Supplementary Material: α-helical Topology Prediction and Generation of Distance Restraints in Membrane Proteins

Scott R. McAllister and Christodoulos A. Floudas*
Department of Chemical Engineering,
Princeton University,
Princeton, NJ 08544-5263, U.S.A.

August 18, 2008

Bilinear Reformulation: Pairwise Model

For clarity, the variables Z^A_{mn} and Z^P_{mn} are introduced in Equations 1 and 2. Given these definitions, the objective function can be further simplified by the introduction of U^A_{mn} and U^P_{mn} in Equations 3 and 4.

$$Z^A_{mn} = \sum_i \sum_j w^{mn}_{ij} \cdot p^A_{ij} \quad \forall (m,n) \tag{1}$$

$$Z^P_{mn} = \sum_i \sum_j w^{mn}_{ij} \cdot p^P_{ij} \quad \forall (m,n) \tag{2}$$

$$U^A_{mn} = y^A_{mn} \cdot Z^A_{mn} \quad \forall (m,n) \tag{3}$$

$$U^P_{mn} = y^P_{mn} \cdot Z^P_{mn} \quad \forall (m,n) \tag{4}$$

Equations 5 through 8 are equivalent to the definitions of Equations 3 and 4, but are used to linearize the system with an equivalent representation (1).

$$0 \leq U^A_{mn} \leq Z^A_{mn} \quad \forall (m,n) \tag{5}$$

$$Z^A_{mn} - \left(Z^A_{mn}\right)^U \left(1 - y^A_{mn}\right) \leq U^A_{mn} \leq \left(Z^A_{mn}\right)^U \cdot y^A_{mn} \quad \forall (m,n) \tag{6}$$

$$0 \leq U^P_{mn} \leq Z^P_{mn} \quad \forall (m,n) \tag{7}$$

$$Z^P_{mn} - \left(Z^P_{mn}\right)^U \left(1 - y^P_{mn}\right) \leq U^P_{mn} \leq \left(Z^P_{mn}\right)^U \cdot y^P_{mn} \quad \forall (m,n) \tag{8}$$

*Author to whom all correspondence should be addressed; Tel: +1-609-258-4595; Fax: +1-609-258-0211. *E-mail*: floudas@titan.princeton.edu

The addition of the above equations to the model simplifies the objective function to the linear form of Equation 9, allowing the use of mixed-integer linear programming (MILP) solvers.

$$\max \sum_{m} \sum_{n;m<n} U^{A}_{mn} + U^{P}_{mn} \tag{9}$$

Bilinear Reformulation: Triplet Model

The constraints below must also be included in the triplet model for the given definitions of $Z^{A}_{m,n}$ and $Z^{P}_{m,n}$.

$$OBJ = \sum_{m} \sum_{n;m<n} U^{A}_{mn} + U^{P}_{mn} \tag{10}$$

$$Z^{A}_{mn} = \sum_{i} \sum_{j} \sum_{t} w^{mn}_{ijt} \cdot p^{A}_{ijt} \quad \forall (m,n) \tag{11}$$

$$Z^{P}_{mn} = \sum_{i} \sum_{j} \sum_{t} w^{mn}_{ijt} \cdot p^{P}_{ijt} \quad \forall (m,n) \tag{12}$$

$$U^{A}_{mn} \leq Z^{A}_{mn} \quad \forall (m,n) \tag{13}$$

$$U^{A}_{mn} \leq \left(Z^{A}_{mn}\right)^{U} \cdot y^{A}_{mn} \quad \forall (m,n) \tag{14}$$

$$U^{A}_{mn} \geq Z^{A}_{mn} - \left(Z^{A}_{mn}\right)^{U} \cdot \left(1 - y^{A}_{mn}\right) \quad \forall (m,n) \tag{15}$$

$$U^{P}_{mn} \leq Z^{P}_{mn} \quad \forall (m,n) \tag{16}$$

$$U^{P}_{mn} \leq \left(Z^{P}_{mn}\right)^{U} \cdot y^{P}_{mn} \quad \forall (m,n) \tag{17}$$

$$U^{P}_{mn} \geq Z^{P}_{mn} - \left(Z^{P}_{mn}\right)^{U} \cdot \left(1 - y^{P}_{mn}\right) \quad \forall (m,n) \tag{18}$$

1h2sB: A Bundle of 2 Helices

The first membrane protein selected for evaluation, 1h2sB, is a signal transducer protein that consists of 2 transmembrane helices (4). The two α-helices exist in an antiparallel orientation and bind tightly to a receptor membrane protein. The entire transducer protein is composed of a 28 residue α-helix, a 2 residue loop region, and a 30 residue α-helix. The proposed models for interhelical pair and three-body contact prediction are applied for a subtract parameter value of 0, a max_contact parameter range of 1-2, and a total of 20 iterations. These runs are performed for both the probability sets derived from the data set in Section 3.2 and the probabilities based on the work of Liang and co-workers (2, 3). It is notable that for all four probability sets, the best result after 5 iterations remains unchanged after 10 and 20 iterations of the model. The best average contact distances for these runs are displayed in Table 1. One particularly good contact prediction is presented in Table 2 and has an average contact distance value of 8.93 Å over 2 three-body contacts. This contact prediction is illustrated in Figure 1.

Table 1: The best average contact distances of the 1h2sB protein predictions using four probability sets. All distances are in Å.

PAIR MIN-1	PAIR AL-P	TRIP MIN-2	TRIP AL-T
10.45	8.75	8.80	10.50

Table 2: A high-scoring set of contact predictions for 1h2sB using the triplet model and the AL-T probability set

Three-body Contact	Three-body Distances (Å)	Helix Pair
25V-(77L,78V)	10.0,8.1	1-2
46A-(55A,56V)	9.3,8.3	1-2

Model Complexity

Table 3: A comparison of the complexity of both optimization models applied to bovine rhodopsin (1f88A) with a $s$ubtract value of 0 and a $m$ax_contact value of 2

Model	Probability set	Binary Variables	Constraints	CPU Time
Pair	AL-P	1350	80377	510s
Pair	MIN-1	3368	170934	87s
Triplet	AL-T	317	23706	<1s
Triplet	MIN-2	112	7448	<1s

	A	F	G	I	L	M	P	V	W
A	0.009554	-	-	-	-	-	-	-	-
F	0.012739	0.000000	-	-	-	-	-	-	-
G	0.012739	0.009554	0.003185	-	-	-	-	-	-
I	0.012739	0.009554	0.009554	0.003185	-	-	-	-	-
L	0.022293	0.003185	0.009554	0.009554	0.012739	-	-	-	-
M	0.009554	0.006369	0.000000	0.000000	0.006369	0.000000	-	-	-
P	0.012739	0.000000	0.012739	0.000000	0.003185	0.000000	0.000000	-	-
V	0.009554	0.003185	0.022293	0.003185	0.009554	0.009554	0.006369	0.006369	-
W	0.006369	0.003185	0.003185	0.003185	0.000000	0.000000	0.000000	0.000000	0.003185
C	0.003185	0.000000	0.003185	0.003185	0.000000	0.000000	0.000000	0.000000	0.000000
N	0.006369	0.003185	0.003185	0.003185	0.003185	0.000000	0.000000	0.000000	0.000000
Q	0.000000	0.003185	0.003185	0.000000	0.003185	0.000000	0.000000	0.000000	0.000000
S	0.003185	0.006369	0.006369	0.006369	0.003185	0.000000	0.006369	0.003185	0.003185
T	0.000000	0.003185	0.009554	0.003185	0.003185	0.000000	0.003185	0.000000	0.000000
Y	0.000000	0.003185	0.000000	0.000000	0.000000	0.000000	0.000000	0.006369	0.000000

Table 4: Parallel PRIMARY MIN-1 pairwise probabilities

	A	F	G	I	L	M	P	V	W
A	0.012739	-	-	-	-	-	-	-	-
F	0.012739	0.009554	-	-	-	-	-	-	-
G	0.022293	0.019108	0.022293	-	-	-	-	-	-
I	0.012739	0.006369	0.012739	0.000000	-	-	-	-	-
L	0.022293	0.009554	0.015924	0.003185	0.012739	-	-	-	-
M	0.006369	0.000000	0.000000	0.003185	0.000000	0.003185	-	-	-
P	0.012739	0.003185	0.012739	0.009554	0.000000	0.003185	0.000000	-	-
V	0.025478	0.009554	0.009554	0.003185	0.009554	0.009554	0.003185	0.009554	-
W	0.000000	0.006369	0.006369	0.000000	0.000000	0.000000	0.003185	0.000000	0.000000
C	0.006369	0.003185	0.000000	0.003185	0.006369	0.003185	0.000000	0.003185	0.000000
N	0.012739	0.000000	0.003185	0.000000	0.000000	0.006369	0.003185	0.003185	0.006369
Q	0.000000	0.003185	0.000000	0.000000	0.000000	0.000000	0.000000	0.003185	0.000000
S	0.025478	0.009554	0.012739	0.006369	0.003185	0.000000	0.003185	0.009554	0.006369
T	0.019108	0.012739	0.009554	0.012739	0.015924	0.000000	0.006369	0.003185	0.000000
Y	0.006369	0.006369	0.006369	0.006369	0.009554	0.000000	0.006369	0.000000	0.003185

Table 5: Antiparallel PRIMARY MIN-1 pairwise probabilities

	A	F	G	I	L	M	P	V	W
A	0.708333	-	-	-	-	-	-	-	-
F	0.472222	0.000000	-	-	-	-	-	-	-
G	0.750000	0.625000	0.500000	-	-	-	-	-	-
I	0.687500	1.000000	0.562500	1.000000	-	-	-	-	-
L	0.583333	0.750000	0.583333	0.416667	1.000000	-	-	-	-
M	0.583333	0.500000	0.000000	0.000000	0.416667	0.000000	-	-	-
P	0.466667	0.000000	0.900000	0.000000	1.000000	0.000000	0.000000	-	-
V	0.666667	0.375000	0.589286	0.500000	0.555556	0.314286	0.250000	0.685714	-
W	0.500000	1.000000	0.500000	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000
C	0.000000	0.000000	0.625000	0.800000	0.000000	0.000000	0.000000	0.000000	0.000000
N	0.125000	0.000000	0.000000	0.333333	1.000000	0.000000	0.000000	0.000000	0.000000
Q	0.000000	0.750000	0.250000	0.000000	0.250000	0.000000	0.000000	0.000000	0.000000
S	1.000000	0.625000	0.000000	1.000000	0.500000	0.000000	0.000000	1.000000	1.000000
T	0.000000	0.250000	0.500000	0.000000	0.714286	0.000000	0.000000	0.000000	0.000000
Y	0.000000	0.800000	0.000000	0.000000	0.000000	0.000000	0.000000	0.250000	0.000000

Table 6: Parallel WHEEL MIN-1 pairwise probabilities

	A	F	G	I	L	M	P	V	W
A	0.843750	-	-	-	-	-	-	-	-
F	0.250000	0.437500	-	-	-	-	-	-	-
G	0.475000	0.416667	0.500000	-	-	-	-	-	-
I	0.791667	0.625000	0.486111	0.000000	-	-	-	-	-
L	0.472619	0.674603	0.475000	0.666667	0.666667	-	-	-	-
M	0.541667	0.000000	0.000000	0.250000	0.000000	0.000000	-	-	-
P	0.531250	0.000000	0.580357	0.583333	0.000000	0.750000	0.000000	-	-
V	0.520833	0.750000	0.500000	0.500000	0.583333	0.575000	1.000000	0.619048	-
W	0.000000	0.589286	0.666667	0.000000	0.000000	0.000000	0.500000	0.000000	0.000000
C	0.425000	0.500000	0.000000	0.000000	0.875000	0.500000	0.000000	0.500000	0.000000
N	0.250000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.416667
Q	0.000000	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.250000	0.000000
S	0.468750	0.500000	0.687500	0.875000	0.250000	0.000000	0.500000	0.327778	0.500000
T	0.513889	0.666667	0.700000	0.625000	0.708333	0.000000	0.687500	0.000000	0.000000
Y	0.500000	0.500000	0.083333	0.500000	0.555556	0.000000	0.625000	0.000000	0.000000

Table 7: Antiparallel WHEEL MIN-1 pairwise probabilities

	A	F	G	I	L	M	P	V	W
A	0.000000	-	-	-	-	-	-	-	-
F	0.008398	0.012146	-	-	-	-	-	-	-
G	0.000000	0.005580	0.000000	-	-	-	-	-	-
I	0.000000	0.008495	0.000000	0.000000	-	-	-	-	-
L	0.009855	0.021224	0.005830	0.013215	0.014242	-	-	-	-
M	0.000000	0.007829	0.000000	0.000000	0.007621	0.000000	-	-	-
P	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	-	-
V	0.000000	0.009134	0.000000	0.005872	0.013506	0.000000	0.000000	0.000000	-
W	0.000000	0.012673	0.000000	0.006455	0.012243	0.000000	0.000000	0.005677	0.000000
C	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
N	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
Q	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
S	0.000000	0.000000	0.000000	0.000000	0.006705	0.000000	0.000000	0.000000	0.000000
T	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
Y	0.000000	0.000000	0.000000	0.000000	0.009620	0.000000	0.000000	0.000000	0.000000

Table 8: Parallel PRIMARY AL pairwise probabilities

	A	F	G	I	L	M	P	V	W
A	0.000000	-	-	-	-	-	-	-	-
F	0.015616	0.022557	-	-	-	-	-	-	-
G	0.000000	0.010369	0.000000	-	-	-	-	-	-
I	0.000000	0.015783	0.000000	0.000000	-	-	-	-	-
L	0.018309	0.039436	0.010827	0.024542	0.026443	-	-	-	-
M	0.000000	0.014533	0.000000	0.000000	0.014172	0.000000	-	-	-
P	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	-	-
V	0.000000	0.016963	0.000000	0.010910	0.025083	0.000000	0.000000	0.000000	-
W	0.000000	0.023542	0.000000	0.012007	0.022737	0.000000	0.000000	0.010550	0.000000
C	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
N	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
Q	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
S	0.000000	0.000000	0.000000	0.000000	0.012451	0.000000	0.000000	0.000000	0.000000
T	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
Y	0.000000	0.000000	0.000000	0.000000	0.017865	0.000000	0.000000	0.000000	0.000000

Table 9: Antiparallel PRIMARY AL pairwise probabilities

Triplet	Prob	Triplet	Prob	Triplet	Prob	Triplet	Prob
AAF	0.011450	AVG	0.007634	GGV	0.011450	LLL	0.007634
AAL	0.015267	AVV	0.007634	GGY	0.007634	LSV	0.007634
AAV	0.007634	FAI	0.007634	GLL	0.011450	LVP	0.007634
AGI	0.007634	FLV	0.007634	IAL	0.007634	LVW	0.007634
AGT	0.007634	FRI	0.007634	IAM	0.007634	LVY	0.007634
ALG	0.007634	FVL	0.011450	IGV	0.007634	MGR	0.007634
ALL	0.007634	GAL	0.007634	LAL	0.007634	PVY	0.007634
ALP	0.007634	GGI	0.011450	LAV	0.011450	RMT	0.007634
ALY	0.007634	GGL	0.007634	LFY	0.007634	TLV	0.007634
AMI	0.007634	GGT	0.007634	LIV	0.007634	TMV	0.007634
AVA	0.007634						

Table 10: Parallel MIN-2 three-body probabilities

Triplet	Prob	Triplet	Prob	Triplet	Prob	Triplet	Prob
AAI	0.007634	FAF	0.007634	GHS	0.007634	LGV	0.011450
AAL	0.007634	FAN	0.011450	GLL	0.007634	LHV	0.007634
AAV	0.007634	FGH	0.007634	GNT	0.011450	LLT	0.015267
AAY	0.007634	FGL	0.022901	GPT	0.007634	LLV	0.011450
AFL	0.007634	FGY	0.007634	GVM	0.007634	LLY	0.007634
AGG	0.019084	FII	0.007634	GYS	0.007634	LRV	0.007634
AGI	0.011450	FSS	0.007634	HVV	0.007634	LVP	0.007634
AGL	0.015267	FTL	0.007634	IFL	0.007634	LVV	0.022901
AIL	0.011450	GAM	0.007634	IFV	0.007634	NNP	0.007634
ALA	0.007634	GAS	0.011450	IIM	0.007634	NYY	0.007634
ALL	0.007634	GAT	0.007634	IVL	0.007634	QAV	0.007634
ANP	0.007634	GCI	0.007634	LAL	0.007634	SFY	0.007634
ASE	0.007634	GFT	0.007634	LAS	0.007634	TGV	0.007634
ATP	0.007634	GGI	0.007634	LAT	0.015267	TLV	0.007634
AVA	0.015267	GGL	0.011450	LAV	0.011450	TNV	0.007634
AVI	0.007634	GGT	0.007634	LAW	0.015267	VGY	0.007634
AYF	0.007634	GGV	0.011450	LGQ	0.007634	VSV	0.011450
ENV	0.007634						

Table 11: Antiarallel MIN-2 three-body probabilities

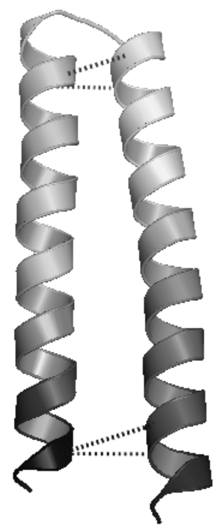

Figure 1: A high-scoring set of contact predictions (blue dashes) for 1h2sB using the triplet model and the AL-T probability set. This figure has been created using PyMol (5).

Triplet	Prob	Triplet	Prob	Triplet	Prob	Triplet	Prob
AAA	0.006857	AGS	0.017143	APY	0.006286	GHT	0.009143
AAF	0.021143	AGT	0.015429	ARG	0.010286	GIL	0.022857
AAG	0.028571	AGV	0.025143	ASS	0.010286	GLL	0.030857
AAI	0.021714	AGY	0.012571	ASV	0.010857	GLM	0.018857
AAL	0.025714	AHS	0.012000	ATV	0.014857	GLV	0.030286
AAM	0.014857	AHT	0.009143	CGF	0.007429	GMF	0.016000
AAP	0.005714	AIM	0.014286	GFF	0.021143	GMS	0.011429
AAS	0.007429	AIP	0.008571	GFS	0.008571	GMV	0.016000
AAV	0.010857	AIT	0.013143	GGF	0.017143	GMW	0.009143
AAW	0.010857	ALL	0.036000	GGG	0.005714	GST	0.011429
ACI	0.008571	ALM	0.021143	GGI	0.009714	GSV	0.005714
AFP	0.010286	ALS	0.026857	GGL	0.013143	GTV	0.008571
AGF	0.042857	AMF	0.019429	GGM	0.009714	GVV	0.006286
AGG	0.037143	AMV	0.013714	GGS	0.009143	IIM	0.010286
AGI	0.025143	ANG	0.006857	GGV	0.009714	NGS	0.006286
AGL	0.050857	ANY	0.006857	GHS	0.006286	STV	0.008571
AGM	0.013143	APT	0.008000				

Table 12: AL-T three-body probabilities

References

1. Floudas, C. A., 1995. Nonlinear and Mixed-integer Optimization: Fundamentals and Applications. Oxford University Press.

2. Adamian, L., and J. Liang, 2001. Helix-Helix Packing and Interfacial Pairwise Interactions of Residues in Membrane Proteins. *J. Mol. Biol.* 311:891–907.

3. Adamian, L., R. Jackups Jr., T. A. Binkowski, and J. Liang, 2003. Higher-order Interhelical Spatial Interactions in Membrane Proteins. *J. Mol. Biol.* 327:251–272.

4. Gordeliy, V. I., J. Labahn, R. Moukhametzianov, R. Efremov, J. Granzin, R. Schlesinger, G. Büldt, T. Savopol, A. J. Scheidig, J. P. Klare, and M. Engelhard, 2002. Molecular basis of transmembrane signalling by sensory rhodopsin II-transducer complex. *Nature* 419:484–487.

5. DeLano, W. L., 2002. The PyMol Molecular Graphics System. DeLano Scientific, Palo Alto, CA USA. http://www.pymol.org.