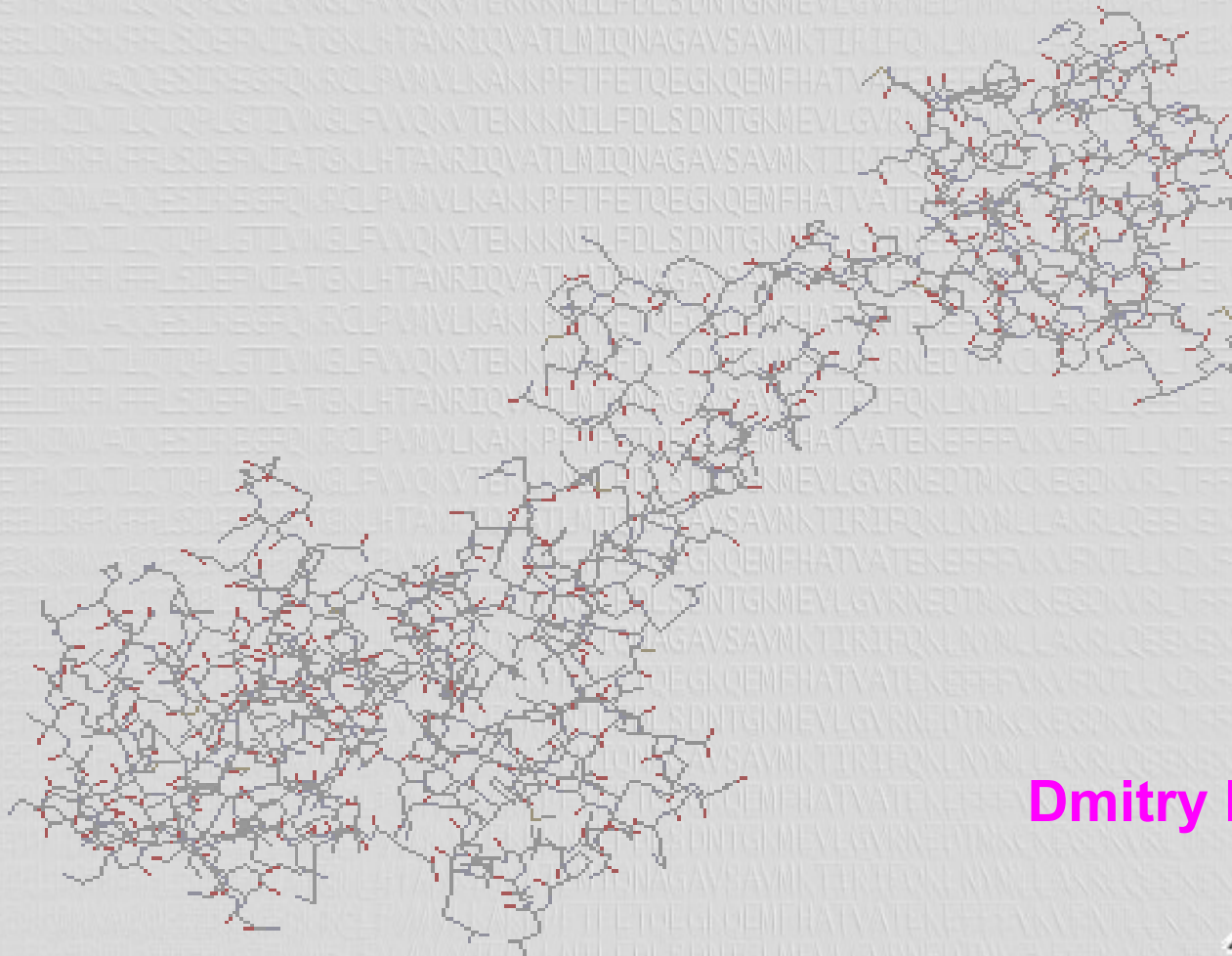


Absent In Melanoma 2 (AIM2) Homology Modeling

Structural Bioinformatics



Dmitry Repchevsky

AIM2 belongs to a class of proteins called inflammasomes, which are multi-protein complexes that play major roles as guardians against both viral and bacterial infections. Inflammasomes also detect dangerous self-molecules associated with tissue damage.

Thomas Jefferson University (2009, January 26).

Key Protein Regulator Of Inflammation And Cell Death Discovered.

sp-O14862-AIM~



(343 aa)

The DAPIN (Domain in Apoptosis and INterferon response) domain is an 80-100-residue domain which is found in the N-terminus of diverse vertebrate and vertebrate-specific viral proteins involved in apoptosis, cancer, inflammation, and immune response.

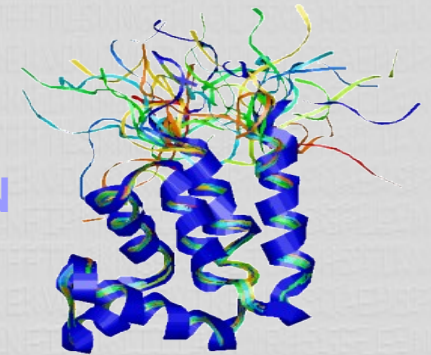
The death domain (DD) is a conserved region of about 80 residues found on death receptors, and which is required for death signalling, as well as a variety of non-apoptotic functions

HIN-200 Interferons (IFNs) are cytokines that regulate host resistance against infections by exerting antimicrobial, immunomodulatory and cell growth regulatory functions.

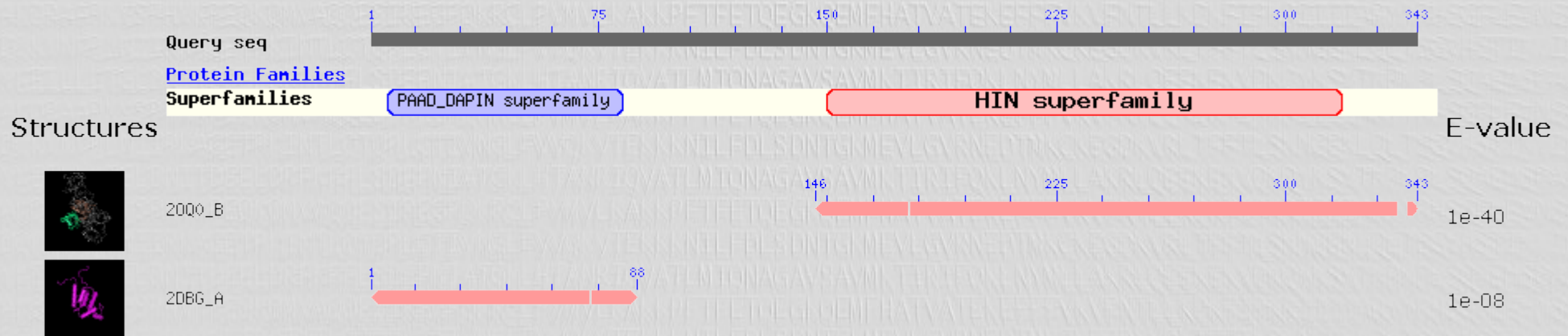
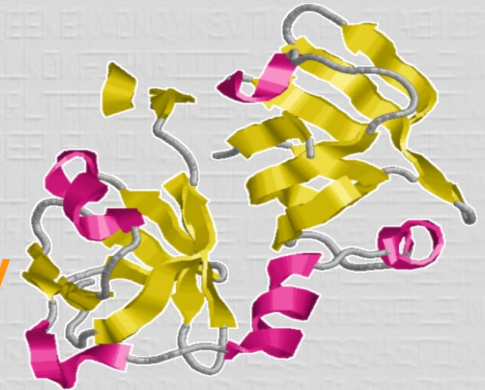
PSI-BLAST Search for homologies

Alignment	DB:ID	Source	Length	Score	Identity%	Positives%	E()
1 <input checked="" type="checkbox"/> New	PDB:2OQ0_D	mol:protein length:206 Gamma-interferon-inducible protein Ifi-16	206	795	43	59	1e-84
2 <input checked="" type="checkbox"/> New	PDB:2OQ0_C	mol:protein length:206 Gamma-interferon-inducible protein Ifi-16	206	795	43	59	1e-84
3 <input checked="" type="checkbox"/> New	PDB:2OQ0_B	mol:protein length:206 Gamma-interferon-inducible protein Ifi-16	206	795	43	59	1e-84
4 <input checked="" type="checkbox"/> New	PDB:2OQ0_A	mol:protein length:206 Gamma-interferon-inducible protein Ifi-16	206	795	43	59	1e-84
5 <input checked="" type="checkbox"/> New	PDB:3B6Y_B	mol:protein length:200 Gamma-interferon-inducible protein Ifi-16	200	620	47	64	2e-64
6 <input checked="" type="checkbox"/> New	PDB:3B6Y_A	mol:protein length:200 Gamma-interferon-inducible protein Ifi-16	200	620	47	64	2e-64
7 <input checked="" type="checkbox"/> New	PDB:2DBG_A	mol:protein length:103 Myeloid cell nuclear differentiation antigen	103	336	28	49	2e-31
8 <input checked="" type="checkbox"/> New	PDB:2YU0_A	mol:protein length:94 Interferon-activable protein 205	94	292	25	48	3e-26

PAAD-DAPIN



HIN superfamily



DAPIN DOMAIN

T-Coffee

```
014862 -----ME---SKYKEILLTGLDNITDEELDRFKFFLSD--EFNIATGKLHTANRIQVATLMIQNAGAVSAVMKTIRIFQKLN-YMLL
2DO9_A GSSGSSGMALARANSPQEALLWALNDLEENSFKTLKFHLRDVTQFHRLARGELESLSQVDLASKLISMYGAQEAVRVVSRSLAMN-LMEL
2DBG_A GSSGSSGMV---NEYKKILLKGFELMDDYHFTSIKSLLAY--DLGLTTKMQEEYNRIKITDLMEKKFQGVACLDKLIELAKDMPSLKNL
```

```
cons          *      :  :  **  .: : : :  :  : *  *      : : : : .  .: : : : :  :  .  . :  .  :  *
```

```
014862 AKRLQEEKEKVDKQYKSVTKPKPLSQAEMSPAASAAIRNDVAKQRAAPKVS PHVKPEQKQMQVAQQESIREGFQKRCLPVMVLKAKKPFTF
2DO9_A VDYLNQVC---LNDYREI-----
2DBG_A VNNLRKEKS-----
```

```
014862 ETQEGKQEMFHATVATEKEFFFVKVFNTLLKDKFIPKRIIIIIARYYRHSGFLEVNSASRVLDAESDQKVNPLNIIRKAGETPKINTLQT
2DO9_A -----
2DBG_A -----
```

```
014862 QPLGTIVNGLFVVQKVTEKKKNILFDLSDNTGKMEVLGVRNEDTMKCKEGDKVRLTFFTL SKNGEKLQLTSGVHS TIKVIKAKKKT
2DO9_A ----- YREHVSGPSSG
2DBG_A ----- --KVASGPSSG
```

```
cons          :  . . .
```

SCORE = 60

We can see that sequence alignment is not good in this particular case that is explained by not very good e-value in our psi-blast results. In fact blast doesn't find these homologues at all.

H1N DOMAIN

T-Coffee

We can see this domain is well conserved

20Q0_A -----
20Q0_B -----
20Q0_C N-----
20Q0_D -----
3B6Y_A -----
3B6Y_B -----
O14862 MESKYKEILLLTGLDNI TDEELDRFKFFLSDEFNIATGKLTANRIQVATLMIQNAGAVSAVMKTIRIFQKLN YMLLAKRLQEEKEKVDK

20Q0_A -----VLQKRPVIVKVLSTTKPFYETPE--KKIMFHA
20Q0_B -----VLQKRPVIVKVLSTTKPFYETPEMEKKIMFHA
20Q0_C -----VLQKRPVIVKVLSTTKPFYETPEMEKKIMFHA
20Q0_D -----LQKRPVIVKVLSTTKPFYETPEMEKKIMFHA
3B6Y_A -----D----LKEVMVLNATESFVYEPKE--QKKMFHA
3B6Y_B -----LKEVMVLNATESFVYEPKE--QKKMFHA
O14862 QYKSVTKPKPLSQAEMSPAASAAIRNDVAKQRAAPKVS PHVKPEQKQMV AQQESIREGFQKRC L PVMVLKAKKPFTFETQEG-KQEMFHA

cons * ** * : * . * :: ****

20Q0_A TVATQTQFFHV KVLNTSLKEKFN GKKIIII SDYLEYDLSLEVNEESTVSEAGPNQTFEVPNKI INRAKETLKIDILHKQASGNIVYGVFM
20Q0_B TVATQTQFFHV KVLNTSLKEKFN GKKIIII SDYLEYDLSLEVNEESTVSEAGPNQTFEVPNKI INRAKETLKIDILHKQASGNIVYGVFM
20Q0_C TVATQTQFFHV KVLNTSLKEKFN GKKIIII SDYLEYDLSLEVNEESTVSEAGPNQTFEVPNKI INRAKETLKIDILHKQASGNIVYGVFM
20Q0_D TVATQTQFFHV KVLNTSLKEKFN GKKIIII SDYLEYDLSLEVNEESTV-----SFEVPNKI INRAKETLKIDILHKQASGNIVYGVFM
3B6Y_A TVATENEVFRVKVFENIDLKEKFTP KKI IAIANYVCRNGFLEVYPFTLVADVNADR NMEI PKGLIRSASVTPKINQLCSQTKGSFVNGVFE
3B6Y_B TVATENEVFRVKVFENIDLKEKFTP KKI IAIANYVCRNGFLEVYPFTLVADVNADR NMEI PKGLIRSASVTPKINQLCSQTKGSFVNGVFE
O14862 TVATEKEFFV KVFENTLLKDKFI PKRIII IARYYRHSGFLEVNSASRVLDAESDQKVN VPLNI IRKAGETPKINTLQ TQPLGTIVNGLVF

cons ***** . . . * ** : * ** : * * : * * . . . : * : * . * * ** : * . * . * : * * :

20Q0_A LHKKT VNQKTTIYEIQDDR GKMDVVG TGQCHNIPCEEGDKLQLFCFRLR KKNQMSKLISEMHSFIQI-----K
20Q0_B LHKKT VNQKTTIYEIQDDR GKMDVVG TGQCHNIPCEEGDKLQLFCFRLR KKNQMSKLISEMHSFIQIK-----K
20Q0_C LHKKT VNQKTTIYEIQDDR GKMDVVG TGQCHNIPCEEGDKLQLFCFRLR KKNQMSKLISEMHSFIQI-----K
20Q0_D LHKKT VNQKTTIYEIQDDR GKMDVVG TGQCHNIPCEEGDKLQLFCFRLR KKNQMSKLISEMHSFIQIK-----K
3B6Y_A VHKK NVRGEFTYYEIQDNTGKMEVVVHGRLTTINCEE GDKLKLTCFELAPKSGT GELRSVIHSHIKV-----I
3B6Y_B VHKK NVRGEFTYYEIQDNTGKMEVVVHGRLTTINCEE GDKLKLTCFELAPKSGT GELRSVIHSHIKV-----I
O14862 VQKVTEKKNILFDLSDNTGKMEV LGVRNEDTMKCKEGDKVRLTFFFTLSKNGEKLQLTSGVHSTIKVIKAKKKT

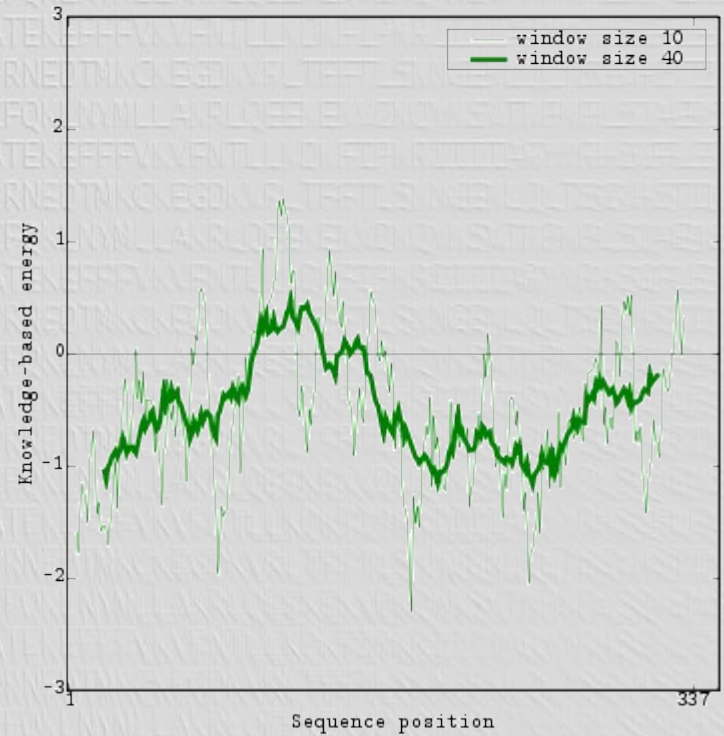
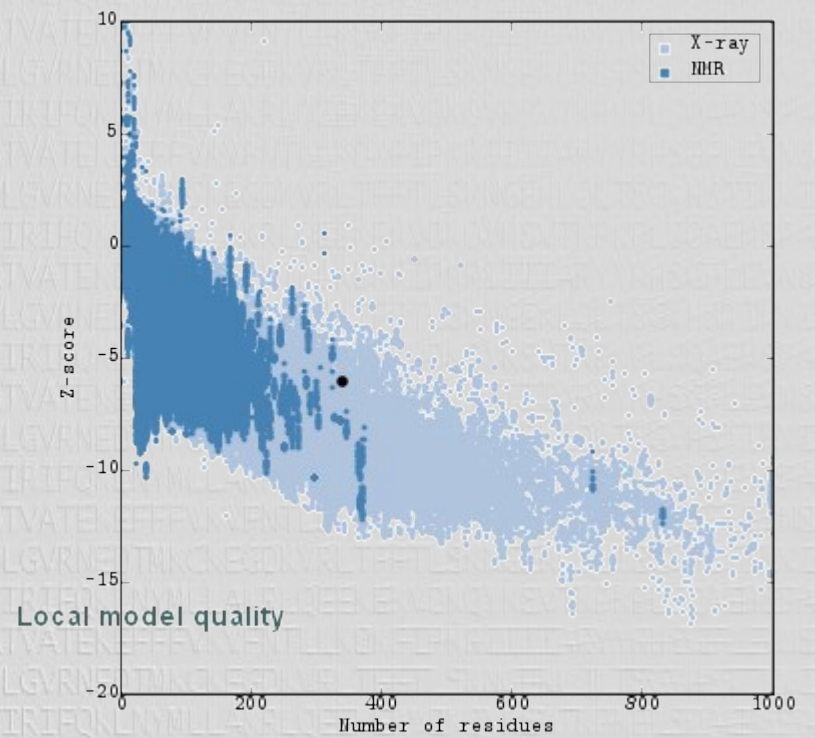
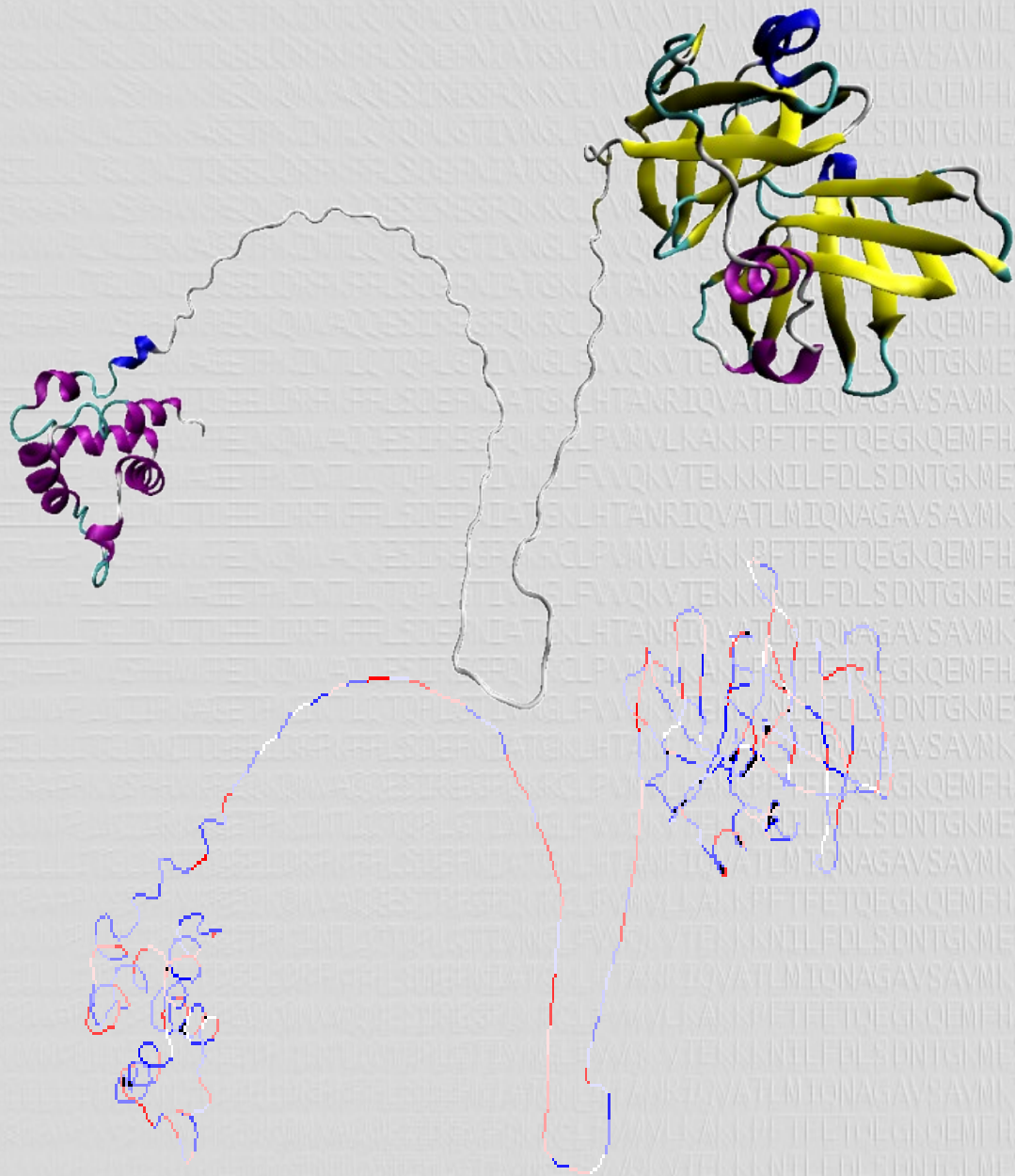
cons : * . . . : : . . * : * * * : * . . . : * * : * * * :

So after a quick look in the studied protein we can see that it consists in two domains. There is no homologue that cover all the protein. Moreover intent to align the sequence with all selected homologues failed (clastalw / t-coffee).
The decision to model domains separately using structural alignment is taken.

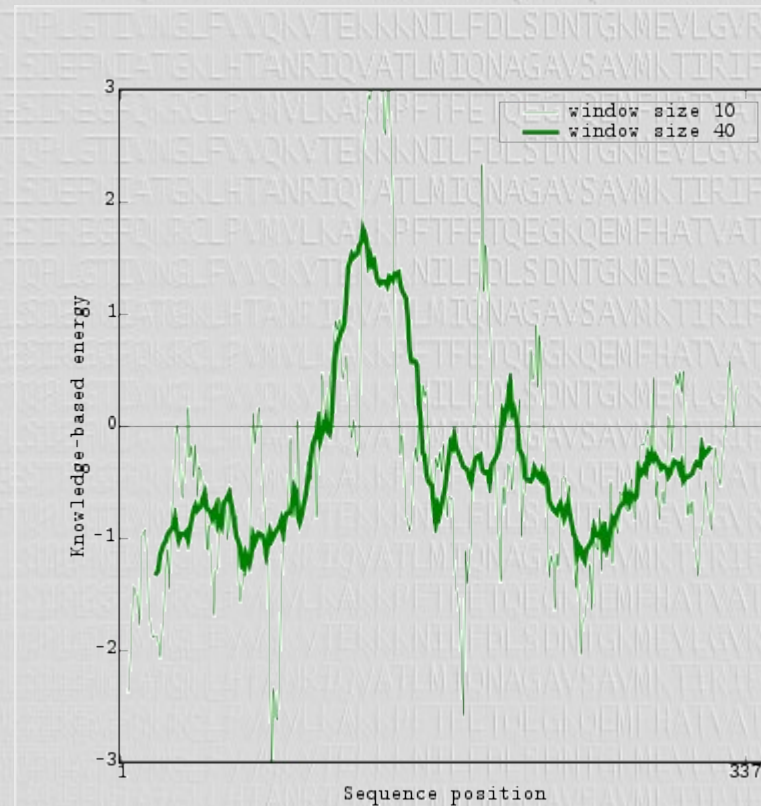
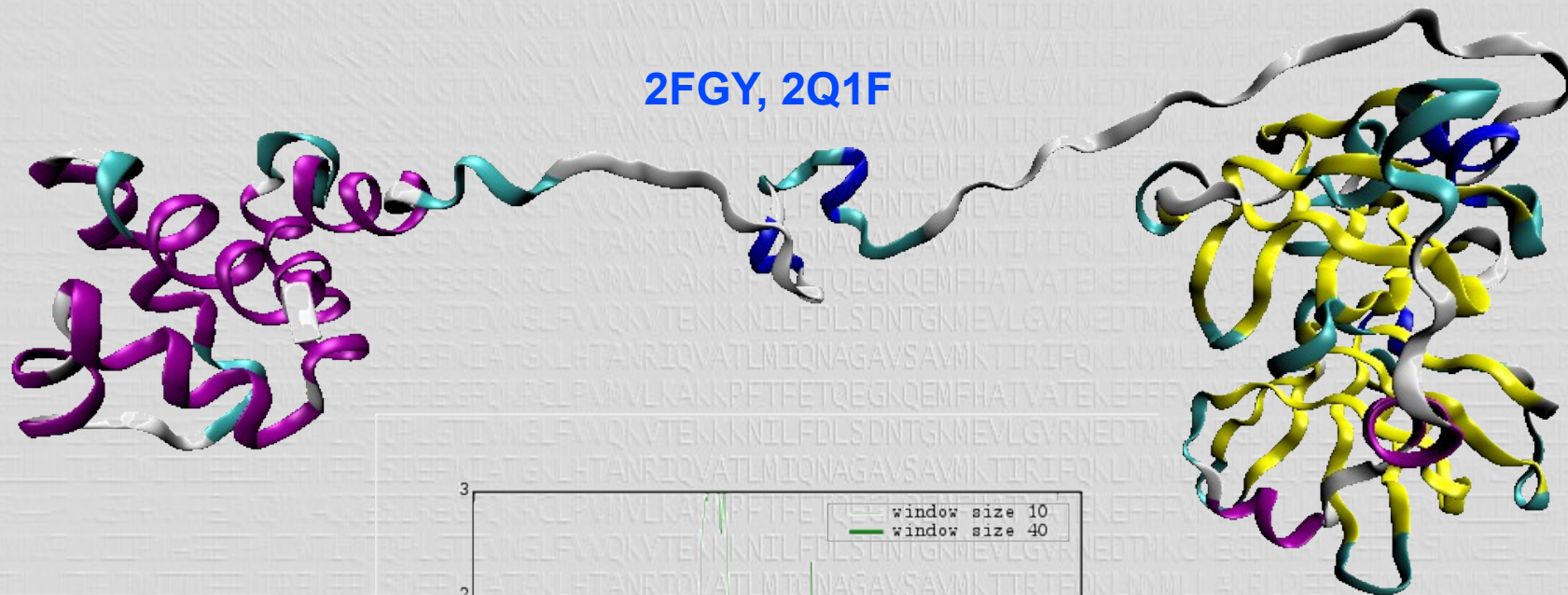
Both HIN and DAPIN domains have been separately aligned using SAP
 Then HMM profiles have been built to align AIM2 (O14862) target sequence

O14862/1-284	1	-----ME-----SKYKEILLLTGLDNITDEELDRFKFFLS-----DEFNIATGKLHTANRIQVATLMIQNAGAVS	60
2OQ0_A/1-189		-----	
3B6Y_A/1-185		-----	
2DO9/1-115	1	GSSGSSGMALARANSPQEALL-WALNDLEENSFKTLKFHLR-DVTQFHLARGELESLSQVDLASKLISMYGAQE	72
2DBG/1-103	1	GSSGSSGMV-----NEYKKILLLLKGFELMDDYHFTSIKSLLAYDLGLTTKMQEEYNRIKITDLMEKKFQGVACLD	70
O14862/1-284	61	AVMKTIRIFQKLN YMLLAKRLQ-----EEKEKVDKQYKSV-----	95
2OQ0_A/1-189		-----V	1
3B6Y_A/1-185		-----D	1
2DO9/1-115	73	AVRVVSRSLLAMNLMELVDYLN--QVCLNDYREIYREHVSGPS SG-----	115
2DBG/1-103	71	-----KLI ELAKDMP SLKNLVN NLRKEKSKVASGPS SG-----	103
O14862/1-284	96	FQKRCLPVMVLKAKKPF TFE TQE -GKQEMFHATVATEKEFF FVKVFN TLLKDKFIPKR I I I IARYYRHS GFLEEV	168
2OQ0_A/1-189	2	LQKRPIVIVKVLSTTKPF EYETPE--KKIMFHATVATQTQFFHV KVLN TSLKEKFN GKK I I I I SDYLEYDSLLEEV	73
3B6Y_A/1-185	2	LK-----EVMVLNATESFVYEPKE--QKKMFHATVATENEVFRVKVFNIDLKEKFTPKK I I A I ANYVCRN GFLEEV	69
2DO9/1-115		-----	
2DBG/1-103		-----	
O14862/1-284	169	NSASRVLD AESDQKVNVP LN I IRKAGETPKINTLQTQPLGTIVNGLFVVQKVTEKKN I LFDLS DNTGKMEVVG	242
2OQ0_A/1-189	74	NEESTVSEAGPNQTFEVPNK I INRAKETL KIDILHKQASGNIVYGVFMLHKKTVNQKTTIYEIQDDR GKMDVVG	147
3B6Y_A/1-185	70	YPFTLVADV NADR NME I PKGLIR SASVTPKINQLCSQTKGSFVNGVFEVHKKNV RGEFTYYEIQDNTGKMEVVV	143
2DO9/1-115		-----	
2DBG/1-103		-----	
O14862/1-284	243	VRNEDTMKCKEGDKVRLTFF T LSKNGEKLQLTSGVHSTIKV- I	284
2OQ0_A/1-189	148	TGQCHNI PCEEGDKLQLFCFRLRKKNQMSKLI SEMHSFIQI-K	189
3B6Y_A/1-185	144	HGRLTTINCEEGDKLKLTCFELAPKSGTGELRSVIHSHIKV- I	185
2DO9/1-115		-----	
2DBG/1-103		-----	

Modeller v9.5 built model

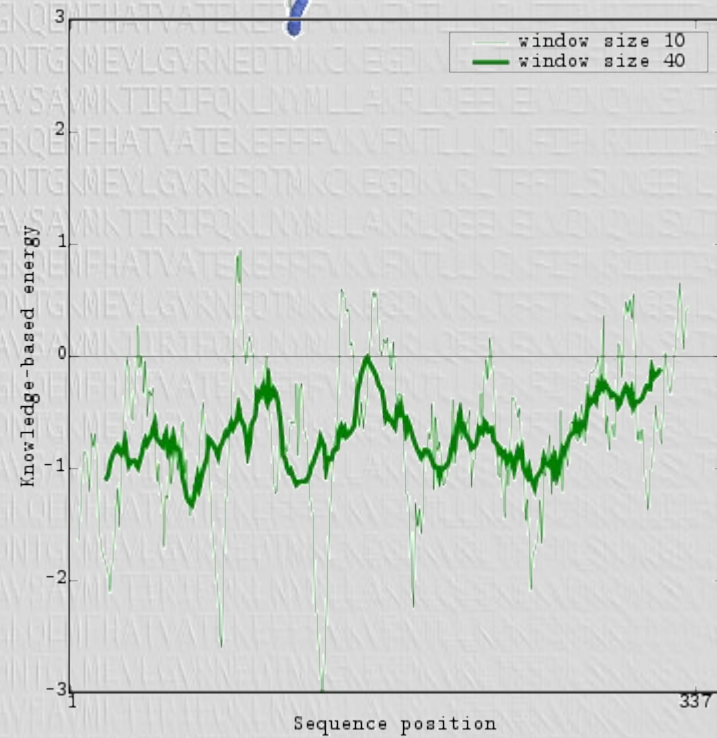
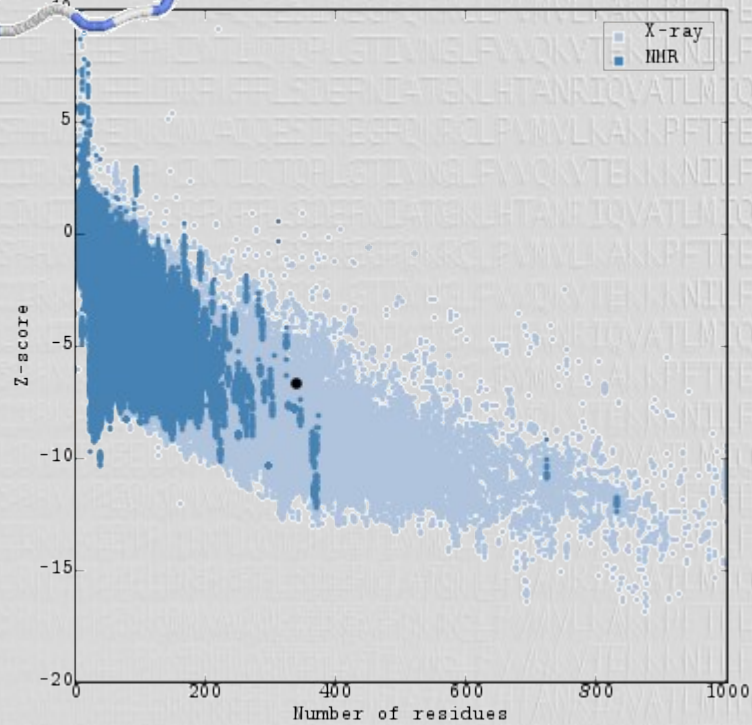
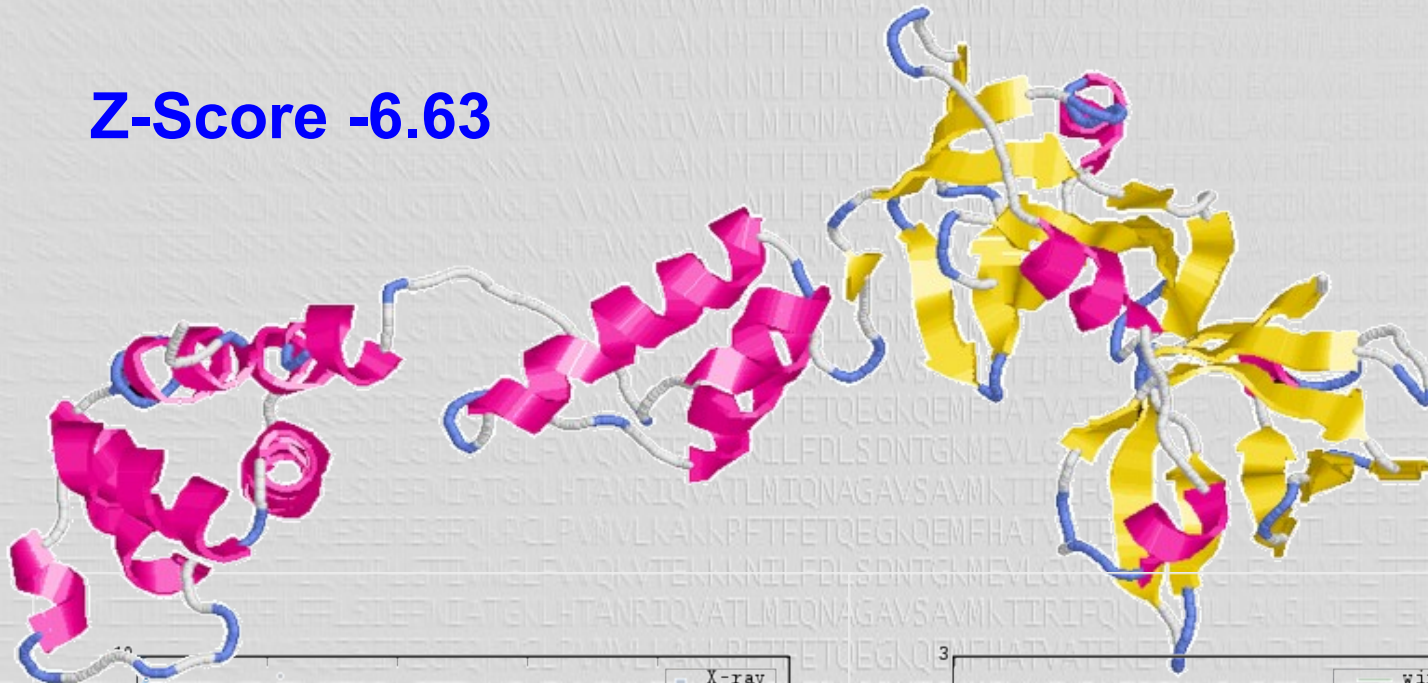


Try to remodel missed part using VERY remote proteins ... failed...



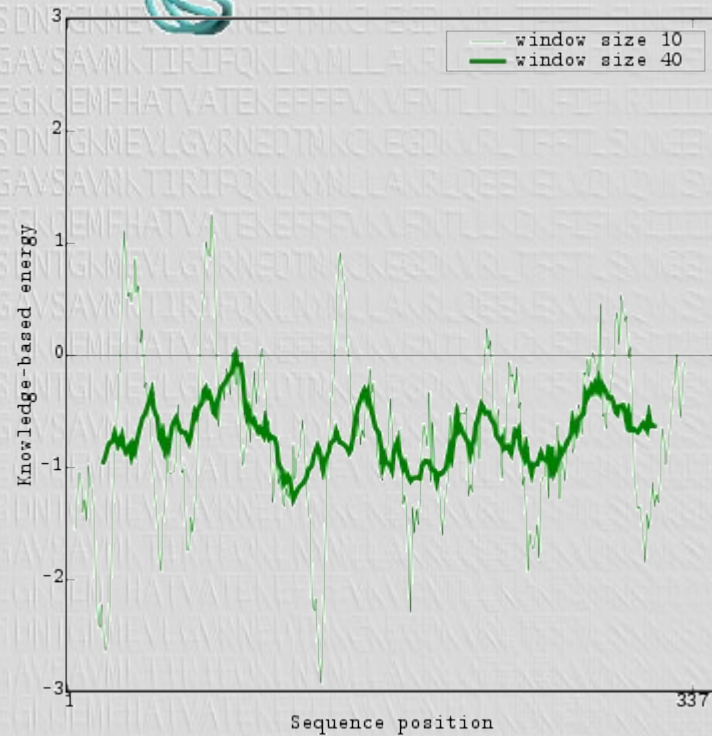
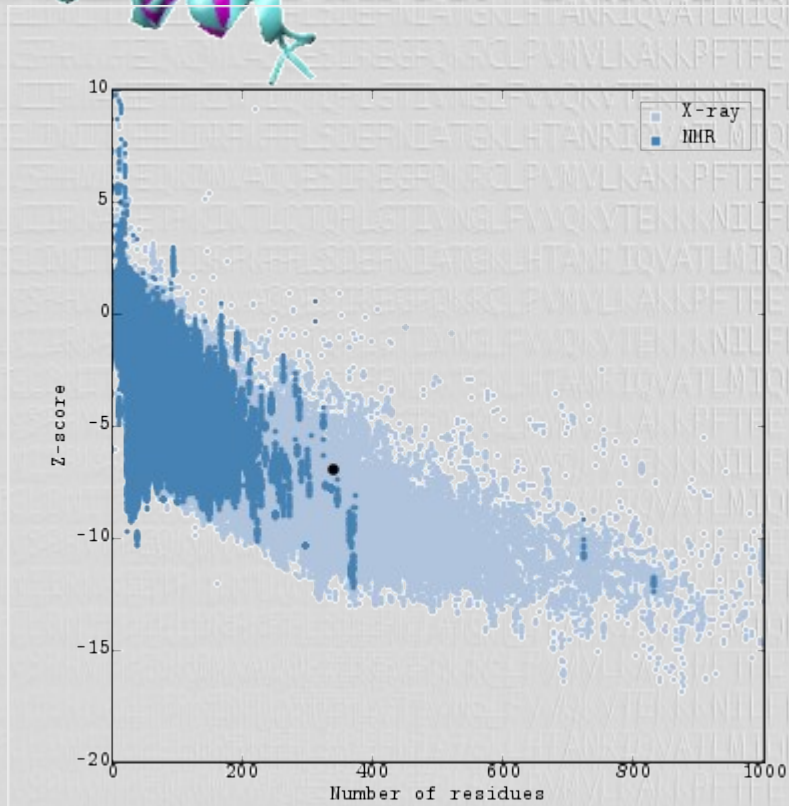
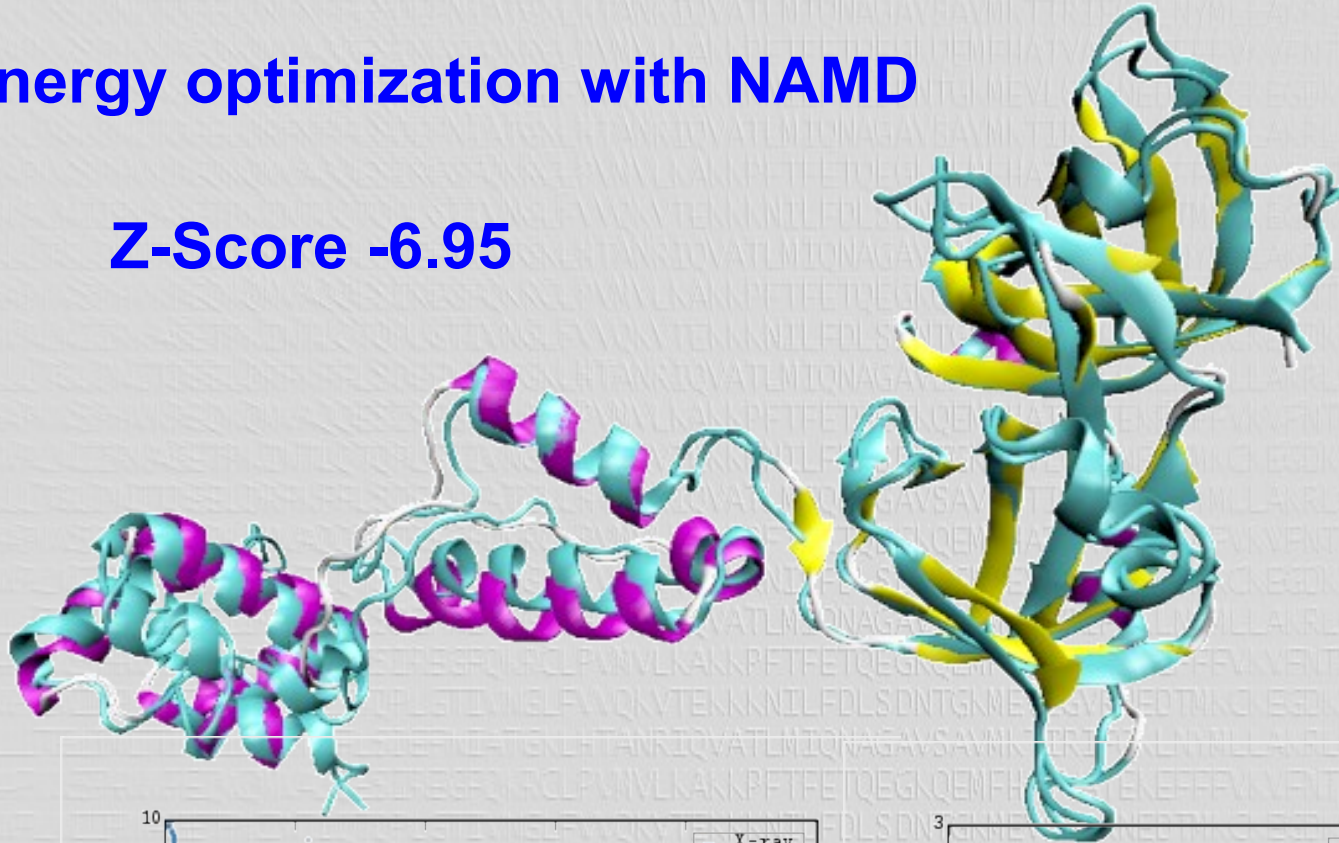
Try to remodel missed part using SAM_T08

Z-Score -6.63



Energy optimization with NAMD

Z-Score -6.95



THANK YOU

Big thanks to my colleagues Laia Codó and Romina Royo for their advices, support when I almos gave up with this witchcraft...