**如何理解蛋白质-小分子的识别过程？**

蛋白质与小分子的相互作用对生物体内主要的生命过程都至关重要，酶催化，信号传导，物质转运等等，了解和探究蛋白质与小分子相互作用的机理一直是重要的研究课题。

**一. 热动力学角度理解蛋白质-小分子相互作用**

1. 热动力学原理：整个热力学系统包括蛋白质，小分子以及溶剂（水以及离子）。

∆G = ∆H - T∆S， 其中∆G代表系统的自由能变化，当自由能变化为负值时可以自发结合。

∆H： 代表系统焓值的变化，焓是表征整个系统能量的值，焓值的降低有利于结合发生。通常小分子与蛋白质之间非共价键的形成（疏水作用，静电作用，氢键等）可以造成焓值下降。反之，非共价键的破坏则会使焓值上升。

∆S: 代表系统熵变化，熵是代表整个系统混乱度的值，自由度越大，越混乱，熵值增加，反之有序的系统熵值降低。而蛋白质与小分子的非共价键形成会造成熵值下降。

∆S = ∆Ssolv + ∆Sconf +  ∆Sr/t ∆Ssolv: 表示溶剂的变化带来的熵变化。通常蛋白质与小分子的结合会导致，口袋部分受限制溶剂的释放，这一过程的熵值会大幅增加，是结合的重要驱动力。∆Sconf： 表示结合过程导致的构象变化带来的熵值改变，对结合的贡献可正可负，取决于结合与非结合状态构象的差别。 ∆Sr/t： 表示蛋白质和小分子平移和旋转自由度的变化带来的熵值改变，由于结合后粒子数目的减少导致上改变为为负值。

另外，‘熵焓补偿’的概念指，在蛋白质与小分子作用过程中，焓值的降低也会导致熵值的下降，从而使∆G 的变化某种程度上不会过于巨大。

2. 解释蛋白质-小分子相互作用的模型（锁钥模型，诱导契合，构象选择）

锁钥模型：假定蛋白质小分子均是刚性结构，在形状上完全互补。 该模型是理想的条件下，由于不考虑构象问题，仅仅依靠几何形状上的互补，这一过程是熵主导的。

诱导契合IF ([Koshland, 1958](#_ENREF_7))：小分子的结合诱发蛋白质分子构象改变，从而实现最优结合。这一过程中，起主导作用的是焓值变化。

构象选择CS ([Csermely, et al., 2010](#_ENREF_5); [Tobi and Bahar, 2005](#_ENREF_13))： 蛋白质分子以多种构象的形式存在，选择最优的构象与小分子结合。这一过程中，熵变化先主导结合，后焓变化驱动构象改变。

具体分析见综述([Xing Du, 2016](#_ENREF_16))

同时，关于三种模型之间的关系： 锁钥模型可以被认为是极端情况下的构象选择。 在诱导契合和构象选择之间存在着很多争论，这些争论主要在于： 蛋白质的构象变化是否需要小分子的结合来诱导。

([Changeux and Edelstein, 2011](#_ENREF_2))总结了50年以来的实验数据和模拟数据，认为在信号传导上的小分子基本上都是支持构象选择的。

而([Greives and Zhou, 2014](#_ENREF_6))的工作得出结论认为蛋白质构象的变化速率以及小分子的浓度都可以改变主导结合的原理，低的变化速率和小分子浓度一般对应CS, 而高的变化速率和小分子浓度则对应IF.

**二. 决定蛋白质-小分子结合的具体的相互作用分析**

以上，是从热动力学的角度去阐释蛋白质-小分子的作用机理。而这些热动力学的机制最终归根于蛋白质-小分子具体的相互作用，落实到具体的相互作用上，又提出以下三个重要的科学问题：第一，是什么相互作用（疏水，氢键，静电相互作用，范德华力）主导了结合的过程，第二，是哪些相互作用决定了蛋白质与小分子结合的亲合力，强弱结合如何在相互作用上得以区分？ 第三，最终这些相互作用又如何在热动力学上得以解释。

1. 哪些相互作用驱动了蛋白质与小分子的识别和结合

蛋白质-小分子结合实质上是蛋白质与小分子之间非共价键的形成和破坏导致的‘熵-焓补偿’的微妙平衡。而具体的这些非共价相互作用对于小分子识别过程的贡献则一直存在争论，并没有明确的结论，从不同的立场考虑可能得到不同的结果，so it is still unclear.

(1).有研究认为是疏水相互作用和范德华力驱动和主导了蛋白质与小分子的识别过程。([Kuntz, et al., 1999](#_ENREF_8); [Tom Young, 2006](#_ENREF_14)) (2). 就有研究认为静电相互作用在结合过程中起到重要作用。([Chong, et al., 1998](#_ENREF_4); [Olson, et al., 2001](#_ENREF_10))

以上研究将蛋白质与小分子的结合当成一个完整的过程来研究，从不同的角度看，可能得到不同的结论。这里，我们将蛋白质与小分子的结合过程分成二个步骤考量，第一是结合过程，即蛋白质与小分子通过怎样的方式识别彼此，然后实现结合过程。第二是解离过程，即小分子离开蛋白质扩散到溶剂当中，这一步骤决定了小分子与蛋白质的亲和力。小分子越难以解离说明亲和力越大，反之则亲和力越小，参考于([Seo, et al., 2014](#_ENREF_12))。利用系统生物学的手段可以对这两个过程加以探究。

关于结合过程，PNAS上的论文有重要参考价值：([Wang, et al., 2011](#_ENREF_15))认为‘presence of dry regions in the receptor has a nontrivial effect on ligand binding affinity, and suggest that such regions may represent a general motif for molecular recognition between the dry region in the receptor and the hydrophobic groups in the ligands’。 JACS 上的另外一篇论文([Raman and MacKerell, 2015](#_ENREF_11))认为' the entropy of water reorganization is found to favor binding in agreement with the classical view ofthe “hydrophobic effect"

**疏水相互作用和范德华力是主要识别驱动力。**

关于亲和力的大小，有很多关于研究亲和力大小的结构基础（包括蛋白质结构和小分子结构）的论文。

(1) ([Seo, et al., 2014](#_ENREF_12))通过MBP 的例子说明是蛋白质构象的动力学变化决定了与小分子的亲和力大小，而这些构象变化对于结合过程的影响很小。

(2) ([Chen, et al., 2016](#_ENREF_3))通过调节蛋白质与小分子氢键的配对可以改变结合的亲合力大小，说明氢键在某些情况下对于结合的亲和力也有重要的影响。

(3) ([Nayak, et al., 2016](#_ENREF_9))证明四个氨基酸残基的改变可以将蛋白质功能改变（亲和力大小不同，有30倍差别）（亲和力大小由少数的氨基酸残基决定）

(4) ([Carlson, et al., 2008](#_ENREF_1))认为‘ The differences in ligand efficiencies do not appear to come

from the ligands; instead, the pockets yield different amino acid compositions despite very similar distributions of amino acids in the overall protein sequences’： 口袋结构的差异决定了分子的亲和力大小。

(5) ([Kuntz, et al., 1999](#_ENREF_8)) 提出ligand efficiency的概念用于描述小分子的亲和能力。

总结起来，静电相互作用，氢键等可能在解离过程中起重要作用，这些相互作用会引发蛋白质构象的变化。extended conformation selection： Conformation selection followed by induced fit?

**三. 我们自己的实验思路是啥？**

1. 对所有满足去冗余后超过八个结合蛋白的小分子，抽取保守程度不同的结合模块。对这些保守模块的氨基酸组成进行分析，看哪些氨基酸占据主导地位，即可从系统生物学的角度分析出是什么相互作用主导了蛋白质与小分子的识别，保守程度越高的模块表明在结合过程起到关键作用。

2. 分析所有的有亲和力数据的小分子（根据不同的亲和力分组），总结出不同亲和力下，对应的结构特征（模块特征）

3.对同一种小分子（ATP等）构建不同亲和力下结合蛋白的数据集，比较不同亲和力下结合模块的差别，结合1,2 的分析可知，不同类型的相互作用，对于结合，解离二个过程的不同程度的贡献。

4. 结合这些信息，看能否对热动力学的模型有更加准确的理解。

Carlson, H.A.*, et al.* (2008) Differences between High- and Low-Affinity Complexes of Enzymes and Nonenzymes, *Journal of medicinal chemistry*, **51**, 6432-6441.

Changeux, J.P. and Edelstein, S. (2011) Conformational selection or induced fit? 50 years of debate resolved, *F1000 biology reports*, **3**, 19.

Chen, D.*, et al.* (2016) Regulation of protein-ligand binding affinity by hydrogen bond pairing, *Science advances*, **2**, e1501240.

Chong, L.T.*, et al.* (1998) Computation of electrostatic complements to proteins: a case of charge stabilized binding, *Protein science : a publication of the Protein Society*, **7**, 206-210.

Csermely, P., Palotai, R. and Nussinov, R. (2010) Induced fit, conformational selection and independent dynamic segments: an extended view of binding events, *Trends in biochemical sciences*, **35**, 539-546.

Greives, N. and Zhou, H.X. (2014) Both protein dynamics and ligand concentration can shift the binding mechanism between conformational selection and induced fit, *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 10197-10202.

Koshland, D.E. (1958) Application of a Theory of Enzyme Specificity to Protein Synthesis, *Proceedings of the National Academy of Sciences of the United States of America*, **44**, 98-104.

Kuntz, I.D.*, et al.* (1999) The maximal affinity of ligands, *Proceedings of the National Academy of Sciences of the United States of America*, **96**, 9997-10002.

Nayak, T.K.*, et al.* (2016) Structural correlates of affinity in fetal versus adult endplate nicotinic receptors, *Nature communications*, **7**, 11352.

Olson, C.A.*, et al.* (2001) Cooperative helix stabilization by complex Arg-Glu salt bridges, *Proteins*, **44**, 123-132.

Raman, E.P. and MacKerell, A.D. (2015) Spatial Analysis and Quantification of the Thermodynamic Driving Forces in Protein-Ligand Binding: Binding Site Variability, *J Am Chem Soc*, **137**, 2608-2621.

Seo, M.H.*, et al.* (2014) Protein conformational dynamics dictate the binding affinity for a ligand, *Nature communications*, **5**, 3724.

Tobi, D. and Bahar, I. (2005) Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state, *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 18908-18913.

Tom Young, R.A., Byungchan Kim, Bruce J. Berne\*, and Richard A. Friesner (2006) Motifs for molecular recognition exploiting hydrophobic enclosure in protein–ligand binding, *PNAS*.

Wang, L., Berne, B.J. and Friesner, R.A. (2011) Ligand binding to protein-binding pockets with wet and dry regions, *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 1326-1330.

Xing Du, Y.L., Yuan-Ling Xia, Shi-Meng Ai,Jing Liang,Peng Sang, Xing-Lai Ji and Shu-Qun Liu (2016) Insights into Protein–Ligand Interactions: Mechanisms, Models, and Methods, *Int. J. Mol. Sci*.

Results

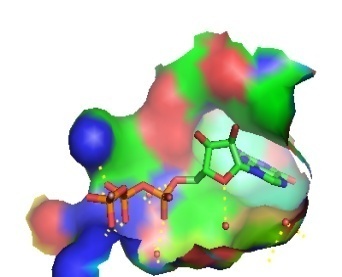
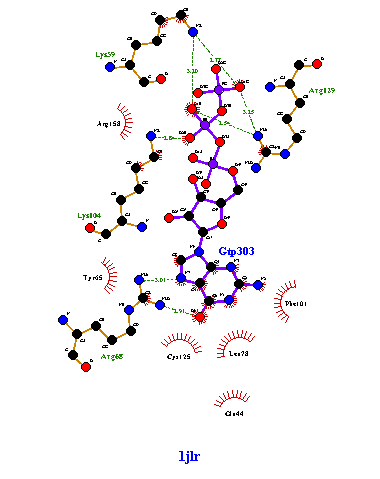
ATP:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| PDBCode (high-affinity) | Protein description | Kd value  (uM) | PDBCode  (low-affinity) | Protein description | Kd value  (uM) |
| 1B38\_A | cyclin-dependent kinase 2 | 0.254 | 1GZ3\_A | NAD (P)+-Dependent Malic Enzyme | 200 |
| 1PK8\_A | rat synapsin I | 0.120 | 1M83\_A | tryptophanyl-tRNA synthetase | 400 |
| 1R0X\_A | nucleotide-binding domain 1 | 0.149 | 1L2T\_A | ABC transporter | 800 |
| 1Y8P\_A | pyruvate dehydrogenase kinase 3 | 0.840 | 1VJC\_A | 3-phosphoglycerate kinase | 263 |
| 2FGH\_A | gelsolin bound to ATP | 0.280 | 2CBZ\_A | Human Multidrug Resistance Protein 1 | 118 |
| 2HMU\_A | RCK Domain of the KtrAB K(+) | 0.020 | 2J9L\_A | Human Chloride Transporter Clc-5 | 100 |
| 2P09\_A | a non-biological protein | 0.450 | 2KMX\_A | Human ATP7A | 83 |
| 2W5G\_A | ribonuclease A | 0.860 | 2VHQ\_A | P4 protein of bacteriophage phi12 | 140 |
| 3D2E\_A | complex of Sse1p and Hsp70 | 0.380 | 3DNT\_A | Methyltransferase-MM\_2633 | 18 |
| 3DGL\_A | A synthetic protein | 0.250 | 3B2Q\_A | A1Ao ATP synthase | 28 |
| 3HGM\_A | universal stress protein TeaD | 0.800 | 3QX9\_A | Human AGO2 MID domain | 90 |
| 3DGO\_A | Tyr-Phe mutation of 3DGL | 0.180 | 1XSC\_A | nudix enzyme AP4A hydrolase | 50 |

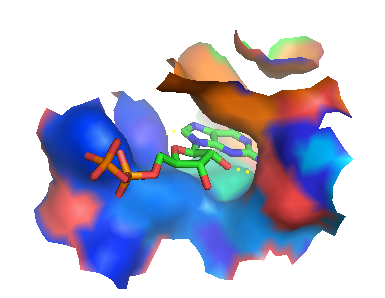
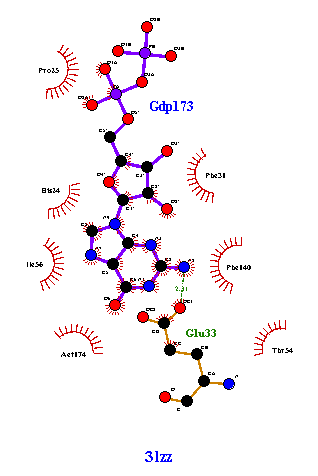
FMN:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| PDBCode (high-affinity) | Protein description | Kd value  (nM) | PDBCode  (low-affinity) | Protein description | Kd value  (nM) |
| 1AG9\_A | oxidized flavodoxin | 1.000 |  |  |  |
| 1AKQ\_A | D95A oxidized flavodoxin mutant | 0.790 |  |  |  |
| 1AZL\_A |  |  |  |  |  |
| 1C7F\_A |  |  |  |  |  |
| 1CZR\_A |  |  |  |  |  |
| 1OBV\_A |  |  |  |  |  |
| 3A20\_A | FMN-binding protein | 1.120 |  |  |  |
| 3A6R\_A | E13Q FMN-binding protein mutant | 1.390 |  |  |  |
| 2V5V\_A | W57E flavodoxin mutant | 4.800 |  |  |  |
| 1FLM\_A | Dimer of FMN-binding protein | 0.430 |  |  |  |

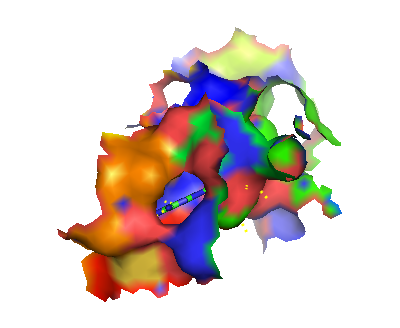
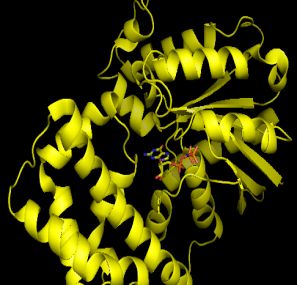
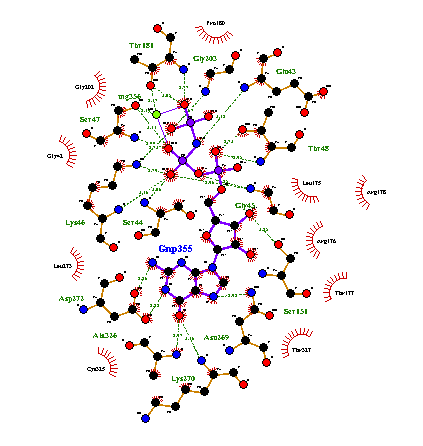
1JLR: Transferase GTP complex. (kd = 0.465mM)



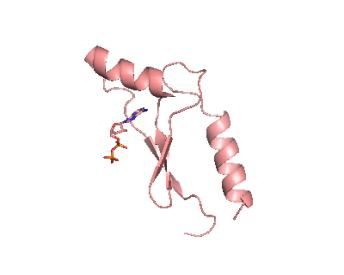
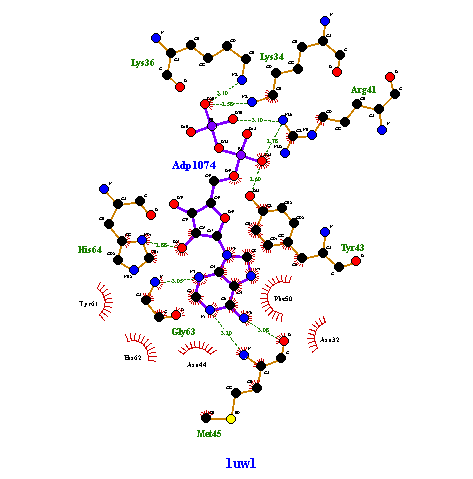
3LZZ: Cupin superfamily BbDUF985 ( kd = 0.146mM)



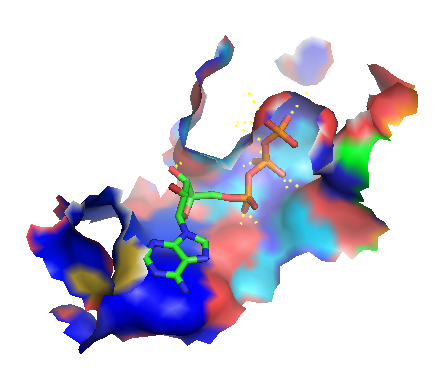
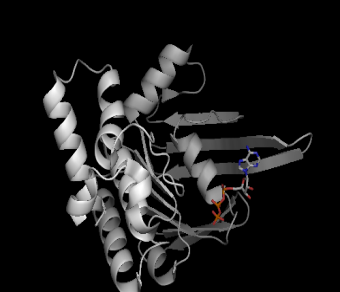
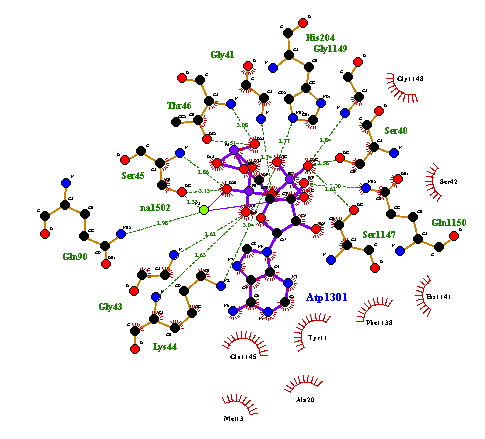
1SVS Hydrolase signaling protein (kd = 2.3nM)



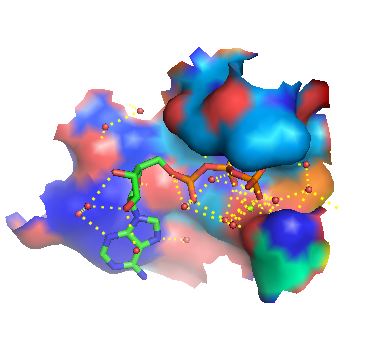
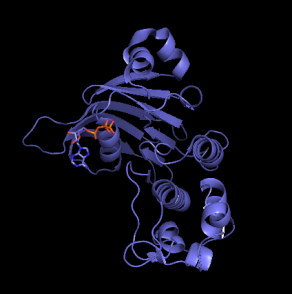
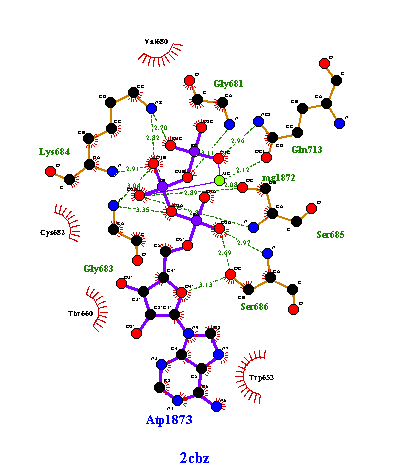
1UW1 novel designed protein ( kd = 0.1uM )

![C:\Users\hwkobe\AppData\Roaming\Tencent\Users\827462509\QQ\WinTemp\RichOle\H8(_4MASH@EKJ`](1@3YET4.png](data:image/png;base64,)

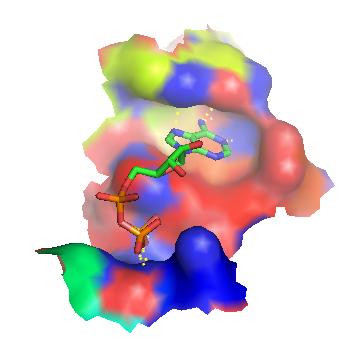
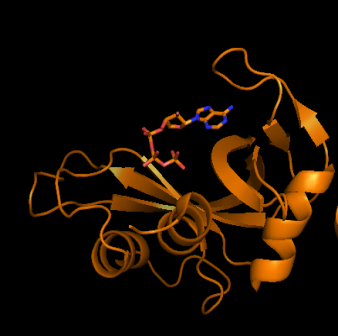
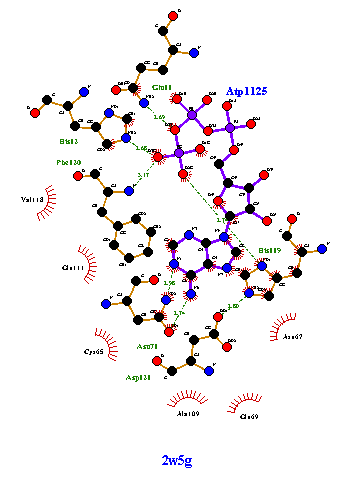
1L2T Transport Protein (kd = 0.8mM)



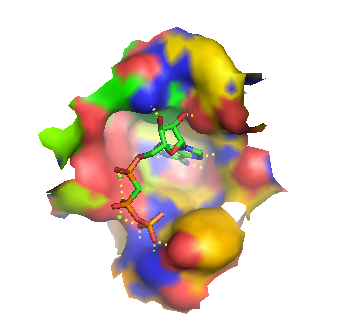
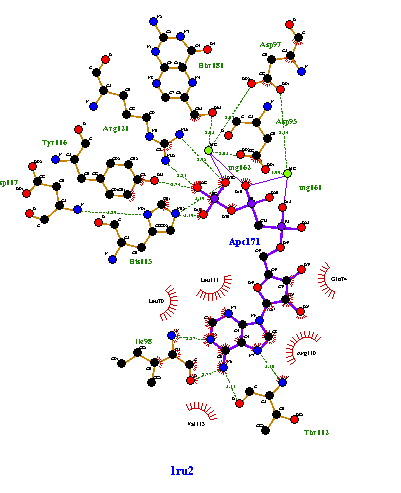
2CBZ Transport Protein (kd = 0.118mM)



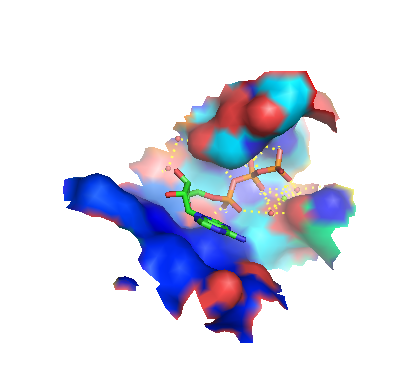
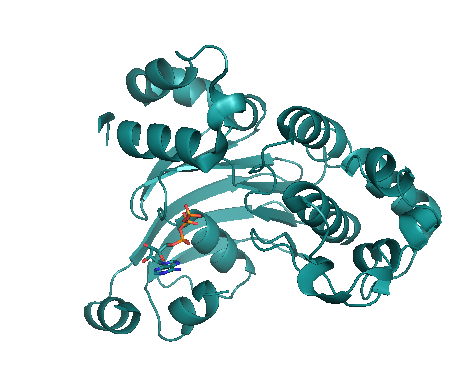
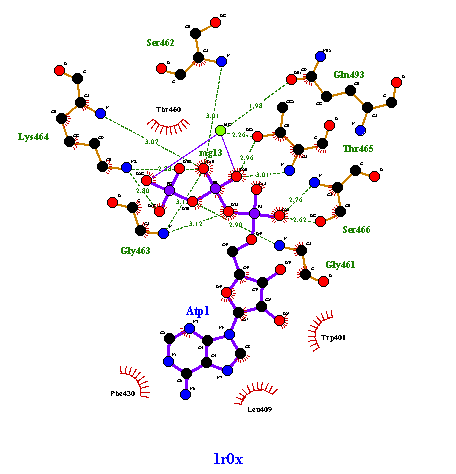
2W5G Hydrolase (ki = 0.86uM）



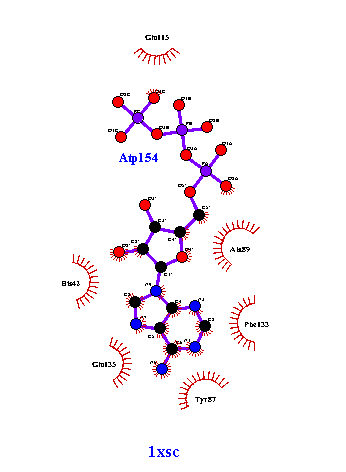
1RU2 Transferase (kd= 0.32uM)



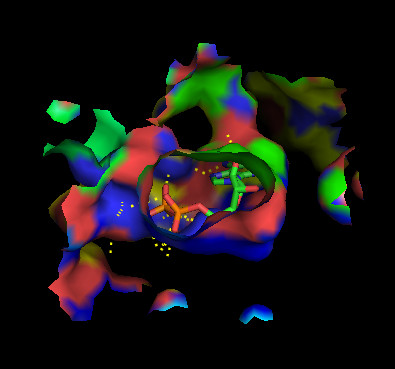
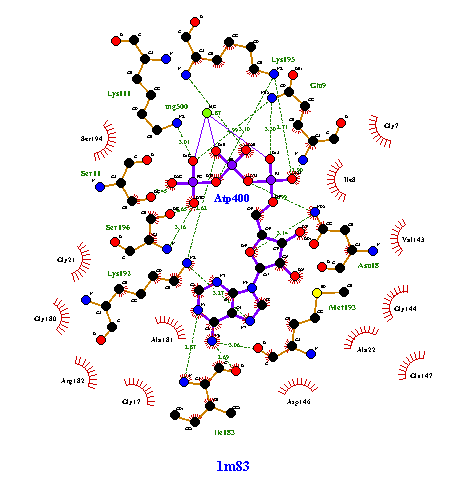
1R0X Transport Protein (kd = 0.149mM)



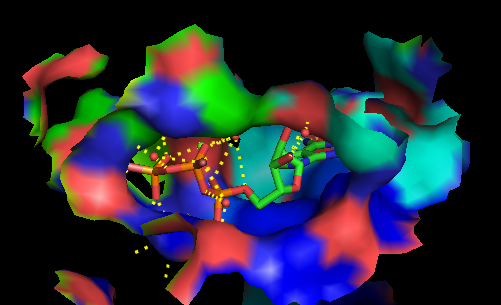
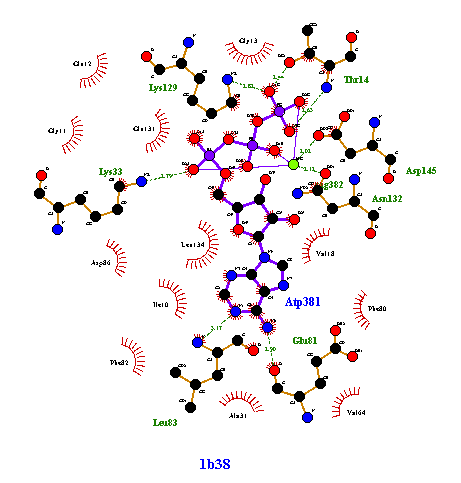
1XSC Hydrolase



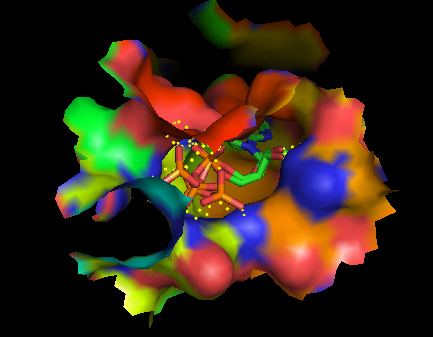
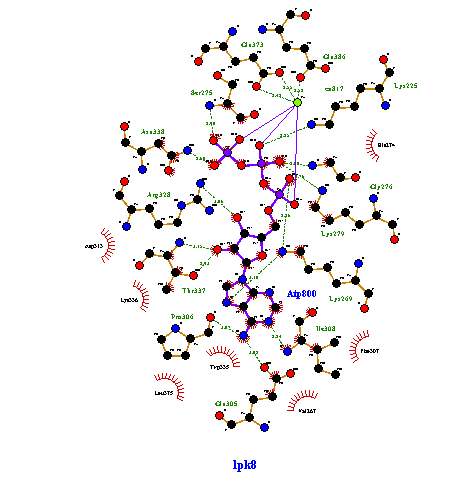
1M83 Ligase (kd=0.4mM）



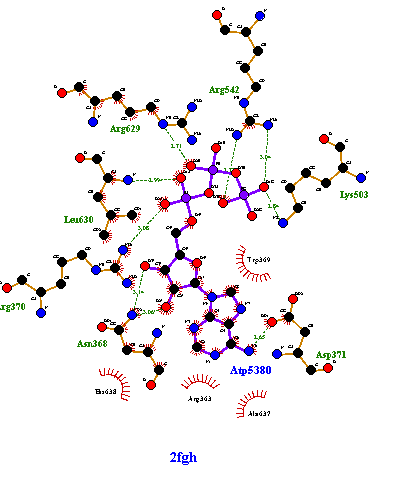
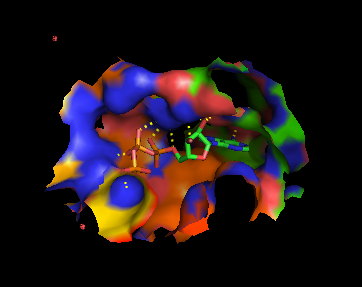
1B38 Transferase



1PK8 Membrane protein



2FGH

MLIKVKTLTGKEIEIDIEPTDKVERIKERVEEKEGIPPQQQRLIYSGKQMNDEKTAADYKILGGSVLHLVLALRGGGGLRQKHHHHHH

VLENTAPSSTNLLIRFEIDREDPPIVLVKTKNENFNPETAVKNKIYLLENKLYFIDKMGNLFNLGPGKKKCTQLFNAIGDSAEYSLCDPFVLEEPEKPEDFAISEIVDIFNEQKERFDFWI

0.00797 -7.09184 -7.77611

0.01797 -7.84856 -8.22011

0.02583 -8.11770 -8.53073

0.03276 -8.13116 -8.35849

0.03904 -7.84311 -8.05600

0.04485 -7.61563 -7.83230

0.05040 -7.63146 -7.73430

0.05563 -7.58764 -7.74186

0.06065 -7.57451 -7.78420

0.06556 -7.56254 -7.83475

0.07018 -7.71678 -7.88193

0.07478 -7.74744 -7.94132

0.07925 -7.80858 -8.00243

0.08349 -7.79035 -8.06597

0.08773 -7.85136 -8.14296

0.09194 -8.07275 -8.23475

0.09597 -8.07609 -8.32713

0.09984 -8.22449 -8.41141

0.10373 -8.26723 -8.48540

0.10768 -8.36749 -8.55959

0.11144 -8.50300 -8.62540

0.11508 -8.44849 -8.69320

0.11865 -8.74310 -8.75817

0.12232 -8.73866 -8.82037

0.12594 -8.72586 -8.87852

0.12945 -8.80487 -8.93077

0.13283 -8.90732 -8.97547

0.13622 -8.97976 -9.01328

0.13968 -9.04100 -9.04297

0.14299 -9.09834 -9.07261

0.14639 -9.14571 -9.10074

0.14969 -9.17701 -9.13132

0.15281 -9.20999 -9.16465

0.15604 -9.32436 -9.19683

0.15924 -9.25820 -9.21876

0.16240 -9.29596 -9.23653

0.16563 -9.23147 -9.25633

0.16869 -9.34132 -9.28084

0.17174 -9.46119 -9.29237

0.17467 -9.36107 -9.30674

0.17777 -9.33274 -9.32185

0.18076 -9.51492 -9.33290

0.18375 -9.45226 -9.34919

0.18673 -9.41871 -9.36292

0.18976 -9.46598 -9.37522

0.19265 -9.51669 -9.38740

0.19544 -9.56367 -9.39860

0.19831 -9.36570 -9.41399

0.20117 -9.52094 -9.43588

0.20392 -9.63256 -9.45290

0.20686 -9.62979 -9.47396

0.20968 -9.66788 -9.49742

0.21254 -9.62847 -9.52066

0.21528 -9.79447 -9.54646

0.21801 -9.72139 -9.57297

0.22075 -9.75131 -9.58779

0.22336 -9.84515 -9.60246

0.22610 -9.77377 -9.61906

0.22882 -9.87946 -9.64521

0.23160 -9.91372 -9.67923