第４２回生物物理学会年会（於国立京都国際会館）
（２００４，Ｄｅｃ．１３－１５）

球状蛋白質分子の構造量におけるカオス的性質の検出

(Detection of Chaotic Aspect s in Structural Quantities for Globular Proteins)

○ 窪田 綏（Advanced Tech. Inst., Inc.）、土屋 尚（明星大・情報学部）

〔要旨〕

Previously we showed [Y.Kubota, et al., Biphysica. Acta, Vol.1079, 73-78 (1991)] the distance ρ of each Cα atom from the center of mass and the quantity called N14 [K.Nishikawa and T.Ooi, J. Biochem. Vol.100,1043-1047 (1986)], which represents the number of all the Cα atoms within a sphere of 14Å radius centered on each Cα atom, are the structural quantities which reflects the tertiary structure of a protein molecule from its amino acid sequence by the method proposed by Grassberger and Procaccia (the G-P method) [P.Grassberger and I.Procaccia, Physica D, Vol.9, 189-208 (1983)] in the chaos theory. That is , the correlation dimension D2 and the embedding dimension d become 4-7 and 20, respectively for the sequences of ρ and N14 of globular proteins. According to our remark, the experimentally obtained sequence of the quantities has been generated by a certain deterministic mechanism. In other words, it is suggested to be deterministic chaos.

Here, we try another approach than the above G-P method, in order to detect a chaotic behavior of protein structures. Briefly, we construct a predictor from the data of K residues long from the sequence of the structural quantity by using RBF (Radial Basis Function) method proposed by Casdagli [M. Casdgli, Physica D Vol.35, 335-356 (1989)]. By applying this predictor, we can find the short-term predictability and the long-term unpredictability in globular proteins, which is a characteristic feature of the deterministic chaos.

Furthermore, we can generate a data over the number of residues of a protein by iterate to use the predictor, recursively (free running). Interestingly, the data which is obtained in such a way shows a similar attractor to the original one. And also the application of the G-P method to it reveals the same the correlation dimension D2 and the embedding dimension d to the native ones of a protein.

These results suggest the existence of chaotic aspect s in globular proteins.

〔序論〕

1．　以前我々〔1〕は、球状蛋白質の残基がどの程度表面－内側に位置するかの度合いを示す指標として各々C^{α} 原子を中心とした半径１４Åの球内に含まれる全てのC^{α} 原子の数で定義される量N_{14} 〔2〕や重心からの各C^{α} 原子の距離ρ がアミノ酸配列に沿った量の（1次元）数列として、物理化学的な決定論的メカニズムによって生成されると考えられている立体構造を反映するStructural Quantityであることをカオス理論の手法であるGrassberger と Procacciaによる相関積分の方法(G－P法)〔3〕により示した。つまり球状蛋白質におけるN_{14}やρ 等の量の列はランダムというよりカオス的振舞いを見せることを示した。従って我々は蛋白質の構造形成にはある決定論的プロセスによって制御されているに違いないと考える。

2．　今回この考えを更に裏付けるために、我々は別な方法である Casdagli の動径基底函数（Radial Basis Function: RBF）による手法〔4〕を使ってこれらの量にカオス的性質が存在することを示すことが出来た。即ち、RBF によりこれらの量を Time Series と見做して予測を試みた時、短期予測は可能であるが長期予測は不可能という意味においてカオス的性質があることがわかった。

3．　また残基数Nの蛋白質においてある学習データから更にその先N残基分フリーランさせてそのアトラクターを求めたところそのオリジナルアトラクターと定性的に究めて類似のアトラクターを得ることができ、更にこれを定量的にG-P法による相関積分によってCorrelation Exponentを求めそこから埋め込み次元 d と相関次元 D_2を求めたところオリジナル蛋白質の場合とほぼ同じプロフィールと次元が得られ、典型的なカオス系の場合と同様にストレンジアトラクタの再構築ができた

4．　このことからN_{14}や重心からの各C^{α} 原子の距離ρ のような構造量にカオス的性質が存在することが分かり、益々決定論的なメカニズムで蛋白質構造が形成されていることが支持されたと考える。

〔方法〕

○*Casdagli による非線形予測法*

Casdagli　は動径基底函数（Radial　Basis　Function:　RBF）法をカオス的時系列の予測問題に応用した。即ち；

$v(t)$ を再構成されたアトラクターの軌道の１点，c_i をセンターと呼ばれる点、とした時

１期先の予測を

$$\hat{y}(t+1) = \hat{F}(v(t) = \sum_{i}^{M} \vec{\gamma}_i \Phi(\|\vec{v}(t) - \vec{c}_i\|) \qquad (1)$$

により表す。ここで、γ_i は重み、$\Phi(\|\vec{v}(t) - \vec{c}_i\|)$ は基底函数と呼ばれる。基底函数としては；

ガウシアン函数：　$\Phi(r) = \exp(-r^2 / b)$　（２）

ベル型函数：　$\Phi(r) = \dfrac{1}{1 + \cosh(\sigma_h r)}$　（３）

が用いられる。

さて、通常データは１次元の離散的データとして得られるので、Eqn.1 を行列・ベクトルの形で表現すると便利である。即ち、各ベクトル$\vec{x}$をｍ次元ベクトルして；

$$\mathbf{y} = (\vec{y}_2, \cdots\cdots, \vec{y}_{M+1})^{\mathrm{t}} \qquad (4)$$

$$\boldsymbol{\gamma} = (\vec{\gamma}_1, \cdots\cdots, \vec{\gamma}_M)^{\mathrm{t}} \qquad (5)$$

$$\mathbf{P} = \lfloor p_{ij} \rfloor = \Phi(\|\vec{v}(i) - \vec{c}(j)\|) \qquad (6)$$

とすると

$$\mathbf{y} = \mathbf{P}\boldsymbol{\gamma} \qquad (7)$$

となり、予測器の作成は Eqn.7　の $\boldsymbol{\gamma}$　を求め、これを次々に再帰的に繰り返し用いることで長期に渡って時間発展（Free　Running）させて、予測時系列を作成することができる。こうして、与えられた時系列の決定論的力学系としての挙動の特徴を捉えることができる。

なお、今後予測器（ *predictor*）とは Eqn..1 またはＥｑｎ．７を指して言う。

○*相関積分法*

Grassberger　と　Procaccia らは、解析対象となるシステムの有する自由度であるアトラクターの次元を推定する相関積分の手法を提案した〔３〕。即ち、今１次元時系列

を｛ｘ（t_i)｝　とし、この１次元データからｍ次元ベクトルを；

$$\vec{X}_i = (x(t_i), x(t_i+\tau), ..., x(t_i+(m-1)\tau) \qquad (8)$$

を構成する。ここでτ　は時間ラグである。この時、相関積分C（ｒ）を；

$$C^m(r) = \frac{1}{N^2}\sum_{i,j=1}^{N} \vartheta(r - |X_i - X_j|) \qquad (9)$$

と定義する。ここで、Nはデータ数、Θ　はヘビサイド函数である。この$C^m(r)$が；

$$C^m(r) \propto r^{\nu(m)} \qquad (10)$$

と表わされる時、ν　（m）を相関指数と呼ぶ。このν　（m）はEqn.10の両辺の対数；

$$\log C^m(r) \propto \nu(m) \log r \qquad (11)$$

をとり、横軸にlog r、縦軸に log $C^m(r)$　をプロットした時、適切な　r　の範囲内での直線部分の勾配として求められる。そして相関次元　D_2　は、空間次元m　を漸次増加させて行った時、もはやν（m）が増加しない縦軸の点ν　として求められる。またνがこれ以上変化しない最小のmは埋め込み次元dとよばれる。

〔結果〕

Fig.1は典型的な球状蛋白質の例として　porcine elastase (PDB code: 1est　残基数N=240)に対して得られた西川及び大井によるパラメータN_{14}　のプロット〔２〕である。

さて、Fig.1に示された１次元時系列に前項で導入したCasdagliの予測法を適用して見る。即ち、Eqn.1における学習データ長Mを194とし、また前回我々によって得られた結果を参考にして再構成されたアトラクターの次元mの値を６として〔１〕、予測器、Ｅｑｎ．１をFree Running　させた結果をFig.2に示す。残基番号２０１（＝M+m+1）以降がフリーラン予測となるが、残基２０１～２２０付近の予測精度は高く、それ以降となるとその予測精度は漸減しており、カオスとしての特徴である短期予測可能性と長期予測不能性と言うカオス的挙動を良く呈していると考えられる。

さて、この予測モデルがオリジナルなN_{14}　の挙動を良く再現しているかどうか、まず得られたアトラクターの様子から定性的に見てみよう。Fig.3aはオリジナルなアトラクターであり、Fig.3bはフリーランによるそれである。この予測モデルが比較的良くオリジナルの振る舞いを再現していることが分かる。更に定量的に見るためにG-P法による相関積分を計算することにより、相関次元D_2及び埋め込み次元dを求めて見る。

Ｆｉｇ．４ａはオリジナルN_{14}(Fig.1) 、Ｆｉｇ．４bは２０１残基からフリーランさせたものについて（Fig.2） ｍ＝１～３０に対する相関積分（Ｅｑｎ．９）である。
Ｆｉｇ．４ａ（オリジナル）及びＦｉｇ．４ｂ（フリーラン）からそれぞれの相関次元D_2及び埋め込み次元 ｄ を求めるために、これらの図のカーブの直線部分の勾配ν（ｍ）を計った（Ｆｉｇ．５）。
Fig. ５から示されるようにオリジナル（点線）、フリーラン（実線）共に相関次元D_2及び埋め込み次元 ｄ がそれぞれ５と２０となっており、この予測モデルが良くそのオリジナルな振る舞いの特徴を再現していると言える。また重心からの各C^{α}原子の距離ρについても類似の結果が得られた。
以上のことからN_{14}や重心からの各C^{α}原子の距離ρ のような構造量にカオス的性質が存在すると考えられ、決定論的なメカニズムで構造が形成されていると考える。このことは、蛋白質の構造形成過程の背後に、恐らくは自由度が２０であるような微分方程式或るいは差分方程式の存在を暗示していると思われる。

〔参考文献〕

〔1〕Kubota, Y. and Tsuchiya, T.：“Chaos-Theoretical Analysis of Possible Structural Quantities for Globular Proteins,” Biophysica Acta, **1079**, 73-78 (1991)
〔2〕Nishikawa, K. and Ooi, T.：“Radial Locations of Amino Acid Residues in a Globular Proteins: Correlation with the Sequence” J. Biochem, **100**, 1043-1047 (1986)
〔3〕Grassberger, P. and Procaccia, I.: “Measuring the Strangeness of Strange Attractors” Physica D **9**, 189-208 (1983)
〔4〕Casdagli, M. :”Nonlinear Prediction of Chaotic Time Series”, Physica D **35**, 335-356 (1989)

Figure Legends

Fig.1 Ｐｏｒｃｉｎｅ　Ｅｌａｓｔａｓｅ　（残基数２４０、ＰＤＢ　Ｃｏｄｅ：1est）のN_{14}プロフィール。各残基に対してＰＤＢ座標データから１４Å以内のC^{α}原子がカウントされている。表面に近い残基ほどN_{14}　の値は高く、内側のそれは低く、丁度各残基のC^{α}原子の重心からの距離ρ　とほぼ逆相関の関係になっている。

Fig.2　Ｆｉｇ．１において、残基番号１～２００のデータを使って予測器 Eqn.1 を作成し、２０１番目の残基からもう２４０先までフリーランさせた。実線はフリーラン、破線はオリジナル。２０１～２２０番目付近までは良く予測されており（短期予測可能性)、それ以降の予測ははずれており（長期予測不能性)、決定論的カオスである特徴を呈している。

Fig.3a－b　**(a)** N_{14}プロフィール（Ｆｉｇ．１）から得られたオリジナルなアトラクター,
（b） フリーラン（Ｆｉｇ．２）から得られたアトラクター。定性的には、オリジナルアトラクターの振る舞いを再現する予測モデルが構築できていることが分かる。

Ｆｉｇ．４ａ－ｂ　再構成空間次元ｍを１から３０まで変化させた時の相関積分Ｅｑｎ９。
(a) オリジナル　**(b)**　フリーラン

Ｆｉｇ．５　各ｍに対してＦｇｉ．４ａ，４ｂの直線部分の勾配から得られた相関指数ν（m）のプロット。実線はフリーラン、点線はオリジナル。このグラフから共に、相関次元D_2＝５、埋め込み次元　ｄ≅２０と求められた。

Fig.1

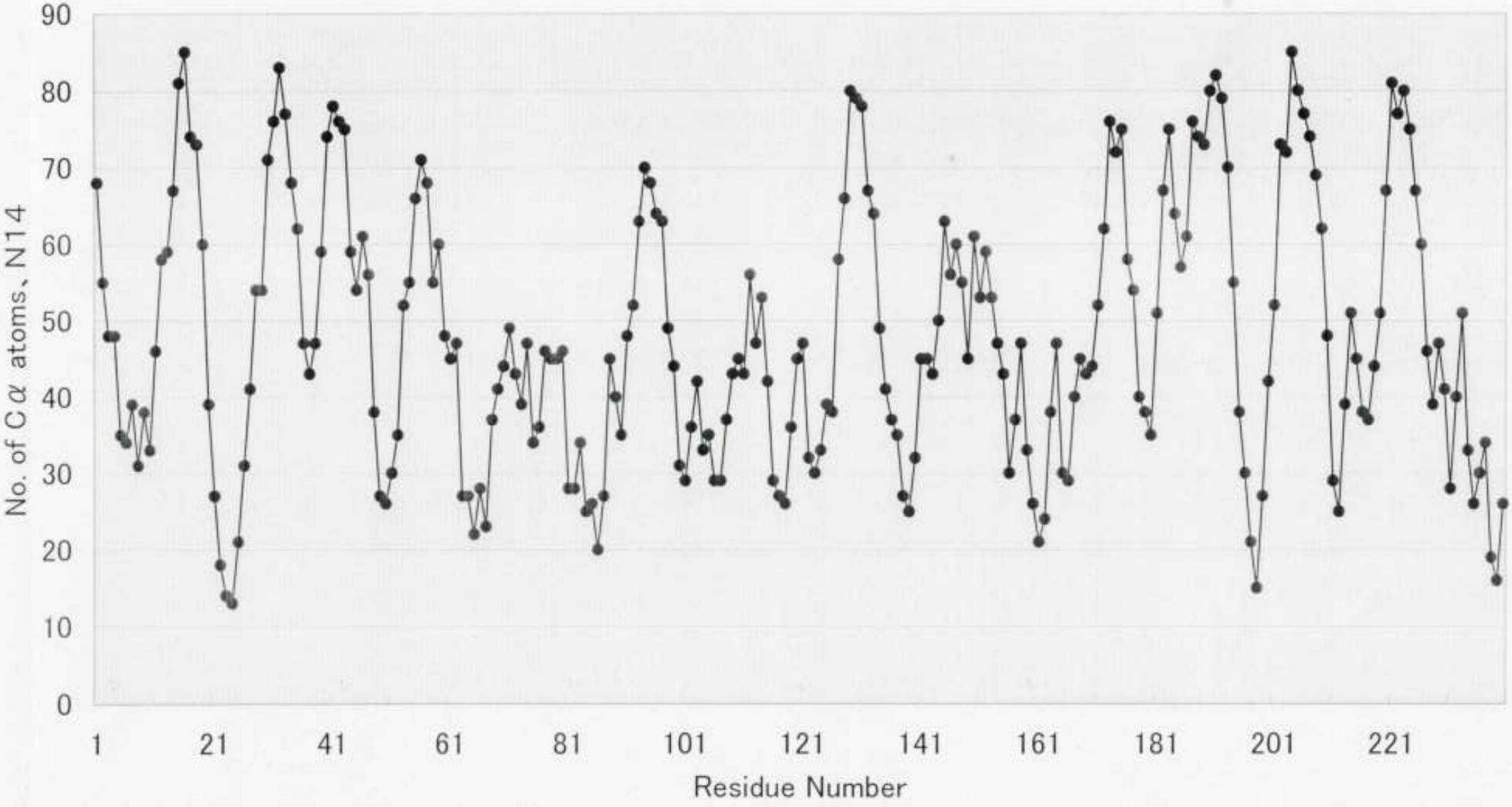

90
80
70
60
50
40
30
20
10
0
No. of Cα atoms、N14
1
21
41
61
81
101
121
141
161
181
201
221
Residue Number

Fig.2

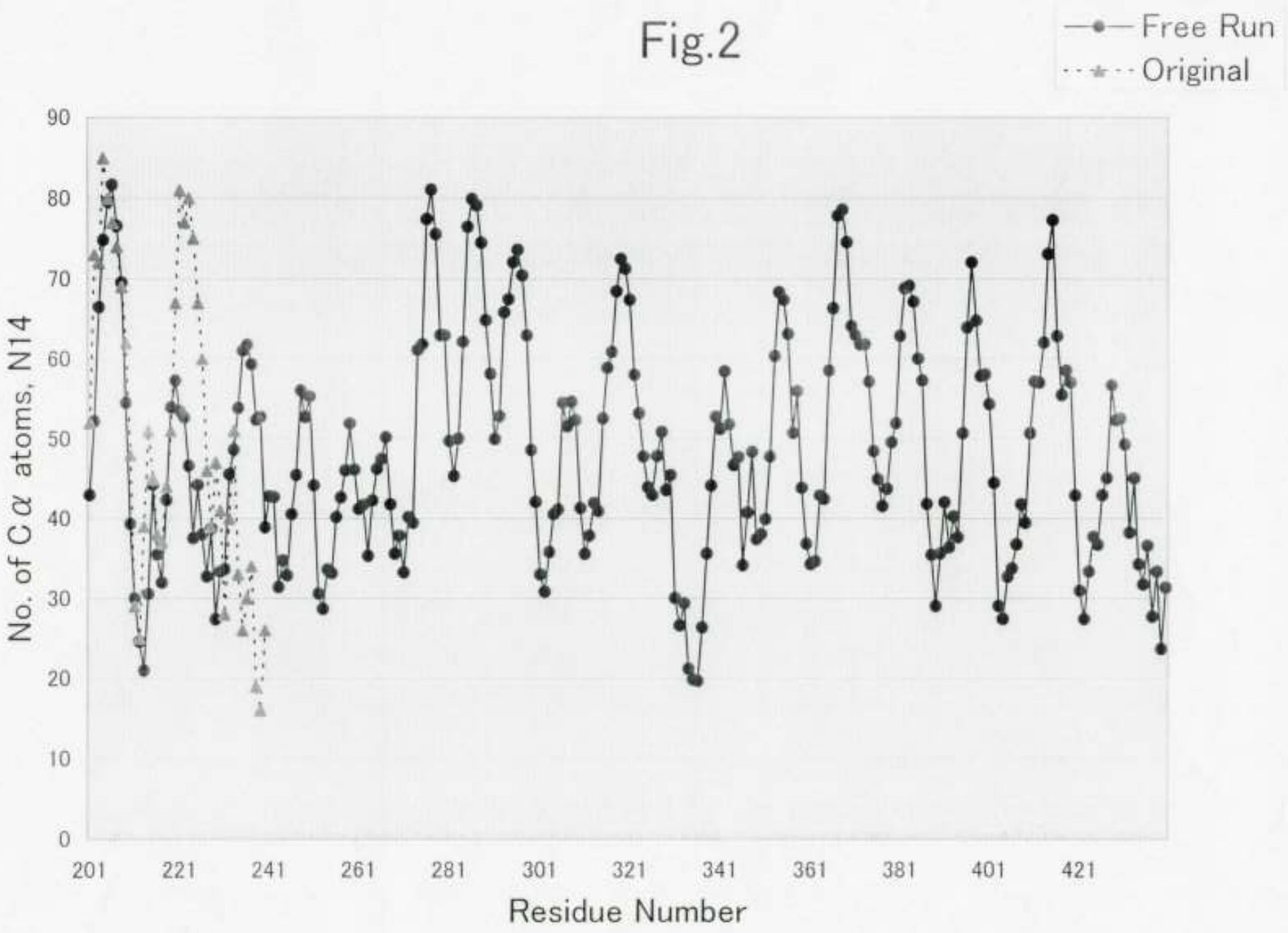

Free Run
Original
No. of Cα atoms, N14
0
10
20
30
40
50
60
70
80
90
201
221
241
261
281
301
321
341
361
381
401
421
Residue Number

Fig.3a (Original)

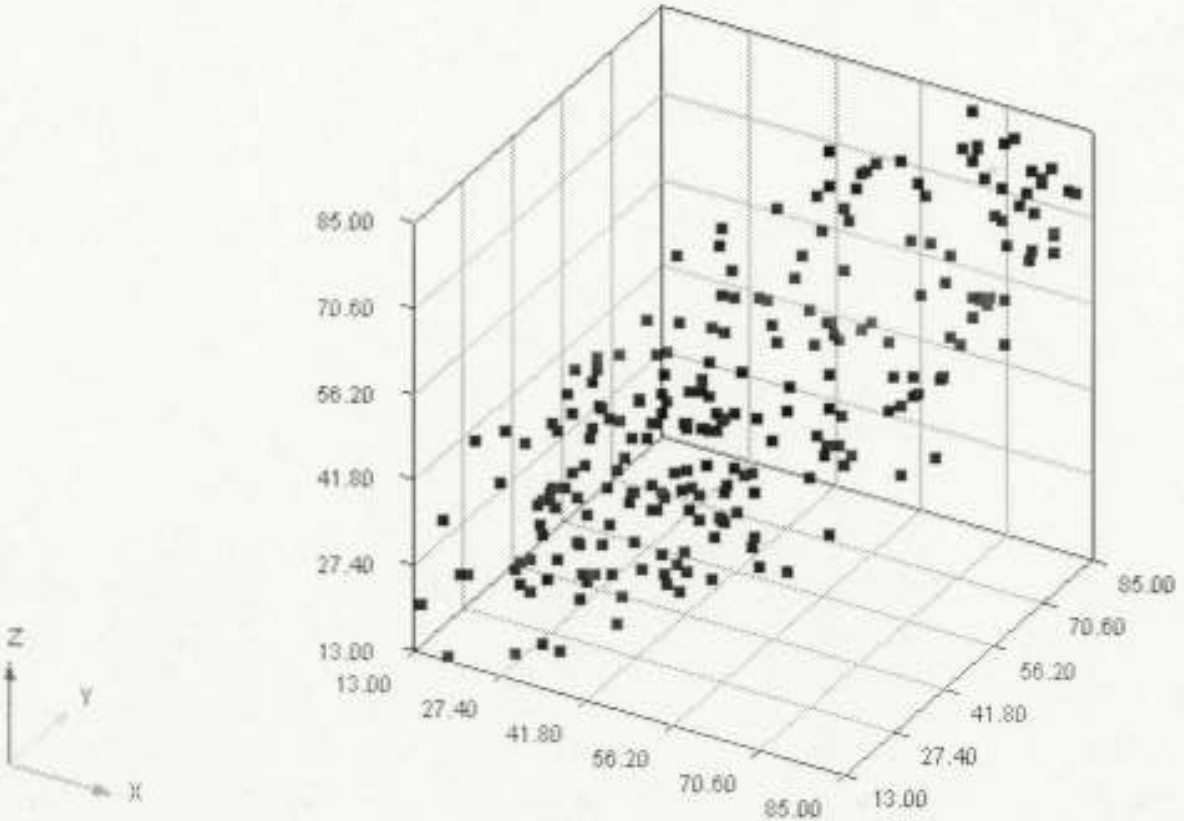


Fig.3b (Free Run)

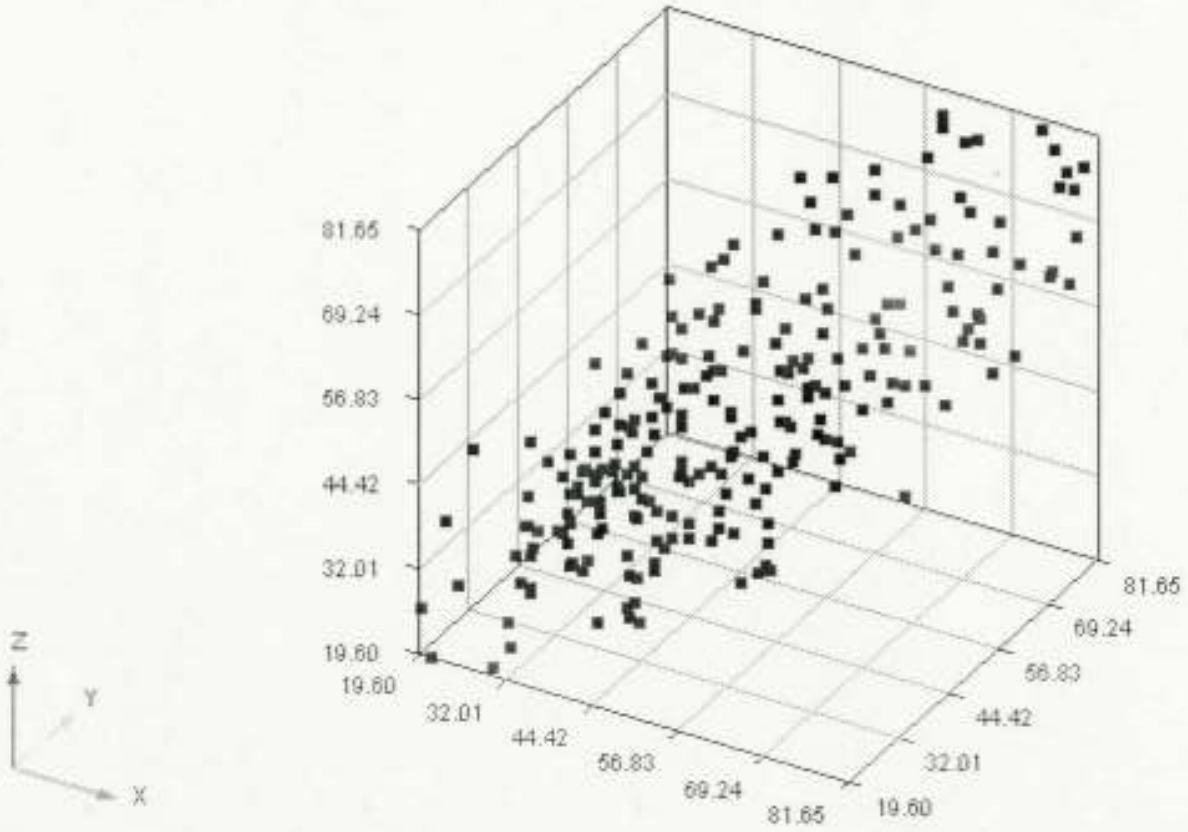


Fig. 4a

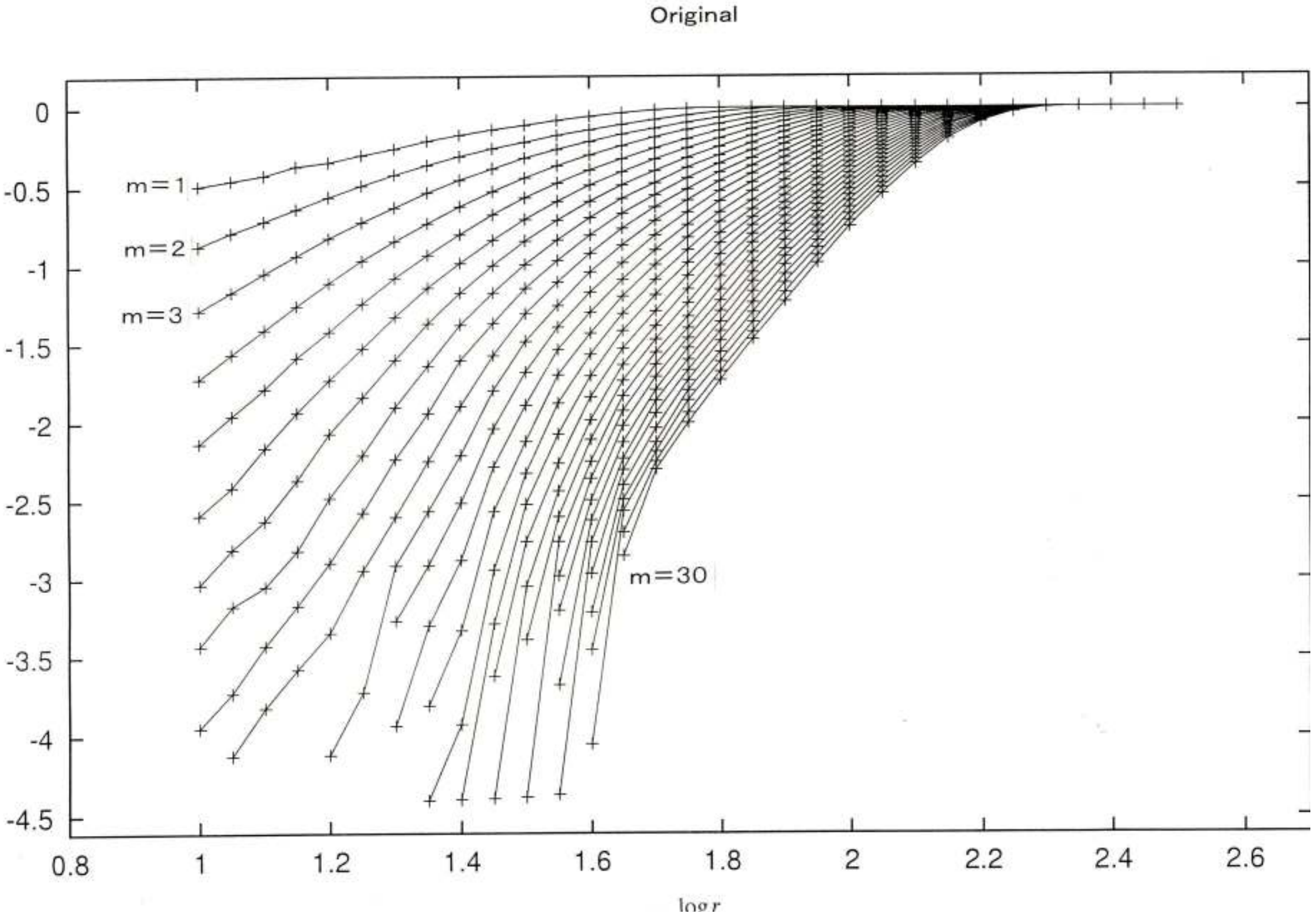

Original
0
-0.5
-1
-1.5
-2
-2.5
-3
-3.5
-4
-4.5
m=1
m=2
m=3
m=30
0.8
1
1.2
1.4
1.6
1.8
2
2.2
2.4
2.6
log r

Fig. 4b

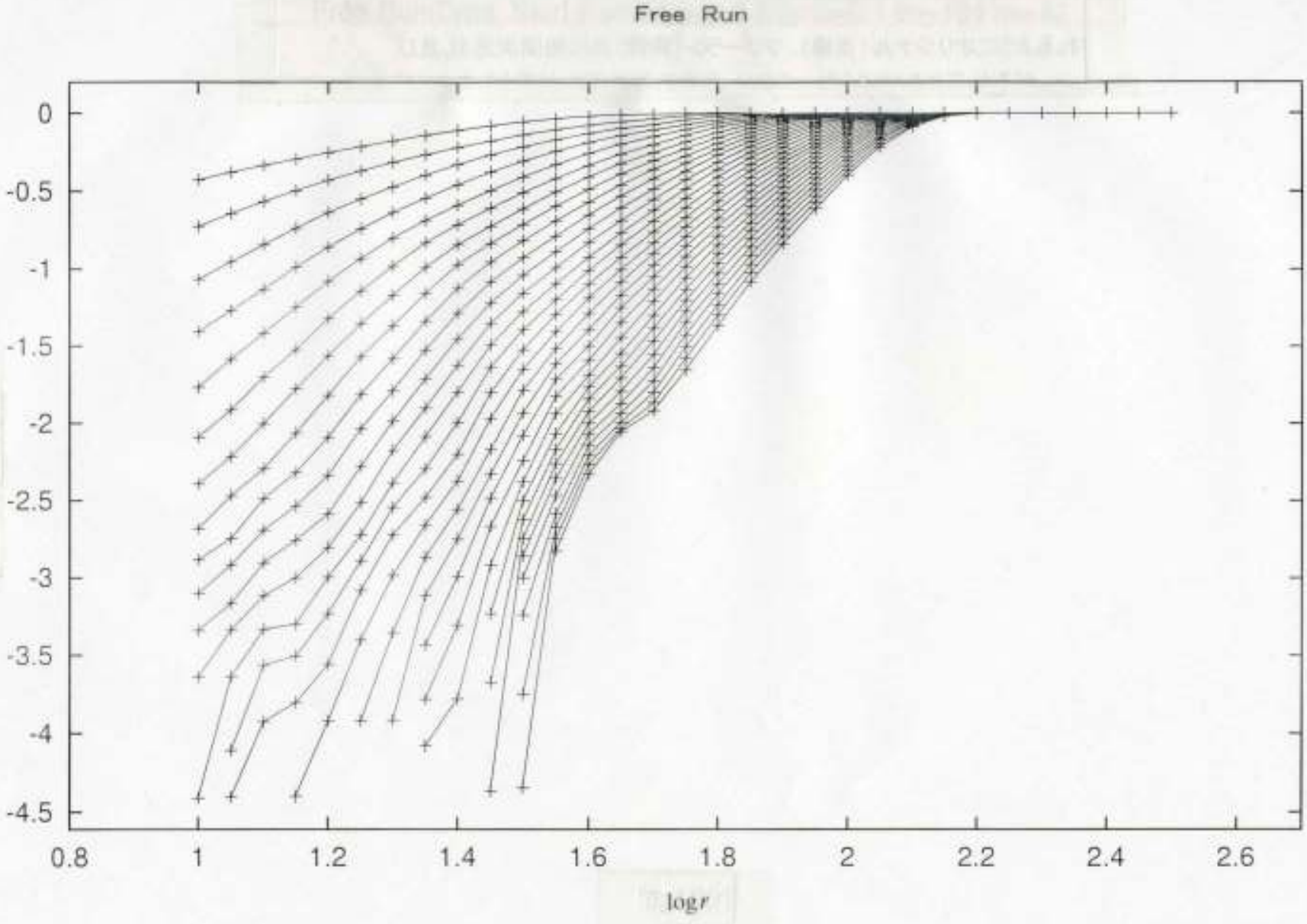
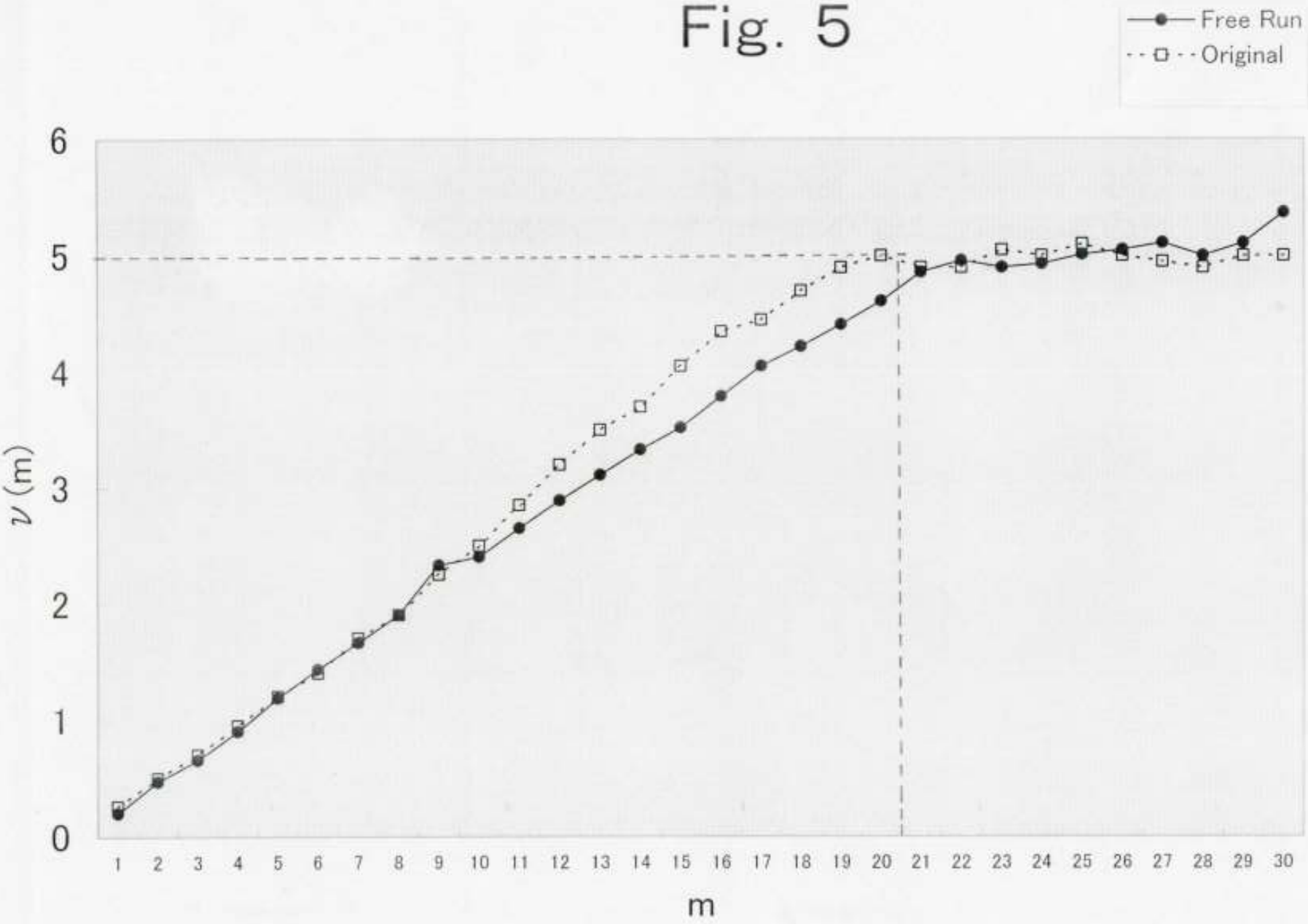

Free Run
0
-0.5
-1
-1.5
-2
-2.5
-3
-3.5
-4
-4.5
0.8
1
1.2
1.4
1.6
1.8
2
2.2
2.4
2.6
$\log C^m(r)$
$\log r$

Fig. 5


Free Run
Original
6
5
4
3
2
1
0
ν (m)
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
m