

## Discovery Studio HT-XPIPE : ハイスループットX線結晶解析によるリード化合物最適化のためのアプリケーション

Igor Mochalkin<sup>1</sup> and Tadahiro Ohmura<sup>2</sup>

<sup>1</sup> Accelrys Inc. 9685 Scranton Road, San Diego, CA 92121-3752, USA

<sup>2</sup> Accelrys-KK. Nakarín-Auto Bldg. 5F, 2-8-4 Shinkawa Chuo-Ku, Tokyo 104-0033,

### 序論

X線結晶解析はリガンド分子と相互作用する生体高分子の生物学的機能に重要な見解を提供することで現在の創薬において中心的なツールになっています。過去数年で、生体高分子のX線結晶解析の方法(タンパク質精製、結晶化、X線データ収集、構造決定など)には多くの技術的進歩がありました。これらの進歩によって、多くのタンパク質の結晶構造や複合体の構造が正確にすばやく決定できるようになります。この点で、構造遺伝学、プロテオミクスの分野において、タンパク質データバンク(PDB: Rutgers University, USA)のタンパク質構造登録数の指数的増加を導きます(Berman, H.M., et al 2000)。2003年5月までに、約20800個の構造がPDBに登録され、約3分の1が野生型タンパク質の構造で、残りが同じものか、変異体、タンパク質-リガンド複合体です。

構造遺伝学やプロテオミクスの分野における技術進歩は結晶化やX線解析に最適な質的、量的なタンパク質のハイスループットな発現、精製をも含んでいます(Lesley, S. A. 2001)。次世代ロボット化結晶化系を使用したハイスループット結晶化法は結晶化に成功した沈殿剤、試薬、緩衝液とpHの知識を活用し、結晶化への試みを多次元的に拡張できます。この自動化された結晶化は、タンパク質量をナノスケールレベル(ナノリットル)に減らすことができ、1日あたり最高10<sup>5</sup>回のトライアルを行い、結晶をより早く成長させることができます(Stevens, R.C., 2000, Mueller U., et al. 2001)。得られた結晶は、手動で、あるいは自動化されたサンプル交換システムを使用して、X線装置あるいは放射光のX線ビームラインで解析されます。

タンパク質の発現、精製、結晶化、X線データ収集の自動化における劇的な進歩に対応して、ハイスループットX線結晶解析のための新しいソフトウェアの開発は非常に重要になっています。複数のソフト開発ベンチャーの結果として、最新のコンピュータ処理方法は近年劇的に進歩し、現在、構造ベースの創薬のハイスループット化は現実的になってきました。薬学や生物工学分野の研究所のハイスループットX線結晶解析への大きくなる期待に対して、ソフトウェアと情報科学支援の2つの試みで提供されることになるでしょう。すなわち、第1にデータ収集からモデル構築、精密化を通して分析までのすべての構造決定ステージの自動化プロトコルの開発であり、第2に重要な化学的データ、生物学的データおよび実験データの分析とのリンクを提供し、X線結晶解析プロジェクトの進行を追跡する情報科学環境を提供することです。

ここで、我々は、単一のコマンドによる複数のタンパク質-リガンド系の自動結晶構造決定および精密化の方法のプロトコル開発とソフトウェアについて概説します。この手法は、オラクルデータベースを使用することで様々なデータ（実験プロトコル、入力データ、出力データ、途中結果）を他の研究者と共有でき Discovery Studio™ (DS) 環境において実行されます。

## プロトコル開発

構造決定の自動化プロトコルの開発は、方策決定機能を備えた自由度の高いパイプラインに統合できる強力なアルゴリズムの有効性に大きく依存します（図1）。CNX (Brunger, A.T., 1992、Badger, J. et al 1999)、および QUANTA X-LIGAND (Oldfield, T., 2001)のテクノロジーが、分子置換法、初期構造精密化（タンパク質構造のみ）、リガンド配置、タンパク質-リガンド複合体精密化、水分子配置、最終構造精密化からなるタンパク質-リガンド複合体の結晶構造決定の自動化プロトコルを開発するために使用されました。これらのプロトコルは DS HT-XPIPE で提供されます。

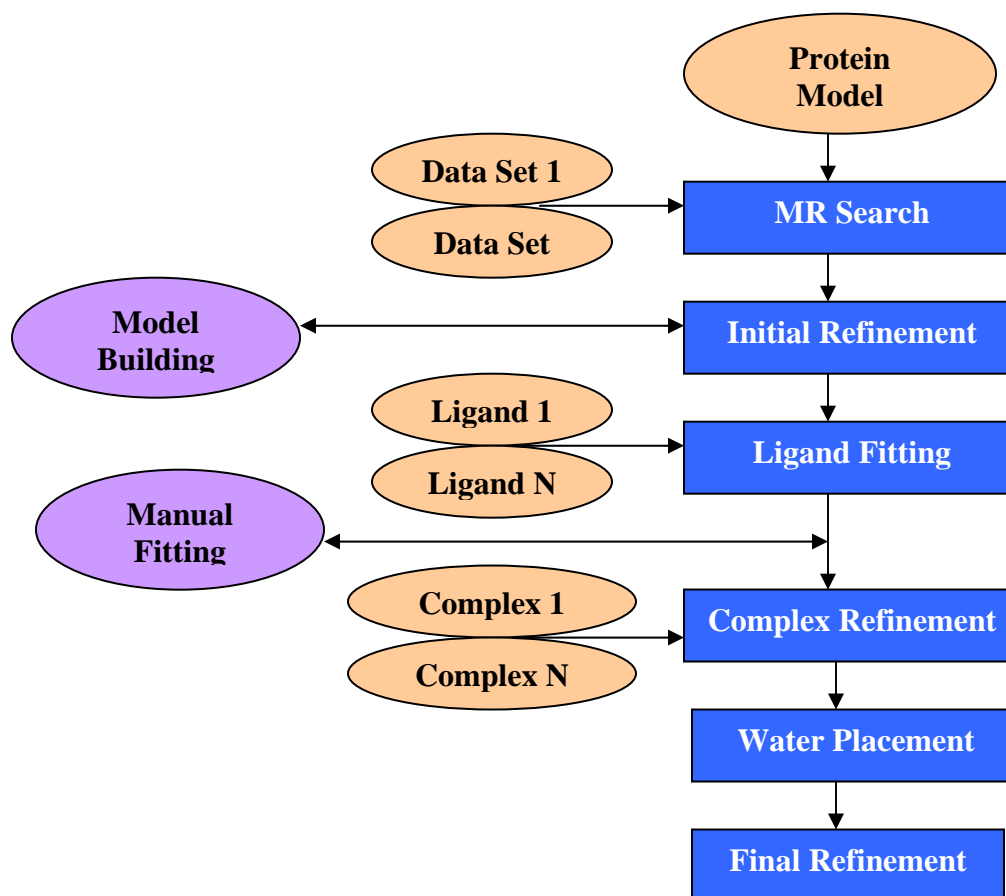


図1：タンパク質-リガンド複合体の構造決定のプロセスは、分子置換法、初期構造精密化、リガンド配置、複合体精密化、水分子配置、および最終構造精密化から成っています。初期タンパク質モデル構造、複数のX線データセット、リ

ガンド構造あるいは複合体構造を入力データとします。プロトコルや入力項目などは全てのステップでユーザー定義かオプション設定になります。手動のモデル構築とリガンドフィッティングは自動化されたパイプラインのリガンドフィッティングステップの前後で実行できます。

## モデル選択および分子置換検索

計算された構造因子と観察された構造因子との間の最も良い相関係数を得るために同一あるいは相同分子のタンパク質構造が単位格子中に配置される検索分子として利用されるときに、分子置換法による位相の決定が行われます。DS HT-XPIPEは、分子置換が必要かどうかを決定するために、検索モデル構造とタンパク質-リガンド複合体の構造因子から得られる空間群と単位格子パラメータを比較します。分子置換を実行することをユーザーが選択しない限り、検索モデル構造から2%以内の誤差の単位格子パラメータを持つ同じ空間群で結晶化したタンパク質-リガンド複合体に対してソフトウェアは分子置換検索を省略し、剛体精密化ステップに進みます。構造決定プロセスを自動化するために、DS HT-XPIPEの分子置換法は、単位格子中(非結晶学的対称をもっているかもしれませんが)へ検索モデル構造の複数コピーを自動配置します。非対称単位(ASU)毎の分子数は、検索モデル構造の座標および結晶学的パラメーターを含んだ反射データの読み込み中に見積もられます。

分子置換法用の自動化スクリプトは、修正されたCNXコマンド (Brunger, A.T., 1992)で構築され、変換マトリックスの回転・並進成分を見つけるためにパターン相関法を使用します。ターゲット構造が非結晶学的対称を持っている場合、ASU中の全分子位置は単一の回転項探索で決定されます。一方、並進操作はいくつかの並進関数探索によって連続して決定されます。さらに、異なる検索モデル構造(モデル構造そのもの[complete model]、ポリアラニン構造[poly-alanine backbone]、あるいは炭素原子鎖のみ[C-atoms only])を選択して回転・並進検索を行うことも可能です。またコンフォメーションのフレキシブルな領域(高い温度因子を持つ残基)を削除したターンケートモデルの自動作成も可能です。プログラムは、単位格子内にモデル構造をすべて敷き詰めた後に、1つの任意に選択された分子に対して分子間の埋もれた表面が最大限になるように非対称単位分子を残すための結晶学的変換を見つけます。

## 初期構造精密化(タンパク質構造のみ)

ターゲットタンパク質のモデル構造を単位格子中へ正確に配置した後に、剛体精密化、構造精密化および温度因子精密化計算が実行されます。初期構造精密化は、実験データ(観察された構造因子)と計算データ(モデル構造から計算された構造因子)の一致を最適化し、リガンド分子配置に適した質の良い電子密度マップを得るために実行されます。付加的な拘束条件を加えることも可能です。計算

された構造因子と観察された構造因子間の一致度は、収束値（R 因子）の計算により示されます。

## リガンド配置

DS HT-XPIPE はモンテカルロリアルスペース回転角精密化法を使用して電子密度へのリガンド分子の自動配置が可能です。これには QUANTA X-LIGAND (Oldfield, T., 2001) と同じ手法を採用し、以下のオプションを付加しています；(1)電子密度同定と初期リガンド配置に対する自動シグマ閾値調節、(2)活性部位残基同定、(3)タンパク質原子との重なりを持つ電子密度を除かないためのオプションとリガンド分子の初期コンフォメーションをランダムにしないためのオプション、(4)Z-weight (Džakula, Ž., to be published)。

自動リガンドフィッティングは、リガンド部位検索、回転可能結合解析、コンフォメーション検索、リアルスペース回転角精密化の4つの過程から成り立っています。リガンドフィッティングステップは、まずタンパク質表面周辺6 Å以内の領域に対称分子の生成を要求します。次に、ユーザー定義された閾値以上のタンパク質原子の入っていない単位格子中の全電子密度を検索した後、リガンド結合可能な部位のリストを生成します。DS HT-XPIPE で新規にセットされたオプションを使用することで、タンパク質あるいはそれらの対称分子の原子とオーバーラップする電子密度を除去できます。リガンド結合可能部位の全電子密度リスト後、体積順に並べられたリガンド配置部位の独立な電子密度リストを作成します。閾値調整オプションを使用すると、プログラムは選択部位とリガンドの体積を適合させるためにシグマ閾値を増減させます。タンパク質モデルが完全に電子密度にフィッティングしている場合に、満たされていない電子密度のもっとも大きい部位はしばしばリガンド結合部位とみなされ、その一方、残りの部位は通常水分子としてみなされます。

コンフォメーション探索中に、プログラムは、自動的に分子構造上における効果によって、それぞれの回転可能な結合に優劣をつけます。構造変化に大きく影響を及ぼす結合については、詳細な検索を行うために回転角度は1度づつに設定されます。構造変化にほとんど影響のない結合は回転角度が180度づつに設定されます。その他の結合は、構造に与える効果の程度にあわせて検索範囲が設定されます。そして、可能なコンフォメーションは約2000個/秒の割合で電子密度に適合させることができます。検索の結果、最も良いコンフォメーションのうち20個が保存され、コンフォメーション探索に続いて最小二乗法によって電子密度にリガンド分子をフィッティングさせるためのリアルスペース精密化を実行できます。

## タンパク質-リガンド複合体精密化と水分子配置

精密化されたタンパク質-リガンド複合体の精密化は初期構造精密化と同様の方法で実行されます。精密化された構造は水分子配置を援助するための差電子密度マップを生成するために使用されます。差電子密度から、ある閾値以上の電子

密度ピークを選択し、潜在的な水分子として認識させ、削除基準によって不要ピークを削除後、残った水分子ピークの座標および温度因子が精密化されます。

## 情報科学環境

ハイスループットX線結晶解析により解析する多くの結晶構造を扱う効率を最大限にするために、X線結晶解析のアプリケーションと統合したタンパク質シーケンスおよび構造の集約データベースが開発されることを必要とします。この検索可能なデータベースシステムは、生物分子の結晶構造と、構造因子および結晶化条件を含む対応する実験データを管理するために製薬産業の要求を満たします。

## 認証研究の結果

トロンビン(分子量、約36000ダルトン)は血液凝固とフィブリノーゲン化に重要な役割をもつ多機能セリンプロテアーゼです。目立った血液凝固として、トロンビンはフィブリノーゲンを分解し、活性化血小板因子V、VII、VIII、XIII(FV、FVII、FVIII、FXIII)と同様にフィブリンを生成します(Colman et al., 1994)。また、スロンボモジュリンに結びつくと、トロンビンはコファクタープロテインSの存在下で活性化血液凝固因子Va(FVa)と活性化血液凝固因子VIIIa(FVIIIa)を不活性体に変換する活性化プロテインCによって抗血液凝固を促進します(Esmon, 1993)。トロンビンは4つの独立した作用部位(活性部位、フィブリノーゲン認識部位[エキソサイトI]、ヘパリン結合部位[エキソサイトII]、ナトリウムイオン結合部位)でその多機能性を発揮します(Tulinsky, 1996)。他のトリプシン様ファミリーメンバーと同じように、トロンビンは2つの直交ループを形成するSer214-Cys220構造領域に囲まれた活性部位にHis57-Asp102-Ser195(キモトリプシノーゲンのナンバリング、Bode et al., 1992)の3つの触媒残基を持っています。Ser214-Gly216の軸は逆並行シートを形成したり(基質結合)(Bode et al., 1989; Mathews et al., 1994)、あるいは並行シートを形成したりすること(逆結合)(Taberner et al., 1995; Mochalkin & Tulinsky 1999)によって基質やペプチド阻害剤の主鎖との結合を可能にしています。トロンビンの活性部位阻害剤はトロンビンのいくつかの残基(Ser214-O、Gly216-N、Gly216-O)に対して、2つ、あるいは3つの水素結合を持っています。Glu217-Cys220の軸はトロンビンの基質認識のキー残基であるAsp189を含んでいる特異的ポケット下部を固定します。上部で活性部位を部分的に塞いでいる残基60の挿入ループは、トリプシンやFXaと比較して、いくつかの加算的に限定されたトロンビンの活性部位特異性の原因になっています(Locht et al., 1997)。

トロンビンの結晶構造が決定されて以来(Bode et al., 1989)、様々なトロンビン阻害剤が発見されました(Balasubramanian, 1995)。トロンビンの選択的阻害は、基本的に適切でも便利でも扱いやすくないヘパリンの臨床使用と置き換えることを目的としたこれらの抗トロンビン薬物デザインの初期ターゲットになっています。このように多くの阻害剤を持ち、その複合体の構造解析が終了しているトロンビン複合体をDS HT-XPIPEの検証に使用します。

4つのトロンビン複合体を DS HT-XPIPE 検証のためにタンパク質データバンクから選択しました(表1)。1AD8 (Malikayil, J. A. et al., 1997)、1BCU (Conti, E. et al., 1998)、および 1BHX (Wagner, J. et al., 1998)複合体は、単斜晶の空間群 C2 に属しています。1DOJ 複合体は、P21212 空間群で結晶化されています(Recacha, R. et al, 2000)。

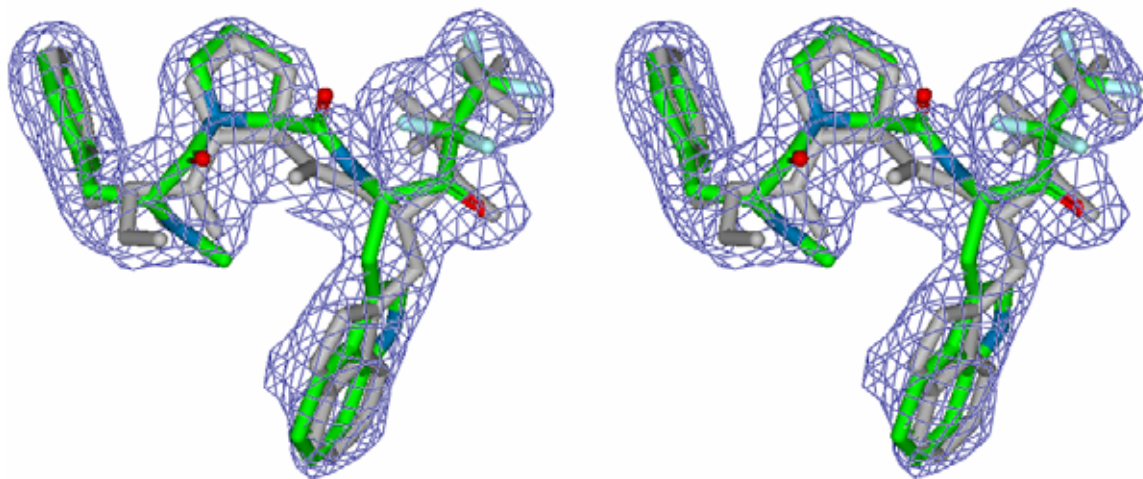
<表 1 >

PDBID	空間群	格子定数 (Å, degree.)	分解能, Å	R 値
1AD8	C2	a=70.55 b=72.05    beta=101.00 c=73.00	2.0	0.205
1BCU	C2	a=71.15 b=72.25    beta=100.61 c=73.67	2.0	0.189
1BHX	C2	a=71.00 b=72.20    beta=100.70 c=73.10	2.3	0.187
1DOJ	P21212	a=62.98 b=117.52 c=47.99	1.7	0.174

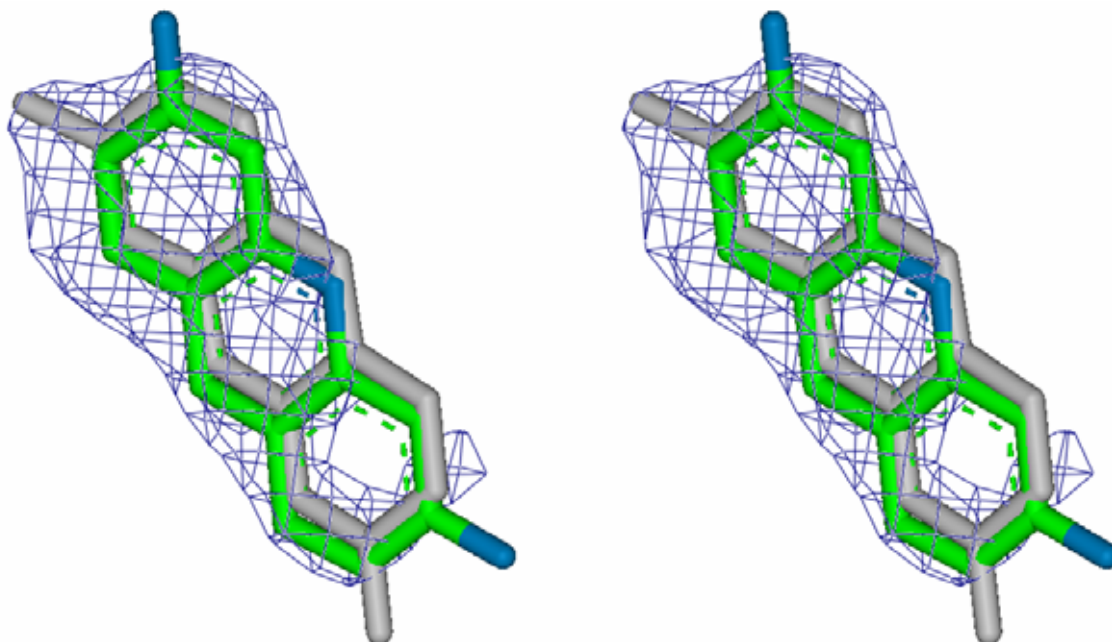
トロンビン-ヒルゲン複合体の初期検索モデル構造は PDB エントリー ID 1D6W を使用しています (Krishnan et al., 2000)。全てのリガンドの初期構造は任意に作成・配置しました。タンパク質 - リガンド複合体のタンパク質の初期座標は分子置換検索によって決定され、必要ならば、DS HT-XPIPE の分子置換および初期構造精密化オプションを使って 500.0 -3.5 の分解能範囲で剛体精密化によって最適化されます。すべての複合体でほとんどのトロンビン残基に対する電子密度は、A 鎖の末端領域 (Thr1H-Ala1B、Ile14K-Arg15、Trp148 - Lys149E 周辺の自己消化ループ、これらは他の同型トロンビン構造で典型的に消失しているもの) を除いて、十分にフィッティングできました。精密化されたタンパク質構造から計算された Fo-Fc 電子密度マップによって、1AD8、1BCU、1BHX、および 1DOJ 複合体のリガンド分子に合う電子密度が観察されました。リガンド配置、複合体精密化、水分子配置、および最終的な構造精密化は、DS HT-XPIPE によってスムーズに自動実行されました。複合体の最終的な R 値 / R-free 値は、0.222 / 0.264 (1AD8)、0.195 / 0.232 (1BCU)、0.194 / 0.253 (1BHX) and 0.243 / 0.274 (1DOJ) でした。これらの計算に要した時間は IBM ThinkPad A31p laptop (1.7 GHz Intel Pentium 4 processor) を使用して、それぞれ 56 分(1AD8)、51 分(1BCU)、37 分(1BHX)、81 分(1DOJ) でした。

DS HT-XPIPE のリガンドフィッティング方法の精度を分析するために、DS HT-XPIPE で解析された構造と PDB に登録された構造を重ね合わせしました

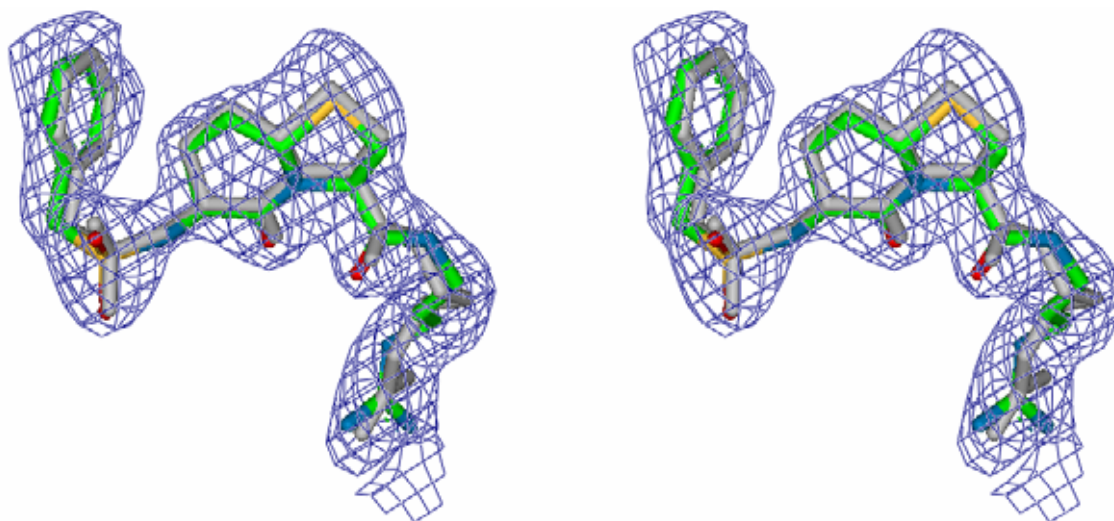
(図2)。その結果、対応したリガンドの原子位置は、ほぼ同じ位置に存在していました。1BCUと1DOJではフィッティングに追加の条件(オリエンテーション修正、トーシヨン角修正)が必要でしたが、電子密度への全自動リガンドフィッティングは正しく実行されたと言えるでしょう。DS HT-XPIPEは複合体精密化を完全に自動化して実行できたことを考慮すると、このプログラムの使用は、複合体の結晶構造決定のための時間短縮に関して効率的であると証明できます。



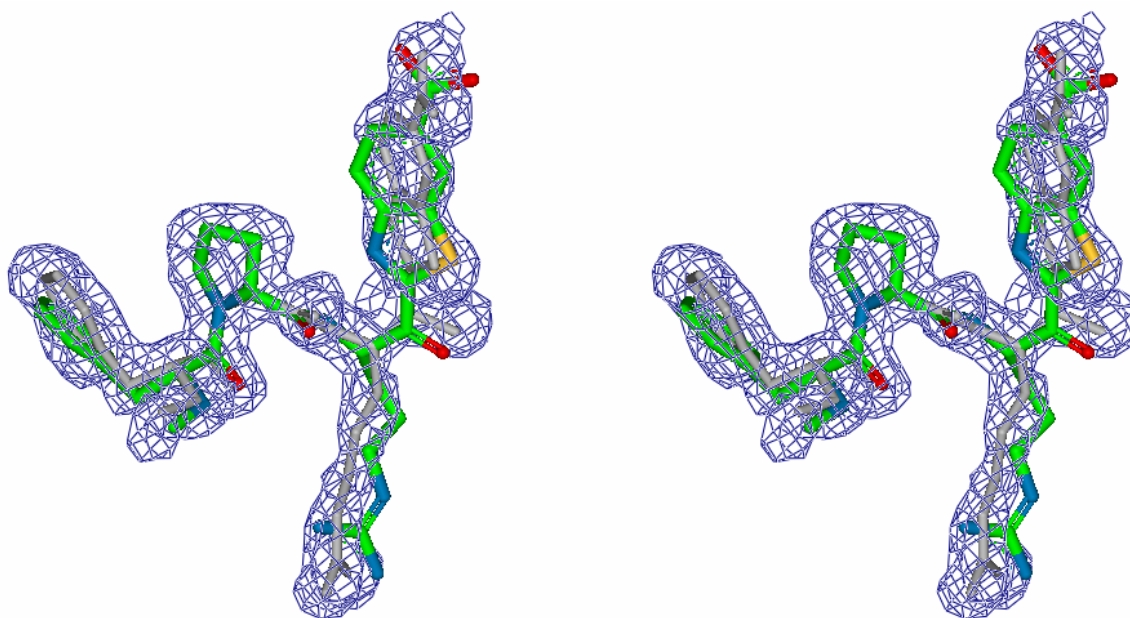
(a) 1AD8



(b) 1BCU



(c) 1BHX



(d) 1DOJ

図2： トロンビン複合体の各阻害剤 1AD8 (a)、1BCU (b)、1BHX (c)、および 1DOJ (d) の最終的な 2Fo-Fc 電子密度マップ (1シグマレベル表示) のステレオ図。DS HT-XPIPE で解かれた 1AD8、1BCU、1BHX、および 1DOJ 複合体は、各原子の色で表示され、PDB の最終リガンド座標は灰色で表示されています。



## 結論

タンパク質産生、結晶化およびX線データ収集の自動化における劇的な進歩によって、1つの指令(コマンド)で多数のタンパク質-リガンド複合体の結晶構造を決定することができるハイスループットX線結晶解析用の新しいソフトウェアの開発を要求されます。Discovery Studio HT-XPIPEは、人間の介在を最小限に抑えて結晶構造決定を実行するために、分子置換、精密化、リガンド配置および水分子配置に対して完全に自動化され合理化されたプロトコルを提供します。異なるトロンビン複合体で実行された確認研究の結果は、重大なコンフォメーション変化を受けない、またリガンドの電子密度領域がつながっている十分に精密化されたタンパク質系でハイスループットな構造決定が可能であることを指し示しています。

## 参考文献

- Badger, J., Ajay Kumar, R., Yip, P., Szalma, S. (1999) *Proteins* 35, 25-33.
- Balasubramanian, B. N. Ed. (1995). *Biorg. & Med. Chem.* 3, 499-1156.
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. and Bourne, P. E. (2000) *Nucleic Acids Research*, 28, 235-242.
- Bode, W., Mayr, I., Baumann, U., Huber, R., Stone, S. & Hofsteenge, J. (1989) *EMBO J.* 8, 3467-3475.
- Bode, W., Turk, D. & Karshikov, A. (1992) *Protein. Sci.* 1, 426-471.
- Brunger, A.T. (1992) X-PLOR Version 3.1: A System for Crystallography and NMR. Yale University Press, New Haven.
- Colman, R. W., Marder, V. J., Salzman, E. W. & Hirsch, J. (1994) Hemostasis and Thrombosis, 3<sup>rd</sup> ed., edited by Colman, R.W., Hirsch, J., Marder, V. J. & Salzman, E. W. pp. 3-18, J. B. Lippincott Co., Philadelphia.
- Conti, E., Rivetti, C., Wonacott, A., Brick, P. (1998) *FEBS Lett* 425, 229-233
- De Simone, G., Balliano, G., Milla, P., Gallina, C., Giordano, C., Tarricone, C., Rizzi, M., Bolognesi, M., Ascenzi, P. (1997) *J Mol Biol.* 269, 558-569.
- Džakula, Ž. to be published.
- Esmon, C. T. (1993) *Thrombo. & Haemost.* 70, 29-35.

- Krishnan R., Mochalkin. I., Arni, R. and Tulinsky, A. (2000) *Acta Cryst.* D56, 294-303.
- Lesley, S. A. (2001) *Protein Exp. Purif.* 22, 159–164.
- Locht, A., Bode, W., Huber, R., Bonniec, B. F., Stone, S. R., Esmon, C. & Stubbs, M. T. (1997) *EMBO J.* 16, 2977-2984.
- Malikayil, J. A., Burkhart, J. P., Schreuder, H. A., Broersma Jr., R. J., Tardif, C., Kutcher, Mehdi, S., Schatzman, G. L., Neises, B., Peet, N. P. (1997) *Biochemistry* 36, 1034-1040.
- Mathews, I. I., Padmanabhan, K. P., Ganesh, V., Tulinsky, A., Ischii, M., Chen, J., Turck, C. W., Coughlin, S. R. & Fenton, J. W. (1994) *Biochemistry* 33, 3266-3276.
- Mochalkin, I. & Tulinsky, A. (1999) *Acta Cryst.* D55, 785-793.
- Mueller U, Nyarsik L, Horn M, Rauth H, Przewieslik T, Saenger W, Lehrach H. and Eickhoff H. 2001 Development for automation and miniaturisation of protein crystallisation. *J. Biotechnol.* 85, 7–14.
- Oldfield, T. (2001) *Acta Cryst.* D57, 696-705.
- Recacha, R., Costanzo, M. J., Maryanoff, B. E., Carson, M., Delucas, L., Chattopadhyay, D. (2000) *Acta Cryst.*, D56, 1395-1400.
- Stevens, R.C. (2000) *Curr. Opin. Struct. Biol.* 10, 558–563.
- Taberner, L., Chang, C. Y., Ohringer, S. L., Lau, W. F., Iwanowica, E. J., Han, W. C., Wang, T. C., Seiler, S. M., Roberts, D. G. M. & Sack, J. S. (1995) *J. Mol. Biol.* 246, 14-20.
- Tulinsky, A. (1996). *Seminars in Thrombosis and Hemostasis.* 22, 117-124.
- Wagner, J., Kallen, J., Ehrhardt, C., Evenou, J. P., Wagner, D. (1998) *J Med Chem.* 41, 3664-3674.