

ドメイン、Pfam、Rasmol : マルチプルアラインメントから構造機能部位の特定へ

タンパク質の機能にはまだよくわかっていない部分も多いが、ここではタンパク質は「他の分子との相互作用」により、その機能を発現していると考えことにしよう。他の分子としては、1) ホルモン、生体構成分子などのいわゆる低分子などである場合、2) DNA, RNAなどの核酸である場合、3) 他のタンパク質分子である場合、がある。

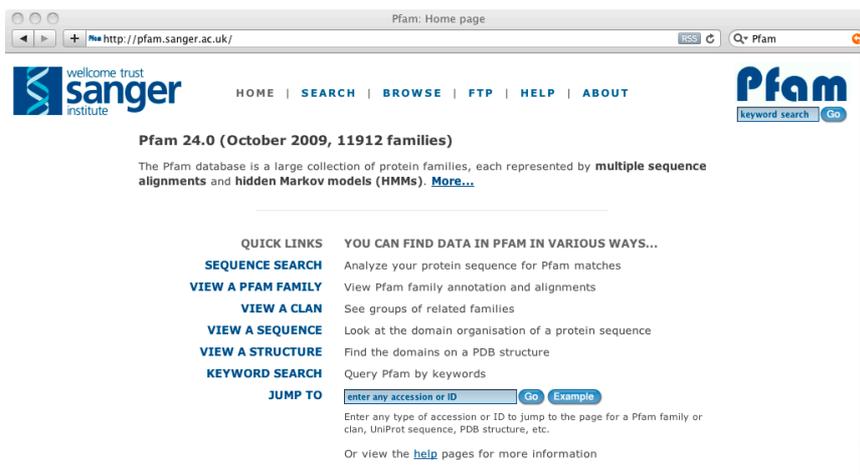
例えば、酵素タンパク質は、反応前の化合物（基質）を活性部位（活性サイト）へ取り込み、化学的な修飾を加えた反応後の化合物（反応物）を放出するサイクルを繰り返す。また、特定のシグナル化合物が膜タンパクの細胞外に露出したドメインに結合することにより、細胞内で別の酵素やシグナル系が活性化される。こうしたシグナル伝達の経路ではタンパク質間の相互作用が重要な役割を果たす。核酸に結合するタンパク質は例えば、結合部位の核酸配列に応じて、結合強度をコントロールすることで、転写、翻訳の制御を行っている。

多くの場合、これらの相互作用において最も重要になるのは、これらの2つ（あるいはそれ以上）の分子が会合する、立体構造上の「界面」（接触面）である。この、広い意味での「活性サイト」を特定することがタンパク質の機能を知り、さらには制御してゆく上での第一歩となる。

タンパク質による化合物の認識は一般に、「特異性」が極めて高い。すなわち、似通った構造を持つ分子の混合物から、特定の構造を持つ分子だけを選び出して、相互作用を行う能力があると考えられている。こうした特異的な相互作用の分子レベルでの詳細な描像がわかれば、その知識を利用していろいろなことが可能になる。例えば、ある酵素の活性サイトの構造を調べ、その酵素の本来の基質よりも強く結合できる分子をデザインできれば、酵素の触媒する反応を阻害することが出来る。（阻害剤：インヒビタ）実際には、こうした阻害剤を薬として用いるためには、副作用への対処など対処しなければならない数多くの事柄があるが、薬の開発の初期の段階では有望な化合物を絞り込むために有効に用いられる手法となっている。

今日は、配列、立体構造の関係から配列上の機能部位を推定してゆくプロセスを理解することを目指す。例として、ETSドメインのDNA認識部分を探す。

Pfamは、タンパク質の配列ドメインのデータベースである。配列ドメインはマルチプルアラインメントの形で保存されている。一般に、特定のドメインは特定の機能と深く結びついている。すでに、立体構造のわかっている配列が含まれる場合には、構造、配列、機能を関連づけられる場合もある。立体構造が未知の場合でも、機能上重要な部位は保存されやすい（変異を受けると機能を失い退化する）ことを考えると、機能部位が推定される場合もある。



ページ右上のkeyword search(検索欄)で、ETSを探す、

Section	Description	Number of hits
Pfam	Text fields for Pfam entries	2
Seq_info	Sequence description and species fields	441
Pdb	HEADER and TITLE records from PDB entries	28
GO	Gene ontology IDs and terms	0
Interpro	InterPro entry abstracts	4

Accession	ID	Description	Pfam	Seq_info	Pdb	Interpro
PF00178	Ets	Ets-domain	✓	✓	✓	✓
PF04621	ETS_PEA3_N	PEA3 subfamily ETS-domain transcription factor N terminal domain	✓	✓		
PF02198	SAM_PNT	Sterile alpha motif (SAM)/Pointed domain		✓	✓	✓
PF00005	ABC_tran	ABC transporter		✓		
PF00009	GTP_EFTU	Elongation factor Tu GTP binding domain		✓		

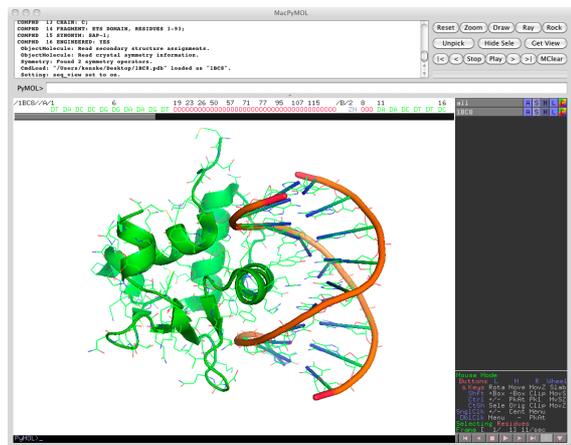
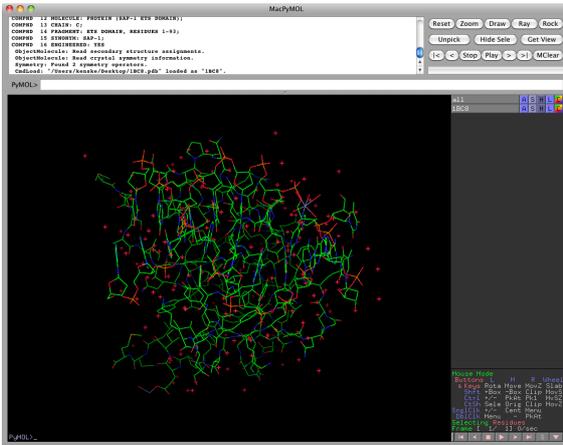
出力される候補リストの一番目PF00178をクリックするとETSドメインに関する情報が示される。

右側に立体構造が示されていれば、このファミリーで構造のとかれているものがあることがわかる。図の下にあるリストボックスを操作するとファミリー内の他の構造が示される。このサマリーページには文献情報やファミリーに関する情報が簡潔にまとめられている。左側のメニューでAlignmentsを選ぶと以下のようなアラインメントが得られる。Seedとはこのファミリーを代表する配列として人間が集めた配列で、これらをもとにしてHMMプロファイルを作成し配列データベースの中から類似した（ホモロジーのある）配列をすべて収集したものがFullである。ETSについては15の配列をもとに、762本の配列がファミリーとして集められていることになる。下のViewボタンを押せばこれらの配列のアラインメントを見ることができる。

```

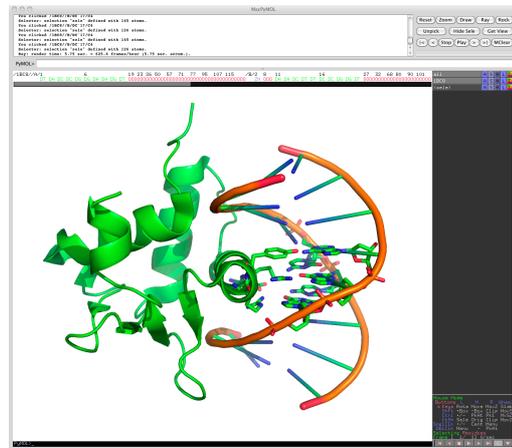
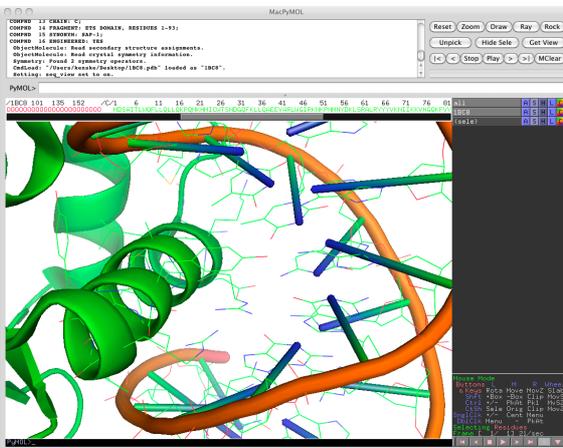
ETY6_HUMAN/338-422    GRLLWDYYQLLSD..SRYENFIWEDKESKIFRIV..DPNGLARLWGNHKN.RTNMTEKMSRALRHYKLNIRKEP.GQRLLFRFMKT
POK_DROME/395-481    GRLLWDFLQQLLNDRNQKYSDLIAWKCRDTGVFKIV..DPAGLAKLWGIQKN.HLSMNYDKMSRALRYRYRNILRKYVQ.GERHCYQFLRN
ETY3_HUMAN/56-140    QIQLWHFLELLQK..EEFRHYIAWQQGEYGFYIK..DPDEVARLWGRKCK.KPQMNYDKLSRALRYRYNKRILHKT.KGRFTYKFNFN
GABA_HUMAN/319-402   QIQLWQFLLELLTD..KDARDCCISWYG.DEGEFKLN..QPELYAQKWGRKKN.KPTMNYEKLRSALRYRYDGMICKYVQ.GKRFYKFDYV
ELG_DROME/345-428    QVQLWQFLLELLTD..CEHTDVIWYWG.TEGEFKLT..DPDRVARLWGERKKN.KPAMNYEKLRSALRYRYDGMICKYVQ.GKRFYKFDYV
ETS1_HUMAN/334-417   PIQLWQFLLELLTD..KSCQSFISWTG.DGWEFKLS..DPDEVARRWGRKKN.KPKMNYEKLRSALRYRYDGMICKYVQ.GKRYRYFVCD
ERG_HUMAN/293-376    QIQLWQFLLELLSD..SSNSSCITWEG.TNGEFKMT..DPDEVARRWGERKKS.KPNMNYDKLSRALRYRYDGMICKYVQ.GKRYRYFVCD
ETV2_MOUSE/233-316  PIQLWQFLLELLQD..GARSSCIRWTG.NSREFQLC..DPKEYARLWGERKKN.KPAMNYEKLRSALRYRYDGMICKYVQ.GKRYRYFVCD
ETV4_HUMAN/407-490  ALQLWQFLYALLDD..PTNAHFIAWTG.RGMEFKLI..EPEEARLWGIQKN.RPAMNYDKLSRSALRYRYDGMICKYVQ.GERYRYFVCE
ELK3_MOUSE/4-87      AITLWQFLHLLLD..QKHEHLICWTS.NDGEFKLL..KAEVYAKLWGLRKN.KTNMNYDKLSRALRYRYDGMICKYVQ.GQKFYKFSYF
ELK1_HUMAN/4-88      SYTLWQFLQLLRE..QGNHGISWTSRDGGEFKLV..DAEEVARLWGLRKN.KTNMNYDKLSRALRYRYDGMICKYVQ.GQKFYKFSYF
E74A_DROME/732-817  TTYLWQFLKLLQDR.EYCPFRFKWTRNREKGVFKLV..DSKAVSRLWGMHKN.KPDMNYETMGRALRYRYDGMICKYVQ.GQRYVYHFVDY
ETV4_HUMAN/407-490  HIHLWQFLKELLASP.QYNGTAIRWIDRSKGIKIE..DSYRYAKLWGRKKN.RPAMNYDKLSRSIRQYKKGIMKKTERTSQRLLYQFCHP
SPIB_HUMAN/168-254  KLRLYQFLGLLTR..GDMRECVWYEPGAGYVGFSSKHKELLARRWQOQGNRKRMTYQKLARALRNYAKTGEIRKVK..RKLTYQFDSA
PUI_HUMAN/163-249    KIRLYQFLDLLRS..GDMKDSIWWVDKGTGFSSKHKELALHRWGIQKGNRKKMTYQKMARALRNYGKTGEVKKVK..KCLTYQFSGE
  
```

ETSドメインがどのようにDNAと相互作用しているかを観てみよう。Pfamのリストの中からここでは1bc8を選択した。PDBのページから1BC8のデータファイルをダウンロードしPymolで開く。背景を白とし(display->background->white) 配列を表示(display->Sequence On)、水を消して(all->H->Waters)、全体をCartoon表示(all->S->Cartoon)すると右側の図のようになる。



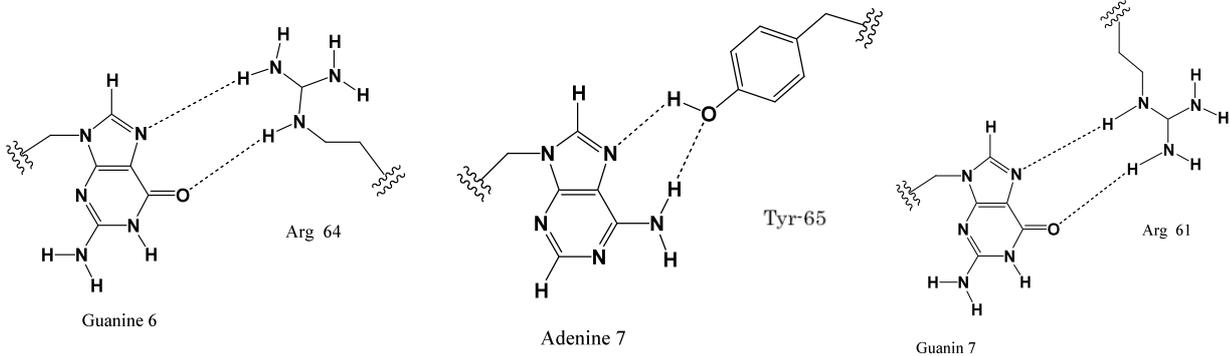
DNA二本鎖の広い溝 (major groove)に対してETSのヘリックスが食い込んでいるのがわかる。このヘリックスはAsn-53からVal-68にかけての部分であるが、この部分とDNAの相互作用をより詳しく観てみよう。

一旦、線面表示をすべて消し(all -> H -> lines)、上の配列表示から、53から68番の残基とDT, DA, DC, DGなどと示されたDNA塩基を選択しこの部分を線面表示とする(sele -> S -> lines) この状態で、近くにある残基と塩基を探してみよう。下、左図



たとえば、Tyr-65とDNA-A鎖の7番目のA、Arg-61とDNA-A鎖の6番目のGおよびB鎖の6番目のC、Asp57とArg-64とDNA-A鎖の5番目のGとB鎖の7番目のCなどの間に水素結合を含む相互作用が認められる。これらの残基を選択し、Stick表示として、再び線面を消すと、上の右側の図のような表示が得られる。

これらの残基の位置関係を見れば、化学式で示したような水素結合が、Arg61, Arg64, Tyr65によるGGAという配列の認識に関わっていることが了解される。



今一度、5ページのマルチプルアラインメントに戻ってみよう。シーケンスを見比べればDNAとの相互作用に関与していた $\alpha$ ヘリックスは配列上後半の部分にあることがわかる。最も重要なArg61, Arg64はRALRといった並びの中にある。全配列を通してこの二つのアルギニンは保存されている。次の、Tyr65については、大部分で保存されているが、一本目がHis、最後の方で、Asn, Glnなどに置換されているようである。このことはどのように解釈されるであろうか？考えてみることに。

```

ETY6_HUMAN/338-422  CRLLWDYYYQLLSD..SRYENFIRWEDKESKIFRIY..DPNGLARLWGNHKN.RTNMTYEKMSRALRHYKLNIIIRKEP.GQRLFRFMKT
POK_DROME/395-481  GRLLWDFLQQLNDRNQKYSDLIAWKCRDTGVFKIV..DPAGLAKLWGIQKN.HLSMNYDKMSRALRYYYRYNIIIRKVV.GERHCYQFLRN
ETY3_HUMAN/56-140  QIQLWHFILELLQK..EEFRHYIAWQQGEYGEFYIK..DPDEVARLWGRKRC.KPQMNYDKLSRALRYYYNKRILHKT.KGRFTYKFNFN
GABA_HUMAN/319-402  QIQLWQFLELLTD..KDARDCISWVG.DEGEFKLN..QPELYAQKWGRKN.KPTMNYEKLRSRALRYYYDGMICKVQ.GKRFYKFFVCD
ELG_DROME/345-428  QVQLWQFLLEILD..CEHTDVIIEWVG.TEGEFKLT..DPDRVARLWGEKKN.KPAMNYEKLRSRALRYYYDGMISKVS.GKRFAYKFDCC
ETS1_HUMAN/334-417  PIQLWQFLELLTD..KSCQSFISWTG.DGWEFKLS..DPDEVARRWGRKKN.KPKMNYEKLRSGLRYYYDKNIIHKTA.GKRYVYRFVCD
ERG_HUMAN/293-376  QIQLWQFLELLSD..SSNSSCITWEG.TNGEFKMT..DPDEVARRWGERKS.KPNMNYDKLSRALRYYYDKNIMTKVH.GKRYAYKDFDH
ETY2_MOUSE/233-316  PIQLWQFLELLQD..GARSSCIRWTG.NSREFQLC..DPKEVARLWGERKR.KPGMNYEKLRSGLRYYYRRDIIYKSG.GRKYTYRFGGR
ETY4_HUMAN/407-490  ALQLWQFLYALLDD..PTNAHFIAWTG.RGMEFKLI..EPDEVARLWGIQKN.RPAMNYDKLSRSLRYYYEKGIMQKVA.GERYYKVFCE
ELK3_MOUSE/4-87    AITLWQFLHLLLD..QKHEHLICWTS.NDGEFKLL..KAEEYAKLWGLRKN.KTNMNYDKLSRALRYYYDKNIIKKVI.GQKFVYKVFVSF
ELK1_HUMAN/4-88    SVTLWQFLQLLRE..QGNHIIISWTSRDGGEFKLV..DAEEVARLWGLRKN.KTNMNYDKLSRALRYYYDKNIIIRKYS.GQKFVYKVFVSY
E74A_DROME/732-817  TTYLWEFLKLLQDR.EYCPFRFKWTNREKGVFKLV..DSKAYSRLWGMHKN.KPDMNYETMGRALRYYYQRGILAKVD.GQRLYYHFVDY
ETS4_DROME/415-501  HIHLWQFLKELLASP.QYNGTAIRWIDRSKGIFKIE..DSYRVAKLWGRKKN.RPAMNYDKLSRSIRQYKKGIMKKTERSORLYYQFCHP
SPIB_HUMAN/168-254  KLRLYQFLGLLTR..GDMRECVWVVEPGAGYVQFSSKHKELLARRWQQQGNRKRMTYQKLARALRNYAKTGEIRKVK..RKLTQYQFDSA
PU1_HUMAN/163-249  KIRLYQFLDLLRS..GDMKDSIWWWYDKGTGFQFSSKHKKEALHRWGIQKGNRKRMTYQKMARALRNYGKTGEVYKVK..KKLTYQFSGE

```

ターゲットの立体構造が解っていても、そこにぴったりはまるような化学構造を作り出すことは実はそれほど容易ではない。今回見た、DNAに対するタンパク残基の特異的な相互作用も、自然が長い時間をかけて見つけたものである。様々な計算手法の進歩に伴い、いろいろな構造間の相対的な結合の安定性を見積もる方法は次第に進歩しつつある。タンパク質のアミノ酸を人工的に変異させて、機能をコントロールすることも可能になりつつある。ランダムなスクリーニングによるコストの削減のためにも、計算機による予測はこれから重要性を増してゆくであろう。