Extending HP Lattice Model with Non-Local Hydration

Jarmo T. Alander
Jarmo.Alander@uwasa.fi
Department of Information Processing and Production Economics,
University of Vaasa, PO Box 700, FIN-65101 Vaasa, Finland

Keywords: global optimisation, hydration, lattice model, protein folding, selforganization

1 Introduction

It seems obvious that neither a pure random optimisation approach nor a pure deterministic one is able to solve the protein folding problem. A stochastic approach is too slow. Most of the effort is lost in generating hopelessly denatured conformations. A pure deterministic approach is able to avoid most of the bad conformations, but it usually has to process much too many conformations before the native one is found. So many, that too long a time is needed to reach the solution.

2 Method and Results

In this work we have attacked the famous protein folding problem by extending a simple HP lattice model [5] by a longer range hydration interaction. The conformation energy optimisation was done using several population based methods: the most fit conformations of a set of randomly generated trial conformations were re-evaluated by slightly modifying them in a Monte Carlo fashion [1,2].

It is proposed that by combining features of both stochastic and deterministic search approaches, we can effectively avoid generating hopelessly bad conformations while not trapping to local extremes. Our approach using in addition a longer range hydration interaction turns out to solve this problem in a way that is not only computationally efficient but also physically consistent [3,4].

The major physical contribution of this work is the modeling and testing of hydration in a lattice model framework. The results indicate that hydration, or similar phenomenon causing longer range interaction, is vital in overcoming the Levinthal's paradox [6]. More specifically hydration interaction brings the distant parts of a sequence into close association. Hence hydration drives the sequence into more compact, molten globule like, conformations and thus dramatically reduces the otherwise intractable number of possible conformations into a tractable one, out of which the random search by thermal motion is able to reveal the native state.

The effect of hydration on conformation search was tested by searching the native conformations of several test sequences (Table 1). As can be seen, in general hydration seems to be beneficial. Without it our algorithm needs considerably more local steps. Similarly it seems to be beneficial to count HH-contacts, but their weight used in the free energy formula does not seem to be so critical. Hence it seems that both local and longer range interactions should be modelled in order to have a realistic fast folding process. Moreover a small population size seems to be better than large. This is easy to explain: the bigger the population size the more time it takes to process the best trials. Population based search is beneficial because search without a population of trials is clearly less efficient than that using a population, however.

References

[1] Alander, J.T., Protein folding problem, an algorithmic lattice model approach, Licenciate thesis, University of Helsinki, Department of Physics, 1999. ftp.uwasa.fi/cs/report99-1/Main.ps.Z

Table 1: The average number of local steps $\hat{n}_e$ needed to find the native state of several test sequences. *Psize* = population size, w_{HH} = HH interaction weight, r_h = hydration radius, σ^2 = variance, and # = number of samples without outlyers.

12-mer = HPPHPHPHPHPH $\max(n_e) = 50,000$

Psize	$\hat{n}_e$	σ	#
1	10934	10192	23
4	7614	8748	23
16	5504	7526	27
64	8671	10078	25
256	14563	5980	25

w_{HH}	$\hat{n}_e$	σ	#
0	12143	9209	11
0.04	8488	8376	23
0.08	9958	10782	25
0.12	8145	6559	23
0.16	9030	7622	19
0.2	9975	10926	22

r_h	$\hat{n}_e$	σ	#
0	15026	12214	20
1	14415	10337	18
2	9818	8775	26
3	6348	5664	29
4	5224	4491	30

16-mer = HPPHHPPHHPPHHPPH $\max(n_e) = 100,000$

Psize	$\hat{n}_e$	σ	#
1	12152	8054	20
4	13295	16788	22
16	18767	21064	25
64	26907	21954	24
256	32588	17638	23

w_{HH}	$\hat{n}_e$	σ	#
0.04	20660	16885	23
0.08	23377	22903	23
0.12	19452	15550	24
0.16	21208	18723	23
0.2	20595	23454	21

r_h	$\hat{n}_e$	σ	#
0	48003	19462	6
1	35101	23126	9
2	25378	21306	25
3	17205	18561	25
4	15510	13166	24
5	14369	12962	25

27-mer = HPPPPHHHHPPHPHPHHHPHPPHHPPH $\max(n_e) = 1,000,000$ for all 27-mers

Psize	$\hat{n}_e$	σ	#
1	285237	239746	7
4	154075	164479	6
16	422304	289279	7
64	136473	90155	7
256	340112	161530	5

w_{HH}	$\hat{n}_e$	σ	#
0.2	15648	NaN	1
0.3	269123	187082	10
0.4	279360	250160	11
0.5	275328	250497	10

r_h	$\hat{n}_e$	σ	#
1	245300	173861	5
2	327373	267465	11
3	178847	151196	8
4	284343	250876	8

27merRep = PHPPHPPHPPHPPHPPHPPHPPHPPHP

Psize	$\hat{n}_e$	σ	#
1	165528	140474	12
4	208232	253014	10
16	211856	225548	7
64	248643	263268	10
256	267002	160215	4

w_{HH}	$\hat{n}_e$	σ	#
0.1	218663	286604	8
0.2	265913	251109	8
0.3	148169	112212	8
0.4	256811	211540	10
0.5	164001	174357	9

r_h	$\hat{n}_e$	σ	#
1	618974	240188	2
2	187828	165019	16
3	97918	68007	12
4	283683	264614	13

27merPivot = PHPPPPPPHPPPPPHPPPPPHPPPPPH

Psize	$\hat{n}_e$	σ	#
1	60870	53076	20
4	73811	69293	21
16	91637	130928	21
64	128170	224794	22
256	123036	185878	22

w_{HH}	$\hat{n}_e$	σ	#
0.1	97998	133274	21
0.2	108720	186891	24
0.3	67664	90892	21
0.4	137013	212493	21
0.5	65940	54743	19

r_h	$\hat{n}_e$	σ	#
1	419645	236684	8
2	109265	179016	23
3	72949	86782	25
4	50149	49238	25
5	50824	53014	25

[2] Alander, J.T., Indexed bibliography of genetic algorithms in chemistry and physics, Report, University of Vaasa, 2000. `ftp.uwasa.fi/cs/report94-1/gaCHEMPHYSbib.ps.Z`

[3] Daura, X., Mark, A.E., and van Gunsteren, W.F., Peptide folding simulations: no solvent required? *Computer Physics Communication*, 123(1-3):97–102, 1999.

[4] Garde, S., Hummer, G., García, A.E., Pratt, L.R., and Paulaitis, M.E., Hydrophobic hydration, *Physical Review E*, 53(5):R4310–R4313, 1996.

[5] Lau, K.F. and Dill, K.A., A lattice statistical mechanics model of the conformational and sequence spaces of proteins, *Macromolecules*, 22(10):3986–3997, 1989.

[6] Levinthal, C., [the proper reference to Levinthal's paradox]. *Proc. of a Meeting held at Allerton House, Monticello, IL*, pages 22–24, Monticello, IL, 1969. University of Illinois Press, Urbana.