SynthImmunol_NMU Notebook No2

Time: 2024.08.29-2024.09.28

---

1. **Experiment:** RFDiffusion-Based Protein Design
2. **Time:** 2024.08.29-2024.09.28
3. **Member:** Xudong Tang, Yang Jin, Binxuan, Zhang, Kaiqing Zhang, Xuantong Liu
4. **Method:**

RFDiffusion is a cutting-edge method for de novo protein structure and function design, leveraging the capabilities of deep learning and diffusion models. It is based on the RoseTTAFold (RF) structure prediction network, fine-tuned on protein structure denoising tasks. The principle behind RFDiffusion is to generate protein backbones by iteratively refining noise-corrupted structures, eventually producing functional and structurally accurate proteins from simple molecular specifications.

(1) Core Components:

① Denoising Diffusion Probabilistic Models (DDPMs): These models are trained to reverse the process of adding Gaussian noise to protein structures, thereby generating new, realistic protein structures.

② Rotational Equivariance: RFDiffusion maintains rotational equivariance, allowing it to model three-dimensional (3D) structures in a global representation frame independent manner.

③ Conditioning Information: The model can be guided towards specific design objectives by providing conditioning information at each step of the generation process.

(2) The procedure of RFDiffusion:

① Model Training
- Data Preparation: Sample protein structures from the Protein Data Bank (PDB) and introduce noise to create training inputs.
- Noise Application: Perturb Cα coordinates with 3D Gaussian noise and apply Brownian motion to residue orientations.
- Model Training: Train the RFDiffusion model by minimizing the mean-squared error (MSE) loss between frame predictions and the true protein structure.

② Protein Design
- Initialization: Start with random residue frames.
- Denoising Iterations: Iteratively refine the protein structure by denoising the noisy input, adding noise at each step to generate the input for the next iteration.
- Sequence Design: Use the ProteinMPNN network to design sequences encoding the generated protein structures.

③ Conditioning for Specific Designs
- Unconditional Design: Generate diverse protein structures without additional input.
- Topology-Constrained Design: Provide secondary structure and/or fold information to guide the design towards specific topologies.
- Symmetric Oligomer Design: Specify point group symmetry to create symmetric oligomeric structures.

④ Experimental Characterization
- Expression and Purification: Express the designed proteins in a suitable host and purify them for further analysis.
- Structural Verification: Use techniques such as circular dichroism (CD) and cryo-electron microscopy (cryo-EM) to verify the structure and stability of the designed proteins.
- Functional Validation: Assess the functionality of the designed proteins through binding assays, enzymatic activity tests, or other relevant functional assays.

5. Result:

Following the identification of active sites on bait proteins, RFDiffusion was employed to design binding proteins targeting these active site regions, with the designed constructs set to 45-65 residues in length. We generated 100 such binding proteins for molecular docking to identify optimal binders for the bait proteins.

Table.1 The Sequence of IL-2β Mimics

β Complex	**Amino Acid Sequence**
B0	MEEKLEELKKKLAELDGKYIYEKCYGTEEEAKKALEELKAALEELAKAEKEAAAAAA
B1	SEEERRREEEEERERREEEEERRKEERLEKMRERDEEIREEEEEEEEEEEEEEEEEEEEEEEEEEE
B2	SLEEALRAELRRRAAEECGALLREAERAAAAFRAATPSEEEAAAFLEAARAEARAACEARFAAL
B3	EELEKKLKELKEKAEEAREKAKEYSAKAVKYLADPSKKEEAEECLKKCEELIKEERKYIEEAKKL
B4	MLEELELKKKLKEKLEKEEEERKKRWEARREEAKAAALAARKAEEEARAAAEAAAAAA
B5	AEELARRAAEFLARAKELDLEMAKKIEEVRKKTGNEEETEKARKELLEELRKEIKRLKEEL
B6	EEEERKKKIEEYKTKAEAHKLDAEQLEAKAAAASPEAAKLYKKLAEKEKELAAEYEKKAKELEEA
B7	SLLEELKRKLECEKRANECSEKLKKKREEEKEKKKKLEEEEEKLEEEEEELEEEEEEEEEEE
B8	SLEKILEELKKKEEEKKLEESLKKLEEEAKKFEEERKAAAEAAAKAAAE
B9	ELEEKAKKEAKEKEREERSKRLREERERRLEEEERRRREEEERRERERRRREEEEA
B10	SLEELLKRLRELKREECLARMRAKAEEERAAEAAAAAAAAAAAEELARRLAELA
B11	SLAAEKAAKIAALTAEAEKKAKELLAKAAAASPEERKKYMEEAEKYLAEIRAEIAAIL
B12	SLEELLKELEEKKKKEEEEKLKKAKENWEKYQKELAERLAAALAALAAAAAALAAALAEAAAAAA
B13	SELKELEEKRKEEEEKLKEETEKRLAEEERRFLEERARREAERRAEEERRRREELERRR
B14	MSERIKTLREALELVRQGVENPATQAELIARGRALAEAATGEAGRALFERELARLEAQKAA
B15	EEEVKALLEELAKLEYEYLKAAKEDKKLAEEYLKKAEEAEKKLLEAKKALEEKKAKE
B16	EEEERKAAEEAARKAAHLKEATERFRERRRRREEEEREREEREEEERREEEEEEEEREREE
B17	AAAAAAREAALKARRAAGDEDAARDRAACEALYAEDPAKGAECLAKVEAEEKAFRAEIDAALAA
B18	SLEEELRREEEEERRREEEERLEELWKKNKEKAEELAKKREKELEEKEEKERKEKLEELKKELE
B19	MSEEELKKLLEEQKKKIEQLRKEGEAKAAALRAAARAAAAALAAAAAAA
B20	MTLEEKLANLQAGKAASLAALEALAAEAAAESPEKAALVRELARRVRAQYDKEIAAVAAELA
B21	EEKEKEEKKKKEEEEKKKKEREEKEKEYIEKVKKEVKEREEKEKAEEEAKKKA
B22	SLEEVKATYEESIEAAKKLGAERVAKAAAEGPEAAEKAAALSKEAIELLEKKKEEELAKL
B23	MEEKEKKEKEKEELKKKAEEAEKKAKELKEKMAKSSAEEAEKLAKEREELEKKRVELKKELE
B24	SLAAELAKKKEREKALEERKKKAEEIKEEEEERKRKAAEAAAAAAAAAAAAAAAAAAAA
B25	AALAALLAALAAAAAAAAAAAAAARAAEAEKRAAELRRRREEELARRLAA
B26	SLEEEKKRKEEEEKRKKRLEEARREWEERLEARRRAEEEERLEEERRR
B27	SLLEELRRALEEREELERELEEAKKRYEEALKKLKEEKEKEEEERKRKEEE
B28	MEEERKEEERKREEEKERERRRAAGRAAREAAAAAAAAAAAAARAEAERRR
B29	SAAAKAELAALEKRAKEAAALAEEAAKKDKKKAKRYKADRDILEKEAKALKAAL
B30	AAAAEAAAAAAALRARQAEREARAREMAAAIAAADGEEGKRKAALLLHYAAVVRARVEAEVAA
B31	SEEERERAEAAARAAAEAARAAARAKAKERYKEELEEIKKKREEEEKKKKEEELK
B32	SLEALAAAAAAAAAAAAAAAAAEAAALEEEYRKRLEEEAEELEEELEEEEEEEEEEEEELE
B33	AEEEERERAEAAAAAAAEAEARLARMRAEDERIRAAKREAAARAAAEAAAAEERRRAE
B34	AAEAARRAARAAFDARLTAAERKYLAAQDDPEAAAAWLAEIAAIEAERTAAERAWAA
B35	AAAAAAAEIAEAARKAEERYKELEKEAEEALKKDKEEGTKKRKEALEESLKLAKELLELRKRLEA
B36	SLAEALAAAAAAAAAAAAEAARRARIEADIAEARRRLEEEEREKEEEEKR
B37	SLLEELLRRLREEEEEERRREKSREEGRRRREAERAALAAAAEAAAAAEAAAREAAEREAA
B38	EEEKRKAELLKQIEEDLKKAEEALALGAATPKHDYYEGLAKSYLARAEELKAILK
B39	EEERREENRKRRKERAEKEIKESKEEEEARRRKEEEEEERRRREEER
B40	EEEEKRRLEEERERAKREAEERAERIKAEIEAERARRAALEALRRELEEL

B41	EEEKREKELKELQEEAERLAKELLEHALPKLLAALLAAAAALAAAAALLAALA
B42	SLLDALLEALFRALVTAHFQAAAAATDRETAVEQARAFAAALRAALAAAAAAAAAALAAALA
B43	SELEEELRRLEELEEESRRQDEEARRRIREREAEEEARKRLEELLRKLLES
B44	SKEEEEKLKKEKRKEELEKRLKEAEEIWKKYIELLKKAKENEENKEKYLEEAKKLKEELEK
B45	AALREALEAEEAACRAALAALIAATRAAPRAERIAALTALVKECAARRAAVLA
B46	AAEIEALLAAADAEIAELRAAGVAEARAAPREERPAIVLETDRRALELRAAARAAAAALAAALA
B47	EEEERRRLEEERRAELERRLAEAEREREEELRRREEEERERЕELERELREEEERRRREA
B48	EEEEKRKLEEERKKAEEKSKELIEKRKKEREERKRRREEEEEREEEEEEEEEEEEELEEELEE
B49	SLLEELERLAAKAAEAARKAARLAAGEAALKARLAAEAAAREAEEERRRREEE
B50	AAAAEAAAAAAAAAAAAAAARAAREAEAAAEARAIAAARKAAEEAAAAAAAAAAAAAAAEAAAAAA
B51	EEERRRQIEALKRAAAAAEYEYALAKELAAKDPAYAPLAEALKAELERLKAELAALEAA
B52	SLEDIEKKIAELKRRLKLFEESCKKIAEERLKKDPEKGKKHKEELEKLNEELRKKVEAEIAELLA
B53	PLLAALRELADRLHREAVRERERARREAERAARAAAAAEAARAAAAA
B54	MSPEQKELQAQRDKYDDEALKLNELALKDPEKAEEYNAQAKKYIEKAYEIRKQIEA
B55	MEAVKKLEAKGEEYFALGKANPEKREEYERLGKLYFESAKLAKERLEKAKAEKLA
B56	EEEEEEELRRLEEEREELEKRREERIAARAAEEAARRAAAAAAAAAAAAAAAAAAAAAAAAAA
B57	AEEERREEEERREEEERRRREEEERLEEERRRRIEEATKERREKLKKEKEEKKKEEEEEKKEKE
B58	ALLEALREAAAAEAARLAALEAENRAKYAALTAALKALCAALAELA
B59	SLLEALKALLEALKALVEKAKKEAEEEKKKAEEKKKKEEEEKAKKELEELK
B60	EEEAARAAAAAAAAAAAAAAAAIAARAAEDAKRERERRKAEKERKELEEELRKK
B61	SEEEERRRRREEEEEERRRRSEEAGAEARAREKAELERRELEELLREL
B62	EEEERERERREEEEERERLEKLRREDEEIMEKLREERRREEERREREELERLR
B63	EEEEEELERERERRRAAEEAAAFRAEAAARRAALEAARAAAAAAAAAAAA
B64	SEELRRRLEELAAAAAAAAAAEAAARAARKAESEKIAAERREKAKKEK
B65	KELEEEKAKREKEREKLLAEAKAEGEKRLKAEEEARRREEEEERE
B66	SEEEREREEEEERARRAAEHREAVRRRAREELERRRAAAAAAAAAAEAARAAA
B67	LPLLLLLLLLLALALLAALLAALAEAQAAREKKLREESEKYYEELRKKKEEERRKEEEE
B68	SEKEALIAALKEKAAKALAELEEKAKEDPEAAAAAAAALKAATERLVARIKA
B69	KEKEEEEKKKEKEKAEEEERKRKAEELKARAAALAAERAAAAAAAAAAA
B70	SAAELAAAAAAEAAARAKALAEAMRAEEEAEREREAEEEARRRREEEERLR
B71	SLLAALAALLAAAAAAAAAAAAAAAGEARSRALHEEKQKKVLEEKLEEEKEKLEEELEKLLE
B72	SLEELLEEERRRAAEEEEERERRRREDIEEAREEALKRAEERKKKAE
B73	SPADALWEAELAAIREAVVAVGAEASKLSPEEAARRWAAATAEAAARMAAAKARRDALLAAA
B74	MEEEEEKKKEEEKAKKEKEKEIEEIKKKGLESAKKSAAEASLAASLAYCLAAAAAAAA
B75	EEELERRREEAERRIAELGRACLRAPEAERPACRAALRAEDRRLREELRREEEAR
B76	SLEELLRALAAAAAEAAEAARREAEARARGAALRAEREARRAEEERRRAEEEAARRAEEEAA
B77	AALAAALAALAAAAAAAAAAAAAAARAAEGEARLAAHRAEYEALLAARAAEAAAAA
B78	DAHLARARAEAAALAAELAAAGKAACGAAAAAALLAAAKALAAANAAAAAAA
B79	AELEEKAAKAAAEEAAKAAALEAELKKKSEEFAKKRREEKKKEEEEKK
B80	EEELQEYRERLQALAEEAAKKGWSPEEVKEAAKKLQEELEQERKER
B81	MEEEKKKKEEEEKKRLEELKKEIAKATDEAERAKKNASDPANKAKMAEAKAKKEAAEKELK
B82	SEEEERRRRLAEELAAQRAREEALKRESEALAEARRRLRELAAALAALLAELA
B83	AEEEAERAAVEREIKAKTNELDAKCTAAANEVAKTEGPEAAEKVRAECEAQRAAEAAAIRAAA
B84	SLLEELLRRLLEELERRRRLEEMEREREEAEARAAAQLAAEAAALA
B85	MAEEAAAAAAAAAALRAEAAAAEAAWAEERRAAAARERERAEAERRRRREEELLE
B86	EEEEERRRRREEEARREAELRAWRERIERETAELRARNEEERRRRREEEEEERRRRE
B87	GLAALAAEALRRALEAEFARDRARLERLRDHYAGDPEKAARVDAALARLEAQRAEALA
B88	SLLAALAELAAAAAAAAAEALRKAVREAAEEERKRRREEEERERREEEEERERE
B89	SLEELLEELLEEARRAAEELERLRREREEEDERERERLLRERREEEERR

B90	AAKEAEEARLARANAIYAAARAARAAREAAAAAAAAAAAAAAAAAAAAAAAAA
B91	AEAALAEAAAAAAAAAAAAAEEERRRRAAEDIERAKKEREEEKREKEEEEERKRR
B92	NEEKLKELEEKAKEYKAKYEAAKAKAAEEKANGSPEEAAAWQREANLYLGKYLVAEKKAKELKEK
B93	SREELEREAREAAEAARRAAEEARRAALGAAIEAARAAAEAAAAAAAAAAAEAEAARAAA
B94	EEEEEELRREEEAEEEERREEEERRRLLEERLREGEEEVEREREEEELREREEELEELERR
B95	SLKEEAERLKEEEENLRKAAEAYEAAGNKEKAKEAKEKAEEAKKKAEEAKKKLEEA
B96	EEEEEEERRRREEEERRREEAERRAREEALREERRRRSEEEARRRREERRRREEREEEEERRR
B97	EEEEEEERRRREEEERKEKEREEAEKEAKELREKADAEEAARAAAREAAAAAAEELRRALE
B98	VKEKAEKEIEELLKEARAVLKEAAATAAAADPATAAAARAEAAKRLAELAKKIREVKKKMKEELA
B99	EAEELAAAALAAAAAAAAAAAAAAAARAARREAIDAAARAAREAAAAAAAAAAAAAAAAAA

Table2 The Sequence of IL-2γ Mimics

γ Complex	**Amino Acid Sequence**
G0	ALEEEERRKAEEEAWLAEVKAKRAELTAAAEAARAAGDSEAADAAREKIRALVEEAIKRDRE
G1	AAAEAAAREAAAAAALAARRAAALEARARARREAEEREEREKEEREREEEER
G2	EEEEKEFEEIRKKTQEMQEKIRELQRLEWEAKKNGSKEKAEELRKKREEVLKELEELRKKRS
G3	DELEKEIEETEELLKKAKEEQAKTGKTEEYGKLIAELQARLEELKAALAAAKAAEAAA
G4	AEVERKVAELKALNAECAARIAAAAAEGPEAAAAAERECNEELRRLVE
G5	MTDKEKLMKELKAKAEEYKAAAAAAAAAGGDPECVATQQARADKYNELLAKLEAS
G6	EEKKREEERKEIQEYVEIADKAGKGEEFLKLLSENKESKEKAKKLAEEMKKKLKEEEEKKKKA
G7	STVEELEANIAANRAAAALTAAAFPETAAAARAEAAAREARNQALIAKEKA
G8	EEEERAKKLEEELKELEKLVEKIKEEGKKDPKAAELAEKLKKELEKLKAE
G9	SAAAKEAAKALATALRLAGTRLFTAGAVAAKIDPAAGAALFAAGAAAFAAAAALEKALA
G10	DAVAAALAEVTAQCRAIVAASEDPEAALAEATALATAFFAQFVGPEEARRLGEEHARAVLAEV
G11	AAAALAELAAAAAEAAEAAALAARRAARAAALAERERRREEEERRREEERR
G12	ELEEEEEEERRRFEEEARRREEKRRREREERLRERIRRELEEELRRLLE
G13	LEAELKALLAELTALAAAAAAAAAAAAAAGDAELAAIWKAQAAKLNALAAKVAAALA
G14	STLALARALRAIGRAVAAALFGLGYAALKAGNVALAALLYALGAAVLAATTAAIRALLAAA
G15	SAKLKEEYEKAKAEAEKAKALAKEAAAKHPEAGKAYQKYADRLEKLAKAIEKQI
G16	MSLAEAIRDAGVAAALASGDPAHLDAAKAAIAAAVSPEEAARWAAVLDEDYARARAAAA
G17	MIEQLENATKLAKEIYEKLKKTGTPEEAKLAKEIYEKYKEKLEEAKAKKEKEEALKKLLEEL
G18	LLLLLLLALLALAALAAAAKVAAMRARRAAAAAARAAAEAARAAAEAAAREAAAA
G19	SLAEAAALAAARAAREARARERARLEAEREARRREEEKEEERRRELE
G20	SKQEALAKARELYEKARELIREGKFEEAEKLIEELEKTEQGKALAKALREELKKEKELLK
G21	LLEELELKAKEEEAKKEYEERKKEREEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEE
G22	EKLKKLQELADKAKKREELAKARAAEAKAKGDAAVAAECEATAAKYKETAENLQKEIEKLK
G23	MTPEELAALRAAAIAANKARLAAEEARRAAREAAALLAEAAAALAAALAEALAA
G24	ALEELRRAAEELIREIRENGEPSKETKERLIAAARRLREELGEEGARVVEELDREIERAYEEFKA
G25	KEEEEEAEKKREELLAKLKELAERNKKLAKEGIAPEEAKKIAEELEKKRKELE
G26	SREEEMERLEEEAKREVEEALRTEGIEAAKAVAERLAAEFEALGYLRVARRVRNYAESLILKAK
G27	MKEKAEKIKKEAIALAKEKGEELAKKVEAANPAEQYRMLKKLKKEKAEEK
G28	GLLLALLALLLLLALLRAAAIAAKGAAARAAAAAASAAYRAAKAAA
G29	SEEERREREEKLREEEERKAKLKAEGKAKREALKKAGEEYKKEREEKKEEEERKKKEEEE
G30	SDADRCAAAAARLRARAAQTEALAAQGLSPECRAAAAAEAARLRALAAELEARRAA

G31	LEEVIDVITRITDAIAQALIAANPVVGGAIAAAIRAAVSAALLALIK
G32	EKEKAILKQEAERAKKRAAGYNALAAEQAANGNTAAAAAAKAKAAALKAKAAEYEEKAK
G33	EEEEERRKREEELRARIAALEAEAAAARALGAAIPALAAAAEGAAEAARRQAARIREYLAR
G34	AAAAAAEAAAAAAAAAAAAAAAAAAARAAVEALKAALAAEKLLKELLS
G35	AAEEEAERLRRAAAELAERLARAALLAALRAALAARLAANALKIAAAAAAALAAAAA
G36	EEEELERELERLREERERFKEEEEKRRAALEAEAAARRAAAERAAEEARRLAE
G37	MTPEERARALLEAALAHTQARLAELKARTAESPEYAAEAEAAAARLARLQAALAA
G38	DKLEKELAETTAEAEALAAAAAKAAAEDPEKAKEYAHRLAVAQRRRDNIEKELEKKK
G39	MEREREEERRRAEEEAERARRAAERLEEKLAARKKEREKLLKERLEKELEELKKELE
G40	SVKEAIEEAIKEIKETEELSAGLSSTERAERIEETRKRAEERIREVIEKKKKE
G41	SLEEVKQEAEARIARLRAEAERICAASSPEECAEQRAIVEAREARIKALVE
G42	YLEEAVAALKKLRDDLAAQLAKAKAAADTPEMKALAAETQALLELATKQLEKAEAKLK
G43	SLEELLKELEELKKKLLEALAALALALRRAANQAPPELREKLLALADALKKLYLSLLK
G44	SLAEALRALAAAAGKALGNIAAGAAFLKALLAALAAAAAAAAAAA
G45	MTEEEKEREEEKKKEAAKREAERALAELARLLAAAAAALALLAALA
G46	SEAEKLVEKMREVRKEYRELAAKKGEEEAKKVKEEMEKELKKLKEEREKVIKEKEEK
G47	ELEKKEEELKKKKAEERAAKAAALLAYALAKALAELLEALAKALA
G48	EEEEAARKKLQEEVEKLEKETKERVKELEEKAKESTPEEAEKYKKEAEEVVEETVKKIEEICKS
G49	LLALAARLAELAARRAAIEARRAEVAARRAARAAAAAAEAAAALA
G50	DLLAELLRAVDAAIRAAAEALAAAAATEEEKLAVLQAGRLAIRLFRFLLALLRALLLS
G51	SPEALREACRAELEAVNKEYEKAKEEAKKLTPEEAKKVLAEQNKRLREAVERCRAAEA
G52	EEEKKAKEKELLIAALKRKAEEETARLKAEGKEAEAEAFRAECEAKIAAAK
G53	EEEELKKAIEEAKKKLAELKEERERRERAARLEEERRRREEEERRRRE
G54	MMEELKELVERLLELIDELLKNKELQPFAEELIKRFKEASKLSLEEFYKKVKELLEEAEKKLKEL
G55	SEEEKLKELAKKRKEELEAAKARREAEVARVLAEESEAAAKAKAAELTAAIVAKHKEFLKKLEE
G56	KEEKIKELEEKAKELEEEFKKAAKELIEAEKAGDKEKAEKYKEKAKELLKKKKEAEKKRKELE
G57	EELARLAEELAAARREALRAELEALRREQEERLREEEEERRRREEEEK
G58	ELEEELAKKLREEREAALKARAAALRAAAEAERAARAERAAAARAAKAEAERERLA
G59	LTEEEIKKLREQALQLREEINRLKSKAALASPEEKAKLEEEIKEKEAEREELIKKIK
G60	AALAAALQAALAAAIDAAIAAAAAARAAAARAEEALKKLLEELAK
G61	SLAEIAARLAANKARAALFKALAAKLAKLTTPELAKKLAAKIVAK
G62	MEKELEEKLKKKEEEKKEKEAEEERKKAKEKAKELEEKKKKEKELKELE
G63	LPLAELLALLAALAAAQLAAKAAYAAKKAAIKAAGEARKAAEAAAAAEEAELL
G64	AQAELEAKRAAAAALEAELLEKRREEEEEEEEEEEEEEEEEEEEEEE
G65	SSLAKKILELRKKALEGLKAGKRTPETKRFQDIVEKTLKEEAEKAAKKALEELK
G66	MELEKKIEEELAKMKELAAKGPEYRPELEKAGAKAREYREKLREEALKKEEEE
G67	ELLKKLILLILKIGRGLNELGKKLRKAGLKKLANKFFKIGRKLYEIAEKL
G68	AEALEDLARRLIAKVREEAEKRAKAGTPEEAKEAWKEAAELRARIEALRAALLAA
G69	AATIAAYAAKLAAHAAQCRALGAAHPEIKAAAEKNAAAIEAARDKALANAAA
G70	SAAEAAFEAELAAAKAAYAAERDAAAAKYKDNLEKREAALREWHRKEAETIAALKAKHAAEKAAA
G71	DVATLKALAAQYRAARAAVREEAARLAAAEPERAAEILAEGAALAAAFDAKAAAAAAAAAAAA
G72	TEEELRRELEARRAAAEEEARREAARKEQAELQAKAAALEAALKAAA
G73	EEEERRRREEEEREARLAALREENERRRAERQAKIAAAALAALLAALLAA
G74	SLEENKKQWEKLQEETEKEVAKLKAEGDARRAALAAEAAARAAAEALAALAAL
G75	SEIKEEAERREREEEEERARRIAEAEAARAAAAAAAAAAALAAAAAEELARA

G76	AEEELRAAEERAAALAEAAARKAAREARKAAEEAAKAAAEAAELERKRAE
G77	EEEEKERLKEELREKLKKLLEEASKLENPQEVSEEAKKVYEEYLAL
G78	SAKECTKLKVETHKKYQELAKKSKPEDLPELVKKKEEELKKIEEKCK
G79	EEEIRKRKEELEKKIAEAKEELEKAKSNPEMAKIAQELLEKLKAWAAEEIAKLEK
G80	AEEKEKRLEELKKEAARLEEEAARYKELAPETGLEARKKAGEAEREREKALEEIRCLEAE
G81	GYAELAKRYEEIAAKLKEQAKKNKEKGISEEKAKYLEEKAAEFEAKAKEAKAIYEA
G82	ETLAQLRAERDRAEARRAALLALPPEERAANAAAIAAAAAAAAETEAGIKKLEEEK
G83	LLEQLATLAALAAALAERAAREARRERLRREREEEERRRREEEERRREEEERRRR
G84	EEEEKKRKEELEKLKAEKEAERAALKAAEEAYRAAREAARREAEERERRER
G85	LSPEELEARRAALRAAEEARRAAERAAREEARAAAAAAEAEARARLEAELAAAA
G86	DPEELKKEAEELAKKAEIYKKLAEEAAAKYSQSAADRLKAKAAEYEAKRKAVEAKLKALE
G87	EALRAAIDALADALLAATDALAAASTPEAAARLRAATAAAIRALYALAE
G88	SEVERLKAEAARLTERIVELAEKAMELAKKSTPEEAKKIMEEAKKERDRLRAERERALAEAEALE
G89	SKEEIEALEKKLAELKALAEKAAAENPVLAAQYRAQALEAEAQLEKLKKEA
G90	LSDLASALAQLALGLELLADPETKEEGLKLIAEALARLAAALEQLARLLAGLAAKAAEAA
G91	AEVAKLKAEAAAAKAKAAAYAAAGNLAAADAARKKALEYEAKANKALEE
G92	EEELAKAALEAERAALKAAREAERAAREAAAAERAAAAAAEEARRAAEAEAERERAA
G93	SEEEIEKIKEEAIKKLKEVKAEAEKKKATSSPEEREKIEKEAKEKMDEILREEREKIEKLKKE
G94	EEEELEEKLKKLLEEKEKKLKKELEERKKKRKEKLEKAEKELKKKLE
G95	LDEAVAAAREAIRAAVEEAKKLYKEDPEKGKELLKAAQAALAALRAARAAAEAAAAAA
G96	SALEEELAKAKAEAEKLAAETEKTGDEETAKKTLEARAKALKLEEEL
G97	SLRELLRLLASLALRLFRALRAAAGAFMAADPALGAALLAAVEALEEAFRALVLAILLS
G98	LAELYKKQAEEKQAQAAAALAQAAADPANAAELQAQAAKLQAQAAELQKKAAEALA
G99	EEEKEVEKKIKELLEKGKKSTPEEAAKYNAKAEYLKIEAEAKKRKEEAEKAKKLKELEEKL