

大学共同利用機関法人 高エネルギー加速器研究機構 物質構造科学研究所

構造生物学研究センター クライオ電顕 特集

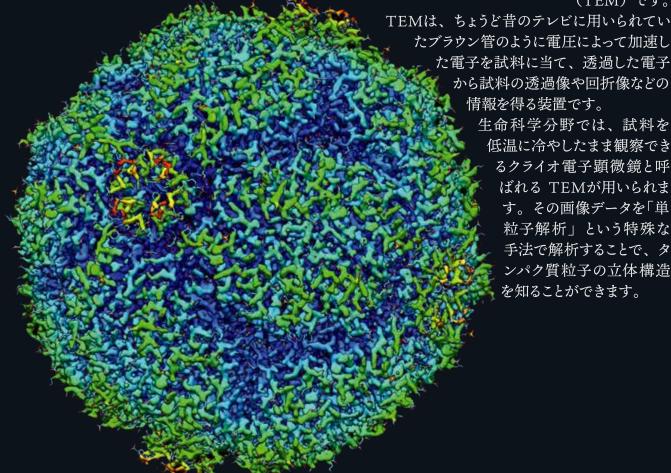
## 物構研第5の量子ビーム、電子線 クライオ電子顕微鏡を用いたタンパク質の単粒子解析とは?

物構研は、放射光・中性子・ミュオン・低速陽電子という4つの量子ビームの 実験施設を運用し、それを活用した研究を行っています。

近年、生命科学分野での必要性から、新しい量子ビーム「電子線」が加わり ました。構造生物学研究センター(SBRC) に導入された透過型電子顕微鏡

(TEM) です。

たブラウン管のように電圧によって加速し た電子を試料に当て、透過した電子 から試料の透過像や回折像などの 情報を得る装置です。 生命科学分野では、試料を 低温に冷やしたまま観察でき るクライオ電子顕微鏡と呼 ばれる TEMが用いられま す。その画像データを「単 粒子解析」という特殊な 手法で解析することで、タ ンパク質粒子の立体構造 を知ることができます。



SBRC での単粒子解析による酸化還元酵素 「硫黄オキシゲナーゼ / レダクターゼ」の構造図 直径はおよそ154 Å

協力: 東京大学大学院 農学生命科学研究科

伏信 進矢 教授

赤や黄色に見える部分は突起しているためチムニー (煙突)と呼ばれ、硫黄の出入り口ではないかと考え られている。色は分解能の目安で、青いほど構造の 揺れが小さいために分解能が高く、赤いほど構造の 揺れが大きい可能性が高いことを示す。東京大学と の共同研究によるこのタンパク質の解析では、全体 の平均で2.05 Åという高い分解能が得られた。

# タンパク質の単粒子解析って どうやるの?

# 二次元の画像データから 三次元の情報が得られるのはなぜ?



### タンパク質のかたちを知りたい!

物構研の構造生物学研究センター(SBRC)では、生命の謎や病気のしくみなどの解明のため、生命体の主要な構成要素であるタンパク質を調べています。タンパク質を構成するアミノ酸は基本 20 種類ですが、その組み合わせによって多様なタンパク質が存在し、それぞれ複雑な立体構造を持ちます。タンパク質の立体構造が分かると、そのタンパク質がどう動くのかが分かり、どんな働きをするのかを知ることができます。だから、構造生物学者たちはタンパク質のかたちを知りたいと考えるのです。

タンパク質の立体構造を知る方法として、同じタンパク質が規則正しく並んだ状態(結晶)にして放射光(X線)を当て、結晶による反射が強めあったり打ち消しあったりする情報(回折像)を検出し解析する方法があります。フォトンファクトリー(PF)では、このX線結晶構造解析によって数々のタンパク質の構造が調べられています。

しかし、タンパク質を結晶にするのは難しく、試行錯誤を重ねてもなかなか結晶にならない例も少なくありません。そこで開発されたのが、結晶化の必要がない「クライオ電子顕微鏡による単粒子解析」で、2017年、研究開発に貢献した3名にノーベル化学賞が授与されました。



### クライオ電子顕微鏡って?

透過型電子顕微鏡(TEM)は、数百 kV の電圧で加速した電子を試料に当て、試料を透かして見たときの濃淡のある影絵のような像(透過像)を得る顕微鏡です。虹色に分かれる可視光で撮った写真ではないので、TEM 写真には色の情報はありません。しかしその分解能は、条件が合えば原子の粒が見

分解能:対象を識別できる能力のこと。例えば、人間の肉眼の分解能は 0.1 mm程度。TEM の分解能は数Å程度。

えるほど高いのです。

量子力学の創成期、ド・ブロイが提唱した物質波の考えによれば、電子も波と考えることができます。波としての電子の波長を計算してみると、加速電圧が 200 kV のときおよそ  $2.7 \times 10^{-12}$  m = 0.027 Åで、可視光の波長  $(3 \times 10^{-7} \sim 8 \times 10^{-7}$  m) に比べてはるかに短いことが分かります。波長が短い波ほど小さいものを識別できるので、Åで測るような細かい構造を知るために電子線は有効なのです。参考までに、PF でよく使われる放射光の波長は、 $10^{-10} \sim 10^{-7}$  m  $(1 \text{ Å} \sim 100 \text{ nm})$  程度です。

ただしX線と違って、電子線は透過力が弱いので、 試料はとても薄くなければならないだけでなく、電子線の通り道も空気中の分子をできるだけ除いた状態(真空)にする必要があります。そのような環境にタンパク質を入れたらどうなるでしょう。真空であることで元の形を保っていられず、干からびてしまうでしょう。また電子線が当たることで化学反応が起こり、タンパク質は壊れていきます。それでは知りたかった元のかたちの観察はできません。そこでタンパク質を極低温の薄い氷に閉じ込めて観察する方法が考案されました。この氷は、水分子が規則正しく並んだ結晶ではなく、不均一なガラスのような構造になっています。このガラス状の氷によってタンパク質は真空から守られ、また低温であるほど化学反応が進みにくくダメージが少ないと言われています。

試料とその周辺を液体窒素で -200℃くらいの低



直径3mmのTEM 用グリッドと1mm 刻みの定規 網状の金属板に カーボン膜が張ら れています。



ま料グリッドを作成しているところ タンパク質の粒が入ったピペットを専用の装置に 差し込み、タンパク質をグリッドに載せ瞬時に凍 らせるための作業をしています。緊張の一瞬です。



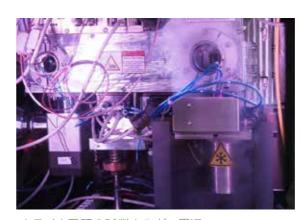
温に冷やしながら観察できる TEM を、クライオ電子顕微鏡、略してクライオ電顕と呼びます。 SBRC では 2018 年の初めに導入しました。



#### 単粒子解析って?

単粒子解析は、その名の通り、ひとつひとつばらばらの粒になったタンパク質の TEM 写真から、その立体構造を知るための方法です。同じタンパク質のたくさんの粒の TEM 写真を撮り、その二次元の情報から三次元像を復元します。

多数の二次元像から三次元構造を再構築するという例では、病院などで行う X 線 CT があります。 X 線源あるいは撮影対象を回転させ、少しずつ角度を変えて、1つの立体物のあらゆる角度からの X 線写真(投影像)を撮ります。二次元の情報は、コンピュータで三次元に再構築されますが、複数の投影



クライオ電顕の試料ホルダー周辺 黄色の三角シールが貼られている筒に試料グリッドが入っています。液体窒素によって冷やされた 空気中の水蒸気が見えています。

情報の辻褄が合うところの信号が強くなり、合わないところは弱くなって、三次元構造が得られていきます。この場合はあらかじめ投影した角度が分かっているのでその情報を再構築に使うこともできます。

さて、タンパク質の TEM 像の場合はどうでしょう。 タンパク質の粒は薄い氷に閉じ込められて固定され ています。しかも、タンパク質の粒の向きは初めか らバラバラ……。!

ここで発想の転換です。1つの粒をあらゆる角度 から撮影することができない代わりに、初めからあ らゆる方向を向いている(と思われる)たくさんの 同じタンパク質の粒を撮影すれば、得られる写真は 同じではないでしょうか?

ただ、違うのは、どの角度から撮った写真なのかが分からないこと。つながりの分からないバラバラの写真だけが多数得られるということです。

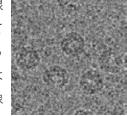


自動撮影をしながらデータ処理をしているところ 奥の部屋にクライオ電顕があり数分に1枚のペースで 自動撮影をしています。手前の大きなモニターがその 状況を刻々と映しています。

ちょっとパズルのような話になってきました。バラ バラの写真の山から推理して元のストーリーに組み 立てる、そんな探偵のような作業が必要なのです。

この探偵業の難しいところはもうひとつあって、 元の写真がくっきりしていないこと。タンパク質のダ

メージを防ぐため弱い電子線 を当てているからです。例えて 言うなら、ろうそくの灯りで撮っ た写真のようなもの。解析には ノイズの多いざらざらした画像 を使わざるを得ないのです。



単粒子解析に用いら れる TEM 写真の例



#### 推理の方法は?

ここからは単粒子解析を行う探偵のような研究 者、自称「画像処理屋」の守屋 俊夫 特任准教授 の手法をご紹介しましょう。



#### 1.データの拾い上げとごみの選別→「原画像」

まず、撮影した TEM 写真から粒子の部分を切 り出します。粒子の大きさにもよりますが、1枚の TEM 写真には 100 ~ 200 個ほどのタンパク質の粒 が写っています。粒同士が距離を保って散らばるよ うに試料を作成しているので、一定の余白をとって 1粒ずつ1枚の画像に切り出すことができます。こ こで活躍するのが粒子ピッカーと呼ばれるソフトウェ アで、ディープラーニング(深層学習)を使って自 動で粒子を認識し拾い上げを行います。

しかしこの中には、氷の結晶やカーボン膜の縁、 壊れたタンパク質のかけらなどが写り込んでいる「ご み画像 | も入っています。ピッカーが拾い上げた後、 基準から外れているものは自動で捨てられていきま すが、その基準を与えるのは人です。基準を厳しく しすぎると、大切なタンパク質の粒画像まで捨てて しまいますから、ごみとそうでないものの判別をど う機械に任せるか、これも腕の見せどころ。こうして およそ 2000 枚の TEM 写真から数十万から百万の 「原画像」が得られます。

### 2. ノイズの除去と二次元平均化

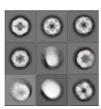
#### →「クラス平均像」

データにノイズはつきもの。このノイズを取り除く には、同じようなデータをたくさん重ねるのが王道 です。ノイズは真の情報ではないので、ランダムな 場所にランダムな強さで現れます。だから、たくさ んのノイズを重ね合わせれば、打ち消し合って弱く なる。これがノイズ除去の考え方です。

重ね合わせるためには、同じ投影像は、原画像 中で同じ角度・同じ位置になければなりません。回 転と位置合わせの作業が必要です。それによって二 次元画像を大まかに合わせて、投影方向ごとに分け ることを「二次元クラス分け」と呼んでいます。

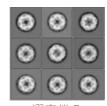
さて、回転・位置合わせと言っても、何か基準が ないと合わせられませんね。こんな投影像になるは ずと基準を与える方法 A もありますが、何もない状 態からクラス分けをさせる方法 B もあります。方法 Bでは、原画像から適当に数枚を選んで「この像に 似たもの集まれ~」と、何度も位置合わせや回転 をかけて、形が似ているものを集めます。基準のか たちに近いものは同じクラス、遠いものは違うクラ

スと原画像を振り分けていくと、似 た画像が集まったクラスが数百でき ます。クラスごとに平均像を計算する と、ノイズが少ない像ができるので、 これを基準にして次の作業に移りま す。



クラス平均像の例

しかし、この時点でもごみが残っ ています。計算は自動で行っても、 どの絵がきれいか、ごみクラスがで きていないか、最終的に選ぶのは人 の目ということになります。この過程 で数百枚のクラス平均像ができます。 クラス平均像の例



選定後の

#### 3. 初期三次元再構築→おおまかな三次元構造

さて、いよいよ三次元の大まかな構造を作り上げ るのですが、それにもいろいろな手法があります。

#### (i) 古典的な方法

まず、このタンパク質はこんな立体構造をしてい るだろうとある程度分かっているとして、適当な三 次元構造を仮定します。その立体構造からシミュ レーションによってあらゆる角度からの投影像を作 ります。それを基準の画像として、原画像と比べ、 類似度を計算していきます。これをプロジェクション マッピングならぬ、「プロジェクションマッチング(投 影像照合法) | と呼びます。

最も類似度が高いものにシミュレーション時の 角度を割り当てるという作業を続けると、全部の原 画像に三次元の角度データを割り振ることができま す。この情報から三次元構造が組み立てられます。

(ii) ab-initio (最初から) at random (手当たり次 第な) 方法

さて上記(i)の方法では、元のかたちを仮定したこ とが結果に強く影響してしまいます。仮定せずにや る方法、それは二次元のところで考えた B の方法(何 もない状態からのクラス分け)の応用です。そうす るとほとんどの場合、得られる三次元構造は球に近 い塊になります。このブロック状の構造に対して投 影像を作成し、原画像とのプロジェクションマッチ ングを行います。もっとも類似度が高いものに投影 角度を割り振り三次元構造を構築します。

#### (iii) ab-initio で三次元構造を複数つくる方法

上記(ii)の方法で、三次元構造を複数(3つくらい) 作るようにプログラムします。原画像がどの三次元 構造の投影像か分類し、それぞれで角度を割り振 ることになります。三次元構造の種類が増えるほど、 その後の計算量は増大します。

上記(iii)の方法で、どの角度から見 ても矛盾のない初期三次元構造が できあがりました。



次は、統計手法を使って三次元角 度と位置合わせを最適化します。

初期三次元構造の例



#### 最適化について 一どこがもっとも類似度が高いのかー

ちょっと想像してみてほしい。あなたはいま濃い 霧に包まれた山の中で山頂を目指している。一歩踏 み出せば登ったのか下ったのかは分かる。だが先は 見通せない。山が単純な円錐形なら頂上に着いたら そこはそのまま最も高い場所だが、現実の山は凸凹 していて、小さな頂上がたくさんある。闇雲に歩い ていても時間がかかるばかり。

いま、標高に例えたのは類似度だ。ノイズによる 小さなピークもたくさんあって、どこが最も類似度 が高いのか分からない。こんなとき、統計的手法が 役に立つ。

まずひとつめは大きなジャンプをしながら進むと いうもの。そんなこと普通の人間にはできないが、 天狗のように小山を飛び越え、飛び降りた先で頂を 目指す。頂上に着いたら高度を確かめ、また大きな ジャンプをする、何度もジャンプをするうちに最も高 い頂上を見つける、という方法だ。歩くよりは速そ うだが、力勝負には違いない。ジャンプの方法にコ ツがありそうだ。

もう少し賢い次の手は、シミュレータアニーリン グと言う。アニーリングとは、もともとは融けた金 属をゆっくり冷やしてきれいな結晶を得る「焼き鈍し 法」のこと。鉄をゆっくり冷やしながら打っていくと 硬くなるという物理現象を活かした鍛鉄方法の名前 が、データ解析の分野でも使われている。初めラン ダムにバンバン大きく飛び、たまに下ってもいいが、 進むごとに下ってもいい確率を減らしていく。だんだ ん絞っていくその確率を焼き鈍し法に倣って「温度因 子」と呼ぶ。まるで実際に熱い金属を扱っているか のようだ。「温度」が高いときは下る回数が多くて、「温 度」が下がってくると下る回数が減り、ゆっくり温度 を下げると確実に収束し、いいところに落ち着く。

このような最適化アルゴリズムは画像処理に限ら す普遍的なものである。分野や研究の目的に合わせ て、例えば処理の時間がかかってもより最適な解を 得られる方法をとったり、得られる解の精度が悪く なる確率が高くなっても速い方法を使ったり、どのア ルゴリズムを使うかは両者のバランスで決まる。ど うやって霧を晴らすか、は解析者次第だ。

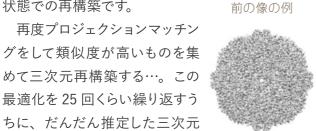
#### 4. 三次元クラス分け→原画像の分類

できあがった複数の三次元構造と、原画像とのプ ロジェクションマッチングを行ないます。プロジェク ションマッチングを繰り返し、「この原画像は、どの 立体構造のどの角度との類似度が最も高いかしを探 して、原画像を立体構造ごとに分類していきます。

#### 5. 三次元精密化→詳細な三次元構造

立体構造ごとに分けた原画像のグループの中で、 より精密な構造を再構築していきます。3.では、同

じタンパク質でも構造が違うもの が混ざっている(かもしれない) 状態で三次元再構築をしました が、今度は同じ構造を持つと思 われる原画像だけを抜き出した 状態での再構築です。



めて三次元再構築する…。この 最適化を25回くらい繰り返すう ちに、だんだん推定した三次元 角度が正確になっていき三次元 像がきれいになってきます。いい

三次元精密化 後の像の例

データセットの場合は、数日の計算で3Åくらいの 分解能で構造を解析できます。

#### 単粒子解析で複数の構造が見えたら

初期三次元再構築や三次元クラス分けで、構造 が複数あるとするのはどうしてでしょう。試料の中 に別のタンパク質やごみが混ざっている場合の受け 皿として? いえ、それだけではありません。

タンパク質は本来、複数の異なる構造を持つもの です。タンパク質は生命体の中で様々な働きをしま すが、文字通り動いているので、その構造も変化す るのです。もしクライオ電顕による単粒子解析で同 じタンパク質の複数の構造が捉えられたとしたら、 それはタンパク質の粒が氷に閉じ込められる瞬間の 構造の違いを捉えたということです。

タンパク質が、例えば $\alpha$ という構造から $\beta$ という 構造に姿を変えるとします。αからβへの過渡期に はαともβとも違う構造があるはずですが、タンパク 質は安定した位置まで短時間に一気に動く性質があ るので、途中の状態で固定される可能性はとても小 さいと考えられます。また、タンパク質の外側には ループと呼ばれる柔らかい構造がありますが、ルー プは動きやすいことが多いため平均化によって相殺 されて見えないことがよくあります。

1つのタンパク質結晶から得た結晶構造解析デー タでは複数の構造を捉えることはできないので、複 数の構造が見えることは単粒子解析の利点のひとつ とされています。

#### 単粒子解析の分解能

単粒子解析の分解能は、通常の分解能の定義と は違い、FSC (Fourier Shell Correlation、フーリ エシェル相関)分解能と呼ばれ、「特定のデータセッ トを2つに分けて、それぞれ別々に処理してもその 分解能(周波数の逆数)の情報までは、いつも一 致して再構築できる」という指標です。

また、単粒子解析では、分解能は構造の局所ご とに異なります。表紙のタンパク質の構造図は、分 解能によって色分けをしたもので、基本的に、よく 動く場所ほど分解能が低く、構造が安定している場 所は高く表示されています。

単粒子解析の世界最高分解能は1.15 Åです。 SBRC では平均 2.5 ~ 4.0 Å程度、最高 2.05 Åです。

SBRC のクライオ電顕による単粒子解析は共同利 用の軌道に乗ったばかり。大学や企業などから新た なユーザーが続々と訪れています。経験と知見を積 み重ね、分解能向上はこれから大いに期待できます。



守屋氏が主宰する単粒子解析の研究会のようす 2020 年 10 月 KEK つくばキャンパス COI 棟にて

Q. クライオ電顕で撮影をしたら全てのデータを単粒子解析するの? A. いま SBRC ではスクリーニングで 500 枚程度撮り、そこでいい感触が得られれば本撮影を行います。 本撮影では 2000 枚程度の TEM 写真を撮り、それが単粒子解析の対象となります。

画像処理屋が KEK で単粒子解析をやるということ

解析ってコンピュータがやるんでしょ、と思う人は 少なくないだろう。

しかし守屋氏は「単粒子解析はやる人によって質 が変わる」と言う。「注意して見てデータを扱う人が やらないと新しい知見は生まれないし、いいアルゴ リズムはできない」というわけだ。たくさんのデー 夕に触れ、こういう傾向があるからこういうアルゴ リズムができると発想する、工夫する。それが結果 に直結することがある。

例えば、粒子の切り出しのときに問題となる2つ のパラメータがある。切り出しの枠の大きさと、粒 子範囲 (粒子の中心から半径いくらのところまでを 粒子、その外は背景と指定するか)の大きさだ。切 り出しの枠については大きくした方がいい、でも、 大きくしすぎると回転と位置合わせのときに合わせ にくいし処理量が増えて無駄になる、と言われてい た。一方、粒子範囲についてはこれまで議論はなく、 多くの人が見た目の粒子ぎりぎりのところを境界と していた。

最近、守屋氏はそれらについて最適化を試みた。 電子顕微鏡の理論を考え、粒子範囲を切り出しの枠 ぎりぎりまで大きくしてみたのだ。すると、それだけ で、一見して分かるほど分解能がよくなった。

実は TEM 撮影の過程で目に見える粒子の像の外 にも情報が滲みだしているのだが、その情報を粒子 範囲の設定で切り捨ててしまっていたのだ。ソフト ウェア開発研究者も手法論者も認識していなかった 「切り出しの枠をいくら大きくしても粒子範囲のパラ メータがその効果を消していた」という事実が初め て判明したのだった。

様々な環境に身を置いてきた守屋氏だが、KEKは 自分に合ったいい職場だと思うと語る。その理由は 所属する物構研 SBRC のメンバーの考え方だという。 「大学共同利用機関としてユーザーをサポートして あげたい、というのは大学や研究所ではなかなか見 られない考え方。みんなのためにやると自分の成果 にならないのにもかかわらず、率先して頑張ってい るのがすごいな、と」

守屋氏はソフトウェア開発はみんなのための仕事 になると考えている。そのような形での貢献を認め てくれる研究仲間と共に活躍していきたい、と目を 輝かせた。

KEK は装置も計算機システムも人材も余裕がある とは決して言えないが、工夫次第で分解能が上がる 余地はまだまだある。この記事を準備する間にも最 高分解能が更新されたという知らせがあったように。

#### 守屋 俊夫さんのプロフィール

タンパク質の二次元の電顕像から三次元の構造を再構築 する手法の研究をしている。アルゴリズムを開発するには 全体の流れを把握する必要があり解析実務も担当。

学生時代は神経学や認知神経心理学を専攻し、人工知能 の研究をしていた。卒業後、電子顕微鏡の開発会社で画像 解析のソフトウェア開発をしているときに単粒子解析に出 会う。前職のマックスプランク研究所では解析ソフトウェ アの開発に従事。KEK 着任はクライオ電顕の立ち上げ期間 にあたる 2018 年。 クライオ電顕のユーザー受入れが始まっ てからは、実習型のセミナーなどデータ解析の指導にも力 を入れている。パズルが大好き。



KEK つくばキャンパス COI 棟 クライオ電子顕微鏡前にて



物構研 構造生物学研究センター(SBRC)は、日本国内の研究者がクライオ電顕を利用し て高分解能構造解析を行うための「BINDS プロジェクト クライオ電顕ネットワーク」におい て中心的な役割を果たしています。守屋准教授が講師を務める講習会も開催しています。 https://www2.kek.jp/imss/sbrc/beamline/cryoem.html



# クライオ電子顕微鏡によるタンパク質の単粒子解析成果

クライオ電子顕微鏡 (クライオ電顕) は特に分子量の大きなタンパク質複合体の観察に適しています。 学術機関のみならず企業ユーザーも増え、構造生物学研究センター (SBRC) と共同利用者との共同研究成果も続々と出ています。

創薬において副作用の原因となりやすい hERG の構造図

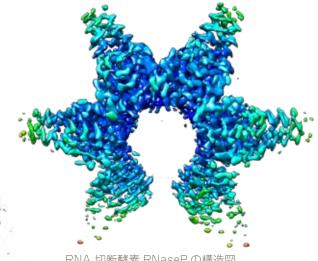
クライオ電顕勉強会

(千葉大学 村田 武士教授・Axcelead Drug Discovery Partners 株式会社・アステラス製薬株式会社・小野薬品工業株式会社・第一三共RDノバーレ株式会社・武田薬品工業株



式会社・武田薬品工業株式会社)との共同研究

https://www.kek.jp/wp-content/uploads/2021/01/PR20210115.pdf

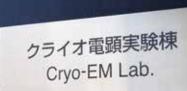


RNA 切断酵素 RNaseP の構造図 九州大学 角田 佳充 教授との共同研究

## クライオ電顕実験棟を新設

2022 年春、KEK つくばキャンパス内にクライオ電顕専用の建物を建設、移転しました。タンパク質の精製のための実験室も集約し、精製したてのタンパク質の観察から解析までを一貫して行えるようになります。これによって、より質の高い解析が可能になります。また、作業スペースが広くなるので不安定な試料を扱いやすくなり、共同利用者にとっても使いやすい環境となりました。









物質構造科学研究所 構造生物学研究センター The Structural Biology Research Center (SBRC) TEL: 029-879-6178 FAX: 029-879-6179 センター長:千田 俊哉(物質構造科学研究所 研究主幹・教授) 生物科学・構造生物学・X 線結晶構造解析

e-mail:toshiya.senda@kek.jp



編集:物構研 広報室 (瀬戸 秀紀、足立 伸一、安達 成彦、阿部 仁、市村 規子、岩野 薫、宇佐美 徳子、 大島 寛子、神田 聡太郎、瀬谷 智洋、中村 惇平、深堀 協子、山田 悟史)

<sup>発行:</sup> 高エネルギー加速器研究機構 物質構造科学研究所

〒305-0801 茨城県つくば市大穂1-1 TEL: 029-864-5602 https://www2.kek.jp/imss/ e-mail: imss-pr@ml.post.kek.jp (物構研 広報室) 禁無断転載 @All rights reserved by High Energy Accelerator Research Organization (KEK)

