Supplementary Material: Enhanced Inter-helical Contact Prediction in Transmembrane Proteins

Y. Wei, C. A. Floudas*
Department of Chemical and Biological Engineering,
Princeton University,
Princeton, NJ 08544-5263, U.S.A.

February 11, 2011

---

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

1 Description of the Constraints for the TRIPLET model

This part of the supplementary material aims at providing a complete description of the constraints for the TRIPLET model. The constraints are grouped into five categories as they were for the PAIRWISE model in this paper.

Basic model constraints. There are two basic model constrains for the TRIPLET model. Eq. 1 states that a residue-residue contact can happen only when the two helices containing these two residues are contacting with each other. Eq. 2 expresses that when all the possible residue pairs on helix m and helix n are not contacting, the residue contact between helices m and n is disallowed.

$$\sum_j \sum_t w_{ijt}^{mn} \leq y_{mn}^P + y_{mn}^A, \quad \forall (i, m, n) \tag{1}$$

$$y_{mn}^P + y_{mn}^A - \sum_j \sum_t w_{ijt}^{mn} \leq 0, \quad \forall (m, n) \tag{2}$$

Geometrical constraints. The first geometrical constraint is that a helical pair can only interact either in antiparallel or parallel fashion, and the expression is the same as Eq. 4 for the PAIRWISE model in the paper. This is expressed here in Eq. 3:

$$y_{mn}^P + y_{mn}^A \leq 1, \quad \forall (m, n) \tag{3}$$

The second geometrical constraint deals with the contacting fashion between the first and last helices of a protein, and this idea is the same as in the PAIRWISE model. The constraint is expressed in two equations in the paper, Eq. 4 and Eq. 5. In these two equations, N is the total number of helices. Eq. 4 disallows the parallel contact between the first and last helices if N is even, but is not enforced if N is odd; Eq. 5 disallows the antiparallel contact between the first and last helices if N is odd, but is not enforced if N is even.

$$y^P_{1N} \leq MOD(N,2), \quad N > 2 \tag{4}$$

$$y^A_{1N} \leq 1 - MOD(N,2), \quad N > 2 \tag{5}$$

In order to prevent kinks between interacting helices, a similar equation of Eq. 7 is written for TRIPLET model as follows:

$$\begin{aligned} \sum_t (w^{mn}_{ijt} + \sum_l w^{mn}_{klt}) \leq 1 \\ \forall (i,j,k),\ \ (|diff(i,k)| - |diff(j,l)|) > 3 \\ or\ |diff(i,k)| < 5,\ or\ |diff(j,l)| < 5. \end{aligned} \tag{6}$$

The fourth geometrical constraint is expressed in Eq. 7 and Eq. 8. For a helical pair m and n contacting in parallel, Eq. 7 states if there are two residues pairs (i,j) and (k,l) on them forming contacts, j must be greater than l given i is greater than k. Eq. 8 describes the case when two helices are in antiparallel contact.

$$\begin{aligned} w^{mn}_{ij} + \sum_l w^{mn}_{kl} + y^A_{mn} \leq 2 \\ \forall (i,j,k) | l > j,\ |(|diff(i,k)| - |diff(j,l)|)| < 3 \end{aligned} \tag{7}$$

$$\begin{aligned} w^{mn}_{ij} + \sum_l w^{mn}_{kl} + y^P_{mn} \leq 2 \\ \forall (i,j,k) | l < j,\ |(|diff(i,k)| - |diff(j,l)|)| < 3 \end{aligned} \tag{8}$$

The TRIPLET model also includes that constraints dealing with the case when a small helix and a long helix is connected by a short loop. These constraints are expressed in a similar fashion as those for PAIRWISE model. The equations are as follows in Eqs. 9 and 10:

$$\begin{aligned} w_{ijt}^{mn} &= 0 \\ \forall(i,j,t) \mid (3.0 \times (loop\_length(m,n) - 1) \\ +j \times 6.0/4) \leq (len(m) - i + 1) \times 6.0/4 \end{aligned} \tag{9}$$

$$\begin{aligned} w_{ijt}^{mn} &= 0 \\ \forall(i,j,t) \mid (3.0 \times (loop\_length(m,n) - 1) \\ +(len(n) - i + 1) \times 6.0/4 \leq j \times 6.0/4) \end{aligned} \tag{10}$$

Model complexity constraints. The first constraint in this category is dealing with the total number of contacts for each helix. A parameter, counth(m), is used to limit the allowed number of helical contacts for helix m. The functional form of this constraint is same as the one for PAIRWISE model. This is written in Eq. 11.

$$\sum_n (y_{mn}^P + y_{mn}^A) \leq counth(m), \quad \forall(m) \tag{11}$$

The second constraint of this category, however, is not easily transferable from the one of PAIRWISE model. This constraint basically states that a residue can mostly contact with one other residue. This is the main difference between the TRIPLET and PAIRWISE models. In order to have the similar constraint as in Eq. 11 for the PAIRWISE model, three rules are established. The first rule states that a MAIN residue in a TRIPLET contact can not server either as a MAIN or SECONDARY residue in another TRIPLET contact; the second rule says two SECONDARY residues of a TRIPLET contact cannot participate in TRIPLET contacts with multiple residues; the third rule states that a SECONDARY residue is disallowed to form more than two contacts. These rules are implemented through several constraints in the following.

The first rule is constrained by the following three equations, Eqs 12, 13 and 14.

$$w^{mn}_{i,j,t=2} + \sum_k (w^{np}_{j,k,t=2} + w^{np}_{j+1,k,t=2}) \leq 1, \quad \forall (i,j) \tag{12}$$

This equation restricts that if residue j or j+1 serves as a MAIN residue on helix n in a TRIPLET contact between helix n and helix p, j or j+1 cannot serve as a SECONDARY residue in another TRIPLET contacts.

$$w^{mn}_{i,j,t=1} + \sum_k w^{np}_{j,k,t=2} \leq 1, \quad \forall (i,j) \tag{13}$$

Eq. 13 basically states that if j serves as a MAIN residue of a TRIPLET contact, it cannot serves as another MAIN residue in a second TRIPLET contact. That is $w^{mn}_{i,j,t=1}$ =0 if $\sum_k w^{np}_{j,k,t=2}$ =1; and $\sum_k w^{np}_{j,k,t=2}$ =0 if $w^{mn}_{i,j,t=1}$ =1.

Equation 14 has the similar effect as Eq. 12. It states that if residue j or j-1 serves as a MAIN residue on helix n in a TRIPLET contact between helix n and helix m, j or j-1 cannot serve as a SECONDARY residue in another TRIPLET contacts.

$$w^{mn}_{i,j,t=1} + \sum_k (w^{np}_{j,k,t=1} + w^{np}_{j-1,k,t=1}) \leq 1, \quad \forall (i,j) \tag{14}$$

Eq. 15 is used to impose the second rule. It states that if residue j and its neighbor residue serve as SECONDARY residues in a TRIPLET contact, then they cannot participate in other TRIPLET contact as a SECONDARY residue pair.

$$w^{mn}_{i,j,t=2} + \sum_k w^{np}_{j,k,t=1} \leq 1, \quad \forall (i,j) \tag{15}$$

The final rule is implemented in Eq. 16. This limits the overlapping TRIPLETs on one helix. This equation says if residue j is a SECONDARY residue on helix n of a TRIPLET between helix m and n, only one of its neighbor residue, j+1 or j-1, can participate in a TRIPLET contact.

$$w^{mn}_{i,j,t=2} + \sum_{k}(w^{np}_{j-1,k,t=1} + w^{np}_{j+1,k,t=1}) \leq 1, \quad \forall(i,j) \tag{16}$$

Membrane protein observations. The first constraint in this category limits the neighboring helices to form anti-parallel contacts, which is written as:

$$\sum_{mn;}\sum_{(n-m)=1} y^{P}_{mn} \leq 0 \tag{17}$$

The second constraint prevents the helical contact prediction resulting in the overlap between two helices less than 90% of the shorter helix, and this is expressed in the following equations, Eq. 19 and Eq. 20.

$$\begin{aligned} w^{mn}_{ij} + y^{A}_{mn} &\leq 1, \quad \forall(i,j,m,n) \\ if\ overlap\ between\ helices\ (m,n)\ is &< 90\%\ of\ the\ shorter\ helix \end{aligned} \tag{18}$$

$$\begin{aligned} w^{mn}_{ij} + y^{P}_{mn} &\leq 1, \quad \forall(i,j,m,n) \\ if\ overlap\ between\ helices\ (m,n)\ is &< 90\%\ of\ the\ shorter\ helix \end{aligned} \tag{19}$$

Model features. Max_contact and subtract parameters in TRIPLET model have the same effect as in PAIRWISE model. Max_contact is used to limit the maximum number of residue-residue contacts for each contacting helical pair. For TRIPLET model, this is expressed in Eq. 20.

$$\sum_i \sum_j \sum_t w_{ijt}^{mn} \leq max\_contact \times (y_{mn}^P + y_{mn}^A), \quad \forall (m,n) \tag{20}$$

In a similar fashion, subtract value is used here to reduce the number of helical contact predictions, thus allowing various different degrees of helix packing. The functional form of this constraint is as follows:

$$\sum_{mn;} \sum_{m<n} (y_{mn}^P + y_{mn}^A) \leq \sum_m counth(m)/2 - subtract \tag{21}$$

Finally, in order to generate a rank-order list, the same strategy of using integer cut is used to generate multiple solutions, see Eq. 49 of McAllister and Floudas (2008) [1]. The functional form is expressed as in Eq. 22:

$$\begin{gathered}\sum_{(m,n)\in A} (y_{mn}^P + y_{mn}^A) + \sum_{(i,j,t)\in A} w_{ijt}^{mn} - \\ \sum_{(m,n)\in I} (y_{mn}^P + y_{mn}^A) - \sum_{(i,j,t)\in I} w_{ijt}^{mn} \leq card(A) - 1\end{gathered} \tag{22}$$

For additional information about the model, the readers are referred to the paper by McAllister and Floudas [1].

2 Probability Sets

This section contains the tables for the developed probability sets. Since AL-T and AL-P probability sets are directly obtained from the work by McAllister and Floudas [1], they are not included in this section. The tables start from the next page.

	A	F	G	I	L	M	P	V	W
A	0.008902	0.004946	0.014837	0.013848	0.014837	0.003956	0.005935	0.012859	0.005935
F	0.004946	0.001978	0.006924	0.004946	0.008902	0.002967	0.001978	0.004946	0.001978
G	0.014837	0.006924	0.005935	0.006924	0.014837	0.003956	0.006924	0.016815	0.001978
I	0.013848	0.004946	0.006924	0.000989	0.011869	0.002967	0.001978	0.008902	0.001978
L	0.014837	0.008902	0.014837	0.011869	0.004946	0.004946	0.003956	0.009891	0.000989
M	0.003956	0.002967	0.003956	0.002967	0.004946	0.000989	0.000989	0.005935	0.000989
P	0.005935	0.001978	0.006924	0.001978	0.003956	0.000989	0.000000	0.005935	0.000000
V	0.012859	0.004946	0.016815	0.008902	0.009891	0.005935	0.005935	0.005935	0.000989
W	0.005935	0.001978	0.001978	0.001978	0.000989	0.000989	0.000000	0.000989	0.000000
C	0.003956	0.000000	0.000000	0.001978	0.000000	0.000989	0.000000	0.000989	0.000000
N	0.005935	0.001978	0.001978	0.004946	0.000989	0.000000	0.003956	0.000000	0.000000
Q	0.001978	0.001978	0.000989	0.000000	0.001978	0.000000	0.000000	0.000989	0.000000
S	0.008902	0.001978	0.006924	0.008902	0.002967	0.000000	0.001978	0.003956	0.001978
T	0.004946	0.003956	0.004946	0.003956	0.002967	0.000000	0.003956	0.000989	0.000000
Y	0.001978	0.002967	0.000989	0.000000	0.003956	0.001978	0.000000	0.004946	0.000000

Table 1: Parallel PRIMARY MIN-1N PAIRWISE probabilities.

	A	F	G	I	L	M	P	V	W
A	0.022750	0.017804	0.013848	0.022750	0.022750	0.007913	0.008902	0.023739	0.002967
F	0.017804	0.004946	0.015826	0.006924	0.008902	0.006924	0.002967	0.008902	0.003956
G	0.013848	0.015826	0.019782	0.015826	0.015826	0.001978	0.008902	0.010880	0.004946
I	0.022750	0.006924	0.015826	0.006924	0.008902	0.006924	0.005935	0.012859	0.003956
L	0.022750	0.008902	0.015826	0.008902	0.017804	0.004946	0.004946	0.017804	0.000989
M	0.007913	0.006924	0.001978	0.006924	0.004946	0.001978	0.002967	0.002967	0.001978
P	0.008902	0.002967	0.008902	0.005935	0.004946	0.002967	0.000000	0.004946	0.002967
V	0.023739	0.008902	0.010880	0.012859	0.017804	0.002967	0.004946	0.008902	0.000989
W	0.002967	0.003956	0.004946	0.003956	0.000989	0.001978	0.002967	0.000989	0.000000
C	0.003956	0.002967	0.000000	0.001978	0.003956	0.000989	0.000000	0.001978	0.000000
N	0.008902	0.000000	0.002967	0.001978	0.000989	0.000989	0.000000	0.004946	0.002967
Q	0.001978	0.001978	0.000989	0.000989	0.000989	0.000000	0.000000	0.000989	0.000000
S	0.015826	0.010880	0.013848	0.008902	0.005935	0.004946	0.002967	0.005935	0.000000
T	0.017804	0.010880	0.006924	0.009891	0.009891	0.001978	0.004946	0.011869	0.001978
Y	0.009891	0.004946	0.003956	0.001978	0.003956	0.002967	0.003956	0.004946	0.000989

Table 2: Antiparallel PRIMARY MIN-1N PAIRWISE probabilities.

	A	F	G	I	L	M	P	V	W
A	0.597222	0.393750	0.736667	0.701389	0.631410	0.666667	0.530000	0.597222	0.002967
F	0.393750	0.200000	0.687500	0.666667	0.484694	0.555556	1.000000	0.468750	0.003956
G	0.736667	0.687500	0.630952	0.726191	0.634325	0.333333	0.752381	0.612180	0.004946
I	0.701389	0.666667	0.726191	0.000000	0.540476	0.750000	0.000000	0.654762	0.003956
L	0.631410	0.484694	0.634325	0.540476	0.733333	0.478571	0.625000	0.663889	0.000989
M	0.666667	0.555556	0.333333	0.750000	0.478571	0.600000	1.000000	0.355238	0.001978
P	0.530000	1.000000	0.752381	0.000000	0.625000	1.000000	0.000000	0.353333	0.002967
V	0.597222	0.468750	0.612180	0.654762	0.663889	0.355238	0.353333	0.700794	0.000989
W	0.562500	0.500000	0.500000	0.000000	0.400000	0.250000	0.000000	0.500000	0.000000
C	0.722222	0.000000	0.000000	0.483333	0.000000	0.500000	0.000000	0.714286	0.000000
N	0.406250	1.000000	0.500000	0.610714	1.000000	0.000000	0.312500	0.000000	0.002967
Q	0.333333	0.750000	0.250000	0.000000	0.250000	0.000000	0.000000	0.500000	0.000000
S	0.708333	0.625000	0.458333	0.708333	0.833333	0.000000	0.000000	0.569444	0.000000
T	0.350000	0.226191	0.666667	0.488095	0.738095	0.000000	0.777778	0.000000	0.001978
Y	0.875000	0.550000	0.500000	0.000000	0.250000	0.583333	0.000000	0.390000	0.000989

Table 3: Parallel WHEEL MIN-1N PAIRWISE probabilities.

	A	F	G	I	L	M	P	V	W
A	0.517045	0.422619	0.417208	0.518128	0.488095	0.698980	0.578125	0.472457	0.002967
F	0.422619	0.833333	0.507143	0.423611	0.711309	0.388889	0.583333	0.535714	0.003956
G	0.417208	0.507143	0.590278	0.555736	0.362179	0.125000	0.711640	0.527056	0.004946
I	0.518128	0.423611	0.555736	0.278571	0.750000	0.260417	0.465278	0.542582	0.003956
L	0.488095	0.711309	0.362179	0.750000	0.583333	0.333333	0.416667	0.460417	0.000989
M	0.698980	0.388889	0.125000	0.260417	0.333333	0.250000	0.527778	0.575000	0.001978
P	0.578125	0.583333	0.711640	0.465278	0.416667	0.527778	0.000000	0.562500	0.002967
V	0.472457	0.535714	0.527056	0.542582	0.460417	0.575000	0.562500	0.498639	0.000989
W	1.000000	0.598214	0.654762	0.708333	0.000000	0.250000	0.222222	1.000000	0.000000
C	0.450000	0.611111	0.000000	0.500000	0.700000	0.500000	0.000000	0.500000	0.000000
N	0.466667	0.000000	0.666667	0.000000	1.000000	0.500000	0.000000	0.250000	0.002967
Q	0.625000	0.187500	0.500000	0.333333	0.500000	0.000000	0.000000	0.500000	0.000000
S	0.461667	0.537037	0.718750	0.833333	0.339286	0.468750	0.833333	0.518056	0.000000
T	0.599868	0.528125	0.580556	0.566667	0.512649	0.812500	0.375000	0.434848	0.001978
Y	0.611111	0.413333	0.333333	1.000000	0.166667	0.250000	0.430556	0.241667	0.000989

Table 4: Antiparallel WHEEL MIN-1N PAIRWISE probabilities.

TRIP	Prob	TRIP	Prob	TRIP	Prob	TRIP	Prob
AAL	0.001445	ADQ	0.001445	MME	0.002168	CLR	0.001445
AAM	0.002168	AEC	0.001445	MMH	0.004335	CMC	0.001445
AAV	0.001445	AHA	0.001445	MMR	0.001445	CMS	0.001445
AAC	0.004335	AHL	0.002168	MVH	0.001445	CMY	0.001445
AAN	0.001445	AHM	0.002168	MCM	0.001445	CME	0.002168
AAQ	0.001445	AHV	0.00289	MCC	0.00578	CMK	0.001445
AAS	0.001445	AHC	0.002168	MCN	0.002168	CVC	0.002168
AAY	0.001445	AHS	0.001445	MCS	0.001445	CVQ	0.001445
AAD	0.002168	AHT	0.001445	MNC	0.001445	CVY	0.001445
AAH	0.004335	AHH	0.00289	MNE	0.001445	CVH	0.002168
ALA	0.001445	GVC	0.002168	MQY	0.001445	CVK	0.002168
ALC	0.001445	IGS	0.001445	MSC	0.002168	CCC	0.001445
ALN	0.001445	IMH	0.001445	MSQ	0.001445	CSH	0.001445
AMA	0.002168	IVP	0.001445	MDD	0.001445	CYH	0.001445
AML	0.001445	IVC	0.002168	MHV	0.001445	CDH	0.002168
AMV	0.00289	ICV	0.002168	MHC	0.002168	CHE	0.001445
AMC	0.001445	ICC	0.001445	MKC	0.001445	CHH	0.001445
AMN	0.001445	IEC	0.002168	PAC	0.001445	CHK	0.002168
AME	0.001445	IHS	0.001445	VAC	0.001445	NMY	0.001445
AMH	0.001445	LAM	0.00289	VAE	0.001445	NHH	0.001445

Table 5: Parallel MIN-2N TRIPLET probabilities (1 of 2).

TRIP	Prob	TRIP	Prob	TRIP	Prob	TRIP	Prob
AVL	0.001445	LAV	0.001445	VMN	0.001445	QVH	0.001445
AVV	0.001445	LAC	0.002168	VVC	0.001445	SAE	0.001445
AVC	0.001445	LLC	0.002168	VVH	0.001445	SVH	0.001445
AVN	0.002168	LML	0.002168	VCC	0.001445	SHR	0.001445
AVH	0.003613	LMC	0.001445	VCN	0.001445	THR	0.002168
AVK	0.001445	LMN	0.002168	VCH	0.001445	YCK	0.001445
ACA	0.001445	LMS	0.002168	VSN	0.001445	YNE	0.001445
ACM	0.00289	LME	0.003613	VDV	0.001445	DWE	0.001445
ACC	0.004335	LMK	0.002168	VHC	0.00289	DCR	0.001445
ACS	0.002168	LVD	0.001445	VHR	0.001445	EAH	0.001445
ACH	0.00289	LVH	0.002168	WLT	0.001445	EMH	0.001445
ACR	0.002168	LSM	0.001445	WVC	0.001445	EWE	0.001445
ANV	0.001445	LHC	0.001445	WCN	0.001445	ESE	0.002168
ANC	0.001445	MAC	0.003613	CAC	0.003613	EDH	0.001445
AQG	0.001445	MAN	0.001445	CAE	0.001445	EHH	0.001445
AQV	0.001445	MFH	0.001445	CAH	0.005058	HAH	0.001445
ASS	0.001445	MLD	0.001445	CFD	0.001445	HHR	0.001445
ASH	0.001445	MMV	0.00289	CFR	0.001445		
ADG	0.001445	MMC	0.002168	CLQ	0.001445		
ADL	0.001445	MMN	0.001445	CLK	0.001445		

Table 6: Parallel MIN-2N TRIPLET probabilities (2 of 2).

TRIP	Prob	TRIP	Prob	TRIP	Prob	TRIP	Prob
AAG	0.002168	GAL	0.001445	MVH	0.001445	CME	0.001445
AAI	0.002168	GAC	0.001445	MVK	0.001445	CMH	0.00578
AAL	0.00578	GQN	0.001445	MWQ	0.001445	CMK	0.001445
AAM	0.002168	GQE	0.001445	MCM	0.001445	CPH	0.001445
AAV	0.006503	IAE	0.001445	MCV	0.00289	CVD	0.00289
AAW	0.001445	IMM	0.001445	MCC	0.007225	CVK	0.001445
AAC	0.005058	IMS	0.001445	MCH	0.001445	CVR	0.001445
AAN	0.004335	IVV	0.001445	MCK	0.001445	CCC	0.00578
AAD	0.002168	ICH	0.001445	MQE	0.004335	CCQ	0.001445
AAE	0.002168	IQL	0.001445	MSC	0.002168	CCS	0.001445
AAH	0.006503	ISC	0.002168	MSE	0.002168	CCT	0.001445
AAK	0.001445	IDC	0.001445	MSH	0.001445	CCY	0.003613
AAR	0.003613	LAL	0.001445	MTE	0.001445	CCD	0.004335
AFF	0.001445	LAM	0.002168	MYE	0.001445	CCE	0.003613
AGM	0.002168	LAV	0.001445	MDM	0.001445	CCH	0.00578
AGV	0.001445	LAC	0.003613	MDC	0.001445	CCK	0.001445
AGR	0.001445	LAQ	0.00289	MDD	0.001445	CCR	0.001445
AIG	0.001445	LAS	0.001445	MDE	0.002168	CNS	0.001445
AII	0.001445	LAT	0.002168	MHV	0.001445	CQT	0.001445
ALA	0.004335	LAY	0.002168	MHC	0.004335	CQD	0.002168
ALL	0.001445	LAH	0.001445	MHN	0.002168	CQE	0.001445
ALC	0.002168	LFC	0.001445	MHH	0.001445	CSH	0.001445
ALY	0.001445	LFE	0.001445	MRD	0.00289	CTS	0.001445
ALH	0.002168	LID	0.001445	PAK	0.001445	CYH	0.001445
AMI	0.002168	LLL	0.001445	PMV	0.001445	CDN	0.002168
AMM	0.007225	LLM	0.002168	PDS	0.001445	CDQ	0.002168
AMV	0.007948	LLC	0.00289	PHH	0.001445	CDD	0.00289

Table 7: Antiparallel MIN-2N TRIPLET probabilities (1 of 3).

TRIP	Prob	TRIP	Prob	TRIP	Prob	TRIP	Prob
AMC	0.004335	LMV	0.001445	VAC	0.002168	CDH	0.001445
AMN	0.001445	LMC	0.004335	VAS	0.00289	CEC	0.001445
AMD	0.001445	LMD	0.001445	VAE	0.003613	CEH	0.001445
AME	0.002168	LVM	0.001445	VAH	0.001445	CEK	0.001445
AMH	0.003613	LVV	0.00289	VIC	0.001445	CER	0.001445
AMK	0.001445	LVC	0.001445	VLC	0.002168	CHC	0.001445
AVM	0.003613	LVD	0.002168	VLE	0.001445	CHQ	0.00289
AVV	0.004335	LVH	0.001445	VLH	0.002168	CHS	0.002168
AVC	0.005058	LCV	0.003613	VMC	0.003613	CHD	0.002168
AVN	0.001445	LCC	0.001445	VMS	0.001445	CHE	0.001445
AVE	0.001445	LCS	0.002168	VMD	0.002168	CHH	0.007225
AVH	0.001445	LCY	0.002168	VMH	0.001445	CHR	0.001445
AWT	0.001445	LNH	0.001445	VMK	0.001445	NAH	0.00289
ACA	0.002168	LQE	0.002168	VVV	0.001445	NAR	0.001445
ACF	0.00289	LQK	0.002168	VVC	0.004335	NME	0.00289
ACI	0.001445	LSK	0.001445	VVN	0.00289	NVE	0.001445
ACL	0.002168	LYC	0.001445	VVD	0.001445	NCE	0.001445
ACM	0.003613	LDY	0.001445	VVH	0.002168	NHE	0.001445
ACP	0.001445	LDD	0.001445	VWS	0.002168	NRY	0.001445
ACV	0.003613	LDE	0.001445	VCS	0.001445	QAS	0.001445
ACC	0.004335	LEM	0.002168	VCH	0.001445	QQS	0.00289
ACE	0.002168	LEP	0.001445	VCR	0.001445	QDH	0.00289
ACH	0.003613	LEV	0.00289	VNC	0.001445	QHH	0.001445
AQC	0.001445	LEC	0.00289	VSD	0.002168	SAE	0.001445
AQS	0.002168	LEH	0.001445	VYS	0.001445	SAH	0.001445
ASV	0.001445	LHM	0.001445	VYD	0.002168	SME	0.002168
ASC	0.002168	LHV	0.004335	VDN	0.002168	SVD	0.001445

Table 8: Antiparallel MIN-2N TRIPLET probabilities (2 of 3).

TRIP	Prob	TRIP	Prob	TRIP	Prob	TRIP	Prob
ASR	0.002168	LHC	0.001445	VED	0.00289	SCH	0.001445
AYH	0.002168	LHQ	0.001445	VEE	0.001445	SHH	0.001445
ADL	0.00289	LHE	0.001445	VHC	0.003613	SHR	0.001445
ADV	0.001445	LRM	0.001445	VHN	0.001445	TAH	0.002168
ADC	0.001445	MAC	0.00289	VHD	0.00289	TCH	0.001445
ADN	0.002168	MAD	0.003613	VHE	0.001445	TCR	0.001445
ADQ	0.002168	MAE	0.005058	VHH	0.001445	TRY	0.001445
ADY	0.001445	MAH	0.00289	WMC	0.001445	YCH	0.001445
ADE	0.002168	MFV	0.001445	WVD	0.001445	DAD	0.002168
ADH	0.001445	MGC	0.001445	WCC	0.002168	DAH	0.002168
AEV	0.001445	MIV	0.001445	CAN	0.002168	DLK	0.001445
AEC	0.002168	MLM	0.001445	CAS	0.001445	DLR	0.001445
AES	0.002168	MLV	0.001445	CAT	0.003613	DNH	0.001445
AHA	0.004335	MLC	0.001445	CAY	0.001445	DDE	0.001445
AHL	0.002168	MLN	0.002168	CAD	0.002168	EAH	0.002168
AHV	0.003613	MLE	0.001445	CAE	0.002168	EAR	0.001445
AHC	0.003613	MLH	0.002168	CAH	0.00578	ELH	0.001445
AHT	0.001445	MMV	0.00289	CAK	0.004335	EVH	0.001445
AHD	0.003613	MMC	0.00578	CAR	0.001445	ECH	0.001445
AHE	0.002168	MMD	0.00289	CLD	0.00289	EHH	0.002168
AHH	0.00289	MME	0.001445	CLE	0.002168	ERH	0.001445
AHK	0.001445	MMH	0.002168	CLH	0.002168	HMH	0.001445
AKT	0.001445	MPC	0.00289	CLR	0.001445	HSR	0.001445
ARL	0.001445	MPD	0.002168	CMS	0.001445	HDH	0.002168
ARC	0.001445	MPE	0.003613	CMT	0.00289	RLR	0.001445
FAC	0.001445	MVC	0.002168	CMY	0.001445		
FCC	0.001445	MVD	0.001445	CMD	0.002168		

Table 9: Antiparallel MIN-2N TRIPLET probabilities (3 of 3).

PDB ID	2ZY9A	2W1PA	3CN5A
No. of Helices	5	6	6
Helical Segments	265R-284F 290A-314T 336K-360G 364L-388F 398A-422R	42F-53F 82F-94Y 124A-136G 165E-176V 194I-203L 238G-248S	15I-26A 49A-60C 96C-107K 137E-148V 167A-178V 217G-226A
PDB ID	3KCUA	3IXZA	3K3FA
No. of Helices	6	10	10
Helical Segments	8Y-31T 40K-64S 85N-109E 125A-148C 163K-187F 220M-244L	66A-79W 108L-119C 271D-282A 305C-317A 754L-766I 780T-789D 830Q-842D 893Y-902M 932A-943L 966P-977D	25G-45Y 48T-67L 78G-97Y 100A-120L 127G-150S 187M-206S 210A-229S 233P-256Y 259I-279F 282L-306G

Table 10: Transmembrane helical segments predicted by MEMSAT3 method for the independent test set (See David T. Jones, "Improving the accuracy of transmembrane protein topology prediction using evolutionary information", Bioinformatics, 2007;23(5):538-544).

PDB ID	2ZY9A	2W1PA	3CN5A
No. of Helices	5	6	6
Helical Segments	260W-283G	38M-56A	10A-27T
	295A-313A	77M-94Y	45G-62A
	334L-359D	121I-141M	88A-106V
	366L-387L	160R-179L	132T-151A
	400V-421A	190A-206I	167A-183I
		223W-250W	212W-229H

PDB ID	3KCUA	3IXZA	3K3FA
No. of Helices	7	10	10
Helical Segments	8Y-31T	74L-88A	35T-48T
	35P-55V	103N-117T	50L-62T
	87L-113A	272L-286C	82F-94F
	118G-151V	294A-309G	100A-116M
	165F-176A	752Y-769T	132T-146V
	182S-198F	776L-792P	193V-203L
	217L-241L	827S-841T	210A-227M
		896F-909L	237G-249A
		930V-944C	264Y-283L
		963W-977D	289P-302F

Table 11: Transmembrane helical segments obtained from PDBTM(See Tusnády GE, Dosztányi Z, Simon I. Transmembrane proteins in the Protein Data Bank: identification and classification. Bioinformatics. 2004;20(17):2964-72).

Res1	Res2	Distance	Res1	Res3	Distance
276V	300V	19.00	276V	301L	17.40
275M	348L	27.10	275M	349S	27.10
272L	301L	21.30	272L	302L	17.60
360G	290A	17.10	360G	291V	13.50
298V	352L	12.00	298V	353V	13.60
348L	301L	10.10	348L	302L	12.80
338M	384F	9.20	338M	385E	11.10
370V	352L	10.70	370V	353V	9.40
372L	351L	10.40	372L	352L	8.60
382V	405V	7.90	382V	406A	10.50
368V	420V	9.10	368V	421A	8.40
367V	420V	7.00	367V	421A	6.10
275M	408L	10.80	275M	409S	9.90

Table 12: The proposed contact prediction for 2ZY9A from our method. Res1 is the MAIN residue and Res2 and Res3 are the SECONDARY residues in the TRIPLET contact.

Res1	Res2	Distance	Res1	Res3	Distance
135A	51S	9.90	135A	52A	9.90
133A	51S	12.00	133A	52A	13.10
134A	51S	8.30	134A	52A	9.40
87M	132M	20.20	87M	133A	19.80
132M	86V	16.60	132M	88V	19.70
			132M	88V	19.70
170A	129V	23.30	170A	130A	22.60
134A	165E	21.50	134A	166A	23.80
135A	165E	20.50	135A	166A	23.10
198L	172L	6.00	198L	173C	9.60
245A	173C	8.80	245A	174L	8.50
197A	245A	15.50	197A	246A	13.00
241L	201A	14.90	241L	202A	17.30

Table 13: The proposed contact prediction for 2W1PA from our method. Res1 is the MAIN residue and Res2 and Res3 are the SECONDARY residues in the TRIPLET contact.

Res1	Res2	Distance	Res1	Res3	Distance
49A	25V	8.90	49A	26A	9.70
57L	18L	10.50	57L	19L	8.40
25V	105L	9.70	25V	106V	6.90
96C	59Y	20.10	96C	60C	19.00
145V	100I	23.00	145V	101C	24.80
143V	100I	25.30	143V	101C	27.30
140G	174A	8.90	140G	175V	8.50
145V	223A	8.80	145V	224V	9.10
143V	223A	11.70	143V	224V	10.30
172G	223A	19.90	172G	224V	20.20
221G	174A	13.90	221G	175V	13.80

Table 14: The proposed contact prediction for 3CN5A from our method. Res1 is the MAIN residue and Res2 and Res3 are the SECONDARY residues in the TRIPLET contact.

Res1	Res2	Distance	Res1	Res3	Distance
10A	59A	6.20	10A	60D	4.70
15V	54C	9.60	15V	55V	13.40
95V	56V	21.00	95V	57C	18.40
49L	102L	20.60	49L	103L	18.80
107S	40K	24.80	107S	41L	26.50
98L	138C	25.20	98L	139L	22.70
89V	147V	15.90	89V	148C	15.70
95V	137V	23.00	95V	138C	25.50
173M	136A	17.10	173M	137V	17.60
			125A	185N	8.10
184A	125A	10.90	184A	126D	14.30
168V	241I	19.70	168V	242L	22.10
233G	175V	15.50	233G	176A	18.20
230V	175V	14.10	230V	176A	16.90
9L	240G	25.40	9L	241I	35.10

Table 15: The proposed contact prediction for 3KCUA from our method. Res1 is the MAIN residue and Res2 and Res3 are the SECONDARY residues in the TRIPLET contact.

Res1	Res2	Distance	Res1	Res3	Distance
57G	109G	1.50	57G	110A	3.50
102L	50L	7.50	102L	51G	6.80
63L	117A	11.20	63L	118S	12.90
195G	137L	6.00	195G	138A	9.00
203L	130A	3.00	203L	131L	1.40
187M	145G	9.70	187M	146V	12.50
216G	265A	8.00	216G	266L	7.30
272G	224A	9.30	272G	225L	12.10
221L	271L	7.20	221L	272G	6.60
42G	290A	10.10	42G	291L	19.30
25G	305A	21.00	25G	306G	99.90
296I	36G	13.50	296I	37L	15.60

Table 16: The proposed contact prediction for 3K3FA from our method. Res1 is the MAIN residue and Res2 and Res3 are the SECONDARY residues in the TRIPLET contact.

Res1	Res2	Distance	Res1	Res2	Distance
375I	413G	5.9	350F	377F	14.6
348L	376V	8.5	280I	293A	17.4
375I	416I	6.8	280I	365L	22.2
337E	380N	7.4	345G	373V	4.3
353V	369G	5.7	337E	407T	20.7
357Y	365L	6.6	352L	368V	8.1
349S	369G	5.4	354G	368V	10.6
356V	365L	6.2	337E	410D	20.2
367V	420V	7	267L	407T	10.2
370V	420V	8.4	267L	369G	31.3
356V	368V	7.7	271I	377F	23.8
337E	381L	6.7	355K	365L	9.3
342L	377F	5.6	280I	339G	28.9
352L	372L	8.6	301L	400V	21.6
345G	376V	5.3	267L	377F	26.1
371S	417Y	5.3	356V	366P	9.6
310N	405V	6.2	275M	377F	23.1
309G	405V	6	271I	369G	27.9
341G	377F	4.6	353V	377F	16.5
349S	373V	6.6	280I	364L	20.2
371S	420V	7	371S	412T	11.4
367V	421A	6.1	370V	412T	14.9
356V	364L	7.9	374L	412T	10.4
306G	409S	5.1	302L	410D	5.7
346L	377F	9.6	337E	372L	16.3
353V	368V	7	280I	290A	16.7
349S	372L	5.9	268V	412T	19.1
305G	409S	5.8	302L	413G	5.2
358W	365L	10.4	349S	377F	11.5
280I	400V	23	350F	372L	9.6

Table 17: The proposed contact prediction for 2ZY9A from TMhit(L/2). The predictions of TMhit(L/5) are the first 24 contacts in this table (vertical counting first).

Res1	Res2	Distance	Res1	Res2	Distance
169T	246A	8.3	130A	165E	21.2
45T	246A	21.5	202H	246A	17.5
169T	242G	4.9	125F	197A	23.5
172L	241L	12.2	131G	246A	23.5
165E	246A	12.9	47L	241L	26.1
44G	93T	6.1	90V	126T	15.2
84F	241L	24.3	48F	169T	16.8
127Q	246A	20.3	125F	241L	19
48F	242G	20.8	48F	165E	19.2
48F	246A	22.7	200I	241L	18.3
48F	126T	13.7	169T	203L	12.5
48F	130A	9.7	166A	244F	9.2
168G	241L	10.3	170A	244F	7.7
170A	197A	9.8	48F	129V	13.3
169T	199L	11.4	45T	173C	16.1
128I	166A	19	176V	241L	15.7
195G	246A	16.6	47L	126T	13.2
131G	165E	18.2	45T	165E	19.4

Table 18: The proposed contact prediction for 2W1PA from TMhit(L/2). The predictions of TMhit(L/5) are the first 15 contacts in this table (vertical counting first).

Res1	Res2	Distance	Res1	Res2	Distance
141T	225A	8.4	98G	137E	21.8
17T	225A	22.5	53M	137E	19.7
141T	221G	5	142F	223A	8.2
144L	220I	12.3	96C	223A	19.7
98G	225A	25.6	16A	98G	7.5
137E	225A	13	20F	221G	21.3
20F	102G	7.3	19L	220I	26.1
172G	225A	17	21L	141T	17.8
147T	174A	7	140G	220I	10.7
52G	220I	24.7	168P	225A	17.9
17T	137E	19.6	138I	223A	9.7
105L	220I	30.5	20F	141T	17.2
99A	225A	24.2	16A	178V	20.4
99A	137E	18.8	144L	174A	5.6
20F	225A	23.4	144L	167A	12.6
98G	221G	22.9	20F	146Y	21
21L	60C	15.8	20F	137E	19.2
17T	107K	16			

Table 19: The proposed contact prediction for 3CN5A from TMhit(L/2). The predictions of TMhit(L/5) are the first 14 contacts in this table (vertical counting first).

Res1	Res2	Distance	Res1	Res2	Distance
46C	237G	23.3	9L	60D	5.4
170P	236I	18.2	139L	180E	5.2
92G	236I	20.7	46C	138C	23.9
55V	238G	17.1	92G	166I	26.5
9L	180E	22.1	9L	56V	12.5
46C	184A	13.7	91F	142L	22
170P	240L	17.2	104M	220M	21.8
100F	176A	22.3	46C	147V	14.7
170P	232I	20.8	53L	236I	24.5
48S	97A	16.4	107S	134I	36.3
55V	97A	17.4	40K	136A	19.1
31T	184A	21	41L	141I	19.9
41L	231T	23.8	52I	141I	21
128K	184A	11.2	48S	238G	18.6
46C	242V	27.8	85N	175V	25.8
54C	97A	13.7	92G	237G	21.1
18S	48S	11.3	169L	232I	24.3
92G	232I	17.9	59A	147V	14.2
40K	140G	18	31T	126D	17.6
54C	238G	17.3	92G	147V	18.4
40K	93N	24.6	16F	43G	15.6
46C	232I	26.1	11I	244L	29.6
53L	232I	24.7	60D	184A	13.4
88N	232I	16.4	40K	233G	22.1
40K	221N	24.8	101V	168V	24.4
16F	237G	21.5	103L	220M	22.5
46C	236I	26.2	48S	243G	25.8
31T	134I	29.6	48S	235I	20.3
44G	175V	4.9	46C	229P	25.9
40K	234N	21.2	52I	231T	21.7
91F	232I	15.2	91F	146M	21.8
14G	238G	18.4	14G	242V	23.1
49L	146M	14.7	53L	220M	30.4
92G	224T	20.1	101V	232I	23.5
15V	97A	5.4	128K	229P	11.5
31T	131H	23.4	167M	244L	20.3
40K	181H	15.7	48S	90Y	19.7

Table 20: The proposed contact prediction for 3KCUA from TMhit(L/2). The predictions of TMhit(L/5) are the first 30 contacts in this table (vertical counting first).

Res1	Res2	Distance	Res1	Res2	Distance
305Y	761L	24.8	838A	969F	80.6
305Y	757N	34.3	839G	935F	21.9
75Q	114V	48.9	763P	935F	66.5
114V	836S	16.1	280F	766I	43.1
280F	935F	33.3	280F	939I	36
842D	935F	82.2	273A	835Q	16.1
763P	838A	32.3	273A	935F	37.9
763P	835Q	25.2	766I	837F	59.5
766I	935F	33.5	79W	836S	41.2
280F	835Q	95.1	273A	758I	32.7
114V	312A	76.1	282I	834I	7.4
115V	836S	19.1	899S	940G	25.9
838A	935F	84.8	115V	279T	22.6
114V	761L	13.5	766I	939I	33
279T	837F	3.8	842D	899S	46.8
274I	836S	99.3	766I	834I	55.5
839G	896F	17.4	114V	271G	15.6
278A	837F	85.8	271G	974F	21.9
280F	762T	38.8	766I	835Q	54.4
759P	835Q	37.8	759P	838A	41.2
112A	276F	25.6	109A	279T	53.2
788T	835Q	49.7	274I	941C	17.8
762T	935F	52.7	278A	836S	84.3
118G	836S	39.3	115V	275L	29.9
766I	943L	38.7	114V	967M	27.5
788T	935F	68.2	271G	836S	101.1
112A	935F	58.2	75Q	785E	29.8
766I	838A	59.7	842D	969F	56.5
273A	762T	29.4	75Q	977D	44.7
112A	279T	33.3	763P	900I	56
75Q	305Y	33.4			

Table 21: The proposed contact prediction for 3IXZA from TMhit(L/2). The predictions of TMhit(L/5) are the first 25 contacts in this table (vertical counting first).

Res1	Res2	Distance	Res1	Res2	Distance
200V	215I	14.4	50L	254F	30.1
28M	194T	14.1	67L	263L	42.4
38L	275A	29.5	200V	219V	19.5
32N	51G	18.3	27V	194T	14.5
44F	195G	12.2	42G	215I	20
38L	271L	29.4	111F	275A	25.7
64T	212F	32.2	202I	210A	5.3
27V	246C	12.2	42G	194T	4.5
53C	275A	28.4	53C	254F	29.8
38L	60A	20	92P	246C	25.1
38L	200V	8.2	27V	265A	18.5
67L	194T	32.7	104Y	197I	28.3
42G	87A	17.1	271L	283L	18.2
200V	274I	27.3	53C	197I	18.8
41A	215I	23	42G	51G	8.6
32N	80F	14.1	79L	194T	22.8
197I	275A	26.9	130A	194T	25.9
44F	51G	4.8	194T	254F	26.1
216G	283L	23.9	41A	197I	8
27V	87A	11	88G	275A	24.4
35T	254F	19.2	200V	244V	14.3
27V	84G	9.3	111F	219V	32.7
32N	81G	12.4	44F	210A	20.7
58T	195G	19.9	113T	246C	22.3
92P	194T	17.7	104Y	271L	33.1
32N	210A	13.4	203L	210A	5
35T	190D	19.4	219V	283L	23.8
60A	219V	34.4	268C	283L	21.8
200V	218L	16.7	42G	271L	28
67L	210A	34	67L	261T	39.2
27V	195G	13.7	141L	195G	28.2
38L	219V	24.7	53C	279F	26.2
60A	215I	32.7	212F	283L	28.5
202I	246C	14.7	41A	219V	26.4
247G	275A	14.5	32N	236L	27.7
38L	223T	29.3	53C	204V	24.2
67L	259I	44	113T	195G	26.7

Table 22: The proposed contact prediction for 3K3FA from TMhit(L/2). The predictions of TMhit(L/5) are the first 44 contacts in this table (vertical counting first).

Res1	Res2	Distance	Res1	Res2	Distance
146V	194T	29.7	104Y	275A	29
35T	212F	15.8	39F	201A	6.5
42G	275A	27.4	41A	271L	31.1
32N	237G	25.1	139T	275A	19.1
32N	213A	16	190D	212F	20
38L	215I	20.6	38L	87A	18.6
35T	283L	32.9	61S	194T	23.7
31N	212F	16.3	104Y	223T	36.1
67L	195G	33.7	143M	195G	25.9
88G	271L	27.5	194T	283L	28
43I	51G	5.4	212F	275A	18.9
38L	201A	5.5	32N	127G	22.3
91L	212F	27.9	29F	194T	17.6
141L	200V	33.4	27V	266L	20.5
42G	246C	21	67L	101M	22.5
32N	284S	36.6	58T	296I	19.9
60A	212F	30	53C	283L	25.4

Table 23: Continue of Table 22. The proposed contact prediction for 3K3FA from TMhit(L/2). The predictions of TMhit(L/5) are the first 44 contacts in this table (vertical counting first).

References

[1] McAllister SR, Floudas CA, alpha-helical topology prediction and generation of distance restraints in membrane proteins. Biophys J 2008;95:5281-5295.