Appendix - A multi-objective approach for the protein structure prediction problem

Sebastián Aliaga-Rojas*, Manuel Villalobos-Cid*, Marcio Dorn†§¶, Mario Inostroza-Ponta*‡

*Departamento de Ingeniería Informática, Universidad de Santiago de Chile (USACH)
†Institute of Informatics, Federal University of Rio Grande do Sul, Porto Alegre, Brazil
§Center of Biotechnology, Federal University of Rio Grande do Sul, Porto Alegre, Brazil
¶National Institute of Science and Technology - Forensic Science, Porto Alegre, RS, Brazil.
{*sebastian.aliaga.r, manuel.villalobos, mario.inostroza*}@usach.cl; mdorn@inf.ufrgs.br
‡ Corresponding author

I. APPENDIX

Algorithm 1 Local Search 1

1: **function** LS1($pop,res\_num,end\_condition,energy\_evals$)
2:     $stop\_condition \leftarrow energy\_evals < end\_condition$
3:     **while** $energy\_evals < end\_condition$ **do**
4:         **for** $i \leftarrow 1$ to $length(pop)$ **do**
5:             **for** $j \leftarrow 1$ to res_num **do**
6:                 $res \leftarrow$ SELECT-RANDOM-RESIDUE
7:                 **for** $k \leftarrow 1$ to 2 **do** //for phi and psi
8:                     $continue \leftarrow True$
9:                     **while** $continue$ & $stop\_condition$ **do**
10:                         $sol \leftarrow$ CHANGE_ANGLE($pop[i]$, res_k)
11:                         $result \leftarrow$ COMPARE($pop[i], sol$)
12:                         **if** $result == dominates$ **then**
13:                             $pop[i] = sol$
14:                         **else**
15:                             $continue = False$
16:                         **end if**
17:                     **end while**
18:                 **end for**
19:             **end for**
20:         **end for**
21:     **end while**
22: **end function**
    Output pop

Algorithm 2 Local Search 2

1: **function** LS2($pop,end\_condition, energy\_evals$)
2:     $stop\_condition \leftarrow energy\_evals < end\_condition$
3:     **while** $stop\_condition$ **do**
4:         **for** $i \leftarrow 1$ to $length(pop)$ **do**
5:             $continue \leftarrow True$
6:             **while** $continue$ & $stop\_condition$ **do**
7:                 $sol \leftarrow$ CHANGE_ANGLES($pop[i]$)
8:                 $result \leftarrow$ COMPARE($sol, pop[i]$)
9:                 **if** $result == dominates$ **then**
10:                     $pop[i] = sol$
11:                 **else**
12:                     $continue = False$
13:                 **end if**
14:             **end while**
15:         **end for**
16:     **end while**
17: **end function**
    Output pop

TABLE I: Second stage parametrisation results

	RMSD					
	NSGA-II			NSGA-II + PLS		
Protein	Run	Best	AV (STD)	Run	Best	AV (STD)
1AB1	3	4.49	7.58 (1.5)	**12**	**2.90**	**8.14 (1.9)**
1ACW	6	3.21	5.47 (1.5)	**27**	**2.51**	**5.51 (1.3)**
1CRN	17	3.62	8.01 (1.5)	**28**	**3.02**	**8.06 (1.7)**
1D5Q	12	2.15	4.84 (1.5)	**31**	**1.99**	**4.52 (1.4)**
1ENH	28	3.26	8.07 (2.2)	**10**	**3.19**	**8.28 (2.0)**
1K43	**25**	**0.84**	**2.73 (1.3)**	6	1.14	2.64 (1.1)
1Q2K	10	2.59	5.53 (1.8)	**29**	**2.40**	**5.22 (1.3)**
1WQC	9	2.71	5.16 (0.9)	**13**	**2.70**	**4.88 (1.0)**
2MR9	25	3.02	6.54 (1.8)	**29**	**2.85**	**6.56 (1.8)**
2MTW	**29**	**2.13**	**5.50 (1.1)**	22	2.37	5.45 (1.0)
2P5K	**7**	**3.33**	**10.20 (2.2)**	27	5.71	10.57 (1.6)
2P81	24	3.61	8.83 (1.2)	**23**	**2.87**	**8.64 (1.3)**
	GDT					
	NSGA-II			NSGA-II + PLS		
Protein	Run	Best	AV (STD)	Run	Best	AV (STD)
1AB1	16	57.83	40.25 (6.0)	**12**	**63.91**	**40.25 (6.3)**
1ACW	21	67.59	47.97 (6.6)	**27**	**73.79**	**47.02 (6.9)**
1CRN	**16**	**62.17**	**40.70 (6.1)**	3	60.87	40.17 (6.3)
1D5Q	27	78.52	55.63 (8.5)	**10**	**80.74**	**57.10 (9.1)**
1ENH	28	64.44	39.37 (6.8)	**22**	**64.81**	**38.62 (6.2)**
1K43	**25**	**90.00**	**71.20 (8.8)**	7	90.00	70.19 (9.6)
1Q2K	**10**	**79.35**	**53.47 (6.9)**	29	75.48	54.10 (6.5)
1WQC	**9**	**75.38**	**51.83 (5.9)**	13	72.31	53.39 (5.2)
2MR9	19	71.82	46.29 (7.3)	**6**	**74.09**	**46.29 (7.9)**
2MTW	**8**	**85.00**	**51.07 (10.3)**	4	83.00	51.78 (10.5)
2P5K	**7**	**62.22**	**31.78 (5.3)**	4	46.35	30.88 (3.5)
2P81	1	68.64	42.42 (7.2)	**23**	**72.27**	**43.00 (7.4)**

TABLE II: Hypervolume results of second stage parametrisation

Top 5 Hypervolume Configurations				
LS Position	Energy eval. %	Res. mod. %	HV AVG	HV Variance
End	60	40	3.2609	0.0023
Intermediate	**70**	**10**	**3.2600**	**0.0022**
End	70	20	3.2592	0.0025
End	70	30	3.2608	0.0035
End	80	70	3.2612	0.0030

TABLE III: Best RMSD and GDT Results

	Best RMSD			Best GDT		
Protein	Run	RMSD	AV(STD)	Run	GDT	AV(STD)
1AB1	11b	4.19	7.72 (1.5)	26b	57.83	41.26 (5.7)
1ACW	27b	2.85	5.27 (1.5)	27b	73.79	48.50 (8.0)
1CRN	23a	4.19	8.01 (1.7)	3a	61.74	40.33 (6.2)
1D5Q	22a	2.03	4.29 (1.5)	21a	79.26	58.62 (9.5)
1ENH	14a	2.81	7.90 (2.0)	1a	64.44	39.25 (6.8)
1K43	4a	1.05	2.39 (0.9)	13a	88.57	72.10 (7.9)
1L2Y	16a	1.20	3.19 (1.0)	26a	91.00	67.42 (9.2)
1Q2K	26a	2.21	5.26 (1.6)	26a	81.94	55.14 (7.6)
1ROP	16a	2.33	5.68 (2.4)	16a	74.29	48.45 (7.1)
1UTG	5a	3.61	10.13 (2.3)	5a	54.29	31.90 (4.7)
1WQC	2a	2.15	4.69 (1.0)	3a	80.00	55.65 (6.9)
1ZDD	17a	2.85	6.06 (1.2)	17a	69.41	45.59 (6.7)
2MR9	3b	2.70	6.66 (1.7)	3b	72.27	45.89 (7.1)
2MTW	26b	2.33	5.40 (1.0)	26b	81.00	52.48 (10.1)
2P5K	7b	5.30	10.59 (1.8)	2b	46.67	32.30 (3.9)
2P6J	1b	3.93	7.70 (2.2)	1b	62.31	38.20 (5.3)
2P81	24b	3.33	8.65 (1.1)	6b	69.09	42.68 (7.1)
2PMR	23b	3.50	8.58 (3.5)	23b	64.21	38.39 (6.3)
3V1A	26b	2.14	5.71 (2.2)	26b	75.00	46.28 (7.6)

TABLE IV: Best Talaris2013 Results

	Best Talaris2013				
Protein	Run	Talaris	SASA	RMSD	GDT
1AB1	22b	2996.68	3632.25	11.28	31.30
1ACW	12b	2337.38	2786.70	10.04	35.86
1CRN	6a	2940.83	3854.56	11.40	32.61
1D5Q	28a	-14.06	2955.70	7.57	54.07
1ENH	8b	1544.07	6100.20	17.28	35.93
1K43	19a	2.88	1951.53	2.10	75.71
1L2Y	24a	1683.07	1678.09	4.66	49.00
1Q2K	18a	12.17	3594.09	10.71	52.26
1ROP	6a	63.84	5791.26	21.76	47.50
1UTG	1a	8731.70	6841.96	18.95	28.29
1WQC	14a	1745.45	2957.19	8.14	48.46
1ZDD	27a	1347.83	3743.88	9.00	50.59
2MR9	3b	2351.47	4237.64	9.05	45.91
2MTW	6b	-6.90	2336.00	2.48	77.00
2P5K	28b	2236.27	6894.15	18.04	28.25
2P6J	7b	578.22	5915.96	16.18	32.69
2P81	22b	192.40	5351.45	14.68	39.09
2PMR	6b	556.72	7772.89	21.01	35.53
3V1A	14b	1451.38	4982.18	14.91	43.75

TABLE V: Best SASA Results

	Best SASA				
Protein	Run	Talaris	SASA	RMSD	GDT
1AB1	3b	149253.56	1420.41	9.02	28.26
1ACW	15b	56975.89	1328.49	6.22	39.31
1CRN	21a	140885.90	1470.26	8.99	34.35
1D5Q	30a	47754.19	1109.76	4.80	52.59
1ENH	6a	291005.76	1572.09	9.35	33.70
1K43	26a	38320.48	895.70	3.20	67.14
1L2Y	1a	41493.60	967.32	5.22	54.00
1Q2K	24a	64140.67	1295.78	8.03	44.52
1ROP	18a	149993.75	2224.37	5.36	50.71
1UTG	14b	333410.11	2004.18	10.40	29.43
1WQC	25a	47000.00	1201.27	5.21	51.54
1ZDD	24a	68512.43	1443.07	7.78	35.29
2MR9	22b	122916.37	1443.22	7.58	38.18
2MTW	18b	38763.87	1080.14	8.48	34.00
2P5K	4b	250678.41	1848.32	9.22	26.67
2P6J	19b	203751.83	1591.80	7.87	33.08
2P81	13b	136546.05	1536.45	10.06	29.55
2PMR	29b	324657.27	2344.35	5.66	36.05
3V1A	3b	85568.72	1797.34	10.61	38.33

TABLE VI: Evaluated configurations in the first stage of parametrisation. Configurations without letter b in their names, start the LS at the first generation, and configurations with letter b in their names, start at 80% of energy evaluations (The value of this parameter is actually defined in the third parametrisation stage)

Configuration	Crossover	Mutation	Local Search	Residues Modified
1	0.5	0.1	None	0
2	0.5	0.1	LS1	0.5
3	0.5	0.1	LS2	0.5
4	0.5	0.1	PLS	0.5
5	0.9	0.1	None	0
6	0.9	0.1	LS1	0.5
7	0.9	0.1	LS2	0.5
8	0.9	0.1	PLS	0.5
9	0.5	0.2	None	0
10	0.5	0.2	LS1	0.5
11	0.5	0.2	LS2	0.5
12	0.5	0.2	PLS	0.5
13	0.9	0.2	None	0
14	0.9	0.2	LS1	0.5
15	0.9	0.2	LS2	0.5
16	0.9	0.2	PLS	0.5
2b	0.5	0.1	LS1	0.5
3b	0.5	0.1	LS2	0.5
4b	0.5	0.1	PLS	0.5
6b	0.9	0.1	LS1	0.5
7b	0.9	0.1	LS2	0.5
8b	0.9	0.1	PLS	0.5
10b	0.5	0.2	LS1	0.5
11b	0.5	0.2	LS2	0.5
12b	0.5	0.2	PLS	0.5
14b	0.9	0.2	LS1	0.5
15b	0.9	0.2	LS2	0.5
16b	0.9	0.2	PLS	0.5

Config.	1WQC				2MTW				2P8I				Totales				
	HV	STD	ranking 1	ranking 2	HV	STD	ranking 1	ranking 2	HV	STD	ranking 1	ranking 2	Total Ranking 1	Total Ranking 2	total HV	STD	Score
1	3.05	0.03	0	0	2.92	0.04	0	0	3.27	0.03	0	3	0	3	3.08	0.03	3.77
2	3.02	0.03	0	0	2.88	0.04	0	0	3.22	0.03	7	0	7	0	3.04	0.03	6.28
3	3.04	0.03	0	0	2.90	0.06	0	0	3.24	0.01	1	0	1	0	3.06	0.04	4.77
4	3.02	0.04	0	0	2.88	0.06	0	0	3.21	0.03	16	0	16	0	3.04	0.04	8.53
5	3.07	0.04	0	4	2.93	0.09	0	1	3.28	0.04	0	5	0	10	3.09	0.05	2.02
6	3.03	0.04	0	0	2.86	0.05	3	0	3.24	0.03	2	0	5	0	3.04	0.04	5.78
7	3.04	0.03	0	0	2.90	0.12	0	0	3.25	0.05	0	0	0	0	3.06	0.07	4.53
8	3.01	0.03	7	0	2.84	0.06	5	0	3.23	0.04	2	0	14	0	3.03	0.05	8.03
9	3.06	0.05	0	1	2.93	0.07	0	1	3.27	0.04	0	1	0	3	3.09	0.05	3.77
10	3.00	0.05	7	0	2.86	0.07	3	0	3.23	0.05	5	0	15	0	3.03	0.06	8.29
11	3.05	0.04	0	0	2.92	0.10	0	0	3.23	0.03	5	0	5	0	3.06	0.06	5.78
12	3.01	0.03	7	0	2.84	0.05	11	0	3.22	0.05	7	0	25	0	3.02	0.04	10.78
13	**3.07**	**0.06**	**0**	**4**	**2.97**	**0.10**	**0**	**6**	**3.28**	**0.04**	**0**	**8**	**0**	**18**	**3.11**	**0.06**	**0.02**
14	3.02	0.04	0	0	2.87	0.04	3	0	3.21	0.04	11	0	14	0	3.03	0.04	8.03
15	3.038	0.041	0	0	2.94	0.13	0	1	3.23	0.03	5	0	5	1	3.07	0.07	5.53
16	3.00	0.05	11	0	2.85	0.07	4	0	3.22	0.03	11	0	26	0	3.02	0.05	11.04
2b	3.06	0.62	0	0	2.94	0.35	0	0	3.27	0.88	0	3	0	4	3.09	0.62	3.66
3b	3.06	0.84	0	1	2.93	0.33	0	0	3.27	0.65	0	3	0	4	3.09	0.60	3.66
4b	3.06	0.70	0	4	2.95	0.43	0	1	3.28	1.21	0	8	0	13	3.10	0.78	1.45
6b	3.06	0.67	0	1	2.99	0.26	0	6	3.27	1.12	0	1	0	8	3.10	0.68	2.67
7b	3.04	0.90	0	0	2.97	0.35	0	6	3.29	0.78	0	10	0	16	3.10	0.68	0.67
8b	**3.07**	**0.78**	**0**	**4**	**2.95**	**0.33**	**0**	**2**	**3.29**	**0.89**	**0**	**11**	**0**	**17**	**3.10**	**0.66**	**0.42**
10b	3.06	0.59	0	0	2.94	0.34	0	0	3.27	0.66	0	5	0	5	3.09	0.53	3.39
11b	3.05	0.64	0	0	2.91	0.40	0	0	3.27	1.07	0	3	0	3	3.08	0.71	3.94
12b	3.07	0.68	0	4	2.94	0.43	0	0	3.27	0.64	0	1	0	5	3.09	0.58	3.40
14b	3.07	0.68	0	4	2.95	0.32	0	1	3.27	0.64	0	1	0	6	3.09	0.55	3.14
15b	3.06	0.65	0	4	2.96	0.50	0	3	3.28	0.79	0	8	0	15	3.10	0.65	0.92
16b	3.05	0.75	0	0	2.95	0.39	0	1	3.27	0.82	0	1	0	2	3.09	0.65	4.17

TABLE VII: Hypervolume results of first stage parametrisation