><td>7e-66</td><td>42.92%</td><td>242</td><td>6T7P_A</td></tr><tr><td><input checked="" type="checkbox"/></td><td>Crystal structure of extracellular domain of MSPL in complex with furin inhibitor [Homo sapiens]</td><td>Homo sapiens</td><td>214</td><td>214</td><td>47%</td><td>8e-66</td><td>45.38%</td><td>261</td><td>6KD5_B</td></tr><tr><td><input checked="" type="checkbox"/></td><td>Structure of human plasma kallikrein [Homo sapiens]</td><td>Homo sapiens</td><td>211</td><td>211</td><td>48%</td><td>1e-64</td><td>42.56%</td><td>263</td><td>5TJX_A</td></tr><tr><td><input checked="" type="checkbox"/></td><td>The crystal structure of human Plasma Kallikrein in complex with its peptide inhibitor pkalin-2 [Homo sapiens]</td><td>Homo sapiens</td><td>209</td><td>209</td><td>47%</td><td>2e-64</td><td>42.50%</td><td>239</td><td>5F8T_A</td></tr><tr><td><input checked="" type="checkbox"/></td><td>Expression, crystallization and three-dimensional structure of the catalytic domain of human plasma kallikrein: Implic...</td><td>Homo sapiens</td><td>209</td><td>209</td><td>47%</td><td>2e-64</td><td>42.50%</td><td>241</td><td>2ANW_A</td></tr><tr><td><input checked="" type="checkbox"/></td><td>Expression, Crystallization and the Three-dimensional Structure of the Catalytic Domain of Human Plasma Kallikrein:...</td><td>Homo sapiens</td><td>209</td><td>209</td><td>47%</td><td>4e-64</td><td>42.50%</td><td>241</td><td>2ANY_A</td></tr><tr><td><input checked="" type="checkbox"/></td><td>Crystal structure of the extracellular region of the transmembrane serine protease hepsin with covalently bound prefe...</td><td>Homo sapiens</td><td>211</td><td>211</td><td>67%</td><td>2e-63</td><td>34.10%</td><td>372</td><td>1Z8G_A</td></tr><tr><td><input checked="" type="checkbox"/></td><td>Human hepsin protease in complex with the Fab fragment of an inhibitory antibody [Homo sapiens]</td><td>Homo sapiens</td><td>211</td><td>211</td><td>67%</td><td>2e-63</td><td>34.10%</td><td>372</td><td>3T2N_A</td></tr><tr><td><input checked="" type="checkbox"/></td><td>Full length human plasma kallikrein with inhibitor [Homo sapiens]</td><td>Homo sapiens</td><td>217</td><td>217</td><td>51%</td><td>6e-63</td><td>41.92%</td><td>638</td><td>6O1G_A</td></tr><tr><td><input checked="" type="checkbox"/></td><td>The Serine Protease Domain Of Enteropeptidase Bound To Inhibitor Val-Asp-asp-asp-lys-chloromethane [Bos t...</td><td>Bos taurus</td><td>202</td><td>202</td><td>46%</td><td>1e-61</td><td>42.31%</td><td>235</td><td>1EKB_B</td></tr><tr><td><input checked="" type="checkbox"/></td><td>Dissecting and Designing Inhibitor Selectivity Determinants at the S1 site Using an Artificial Ala190 Protease (Ala190...</td><td>Homo sapiens</td><td>202</td><td>202</td><td>47%</td><td>1e-61</td><td>40.98%</td><td>255</td><td>1O5E_H</td></tr><tr><td><input checked="" type="checkbox"/></td><td>Crystal Structure of the Catalytic Domain of Coagulation Factor Xla in Complex with Double Bridged Peptide F19 [Ho...</td><td>Homo sapiens</td><td>202</td><td>202</td><td>47%</td><td>2e-61</td><td>42.21%</td><td>262</td><td>6TWB_A</td></tr></tbody></table>												Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession	<input checked="" type="checkbox"/>	human plasmakallikrein protease domain in complex with active site directed inhibitor [Homo sapiens]	Homo sapiens	213	213	47%	7e-66	42.92%	242	6T7P_A	<input checked="" type="checkbox"/>	Crystal structure of extracellular domain of MSPL in complex with furin inhibitor [Homo sapiens]	Homo sapiens	214	214	47%	8e-66	45.38%	261	6KD5_B	<input checked="" type="checkbox"/>	Structure of human plasma kallikrein [Homo sapiens]	Homo sapiens	211	211	48%	1e-64	42.56%	263	5TJX_A	<input checked="" type="checkbox"/>	The crystal structure of human Plasma Kallikrein in complex with its peptide inhibitor pkalin-2 [Homo sapiens]	Homo sapiens	209	209	47%	2e-64	42.50%	239	5F8T_A	<input checked="" type="checkbox"/>	Expression, crystallization and three-dimensional structure of the catalytic domain of human plasma kallikrein: Implic...	Homo sapiens	209	209	47%	2e-64	42.50%	241	2ANW_A	<input checked="" type="checkbox"/>	Expression, Crystallization and the Three-dimensional Structure of the Catalytic Domain of Human Plasma Kallikrein:...	Homo sapiens	209	209	47%	4e-64	42.50%	241	2ANY_A	<input checked="" type="checkbox"/>	Crystal structure of the extracellular region of the transmembrane serine protease hepsin with covalently bound prefe...	Homo sapiens	211	211	67%	2e-63	34.10%	372	1Z8G_A	<input checked="" type="checkbox"/>	Human hepsin protease in complex with the Fab fragment of an inhibitory antibody [Homo sapiens]	Homo sapiens	211	211	67%	2e-63	34.10%	372	3T2N_A	<input checked="" type="checkbox"/>	Full length human plasma kallikrein with inhibitor [Homo sapiens]	Homo sapiens	217	217	51%	6e-63	41.92%	638	6O1G_A	<input checked="" type="checkbox"/>	The Serine Protease Domain Of Enteropeptidase Bound To Inhibitor Val-Asp-asp-asp-lys-chloromethane [Bos t...	Bos taurus	202	202	46%	1e-61	42.31%	235	1EKB_B	<input checked="" type="checkbox"/>	Dissecting and Designing Inhibitor Selectivity Determinants at the S1 site Using an Artificial Ala190 Protease (Ala190...	Homo sapiens	202	202	47%	1e-61	40.98%	255	1O5E_H	<input checked="" type="checkbox"/>	Crystal Structure of the Catalytic Domain of Coagulation Factor Xla in Complex with Double Bridged Peptide F19 [Ho...	Homo sapiens	202	202	47%	2e-61	42.21%	262	6TWB_A
	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession																																																																																																																																			
<input checked="" type="checkbox"/>	human plasmakallikrein protease domain in complex with active site directed inhibitor [Homo sapiens]	Homo sapiens	213	213	47%	7e-66	42.92%	242	6T7P_A																																																																																																																																			
<input checked="" type="checkbox"/>	Crystal structure of extracellular domain of MSPL in complex with furin inhibitor [Homo sapiens]	Homo sapiens	214	214	47%	8e-66	45.38%	261	6KD5_B																																																																																																																																			
<input checked="" type="checkbox"/>	Structure of human plasma kallikrein [Homo sapiens]	Homo sapiens	211	211	48%	1e-64	42.56%	263	5TJX_A																																																																																																																																			
<input checked="" type="checkbox"/>	The crystal structure of human Plasma Kallikrein in complex with its peptide inhibitor pkalin-2 [Homo sapiens]	Homo sapiens	209	209	47%	2e-64	42.50%	239	5F8T_A																																																																																																																																			
<input checked="" type="checkbox"/>	Expression, crystallization and three-dimensional structure of the catalytic domain of human plasma kallikrein: Implic...	Homo sapiens	209	209	47%	2e-64	42.50%	241	2ANW_A																																																																																																																																			
<input checked="" type="checkbox"/>	Expression, Crystallization and the Three-dimensional Structure of the Catalytic Domain of Human Plasma Kallikrein:...	Homo sapiens	209	209	47%	4e-64	42.50%	241	2ANY_A																																																																																																																																			
<input checked="" type="checkbox"/>	Crystal structure of the extracellular region of the transmembrane serine protease hepsin with covalently bound prefe...	Homo sapiens	211	211	67%	2e-63	34.10%	372	1Z8G_A																																																																																																																																			
<input checked="" type="checkbox"/>	Human hepsin protease in complex with the Fab fragment of an inhibitory antibody [Homo sapiens]	Homo sapiens	211	211	67%	2e-63	34.10%	372	3T2N_A																																																																																																																																			
<input checked="" type="checkbox"/>	Full length human plasma kallikrein with inhibitor [Homo sapiens]	Homo sapiens	217	217	51%	6e-63	41.92%	638	6O1G_A																																																																																																																																			
<input checked="" type="checkbox"/>	The Serine Protease Domain Of Enteropeptidase Bound To Inhibitor Val-Asp-asp-asp-lys-chloromethane [Bos t...	Bos taurus	202	202	46%	1e-61	42.31%	235	1EKB_B																																																																																																																																			
<input checked="" type="checkbox"/>	Dissecting and Designing Inhibitor Selectivity Determinants at the S1 site Using an Artificial Ala190 Protease (Ala190...	Homo sapiens	202	202	47%	1e-61	40.98%	255	1O5E_H																																																																																																																																			
<input checked="" type="checkbox"/>	Crystal Structure of the Catalytic Domain of Coagulation Factor Xla in Complex with Double Bridged Peptide F19 [Ho...	Homo sapiens	202	202	47%	2e-61	42.21%	262	6TWB_A																																																																																																																																			

human plasmakallikrein protease domain in complex with active site directed inhibitor [Homo sapiens]

Sequence ID: [6T7P_A](#) Length: 242 Number of Matches: 1

Range 1: 1 to 237 [GenPept](#) [Graphics](#)

[▼ Next Match](#) [▲ Previous Match](#)

Score 213 bits(543)	Expect 7e-66	Identities 103/240(43%)	Positives 149/240(62%)	Gaps 8/240(3%)
------------------------	-----------------	----------------------------	---------------------------	-------------------

Query 25 Crystal structure of the extracellular region of the transmembrane serine protease hepsin with covalently bound preferred substrate. [Homo sapiens]

Sbjct 1 Sequence ID: [1Z8G_A](#) Length: 372 Number of Matches: 1

Query 31 Sbjct 59 Range 1: 18 to 360 [GenPept](#) [Graphics](#)

[▼ Next Match](#) [▲ Previous Match](#)

Score 211 bits(537)	Expect 2e-63	Identities 118/346(34%)	Positives 174/346(50%)	Gaps 17/346(4%)
------------------------	-----------------	----------------------------	---------------------------	--------------------

Sbjct 11 Query 158 LQVYSSQRKSWHPCVQDDWNENYGRAACRDMGYKNNFYSSQGIVDDSGSTSFMKLN-TSA 216
L V+ +W +C N +C +MG+ S+ V +G+

Query 43 Sbjct 18 LMVFDKTEGTWRLLCSSRSNRARVAGLSCEEMGFLRALTHSELDVRTAGAAGTSGFFCVD 77

Sbjct 17 Query 217 GNVDIYKKLYHSDA---CSSKAVVSLRCIACGVNLNSSRQSRIVGGESALPGAWPWQVSL 273
G + ++L + C ++ C CG RIVGG G WPWQVSL

Sbjct 78 GRLPHTQRLLEVISVCDCPGRGRFLAACIQDCG--RRKLPVDRIVGGRTDSLGRWPWQVSL 135

Query 274 HVQNVHVCGGSIIITPEWIVTAAHCVEKPLNNPWHWTAFAGILRQSFMFYGAGYQVEKVIS 333
H+CGGS+++ W++TAAH + W FAG + Q+ +G V+ V+

Sbjct 136 RYDGAHLCGGSSLSDGWVL TAAHCFCPERNRVLSRWRVFAGAVAQASP-HGLQLGVQAVVY 194

Query 334 HPNY-----DSKTKNNDIALMKLQKPLTFNDLVKPVCLPNPGMMLQPEQLCWISGWGAT 387
H Y +S+ +NDIAL+ L PL + ++PVCLP G L ++C ++GWG T

Sbjct 195 HGGYLPFRDPNSEENSNDIALVHLSSPLPLTEYIQPVCLPAAGQALVDGKICTVTGWGNT 254

Query 388 EEKGKTSEVLNAAKVLLIETQRCNSRYVYDNLITPAMICAGFLQGNVDSCQGDGGPLV- 446
+ G+ + VL A+V + I CN Y N I P M CAG+ +G +D+CQGDSGGP V

Sbjct 255 QYYGQQAGVLQEARNPIISNDVCNGADFYGNQIKPKMFCAGYPEGGIDACQGDGGPFVC 314

Query 447 ---TSKNNIWWLIGDTSWGSGCAKAYRPGVYGNVMVFTDWIYRQMR 489
S+ W L G SWG+GCA A +PGVY V F +WI++ ++

Sbjct 315 EDSISRTPRWRLCGIVSWGTCALAOKPGVYTKVSDFREWIFQAIK 360

Related Information

[Structure](#) - 3D structure displays

Crystal structure of the extracellular region of the transmembrane serine protease hepsin with covalently bound preferred substrate. [Homo sapiens]

Sequence ID: [1Z8G_A](#) Length: 372 Number of Matches: 1

Range 1: 18 to 360 GenPept Graphics

▼ Next Match ▲ Previous Match

Score	Expect	Identities	Positives	Gaps
211 bits(537)	2e-63	118/346(34%)	174/346(50%)	17/346(4%)

Query 158 LQVYSSQRKSWHPVCQDDWNENYGRAACRDMGYKNNFYSSQGVDDSGSTSFMKLN-TSA 216
L V+ +W +C N +C +MG+ S+ V +G+
Sbjct 18 LMVFDKTEGTWRLLCSSRSNARVAGLSCCEMGLRALTHSELDVRTAGAGTSGFFCVDE 77

Query 217 GNVDIYKKLYHSDA---CSSKAVVSLRCIACGVNLNNSRQSRIVGGESALPGAWPQVSL 273
G + ++L + C ++ C CG RIVGG G WPQVSL
Sbjct 78 GRLPHTQRLLEVISVCDCPGRFLAAICQDCG--RRKLPVDRIVGGRTDSLGRWPQVSL 135

Query 274 HVQNVHVGGSIIITPEWVTAACVEKPLNNPWHNTAFAGILRQSFMFYGAGYQVEKVIS 333
H+CGGS+++ +W++TAAHC + W FAG + Q+ +G V+ V+
Sbjct 136 RYDGAHLCGGSLLSGDWLTAHHCFPERNRVLSRVRVFAGAVAQASP-HGLQLGVQAVVV 194

Query 334 HPNY-----DSKTKNDIALMKLQPLTFNDLVPKVCLPNPGMMLQPEQLCWISGWGAT 387
H Y +S+ +NDIAL+ L PL + ++PVCLP G L ++C ++GWG T
Sbjct 195 HGGYLPFRDPNSENNSNDIALVHLSSPLPLTEYIQPVCLPAAGQALVDGKICTVTGWGNT 254

Query 388 EEKGKTSEVLNAAKVLLIETQRCSNSRYVYDNLTTPAMICAGFLQGNVDSQCGDGGPLV- 446
+ G+ + VL A+V +I CN Y N I P M CAG+ +G +D+CQGDSGGP V
Sbjct 255 QYYQQAGVLQEAVPIIISNDVNCNGADFYGNQIKPKMFCAKYPEGGIDAQCGDGGPFVC 314

Related Information

[Structure](#) - 3D structure displays

Human hepsin protease in complex with the Fab fragment of an inhibitory antibody [Homo sapiens]

Sequence ID: [3T2N_A](#) Length: 372 Number of Matches: 1

[See 2 more title\(s\)](#) ▼ [See all Identical Proteins\(IPG\)](#)

Range 1: 18 to 360 GenPept Graphics

▼ Next Match ▲ Previous Match

Score	Expect	Identities	Positives	Gaps
211 bits(537)	2e-63	118/346(34%)	174/346(50%)	17/346(4%)

Query 158 LQVYSSQRKSWHPVCQDDWNENYGRAACRDMGYKNNFYSSQGVDDSGSTSFMKLN-TSA 216
L V+ +W +C N +C +MG+ S+ V +G+
Sbjct 18 LMVFDKTEGTWRLLCSSRSNARVAGLSCCEMGLRALTHSELDVRTAGAGTSGFFCVDE 77

Query 217 GNVDIYKKLYHSDA---CSSKAVVSLRCIACGVNLNNSRQSRIVGGESALPGAWPQVSL 273
G + ++L + C ++ C CG RIVGG G WPQVSL
Sbjct 78 GRLPHTQRLLEVISVCDCPGRFLAAICQDCG--RRKLPVDRIVGGRTDSLGRWPQVSL 135

Query 274 HVQNVHVGGSIIITPEWVTAACVEKPLNNPWHNTAFAGILRQSFMFYGAGYQVEKVIS 333
H+CGGS+++ +W++TAAHC + W FAG + Q+ +G V+ V+
Sbjct 136 RYDGAHLCGGSLLSGDWLTAHHCFPERNRVLSRVRVFAGAVAQASP-HGLQLGVQAVVV 194

Query 334 HPNY-----DSKTKNDIALMKLQPLTFNDLVPKVCLPNPGMMLQPEQLCWISGWGAT 387
H Y +S+ +NDIAL+ L PL + ++PVCLP G L ++C ++GWG T
Sbjct 195 HGGYLPFRDPNSENNSNDIALVHLSSPLPLTEYIQPVCLPAAGQALVDGKICTVTGWGNT 254

Query 388 EEKGKTSEVLNAAKVLLIETQRCSNSRYVYDNLTTPAMICAGFLQGNVDSQCGDGGPLV- 446
+ G+ + VL A+V +I CN Y N I P M CAG+ +G +D+CQGDSGGP V
Sbjct 255 QYYQQAGVLQEAVPIIISNDVNCNGADFYGNQIKPKMFCAKYPEGGIDAQCGDGGPFVC 314

Query 447 ---TSKNNIWWLIGDTSGSCAKAYRPGVYGNVMFTDWIYQMR 489
S+ W L G SWG+GCA A +PGVY V F +W+ +
Sbjct 315 EDSISRTPRWRRLCGIVSNGTGCALAQPKGVYTKVSDFREWFQAIK 360

Related Information

[Structure](#) - 3D structure displays

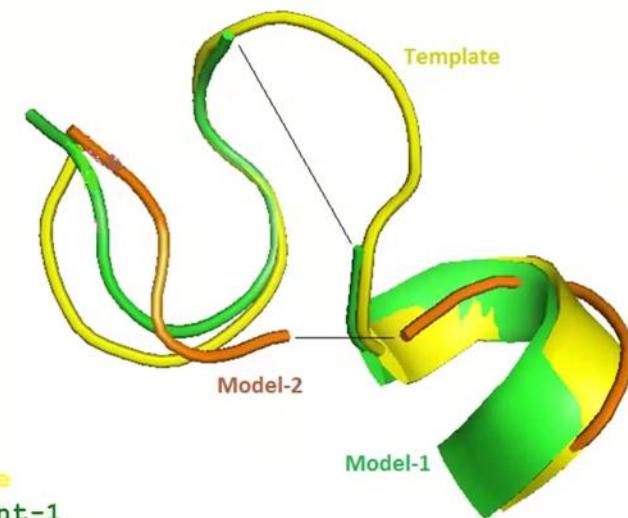
[Identical Proteins](#) - Identical

proteins to 3T2N_A

Step2- Correct sequence alignment

- Sometimes it may be difficult to align two sequences in a region where the percentage sequence identity is very low.
- A simple sequence alignment gives the higher score based on matching/pairing of residues on the two sequences

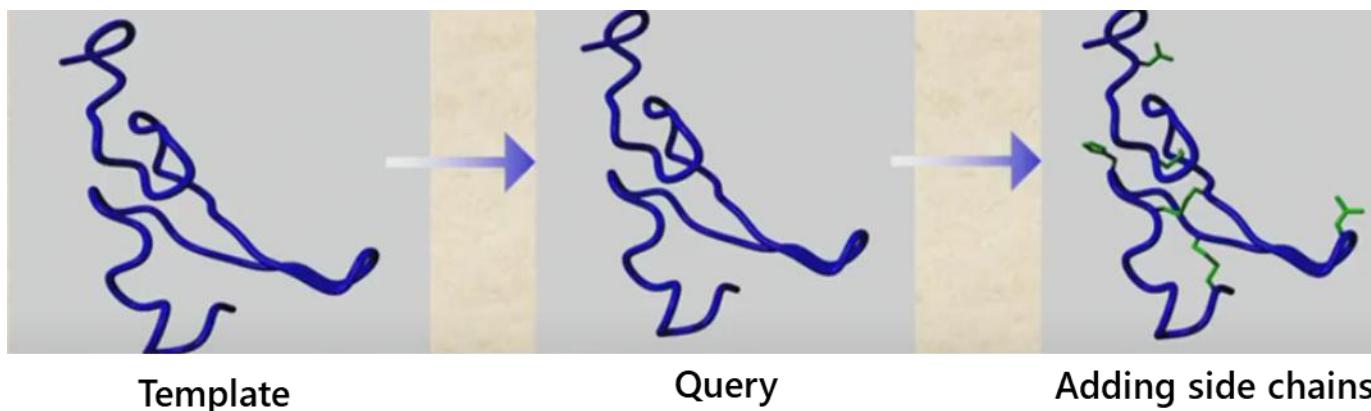
1	2	3	4	5	6	7	8	9	10	11	12	13	14
ALA	SER	ILE	CYS	ARG	LEU	PRO	GLY	SER	ALA	GLU	GLY	VAL	CYS
Template													
ALA	ASN	VAL	CYS	ARG	THR	PRO	---	---	---	GLU	GLY	ILE	CYS
Alignment-1													
ALA	ASN	VAL	CYS	ARG	---	---	---	THR	PRO	GLU	GLY	ILE	CYS
Alignment-2													



- Model-2 for alignment-2 is correct, because it leads to a small gap and can be corrected by small backbone shifts.

Step3- Generating the backbone

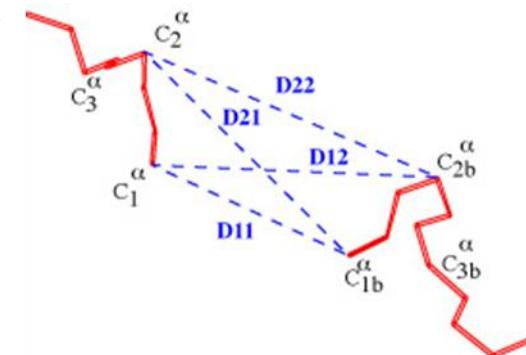
- Generate backbone coordinates from the template for the aligned regions.
- If most of aligned residues are identical, back bone too will be identical
- Template that generates a back bone with least errors is chosen.
- Back bone of secondary structural elements like helices and beta sheets are usually maintained, while variations may be seen in the loops and turns,



Step4- Loop modeling

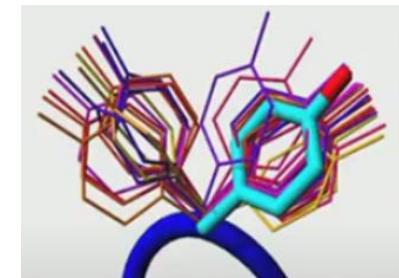
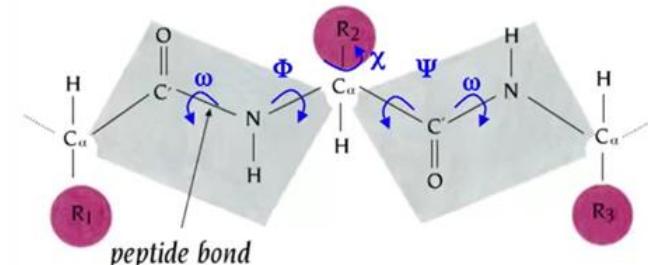
- Region of query sequence that is not aligned to template is modeled as a loop.
- Loop modeling is done by knowledge-based method where PDB is searched for known loops, or by Energy based method where long chains are built by sampling Ramachandran conformations randomly.
- For deletions, residues from the template is omitted, creating a void in the model that is filled by bringing together the adjacent residues' coordinates.
- For insertions, the coordinates for the missing residues is given a loop conformation in most cases and simulated for it to take up a secondary structure in accordance with the adjacent structures in the template.

Selecting protein fragments for loop modelling: In the example shown here, the gap in the framework is characterized by 4 distances (D11, D12, D21 and D22) between the 2 C α preceding the gap and the 2 C α following the gaps. The library of protein structure fragments of appropriate size (i.e. Ngap + 4, where Ngap is the number of residues in the gap and 4 corresponds to the 2 anchoring residues on each side of the gap) is scanned and the fragments with the proper geometry (i.e. such that the distances between the 2 first and last residues match the 4 characteristic distances of the gap) are retained has potential loops to fill in the gap.



Step5- Side chain modeling

- Involves placing side chains coordinates either obtained from a template structure or generated from *ab initio* modeling simulations or both.
- Side chain remains more flexible on the surface and tend to adopt multiple conformations as compared to the core where residues have restricted environment and limited conformational liberty.
- At high levels of sequence identity χ_1 -angles adopt positions from template structure and for sequence identity <35% combinatorial approach using rotamer libraries are used.
- Wrong backbone will affect the side chain building process.
- Prediction accuracy is highly dependent on scoring functions.



TYR

Step6- Model optimization and structure refinement

- Model optimization tries to bring it closer to real in-vivo conformation.
- Commonly achieved by Molecular Dynamics Simulation.
- Simulation follows the motions of the protein on a femto-second timescale and mimics true folding process.
- This helps to get a true structural model having minimum energy configuration

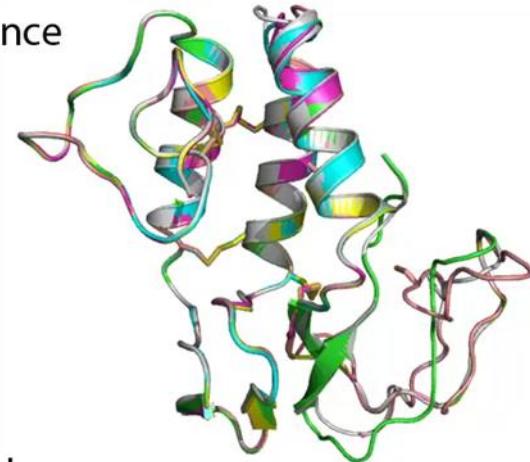
Step6- Validation of the developed model

Root mean square deviation (RMSD)

- Root mean square deviation is the measure of the average distance between the atoms of two superimposed proteins.

$$RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^{i=N} \delta_i^2}$$

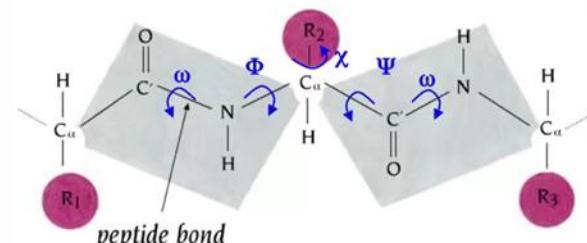
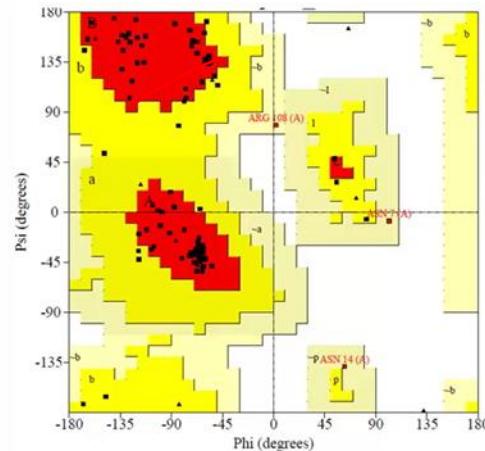
- δ is the distance between number of pairs of equivalent atoms
- 0.0-0.5 Å : Essentially Identical; <1.5 Å: Very good fit; < 5.0 Å: Moderately > 7.0 Å: Dubious; > 12.0 Å: Completely unrelated



Step6- Validation of the developed model

Ramachandran Plot

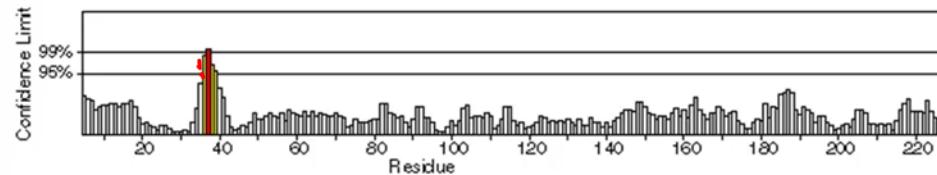
- Atoms are treated as hard spheres with dimensions corresponding to their van der Walls radii
- Residues with phi and psi angles which cause atoms to collide are considered to be in sterically disallowed conformations
- Most allowed region occupied by residues with no steric clashes
- Dis-allowed region with residues whose atoms come closer than their van der Wall radii.
- Additionally allowed region with residues have atoms with inter-atomic distance slightly shorter than van der Walls radii.



Step6- Validation of the developed model

ERRAT Plot

- Of three heavy atoms carbon, nitrogen, and oxygen, there exist six different combinations of pairwise non-covalently bonded interactions (CC, CN, CO, NN, NO, and OO).
- In a database of 96 reliable protein structures, these set of pairwise interactions are characterized; ERRAT finds regions of model structures that have incorrect pattern of non-bonded interactions that result in incorrectly folded regions

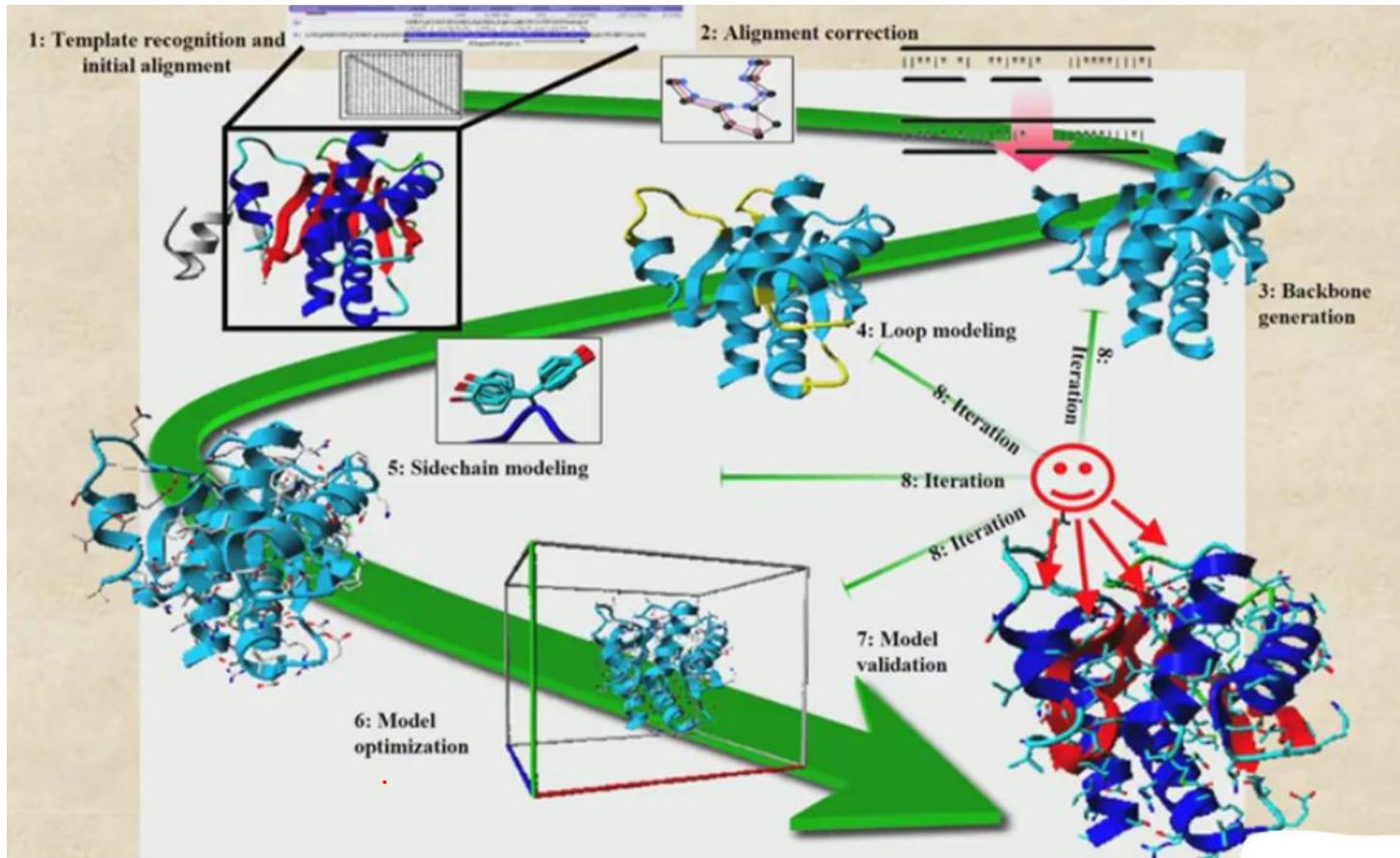


Regions of the structure that can be rejected at the 95% confidence level are shown in yellow and regions that can be rejected at the 99% level are shown in red.

Step6- Validation of the developed model

- WHAT IF <http://www.cmbi.kun.nl/gv/servers/WIWWWI/>
- SOV <http://predictioncenter.llnl.gov/local/sov/sov.html>
- PROVE <http://www.ucmb.ulb.ac.be/UCMB/PROVE/>
- ANOLEA <http://www.fundp.ac.be/pub/ANOLEA.html>
- ERRAT <http://www.doe-mbi.ucla.edu/Services/ERRATv2/>
- VERIFY3D http://shannon.mbi.ucla.edu/DOE/Services/Verify_3D/
- BIOTECH <http://biotech.embl-ebi.ac.uk:8400/>
- ProsaII <http://www.came.sbg.ac.at>
- WHATCHECK <http://www.sander.embl-heidelberg.de/whatcheck/>

Revisiting the protocol for homology modeling



Advantages and disadvantages of homology modeling

Advantages

- Requires amino acid sequence and a 3D structure of a similar protein
- Does not involve any bench work or consumables
- Less time consuming and less costly
- Most of the softwares are freely available online
- Most promising method among the three computational methods

Disadvantages

- Loop regions are difficult to model
- Experimentally derived structure for homologous protein is essential
- Does not explain aspects of protein folding.

Programs available for homology modeling of proteins

- SWISS Model : <http://www.expasy.org/swissmod/SWISS-MODEL.html>
- WHAT IF : <http://www.cmbi.kun.nl/swift/servers/>
- The CPHModels Server : <http://www.cbs.dtu.dk/services/CPHmodels/>
- 3D Jigsaw : <http://www.bmm.icnet.uk/~3djigsaw/>
- SDSC1 : <http://cl.sdsc.edu/hm.html>
- EsyPred3D :<http://www.fundp.ac.be/urbm/bioinfo/esypred/>

- COMPOSER
<http://www.tripos.com/sciTech/inSilicoDisc/bioInformatics/matchmaker.html>
- MODELER <http://salilab.org/modeler>
- InsightII <http://www.msi.com/>
- SYBYL <http://www.tripos.com/>