ChiRotor and Looper for Side-Chain and Loop Optimization

Deepak Singh, Tai-Sung Lee, Tina Yeh, Velin Z. Spassov, Lisa Yan, and Dana Haley-Vicente
Accelrys Inc., 10188 Telesis Court, Suite 100, San Diego, CA 92121, USA

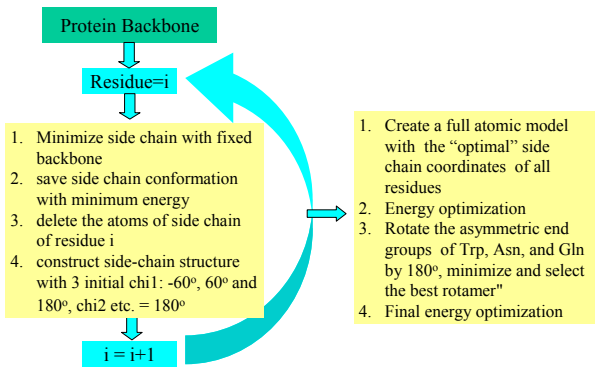
Abstract

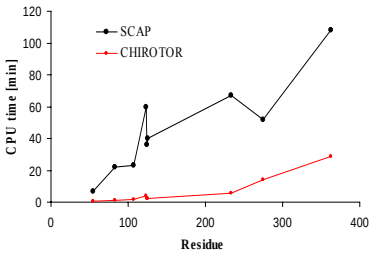
CASP6 target homology models were predicted using a suite of tools available in Discovery Studio® (DS) Modeling and Insight II® modeling and simulations packages (Accelrys, Inc). Some models have been further optimized by two new methods developed at Accelrys, ChiRotor and Looper, for side-chain and loop optimization, respectively. ChiRotor is a fast conformational search algorithm that combines rotamer searches with an energy evaluation to calculate optimal side-chain conformations for all or part of a protein with an average RMSD of ~1Å for the core residues. Looper is a fast algorithm that performs a hierarchical search of low energy loop structures to provide a ranked list of loop fragments with a high level of accuracy.

ChiRotor

ChiRotor is a fast conformational search algorithm that combines discrete rotamer searches with an energy evaluation implemented as a set of CHARMm scripts. The program calculates optimal side-chain conformations for all or part of a protein with an average RMSD of ~1Å for the core residues. All backbone atoms and other protein atoms with "known" coordinates are fixed and the scoring function is based on CHARMm force fields, either charmm19 or CHARMm (Momany & Rone) with an optional electrostatic term. In this poster, we have compared the ChiRotor results (RMSD values) to a program called SCAP (Xiang, Z., Honig, B., *J. Mol. Biol.*, 311:421-430 2001), and we demonstrate the usage of ChiRotor optimization on two CASP6 models for targets 231 and 246 produced by the MODELER (Accelrys Inc.) program.

ChiRotor Program





CPU time used to predict the side-chain conformation of all residues in proteins of different chain lengths. The calculations were carried out on an SGI Indigo2 computer (R10K 175Mhz).

The RMSD values of predicted side-chain atomic coordinates for a set of 24 proteins using ChiRotor in fast (fst) and slow (slw) mode. All X-ray structures have a resolution less than 1Å and similarity less than 20%.

PDB code	res.	ChiRotor						SCAP fast mode	
		charmm19		MSI polar					
		core fst	all fst	core fst	core slw	all fst	all slw	core	all
1ejg	46	0.54	1.69	0.62	0.69	1.59	1.65	0.26	0.99
1rb9	53	0.37	1.87	0.20	0.51	1.49	1.76	0.51	1.26
2fdn	55	0.38	1.69	0.35	0.34	1.61	1.60	1.66	2.68
1g6x	58	0.88	2.02	1.04	0.30	1.73	1.83	2.00	1.97
1f94	63	0.31	2.2	0.31	0.27	2.23	2.23	1.07	2.03
1aho	64	1.51	1.79	1.09	0.57	2.11	1.71	0.81	1.90
1c75	71	0.13	1.43	0.19	0.23	1.42	1.64	0.60	1.56
1iqz	81	0.29	1.53	0.59	0.25	1.61	1.52	0.45	1.27
1ua	83	1.18	1.89	1.10	0.67	1.56	1.43	0.38	1.47
2pvb	107	1.56	2.00	1.65	1.13	2.05	2.02	0.40	1.26
1dy5	123	1.00	2.12	0.97	0.89	1.98	1.83	0.88	1.90
1g4i	123	0.95	1.61	1.01	0.58	1.53	1.68	0.65	1.93
3pyp	125	1.05	1.5	0.99	0.62	1.57	1.53	1.56	1.71
3lzt	129	1.27	1.87	0.91	0.64	1.83	1.93	0.89	1.63
1gb6	207	1.29	1.77	1.72	1.25	1.80	1.52	0.65	1.34
1byi	224	1.81	2.18	1.66	1.57	2.09	2.11	0.90	1.67
1fn8	224	0.68	1.54	1.21	0.60	1.54	1.33	0.62	1.13
1k4i	233	1.16	2.20	1.33	1.06	2.14	2.28	1.00	1.90
1nls	237	1.28	1.88	1.47	1.42	2.22	1.90	0.75	1.55
1gci	269	1.05	1.92	0.88	0.80	1.50	1.61	0.78	1.38
7a3h	303	1.66	2.06	1.17	1.31	1.83	1.79	1.16	1.56
1ixh	321	1.45	1.98	1.39	1.04	1.91	1.58	1.20	1.68
1hxo	323	1.33	1.60	1.19	1.35	1.52	1.65	0.53	1.10
1kwf	363	1.73	1.84	1.47	1.27	1.66	1.50	1.30	1.70
Average		1.03	1.84	1.02	0.80	1.77	1.73	0.88	1.61

Looper

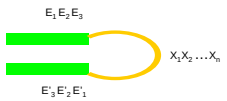
Looper is a program based on a hierarchial search method of low energy loop structures using a minimal representative set of the possible backbone conformers in combination with ChiRotor side-chain prediction and CHARMm energy minimization. Below are the RMSD values of backbone atomic coordinates of predicted structures of loop region for selected proteins. We demonstrate the usage of Looper optimization on two loops within a CASP6 model for Target 246 produced by the MODELER program.

PDB code	loop	length	best	RMSD		rank (energy)		RMSD of the loop with lowest energy		
2apr	76-83	8	0.98	0.62	0.94	3	2	1.65	5.16	1.31
8abp	203-208	6	0.24	0.28	0.24	1	32	0.24	0.28	0.38
2act	98-205	8	1.42	1.47	1.93	8	392	3.00	1.58	2.04
8tln	E32-E38	7	1.37	1.45	0.93	9	314	6.00	3.70	2.03
5cpa	231-237	7	0.76	1.18	1.00	1	244	0.76	2.14	0.95
2fb4	H26-H32	7	2.18	0.36	0.52	26	148	4.39	1.62	4.20
8tln	E248-E255	8	1.61	0.61	0.70	48	246	8.45	1.83	0.87
3sgb	E199-E211	9	0.43	1.12	0.29	1	3	0.43	1.79	0.28
3dfr	20-23	4	0.91	0.44	0.35	2	844	1.01	2.64	1.15
3dfr	89-93	5	0.80	0.93	0.87	4	690	2.90	1.62	1.02
3dfr	20-124	5	0.47	0.33	0.23	2	484	0.55	0.47	0.26
3blm	131-135	5	0.17	0.65	0.16	1	75	0.17	0.82	0.16

Color Code
- Results from LOOPER
- Results from Vlijmen H.W.T, Karplus M(1997) J.Mol.Biol.,267,975-1001.
- Results from MODELER, Fiser A, Do R, Sali A (2000) Protein Sci.,9,1753-1773

Loop Definition

- Loop region $[X_1X_2\ldots X_n]$. Extend the loop to include 3 flexible residues at each end: $E_1E_2E_3X_1X_2\ldots X_nE'_1E'_2E'_3$
- Set the peptide in extended conformation
- Minimize and set the reference energy: E_0 = Epeptide



Test Conformers

- From residue E_3, generate N_c loop conformers by combining maximum four basic Ramachandran states of each X_i
- Gly: **B, R, L** and **P** ; Pro: **B** and **R**
- Other residues:
- Level 0: **B** and **R**
- Level 1: level 0 + one **L** per loop on any non Gly or Pro residue

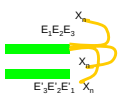
Four basic Ramachandran states:

B: Φ = -120, Ψ = 120° ; R: Φ = -65, Ψ = -40°
L: Φ = 60, Ψ = 30° ; P: Φ = 120, Ψ = -120°

Total number of conformers:

Level 0: $N_c = 2^{NX} * 4^{NGly}$

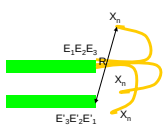
Level 1: $N_c = (2+N_x)\ 2^{(NX-1)} * 4^{NGly}$



Loop Closure

For each initial conformer:

- 20 steps minimization of the loop $X_1\ldots X_N$ region
- Calculate distance R between C atom of X_N and N atom of E'_1.
- If R < Rmax, minimize the peptide with stem residues fixed
- Calculate CHARMm energy E.
- If $E-E_0 < E_{tp}$, accept the structure
- Minimize the accepted conformers in the environment of the rest of protein



Results of Optimizing CASP6 Models

We demonstrate the usage of two effective protocols for loop (Looper) and side chain (ChiRotor) optimization on models produced by the MODELER program. For the CASP6 target model T0231 with high sequence similarity to the template (76%), the MODELER model seems to give very good loop and side chain conformations thus no significant improvement is observed by using Looper and ChiRotor. For the CASP6 target model T0246 with moderate similarity template (58%), we found some improvements in applying both Looper and ChiRotor, more significantly within the loop region, residues 183-192. Compared to other optimization tools, ChiRotor and Looper are faster and more feasible for refining a large number of models. We are currently applying Looper and ChiRotor to models covering a wide range of sequence similarity to clearly understand the applicability of both methods.

ChiRotor

T0231 (76%)	RMSD			Correct prediction of χ_1 (%)*	
	All	Side Chain	Buried	All	Core
MODELER	2.6	3.4	1.5	61	81
ChiRotor	2.6	3.4	2.1	62	71
SCAP	2.6	3.4	1.9	63	66

T0246 (58%)	RMSD			Correct prediction of χ_1 (%)*	
	All	Side Chain	Buried	All	Core
MODELER	2.8	3.8	2.3	58	67
ChiRotor	2.8	3.6	2.3	61	68
SCAP	2.8	3.6	2.3	61	72

*Within 40 degrees.

Looper

RMSD

T0246	RMSD from MODELER		RMSD from Looper	
	Cα only	All	Cα only	All
Region (29-35)	1.24	1.17	1.75	1.64
Region (183-192)	2.20	2.04	1.93	0.98

* in Å

Ramachandran State

T0246	Ramachandran State	
	Region (29-35)	Region (183-192)
X-Ray	RRLBBBB	BBBBBBBLBR
Looper	RRBBLBB	BBBBBBRRBR
Modeler	RRPBRBB	BBBBBBBBBB

Loop region (183-192) modeled by MODELER (green) and LOOPER (red), compared to the X-Ray crystal structure results (blue).