SUPPLEMENTARY DATA

Molecular dynamics

Molecular modeling. A fragment sequence of 24 residues encompassing the region of interest of WT-KISS1R, *i.e*. the last intracellular domain (Figure S1a), has been taken (Figure S1b). The extra repetition adds 3 residues leading to a fragment length of 27 residues for PRR-KISS1R (Figure S1c). I-Tasser webserver (http://zhanglab.ccmb.med.umich.edu/I-TASSER/, Figure S1d) was used to generate potential structural models of these two protein fragments (1, 2). I-Tasser has been successfully assessed at the last four CASP meeting as the best *de novo* approach. It is a complex combination of multiple approaches including the prediction of secondary structures, protein contacts and solvent accessibility, threading approach LOMETS, of potential binding sites … Hundreds of structural models are generated, optimized and then clustered. The best five structural models were selected based on the lower C-scores (2). Using Modeller software v9.8 (3), the repetitions were constrained to adopt a PPII conformation ((4), Figure S1e), lower DOPE scores were used to select the best conformations each time. This led us to obtain 10 structural models for each of the two protein fragments. SCWRL4 was used to optimize the side-chain positions (5).

Molecular simulations. Molecular dynamics (MD) simulations were performed with GROMACS 4.0.5 software (5-9) using OPLS-AA force field (10) for proteins and the TIP 4P model for water was used (11) (Figure S1f). The structure was then immersed in a periodic water box neutralized with Na^+ or Cl^- counterions. Each system was then energy-minimized with the steepest-descent algorithm for 1000 steps. Once the system was heated at 300K, MD simulation was performed in NPT together, with both temperature and pressure kept constant at 300 K and 1 bar, respectively, using the Berendsen algorithm (12). The coupling time constants were t=0.1 ps and t=0.5 ps for temperature and pressure, respectively. Bond lengths were constrained with the LINCS algorithm (13), which allowed an integration step of 2fs. The generalized reaction field algorithm (14) was used for long-range electrostatic interactions using a dielectric constant of 54 and a cut-off of 1.4 nm for non-bonded interactions. A MD simulation was first performed for 100 ps, with protein atom positions constrained while both ions and water molecules were free. Then, the simulation was fully relaxed and

equilibrated for 5 or 50 ns. The coordinates were recorded at every ps interval. The MD was checked and analyzed using Gromacs tools.

Local protein structure analysis. The 20 structure model simulations were analyzed using secondary structure assigned by the DSSP software (15). PPII were assigned using a recently published development based on DSSP (16).

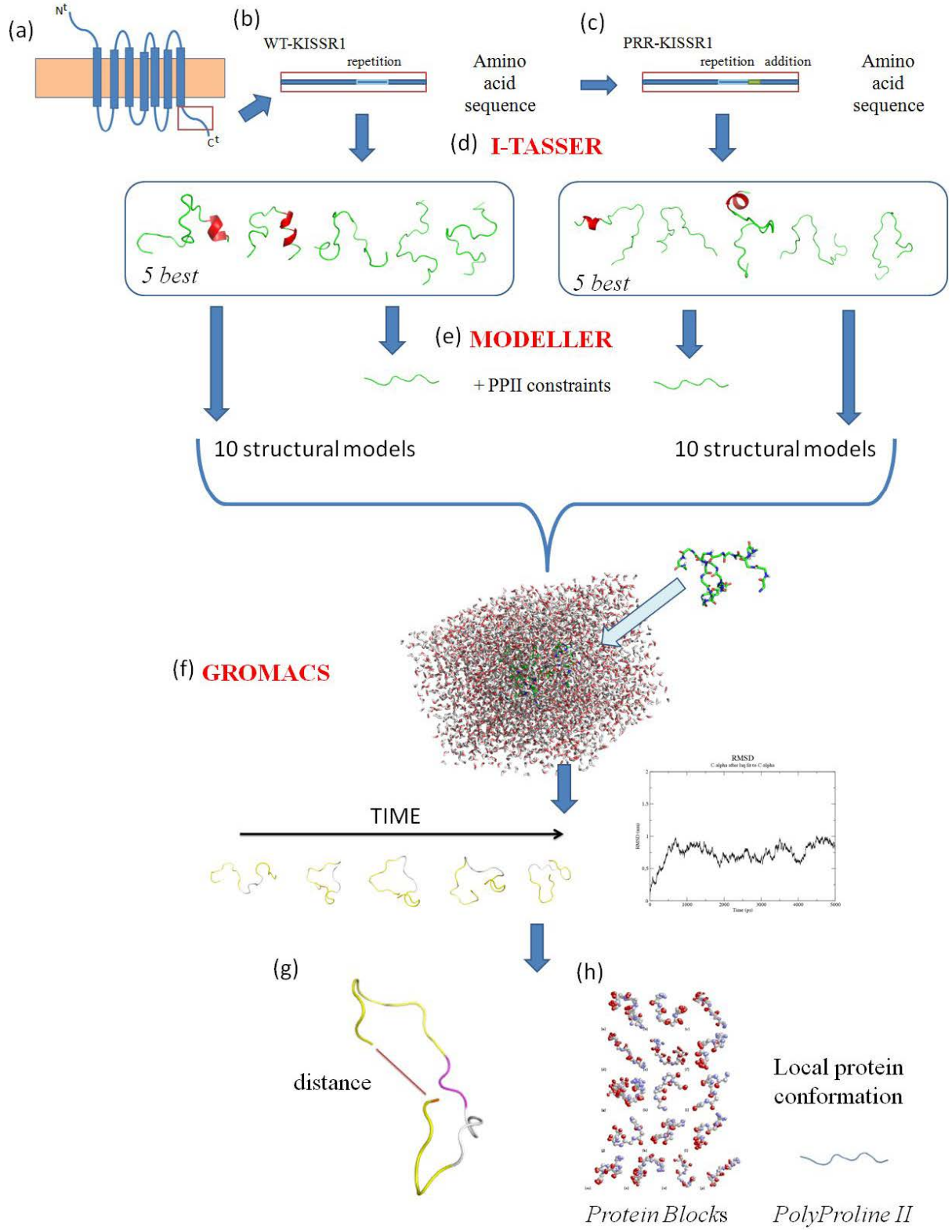
In the same way, Protein Blocks (PBs) analyses were performed (17). PBs (18) correspond to a set of 16 local prototypes, labeled from *a* to *p* (see Figure S1h), of 5 residue length, clustered based on F, Y dihedral angles description. They were obtained using an unsupervised classifier similar to Kohonen Maps (19) and Hidden Markov Models (20). The PBs *m* and *d* can be roughly described as prototypes for central alpha-helix and central beta-strand, respectively. PBs *a* through *c* primarily represent the N-cap region of beta-strand while PBs *e* and *f* correspond to the C-caps; PBs *g* through *j* are specific to coils, *k* and *l* correspond to the N cap region of alpha-helix, and PBs *n* through *p* to that of C-caps. PPII helices correspond mainly to PB *d* (16). This structural alphabet allows a reasonable approximation of local protein 3D structures (17) with an average root mean square deviation (*rmsd*) of 0.42 Å (21). PB assignment was carried out using an in-house program as done by iPBA web server (22). PB assignments were performed for every residue of the proteins and for every frame extracted from the MD simulations. The equivalent number of PBs (17) (N_{eq}) is a statistical measurement similar to an entropy that represents the average number of PBs a given residue takes. The N_{eq} is calculated as:

$$Neq = exp(-\sum_{x=1}^{16} f_x \, ln \, f_x)$$

Where f_x is the probability of PB *x*. A N_{eq} value of 1 indicates that a single PB is observed; on the opposite a value of 16 is equivalent to a random distribution.

Figure S1. *Proposition of structural models, molecular dynamics and their analyses*. (a) A fragment encompassing the region of interest is selected. (b) It corresponds to a 24 length residue fragment for $(PRR)_3$ sequence and (c) 27 for the $(PRR)_4$ sequence. (d) 5 structural models have been selected using I-Tasser software. (d) The different repetitions are constrained to adopt PPII helix conformation thanks to Modeller software. (f) The 20 structural models are used to perform Molecular Dynamics. (g) Distance between the extremities and (h) local protein conformations are analyzed.

Figure S1


(a)
Nt
Ct
(b)
WT-KISSR1
repetition
Amino acid sequence
(c)
PRR-KISSR1
repetition addition
Amino acid sequence
(d) I-TASSER
5 best
5 best
(e) MODELLER
+ PPII constraints
10 structural models
10 structural models
(f) GROMACS
TIME
RMSD
(g)
distance
(h)
Local protein conformation
Protein Blocks
PolyProline II

References

1. Roy A, Kucukural A, Zhang Y 2010 I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc 5:725-738
2. Zhang Y 2008 I-TASSER server for protein 3D structure prediction. BMC Bioinformatics 9:40
3. Sali A, Blundell TL 1993 Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 234:779-815
4. Fitzkee NC, Fleming PJ, Gong H, Panasik N, Jr., Street TO, Rose GD 2005 Are proteins made from a limited parts list? Trends Biochem Sci 30:73-80
5. Wang Q, Canutescu AA, Dunbrack RL, Jr. 2008 SCWRL and MolIDE: computer programs for side-chain conformation prediction and homology modeling. Nat Protoc 3:1832-1847
6. Berendsen HJC, van der Spoel D, van Drunen R 2005 GROMACS: A message-passing parallel molecular dynamics implementation. Comp Phys Comm 91:43-56
7. Hess B, Kutzner C, van der Spoel D, Lindahl E 2008 GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. J Chem Theor Comp 4:435-447
8. Lindahl E, Hess B, van der Spoel D 2001 GROMACS 3.0: A package for molecular simulation and trajectory analysis. J Mol Mod 7:306-317
9. van der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC 2005 GROMACS: Fast, Flexible and Free. J Comp Chem 26:1701-1718
10. Jorgensen WL, Maxwell DS, Tirado-Rives J 1996 Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. J Am Chem Soc 118:11225–11236
11. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML 1983 Comparison of simple potential functions for simulating liquid water. J Chem Phys 79:926-935
12. Berendsen HJC, Postma JPM, van Gunsteren WF, Di Nola A, Haak JR 1984 Molecular dynamics with coupling to an external bath. J Chem Phys 81:3684–3690
13. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM 1997 LINCS: a linear constraint solver for molecular simulations. J Comput Chem 18:1463–1472
14. Tironi IG, Sperb R, Smith PE, van Gunsteren WF 1995 Generalized reaction field method for molecular dynamics simulations. J Chem Phys 102:5451–5459
15. Kabsch W, Sander C 1983 Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22:2577-2637
16. Mansiaux Y, Joseph AP, Gelly JC, de Brevern AG 2011 Assignment of PolyProline II conformation and analysis of sequence--structure relationship. PLoS One 6:e18401
17. de Brevern AG, Etchebest C, Hazout S 2000 Bayesian probabilistic approach for predicting backbone structures in terms of protein blocks. Proteins 41:271-287
18. Joseph AP, Agarwal G, Mahajan S, Gelly JC, Swapna LS, Offmann B, Cadet F 2010 A short survey on protein blocks. Biophys Rev 2:137-145
19. Kohonen T 2001 Self-Organizing Maps (3rd edition): Springer
20. Rabiner LR 1989 A tutorial on hidden Markov models and selected application in speech recognition. Proceedings of the IEEE 77:257-286
21. de Brevern AG 2005 New assessment of a structural alphabet. In Silico Biol 5:283-289
22. Gelly JC, Joseph AP, Srinivasan N, de Brevern AG 2011 iPBA: a tool for protein structure comparison using sequence alignment strategies. Nucleic Acids Res 39:W18-23