# Популярное описание методов решения фазовой проблемы

ВНИМАНИЕ: этот текст задуман для того, чтобы помочь разобраться в материале, а не для того, чтобы заменить собой информацию из презентаций. Его не надо рассматривать в качестве единственного и авторитетного источника знаний.

Картинок и формул тут мало, но они подразумеваются. Их можно посмотреть в презентациях Лунина.

1. **Краткое описание фактов, которые надо понимать перед тем, как разбираться с фазовой проблемой**. Все они разобраны в лекциях и на предыдущих семинарах.
   1. Белок можно закристаллизовать. Кристалл состоит из многих плотно упакованных элементарных ячеек. Каждая ячейка имеет форму параллелепипеда. Следовательно, чтобы описать ячейку надо указать три вектора - **a**, **b** и **c**. Длины этих векторов и углы между ними описывают строение кристалла.
   2. Если на кристалл белка падает когерентное рентгеновское излучение с определенной (около 1Å) длиной волны, то он начинает рассеивать в определенных направлениях. Направление - это значит что-то типа "5 градусов вверх и три градуса вправо от направления падения излучения, которым облучали кристалл". Для каждого направления можно измерить положение рефлекса (точки) на детекторе (то есть эти самые градусы) и интенсивность рассеяния.
   3. Из положения рефлексов на детекторе можно вычислить:
      1. Параметры кристаллической ячейки - векторы **a**, **b**, **c**, то есть длины этих векторов и углы между ними.
      2. Три индекса для каждого рефлекса - целые числа h, k и l. Каждый рефлекс обладает уникальной комбинацией таких чисел (любые два рефлекса отличаются хотя бы одним индексом). Каждой тройке целых чисел из определенного диапазона соответствует хотя бы один рефлекс (правда, некоторые не удается померить).
   4. Для того, чтобы установить строение закристаллизованного белка используется функция электронной плотности ρ(**r**). Эта функция имеет конкретное значение в каждой точке кристалла (некое число). В центрах атомов значение функции э.п. максимально. Поэтому, зная функцию электронной плотности, можно установить положение атомов.
   5. Функцию электронной плотности, как и любую другую функцию, можно представить в виде т.н. ряда Фурье: ρ(**r**)=∑Fhklcos(2π(hx+ky+lz+ϕhkl)). (Кстати, x, y и z тут - это относительные координаты, выраженные в долях длин векторов **a**, **b**, **c**. Поэтому увеличение, например, x на единицу – то есть переход в соседнюю кристаллическую ячейку – соответствует увеличению значения под косинусом на целое число периодов. То есть мы получаем тоже самое значение, как и положено для кристалла.) Каждая косинусоида называется гармоникой, у нее есть амплитуда и фаза. Каждая косинусоида однозначно определяется тремя целыми числами - h, k и l (это значит, что каждой косинусоиде соответствует уникальная тройка целых чисел, а каждой тройке целых чисел обязательно соответствует какая-то одна косинусоида).
   6. Факт жизни состоит в том, что, измерив интенсивности рефлекса с индексами h, k и l, можно вычислить амплитуду гармоники с соответствующими индексами: измеряемая интенсивность рефлекса есть квадрат этой амплитуды Ihkl=Fhkl2. Это число называется модулем структурного фактора.
   7. Таким образом, измерив интенсивности рассеяния от кристалла белка можно определить:
      1. Параметры кристаллической ячейки - векторы **a**, **b**, **c**, то есть длины этих векторов и углы между ними.
      2. Амплитуды гармоник в Фурье-разложении электронной плотности (они же модули структурных факторов).
   8. В теории рассеяния рассматриваются две основные задачи:
      1. Прямая задача теории рассеяния: известное строение белка (положение атомов) и параметры кристалла, требуется найти интенсивности рассеяния (модули структурных факторов). Эта задача легко решается в два этапа:
         1. Поскольку распределение электронов вокруг атома каждого химического элемента хорошо известно, то зная положение атомов можно вычислить значение функции электронной плотности в каждой точке кристалла.
         2. Зная из предыдущего пункта функцию электронной плотности, можно разложить ее в ряд Фурье. Это не требует каких-то измерений или наблюдений, это чисто вычислительная процедура. Амплитуды гармоник дают нам искомые интенсивности рассеяния (модули структурных факторов).

Кстати, решив прямую задачу теории рассеяния можно сравнить вычисленные модули структурных факторов Fр с наблюдаемыми Fн. Если отличия невелики - то это прямое свидетельство в пользу того, что строение белка и параметры кристалла, которые использовались для расчета, близки к реальности.

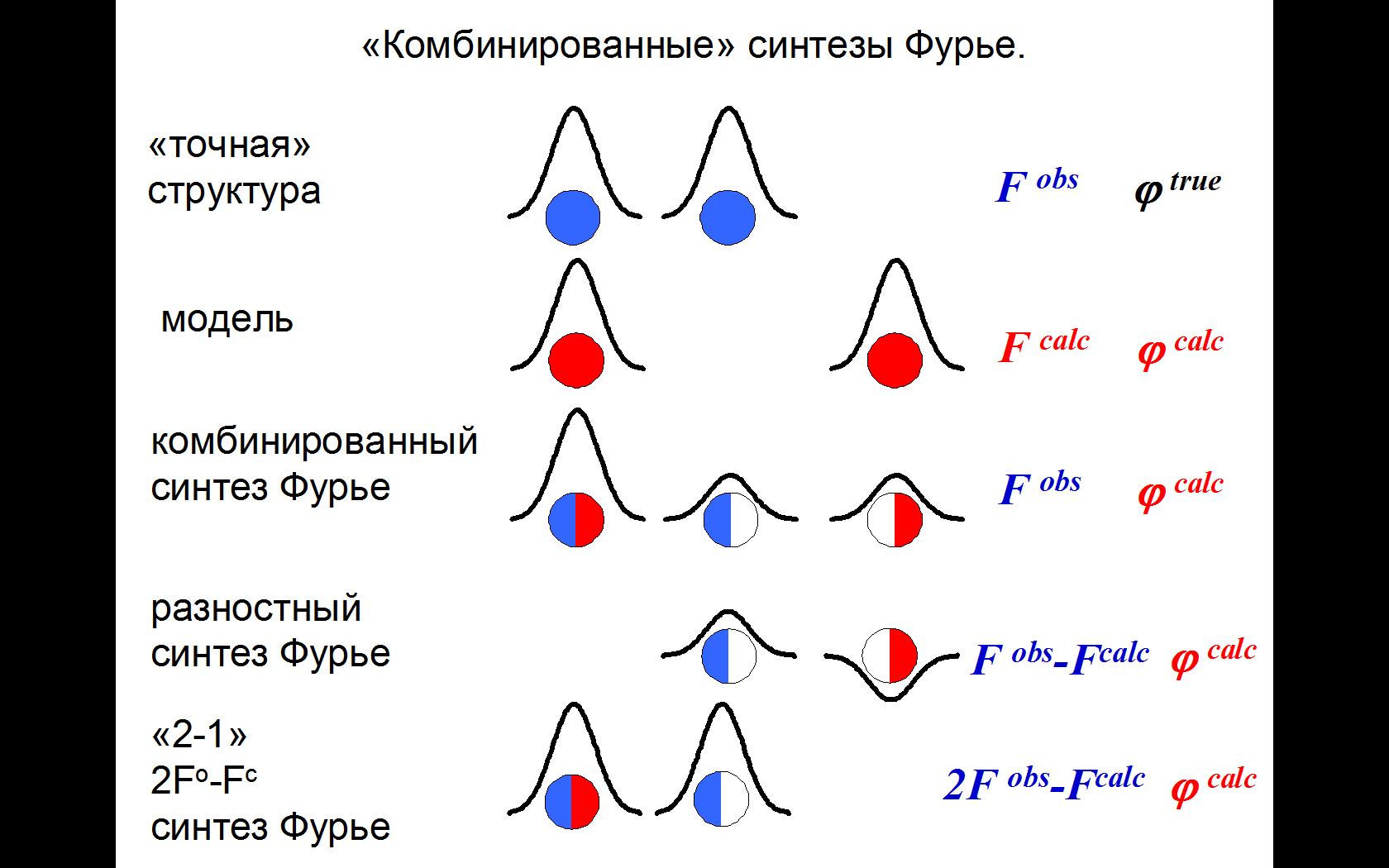
* + 1. Обратная задача теории рассеяния. Это то, в чем, собственно, заключается рентгеноструктурный анализ. Дана картина рассеяния (положение рефлексов и их интенсивности), требуется вычислить параметры кристалла и значение функции электронной плотности. Эта задача решается так:
       1. По положению рефлексов вычислим параметры кристалла.
       2. По интенсивностям рефлексов узнаем амплитуды гармоник в Фурье-разложении функции электронной плотности.
       3. Теперь неплохо бы рассчитать функцию электронной плотности по полученным данным - произвести так называемое обратное преобразование (синтез) Фурье. Действительно, Фурье-разложение состоит из множества гармоник - по одной гармонике для каждой тройки целых чисел h, k, l. Гармоника - это некая косинусоида (трехмерная). Вычислить ее значение в каждой точке можно, если мы знаем ее h, k, l, амплитуду и фазу. Амплитуду для гармоники (hkl) мы можем измерить в эксперименте. Если мы узнаем фазы, то мы решим обратную задачу. Но фазы мы не умеем непосредственно измерять в эксперименте. В этом и заключается ФАЗОВАЯ ПРОБЛЕМА.

1. **Фазовая проблема в обратной задаче теории рассеяния**: Для того, чтобы вычислить функцию электронной плотности, требуется значить амплитуды и фазы гармоник; амплитуды измеряются по картине рассеяния, требуется где-то найти фазы. Существует 4 основных метода решения этой проблемы:
   1. Метод молекулярного замещения. Этим методом сейчас решено большинство структур. Для того, чтобы его использовать, надо знать, хотя бы приблизительно, как белок устроен. То есть где-то в PDB должны лежать трехмерные координаты белка со сходной структурой. Это может быть
      1. Гомолог белка, структуру которого мы пытаемся установить. Если последовательности достаточно сходны, то можно ожидать высокого сходства структур.
      2. Тот же самый белок, но закристаллизованный в других условиях (например, с другим низкомолекулярным лигандом).
      3. Какой-нибудь мутационный вариант.
      4. Теоретическая модель (структура белка, предсказанная биоинформатическими методами). Ну, по крайней мере, теоретически такое возможно.

Допустим, мы хотим решить структуру белка 1XYZ, используя в качестве модели структуру 1ABC. Идея метода состоит в том, чтобы использовать модули структурных факторов от 1XYZ, а фазы вычислить из решения прямой задачи теории рассеяния для 1ABC. Казалось бы, можно было бы связаться с авторами структуры 1ABC и спросить фазы у них, - они же как-то решили фазовую проблему. Но это нам мало поможет, - дело в том, что параметры кристаллов этих двух белков могут быть совершенно разными. Так что фазы придется находит самим, решая прямую задачу для 1ABC. Точнее – для воображаемого кристалла, параметры **a**, **b** и **c** которого такие же, как у реального кристалла 1XYZ, но вместо белка 1XYZ там – 1ABC. Метод молекулярного замещения состоит из следующих стадий:

* + 1. По положению рефлексов от 1XYZ определить параметры его кристалла.
    2. Перебрать все возможные положения 1ABC в ячейке с параметрами, полученными в предыдущем пункте. Всего положение белка в ячейке характеризуется, грубо говоря, шестью числами (на самом деле поменьше) - надо перебрать все их комбинации.
    3. Для каждого возможного положения белка 1ABC в кристалле решить прямую задачу теории рассеяния.
    4. Из каждого решения мы получаем значения рассчитанных структурных факторов Fр. Чем больше они похожи на экспериментальные Fн для 1XYZ, - тем больше похожи кристаллы.
    5. Выбрать положение 1ABC с минимальным значением ∑(Fhklн- Fhklр)2 . Это такое положение, при котором расчетные (для 1ABC) структурные факторы наиболее похожи на экспериментальные (для 1XYZ). Следовательно, эти два кристалла - реальный, который у нас есть (1XYZ), и теоретически представленный нами (1ABC) наиболее сходны. То есть это такое положение 1ABC в кристалле, которое больше всех остальных напоминает расположение 1XYZ в том реальном кристалле, который у нас в руках.
    6. Теперь можно взять рассчитанный кристалл 1ABC и решить для него прямую задачу теории рассеяния. Это и есть искомые фазы.

Тут есть одна проблема – фазы, которые мы рассчитали из решения прямой задачи для 1ABC, вносят свой значительный вклад в результат синтеза Фурье, который мы получим. (Более того, фазы на результат влияют больше, чем модули структурных факторов.) Для того, чтобы приблизить получаемую функцию электронной плотности к ее реальному виду для кристалла 1XYZ, используется такой прием: надо взять фазы, рассчитанные для кристалла 1ABC, а в качестве модулей структурных факторов взять 2Fн-Fр (где наблюдаемые взяты из измерения рассеяния 1XYZ, а расчитанные – из решения прямой задачи для 1ABC). В лекциях Лунина это иллюстрировалось вот такой картинкой:



* 1. Прямые методы. Используются для низкомолекулярных соединений (кристаллы солей, сахаров и т.п.). В принципе, для таких соединений фазовая проблема вообще не возникает. Для них возможно не так уж много конформаций. Просто переберем их все, решим для каждой прямую задачу и выберем ту, для которой значения структурных факторов будут больше всего похожи на реальные. А именно:
     1. По положению рефлексов определить параметры кристалла.
     2. Перебрать все возможные конформации и положения структуры в ячейке с параметрами, полученными в предыдущем пункте.
     3. Для каждого возможной конформации и каждого возможного положения молекулы с такой конформацией решим прямую задачу теории рассеяния.
     4. Из каждого решения мы получаем значения рассчитанных структурных факторов Fр. Чем больше они похожи на экспериментальные Fэ для 1XYZ, - тем больше похожи кристаллы.
     5. Выбрать положение и конформацию с минимальным ∑(Fhklн- Fhklр)2. Это такое строение кристалла, при котором расчетные значения структурных факторов наиболее близки к реальным. Следовательно, это наиболее вероятное строение реального кристалла.

Из решения прямой задачи теории рассеяния тут можно получить и значения структурных факторов. Для простых соединений они не нужны, но понадобятся в следующем пункте.

(В реальности, вместо прямого перебора всех мыслимых конформаций применяется некий хитрый математический прием. Если интересно, - можно посмотреть в презентациях за прошлые годы, там есть такая функция Паттерсона. Поскольку мы занимаемся белками, а не простыми соединениями, то не будем останавливаться на подробностях. Важно понимать, что для простых соединений легко вычислить значения фаз непосредственно из экспериментальных данных. Наиболее понятный прием я тут описал.)

К белкам подобные методы, конечно, неприменимы, так как число возможных конформаций невероятно велико.

* 1. Метод изоморфного замещения. Суть его состоит в том, что изучается рентгеновское рассеяние от двух кристаллов. Первый - это кристалл исследуемого белка. Кроме него, такой же кристалл вымачивают в солях тяжелых металлов. Поскольку кристаллы белков содержат довольно много полостей, то ионы этих металлов могут проникать внутрь кристалла. Если повезло - на поверхности белка есть сайты, к которым такие ионы хорошо прилипают. Задача состоит в том, чтобы получить кристалл т.н. изоморфного производного. Это кристалл нашего белка, к которому в каждой кристаллической ячейке в одном и том же месте (или в нескольких местах) прикреплен ион тяжелого металла.  
       
     Как рассеивает кристалл изоморфного производного? Фактически, это два кристалла в одном - есть кристалл белка, он как-то рассеивает + есть еще кристалл, состоящий из атомов металла, который тоже как-то рассеивает. Причем важно, что параметры элементарной ячейки у них совпадают. Это значит, что рассеяние происходит в одних и тех же направлениях (эти направления зависят только от векторов **a**, **b**, **c**). Следовательно, рассеяния от этих двух совмещенных кристаллов складываются, и мы можем измерить их сумму.   
       
     Поэтому, получаются следующие стадии:
     1. Измерить рассеяние кристалла белка, по положению рефлексов определить параметры кристалла. Запомнить интенсивности рассеяния.
     2. Выдержать кристалл белка в соли тяжелого металла, получить изоморфное производное, измерить рассеяние от такого кристалла. Запомнить интенсивности рассеяния.
     3. Интенсивности, полученные в п. ii являются суммой интенсивностей рассеяния кристалла белка (из п. i) и интенсивностей рассеяния воображаемого кристалла, состоящего только из тяжелых атомов. Iпр=Iб+Iм. Следовательно, можно вычислить Iм – интенсивности рассеяния воображаемого кристалла из тяжелых атомов.
     4. Решить для воображаемого кристалла из тяжелых атомов фазовую проблему прямыми методами.
     5. Решить для него прямую задачу теории рассеяния. Отсюда можно получить искомые значения амплитуд и фаз всех гармоник для кристалла из тяжелых атомов.
     6. На самом деле, в предыдущем пункте получается не одно, а два возможных значения фазы для каждого структурного фактора. Поскольку структурных факторов много (десятки тысяч), то всему кристаллу белка соответствует порядка 2^10000 комбинаций значений фаз. Это очень много. Это нельзя перебрать. Поэтому для того, чтобы решить фазовую проблему, надо получить не одно, а два изоморфных производных (с разными тяжелыми металлами). Каждое из них дает два возможных значения фазы для каждого структурного фактора, - надо просто выбрать совпадающие.  
          
        Далее следует довольно хитрое математическое объяснение, которое можно опустить. Если кто-то хочет на самом деле понять, как там все происходит, - можете прочитать и написать соответствующие формулы. Если это слишком сложно - просто пропустите это рассуждение.  
          
        Итак. Есть такая штука - комплексные числа. Это точка на плоскости, у нее две координаты – действительная и комплексная часть. Представление таких чисел на плоскости тут удобно потому, что сложение таких чисел сводится к сложению факторов. И есть формула Эйлера, что eiα=cosα+i\*sinα. То есть любое выражение типа Feiα есть комплексное число, или, что тоже самое, вектор длины F, который образует с осью Ox угол α (последнее следует из определения косинуса). И еще одна формула тут понадобится: cosα=0.5\*(eiα+e-iα).

Так вот - вместо любой косинусоиды, например, гармоники, cos(blablabla) можно написать ei\*blablabla+e-i\*blablabla (ну, там еще будут какие-то коэффициенты). И потом вот это e-i\*… можно смело заменить на ei\*…, просто придется еще изменить знаки у h, k и l. Таким образом вместо суммы косинусов можно записать сумму экспонент. Подробные выкладки написаны в конце этого файла.

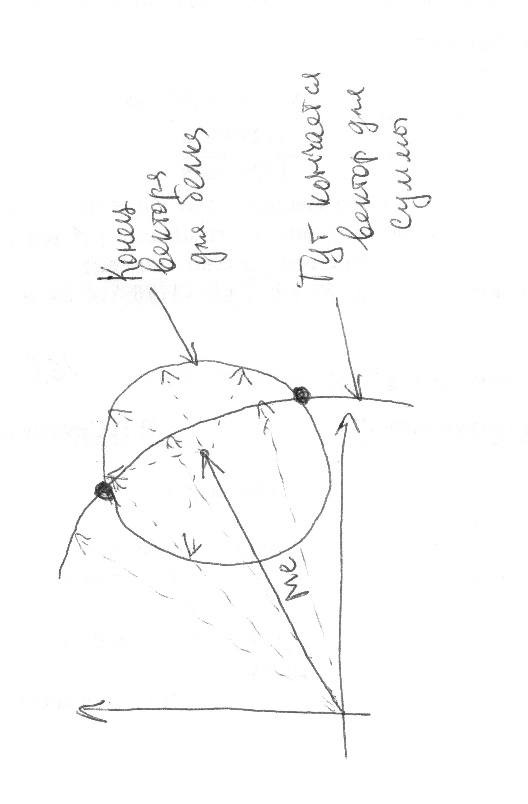
Итак, у нас есть функция электронной плотности ρ(**r**), которая есть сумма таких функций для кристалла белка ρб(**r**) и кристалла тяжелых атомов ρм(**r**). Каждую из этих функций можно разложить в ряд Фурье, в том числе записанный через комплексные экспоненты. Сами функции, конечно, имеют действительные значения, а ряд этот так устроен, что сумма мнимых частей для каждого ряда равна нулю (это связано с тем, что каждая косинусоида соответствует двум экспонентам, каждая экспонента – комплексному числу, мнимая часть которого есть синус, под этими синусами будут значения равные по модулю, но с разными знаками, а синус – нечетная функция).

Удобство комплексных экспонент в том, что фазу можно выделить в отдельную экспоненту. Там, где под косинусом у нас была сумма координат и фазы, такая же сумма будет под знаком экспоненты, а экспонента суммы – это произведение экспонент. Поэтому формулу ρ(**r**)=ρб(**r**)+ρм(**r**) легко можно переписать в виде  
∑Fhkle-i2πϕe-i2π(…)=∑Fhklбe-i2πϕ(б)e-i2π(…)+∑Fhklмe-i2πϕ(м)e-i2π(…)Выражения вида e-i2π(…) – это гармоники, которые зависят только от координат, и они одинаковы у всех трех Фурье-разложений. Поэтому для каждых h, k и l можно записать

Fhkle-i2πϕ(hkl)= Fhklбe-i2πϕ(hkl, б)+Fhklмe-i2πϕ(hkl,м)

А это уже просто сумма двух комплексных чисел, которая равна некоему третьему комплексному числу.

А комплексные числа можно складывать как векторы по правилу треугольника. Итак, для каждой гармоники (для каждой тройки hkl) одно из этих чисел – для кристалла из тяжелых атомов – мы знаем. Пусть будет число Pм. А для двух других знаем только их модули F (из измерений интенсивностей рассеяния кристалла белка и изоморфного производного). И мы знаем, что P=Pб+Pм. Это позволяет решить простую задачу – надо нарисовать известный нам вектор Pм. Дальше, к нему прибавляется вектор длины Fб, направление которого нам неизвестно. Рисуем окружность такой длины с центром в конце вектора Pм. В результате получается сумма – вектор P из начала Pм в конец Pб, то есть куда-то в какую-то точку на нарисованной окружности. Для P мы тоже знаем только длину F, так что рисуем окружность такого радиуса с центром в начале Pм. Две окружности пересекаются вообще говоря в двух точках, так что у нас получается два возможных значения Pб, и два возможных значения фазы для данного структурного фактора.

  
  
Ну, а используя два разных изоморфных производных, можно уже найти правильные значения фаз.

* + 1. Использовать эти фазы для решения обратной задачи для кристалла белка.
  1. Метод аномального рассеяния. Аномально рассеивающие атомы, в отличие от обычных, рассеивают сферически-несимметрично. В рентгеновской кристаллографии такой атом - селен (Se), входящий в состав селенометионина. В обычных белках селенометионин не встречается, но опыт показывает, что если выращивать бактерию, которая продуцирует требуемый белок и не умеет сама синтезировать MSE, на среде без метионина, но с селенометионином, то можно получить белок, где все MET заменены на MSE. Измеряя рентгеновское рассеяние от кристалла такого белка можно получить значения фаз. Правда, тоже неоднозначно - получается два возможных варианта для каждого структурного фактора. Чтобы выбрать правильный, надо изучать рассеяние при другой длине волны падающего излучения. Преимущество - нужен только один кристалл, а не три, как в методе изоморфного замещения.

PS. Про комплексные числа, ряд Фурье и почему в методе изотропного замещения получается два возможных значения фазы для каждой гармоники.

