分子置換法を用いた X 線結晶構造解析

立命館大学・生命科学部

松村 浩由

Molecular replacement in X-ray crystallography College of Life Sciences, Ritsumeikan University Hiroyoshi Matsumura

(投稿日 2021/6/1、再投稿日 2021/6/7、受理日 2021/6/9)

キーワード:分子置換法、X線結晶構造解析

概要

X線結晶構造解析法によってタンパク質を構造解析するためには、まずそのタンパク質 を結晶化する必要がある。そして結晶が得られれば、その結晶を用いて X 線回折実験を 行い、X線回折強度データを収集する。そこで得られた X 線回折強度データを用いて構造 解析を行うが、この際に位相情報が必要となる。というのも、X 線回折強度データには構 造解析に必要な位相情報が含まれないからである(詳細は、「X 線、位相問題」で検索し 調べていただきたい)。その位相情報を得る手段として、分子置換法(MR 法)、重原子 同型置換法(MIR 法)、多波長/単波長異常分散法(MAD/SAD 法)といった方法が用 いられる。なかでも、類似のタンパク質の立体構造を使って位相情報を得る分子置換法は まず初めに試すべき方法であろう。なぜなら、2021 年 5 月の時点で、Protein Data Bank Japan(https://pdbj.org/)には 17 万を超えるタンパク質の立体構造情報(PDB データ)が登録されているからで、今後構造解析されるタンパク質の多くは、類似のタン パク質の立体構造が登録されている可能性が高いからである。

さて、分子置換法は、1)非対称単位中のタンパク質分子数の見積もり、2)自己回転 関数による非結晶学的回転軸の確認、3)類似タンパク質(モデル分子)の選択と編集、 4)モデル分子の回転/並進、5)結晶中の分子のパッキングの確認という 5 段階に分け られる。これらは、CCP4 Program Suite (1)が導入されている Windows や Mac のパ ソコンさえあれば、1 日で実施・完了できる。ただし、類似タンパク質とのアミノ酸同一 性が低い場合(例えば 20%以下の場合)には、4)の段階で容易に結果が得られずより 多くの時間がかかることをご承知おき頂きたい。

ここでは初めて構造解析を実施する研究者を対象として、CCP4 Program Suite が導入されている Windows のパソコンを用いて、CCP4 Program Suite の Molrep (2) という分子置換法のソフトを用いた手順を解説する。

装置

CCP4 Program Suite (2021 年 5 月現在 v7.1.013) と COOT が導入されている Windows パソコンを準備する。なお、Windows パソコンに CCP4 と COOT を導入する ためには、「CCP4 Windows」で検索し、「CCP4 Program Suite including SHELX and COOT」をダウンロードした後、そのファイルをダブルクリックして「詳細」を選 択した後に実行するとよい。

実験手順

- 1) 非対称単位中のタンパク質分子数の見積もり
- 2) 自己回転関数による非結晶学的回転軸の確認
- 3)類似タンパク質(モデル分子)の選択と編集
- 4) モデル分子の回転/並進
- 5)結晶中の分子のパッキングの確認

実験の詳細

本プロトコールを始める前に

目的タンパク質のアミノ酸配列が既知であり、X 線回折強度データファイル(MTZ File)が得られていることを前提として手順を解説する。

分子置換法の詳細

1) 非対称単位中のタンパク質分子数の見積もり

非対称単位とは、結晶中で行う対称操作により重ねることができない最小の領域のこと を指す。なお、Protein Data Bank に登録されている立体構造情報は、通常は非対称単 位に存在する全ての分子の情報を含んでいる。

非対称単位は、例えば空間群が $P2_1$ の場合には単位胞の 1/2 の大きさを持ち、空間群 が $P4_32_12$ の場合には単位胞の 1/8 となる。このように空間群によって単位胞に対する非 対称単位の割合が異なるので、詳しく知りたい場合には空間群の詳細が記載されている書 籍(例えば、International Tables for Crystallography, Volume A, ISBN: 978-0-470-97423-0)をご参照いただきたい。

さて、分子置換法とは、類似タンパク質を回転・並進させ、非対称単位に存在する目的 分子に対して類似タンパク質を重ねる操作法のことである。そのとき、非対称単位内に複 数の分子が存在すれば、その分子全てに対して類似タンパク質を重ねる必要がある。した がって、まず非対称単位に何分子が含まれているかを以下の手順で見積もることから始め る。

- CCP4i を起動し、Project 名と X 線回折強度データファイル(MTZ File)が存在するフォルダ設定した後に、CCP4iのメインパネルから Molecular Replacement モジュールを選択する(図1)。
- ② Analysis→Cell Content Analysis を選択すると Matthews Cell Content Analysis パネルが起動する(図2)。
- ③ Matthews Cell Content Analysis のパネルにて、MTZ file を Browse から選択す ると、空間群、格子定数、分解能が自動で表示される。
- ④ Molecular weight of protein or nucleic acid の欄に分子量を入力する。
- ⑤ Run Now を選択する。

以下のような結果が表示される(以下は一例)。

Nmol/a	asym Matthews	Coeff %solvent	P(2.60)	P(tot)
1	10.04	87.76	0.00	0.00
2	5.02	75.51	0.00	0.00
3	3.35	63.27	0.10	0.11
4	2.51	51.03	0.58	0.57
5	2.01	38.79	0.30	0.31
6	1.67	26.54	0.00	0.00
7	1.43	14.30	0.00	0.00
8	1.26	2.06	0.00	0.00

通常、表の Matthews Coeff で示されている値は 1.7~3.5 A³/Da 程度とされるため (3)、この場合は、非対称単位中におおよそ 3~5分子が存在すると予想される。この段階 では、分子の数を厳密に決定する必要はなく、ある程度の可能性を検討できれば、次の段 階に進む。

M	Secular Replacement	- Toject Dat	abase oob List - C			nesaProjectDir
Analysi	S				Vie	ew Any File
Model (Seneration				View Files from	n Job
Phaser MR					Search/Sort Da	tabase
Run Molrep	- auto MR				Graphical View	of Project
Run MrBUN	IP				Delete/Archive	Filos
Run Balbes					Delete/Archive	riles
Run AMPLE					Kill Job	
Run SIMBA	D				ReRun Job	
▶ Arcimb	oldo				Edit Job Data	-
AMoRe	Suite				Preferences	
Utilities					System Admini	istration
Phaser Sin	gle Atom MR				1 new up	pdates available
		V <			> Manage Up	dates Exit
Matthe Job title Calculate I Read c	ws - Cell Content A Matthews coefficie rystal parameters	nalysis ent for prote	in only —	1	-	Help
Matthe	Matthews coefficient rystal parameters test up P 2 2 2 0280 b 161.9400	nalysis ent for prote from MTZ file P222.mtz 0 c 214.2930 alp	in only — ha 90.0000 be	ta 90.0000 ga		Browse View
Matthe	Matthews coefficient rystal parameters test up P 2 2 2 .0280 b 161.9400 resolution limit 2.6	nalysis ent for prote from MTZ file P222.mtz 0 c 214.2930 alp	in only ha 90.0000 be	1 ta 90.0000 ga	mma 90.0000	Browse View
Matthe	Matthews coefficient rystal parameters test up P 2 2 2 .0280 b 161.9400 resolution limit 2.6 ular weight	nalysis ent for prote from MTZ file P222.mtz 0 c 214.2930 alp 500 entered in	ha 90.0000 be Daltons	ı ta 90.0000 ga		Browse View
Matthe	Matthews coefficient rystal parameters test up P 2 2 2 .0280 b 161.9400 resolution limit 2.6 ular weight weight of protein of	nalysis ent for prote from MTZ file P222.mtz 0 c 214.2930 alp 500 entered in or nucleic acid 140	in only	u ta 90.0000 ga		Browse View
Matthe	Matthews coefficient rystal parameters test up P 2 2 2 0280 b 161.9400 resolution limit 2.6 ular weight weight of protein o untent analysis	nalysis ent for prote from MTZ file P222.mtz 0 c 214.2930 alp 600 entered in for nucleic a cid 140	ha 90.0000 be	ta 90.0000 ga	mma 90.0000	Browse View
Matthe	Matthews coefficient rystal parameters test up P 2 2 2 .0280 b 161.9400 resolution limit 2.60 ular weight weight of protein of intent analysis Lume : 562	nalysis ent for prote from MTZ file P222.mtz 0 c 214.2930 alp entered in or nucleic a cid 140 22794.000	ha 90.0000 be	د ta 90.0000 ga	mma 90.0000	Browse View
Matthe	Matthews coefficient rystal parameters test up P 2 2 2 .0280 b 161.9400 resolution limit 2.6 ular weight weight of protein of mume : 562 h protein molec a Matthews Coe	nalysis ent for prote from MTZ file P222.mtz 0 c 214.2930 alp 600 entered in for nucleic acid 140 22794.000 sular weight: 14	in only ha 90.0000 be Daltons 0000.0 P(2.60)	ta 90.0000 ga	mma 90.0000	Browse View
Matthe	Watthews coefficient rystal parameters test up P 2 2 2 .0280 b 161.9400 resolution limit 2.60 ular weight weight of protein of materna analysis lume : 562 h protein moleco a Matthews Coe 10.04	nalysis ent for prote from MTZ file P222.mtz 0 c 214.2930 alp 600 entered in for nucleic a cid 140 22794.000 sular weight: 14 ff %solvent 87.76	in only ha 90.0000 be Daltons 0000.0 P(2.60) 0.00	ta 90.0000 ga	mma 90.0000	Browse View
Matthe	Matthews coefficient rystal parameters test	nalysis ent for prote from MTZ file P222.mtz 0 c 214.2930 alp entered in or nucleic a cid 140 2794.000 cular weight: 14 ff %solvent 87.76 75.51	in only ha 90.0000 be Daltons 0000.0 P (2.60) 0.00 0.00	ta 90.0000 ga P(tot) 0.00 0.00	mma 90.0000	Browse View
Matthe	Matthews coefficient rystal parameters test	nalysis ent for prote from MTZ file P222.mtz 0 c 214.2930 alp entered in or nucleic a cid 140 22794.000 cullar weight: 14 ff %solvent %7.76 75.51 63.27 63.27	in only ha 90.0000 be Daltons 0000.0 Da P(2.60) 0.00 0.00 0.10 0.59	ta 90.0000 ga P(tot) 0.00 0.01 0.11	mma 90.0000	Browse View
Matthe	Watthews coefficient rystal parameters test up P 2 2 2 .0280 b 161.9400 resolution limit 2.60 ular weight ular weight ume : 562 h protein molec a Matthews Coe 10.04 5.02 3.35 2.51 2.01	nalysis ent for prote from MTZ file P222.mtz 0 c 214.2930 alp entered in for nucleic a cid 140 c2794.000 cular weight: 14 ff %solvent %7.76 75.51 63.27 51.03 38.79	in only	<pre>ta 90.0000 ga P(tot) 0.00 0.00 0.11 0.57 0.31</pre>	mma 90.0000	Browse View
Matthe	Watthews coefficient rystal parameters test up P 2 2 2 .0280 b 161.9400 resolution limit 2.60 ular weight of protein of mume : 562 h protein moleco n Matthews Coefficient 10.04 5.02 3.35 2.51 2.01	nalysis ent for prote from MTZ file P222.mtz o c 214.2930 alp c 21	in only ha 90.0000 be Daltons 0000.0 P (2.60) 0.00 0.00 0.10 0.58 0.30	<pre>ta 90.0000 ga P(tot) 0.00 0.00 0.11 0.57 0.31</pre>	mma 90.0000	Browse View

- 2)自己回転関数による非結晶学的回転軸の確認 非対称単位に複数の分子が存在することがわかれば、次に自己回転関数を計算し、非結 晶学的回転軸(NCS)の存在を確認する。
- ① Molecular Replacement モジュール内の Molrep auto MR を選択する。
- Molrep パネルにて、Do Self rotation function を選択し、MTZ file を Browse から 選択する(図3)。
- ③ 必要であれば、SRF options 内の Select Chi sections を選択して、NCS の候補の角度を入力する。例えば、2 量体が 5 つ集まった 10 量体が存在している分子では、分子内に 2 回軸が 5 本と 5 回軸が存在する。初期値で 180 度は設定されているので、設定されていない 5 回軸(360 度を 5 で割った数字 72)を Select Chi sections に入力すると良い。
- ④ Run→Run Now を選択する。



図4:極座標で示された自己回転関数の例。

1)の①で指定したフォルダに .ps という拡張子のついた Postscript ファイルが出力される。このファイルを Adobe Distiller で PDF ファイルに変換した後に開くと(Adobe Illustrator ではそのまま閲覧可能)、極座標で示された図が表示される(図 4)。

図4では、2回軸の存在を示す Chi = 180 のセクションに5つの高いピーク(上図の 赤矢印)が等間隔で存在しており、Chi = 72 のセクションに1つの高いピーク(上図の 青矢印)が存在していることから、2量体が5つ集まった10量体が存在している可能性 が予想される。この自己回転関数の計算結果と、1)で見積もった非対称単位内の分子数 とを比較して、結晶中で分子が何分子どのように配置しているかを予想しておく。

ただし、この 2) は分子内に NCS が存在して、なおかつそれが結晶学的回転軸とは異 なる場合にのみ有用となる。そのため、図 4 のような明解な情報が得られない場合も多 く、その場合は気にせず次の段階に進む。

	PD	B	178229 Biologic Macromolecular	al Structures	E	nter search l	term(s)			۹
PRO	TEIN DAT?	BANK	Research and E	ducation		ced Search [Browse Annotat	ons	788	Help 🖒
9 PDB	-101		taResource	NUCLENC ACHD	forldwide otein Data Bank undation	Ce	lebrating 🛓	YEARS OF Protein	Data Bank 😭	Y DO
Sea	rch History	Brows	e Annotations	MyPDB						
Use th	e Advanced Sea	arch Query F	Builder tool to a	reate compos	site boolean que	ries. See the	e Help page fo	r more detailed i	nformation.	
• Adv	anced Search	h Query B	uilder Ø							Help 🕑
- 744	→ Attribute	i query Di								5
	- Sequence 😧									
<		EKNF IKWCST	KSMLVVNGSFPG			WHGVKOPRN FLGDLYNCSK	PWSDGPEYITO	CPIKPGTNFIYEVIL:		
	PDB ID 1MBN	Target Pro	tein 🗸 🔞 E-Val	ue Cutoff 0.1	Identity (Cutoff 0	% (Integer o	nly) 🔞	Cour	nt Clear
	 Sequence Motif 	0								
	 Structure Simila 	arity 😡								
	Observational Market	0								
	· Structural Motir									\frown
	Chemical									
	Chemical					Display F	Results as 🔞	Polymer Entities	Count	client of

3)類似タンパク質(モデル分子)の選択と編集

まず、Protein Data Bank(<u>https://www.rcsb.org/</u>)に登録されているタンパク質の 中で、目的タンパク質とアミノ酸同一性の高いタンパク質(モデル分子)を選択する。モ デル分子の選択は様々な方法があるが、以下では、Protein Data Bank(<u>https://www.r</u> <u>csb.org/</u>)から同一性の高いものを選択する方法を示す。

- ① <u>https://www.rcsb.org/</u>にアクセスし Search→Sequence Search→Advanced Search Sequence Search を選択する(図 5)。
- ② Sequence? の空欄にアミノ酸配列(一文字記号)を入力する。
- ③ Display Results as で Polymer Entities を選択し、虫眼鏡のマークを選択する。
- ④ Sequence identity と E-value を確認する。私見であるが Sequence identity が 20%

以上あれば分子置換法を試す価値はあると思われる。Download Alignment を選択し て Alignment ファイルをダウンロードしておく。その Alignment ファイルは、 CLUSTALX (http://www.clustal.org/clustal2/) というソフト等で表示できる。

 ⑤ Sequence identity が高く E-value が小さい PDB ID を選択し、Download Files から PDB Format を選択して PDB ファイルをダウンロードし(図 6)、そのファイルを 1)の①で指定したフォルダに保存する。



次に、テキストエディタ(例えば、Windows アクセサリのメモ帳)で PDB ファイル を開いてファイルの内容を確認する。PDB ファイルの主な行を説明すると、HEADER で 始まる行には PDB ID が示され、CRYST1 には単位格子と空間群の情報が書かれている。 ATOM にはタンパク質の座標データが記され、HETATM にはタンパク質に結合している 基質アナログや水分子などの座標データが書かれている。

蛋白質科学会アーカイブ, 14, e103 (2021)

HEADER TITLE COMPND	CE SI MC	ELL CI TAPHYI DL_ID	YCLE LOCO(: 1;	CCI	JS AUREU	JS	FTSZ-GI	OP II	N T A	(ND F)5-NOV R STAI	V-16 TES	5H5G		
COMPND	2 M	10LECU	JLE:	CE	ELL DIVI	IS	ION PRO	FEIN	FTSZ	;					
(哈) CRYST1	43	917	159	02	2 44	0	58 90 (0	92 78	90	00 0	⊃ 1 2 ⁻	1 1	4	
ORTGX1	10.	1.000	1000	. 02	0,00000	• • •)	0.0000	0	2.10	0.0				1	
ORIGX2		0.000	0000	1	.000000))	0.0000	0		0.0	00000				
ORIGX3		0.000	0000	(.000000)	1.00000	0		0.0	00000				
SCALE1		0.022	2770	(.000000)	0.00110)4		0.0	00000				
SCALE2		0.000	0000	C	.006288	3	0.0000	00		0.0	00000				
SCALE3		0.000	0000	C	.000000)	0.02272	24		0.0	00000				
ATOM	1	Ν	MET	А	11		70.851	24	.422	76.	.500	1.00	58.77		Ν
ATOM	2	CA	MET	А	11		69.861	23	.493	75.	.871	1.00	60.16		С
MOTA	3	С	MET	А	11		70.402	23	.013	74.	.526	1.00	54.98		С
MOTA	4	0	MET	А	11		70.878	23	.819	73.	.724	1.00	52.56		0
ATOM	5	CB	MET	А	11		68.489	24	.170	75.	.715	1.00	65.01		С
ATOM	6	CG	MET	А	11		67.383	23	.242	75.	.219	1.00	70.22		С
ATOM	7	SD	MET	А	11		65.810	23	.448	76.	.090	1.00	79.56		S
ATOM	8	CE	MET	А	11		64.628	22	.796	74.	.912	1.00	76.98		С
ATOM	9	N	ALA	A	12		70.336	21	.703	74.	.294	1.00	51.77		Ν
ATOM	10	CA	ALA	A	12		71.092	21	.069	73.	.206	1.00	46.93		C
ATOM		C	ALA	A	12		/0.521	21	.394	/1.	.823	1.00	42.06		C
ATOM	12	0	ALA	A			69.319 71 1.CA		.313	/1.	.612	1.00	42.26		0
ATOM	13	CB	АLА	А	$\perp \angle$		/1.164	19	.568	13.	.419	1.00	4/.58		C
(哈)															
TER	2220			P	1.0		20 445	1.0	760		100	1 0 0	04 67		
ATOM	2229	N	HIS	В	10		30.445	-18	./63	57.	.180	1.00	84.67		N
ATOM	2230	CA	HIS	В	10		29.963	-18	.600 107	50.	.760	1.00	88.20		C
	2231		пто	D	10		30 737	-19	057	53	965	1 00	80 18		
ATOM ATOM	2232	CB	пто	B	10		29 535	-20	216	55	116	1 00	89 27		C
ATOM	2233	CG	HTS	B	10		29.024	-17	028	54	051	1 00	92 35		C
ATOM	2235	00 1	HTS	B	10		27 735	-17	348	53	681	1 00	94 12		N
ATOM	2236	CD2	HIS	В	10		29.628	-16	.553	52.	.937	1.00	93.17		C
ATOM	2237	CE1	HIS	В	10		27.565	-17	.076	52.	400	1.00	93.87		C
ATOM	2238	NE2	HIS	В	10		28.699	-16	.591	51.	.925	1.00	96.73		Ν
ATOM	2239	Ν	MET	В	11		32.176	-18	.498	54.	.738	1.00	78.54		Ν
ATOM	2240	CA	MET	В	11		33.291	-18	.831	53.	.841	1.00	75.35		С
ATOM	2241	С	MET	В	11		34.497	-19	.169	54.	.713	1.00	66.55		С
ATOM	2242	0	MET	В	11		34.821	-18	.415	55.	.632	1.00	67.02		0
ATOM	2243	CB	MET	В	11		33.608	-17	.626	52.	.948	1.00	80.23		С
ATOM	2244	CG	MET	В	11		34.364	-17	.949	51.	.667	1.00	84.89		С
ATOM	2245	SD	MET	В	11		33.314	-18	.347	50.	.247	1.00	90.91		S
ATOM	2246	CE	MET	В	11		32.482	-16	.780	49.	.969	1.00	92.25		С
(略)															
TER	4475		PHE	В	315										
HETATM	4476	PB	GDP	А	401		73.156	10	.830	45.	.052	1.00	29.42		Ρ
HETATM	4477	01B	GDP	А	401		74.218	10	.572	44.	.048	1.00	33.95		0
HETATM	4478	02B	GDP	А	401		72.385	12	.106	44.	.954	1.00	30.71		0
HETATM	4479	03B	GDP	А	401		73.676	10	.647	46.	.448	1.00	36.84		0
HETATM	4480	03A	GDP	Α	401		72.122	9	.629	44.	.803	1.00	30.55		0
G(略)					_	_									
MASTER		319	0		3 20	5	23	0	15	6	4673	2	63	48	
END															

これらの行のうち、分子置換法で必要な情報は主に ATOM の行である(HEADER、 CRYST1、HETATMの行を残しても問題ない)。また先にも述べたが、PDB ファイルは 非対称単位に存在する全ての分子の情報を含んでいるため、複数の分子の情報が含まれて いる場合も多い。分子置換法では、そのうちのどの分子をモデル分子として選択するかが ポイントとなる(補足 1)。例えば、1 分子のみをモデル分子として使いたい場合には、 残りの分子の座標をテキストエディタで削除する(例えば、上に示した赤字の行を削除し END を残す)。そしてそのファイルをセーブした後、COOT で分子を表示し、思惑通り に1分子のみになっているかを確認すると良い。

Job title [No	title given]							_
Do Mol	ecular Replac	ement -	-					
Use MAP file	s for 🗆 se	earch model						
Data	A1 .	P1.mtz					Browse	View
	F SIG	iF		F_A1		SIGF_A1		_
Model	A1 -	5xdv.pdb					Browse	View
Sequence	A1 -						Browse	View
Fixed	A1 -	_					Browse	View
T Automati	c output filena	ame						
Solution	A1 -	5xdv_mo	irep1.pdb				Browse	View
Search Optic	ons							Γ
Experimenta	l Data							Г
Model								Γ
Infrequently	used options							
	Due			Save or	Poetoro -		Close	

4) モデル分子の回転/並進

ここでは、3)で準備したモデル分子を回転/並進し、結晶の非対称単位中にモデル分子を当てはめていく操作を実施する。

- ① Molecular Replacement モジュール内の Molrep auto MR を選択する。
- Molrep パネルの Data の行の Browse から MTZ file を選択し、Model の行の Browse から準備したモデル分子の pdb file を選択する(図 7)。
- ③ 非対称単位中の分子数が判っていれば、Search Options の Number of copies to find に非対称単位中の分子数を入力する(ただし、初期設定で分子を複数当てはめる 設定となっているため、不明であれば入力する必要はない)。
- ④ Run→Run Now を選択する。

Mole	ecular R	eplace	ment		29 May : F	INISHED molr		[No title	^	Directories&	ProjectDir			
Analysis	🔷 qtR	View 1	.16 - Job	5: [No title give	en]								- 0	\times
Model G	File	Edit	Windo	w Help										_
Run Molrep -	-		1	2	()	e	9. O	0					\sim	0
Run MrBUM	Print	t Pl	DF/PS	Refresh	molrep	CCP4 F	ind Bac	k Forwa	rd				Preferences	Exit
Run Balbes	Resu	Its (Log Fil	e										
Dun AMDI E		12		1.849	0.984	1.00	1.00	1.00	0.640 0	.3838	25.40	(0.6035)		f
	INF	0: C	ontras	st and Th	/sig are	e good en	ough. S	top thi	s run					
RUN SIMBAD														
Arcimbo						Summ	ary (VO)						
AMoRe :	+		E TE	thota	nhi	chi	+ v	+17	+	TE/ea	UDfac	Score	+	
Utilities	+		E IE	LIIELA	PIII					IF/Sy	WRIac	SCOLE	+	
Phaser Sing	1	1	1 1	146.22	135.90	4.21	0.000	0.000	0.281	12.51	0.531	0.60353	1	
	i :	2	2 1	180.00	0.00	3.57	0.504	0.000	0.282	8.40	0.572	0.52812	i l	
	1	3	3 10	43.22	-92.26	175.36	0.870	0.000	0.278	1.48	0.636	0.39007	1	
	1	4	4 2	33.93	1.03	157.09	0.622	0.000	0.292	1.63	0.637	0.38946	I	
	1	5	5 2	36.02	-1.64	176.28	0.478	0.000	0.297	1.65	0.638	0.38703	1	
	1	6 1	2 1	71.43	154.12	49.37	0.977	0.000	0.278	1.85	0.640	0.38378	1	
	1	7	6 6	139.35	-164.51	113.89	0.240	0.000	0.220	1.54	0.640	0.38322	1	
	1	8 1	1 4	70.65	140.30	36.15	0.326	0.000	0.278	1.56	0.643	0.37775	1	
		9 1	0 6	2.34	17.99	117.27	0.000	0.000	0.223	1.31	0.643	0.37635		
	1 1	0	8 11	156.70	121.38	81.43	0.566	0.000	0.306	1.44	0.642	0.37616		
	1 1	1	/ 5	42.77	-176.14	136.31	0.597	0.000	0.329	1.60	0.636	0.37598		
	1 1	2	9 3	87.21	-140.37	/0.30	0.329	0.000	0.250	1.35	0.648	0.36549		
	+													- h
	cor	rF =	0.6	5023										
1														

クリックし、Log File タブ(青丸)内のほぼ最後にある Summary を確認する。

⑤ CCP4i メインパネルの JOB が FINISHED となっているのを確認後、その JOB をダブ ルクリックし、Log File タブ内のほぼ最後にある Summary を確認する(図 8)。以 下のような結果が示される。

RF TF theta phi chi tx ty tz TF/sg wRfac Score 1 1 1 146.22 135.90 4.21 0.000 0.000 0.281 12.51 0.531 0.6035 2 2 1 180.00 0.00 3.57 0.504 0.000 0.282 8.40 0.572 0.5281 3 3 10 43.22 -92.26 175.36 0.870 0.000 0.292 1.63 0.637 0.3894 5 5 2 36.02 -1.64 176.28 0.478 0.000 0.297 1.65 0.638 0.3870 6 12 1 71.43 154.12 49.37 0.977 0.000 0.220 1.54 0.640 0.3832 7 6 6 139.35 -164.51 113.89 0.240 0.000 0.220 1.54 0.640 0.3832 8 11 4 70.			Summary (V0)													
1 1 1 146.22 135.90 4.21 0.000 0.000 0.281 12.51 0.531 0.6035 2 2 1 180.00 0.000 3.57 0.504 0.000 0.282 8.40 0.572 0.5281 3 3 10 43.22 -92.26 175.36 0.870 0.000 0.292 1.63 0.637 0.3894 4 4 2 33.93 1.03 157.09 0.622 0.000 0.297 1.65 0.638 0.3870 5 5 2 36.02 -1.64 176.28 0.478 0.000 0.297 1.65 0.640 0.3837 6 12 1 71.43 154.12 49.37 0.977 0.000 0.220 1.54 0.640 0.3837 7 6 6 139.35 -164.51 113.89 0.240 0.000 0.223 1.54 0.640 0.3837 8 11 4 70.65 140.30 36.15 0.326 0.000 0.223 1.31	+		RF	 	theta	phi	 	tx		tz	TF/sg	wRfac	Score			
10 8 11 156.70 121.38 81.43 0.566 0.000 0.306 1.44 0.642 0.3761 11 7 5 42.77 -176.14 136.31 0.597 0.000 0.329 1.60 0.636 0.3759 12 9 3 87.21 -140.37 70.30 0.329 0.000 0.250 1.35 0.648 0.3654		1 2 3 4 5 6 7 8 9	1 2 3 4 5 12 6 11 10	1 10 2 2 1 6 4 6	146.22 180.00 43.22 33.93 36.02 71.43 139.35 70.65 2.34	135.90 0.00 -92.26 1.03 -1.64 154.12 -164.51 140.30 17.99	4.21 3.57 175.36 157.09 176.28 49.37 113.89 36.15 117.27	0.000 0.504 0.870 0.622 0.478 0.977 0.240 0.326 0.000	0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000	0.281 0.282 0.278 0.292 0.297 0.278 0.220 0.278 0.223	12.51 8.40 1.48 1.63 1.65 1.85 1.54 1.56 1.31	0.531 0.572 0.636 0.637 0.638 0.640 0.640 0.643 0.643	0.60353 0.52812 0.39007 0.38946 0.38703 0.38378 0.38322 0.37775 0.37635			
	 	10 11 12	8 7 9	11 5 3	156.70 42.77 87.21	121.38 -176.14 -140.37	81.43 136.31 70.30	0.566 0.597 0.329	0.000 0.000 0.000	0.306 0.329 0.250	1.44 1.60 1.35	0.642 0.636 0.648	0.37616 0.37598 0.36549			

この Summary には、分子を回転(theta, phi, chi)、並進して(tx, ty, tz)した際の TF/sg、wRfac、Score といった情報が示されている。一般に正しい解が得られたとき、 TF/sg と Score が他の解に比べて有意に高く、wRfac が有意に低くなる。この例では、 一番上の行の解が正しい解であると予想される。

そして Log File のほぼ最後の行には、Molrep が選択した解が以下のように表示される。

Nmon RF TF theta phichi txtytzTF/sgwRfacScore11146.22135.904.210.0000.28112.510.5310.603

モデル分子との同一性にもよるが、正しい解が得られている場合に、wRfac は 50%台 になることが多い。ここで有意な解が得られない場合には、モデル分子の変更(補足 1)、 X線回折強度データの質の確認(補足 2)、分解能の変更(補足 3)などを検討すると良 い。また、X線結晶構造解析を始めたばかりのときには思わぬ落とし穴にはまることもあ る(補足 4)。ここに書いたことを一通り試しても解が得られない場合には、X線結晶構 造解析を複数回経験したことのある研究者に相談してみると良い。



5)結晶中の分子のパッキングの確認 得られた解の妥当性を評価するために、結晶中の分子のパッキングを確認する。

① CCP4i メインパネルからその JOB をダブルクリックし、Results タブ内の Output

Files から Coot を選択する(図 9 左)。Coot が起動した後、Nomenclature error などが出る場合には Yes を選択して、分子置換後のタンパク質分子を表示する。

- ② Coot ウインドウ内の Display Manager を選択し、右のカラムの Bonds (Colour by Atom) を C-alphas/Backbone に変更し、Close を選択して Display Manager を閉 じる。
- ③ Coot ウインドウ内の Draw から Cell & Symmetry… を選択し、Master Switch: Symmetry On、Symmetry by Molecule…から Display as CAs を選ぶ。Symmetry Atom Display Radius: Radius を 30~50 程度に設定し、Apply を選択すると、分子 の周りに対称操作で関係づけられた分子が表示される(図9右)。

結晶は固体であるので、結晶内では分子同士が接触しているはずである。したがって、 上述の操作で隣接する分子を表示させた際に、分子同士が接触しているかどうかを確認す る。また、分子が存在しない大きな空間が存在する場合には、その空間に分子が存在する 可能性も検討する(補足 5)。一方で、この時点で分子同士が重なってしまっている場合 には、解が妥当でない可能性がある。そうした場合には、モデル分子の変更(補足 1)、 X線回折強度データの質の確認(補足 2)、分解能の変更(補足 3)を検討し、4)の操 作をもう一度行うことをお薦めする。

分子置換法の操作 1) ~ 5)を実施した後、CCP4i メインパネルから Refinement モジュールを選択し、Run Refmac5 を選択し、rigid body refinement に続いて、restraint refinement を実施して、構造精密化の段階へと進む。

なお、CCP4 Program Suite 内の分子置換法のソフトとして、Molrep 以外にも Phaser (4), MrBUMP (5), Balbes (6) などが準備されていて、それぞれ特徴があるよう である。詳細は Web などで確認して頂き、Molrep でうまく解が見つからない場合には それらのソフトの使用も検討すると良い。

以上、CCP4 Program Suite の Molrep の手順を追いながら分子置換法の基礎につい て述べてきたが、最近では、CCP4 に加えて Phenix (7) (<u>http://www.phenix-online.org/</u>)の利用者も増えていると思われる。Phenix の場合、X 線回折強度データの 評価を phenix.xtriage、分子置換法による位相決定を Phaser-MR、構造精密化を phenix.refine で行う。Phenix には豊富な Tutorial Video があるので、例えば、 Checking data quality with phenix.xtriage

ecking data quality with phenix.xtriage

(https://www.youtube.com/watch?v=IWBDjfVN9D4) 、

How to run Phaser-MR (<u>https://www.youtube.com/watch?v=IJ7t8dgfFjk</u>) 、 Phenix tutorial: phenix.refine using default values (GUI)

(<u>https://www.youtube.com/watch?v=6m9dX8gA5o0</u>) などの動画を参考にしていただくと良いと思う。

工夫とコツ

(補足1) モデル分子の変更

筆者は、図 10 に示すような D2 対称(分 子の中心に 222 対称)を有する 4 量体の構 造解析を分子置換法で行った際、単量体 (水色の分子)、2 量体(赤枠、青枠で示 した 2 通り)、および 4 量体全体をそれぞ れモデル分子として用いて、分子置換法を 試した。このうち正しい解を得たものは、 唯一、赤枠で示した 2 量体のみであった。 このように、多量体を形成しているタンパ ク質については、複数のモデル分子の可能 性を試すことをお薦めする。

また、解が得られない場合には、ポリア ラニンモデルをモデル分子として用いても よい。具体的には、Molrep パネルの Model から Model modification: Polyalanine model を選択すれば良い。ま た、目的タンパク質とモデル分子とのアミ



ノ酸アライメントも確認し、例えば、保存性の低い領域をテキストエディタで削除したも のをモデル分子として使用するなど工夫してみると良い。全て自動処理に任せるのではな く、色々と考えながらモデル分子を選択すると解が得られる場合がある。

SUBSET OF I RESOLUTION LIMIT	NTENSITY D NUMBER OBSERVED	ATA WITH OF REFL UNIQUE	H SIGNAL/NO ECTIONS POSSIBLE	ISE >= <mark>-3.0</mark> A COMPLETENESS OF DATA	S FUNCTION R-FACTOR observed	OF RESOLU R-FACTOR expected	TION COMPARED	I/SIGMA	R-meas 🤇	00(1/2)	Anomal Corr	SigÅno	Nano
4.92 3.49 2.85 2.47 2.21 2.02 1.87 1.75 1.65 total	98920 171766 213517 273739 317396 353114 327769 210131 74057 2040409	12174 21249 27215 32019 36238 39900 43328 45617 34214 291954	12209 21255 27228 32020 36245 39911 43346 46461 49418 308093	99.7% 100.0% 100.0% 100.0% 100.0% 100.0% 98.2% 69.2% 94.8%	5.33 8.4% 25.2% 72.6% 153.2% 331.7% 675.0% 1233.7% 1758.9% 28.1%	4.9% 7.0% 23.1% 72.6% 159.2% 352.0% 735.5% 1387.9% 2011.6% 28.3%	98906 171762 213504 273739 317395 353113 327749 208598 59422 2024188	35.41 25.28 8.53 3.08 1.40 0.58 0.17 0.01 -0.00 4.73	5.7% 8.9% 27.0% 77.3% 162.7% 352.1% 724.4% 1393.0% 2158.0% 30.2%	99.9* 99.6* 97.2* 62.2* 24.8* 5.3 1.6 1.3 99.5*	-6 -7 -4 -2 -1 0 0 0 1 -2	0.871 0.848 0.761 0.677 0.606 0.542 0.513 0.514 0.674	10308 19201 25161 29976 34256 38012 40916 29910 4965 232705
図 11:XI 目安と思れ 囲の R-FA 性や空間群	OS の CC かれるのて CTOR o が間違っ	RREC 、この bserve	T.LP 内に 図のデー ed(R-me	書かれてい 夕では最高 erge)が 10 あろ(その	る統計テ 分解能は 0%を超え 占に関し	ーブルの 2.2 Å 租 えるような	例。赤ラ ≧度を選 \$高い値 のデー ^な	丸で示し 択すべき を示す場	た CC(1, きである。 易合には、 きな問題	/2) は また、 データ	60%り 青丸で そのも	く上の分 で示した っのが悪 あろ)	解能が 低角範 い可能

(補足 2) X 線回折強度データの質の確認

アミノ酸の同一性が高い(例えば 50%以上)にもかかわらず、有意な解が得られない 場合は、X線回折強度データそのものの質をチェックすることも重要である。データ全体 のCC(1/2)やR-merge(R-FACTOR observed)をチェックすると良い(図11)。X 線回折強度データ処理ソフト XDS で処理した場合には、CORRECT.LP あるいは XSCALE.LP というファイルをメモ帳などのテキストエディタで開き、最終行に移動して から「-3.0」を「上へ」で検索すれば統計データが確認できる。 また、X 線回折強度データ処理の段階では、空間群が決定できない場合も多い。そうした場合には、無理に一つの空間群に決定する必要はない。可能性のある全ての空間群(厳密には Laue 対称)を用いて XDS でそれぞれ処理し、それぞれの X 線回折強度データを用いて分子置換を行うと良い。

(補足3)分解能の変更

分子置換法を行うにあたり、分解能を変更すると良い。例えば、2.0 Å 分解能の X 線回 折強度データが得られている場合、筆者は、2.0 Å, 3.0 Å, 4.0 Å, 6.0 Å のように最高分 解能を変えて分子置換法を実施する場合も多い。また、そうした場合に、ある条件のとき のみ、正しい解が得られることもあるので試して頂きたい。具体的には、Molrep パネル の Experimental Data 内の High resolution cut-off にチェックを入れて、最高分解能の 値を入力すれば良い。ただし、分子置換法で解を得た後の Rigid body refinement や構 造精密化では、全てのデータ(この場合は 2.0 Å 分解能)を用いる必要がある。

(補足4)思わぬ落とし穴にはまった例

アミノ酸同一性が高いモデル分子があるにもかかわらず、どうしても解が見つからない と学生さんが言ってきたので、一緒に調べてみたところ、少し難しいケースであったので 参考までにここに示す。

まず、学生さんが持ってきた X 線回折強度データは X 線回折強度データ処理ソフト XDS を使って空間群 P222 で処理されたものであった。これは、回折 X 線強度の対称で ある Laue 対称を mmm として処理したデータであることを意味し、P222 は仮の空間群 である。また厄介なことに補足1で述べたように可能性のあるモデル分子が複数あった。 そこで、可能性のあるモデル分子を全て準備して、それぞれのモデル分子につき Molrep パネルの Search Options から SG to use: Laue class を選択することで、モデル分子と 空間群の候補をすべて試してみた。その結果、図 12 左に示したように、Molrep のログ ファイルにて高い Score 値と Contrast 値を示すモデル分子と空間群の組み合わせが見つ かった(ここでは P22,2,という空間群が候補)。しかし、P22,2,という空間群は厳密に は存在しない(詳細は、空間群の詳細が記載されている書籍等をご参照いただきたい)。 そこで、CCP4iの Reindex を用いて空間群を P2,2,2 に変換した後、その MTZ file を用 いてもう一度 Molrep を流して、ようやく正しい解を得ることができた(図 12 右)。こ うした少し難しいケースもあるので、構造解析を複数回経験したことのある研究者に相談 してみると良いこともある。

(補足5)結晶中の分子のパッキング

分子置換法では、一般に、非対称単位に当てはめる分子の数が多いほど、正しい解を得 ることは困難となり、ソフトで全ての分子を見つけることができない場合もある。そうし た場合には、一部をマニュアルで当てはめることを検討すると良い。例えば、図 13 のよ うに、ソフトで8分子中7分子が見つかったが、8分子目の解がどうしても見つからなか ったケースがあった。ここで、分子のパッキングを確認したところ、もう1つの分子が存 在すると予想される空間が存在した。このとき、タンパク質は2量体構造をとっていたた めに、青色の分子と2量体を形成するようにもう1つの分子をマニュアルで当てはめて、 構造解析できたケースもあった。

蛋白質科学会アーカイブ, 14, e103 (2021)

corrF = 0.4021 Space group : P 21 21 2 TF/sig = 2.84 Final CC = 0.4021 No: 18 Sett: 1 Cell: 133.040 171.020 112.430 90.00 90.00 90.00 Packing_Coef = 1.0000 -Contrast 8,61 --- Rotation function ------ Space Group Checking. ---: 43.34 Radius of gyration Radius of gyracion WARNING: Radius of integration > Radius of integration : 56.22 Resolution : 47.57 -----+ 56.22 Nsg Score Cntr | sa Resolution : 47.57 2.57 WARNING: For this radius integration program uses data included 0.354 0.000 | 1 16 P 2 2 2 between 47.0 mm --- rfcoef for model ------ rfcoef for Fobs ---NCS (from Self rotation Function): 2 NCS_model (from Model Self rotation Function): 3 Program will use NCS_model =: 3 2 17 P 2 2 21 3 1017 P 21 2 2 0.338 0.000 | 0.351 0.000 | 4 2017 P 2 21 2 0.346 0.000 | 0.355 0.000 | 0.367 0.000 | 0.473 40.053 | 0.402 8.613 | P 21 21 2 18 6 2018 P 21 2 21 73018 P 2 21 21 19 P 21 21 21 ----+ --- Peaks of Rotation Function --------No in symlib = 3018 dat SYMINFO file set to C:\CCP4-7\7.0\lib\data\syminfo.lib +-----INFO: space group was changed to No = 18 Sett = theta phi chi Rf/sigma | 1 110.29 -113.37 97.85 2 134.84 -65.16 141.28 24.92 INFO: New_Fobs: molrep_fobs_new.dat SYMINFO file set to C:\CCP4-7\7.0\lib\data\syminfo.lib Time: 2lh 6m 46s Elapsed: 0h 37m 36s 24.85 46.05 162.89 158.33 44.52 63.64 143.68 3 24.62 24.46 112.39 153.94 99.77 6.52 5.87 66.32 -156.07 101.09 6 137.13 -46.64 162.54 5.15 5.12 122.86 -96.85 108.30 8 63.59 -153.55 104.28 9 4.98 136.18 -148.76 177.05 10 4.97 49.99 81.20 124.79 11 4.96 136.57 -147.81 178.25 4.93 12 13 136.62 -34.74 178.27 4.59 120.28 -100.48 106.22 14 4.49 42.79 44.63 166.00 15 4.47 16 136.98 -51.43 157.14 4.37 144.31 170.21 96.60 55.60 94.17 112.28 17 4.36 18 4.35 19.26 134.60 172.38 19 4.33 20 133.44 -74.52 133.37 4.28 21 129.08 -83.97 120.15 4.26 | 22 46.91 157.71 151.05 4.24 図 12: Molrep のログファイルの例。左図において、高い Score 値と Contrast 値を示す空間群(Space group、 ここでは sg と示している)は P22,2,であったが(赤丸)、P22,2,という空間群は厳密には存在しないため、これ を P2,2,2 に変更する必要があった。そこで、CCP4i の Reindex を用いて空間群を P2,2,2 に変更した後、その MTZ file を用いてもう一度 Molrep を流すことで正しい解が得られた(青丸の 4 つ解が正しい解であった)。



図 13:結晶中の分子のパッキングの例。8分子中7分子が分子置換法で見つかったが、分子のパッキングを確認した ところ、赤矢印に示した空間が存在し、もう1つの分子が存在すると予想された。そこで、水色の分子と2量体を形 成するようにもう1つの分子をマニュアルで当てはめて構造解析することができた。

対献

- (1) Winn, MD, et al., Acta Cryst. D67, 235–242 (2011)
- (2) Vagin, A, et al., J. Appl. Cryst. 30, 1022–1025 (1997)
- (3) Matthews, BW, J. Mol. Biol. 33, 491-497 (1968)
- (4) McCoy, J, et al., J. Appl. Cryst. 40, 658–764 (2007)
- (5) Keegan, RM and Winn, MD, Acta Cryst. D63, 447-457 (2007)
- (6) Long, F, et al., Acta Cryst. D64, 125-132 (2008)
- (7) Liebschner, D, et al., Acta Cryst. D75, 861-877 (2019)

