

О физических основах аминокислотного кода для 3D-структуры белков

Шайтан К.

МГУ им. М.В. Ломоносова, Москва, Россия

shaytan49@yandex.ru

Ключевые слова: линейные полимеры; полипептиды; топография многомерных энергетических поверхностей; динамика фолдинга; эффекты вязкости среды; топологии конфигурационного пространства и симметрии относительно перестановок одинаковых мономерных звеньев; температура денатурации

Мотивация и цель: В настоящее время считается установленным, что пространственная структура белков определяется аминокислотной последовательностью и соответствует минимуму свободной энергии макромолекулы [1]. Имеются некие общие представления о том, что стабильность 3D-структуры требует наличия энергетической щели между нативной конформацией и ближайшими метастабильными конформациями цепи [2], а также о некотором относительно гладком устройстве энергетической поверхности (принцип минимальной фрустрации [3]) для преодоления парадокса Левинталя [4]. Описание современного состояния проблемы можно найти в недавнем обзоре [5]. Однако, конкретные физические аспекты, которые определяют процесс фолдинга долгое время оставались не очень понятными ввиду сложности объекта и отсутствия адекватных математических инструментов для описания данного процесса. Популярность получили компьютерные методы предсказания пространственной структуры белков, которые не основаны на физике фолдинга, а используют технологии ИИ [6]. Эти алгоритмы работают быстро, но с переменным успехом. В основном, хорошо для гомологичных структур. Мы в данном случае исходим из презумпции первенства физической картины и закономерностей для биологических процессов сколь сложными они бы не представлялись на данный момент [7]. Раскрытие физических механизмов организации аминокислотного кода для пространственной структуры белков имеет принципиальное значение для понимания физико-химических основ жизни и способно существенно изменить имеющуюся парадигму исследований в структурной биологии. Развиваемые представления имеют отношение не только к полипептидам, но и к формированию пространственных структур других линейных полимеров и биополимеров.

Методы и алгоритмы: Развиваемые представления основаны на строении линейных полимеров, конформация которых определяется поворотами по двугранным углам вокруг химических связей. Вклад колебаний по валентным степеням свободы в изменение конфигурации макромолекулы является пренебрежимо малым и не рассматривается. Топология конфигурационного пространства линейного полимера соответствует топологии многомерного тора. Функция потенциальной энергии (ПЭ) задана на многомерном торе и представлена в виде разложения в многомерный ряд Фурье. Наличие одинаковых или почти одинаковых мономерных звеньев в длинной полимерной цепи приводит к симметрии поверхности ПЭ относительно перестановки соответствующих

угловых переменных. Это приводит к зависимости коэффициентов разложения от инвариантов алгебраических векторов номеров гармоник, что позволяет моделировать конкретные топографии многомерных энергетических поверхностей и вызванные ими кинетические и термодинамические эффекты. Движение репрезентативной точки (РТ) по многомерной поверхности ПЭ описывается уравнениями механики в вязкой среде. В жидких средах (с вязкостью порядка вязкости воды) инерциальными членами для атомных движений можно пренебречь. Методами многомерной геометрии с учетом асимптотических свойств функций, заданных на многомерных сферах можно установить ряд общих закономерностей для решений системы уравнений движения и правил движения РТ по потенциальной поверхности. Полученные правила движения РТ по многомерной поверхности ПЭ подтверждаются расчетом траекторий молекулярной динамики [7].

Результаты: Можно сформулировать следующие правила движения РТ по многомерной поверхности ПЭ макромолекулы в вязкой среде, которые обеспечивают кинематические связи при фолдинге линейных полимеров.

1. Скорость убыли ПЭ на каждом шаге механической траектории максимальна по сравнению с другими виртуальными перемещениями.
 2. Скорость диссипации энергии минимальна по сравнению с другими виртуальными перемещениями. Здесь нет противоречия с п.1 так как утверждения 1 и 2 действуют в разных областях фазового пространства макромолекулы.
 3. Скорость диссипации энергии, а также скорость убыли ПЭ макромолекулы практически равномерно "размазаны" по степеням свободы. Это приводит к гладким траекториям для движения РТ с избеганием резких перепадов рельефа ПЭ по относительно небольшому числу степеней свободы (принцип «начинающего горнолыжника»).
 4. Как следствие, возникает корреляция конформационных движений линейной полимерной цепи с тенденцией формирования спиральных структур.
- Для топографии многомерной поверхности ПЭ линейного полимера с одинаковыми или почти одинаковыми мономерными звеньями возникают следующие закономерности.

1. Поверхность ПЭ для линейных полимеров, построенных из почти одинаковых звеньев, может быть описана в терминах двух обобщенных переменных: суммы квадратов отклонений торсионных углов от точки глобального минимума ПЭ и суммы отклонений углов от точки глобального минимума ПЭ.
2. Топография поверхности ПЭ представляет собой систему вложенных энергетических воронок, разделенных энергетическими барьерами, высота которых убывает по мере удаления от точки глобального минимума. Центральная воронка является самой глубокой.
3. При заданных значениях обобщенных переменных вычисляется энтропия и свободная энергия полимерной цепи. При этом возникает параметр характеристической температуры T_0 , который соответствует удельному выигрышу ПЭ при сворачивании цепи, приходящемуся на 1 конформационную степень свободы в мономерном звене. При $T > 0.26 T_0$ свернутое состояние становится неустойчивым. При температуре денатурации порядка 60С в ситуации альфа-аминокислот (2 степени свободы на мономер в основной цепи) выигрыш в потенциальной энергии на 1 мономер должен составлять примерно 5 ккал/моль, что близко к энергии водородной связи.

Выводы: Вышеперечисленные закономерности топографии поверхности ПЭ и правила движения РТ дают логическое объяснение многим весьма разнородным экспериментальным фактам для кинетики и термодинамики процессов фолдинга и рефолдинга [7]. Что касается физических основ возникновения аминокислотного кода для 3D-структуры белков, то мы формулируем следующую гипотезу. На ранней, добиологической стадии физико-химической эволюции могли возникнуть линейные полимеры, состоящие из почти одинаковых мономерных звеньев. По причинам, изложенным выше, эти линейные полимеры при соответствующих условиях формировали уникальные 3D-структуры. Химическая модификация мономерных звеньев (например, боковых групп) этих полимеров в условиях давления отбора [8], которое направлено на сохранение способности к формированию уникальных 3D-структур могло привести, конечном счете, к набору полипептидов и иных биополимеров, имеющих функциональное значение. Такой механизм формирования молекулярного кода может проектироваться не только на полипептиды, но и на линейные полимеры другой химической природы, включая нуклеотиды или иные структуры, которые оказались, в конечном счете, способными к образованию простейших репликаторов [9].

On the physical basis of the amino acid code for the 3D structure of proteins

Shaitan K.

*M.V. Lomonosov Moscow State University, Moscow, Russia
shaitan49@yandex.ru*

Key words: linear polymers; polypeptides; multidimensional energy surface topography; folding dynamics; effects of viscosity; configuration space topology and symmetry with respect to rearrangements of identical monomer units; denaturation temperature

Motivation and Aim: It is established that the spatial structure of proteins is determined by the amino acid sequence and corresponds to the minimum free energy of the macromolecule [1]. There are some general ideas that the stability of the 3D-structure requires the presence of an energy gap between the native conformation and the nearest metastable conformations of the chain [2], as well as some relatively smooth topography of the multidimensional energy funnel (the principle of minimum frustration [3]) to overcome the Levinthal paradox [4]. A description of the current state of the problem can be found in a recent review [5]. However, the specific physical aspects that determine the folding process have long remained unclear due to the complexity of the object and the lack of adequate mathematical tools to describe the folding process. Computer methods for predicting the spatial structure of proteins, which are not based on the physics of folding, but use AI technologies, have gained popularity [6]. These algorithms work with varying success, mainly for homologous structures. We proceed from the presumption of the primacy of the physical picture and patterns principle for biological processes, no matter how complex they may seem at the moment [7]. Discovering the physical mechanisms of the organization of the amino acid code for the 3D-structure of proteins is of fundamental importance for understanding the physicochemical foundations of life and can significantly change the existing research paradigm in structural biology. The ideas being developed relate not only to

polypeptides, but also to the formation of 3D-structures of other linear polymers and biopolymers.

Methods and Algorithms: The concepts being developed are based on the structure of linear polymers, the conformation of which is determined by rotations along dihedral angles around chemical bonds. The contribution of vibrations in the valence degrees of freedom to the change in configurations of macromolecules is negligible and is not considered. The topology of the configuration space of a linear polymer is a multidimensional torus. The potential energy (PE) function is defined on a multidimensional torus and is represented as an expansion into a multidimensional Fourier series. The presence of identical or almost identical monomer units in a long polymer chain leads to symmetry of the PE surface with respect to the permutation of the corresponding angular variables. This leads to the dependence of the expansion coefficients on the invariants of algebraic vectors of harmonic numbers, which makes it possible to model specific topographies of multidimensional energy surfaces and then the kinetic and thermodynamic effects. The movement of a representative point (RP) along the multidimensional surface of the PE is described by the equations of mechanics in a viscous medium. In liquid media (with a viscosity on the order of that of water), the inertial terms for atomic motions can be neglected. Using the methods of multidimensional geometry and taking into account the asymptotic properties of functions defined on multidimensional spheres, it is possible to establish a number of general patterns for the rules for RP motion on the multidimensional PE surface. The resulting rules for the RP motion are confirmed by the molecular dynamic simulations.

Results: We can formulate the following rules for the RP movement along the multidimensional energy surface in a viscous medium, which based on kinematic connections for linear polymer nodes during the folding [7].

1. The rate of decrease in PE on the mechanical trajectory is maximum compared to other virtual movements.
2. The rate of energy dissipation on the mechanical trajectory is minimal compared to other virtual movements. There is no contradiction here with point 1, since statements 1 and 2 operate in different regions of the phase space of the macromolecule.
3. The rate of energy dissipation, as well as the rate of PE decrease are almost uniformly “spread” over the macromolecule degrees of freedom. This leads to smooth trajectories for the RP motion with the avoidance of sharp changes in the PE relief over a relatively small number degrees of freedom (the “beginner skier” principle [7]).
4. As a consequence, there is a correlation between conformational movements in polymer chain and to form helical structures tendency.

The multidimensional PE surface topography for a linear polymer chain with identical or almost identical monomer units, the following patterns arise [7].

1. The PE surface for linear polymers (built from almost identical units) can be described in terms of two generalized variables: the sum of squared deviations of torsion angles from the PE global minimum point and the sum of the deviations of angles from the PE global minimum point.
2. The topography of the PE surface is a system of nested energy funnels, separated by energy barriers, the height of which decreases with distance from the global minimum point. The central funnel is the deepest.
3. For given values of the generalized variables, the entropy and free energy of the polymer chain are calculated. In this case, a characteristic temperature parameter T_0 arises, which corresponds to the specific negative gain in potential energy during chain

folding per 1 conformational degree of freedom in the monomer unit. At $T > 0.26 T_0$, the folded state becomes unstable. If the denaturation temperature is about 60°C as in the situation of alpha amino acids (2 degrees of freedom per monomer in the main chain), the negative gain in potential energy per 1 monomer should be approximately 5 kcal/mol, which is close to the hydrogen bond energy.

Conclusion: The above-mentioned regularities of the topography of the PE surface and the rules of the RP movement provide a logical explanation for many heterogeneous experimental data for the kinetics and thermodynamics of the folding and refolding processes [7]. As for the physical basis for the emergence of the amino acid code for the 3D structure of proteins, we formulate the following hypothesis. At the early, prebiological stage linear polymers consisting of almost identical monomer units could arise. For the reasons outlined above, these linear polymers formed unique 3D structures under appropriate conditions. Chemical modification of monomer units (e.g. side groups) of these polymers under selection pressure [8], which is aimed at preserving the ability to form unique 3D structures, could ultimately lead to a set of polypeptides and other biopolymers with functional significance. Such a mechanism for the formation of a molecular code can be designed not only for polypeptides, but also for linear polymers of a different chemical nature, including nucleotides or other structures that ultimately turned out to be capable of forming the simplest replicators [9].

Список литературы/References

1. Anfinsen C.B. Principles that Govern the Folding of Protein Chains. *Science*. 1973;181(4096):223-230. doi 10.1016/S0021-9258(18)64176-6
2. Šali A., Shakhnovich E., Karplus M. How does a protein fold? *Nature*. 1994;369(6477):248-251. doi 10.1038/369248a0
3. Onuchic J.N., Wolynes P.G. Theory of protein folding. *Curr Opin Struct Biol*. 2004;14(1):70-75. doi 10.1016/j.sbi.2004.01.009
4. Levinthal C. Are there pathways for protein folding? *J Chim Physique*. 1968;65(1):44-45. doi 10.1051/jcp/1968650044
5. Finkelstein A.V., Bogatyreva N.S., Ivankov D.N., Garbuzynskiy S.O. Protein folding problem: Enigma, paradox, solution. *Biophys Rev*. 2022;14(6):1255-1272. doi 10.1007/S12551-022-01000-1
6. Callaway E. What's next for the AI protein-folding revolution. *Nature*. 2022;604:234-239. doi 10.1038/d41586-022-00997-5
7. Шайтан К.В. Почему белок сворачивается в уникальную 3D-структуру? И не только это... *Химическая физика*. 2023;42(6):40-62. doi 10.31857/S0207401X23060109 [Shaitan K.V. Why Do Proteins Fold into Unique 3D-Structures? And Other Questions... *Russ J Phys Chem. B*. 2023;17(3):550-570. doi 10.1134/S1990793123030259 (in English)]
8. Wong M.L. et al. On the roles of function and selection in evolving systems. *PNAS*. 2023;120(43):e2310223120. doi 10.1073/pnas.2310223120
9. Szostak J. How Did Life Begin? *Nature*. 2018;557:S13-S15. doi 10.1038/d41586-018-05098-w