Journal of Structural Biology 189 (2015) 123–134

Contents lists available at ScienceDirect

Journal of Structural Biology

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

Sparse and incomplete factorial matrices to screen membrane protein 2D crystallization

R. Lasala [a,1], N. Coudray [a,1], A. Abdine [b,1], Z. Zhang [a], M. Lopez-Redondo [c], R. Kirshenbaum [b], J. Alexopoulos [c], Z. Zolnai [d], D.L. Stokes [a,c], I. Ubarretxena-Belandia [a,b,*]

[a] New York Structural Biology Center, 89 Convent Avenue, New York, NY 10027, USA
[b] Department of Structural and Chemical Biology, Icahn School of Medicine at Mount Sinai, 1425 Madison Avenue, New York, NY 10029, USA
[c] Skirball Institute of Biomolecular Medicine and Department of Cell Biology, New York University School of Medicine, 540 First Avenue, New York, NY 10016, USA
[d] Department of Biochemistry, University of Wisconsin-Madison, 433 Babcock Drive, Madison, WI 53706, USA

ARTICLE INFO

Article history:
Received 6 June 2014
Received in revised form 18 November 2014
Accepted 24 November 2014
Available online 3 December 2014

Keywords:
Two-dimensional (2D) crystal
Membrane protein
Electron crystallography
High-throughput screening
Membrane protein reconstitution
Electron cryomicroscopy
96-well format

ABSTRACT

Electron crystallography is well suited for studying the structure of membrane proteins in their native lipid bilayer environment. This technique relies on electron cryomicroscopy of two-dimensional (2D) crystals, grown generally by reconstitution of purified membrane proteins into proteoliposomes under conditions favoring the formation of well-ordered lattices. Growing these crystals presents one of the major hurdles in the application of this technique. To identify conditions favoring crystallization a wide range of factors that can lead to a vast matrix of possible reagent combinations must be screened. However, in 2D crystallization these factors have traditionally been surveyed in a relatively limited fashion. To address this problem we carried out a detailed analysis of published 2D crystallization conditions for 12 β-barrel and 138 α-helical membrane proteins. From this analysis we identified the most successful conditions and applied them in the design of new sparse and incomplete factorial matrices to screen membrane protein 2D crystallization. Using these matrices we have run 19 crystallization screens for 16 different membrane proteins totaling over 1300 individual crystallization conditions. Six membrane proteins have yielded diffracting 2D crystals suitable for structure determination, indicating that these new matrices show promise to accelerate the success rate of membrane protein 2D crystallization.

1. Introduction

Membrane protein electron crystallography was pioneered in the 1970s by Henderson and Unwin through their studies of bacteriorhodopsin (Henderson and Unwin, 1975), and relies on electron cryomicroscopy (cryo-EM) of two-dimensional (2D) crystalline specimens of membrane proteins in a lipid bilayer. This method is thus ideal for studying the structure of membrane proteins in their natural membrane environment (Ubarretxena-Belandia and Stokes, 2010, 2012). As in X-ray crystallography, growing suitable crystals represents one of the major bottlenecks in the application of this technique. 2D crystals are typically grown by reconstitution of purified, detergent-solubilized membrane proteins into lipid bilayers at a high enough density to favor the formation of a regular array (Jap et al., 1992; Kühlbrandt, 1992; Mosser, 2001). Several methods, including dialysis (Kühlbrandt, 1992), controlled dilution (Remigy et al., 2003), adsorption onto a hydrophobic resin (Rigaud et al., 1997) or complexation with cyclodextrins (Signorell et al., 2007b) are generally employed for detergent removal and reconsti-

Abbreviations: β-CD, methyl-β-cyclodextrin; DMPC, dimyristoylphosphatidylcholine; DMPS, dimyristoylphosphatidylserine; DOPA, dioleoylphosphatidic acid; DOPC, dioleoylphosphatidylcholine; DOPE, dioleoylphosphatidylethanolamine; DOPG, dioleoylphosphatidylglycerol; DOPS, dioleoylphosphatidylserine; EggPA, egg-yolk phosphatidic acid; EggPC, egg-yolk phosphatidylcholine; POPC, palmytoyl-oleoylphosphatidylcholine; POPE, palmytyl-oleoylphosphatidylethanolamine; soyPC, soybean phosphatidylcholine; C8E4, tetraethylene glycol monooctyl ether; C12E8, octaethyleneglycol mono-n-dodecylether; C8POE, octyl-polyoxyethylene; CHAPS, 3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate; DDM, n-Dodecyl-β-D-maltopyranoside; DHPC, diheptanoylphosphocholine; DM, n-Decyl-β-D-maltopyranoside; FC12, 1-Dodecanoyl-2-Hydroxy-sn-Glycero-3-Phosphocholine; HG, n-Hexyl-β-D-glucopyranoside; HTG, n-Heptyl-β-D-thioglucopyranoside; LDAO, n-Dodecyl-N,N-Dimethylamine-N-Oxide; OG, n-Octyl-β-D-glucopyranoside; OTG, n-Octyl-β-D-thioglucopyranoside; TDC, taurodeoxycholate; Triton X-100, α-[4-(1,1,3,3-Tetramethylbutyl)phenyl]-ω-hydroxy-poly(oxy-1,2-ethanediyl).

* Corresponding author at: Department of Structural and Chemical Biology, Icahn School of Medicine at Mount Sinai, Icahn Medical Institute Bldg., Room 16-70E, 1425 Madison Avenue, Box 1677, New York, NY 10029, USA. Fax: +1 212 849 2456.
E-mail address: Iban.Ubarretxena@mssm.edu (I. Ubarretxena-Belandia).
[1] Equal contributions.

http://dx.doi.org/10.1016/j.jsb.2014.11.008
1047-8477/

tution of the protein into proteoliposomes. Identifying the conditions for growing 2D crystals requires screening over a wide range of factors including pH, temperature, lipid composition, lipid-to-protein ratio (LPR), detergent, amphiphiles, mono- and divalent-ions, inhibitors and ligands. A systematic screen over all of these factors generates a huge matrix of possible reagent combinations, which should ideally be sampled to cover the majority of 2D crystallization space. For 3D crystallization, a vast portion of crystallization space can be screened efficiently and rapidly using sparse (Jancarik et al., 1991; Rupp and Wang, 2004) and incomplete factorial crystallization matrices (Carter, 1990; Gorrec et al., 2011) in combination with high-throughput approaches. In contrast, factors relevant for 2D crystallization have traditionally been surveyed in a relatively limited fashion, potentially missing truly optimal conditions, or in some cases failing to even obtain crystals. However, the recent development of high-throughput tools for 2D crystallization (Cheng et al., 2007; Vink et al., 2007; Coudray et al., 2008, 2011; Hu et al., 2010; Iacovache et al., 2010; Karathanou et al., 2010; Kim et al., 2010) make it now possible to conduct 2D crystallization trials at a higher pace and reproducibility, and moreover, sufficient amount of data is now available on membrane protein 2D crystallization (reviewed in (Abeyrathne et al., 2012)) to allow the rational design of new and more comprehensive 2D crystallization screens. To this end we first built a 2D crystallization database with information mined from successful 2D crystallization conditions reported in the literature. We analyzed this information to evaluate the effect of the different crystallization factors, and from this analysis we designed new sparse and incomplete factorial matrices to screen membrane protein 2D crystallization. Using these matrices we have been able to grow 2D crystals suitable for structure determination for several membrane proteins.

2. Materials and methods

2.1. 2D crystallization database

To construct a database of 2D crystallization experiments we mined the successful conditions from ~250 2D crystallization screens published in ~200 journal articles. To guide us in our literature search we used a recent review by Abeyrathne et al. (2012), which tabulated all the membrane proteins studied by electron crystallography up to the year 2012. We tabulated all crystallization conditions according to different factors including pH, temperature, lipid composition, LPR, detergent, amphiphiles, and mono- and divalent-ions, along with their respective concentrations. We completed this database with additional fields to describe particular properties of the membrane protein, detergent, and lipid that constitute the initial ternary mixture in a typical 2D crystallization experiment. The 2D crystallization conditions were analyzed by constructing a series of bar charts showing the number of entries in the database as a function of individual crystallization factors.

2.2. Design of a sparse matrix 2D crystallization screen

To design the screen in an unbiased manner we applied the k-means algorithm to form 10 groups using 94 successful 2D crystallization conditions from 57 unique membrane proteins. These conditions were chosen to be as non-redundant as possible. The input fields for the algorithm were: phospholipids characterized by their alkyl chain length and headgroup composition (in percentages of phosphatidylcholine (PC), phosphatidylethanolamine (PE), phosphatidylglycerol (PG), phosphatidic acid (PA), phosphatidylserine (PS), cardiolipin (CA) and *Escherichia coli* polar lipid extract), pH, NaCl and $MgCl_2$ concentration, and temperature characterized by its median value and variation (which is non-zero when temperature cycling is used).

2.3. Design of an incomplete factorial 2D crystallization screen

The design of so called "incomplete factorial screens" or "grid screens" relies on three major rules: (1) the level of the different factors are assigned randomly; (2) first-order interactions of the level should be balanced; (3) redundancy should be avoided so setups are as different as possible (Carter and Carter, 1979). In order to design a comprehensive incomplete factorial 2D crystallization screen we first identified the ten main factors affecting 2D crystallization, each one of which represents an axis in a ten-dimensional crystallization space. Second, we assigned discrete levels for each factor that then represent points along the corresponding axis. We note that the low occurrence of some of these levels in the database may not reflect the fact that they are not successful, but rather that they have been rarely used. Nevertheless, these levels have been included in the design of the incomplete factorial matrix. For instance, lipid composition is represented by one axis and the levels include commonly used lipids, such as DMPC, DOPC and *E. coli* lipid extracts, as well as seldom used lipids like DOPG. This approach ensures coverage of a wide range of lipid headgroups, chain lengths and degree of unsaturation. To be compatible with the conventional high-throughput format of 96 conditions per run, our incomplete factorial matrix consisted of 90 conditions (corresponding to points in the ten-dimensional space) plus 6 spots for controls. The 90 points were randomly selected with two constraints: each level should be represented the same number of times and second-order interactions were balanced to avoid redundancy. For instance, the second constraint forced a given lipid to be combined with a multitude of pHs. In this manner over 20,000 trial matrices were generated. In order to identify one matrix covering the largest amount of ten-dimensional space we computed the standard deviation (SD) of the nearest neighbor-distance (nnd). More specifically, for each of the 90 conditions we computed nnd values relative to the remaining 89 conditions. We note that, because we minimized redundancy during generation of the trial matrices, none of the conditions were identical and thus the nnd values were always >0. As a consequence, if the conditions were equidistant from each other in ten-dimensional space, then the nnd values would be equivalent and the SD(nnd) would be zero. Following this argument, the matrix with the lowest SD(nnd) represents the most dispersed matrix, i.e., the matrix that most effectively samples the ten-dimensional crystallization space.

2.4. Lipids, detergents and proteins for 2D crystallization screens

Lipids supplied as powder were purchased from Avanti Polar Lipids (Alabaster, AL) and detergents were bought from Anatrace (Maumee, OH). Detergent solubilized lipid stocks, at a final lipid concentration of 2 mg/ml, were prepared in distilled water by first resuspending the lipid at 10 mg/ml, and then mixing a 200 μl aliquot from this suspension with 800 μl of aqueous solution containing detergent. As previously described (Kim et al., 2010), turbidity measurements were employed to determine the minimal detergent concentration needed to solubilize each lipid species. These detergent solubilized lipid stocks could be stored at 4 °C for up to 4 days, or frozen at −80 °C for long-term storage.

Sixteen different membrane proteins were expressed and purified either in our own laboratories or in collaborating laboratories. These proteins were usually expressed with either N- or C-terminal affinity-tags to allow purification by affinity chromatography. Generally, the linker region between the tag and the target protein