Bioorganic & Medicinal Chemistry 25 (2017) 4953–4965

Contents lists available at ScienceDirect

Bioorganic & Medicinal Chemistry

journal homepage: www.elsevier.com/locate/bmc

Review article

Recent advances in racemic protein crystallography

Bingjia Yan, Linzhi Ye, Weiliang Xu, Lei Liu *

Tsinghua-Peking Center for Life Sciences, Key Laboratory of Bioorganic Phosphorus Chemistry and Chemical Biology (Ministry of Education), Department of Chemistry, Tsinghua University, Beijing 100084, China

ARTICLE INFO

Article history:
Received 10 March 2017
Revised 3 May 2017
Accepted 9 May 2017
Available online 10 May 2017

Keywords:
Protein enantiomer
Crystallization
Chemical protein synthesis
Solid phase peptide synthesis
Native chemical ligation

ABSTRACT

Solution of the three-dimensional structures of proteins is a critical step in deciphering the molecular mechanisms of their bioactivities. Among the many approaches for obtaining protein crystals, racemic protein crystallography has been developed as a unique method to solve the structures of an increasing number of proteins. Exploiting unnatural protein enantiomers in crystallization and resolution, racemic protein crystallography manifests two major advantages that are 1) to increase the success rate of protein crystallization, and 2) to obviate the phase problem in X-ray diffraction. The requirement of unnatural protein enantiomers in racemic protein crystallography necessitates chemical protein synthesis, which is hitherto accomplished through solid phase peptide synthesis and chemical ligation reactions. This review highlights the fundamental ideas of racemic protein crystallography and surveys the harvests in the field of racemic protein crystallography over the last five years from early 2012 to late 2016.

Contents

* Corresponding author.
E-mail address: lliu@mail.tsinghua.edu.cn (L. Liu).

http://dx.doi.org/10.1016/j.bmc.2017.05.020
0968-0896/

1. Protein crystallography

Proteins are essential biomacromolecules that play versatile roles of catalysts, antibodies, transporters, channels, signal molecules, and structural bricks in biological processes.[1–6] An important goal of protein science is to interpret the protein activities and thus to manipulate them for various applications. For this purpose, the solution of the three-dimensional structures of proteins is a critical step. Several experimental methods have been heavily used in this respect, including nuclear magnetic resonance (NMR) spectroscopy,[7] small-angle X-ray scattering (SAXS),[8] fluorescence resonance energy transfer (FRET) microscopy,[9,10] cryogenic electron microscopy (cryo-EM),[11,12] and X-ray crystallography.[7] Table 1 shows the advantages and limitations of the methods for protein structure determination.

X-ray crystallography with its long application history remains to be an indispensable method in protein structure determination. With protein molecules distributing orderly in crystalline lattice, X-ray crystallography based on X-ray diffraction (XRD) can provide a great amount of detailed information in terms of three-dimensional structures and intra- or intermolecular interactions. Unpredictable crystallization conditions and the generally low success rate of protein crystallization are usually the problems encountered in the use of X-ray crystallography for protein structure determination.[15–21] In this context, one interesting method to overcome the crystallization obstacle is racemic protein crystallography, which utilizes unnatural D-proteins to facilitate protein crystallization. This review focuses on the studies that use and develop racemic protein crystallography.

2. Racemic protein crystallography

Homochirality is one of the distinctive properties of natural proteins,[22] which are all constituted of L-amino acids and achiral amino acid glycine with little exceptions of D-amino acids that are thought to be converted from L-amino acids through post-translational reactions.[23] The crystals of L-proteins are also chiral and therefore devoid of symmetric center, leading to a random phase problem in X-ray diffraction.[15,24] Although the phase problem is no longer urgent ever since the application of multi-wavelength anomalous dispersion methods,[25] racemic protein crystallography was developed to overcome it in the first place. In 1993, Zawadzke and Berg prepared a racemic protein solution (of a 45-residue protein, rubredoxin) containing two enantiomers constituted of either all L- or D-amino acids in an equal amount.[26] The rubredoxin crystal that grew from the solution, for the first time in terms of protein crystal, was achiral, and had symmetric centers, which meant that its phase in X-ray diffraction took exclusively 0 or 180 degree. This work showed the first advantage of racemic protein crystallography in that the symmetric center imposed by the racemic proteins can simplify the steps of phasing diffraction in data analysis.

The second potential advantage, as analyzed by Wukovitz and Yeates in 1995, of racemic protein crystallography is to facilitate crystallization and increase its success rate.[27] According to the theory of space group symmetry, the natural protein crystals always fall into the 65 chiral space groups lacking any second-kind symmetry operations, such as inversion center or mirror reflection.[15,24] Based on the information of space groups symmetries into which numerous distinct protein crystals have been determined, the occurrence frequency of a specific space group symmetry, $P2_12_12_1$, is much higher than those of any other space group symmetry.[27] This phenomenon indicates a preference of crystalline symmetry in protein crystallization.

Wukovitz and Yeates proposed a theory, the entropic model, to explain the preference of protein crystalline symmetry.[27] In their model, for monomeric proteins without any internal symmetry, the occurrence frequencies of space group symmetries correlate with a parameter named the total number of rigid-body degrees of freedom, D, which is independent of the properties of protein molecules but an inherent character of the symmetry itself. A space group symmetry with a larger D value means that the proteins can crystallize into the symmetry in much more ways than those with a smaller D value. Amongst all the chiral space groups, $P2_12_12_1$ has the largest D value (D = 7), resulting in the dominant occurrence frequency of $P2_12_12_1$ in the natural protein crystals, whereas other space groups have a D value ranging from 4 to 6. Expanding the model to 165 achiral space group symmetries, which can be formed by the racemic protein crystals, Wukovitz and Yeates predicted that $P\bar{1}$ (D = 8) would replace the position of $P2_12_12_1$ to become the most frequently occurring space group symmetry for racemic proteins.[27] Besides, Wukovitz and Yeates predicted that racemic protein crystallization would be much easier than that of natural proteins because of the theoretical increase in the number of feasible protein crystalline forms.[27] Indeed, with the accumulation of the experiment data regarding the symmetries of racemic protein crystals, the entropic model is drawing its strength from the accuracy of its prophecy that 12 out of 44 racemic protein crystals determined during the last 5 years fall into the space group $P1$ (see Table 2). Furthermore, accumulating experiment evidence for a higher crystallization rate renders racemic protein crystallography a practically useful method for protein structure determination.

In combination with the developed XRD methods, such as molecular replacement,[24,26,28–30] direct methods,[31–35] experimental methods,[36–38] and anomalous dispersion methods,[24,39] racemic protein crystallography can solve more accessible and well-ordered protein crystals in high resolution.[15] Additionally, racemic crystallography can be used to solve the structures of nuclei acids, such as Pribnow box consensus sequence involved in a non-self-complementary DNA duplex.[40–42]

For some natural proteins with delicate moieties such as glycosylation, total chemical synthesis of their D-enantiomers is a daunting challenge. One solution is to synthesize their quasi-enantiomers and to use quasi-racemates for crystallization. Quasi-racemates or enantiomorphs refer to the pairs of protein molecules that are almost, but not exactly, mirror-image counterparts. Mimicking racemic protein crystallography, quasi-racemates

Table 1
Methods for protein structure determination.

Methods	Advantages	Limitations
NMR	Dynamic information; Atomic resolution	Oxygen information omitted[13]; Protein $\leq$40 kDa[14]
SAXS	Fast; Small sample; Dynamic information	Low resolution (quaternary structure and oligomeric state)[13]
FRET	High spatial and temporal resolution	Local information[9,10]
Cryo-EM	Near-atomic resolution; Various functional states	Proteins usually >200 kDa[11,12]; Expensive equipment[11,12]
X-ray crystallography	Atomic resolution	Low success rate of protein crystallization[7]; Crystalline and static structure[7]