計算統計力学で生命の進化について何が言えるか: 統計力学の手法で稀なものを探す

菊池誠 (阪大サイバー、阪大理)

物理学会新潟支部例会 (2023.12.2)

Outline

1. Introduction: 統計力学と生命現象
2. レアイベント・サンプリング: Multicanonical MC と魔方陣
3. 遺伝子制御ネットワークの変異に対する頑健性による表現型選択
4. タンパク質の機能と折れたたみの共進化
5. まとめ: 生物進化の普遍性探求と統計力学

Introduction: 統計力学と生命現象

Introduction 1

統計力学と生命現象

- 統計力学
  - 与えられた外部条件下で最もありふれた状態を求める
- 生命現象
  - 与えられた環境に適応した状態
  - 物質としては稀な状態
    - 物質の最もありふれた状態は生きていない → 進化の産物
  - 非平衡 (広い意味で) 定常系

なんとなく相性はよくなさそう

Introduction 2

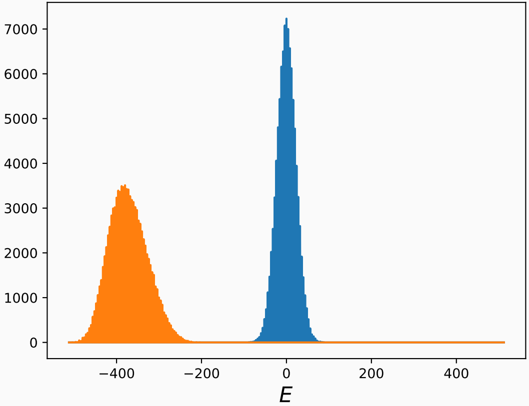
生物物理学の理論は何をするか

- 生体物質の物性と機能
    - 電子状態、生体高分子の統計力学と非線形動力学
        - 進化でできた物質の物性
- 細胞の集団運動や形態形成
    - アクティブ・マター関連で最近盛ん
- 生命の発生と原理
    - シュレディンガー的な問題意識
- 進化の物理学
    - 今日の話

レアイベント・サンプリング: Multicanonical MCと魔方陣

Rare Event Sampling 1

ありふれた状態を求めるはずの統計力学で稀なものを見つける



16^2 Ising model の Random sampling と Markov-Chain Monte Carlo (T_c)

Rare Event Sampling 2

- Random sampling も MCMC も狭い分布しか得られない → 稀な状態は作れない
- MCMC(Metropolis 法) では詳細釣り合いを仮定してカノニカル分布を作る

$$W_{ij}P(E_j) = W_{ji}P(E_i)$$

$$P(E) \propto e^{-\beta E}$$

- 実は $P(E)$ は好きに設定できる

$$P(E) \propto e^{-f(E)}$$

として、任意の分布が生成できる

Rare Event Sampling 3

Multicanonical MC

- $$e^{-f(E)} \sim \frac{1}{\Omega(E)}$$
  $$\text{or } f(E) \propto S(E)$$
  となるように重み $f(E)$ を決める
  $\rightarrow$ あらゆるエネルギーが均等に出現
  $\rightarrow$ 稀な状態も出現する
- $\Omega(E)$ がわかっているなら苦労しない
  - だいたいでよければ学習で決められる
    - B.A. Berg and T. Neuhaus: PRL **68** (1992) 9
    - J. Lee: PRL **71** (1993) 211

Rare Event Sampling 4

Multicanonical MC (cont.)

- MC で得られるエネルギーヒストグラム $H(E)$ から状態密度を推定

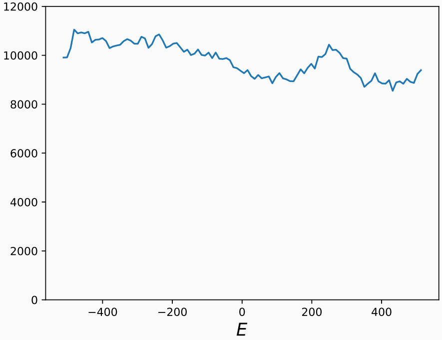
$$\Omega(E) \propto H(E)e^{f(E)}$$

  $\rightarrow$ エントロピーが求められる

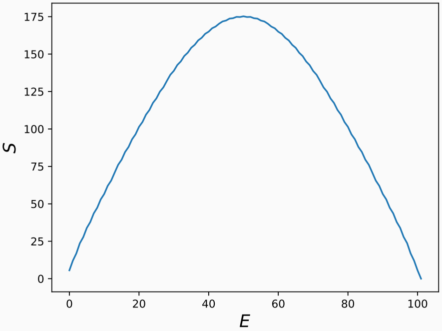
- 学習の方法はなんでも構わない
  - よく使われるのは Wang-Landau 法
    - F. Wang and D.P. Landau: PRL **86** (2001) 2050

Rare Event Sampling 4

16^2 Ising model の計算例



ヒストグラム: 短い計算なのでこの程度の平坦さ



エントロピー

Rare Event Sampling 5

物理的エネルギー以外への応用
なんでもいいのでエネルギーとみなせば適用できる

- ランダム行列の固有値分布
  N. Saito, Y. Iba and K. Hukushima: PRE **82** (2010) 031142
- カオス系の中の周期軌道探索
  A. Kitajima and Y. Iba: Compt. Phys. Comm. **182** (2011) 251
- 結合カオス系の安定性 (遺伝子制御ネットワークを念頭に)
  N. Saito and MK: New J. Phys. **15** (2013) 053037
- 魔方陣の数

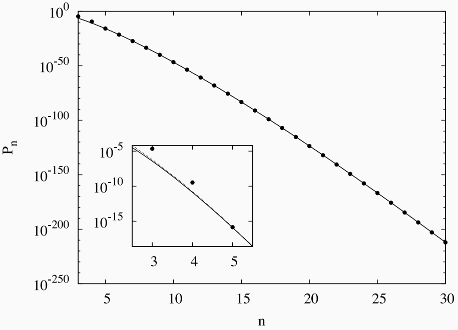
  A. Kitajima and MK: PLOS One **10** (2015) e0125062

Rare Event Sampling: 魔方陣1

4	9	2
3	5	7
8	1	6

- 魔方陣とは: 縦横斜めの数の和が同じ
- 9 マスはこれひとつしかない (回転と裏返しで 8 通り)
- 16 マスは 880
- 25 マスは 275305224
- 36 マスは 17753889189701385264 (preliminary)
    - 正確に数えられる限界。急激に増える!

Rare Event Sampling: 魔方陣2



魔方陣の次数と出現確率

- 魔方陣度 (魔方陣との近さ) を定義してエネルギーとみなし、ランダムな配置から魔方陣までを均等にサンプリング
- 900 マスの魔方陣は 6.6×10^{2056}
- ランダム配置が魔方陣になる確率は 7.8×10^{-213}
  - Numerous but Rare!

Rare Event Sampling: 魔方陣3

693	635	356	574	785	77	125	49	805	469	427	663	151	163	767	579	702	233	183	13	312	129	755	537	634	588	523	782	238	568
506	128	483	831	390	303	61	210	549	596	240	725	268	707	277	794	194	366	58	873	419	336	498	618	119	342	378	865	586	795
715	530	323	87	601	389	224	735	866	587	683	88	684	613	348	566	259	405	880	862	146	130	821	97	722	358	85	624	16	71
254	429	354	686	792	790	69	21	799	875	765	893	585	786	114	511	616	545	676	215	65	2	159	157	304	267	494	333	140	819
694	293	212	623	154	305	768	843	840	869	508	173	817	841	223	83	82	479	554	270	175	728	445	272	852	739	399	11	220	43
103	414	514	98	249	363	538	751	375	359	745	811	110	828	80	166	882	242	655	217	847	836	243	539	505	611	23	133	899	379
55	513	519	756	306	139	377	758	898	391	192	803	352	60	137	464	101	209	150	576	653	877	698	416	213	897	639	640	289	437
752	885	520	661	800	219	704	727	70	669	57	229	53	93	250	4	555	394	132	291	204	597	116	127	816	814	516	625	881	754
170	200	463	78	879	450	134	364	736	155	542	614	284	280	730	829	860	42	779	660	380	544	251	347	478	331	521	647	723	44
592	895	552	395	299	492	226	310	695	468	448	59	871	17	349	633	237	236	748	451	187	629	457	541	180	246	646	529	760	367
160	535	503	241	273	603	848	216	452	214	426	844	729	766	509	761	258	607	33	413	102	853	374	868	10	560	225	369	531	232
446	495	397	796	780	466	425	700	184	673	567	256	563	547	222	462	441	610	682	92	757	124	351	486	387	675	7	30	76	818
262	191	341	753	685	837	350	138	688	645	206	824	257	227	827	261	604	546	812	551	265	421	564	515	50	838	108	288	153	308
485	99	487	810	708	571	230	600	235	526	423	724	195	689	344	118	434	743	309	699	136	64	636	659	317	637	622	476	424	115
595	606	197	826	643	851	371	477	105	383	896	205	208	703	734	444	145	74	443	63	773	605	820	48	248	161	677	31	858	325
791	594	889	415	656	874	275	131	274	500	121	96	91	435	5	630	490	863	393	54	575	631	706	740	433	438	411	301	75	628
572	830	120	244	798	338	510	793	330	807	744	148	3	292	589	430	79	870	386	298	891	287	147	453	697	578	581	1	626	73
340	422	774	38	353	253	228	95	747	185	26	612	189	196	804	252	770	701	608	81	409	279	559	320	772	867	787	886	295	667
141	149	307	678	117	396	591	297	319	168	584	239	527	525	582	650	662	62	732	617	467	642	318	550	737	165	447	454	558	834
857	496	282	15	169	480	403	384	410	381	719	18	769	536	162	883	856	172	177	716	573	158	41	570	72	876	850	854	112	524
710	181	556	504	784	311	615	456	123	286	731	107	835	749	56	94	203	188	501	90	711	801	808	528	313	255	861	126	593	439
171	859	315	329	497	522	360	822	40	475	580	802	720	266	322	32	449	872	283	143	417	37	733	561	218	436	679	652	553	370
548	491	408	202	234	100	28	599	776	14	285	658	900	775	709	334	182	789	783	771	718	326	460	540	674	22	362	156	231	440
19	296	324	27	142	717	759	276	489	372	144	36	644	484	260	420	657	839	24	892	598	888	68	507	632	245	671	855	518	712
690	314	428	373	122	641	649	442	190	741	687	455	532	327	332	51	746	376	66	39	672	461	221	691	474	458	781	278	388	890
281	609	809	470	714	207	849	357	401	186	365	619	167	602	823	488	878	84	788	471	638	174	864	193	89	25	52	368	680	264
742	499	763	402	45	577	583	696	670	152	9	302	46	47	825	726	176	832	762	884	135	721	300	198	565	400	271	269	557	361
668	104	407	666	179	199	842	86	412	385	337	534	321	113	845	713	263	346	316	493	382	459	109	665	846	502	211	651	664	797
111	533	431	465	355	418	692	620	8	806	404	201	750	35	778	705	406	543	590	833	339	482	345	738	343	20	67	764	398	335
392	290	481	472	6	627	681	562	29	178	654	777	894	813	512	432	328	247	12	887	569	294	648	34	815	164	621	517	473	106

遺伝子制御ネットワークの変異に対する頑健性による表現型選択

arXiv: 2310.15729

生命進化の普遍性

Darwinian Evolution の普遍性

- おそらくあらゆる生命はダーウィン進化で作られる
  → ダーウィン進化の理解は生命の普遍性の理解に不可欠

ダーウィン進化: 機能と頑健性の進化

Darwinian Evolution

- 変異と選択によって環境に適応する過程
  - 一種の最適化
- 機能を獲得すると同時に変異に対する頑健性をも獲得する (Wagner)
  - 単なる最適化ではない

ここまでは普通の話

頑健性による表現型選択

- 仮に同じ適応度を持つ二つの表現型 (形質) があるとすると、変異に対して頑健な方が選択されやすいはず
  - 表現型選択の新しいメカニズムとしてあるのではないか

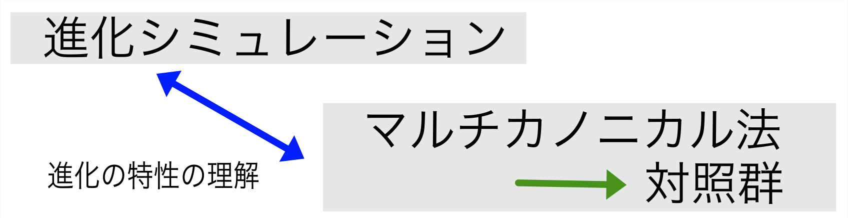
進化の特性をレアイベントサンプリングで調べる

進化過程の特性は何か

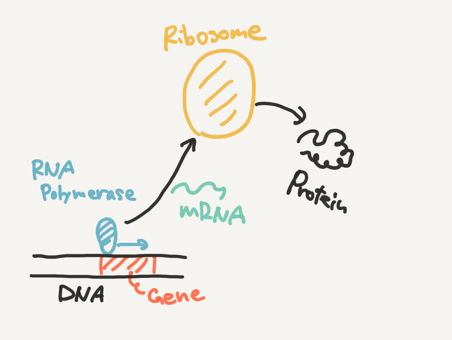
- 進化シミュレーション: 進化で起きることを知る
  - 進化ならではの特性はこれだけではわからない
    $\rightarrow$ 比較対象が必要
- 適切な比較対象: 遺伝子型のランダムサンプリング
  - 進化しないとどのような表現型がどれくらいあるか
- 問題: 適応度が高い遺伝子型は稀
  $\rightarrow$ 単純なランダムサンプリングは無力
- 解決法: レアイベントサンプリング (Multicanonical MC) で比較対象を作る
  - T.Kaneko and MK, PLoS Comput Biol 18 (2022) e1009796

適応度に対する Multicanonical MC

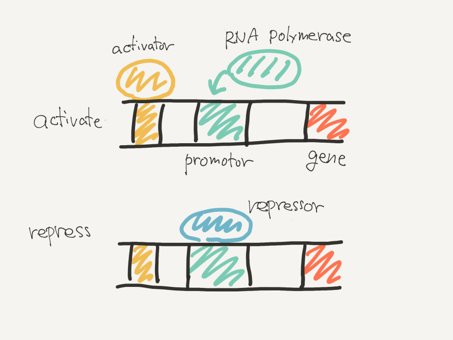
- 適応度をエネルギーとみなす
  - Multicanonical MC により、全適応度範囲から遺伝子型をサンプリング
  - 等重率の原理より、各適応度ごとにランダムサンプリングされている
- McMC の結果を対照系として、進化シミュレーションと比較
  → 違いがあれば、それが進化ならではの特性



遺伝子制御ネットワーク (GRN) 1

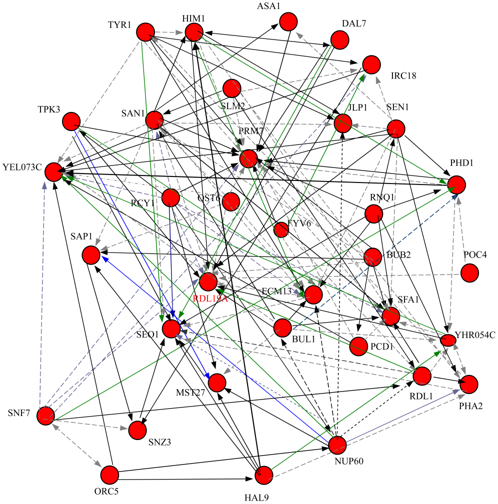


遺伝子発現



遺伝子制御

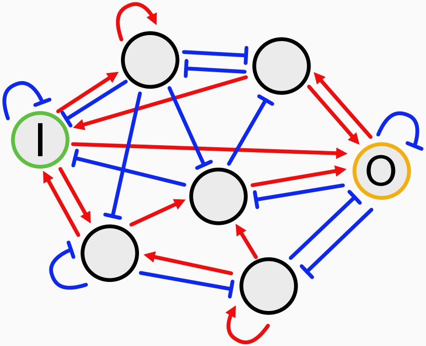
遺伝子制御ネットワーク (GRN) 2



- 細胞内では遺伝子同士が発現制御しあう複雑なネットワークが働いて細胞状態を調節している。
  - 外界の変化への適応
  - 細胞状態の不可逆的変化 (成熟、分化)

Dong *et al.* PLoS ONE 8 (2013) e83263

GRN のモデル 1



小さなモデルの例

- GRN を有向グラフで表す
- ノード: 遺伝子
- エッジ: 制御関係 (活性化、抑制)

詳細

- $N = 40$ ノード、$K = 120$ エッジ
- 1 入力 1 出力遺伝子

モデル 2: ダイナミクス

離散時間ダイナミクス

$$x_i(t+1) = R\left(I\delta_{i,0} + \sum_j J_{ij} x_j(t)\right)$$

$$R(y) = \frac{1}{1 + e^{-\beta(y-\mu)}}$$

- x_i : i 番遺伝子の発現量 ($[0,1]$)
- J_{ij} : j から i への制御 ($0, \pm 1$)
  - $+1$: 活性化, -1: 抑制
- I : 入力 ($[0,1]$)
- $R(y)$: シグモイド型の応答関数

モデル 3: 適応度

遺伝子型

- J_{ij} を遺伝子型とみなす

適応度

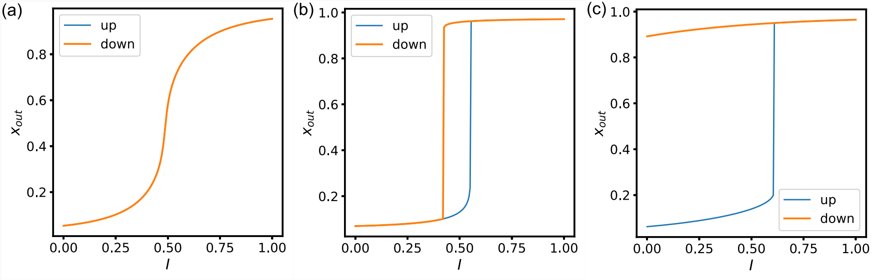
- $I = 0$ の定常状態 $\rightarrow$ $I = 1$ の定常状態と変化させたとき

$$f = x_{out}(1) - x_{out}(0),$$

- $x_{out}(I)$：入力が I のときの定常状態での出力遺伝子の発現量
- $f < 0$ は $f = 0$ とする (逆の反応は適応度が 0)

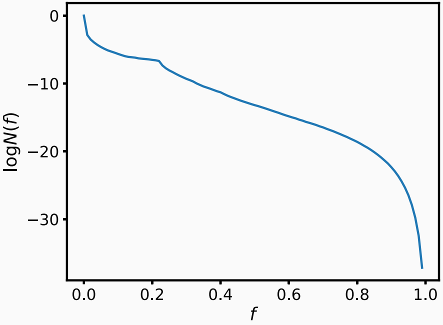
モデルの振る舞い

Monostable　Toggle switch　One-way switch

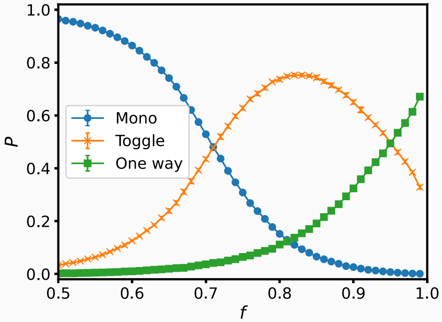


- f が大きい時、入力を準静的に変化させると、遺伝子型の違いにより三通りの応答を示す → 三つの表現型とみなす

Multicanonical による計算



エントロピー: 適応度が高い GRN は稀



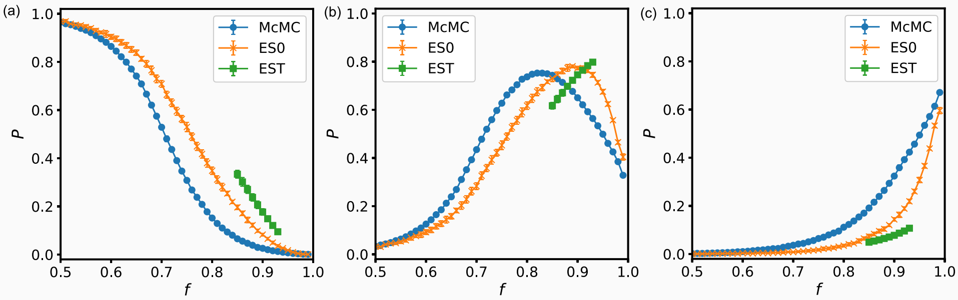
適応度 vs. 表現型の出現確率

2 種類の進化シミュレーション

1. **温度 0 の進化 (transient)**
   - ランダムネットワーク 1000 個からスタート
   - 適応度 f の上位 500 個体を残し、それぞれのコピーを作る
   - コピーに点変異: エッジの繋ぎかえ
2. **有限温度の進化 (steady state)**
   - ランダムネットワーク 1000 個からスタート
   - $\exp(\beta f) \times RND$ の上位 500 個体を残し、それぞれのコピーを作る
   - コピーに点変異: エッジの繋ぎかえ
   - Steady state での平均を計算

各表現型の出現率

Monostable　　Toggle switch　　One-way switch



- 進化の定常状態では One-way switch の出現が大きく抑えられる

変異に対する頑健性の計算

変異に対する頑健性とは

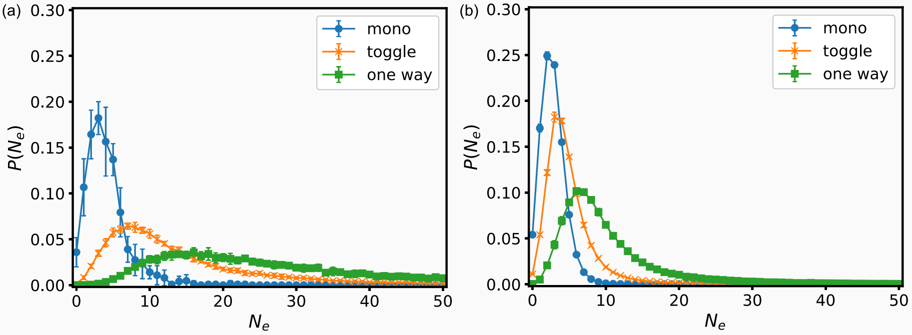
- 生物は突然変異に対して機能を失わない性質を持つ
    - 変異・選択というダーウィン進化の仕組みによって、環境への適応とは別に獲得した性質

本研究での頑健性の指標

- f がある程度大きい場合について、ひとつのエッジを削除した時に $f < 0.5$ となるエッジを必須エッジとする
- 各 GRN に対し、可能なすべての 1 エッジ削除を行い、必須エッジの数を数える
    - 必須エッジが少ないほど、変異に対して頑健

必須エッジ数の分布

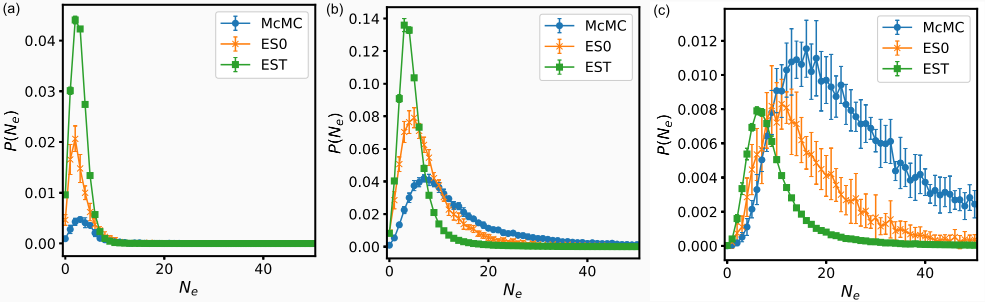
Multicanonical　　定常進化



- $f = 0.9$ での三つの表現型それぞれでの必須エッジ数の分布
  - 進化では必須エッジ数が少ない (頑健な) ものが選ばれる

各表現型での必須エッジの分布

Monostable　　Toggle switch　　One-way switch



- One-way switch では進化によって必須エッジが多い GRN は大きく排除されている

このセクションのまとめ

- 進化では適応度の最適化とは別に変異に対する頑健性による選択が起きる
- 変異に対する頑健性が低い表現型は出現が抑えられる
  - 変異に対する頑健性に起因する表現型選択
- これはダーウィン進化の機構のみによるので、他の系にも普遍的に成立すると考えられる
- この現象は Multicanonical 法と進化シミュレーションを比較することで初めて確認された

タンパク質の機能と折れたたみの共進化

in preparation

タンパク質の折れたたみ

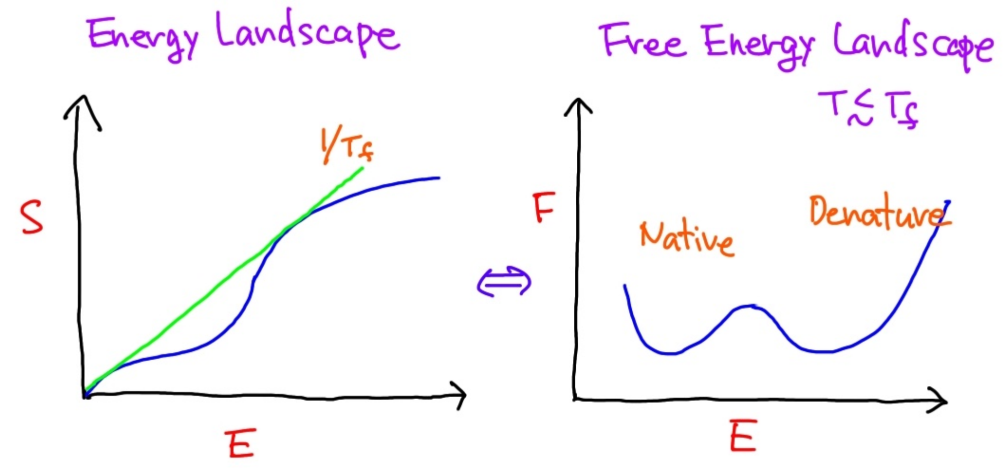
Protein folding

- タンパク質: 20 種のアミノ酸をペプチド結合させた線状高分子
- 生体内で特定の立体構造 (天然構造) に折れたたんで機能する
  - 天然構造は生理条件下での熱平衡状態 (Anfinsen のドグマ)
- 天然構造と変性状態の間には自由エネルギー障壁がある (二状態転移)
- 疎水性アミノ酸を中心に集めることによってコンパクトになる

理論

- 全体構造と局所構造が consistent (Consistency Principle by Go)
- 構造フラストレーションが最小化されている (Minimum Frustration Principle by Wolynes)

(Free) Energy Landscape (Funnel Picture)



- $T < T_f$ で二状態転移するためのエネルギーと自由エネルギー構造

タンパク質の進化

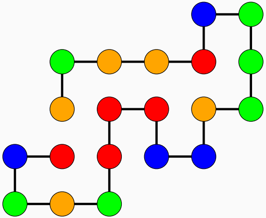
Folding は進化で獲得された

- 唯一の天然構造を低エネルギー構造として持つ
- 二状態転移できるようなエネルギーランドスケープを持つ (ファネル構造)
    - 天然構造だけでなく、エネルギーが高い構造も進化的に決まっている

本研究の狙い

- 問題提起: 進化の選択圧はあくまでも機能のはず
    - 唯一の天然構造やファネル構造は機能の副産物に過ぎないのではないか
- 目標: 機能の強さだけを適応度として進化させると唯一の天然構造やファネル構造を獲得するか

モデル 1

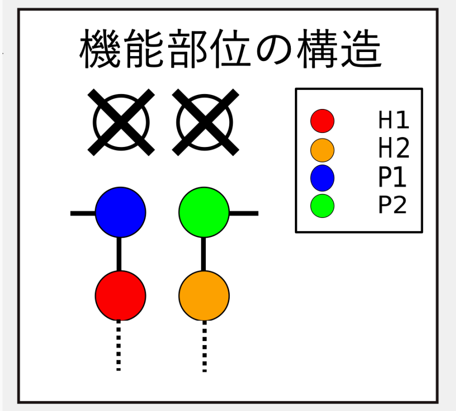


- 二次元格子タンパク質モデル
  - 正方格子上の自己回避ウォーク
- アミノ酸は 4 種
  - 疎水性アミノ酸 H1,H2
  - 親水性アミノ酸 P1,P2
- 再隣接格子点にあるアミノ酸同士が相互作用する
  - 配位ごとにエネルギーが決まる
- 長さ: 20 アミノ酸

モデル 2

	H1	H2	P1	P2
H1	−3.5	−3.1	−1.2	−1.1
H2	−3.1	−2.3	−0.9	−0.8
P1	−1.2	−0.9	−0.2	−0.4
P2	−1.1	−0.8	−0.4	0.0

アミノ酸の相互作用エネルギー



この形があれば活性があるとする

モデル3

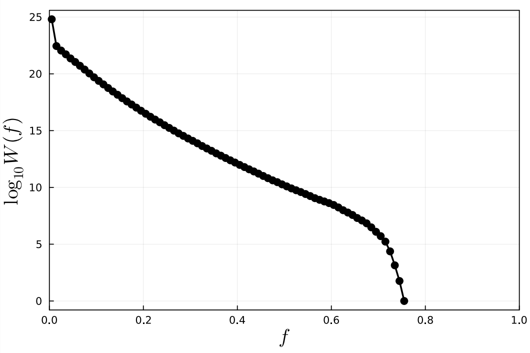
適応度

- Order parameter Ψ: 活性がある配位で 1, ない配位で 0
- 適応度 $f = \langle \Psi \rangle$
  - 有限温度での Order parameter の平均値
  - 温度をパラメーターとする

計算

- 遺伝子型: アミノ酸配列
- Multicanonical 法で全適応度範囲のアミノ酸配列を作り、適応度が高い配列を調べる
- 進化シミュレーションで確認

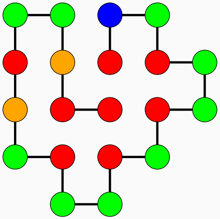
結果 1



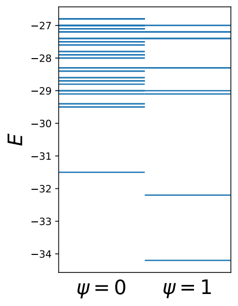
- $T = 1.0$ での適応度 vs エントロピー: 適応度が高い配列は稀
    - 以下 $T = 1.0$ の結果

結果 2-1

McMC で最大適応度 ($f = 0.785$) の配列 (進化でも得られる)

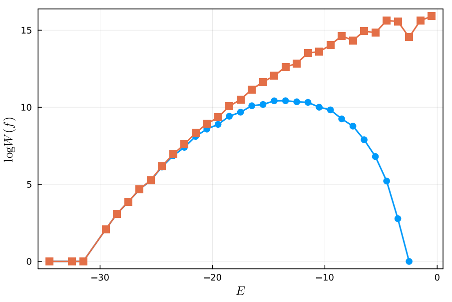


天然構造

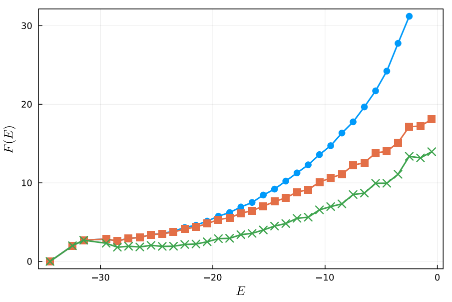


エネルギースペクトル

結果 2-2



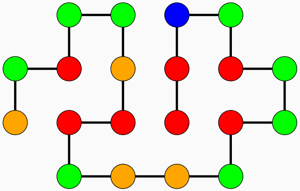
エネルギーランドスケープ: ファネル状



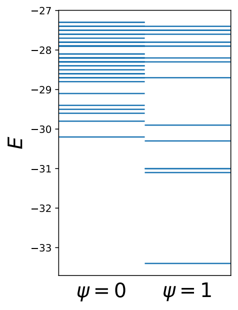
自由エネルギーランドスケープ: 二状態転移

結果 3-1

進化で得られた他の構造 ($f = 0.579$)

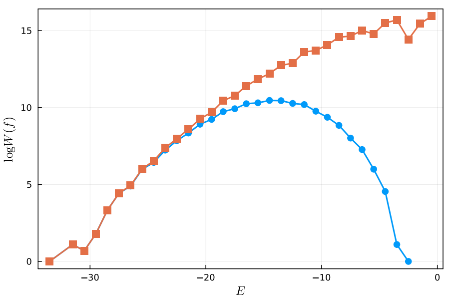


天然構造

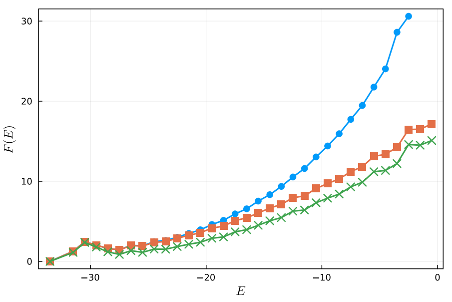


エネルギースペクトル

結果 3-2



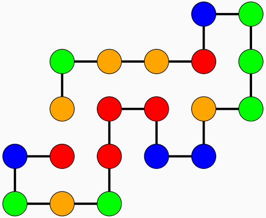
エネルギーランドスケープ: ファネル状



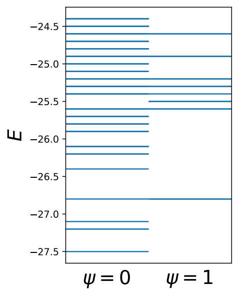
自由エネルギーランドスケープ: 二状態転移

結果 4-1

適応度が低い配列 ($f = 0.103$)

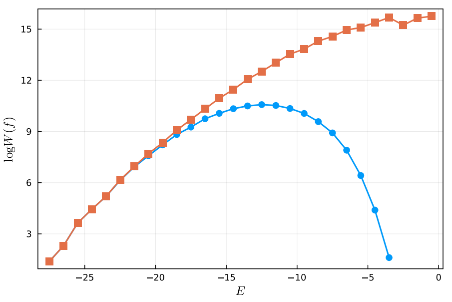


天然構造

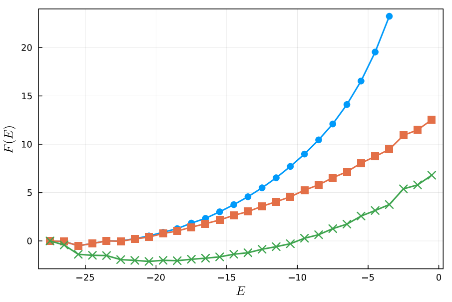


エネルギースペクトル (天然構造で活性がない)

結果 4-2



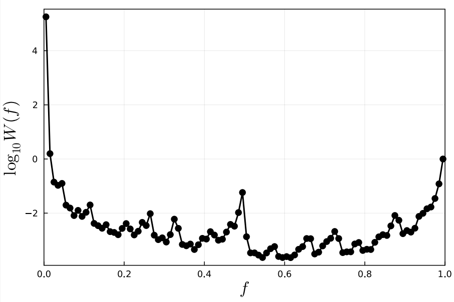
エネルギーランドスケープ: ファネル状でない



自由エネルギーランドスケープ: 二状態転移でない

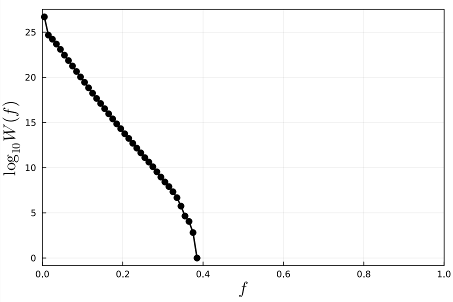
結果 5

低温: $T = 0.1$



基底状態しか関係ないので二状態にならない

高温: $T = 1.4$



適応度が上がらない

このセクションのまとめ

- 機能を適応度とするだけで、適応度が高いアミノ酸配列は唯一の天然構造とファネル状のエネルギーランドスケープ (二状態転移) を自発的に獲得する
  - 機能のみが選択圧となって、折れたたむタンパク質が出現する
- 進化する際の環境温度は重要で、適切な温度範囲でのみ機能を持つタンパク質が出現
- シナリオは普遍的 (アミノ酸の種類などに関係ない)
- この結果は主に Multicanonical MC で示され、進化シミュレーションで確認した

まとめ: 生物進化の普遍性探求と統計力学

Summary

- 本講演では統計力学で開発されたMulticanonical MCが生命の進化を研究する手法として有効であることを示した
  - 進化したものは稀なのでrare event samplingが有効
  - 統計力学は進化で作れないものも網羅できる
  - まだほとんど使われていないので、穴場です
- ダーウィン進化の普遍的な性質を導ける
- 生物進化の普遍性探求には現実的モデルよりもToy modelが有用なことがある

共同研究者
北島顕正 (現国会図書館)、斉藤稔 (現広島大)、永田新太郎 (卒業生)、金子忠宗 (卒業生)、割田詳 (卒業生)、丸山恭史 (M2)