

# Tailoring Contact Based Scoring Functions for Protein Structure Prediction

Rianon Zaman<sup>1</sup>, M. A. Hakim Newton<sup>1,2</sup>, Fereshteh Mataeimoghadam<sup>1</sup>, Abdul Sattar<sup>1,2</sup>

{rianon.zaman, fereshteh.mataeimoghadam}@griffithuni.edu.au, {mahakim.newton,a.sattar}@griffith.edu.au

## Abstract

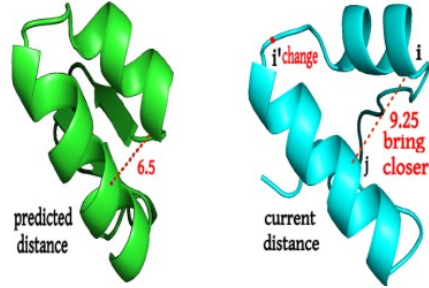
This poster presents a comparison among a number of existing contact based scoring functions within the same Protein Structure Prediction search framework on the same set of benchmark proteins. Moreover, we also propose a number of contact based scoring function variants. We use SSpro8, SpotContact and Spot1D as input for constraints. Our proposed contact based scoring functions help our search algorithm to significantly outperform existing state-of-the-art PSP search algorithms that use contact based scoring functions.

## Context

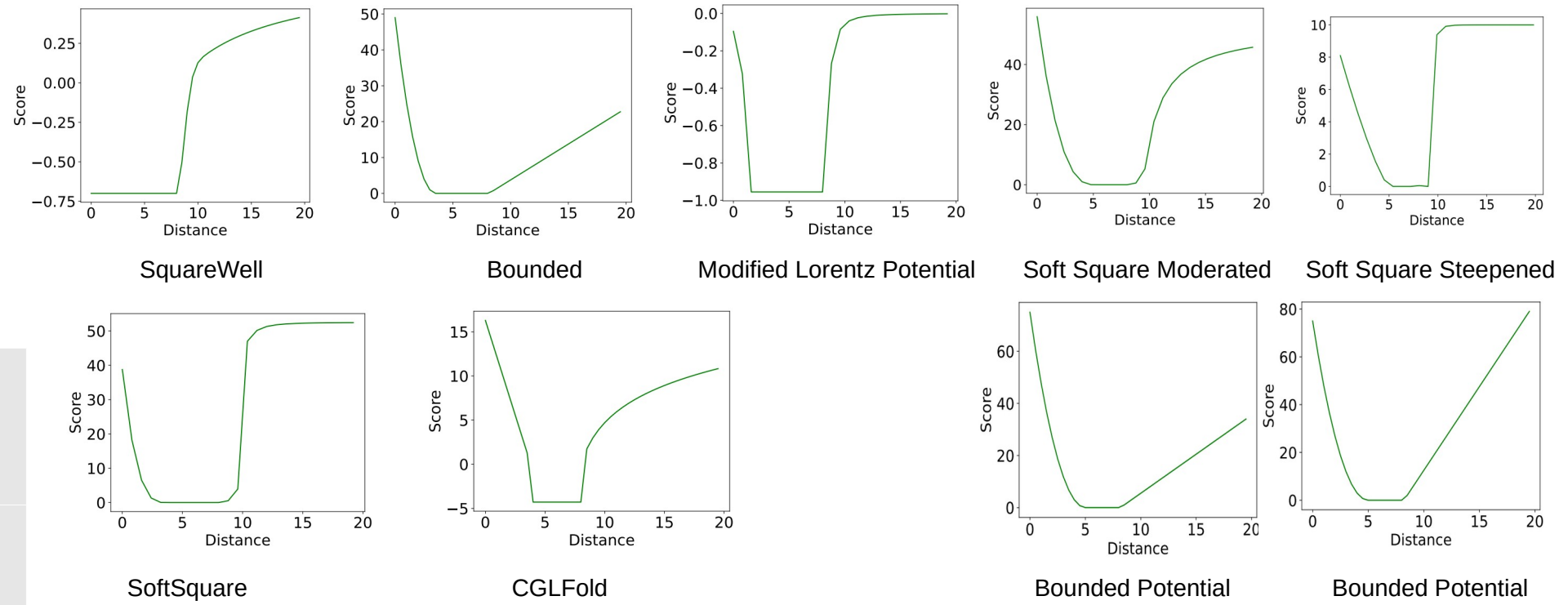
Protein Structure Prediction (PSP) is one of the most challenging issue in the bioinformatics area. PSP is determining the three-dimensional structure of a protein just from its amino acid sequence. There is a huge demand for effective protein structure prediction method as most of the structure is still unknown. From a contact map, we can know whether two residues are within a certain distance or not. In this work, we propose four contact based scoring functions that significantly improves the three dimensional structure of the protein.

## Method Details

17	24	0.925348
16	24	0.924862
16	23	0.889774
13	27	0.875671
16	27	0.870139
13	24	0.860117
13	28	0.732693
13	31	0.683982
9	31	0.559323
17	27	0.540271



Contact Map



Existing Scoring Functions

Proposed Scoring Functions

Number of Protein	Protein Type	Protein Length
39	39 $\alpha$ - 15, $\beta$ - 13, $\alpha\beta$ - 11	42-181

## Experimental Results

Table: Comparison among various existing scoring function

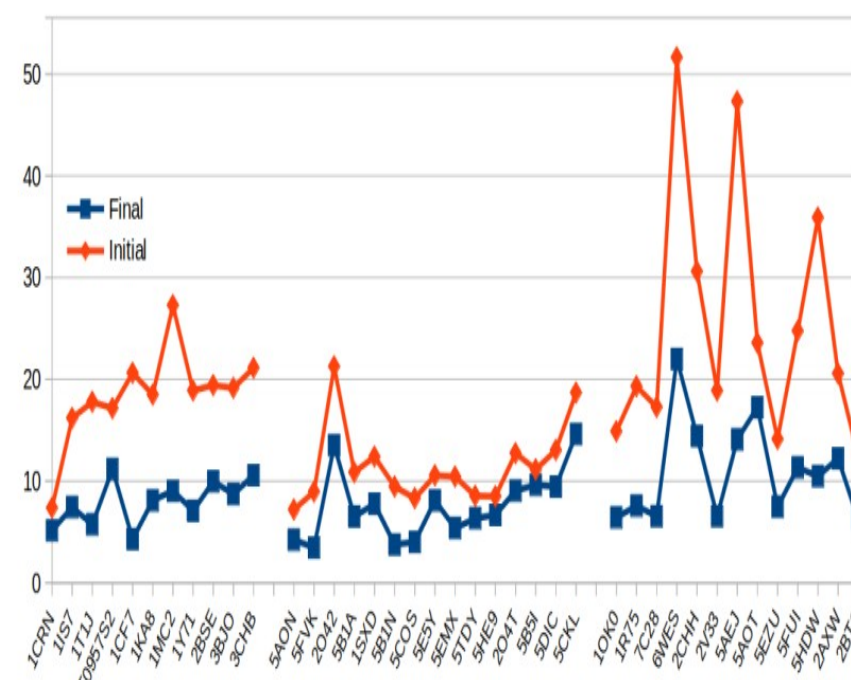
Scoring Function Name	sw	bp	mlp	ss	cf	ssm	sss	bpm	bps
Average RMSD	10.78	9.96	10.79	10.26	10.01	9.84	10.36	9.13	<b>8.78</b>
mean RMSD $\leq 6$	6	5	6	6	5	7	4	5	<b>8</b>
mean RMSD $\leq 9$	14	20	16	15	19	19	18	<b>23</b>	<b>23</b>
mean RMSD $\leq 12$	26	<b>32</b>	26	27	<b>30</b>	29	29	<b>32</b>	<b>32</b>

Table: Mean RMSD and GDT values obtained by our algorithm and state-of-the-art CGLFOLD algorithm.

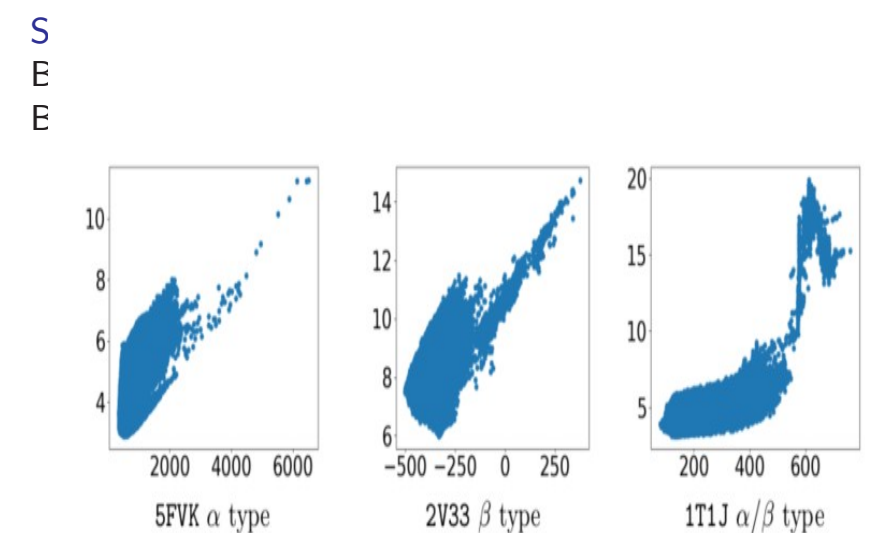
Method Name	Mean RMSD		Mean GDT	
	CGPSPS	CGLFOLD	CGPSPS	CGLFOLD
Mean over all proteins	<b>8.78</b>	9.45	<b>0.409</b>	0.398

Table: Numbers of proteins with mean RMSD values  $\leq$  various threshold values.

Algorithm Name	mean RMSD $\leq 6$				mean RMSD $\leq 9$				mean RMSD $\leq 12$			
	$\alpha$	$\beta$	$\alpha/\beta$	all	$\alpha$	$\beta$	$\alpha/\beta$	all	$\alpha$	$\beta$	$\alpha/\beta$	all
CGPSPS	5	0	3	8	10	6	7	23	13	8	11	32
CGLFOLD	5	0	2	7	8	5	7	20	11	7	10	28

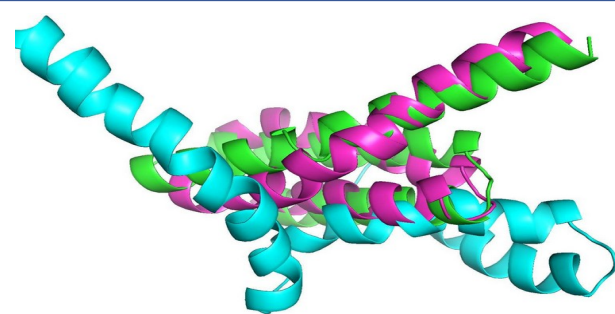


Deviation in mean RMSD values of the initial conformations and the final conformations returned by the search when using the bps scoring function



Scatter plots of bps contact based scores (x-axis) vs RMSD values (y-axis)

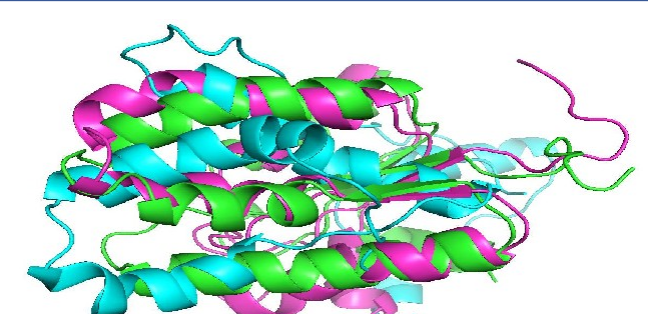
## Protein Structure Generation and Refinement



5FVK Type  $\alpha$  Length 82  
RMSD Init 15.188 Final 3.377

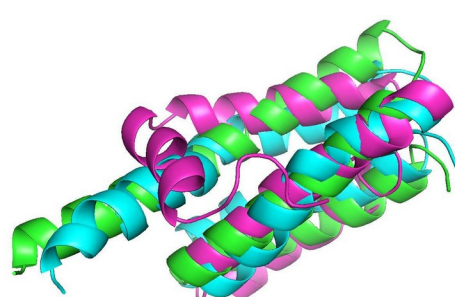


2V33 Type  $\beta$  Length 91  
RMSD Init 18.995 Final 3.602

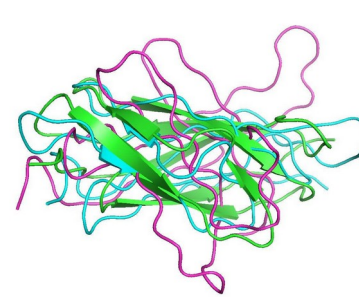


1T1J Type  $\alpha/\beta$  Length 125  
RMSD Init 14.064 Final 4.982

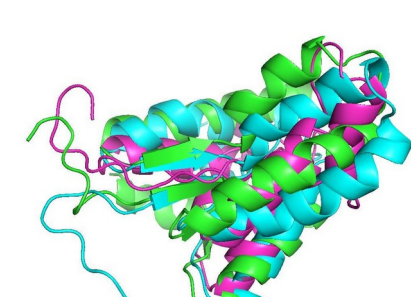
Sample final conformations (magenta) obtained by scoring function bps from initial ones (cyan) w.r.t. native ones (green)



Our RMSD 3.377 GDT 0.613  
CGLFOLD RMSD 4.576 GDT 0.588



Our RMSD 3.602 GDT 0.6036  
CGLFOLD RMSD 6.481 GDT 0.446



Our RMSD 4.982 GDT 0.580  
CGLFOLD RMSD 7.862 GDT 0.4146

Sample best conformations obtained by CGPSPS (cyan) and CGLFOLD (magenta) w.r.t. native conformations (green)