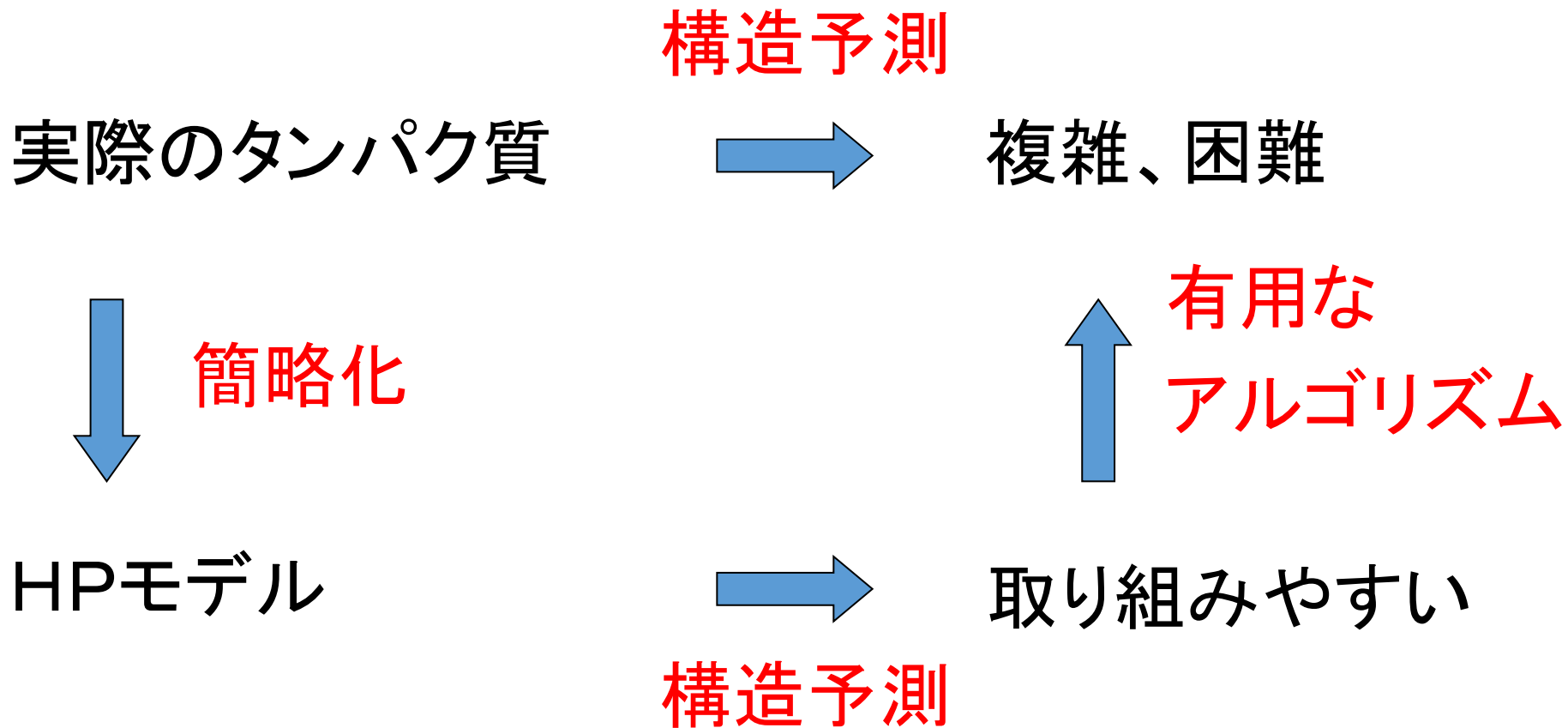


タンパク質構造予測

- タンパク質は、20種類のアミノ酸から成っており、その線形結合で表すことができる。線形結合から、一意に立体構造を決めることができる。
- アミノ酸配列だけではそのタンパク質の生化学的機能を特定するには不十分であり、本来の三次元構造を得ることで生化学的機能が予測できる。
- **AIの手法**やコンピュータネットワークが駆使されている

タンパク質構造予測とHPモデル



タンパク質を構成するアミノ酸

グリシン	Gly	バリン	Val
セリン	Ser	システイン	Cys
グルタミン	Gln	イソロイシン	Ile
フェニルアラニン	Phe	トリプトファン	Trp
グルタミン酸	Glu	リジン	Lys
アラニン	Ala	ロイシン	Leu
スレオニン	Thr	アスパラギン	Asn
チロシン	Tyr	メチオニン	Met
プロリン	Pro	アスパラギン酸	Asp
ヒスチジン	His	アルギニン	Arg



親水性



疎水性

HPモデルのルール

- アミノ酸をH(hydrophobic,疎水性、非極性アミノ酸)とP(polar,親水性、極性アミノ酸)のいずれかに分ける。
- Hは、水を嫌い、互いに引き付けあう
- HとHが隣り合うと、HH結合が生まれる。HH結合はより低いエネルギーを取る。よってHH結合が最も多い構造を最良とする。これをエネルギー最小化法という。

タンパク質→HPモデル

例:

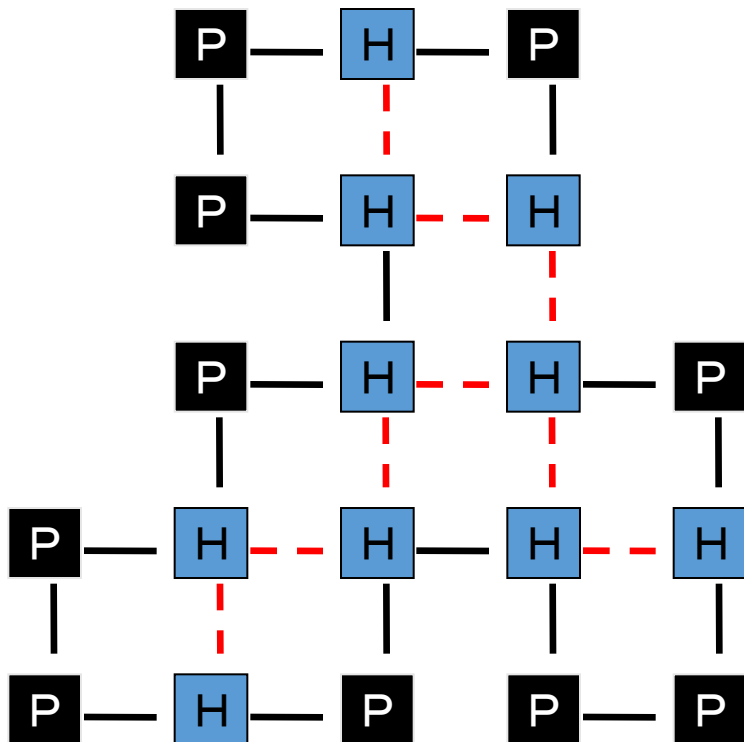
タンパク質: Lys-Val-Arg-Leu- Ile- Asp-Glu-Phe



HPモデル: H - P - H - P - P - H - H - P

HPモデル二次元格子構造の例

H-P-H-P-P-H-H-P-H-P-P-H-P-H-H-P-P-H-P-H



左図では、HH結合
は9個なので、

9点

点数が高い方が
良い結果である。

実験設定

- 配列は長さの違う11種類を用意する。
- 配列は全て最適解が既知のものである。
- GA, GP, ACO, NNなどを用いる。
 - PERM (Pruned Enriched Rosenbluth Method)

実験結果

長さ	最適解	GP	GA	ACO	EMC	PERM
20	9	9	9	9	9	9
24	9	9	9	9	9	9
25	8	8	8	8	8	8
36	14	14	14	14	14	14
48	23	23	23	23	23	23
50	21	20	21	21	21	20
60	36	35	34	36	35	35
64	42	42	37	42	39	38

実験結果

長さ	最適解	GP	GA	ACO	EMC	PERM
85	53	52		51	52	53
99	48			47		48
100	50			47		50

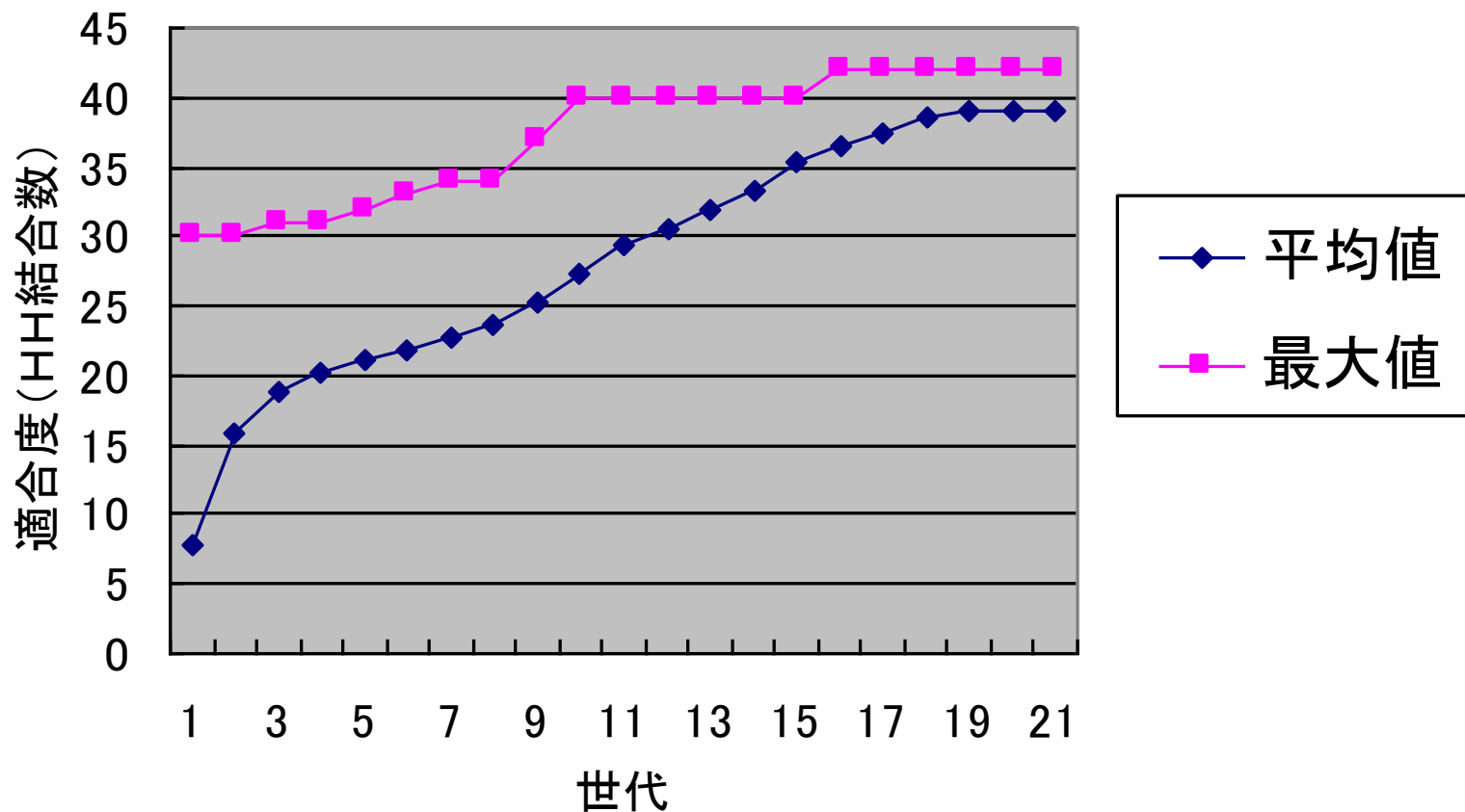
ACO (Ant Colony Optimization)

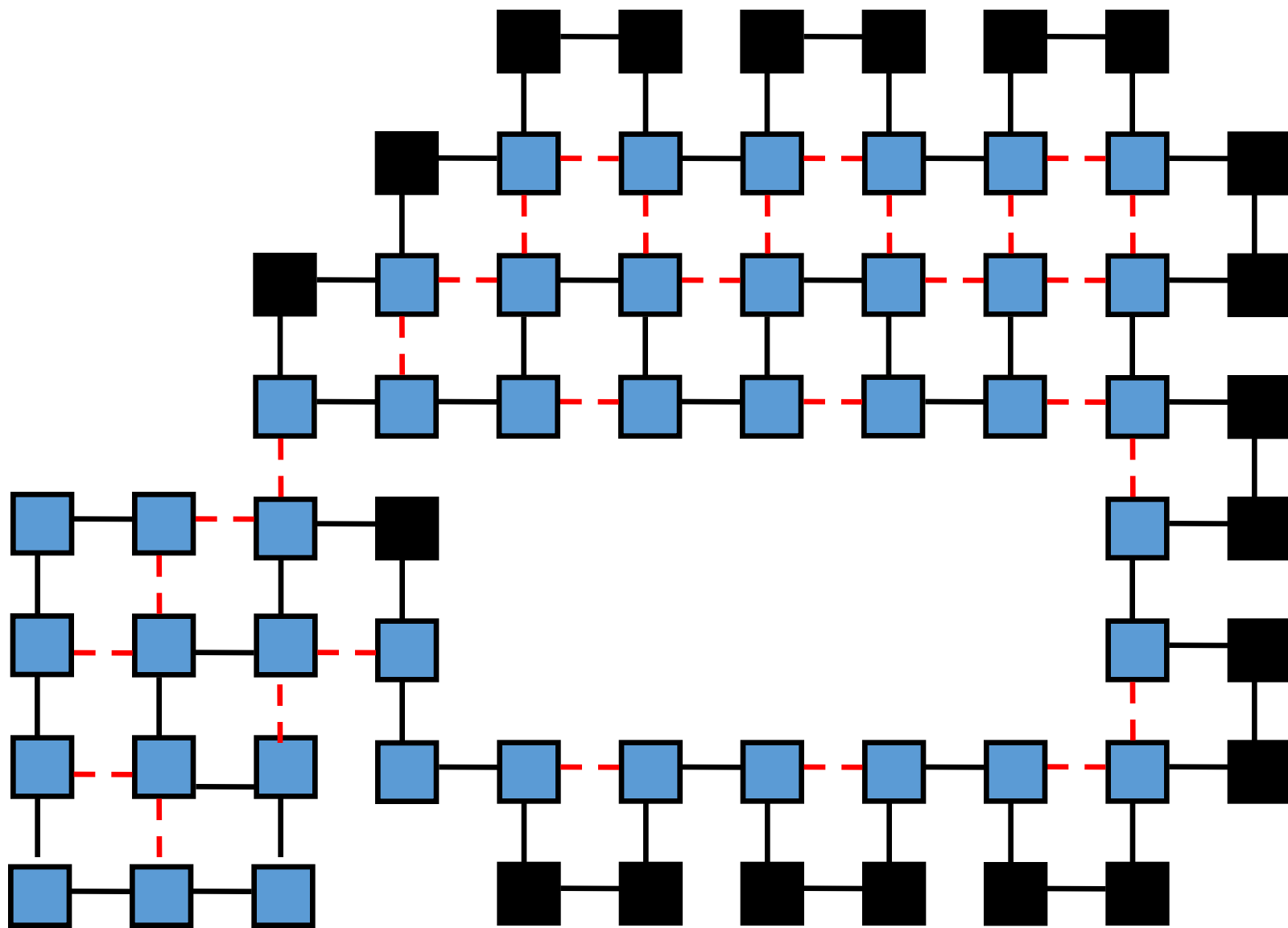
GA (Genetic Algorithm)

EMC (Evolutionary Monte Carlo)

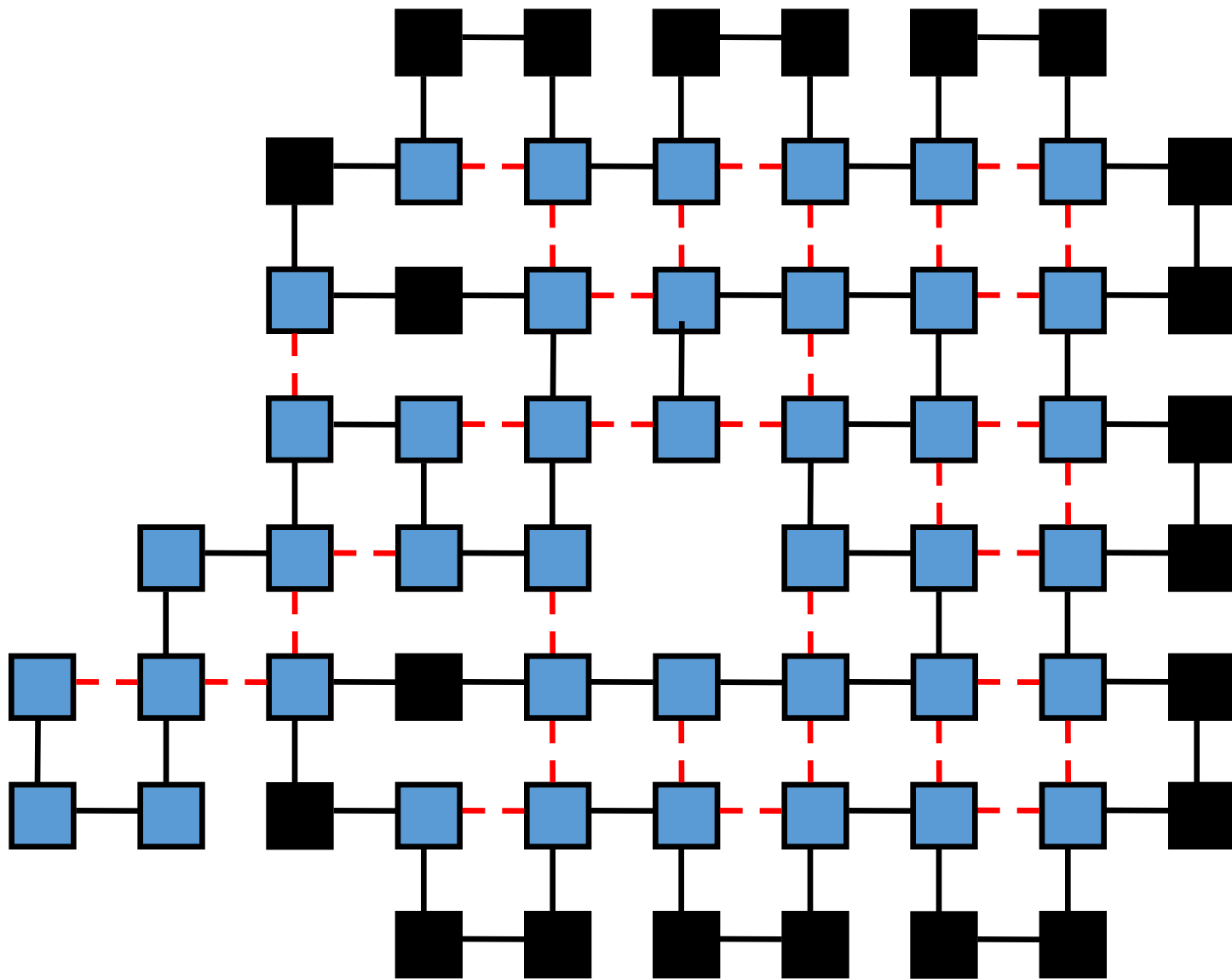
PERM (Pruned Enriched Rosenbluth Method)

学習の様子

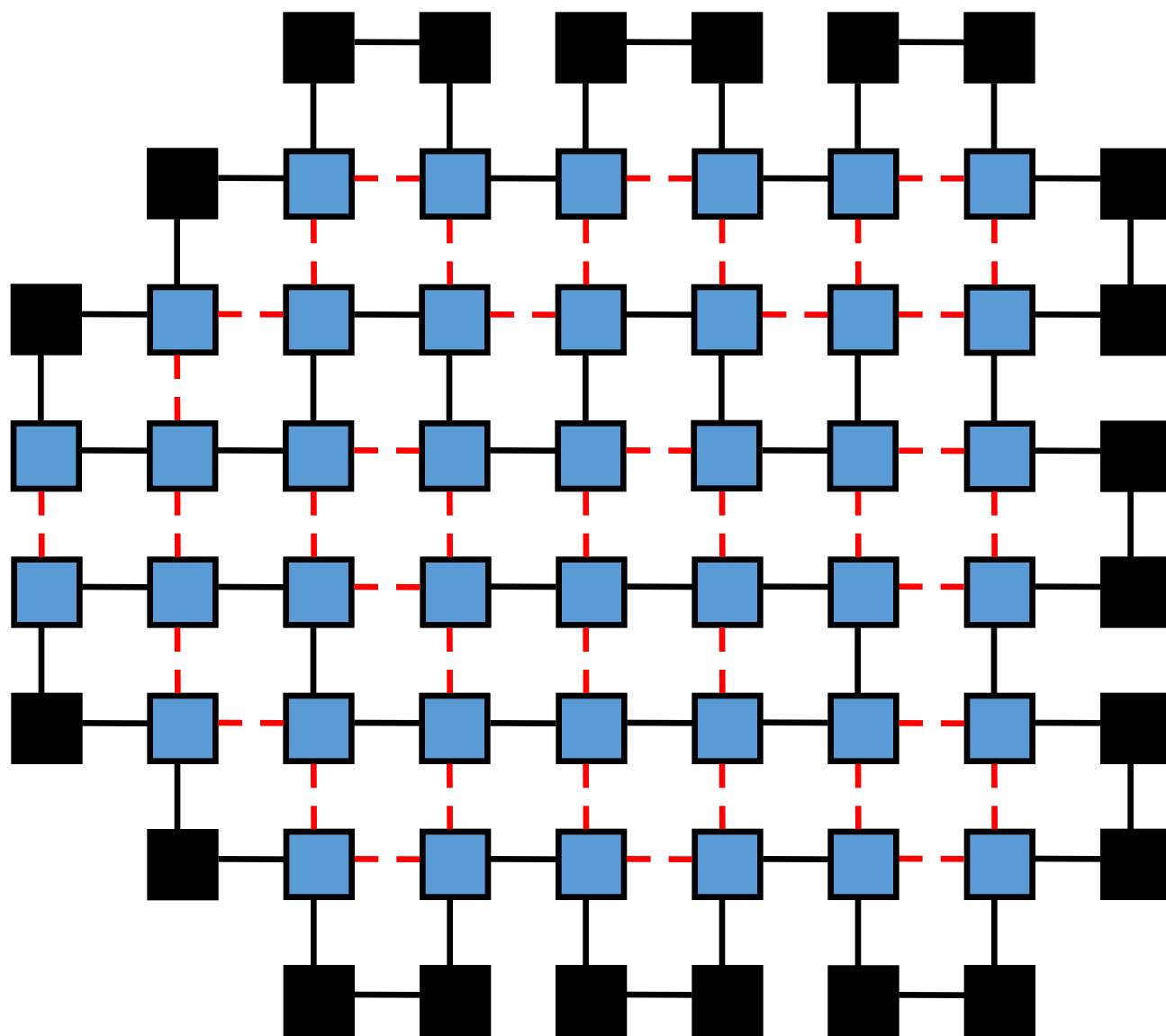




1世代目(30点)



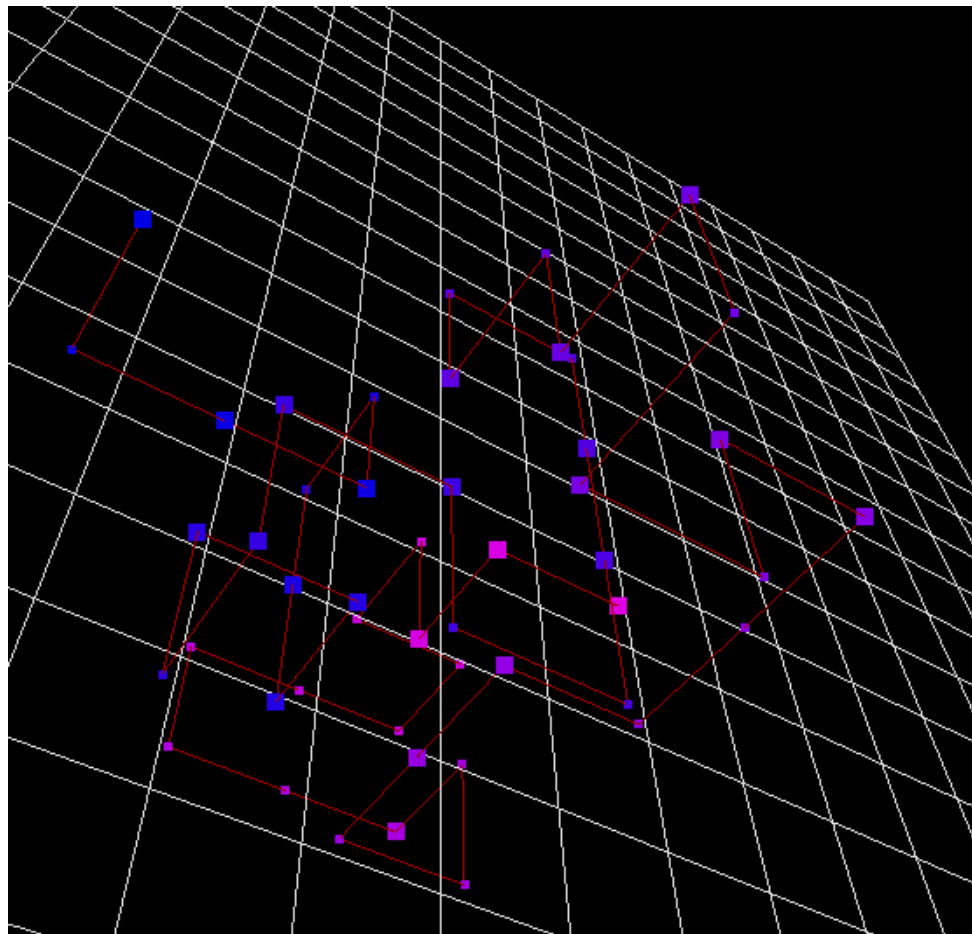
8世代目(34点)



16代目(42点、最適解)

HP問題の3次元版を解いてもよい

- 可視化



配列1について計算途中の構造の例。xy平面上に示した格子とともに表示してある。左上にある青い点が1番の原子で、以降ピンク色が濃いものほど後の番号の原子を表す。点が大きいものはH、小さいものはPを表している。この構造は制約条件を満たしており、HH結合数は21である。