Other methods and integrative approaches

Arnaud.Poterszman@igbmc.fr

The challenge of myriads of complexes

Protein act as complexes. From analysis of ~ 6200 yeast proteins

30 000 binary interactions (by focused small scale experiments)

Affinity purification of 1732 proteins \rightarrow 232 complexes composed of # 7.5 proteins per complex

9 partners per protein and 3.6 partners per domain (not all direct or at the same time)

Distribution of protein complexes in the PDB

Under-representation of structures from large complexes in view of the estimated average of 7.5 protein per complex



Many macromolecules are recalcitrant to main structural biology methods

Structural data on complex systems is often limited to isolated subunits and their domains or to low resolution evelopes by SAXS or EM.

Difficulty to express/reconstitute (incomplete bochemical characterization) and poorly abundant

Conformational heterogenity (prevents cristallisation and high resolution cryo-EM or do not stay intact during analysis (dissocation and/or aggregation on the EM grid)

System complexity

Integrative determination of macromolecular structures

Use structural information from any source

Measurements, physical principles and statistical inferences Resolution: low or high resolution

to obtain a set of models consistent available data



Integrative determination of macromolecular structures

Modeling from experimental structures, comparative modeling and distance constraints

Modeling configurations from connectivity information (native, X-link-MS..)

Propose multi-scale models

Modeling genomic region from 3C data





MOLECULAR STRUCTURE OF
NUCLEIC ACIDSJ. D. WATSONNo. 4356April 25, 1953J. D. WATSON
F. H. C. CRICKJ. D. WATSON
F. H. C. CRICKNATURE



To understand and modulate cellular processes, we need their models. These are best generated by considering all available information.

Other methods and integrative approaches

Structural data on complex systems is often limited to isolated subunits and their domains or to low resolution evelopes by SAXS or EM.

Introduce integrative approaches which allow to combine heterogenous data and propose hybrid models to provide the best possible description of the system.

Summarize mainstream complementary experimental and in silico methods to provide structural information on a macromolecular complex and discuss their pros/cons

Integrating modelling platforms

Integrative modeling platform (IMP)

Russel D, Lasker K, Webb B, Velazquez-Muriel J, Tjioe E, et al. Putting the pieces together:integrative modeling platform software for structure determination of macromolecular assemblies. PLoS Biol 2012;10, e1001244.

Inferential Structure Determination (ISD) framework

Rieping W, Nilges M, Habeck M. ISD: a software package for Bayesian NMR structure calculation. Bioinformatics 2008;24:1104–5.

HADDOCK

van Zundert GC1, Rodrigues JP1, Trellet M2, Schmitz C3, Kastritis PL4, Karaca E4, Melquiond AS5, van Dijk M6, de Vries SJ7, Bonvin AM1. The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. J Mol Biol. 2016 Feb 22;428(4):720-5.

RNABuilder

Flores SC, Sherman MA, Bruns CM, Eastman P, Altman RB. Fast flexible modeling of RNA structure using internal coordinates. IEEE/ACM Trans Comput Biol Bioinform 2011;8:1247–57.

Integrative Modeling Platform (IMP) http://integrativemodeling.org



A Sali

1/ Gathering data 2/ Representing and translating data into restraints 3/ Sampling good scoring configurations 4/ Analysis and assesment

1/ Gathering data

Determine the localization of two subunits of the yeast RNA Polymerase II, Rpb4 and Rpb7 (stalk), hypothesizing that we already know the structure of the remaining 10-subunit complex based on:

2/ Representing and translating data into restraints

- chemical cross-linking coupled with mass spectrometry (CX-MS),
- negative-stain electron microscopy (EM),
- X-ray crystallography data

3/ Sampling good scoring configurations

4/ Analysis and assesment



RNA Pol II is a eukaryotic complex that catalyzes DNA transcription to synthesize mRNA strands Eukaryotic RNA polymerase II contains 12 subunits, Rpb1 to Rpb12

The yeast RNA Pol II dissociates into a 10-subunit core and a Rpb4/Rpb7 heterodimer

Rpb4 and Rpb7 are conserved from yeast to humans, and form a stalk-like protrusion extending from the main body of the RNA Pol II complex

Rpb4/Rpb7







Experimental map of entire complex at 20.9Å resolution (represented with Gaussan mixture models (GMMs))

RX-ray structures of the 10-subunit core of RNA Pol II and of parts of Rbp4 and Rbp7







Experimental map of entire complex at 20.9Å resolution (represented with Gaussan mixture models (GMMs))

RX-ray structures of the 10-subunit core of RNA Pol II and of parts of Rbp4 and Rbp7

Chemical cross-linking coupled with mass spectrometry (CX-MS)







Gathering data Represent and translate data into restraints Sampling good scoring configurations Analysis and assesment

Experimental map of entire complex at 20.9Å resolution (represented with Gaussan mixture models (GMMs))

Macromolecules are represented using high and low resolution spherical beads and 3D gaussians (1 aa/bead and 20 aa/bead). Multi-scale representation

- Missing (unresolved) parts are modelled by low resolution beads
- Resolved regions as rigid bodies, allow unresolved regions to move (floppy bodies)







Define a scoring function, by which the individual structural models will be scored based on the input data

A simple sum of individual restraints

Each restraint maps to one of our input experiments or other physical/statistical information

Sequence connectivity restraint: residues that are adjacent in sequence will also be close in space due to the peptide bond

Excluded volume restraint: one protein cannot occupy the same space as another

EM restraints: A density overlap function to compare the GMM approximation of our model (em_components) with that of the EM map (target_gmm_file)

No electrostatics or stereochemistry; very different to a typical molecular mechanics simulation

Represent and translate data into restraints

Gathering data

Sampling good scoring configurations

Analysis and assesment

Define a scoring function, by which the individual structural models will be scored based on the input data

A simple sum of individual restraints

Each restraint maps to one of our input experiments or other physical/statistical information

Sequence connectivity restraint: residues that are adjacent in sequence will also be close in space due to the peptide bond

Excluded volume restraint: one protein cannot occupy the same space as another

EM restraints: A density overlap function to compare the GMM approximation of our model (em_components) with that of the EM map (target_gmm_file)

Cross-linking restrains: protein and residue numbers for each of the two linked residues (cross linker length,

Represent and translate data into restraints

Gathering data



Analysis and assesment

X-link/MS experiments





Here Monte Carlo is used to sample (not minimize) system (generate many models that satisfy the data)

Need to define a set of movers: rigid_bodies defines the components that will be moved as rigid bodies (in this case, the parts of Rpb4 and Rpb7 for which we have atomic structure). Unstructured regions will move as flexible beads.



https://integrativemodeling.org/

Srb (super rigid body)

Cluster (group by similarity) the sampled models to determine high-probability configurations.

- Chose a reference and align (superpose) all structures
- Calculate distances between structures (RMSD)
- Calculate localization densities for selected subunits

Rmsd_names == Rpb4, Rbp7





Gathering data

Cluster (group by similarity) the sampled models to determine high-probability configurations.

- Chose a reference and align (superpose) all structures
- Calculate distances between structures (RMSD)
- Calculate localization densities for selected subunits



https://integrativemodeling.org/

Gathering data

A founding example: the nuclear pore complex (NPC)

Yeast NPCs are ~50 Mda structures built of multiple copies of some ~30 different proteins (nucleoporins), totalling at least 456 protein molecules

Each NPC is a plastic structure embedded in the nuclear envelope and is composed of eight morphologically similar 'spokes' surrounding a central Tube

Filling this tube and projecting into both the cytoplasmic and nuclear sides are flexible filamentous domains from proteins termed FG (phenylalanine-glycine) repeat nucleoporins; these domains form the docking sites for transport factors that carry macromolecular cargoes through the NPC



Albert et al. 2007, 2008

Integrating spacial restrains from proteomic data



Albert et al. 2007, 2008













Representation:

Atomic Rigid bodies Coarse-grained Multi-scale Symmetry / periodicity Multi-state systems

Scoring:

Density maps EM images Proteomics FRET Chemical and Cys cross-linking Homology-derived restraints SAXS H/D exchange Second harmonic generation Native mass spectrometry Genetic interactions Statistical potentials Molecular mechanics forcefields Bayesian scoring Library of functional forms (ambiguity, ...)

Sampling:

Simplex Conjugate Gradients Monte Carlo Brownian Dynamics Molecular Dynamics Replica Exchange Divide-and-conquer enumeration

Analysis:

Clustering Chimera Pymol PDB files Density maps

Sali et al. From words to literature in structural proteomics. Nature 2003, 422, pp 216-227

Alber et al. Determining the architectures of macromolecular assemblies. Nature. 2007 Nov 29;450(7170):683-94.

F. Alber et al. "Integrating Diverse Data for Structure Determination of Macromolecular Assemblies" Annual Review of Biochemistry 77, 11.1-11.35, 2008.

D. Russel et al. "Putting the pieces together: integrative structure determination of macromolecular assemblies." PLoS Biol. 10, e1001244, 2012.

A. Ward et al. Integrative structural biology. Science 339, 913-915, 2013

Integrating modelling platforms

Integrative modeling platform (IMP)

Russel D, Lasker K, Webb B, Velazquez-Muriel J, Tjioe E, et al. Putting the pieces together:integrative modeling platform software for structure determination of macromolecular assemblies. PLoS Biol 2012;10, e1001244.

Inferential Structure Determination (ISD) framework

Rieping W, Nilges M, Habeck M. ISD: a software package for Bayesian NMR structure calculation. Bioinformatics 2008;24:1104–5.

HADDOCK

van Zundert GC1, Rodrigues JP1, Trellet M2, Schmitz C3, Kastritis PL4, Karaca E4, Melquiond AS5, van Dijk M6, de Vries SJ7, Bonvin AM1. The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. J Mol Biol. 2016 Feb 22;428(4):720-5.

RNABuilder

Flores SC, Sherman MA, Bruns CM, Eastman P, Altman RB. Fast flexible modeling of RNA structure using internal coordinates. IEEE/ACM Trans Comput Biol Bioinform 2011;8:1247–57.

Other methods and integrative approaches

Introduce integrative approaches which allow to combine heterogenous data and propose hybrid models to provide the best possible description of the system.

Summarize mainstream complementary experimental and in silico methods to provide structural information on a macromolecular complex and discuss their pros/cons

Comparative modeling and molecular docking



In silico models are not the product of experimental measurements of a physical sample.

They are generated computationally using various molecular modeling methods and underlying assumptions: comparative modeling, virtual docking of ligand molecules to protein targets, virtual docking of one protein to another, simulations of molecular dynamics and motions and de novo (ab initio) protein modeling.

Structures predictions

In absence of experimental 3D structure, from a sequence (or better from a multiple sequence alignment)

- Secondary structure predictions
- Prediction of 3D structure :
 - With reference to a known parental architecture parente, Homology modelling
 - Without a known parental architecture (still unsolved problem)
 Fold recognition (Threading, Profile recongnition)
 Prediction of a new fold

Homology modeling



http://swissmodel.expasy.org/

MENU

Modeling requests:

- First Approach mode
- Alignment Interface
- Project (optimise) mode
- Oligomer modeling
- GPCR mode

Model Database

• <u>SWISS-MODEL Repository</u>, a database for theoretical protein models.

Interactive tools

- <u>SWISS-MODEL Workspace</u>, an interactive working environment for protein structure modelling and assessment.
- <u>DeepView Swiss-PdbViewer</u>, a tool for viewing and manipulating protein structures and models.
- Lookup ExPDB template codes accessible to SWISS-MODEL.
 Source the SWISS-MODEL Template library
- <u>Search</u> the SWISS-MODEL Template library.
- Examples using SWISS-MODEL and the Swiss-PdbViewer.
 ANOLEA Protein structure quality check (atomic non-local
- ANOLEA Protein structure quality check (atomic non-local environment assessment)
 Nowe from Swiss Model
- <u>News</u> from Swiss Model.

Other links

- <u>Course</u> on protein structure and comparative modeling.
- Other <u>Web-based</u> Comparative Protein Modeling Servers.
- <u>PHYRE</u>, fold recognition server at the ICRF.
- <u>PredictProtein</u>, Burkhard Rost's sequence analysis and structure prediction server.

HELP

- Frequently Asked Questions.
- Visualising 3D models.
- Reliability of models.
- How SWISS-MODEL works.
- How ProModII works
- Modelling of oligomeric proteins.
- Model Confidence factors.
- About model quality.
- Methods and Programs.
 Databases used



An Automated Comparative Protein Modelling Server

SWISS-MODEL is a fully automated protein structure homology-modeling server, accessible via the ExPASy web server, or from the program DeepView (Swiss Pdb-Viewer). The purpose of this server is to make Protein Modelling accessible to all biochemists and molecular biologists World Wide.

The present version of the server is 3.5 and is under constant improvement and debugging. In order to help us refine the sequence analysis and modelling algorithms, please report of possible bugs and problems with the modelling procedure.

SWISS-MODEL is provided by:



History

SWISS-MODEL was initiated in 1993 by Manuel Peitsch, and further developed at Glaxo Wellcome Experimental Research in Geneva and the SIB Swiss Institute of Bioinformatics by Manuel Peitsch, Nicolas Guex and Torsten Schwede. Since 2001, SWISS-MODEL is being developped by Torsten Schwede's <u>Structural Bioinformatics</u> <u>Group</u> at the SIB & Biozentrum (University of Basel). The <u>SWISS-MODEL Repository</u>, a relational database of annotated three-dimensional comparative protein s tructure models, was established in 2004. In 2005, SWISS-MODEL service was extended by <u>SWISS-MODEL Workspace</u>, a web-based work bench for protein structure modelling and assessment. Computational resources for the <u>SWISS-MODEL</u> service ver are provided in collaboration by the Biozentrum (University Basel), the <u>Swiss</u> Institute of Bioinformatics and the <u>Advanced Biomedical Computing Center</u> (NCI Frederick, USA).

Acknowledgements

SWISS-MODEL would not have been possible without a lot of help and support. We are particularly thankful to Nicolas Guex for his many crucial contributions to the development efforts of Swiss-Model and specifically DeepView and to Gale Rhodes of the University of Southern Maine for coordinating the active DeepView user community. We also thank Alexander Diemand, Konstatin Arnold, Jürgen Kopp and Lorenza Bordoli for their many contributions to the development and operations for the modeling platform. Furthermore, we deeply indebted to Jake V. Maizel Jr, Timothy N.C. Wells, Jonathan C.K. Knowles, and Allan Baxter who have provided the necessary environment and resources during various phases of this project. Finally, we thank Stanley K. Burt, Robert W. Lebherz III, Karol Miaskiewicz and Jack R. Collins of the Advanced Biomedical Computing Center at the National Cancer Institute in Frederick Maryland for their support and operating the US mirror of the Swiss-Model server. We gratefully acknowledge the financial support by GlaxoSmithKline, Novartis, the Swiss National Science Foundation, the Biozentrum of the University of Basel and the Swiss Institute of Bioinformatics.

Disclaimer

The result of any modelling procedure is NON-EXPERIMENTAL and MUST be considered with care. This is especially true since there is no human intervention during model building. Carefully read the header section of the files to know what templates and alignments were used during the model building process.



About MODELLER

MODELLER is used for homology or comparative modeling of protein three-dimensional structures (1,2). The user provides an alignment of a sequence to be modeled with known related structures and MODELLER automatically calculates a model containing all non-hydrogen atoms. MODELLER implements comparative protein structure modeling by satisfaction of spatial restraints (3,4), and can perform many additional tasks, including de novo modeling of loops in protein structures, optimization of various models of protein structure with respect to a flexibly defined objective function, multiple alignment of protein sequences and/or structures, clustering, searching of sequence databases, comparison of protein structures, etc. MODELLER is <u>available for download</u> for most Unix/Linux systems, Windows, and Mac.

Several graphical interfaces to MODELLER are commercially available from <u>Accelrys</u>. Teaching licenses are also available to those institutions that acquire and maintain a research license.

1. N. Eswar, D. Eramian, B. Webb, M. Shen, A. Sali. Protein Structure Modeling With MODELLER. in press, 2006.

2. M.A. Marti-Renom, A. Stuart, A. Fiser, R. Sánchez, F. Melo, A. Sali. Comparative protein structure modeling of genes and genomes. Annu. Rev. Biophys. Biomol. Struct. 29, 291-325, 2000.

3. A. Sali & T.L. Blundell. Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. 234, 779-815, 1993.

4. A. Fiser, R.K. Do, & A. Sali. Modeling of loops in protein structures, Protein Science 9. 1753-1773, 2000.

The current release of Modeller is 9v1, which was released on January 22nd, 2007. Modeller is currently maintained by Ben Webb.

Homology modeling

1 (ou several) 3D structure(s)
 1 multiple sequence alignment



htrß 1	209	PTDEEWELIKTVTEAHVATNAOGSHWKOKRKFLPEDIGOAPIVNAPEGGKVD	260
hPPARa	199	ETADLKSLAKRIYEAYLKNFNMNKVKARVILSGKASNNPFFVIHDMETLCMAEKTLVAKLVANGIONKEVE	269
hROR1	269	SMAELEHLAONISKSHLETCOYLREELOOITWOTFLOEEIENYONKOR	316
hVDR	124	LSEEOORIIAILLDAHHKTYDPTYSDFCOFRPPVRVNDGGGSHPSRPNSRH+++LSEEDSDDPSVTLELSO	223
dE75A	339	ELDDOPRLLAAVLRAHLETCEFTKEKVSAMBORARDCPSYSMPTLLACPLNPAPE	403
rNGFI-B	322	PDASPTNLLTSLIRAHLDSGPNTAKLDYSKFOELVLPRFGKED	364
hRARy	182	LSPOLEELITKVSKAHOETFPSLCOLGKYTTNSSAD	223
	· ·		
hRXRα	224	SSANEDMPVERILEAELAVEPKTETYVEANMGLNPSSP	261
rHNF-4	135	YEDSSLPSINALLOAEVLSOOTTSPISGINGDIRAKR	171
dusp	232	NSVSRDFSIERIIEAEORAETOCGDRALTFLRVGPYSTVOPDY	274
hCOUP-TFI	181	GHCYLSGYTSLLLRAEPYPTSRYGSOCMOPNN	212
		AF2-a	
hER	308	ISLTADOMYSALLDAEPPTLYSEYDPTRPFSE	339
hPR	680	DIOLIPPLINLLMSIEPDVIYAGHDNTKPDTS	711
hAR	666		697
hGR	525		556
hMR	731		762
	/31		/01
		H1 (H2)	
		signature	
		h hh ϕ_{AK} hp F I. DO II. h hh	
Ի	261	LEARSHEWET TYDA TYDUNDER KILDWECELDOT TLLKGCOMETMSLEADUR YNDES	320
hppapo	270	UP THE COORDENS THE TREE AT A BASIN IN CARE THE COMPANY AND A COMMUNICATION AND A COMU	320
hPOP1	317	VALUE OCCUPY THE A CONTRACT AND ADDRESS OF THE ADDR	376
hIDD	31/		3/0
IVDR	224		203
CLE/SA	404	LOSEQEF SQRFARVIRGVIDFAGAIFGFQLLTQDDKFTLLKAGLFDALFVRLICAFDSSI	403
rNGFI-B	365	AGDVQQFYDLLSGSLDVIRKWAEKIPGFIELSPGDQDLLLESAFLELFILRLAYRSKPGE	424
hrary	224	LGLWDKFSELATKC <mark>IIKIVEFAKRLPGFTGLSIADQ</mark> ITLLKAACLD <mark>ILML</mark> RICTRYTPEQ	283
nRXRO	262	NDPVTNICQAADKQLFTLVEWAKRIPHFSELPLDDQVILLRAGWNELLIASFSHRSIAVK	321
rHNF-4	172	IASITDVCESMKEQLUVLVEWAKYIPAFCELULDDQVALLRAHAGEHLLLGATKRSMVFK	231
dusp	275	KGAVSALCQVVNKQ <mark>L</mark> FQ <mark>MVEYARMMPHF</mark> AQ <mark>V</mark> FLD DQ VI <mark>LL</mark> KAAWIELLIANVAWCSIVSLDDG+++QP	361
hCOUP-TFI	213	IMGIENIÇELAARL <mark>L</mark> FS <mark>AVEWAR</mark> NIPFFPDLQITDQVŞLLRLTWSELFVLNAAQCSMPLHV	273
hER	340	ÅSMMGLLTNLÅDRELVHMINWAKRVPGFVDLTLHDOVHLLECAWLEILMIGLVWRSMEHP	399
hPR	712	SSLLTSLNQLGERQLLS <mark>VVKWSKSLPGF</mark> RNLHIDDQITLIQYSWMSLMVFGLGWRSYKHVSG	773
hAR	698	AALLSSLNELGEROLVHVVKWAKALPGFRNLHVDDOMAVIOYSWMGLMVFAMGWRSFTNVNS	759
hGR	557	WRIMTTLNMLGGROVIAAVKWAKAIPGFRNLHLDDOMTLLOYSWMFLMAFALGWRSYROSSA	618
hMR	763	ENLLSTLNRLAGKONIOVYKWAKYLPGFKNLPLEDOITLIOYSWMCLSSFALSWBSYKHTNS	824
	100		0
		H3 H4 H5	
		whith h	
hmp@1	221		202
hTRβ1	321	E h h Lh h	383
hTR β 1 hPPAR α	321 330	E h h Lh h ETLTLNG-EMAVTRGQLKNG-GLGVVSDAIFDLGMSLSS-FNLDDTEVALLQAVLINSSDRPGLAC MLVAYGNGFITHEFLKSLKK-FFCDIMEEKFDFAMKFNA-LELDDSDISLFVAAI <mark>IC</mark> GGDRPGLLN	383 393
hTRβ1 hPPARα hROR1	321 330 377	E h h Lh h ETLTLNG-EMAVTRGQLKNG-GLGVVSDAIFDLGMSLSS-FNLDDTEVALLQAVLLMSSDRPGLAC MLVAYGNGFITREFLKSLRK-FFCDIMEPKFDFAMKFNA-LELDDSDISLFVAAIICCGDRPGLLN NTVYFDN-GKYASPDVFKSL-GCEDFISFVFEFGKSLCS-MHLTEDEIALFSAFVLMSADRSVLQE	383 393 438
hTRβ1 hPPARα hROR1 hVDR	321 330 377 284	E h h Lh h ETLTLING-EMAVTRGQLKNG-GLGVVSDAIFDLGMSLSS-FNLDDTEVALLQAVLIMSSDRPGLAC MLVAYGNGFITREFLKSLRK-PFCDIMEPKFDFAMKFNA-LELDDBDISLFVAAIGCGDRPGLLN NTVYFDN-GKXASDPUFKSL-GCEDFISFVFEGKSLCS-MHLTEDEILAFSAFVLMSADRSMLQE MSWTCGNQDYKYRVSDVTKA-GHSLELIEPLIKFQVGLKKLNLHEEEHVLLMAICIVSPDRPGVQD	383 393 438 338
hTRβ1 hPPARα hROR1 hVDR dE75A	321 330 377 284 464	E h h Lh h ETLTLING-EMAVTRGQLKNG-GLGVVSDAIFDLGMSLSS-FNLDDTEVALLQAVLIMSSDRPGLAC MLVAYGNGFITREFLKSLRK-PFCDIMEPKFDFAMKFNA-LELDDSDISLFVAAIIC MTVJFDN-GKYASPDVFKSL-GCEDFISFVFEFGKSLCS-MHLTEDEIALFSAFVIKSADRSWLQE MSWTCGNOPKYRVSDVTKA-GHSLELIEFLIKFQVGLKKLNHEEEHVLIMAICIVSPDRPGVQD NSIICLN-GQVMRRDAIQNG-ANARFLVDSTFNFAERMNSMNLTDAEIGLFCAIVLITPDRPGLRN	383 393 438 338 527
hTRβ1 hPPARα hROR1 hVDR dE75A rNGFI-B	321 330 377 284 464 425	E h h Lh h ETLTLNG-EMAVTRGQLKNG-GLGVVSDAIFDLGMSLSS-FNLDDTEVALLQAVLUNSSDRPGLAC MLVAYGNGFITREFLKSLRK-FFCDIMEPKFDFAMKFNA-LELDDBDISLFVAAIICGDRPGLLN NTVYFDA.GKXASPDVFKSL-GCEDFISFVFEGKSLCS-MLHITEDEIALFSAFVIMSADRSMLQE MSWTCGNQDYKKNVSDVTKA-GHSLELIEPLIKFQVGLKKLNLHEEHVLLMAICIYSPDRPGVQD NSIICLN-GQVMRRDAIQNG-ANARFLVDSTFNFAERMNSMNLTDAEIGLFCAIVLITPDRPGLRN GKLIFCS-GLVLHRLQCRRGFGWIDNILAFSRSLHG-LGVDVPAFACLSALVLIT	383 393 438 338 527 486
hTRβ1 hPPARα hROR1 hVDR dE75A rNGFI-B hRARγ	321 330 377 284 464 425 284	E h h Lh h ETLTING-EMAVTRQQLKNG-GLGVVSDAIFDLGMSLSS-FNLDDTEVALLQAVLIMSSDRPGLAC MLVAYGNGFITREFLKSLRK-FFCDIMEPKFDFAMKFNA-LELDDSDISLFVAAITCGDRPGLLN NTVYFDN-GKYASPDVFKSL-GCEDFISFVFEFGKSLCS-MHLTEDEIALFSAFVLMSADRFGLQ MSWTCGNQDYKYRVSDVTKA-GHSLELIEFLIKFQVGLKKLNLHEEHVLLMAICTVSFDRPGVQD NSIICLN-GQVMRRDAIQNG-ANARFLVDSTFNFARMNSMNLTDAEIGLFCATVLTFPDRPGLRN GKLIFCS-GUVLHRLQCARGFGDWIDNILAFSRSLHS-LGVUPVPAFACLSALVLTPDRHGLQD JTMTFSG-GLTLMRTQMMA-GFGFUTDLVFAFAGQLE-LEMUDTETGLSAICLTCGGDRHDLEE	383 393 438 338 527 486 346
hTRβ1 hPPARα hROR1 hVDR dE75A rNGFI-B hRARγ	321 330 377 284 464 425 284	E h h Lh h ETLTLING-EMAVTRGQLKNG-GLGVVSDAIFDLGMSLSS-FNLDDTEVALLQAVLLMSSDRPGLAC MLVAYGNGFITREFLKSLRK-FFCDIMERKFDFAMKFNA-LELDDBDISLFVAAIICGDRPGLLN NTVYFDN-GKXASPDVFKSL-GCEDFISFVFFEGKSLCS-MHLTEDEIALFSAFVIMSADRSMLQE MSWTCGNQDYKKYVSDVTKA-GHSLELIEPLIKFQVGLKKLNLHEEHVLLMAICIYSPDRPGVQD NSITCLN-GQVMRRDAING-AMARFLVDSTFMFARMNSMILTDAEIGLFCAIVLTPDRPGLRN GKLIFCS-GLVLHRLQCARGFGDMIDMILAFSRSLHS-LGVDVPAFACLSALVLTTDRHGLQD DTMTFSD-GLTLNRTQMHA-GFGPLTDLVFAFAGQLLF-LEMDDTETGLSAICLICGDRMLEE	383 393 438 338 527 486 346
hTRβ1 hPPARα hROR1 hVDR dE75A rNGFI-B hRARγ hRXRα	321 330 377 284 464 425 284 322	E h h Lh h ETLTLING-EMAVTRGQLKNG-GLGVVSDAIFDLGMSLSS-FNLDDTEVALLQAVLÄMSSDRPGLAC MLVAYGNGFITREFLKSLRK-PFCDIMEPKFDFAMKFNA-LELDDBDISLFVAATICGDRPGLLN MTVYFDN-GKXASDPUFKSL-GCEDFISVFFEGKSLC3-MHLTEDEIALFSAFVLMSADRSMLQE MSWTCGNQDYKYRVSDVTKA-GHSLELIEPLIKFQVGLKKLNLHEEEHVLLMAICIVSPDRPGVQD NSIICLN-GQVMRRDAIQNG-ANARFLVDSTRNFARSMNSMNLTDAEIGLFCAIVLTPDRPGLRN GKLIFCS-GUVLRLQCARGFGDVIDNILFSRSLH4-LGVUPVAFACLSALVIITDRPGLRD DGILLAT-GLHVHRQMHAA-GFGPLTDLVFAFAGQLLF-LEMIDTETGLSAICLICGDRMDLEE DGILLAT-GLHVHRNSAHSÅ-GVGAIFDRVLTELVSKMRDMQMDKTELGCLRAIVLFNPDSKGLSN	383 393 438 338 527 486 346 385
hTRβ1 hPPARα hROR1 hVDR dE75A rNGFI-B hRARγ hRXRα rHNF-4	321 330 377 284 464 425 284 322 232	E b b Lh b h ETLTLING-EMAVTRGQLKNG-GLGVVSDAIFDLGMSLSS-FNLDDTEVALLQAVLIMSSDRPGLAC MLVAYGNGFITREFLKSLRK-PFCDIMERKFDFAMKFNA-LELDBDISLFVAATICGDRFGLLN MTVYFDN-GKYASPDVFKSL-GCEDFISFVFERGKSLCS-MHLTEDEIALFSAFVINSADRSMLQE MSWTCGNQDYKRVSDVTKA-GHSELLIEPLIKFQVGLKKLNLHEEHVLLMAICIYSPDRFGVQD MSITCLN-GQVMRRDAIQG-ANAFLVDSTFMFARMNSMNLTDAEIGLFCATVLTPDRFGLRN GKLIFCS-GLVLHRLQCARGFGDMIDMILAFSRSLHS-LGVDVFAFACLSALVLT	383 393 438 338 527 486 346 346 385 297
hTRβ1 hPPARα hROR1 hVDR dE75A rNGFI-B hRARγ hRXRα rHNF-4 dUSP	321 330 377 284 464 425 284 322 232 362	E h h Lh h ETLTLING-EMAVTRGQLKNG-GLGVVSDAIFDLGMSLSS-FNLDDTEVALLQAVLUNSSDRPGLAC MLVAYGNGFITREFLKSLRK-FFCDIMEPKFDFAMKFNA-LELDDBDISLFVAAIICGDRPGLAC MTVYFDN-GKXASPDVFKSL-GCEDFISFVFFGKSLCS-MHLTEDEIALFSAFVLMSADRSWLQE MSWTCGNQDYKKVSDVTKA-GHSLELIEPLIKFQVGLKKLNLHEEHVLLMAICTUSPDRPGVQD NSIICLN-GQVMRDAIQNG-ANARFLVDSTFNFAERMNSMNLTDAEIGLFCAIVLTPDRPGVQD INSICCN-GQVMRDAIQNG-ANARFLVDSTFNFAERMNSMNLTDAEIGLFCAIVLTPDRPGLRN GKLIFCS-GLVLHRLQCARGFGDVIDNILAFSRSLHG-LGVUVPAFACLSAIVLTPDRPGLQD DTMTFSD-GLTLNRTQMHAA-GFGPLTDLVFAFAGQLLI-LEMDDTETGLSAICIICGDRMDLEE DGILLAT-GLHVHRNSAHSÀ-GVGAIFDRVITELVSKMRDMQMIKTELGCIRATVIPNPDŠKGLSN IVVLLGN-DYIVFRICPELA-EMSRVSIRILDELVPGELQIDDEYACLKAITFPDPDAKGLSD QQLFLNG-SFSYHRNSAIKA-GVSAIFDRILESLSVKMRALNLDRELSCLKAITUNPDIRGIKS	383 393 438 338 527 486 346 385 297 427
$\begin{array}{l} hTR\beta 1\\ hPPAR\alpha\\ hROR1\\ hVDR\\ dE75A\\ rNGFI-B\\ hRAR\gamma\\ hRXR\alpha\\ rHNF-4\\ dUSP\\ hCOUP-TFI\\ \end{array}$	321 330 377 284 464 425 284 322 232 362 274	E h h Lh h ETLITING-EMAVTRGQLKNG-GLGVVSDAIFDLGMSLSS-FNLIDEVALLQAVLIMSSDRPGLAC MLVAYGNGFITREFLKSLRK-PFCDIMEPKFPFAMKFNA-LELDDSDISLFVAAICCGDRPGLN MTVYFDN-GKYASPDVFKSL-GCEDFISFVFFGKSLCS-MHLTEDEILFSAFVINSADRSWLQE MSWTCGNQDYKYNSDVTKA-GHSELLIEPLIKFQVGLKKLNLHEEHVLLMAICTVSPDRPGVQD MSITCLN-GQVMRRDAINGG-ANAFLVDSTFNFARMNSMNLTDAEIGFCATVLTPDRPGLRN GKLIFCS-GLVLHRLQCARGFGDWIDNILFSRSLHS-LGVDYPAFACLSALVITDRGLQD DTMTFSD-GLTLNRTQMHAA-GFGPLTDLVFAFAQQLL-LGVDYPAFACLSALVIT	383 393 438 338 527 486 346 385 297 427 337
hTRβ1 hPPARα hROR1 hVDR dE75A rNOFI-B hRARγ hRARγ hRXRα rHNF-4 dUSP hCOUP-TFI	321 330 377 284 464 425 284 322 232 362 274	E h h Lh h ETLTLING-EMAVTRGQLKNG-GLGVVSDAIFDLGMSLSS-FNLIDTEVALLQAVLLMSSDRPGLAC MLVAYGNGFITREFLKSLRK-FFCDIMERKFDFAMKFNA-LELDDBDISLFVAAIICGDRPGLAC MIVYFDN-GKXASPDVFKSL-GCEDFISVVFFGKSLC3MLTEDEIALFSAFVLMSADRSWLQE MSWTCGNQDYKYNVSDVFKA-GHSLELIEPLIKFQVGLKKLNLHEEHVLLMAICTVSPDRPGVQD NSIICLN-GQVMRFNDIQNG-ANARFLVDSTFMFARMNSMNLITDAREIGFCAIVLITSDRPGLRN GKLIFCS-GLVLHRLQCARGFGDVIDINIAFSRSLHS-LGVDVPAFACLSALVLITDRHGLQD DTMTFSD-GLTLNRTQMINA-GFGPLTDLVFAFAGQLLE-LEMIDTETGLSAIVLITDRHGLQD DGILLAT-GLHVHRNSAHSÅ-GVGAIFDRVLTELVSKMNDMQMIKTELGCLRAIVLFNPDRKGLSN QQLFLNQ-SFSYHRNSAIKA-GVSAIFDRILSELSVKMKRLNLDRELSCLKAIIFYNPDAKGLSD QQLFLNQ-SFSYHRNSAIKA-GVSAIFDRILSELSVKMKRLNLDRELSCLKAIILYNPDIRGIKS APLLAAA-GLHASPMSADRV-VAFMIHRIFQEQQVKKLKALHVISAEYSCLKAIVLFTS	383 393 438 338 527 486 346 385 297 427 337
hTRβ1 hPPARα hROR1 hVDR dE75A rNGFI-B hRARγ hRARα rHNF-4 dUSP hCOUP-TFI hER	321 330 377 284 464 425 284 322 232 362 274	E h h Lh h ETLTLING-EMAVTRQQLKNG-GLGVVSDAIFDLGMSLSS-FNLDDTEVALLQAVLIMSSDRPGLAC MLVAYGNGFITREFLKSLRK-PFCDIMEPKFDFAMKFNA-LELDDBDISLFVAATICGDRPGLAC MTVYFDN-GKXASDPUFKSL-GCEDFISVFFERGKSLG-MHLTEDEIALFSAFVLMSADRSWLQE MSWTCGNQDYKYRVSDVTKA-GHSLELIEPLIKFQVGLKKLNLHEEEHVLLMAICIVSPDRPGVQD NSIICLN-GQVMRRDAIQNG-ANARFLVDSTRNFARMNSMNLTDAEIGLFCAIVLTPDRPGLRN GKLIFCS-GUVLRLQCARGFGDUIDNILAFSRSLHG-LGVUPAFACLSALVIITDRPGLRD DGILLAT-GLHVHRNSAHSÅ-GVGAIFDRVLTELVSKMRDMQMDKTELGCLRAIVLFNPDŠKGLSN UVLLGN-DYIVFRICFELA-EMSRVSIRILDELVLPFGELQIDDEYACLKAITFFPPDRGLRSD QQLFING-SFSYRNSALKA-GVSAIFDRILSESVKMRLMLDRRELSCLKAITLFNPDAKGLSD GLLAAA-GLHASPMSADRV-VAFMDHIRIFQEQVEKLKALHVDSAEYSCLKAIVLFTSDAKGLSD GKLLFAP-NLLDRNQGKCVEGMVEIFDMLLÅTSSRFRM-MNLGBEFVCLKSIILLNSGVYŤFLSSTLKSLÉE	383 393 438 527 486 346 385 297 427 337
hTRβ1 hPPARα hROR1 hVDR dE75A rNGFI-B hRARγ hRARγ hRARα rHNF-4 dUSP hCOUP-TFI hER hPR	321 330 377 284 464 425 284 322 232 362 274 400 774	E h h Lh h ETLTLING-EMAVTRGQLKNG-GLGVVSDAIFDLGMSLSS-FNLIDTEVALLQAVLLMSSDRPGLAC MLVAYGNGFITREFLKSLRK-FFCDIMERKFDFAMKFNA-LELDDBJISLFVAAICGDRPGLAC MLVAYGNGFITREFLKSLRK-FFCDIMERKFDFAMKFNA-LELDDBJISLFVAAICGDRPGLAC MSWTCGNQDYKRYSDVTKA-GHSLELIEPLIKFQVGLKKLNLHEEHVLLMATCIVSPDRPGVQD MSWTCGNQDYKRYSDVTKA-GHSLELIEPLIKFQVGLKKLNLHEEHVLLMATCIVSPDRPGVQD MSWTCGNQUMRRDAING-ANARFLVDSTFMFARMNSMINITDAEIGLFCAIVLTPDRPGLRN GKLIFCS-GLVLHRLQCARGFGDWIDMILAFSRSLHS-LGVDVPAFACLSALVLTTDRHGLQD DTMTFSD-GLTLNRTQMINA-GFGELTDLVFAFAGQLE-LEMIDTETGLSAICLICGDRMDLEE IGILLAT-GLHVHRNSAHSÅ-GVGAIFDRVİTELVSKMDMQMIKTELGCLRAIVLIPNPDSKGLSN IVLLGM-DJIVFRICPELA-EMSVSIRILDELVLPFGEQUIDMEYACLKAIIFFDPDAKGLSD QLFINQ-SFSYHRNSAIKA-GVSAIFDRLISELSVLMRKNLNLDRELSCLKAIIFFDPDAKGLSD GKLLFAP-NLLĎRNGKCVEGNVEIFDMLLÅTSSRFRM-MNLGGEEFVCLKSITLNSGVYŤFLSSTLKSLĚE GKLLFAP-DLLLĎROGKCVEGNVEIFDMLLÅTSSRFRM-MNLGGEEFVCLKSITLNSGVYŤFLSSTLKSLĚE	383 393 438 527 486 346 346 385 297 427 337 471 837
hTRβ1 hPPARα hPOR1 hVDR dE75A rNGFI-B hRARγ hRXRα rHNF-4 dUSP hCOUP-TFI hER hPR hAR	321 330 377 284 464 425 284 322 232 362 274 400 774 760	E h h Lh h ETLTLING-EMAVTRGQLKNG-GLGVVSDAIFDLGMSLSS-FNLDDTEVALLQAVLÜMSSDRPGLAC MLVAYGNGFITREFLKSLRK-FFCDIMEPKFDFAMKFNA-LELDDSDISLFVAAIICGDRPGLAC MIVYZDN-GKXASDVFKSL-GCEDFISFVFFFGKSLCS-MLITEDEILAFSAFVLMSADRSMLQE MSWTGGNQDYKKNVSDVTKA-GHSLELIEPLIKFQVGLKKLNLHEEHVLLMAICIVSPDRPGLQN MSIICLN-GQVMRRDAIQNG-ANARFLVDSTFNFAENNSMLIDDEIGLFCAIVLTPDRPGLQN DSIICLN-GUVLRRLQCAR-OFFSDUIDNILFSSLHG-LGVUVPAFACLSALVLTPDRHGLQD DTMTFSD-GLTLNRTQMHA-GFGPLTDLVFAFAGQLLI-LEMIDTETGLSAIVLTPDRHGLQD DILLAT-GLHVHRNSAHSÅ-GVGAIFDRVLTELVSKMRDMQMDKTELGCLRAIVLFNPDSKGLSN UVLLGN-DYIVFRICPELA-EMSRVSINILDELVLFFGELQIDDNEYACLKAITEPPPDSKGLSD QUFFING-SFSYHRSAIKA-GVSAIFDRILSESVKMRLNLDRELSCKAIVLFNPDIKGLSD GKLLFAP-NLLDRNQGKCVEGMVEIFDMLLÅTSSRFRM-MNLQGEEFVCLKSIILLNSGVYFFLSSTLKSLĚE MLYFAP-DLILDRQGMKES-SFYSLCITMWQIPQEFVK-LQVSQEEFLCMKVLLLENTIPLGGLRS	383 393 438 338 527 486 346 346 385 297 427 337 471 837 823
hTRβ1 hFPARα hFOR1 hVDR dE75A rNGFI-B hRARγ hRARα rHNF-4 dUSP hCOUP-TFI hER hPR hAR hAR hAR	321 330 377 284 464 425 284 322 232 362 274 400 774 760 619	E h h Lh h ETLTLING-EMAVTRGQLKNG-GLGVVSDAIFDLGMSLSS-FNLDDTEVALLQAVLLMSSDRPGLAC MLVAYGNGFITREFLKSLRK-FFCDIMERKFDFAMKFNA-LELDDBJISLFVAAICGDRPGLAC MLVAYGNGFITREFLKSLRK-FFCDIMERKFDFAMKFNA-LELDDBJISLFVAAICGDRPGLAC MSWTCGNQDYKTHVSDVTKA-GHSEHIEPLIKFQVGLKKLNLHEEHVLLMATCIVSPDRPGVQD MSWTCGNQDYKTHVSDVTKA-GHSEHIEPLIKFQVGLKKLNLHEEHVLLMATCIVSPDRPGVQD MSWTCGNQOVMRRDAINGA-AHSLELIEPLIKFQVGLKKLNLHEEHVLLMATCIVSPDRPGVQD DTMTFSD-GLTLNRTQMHA-GFGPLTDLVFAFAGQLLF-LEMDDTETGLSAICLICGDRMDLEE IGILLAT-GLHVHRNSAHSÅ-GVGAIFDRVİTELVSKMRDMQMIKTELGCLKAIUTFDRHGLQD ITMTFSD-GLTLNRTQMHA-GFGPLTDLVFAFAGQLLF-LEMDDTETGLSAICLICGDRMDLEE IGILLAT-GLHVHRNSAHSÅ-GVGAIFDRVİTELVSKMRDMQMIKTELGCLKAIIFPDPDSKGLSN IVLLGAD-JTVFRICPELA-EMSVSIRILDELVLFFGEQUIDMEYACLKAIUFPDDAKGLSD QUEFINQ-SFSYHRNSAIKA-GVSAIFDRILSELSVKMRRNLNLDRELSCLKAIUFPDDAKGLSD GKLLFAP-NLLĎRNGKCVEGNVEIFDMLLÅTSSRFRM-MNLIGGEEFVCLKSIILNSGVYFFLSSTLKSLĚE GMLYFAP-DLILNGCRMKES-SFYSICLTMWQIPQEFVK-LQV3QEEFLCMKVLLLNSIFSIFSVDGLKN NLLCFAP-DLINNGCRMES-SFYSICLTMWQIPQEFVK-LQV3DEFLCMKVLLLNSIFSIFS	383 393 438 338 527 486 346 346 385 297 427 337 471 8337 471 823 682
hTRβ1 hPPARα hVDR dE75A rNGFI-B hRARγ hRXRα rNNF-4 dUSP hCOUP-TFI hER hPR hAR hGR hMR	321 330 377 284 464 425 284 322 232 362 274 400 774 760 619 825	E h h Lh h ETLTLNG-EMAVTRGQLKNG-GLGVVSDAIFDLGMSLSS-FNLIDTEVALLQAVLUNSSDRPGLAC MLVAYGNGFITREFLKSLRK-FFCDIMERKFDFAMKFNA-LELDDBJSLFVAAICGDRPGLAC MLVAYGNGFITREFLKSLRK-FFCDIMERKFDFAMKFNA-LELDDBJSLFVAAICGDRPGLAC MSWTCGNQDYKTNVSDVTKA-GHSLHLIPPLIKFQVGLKKLNLHEEHVLLMAICTVSPDRPGLQD MSIICLN-GQVMRDAIQNG-ANARFLVDSTFNFAERMNSMNLTDAEIGLFSAFVINSADRPGVQD NSIICLN-GQVMRDAIQNG-ANARFLVDSTFNFAERMNSMNLTDAEIGLFSAFVINSADRPGLQD DTMTFSD-GLTLNRTQMINA-GFGPUTDLVFFFAGLLI-LGVUVPAFACLSALVLTFDRPGLQD UTMTFSD-GLTLNRTQMINA-GFGPUTDLVFFFAGLLI-LENDDTETGLSAICIICGDRPGLSN QULFLNG-SFSYHRNSAHSA-GVGAIFDRUTELVSKMNDMQMIKTELGCLRAIVIPPDSKGLSN UVLLGN-DYIVFRHCPELA-EMSRVSIRILDELVLPFGEQUIDDNEYACLKAIIFPDPDAKGLSD QULFLNG-SFSYHRNSAIKA-GVSAIFDRILEELSVMMKINLNIRRELSCLKAIVIFTSDAKGLSD GKLLFAP-NLLIDRNQGKCVEGMVEIFDMLLÅTSSRFN-MNLGGEFVCLKSILLNNGVYŤFLSSTLKSLEE GMLYFAP-DLILNGQRMTUFONLLÅTSSRFN-MNLGGEFVCLKSILLNSGYTFLSSTLKSLEE GMLYFAP-DLILNGQRMTHS-RNYSQCVMMRLSQFGM-LQTPQSFLCMKVLLLSSTP	383 393 438 338 527 486 346 385 297 427 337 471 837 823 682 888
hTRβ1 hPPARα hPOR1 hVDR dE75A rN0FI-B hRARγ hRARγ hHR-4 dUSP hCOUP-TFI hER hPR hAR hGR hMR	321 330 377 284 464 425 284 322 232 2362 274 400 774 760 619 825	E b bLh b ETLTLING-EMAVTRGQLKNG-GLGVVSDAIFDLGMSLSS-FNLDDTEVALLQAVLLMSSDRPGLAC MLVAYGNGFITREFLKSLRK-PFCDIMEPKFPAKKINA-LELDBDISLFVAAICGDRFGLAC MTVYFDN-GKYASPDVFKSL-GCEDFISVFFFGKSLCS-MHLTEDEIALFSAFVLMSADRSWLQE MSWTCGNQDYKTVSDVTKA-GHSELLIEPLIKFQVGLKKLNLHEEHVLLMAICIVSPDRFGVQD MSWTCGNQDYKTVSDVTKA-GHSELLIEPLIKFQVGLKKLNLHEEHVLLMAICIVSPDRFGVQD MSITCLN-GQVMRFDAING-ANSHLVDSTFMFARMNSMNLTDAEIGLFCATVLTPDRFGLRN GKLIFCS-GLVLHRLQCARGFGDWIDMILAFSRSLHS-LGUDVFAFACLSALVLTDRHGLQD DGILLAT-GLHVHRNSAHSÅ-GVGAIFDRVLTELVSKNRDMQMDKTELGCLRATVLFNPDSKGLSN UVLLGN-DYIVFRICPELA-EMSRVSINILDELVLPFGELQIDDNEYACLKAIIFFDPDSKGLSN QUEFINQ-SFSYHRNSAIKA-GVSAIFDRLISELSVKMRRLNLDRELSCLKAIIFFDPDAKGLSD GKLLFAP-NLLLDRNQGKCVEGMVEIFDMLLÅTSSRFRM-MNLGGEEFVCLKSIILNNGVYFLSSTLKSLEE GKLLFAP-NLLLDRNMKSS-SFYSLCLTMWQLPQEFVK-LQVSLKSIILLNNGVYFLSSTLKSLEE RMLYFAP-DLIINEQRMKSS-SFYSLCLTMWQLPQEFVK-LQVSLEFLCMKALLLNTIPLEGLRS RMLYFAP-DLIINEQRMTHS-SMYSLCQUMHQISLQFVK-LQVSYEFLCMKATLLIKSIFP	383 393 438 338 527 486 346 385 297 427 337 471 837 682 888
hTRβ1 hPPARα hROR1 hVDR dE75A rNOFI-B hRARγ hRXRα rHNF-4 dUSP hCOUP-TFI hER hPR hAR hGR hMR	321 330 377 284 464 425 284 322 232 362 274 400 774 760 619 825	E h h Lh h ETLTLING-EMAVTRGQLKNG-GLGVVSDAIFDLGMSLSS-FNLIDTEVALLQVLLMSSDRPGLAC MLVAYGNGFITREFLKSLRK-FFCDIMERKFDFAMKFNA-LELDDBJSLFVAAICGDRPGLAC MIVYFDN-GKXASPDVFKSL-GCEDFISVFFEGKSLC3MHITEDEIALFSAFVLMSADRSMLQE MSWTCGNQDYKYRVSDVFKA-GHSLELIEPLIKFQVGLKKLNLHEEHVLLMATCTVSPDRPGLQD MSITCLN-GQVMRRDAIQNG-ANARFLVDSTFMFARMNSMILTDAREIGFCATVLTFPDRPGLQD DTMTFSD-GLTLNRTQMINA-GFGPLTDLVFAFAGQLE-LEMIDTETGLSAICLICGDRHGLQD DTMTFSD-GLTLNRTQMINA-GFGPLTDLVFAFAGQLE-LEMIDTETGLSAICLICGDRHGLQD DTMTFSD-GLTLNRTQMINA-GFGPLTDLVFAFAGQLE-LEMIDTETGLSAICLICGDRHGLQD DTULLGD-DYTVFRICPELA-EMSVSIRILDUVFGFUGLIDNEYACLKAIIFPDP	383 393 438 338 527 486 346 346 385 297 427 337 427 337 421 823 682 888
hTRβ1 hPPARα hPOR1 hVDR dE75A rNGFI-B hRARγ hRARγ hRARγ hCOUP-TFI hER hCOUP-TFI hER hAR hGR hMR	321 330 377 284 464 425 284 322 232 362 274 400 774 760 619 825	$\label{eq:linear} \begin{array}{c c c c c c c c c c c c c c c c c c c $	383 393 438 338 527 486 346 346 385 297 427 337 427 337 471 837 823 888
hTRβ1 hPPARα hROR1 hVDR dE75A rNGFI-B hRARγ hRARα rHNF-4 dUSP hCOUP-TFI hER hPR hAR hAR hAR hMR	321 330 377 284 464 425 284 322 232 362 274 400 774 760 619 825	$ \begin{array}{c} \underline{\mathbf{E}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} \\ \mathbf{h} \\ $	383 393 438 338 527 486 346 346 346 385 297 427 337 427 337 471 837 823 682 888
hTRβ1 hPPARα hOOR1 hVDR dE75A rNGFI-B hRARγ hRXRα rNF-4 dUSP hCOUP-TFI hER hPR hAR hGR hMR	321 330 377 284 462 284 322 232 362 274 300 774 760 774 760 619 825	$\label{eq:linear} \begin{array}{c c c c c c c c c c c c c c c c c c c $	383 393 438 338 527 486 346 346 385 297 427 337 471 837 823 682 888
hTR β 1 hPPAR α hPOR1 hVDR dE75A rNOFI-B hRAR γ hRAR γ hRAR γ hRAR α rHNF-4 dUSP hCOUP-TFI hER hPR hAR hMR hMR	321 330 377 284 464 425 284 322 232 362 274 400 774 760 619 825	$\label{eq:constraint} \begin{array}{cccc} & & & & & & & & & & & & & & & & & $	383 393 438 527 486 346 346 346 347 337 427 337 427 337 471 837 823 888
$\label{eq:hardenergy} \begin{split} & h \mbox{TR}\beta \mbox{1} \\ h \mbox{POR} \mbox{2} \\ h \mbox{OR} \mbox{1} \\ h \mbox{OR} \mbox{2} \\ h \mbox{CA} \mbox{2} \\ h \mbox{2} \\ $	321 330 377 284 465 284 322 232 232 232 274 400 774 760 619 825	$ \begin{array}{c} \underline{\mathbf{E}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} \\ \\ \underline{\mathbf{E}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} \\ \\ \underline{\mathbf{H}} & \underline{\mathbf{A}} & \underline{\mathbf{G}} & \underline{\mathbf{H}} & \underline{\mathbf{F}} & \underline{\mathbf{C}} & \underline{\mathbf{H}} & \mathbf{$	383 393 438 527 486 346 385 297 427 337 427 337 427 337 421 823 682 888
hTR β 1 hPPAR α hPOR1 hVDR dE75A rNOFI-B hRAR γ hRAR γ hHRR α rHNF-4 dUSP hCOUP-TFI hER hPR hAR hGR hMR hMR hTR β 1 hPPAR α hROR1	321 330 377 284 464 425 284 322 232 362 274 400 774 760 619 825 384 394 439	$\begin{array}{cccc} & \underline{\mathbf{E}} & \underline{\mathbf{h}} & \mathbf{$	383 393 438 527 486 346 385 297 427 337 471 823 682 888 888 456 466 511
hTR β 1 hPPAR α hVDR dE75A rNGFI-B hRAR γ hRXR α rNGFI-B hRXR α rNF-4 dUSP hCOUP-TFI hER hPR hAR hGR hMR hMR hTR β 1 hPPAR α hOPAR α hVDR	321 330 377 284 464 425 284 322 232 262 274 400 774 760 825	$ \begin{array}{c} \underline{\mathbf{E}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} \\ \\ \underline{\mathbf{H}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} \\ \\ \underline{\mathbf{H}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} \\ \\ \underline{\mathbf{H}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} \\ \\ \underline{\mathbf{H}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} \\ \\ \underline{\mathbf{H}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} \\ \\ \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} \\ \\ \underline{\mathbf{h}} & \mathbf{$	383 393 438 527 486 385 297 427 337 427 337 427 337 427 337 682 888 456 466 511
$\label{eq:hardborndensity} \begin{split} & h TR\beta 1 \\ h PFPAR\alpha \\ h PFPAR\alpha \\ h CR 1 \\ h VDR \\ dE75A \\ rNoF1-B \\ h RRAF \\ h MRAF \\ h MRAF \\ h COUP-TFI \\ h ER \\ h COUP-TFI \\ h ER \\ h AR \\ h GR \\ h M \\ h M \\$	321 330 377 284 464 425 284 322 232 362 274 400 7760 619 825 384 394 439 3394 328	$\label{eq:construction} \begin{split} & \underline{\mathbf{k}} \underline{\mathbf{h}} \mathbf{b} \\ & \underline{\mathbf{h}} \underline{\mathbf{h}} \mathbf{b} \\ & \underline{\mathbf{h}} \underline{\mathbf{h}} \underline{\mathbf{h}} \\ & \underline{\mathbf{h}} \underline{\mathbf{h}} \underline{\mathbf{h}} \\ & \underline{\mathbf{h}} \underline{\mathbf{h}} \underline{\mathbf{h}} \\ & \underline{\mathbf{h}} \underline{\mathbf{h}} \underline{\mathbf{h}} \\ & \underline{\mathbf{h}} \underline{\mathbf{h}} \underline{\mathbf{h}} \\ & \underline{\mathbf{h}} \underline{\mathbf{h}} \underline{\mathbf{h}} \\ & \underline{\mathbf{h}} \underline{\mathbf{h}} \underline{\mathbf{h}} \\ & \underline{\mathbf{h}} \underline{\mathbf{h}} \underline{\mathbf{h}} \\ & \underline{\mathbf{h}} \underline{\mathbf{h}} \underline{\mathbf{h}} \\ & \underline{\mathbf{h}} \underline{\mathbf{h}} \underline{\mathbf{h}} \\ & \underline{\mathbf{h}} \underline{\mathbf{h}} \underline{\mathbf{h}} \\ & \underline{\mathbf{h}} \underline{\mathbf{h}} \underline{\mathbf{h}} \\ & \underline{\mathbf{h}} \underline{\mathbf{h}} \underline{\mathbf{h}} \\ & \underline{\mathbf{h}} \underline{\mathbf{h}} \underline{\mathbf{h}} \\ & \underline{\mathbf{h}} \underline{\mathbf{h}} \underline{\mathbf{h}} \\ & \underline{\mathbf{h}} \underline{\mathbf{h}} \underline{\mathbf{h}} \underline{\mathbf{h}} \\ & \underline{\mathbf{h}} \underline{\mathbf{h}} \underline{\mathbf{h}} \underline{\mathbf{h}} \\ & \underline{\mathbf{h}} \underline{\mathbf{h}} \underline{\mathbf{h}} \underline{\mathbf{h}} \\ & \underline{\mathbf{h}} \underline{\mathbf{h}} \underline{\mathbf{h}} \underline{\mathbf{h}} \\ & \underline{\mathbf{h}} \underline{\mathbf{h}} \underline{\mathbf{h}} \underline{\mathbf{h}} \\ & \underline{\mathbf{h}} \underline{\mathbf{h}} \underline{\mathbf{h}} \underline{\mathbf{h}} \\ & \underline{\mathbf{h}} \underline{\mathbf{h}} \underline{\mathbf{h}} \underline{\mathbf{h}} \\ & \underline{\mathbf{h}} \underline{\mathbf{h}} \underline{\mathbf{h}} \underline{\mathbf{h}} \\ & \underline{\mathbf{h}} \underline{\mathbf{h}} \underline{\mathbf{h}} \underline{\mathbf{h}} \underline{\mathbf{h}} \\ & \underline{\mathbf{h}} \underline{\mathbf{h}} \underline{\mathbf{h}} \underline{\mathbf{h}} \underline{\mathbf{h}} \underline{\mathbf{h}} \underline{\mathbf{h}} \underline{\mathbf{h}} \\ & \underline{\mathbf{h}} $	383 393 438 527 426 346 346 346 346 486 346 486 346 483 488 888 456 466 511 424 426 602
$hTR\beta 1hPPAR\alphahROR1hVDRdE75ArNGFI-BhRAR\gammahRXRArNNF-4dUSPhCOUP-TFIhERhPRhARhGRhMRhTR\beta 1hPPAR\alphahOR1hVDRdZ75ArNOFI-B$	321 330 377 284 464 425 284 322 232 262 274 400 774 760 825 384 394 4394 4394 4399 528	$ \begin{array}{c} \underline{\mathbf{E}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} \\ \mathbf{E} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} \\ \mathbf{E} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} \\ \mathbf{E} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} \\ \mathbf{E} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} \\ \mathbf{E} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} \\ \mathbf{E} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} \\ \mathbf{E} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} \\ \mathbf{E} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} \\ \mathbf{E} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} \\ \mathbf{E} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} \\ \mathbf{E} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} \\ \mathbf{E} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} \\ \mathbf{E} & \underline{\mathbf{h}}$	383 393 393 385 527 486 346 346 346 346 346 427 337 471 823 682 888 426 511 424 602 563
hTR β 1 hPPAR α hPOR1 hVDR dE75A rNGFI-B hRAR γ hRXR α rNF-4 dUSP hCOUP-TFI hER hPR hAR hGR hMR hMR hPPAR α hROR1 hVDR dE75A rNGFI-B hRAR γ	321 330 377 284 464 425 284 322 232 362 274 3 22 274 4 00 774 776 619 825	$\label{eq:product} \begin{split} & \underline{\mathbf{k}} \underline{\mathbf{h}} \mathbf{k} \\ & \underline{\mathbf{k}} \underline{\mathbf{k}} \underline{\mathbf{k}} \underline{\mathbf{k}} \underline{\mathbf{k}} \underline{\mathbf{k}} \underline{\mathbf{k}} \\ & \underline{\mathbf{k}} $	383 393 527 486 346 346 346 346 427 337 427 823 337 471 837 823 808 808 456 456 511 424 602 553 418
$\label{eq:hardbarrendom} \begin{split} & h TR\beta 1 \\ h PFPAR\alpha \\ h PFPAR\alpha \\ h CR1 \\ h VDR \\ dE75A \\ r MOFI-B \\ h RAR \\ \gamma \\ h RAR \\ h RAR \\ h COUP-TFI \\ h RAR \\ h COUP-TFI \\ h RAR \\ h M \\ h M$	321 330 377 284 464 425 284 322 232 362 274 400 774 760 619 825 384 394 439 339 528 487 347	$\label{eq:product} \begin{array}{cccc} & \underline{\mathbf{b}} & \mathbf{b$	383 393 438 338 527 486 346 346 346 427 337 471 837 823 682 888 456 466 466 466 466 466 466 466 418
hTR β 1 hPPAR α hPOR1 hVDR dE75A rNGFI-B hRAR γ hRXR α rNHF-4 dUSP hCOUP-TFI hER hPR hAR hGR hMR hMR hMR hTR β 1 hPPAR α hROR1 hVDR dE75A rNGFI-B hRAR γ hRXR α	321 330 377 284 464 425 232 232 232 232 274 400 774 619 825 384 394 439 339 439 339 4487 339 528 487 339	$ \begin{array}{c} \underline{\mathbf{E}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} \\ \mathbf{E} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} \\ \mathbf{E} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} \\ \mathbf{E} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} \\ \mathbf{E} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} \\ \mathbf{E} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} \\ \mathbf{E} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} \\ \mathbf{E} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} \\ \mathbf{E} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} \\ \mathbf{E} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} \\ \mathbf{E} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} \\ \mathbf{E} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} \\ \mathbf{E} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} \\ \mathbf{H} & \underline{\mathbf{h}}$	383 393 393 385 527 427 336 297 427 337 427 337 427 823 682 888 888 456 456 511 424 602 563 418 457
$\label{eq:hardensity} \begin{split} & h TR\beta 1 \\ h PFPAR\alpha \\ h PFPAR\alpha \\ h CR1 \\ h VDR \\ dE75A \\ rMOFI-B \\ h RAR\gamma \\ h RAR\gamma \\ h RAR\gamma \\ h COUP-TFI \\ h COUP-TFI \\ h COUP-TFI \\ h COUP-TFI \\ h RAR \\ h M \\ h MR \\ h M \\ h$	321 330 377 284 464 425 222 284 222 232 362 274 400 774 60 619 825 384 394 439 339 528 487 347 386	$ \begin{array}{c} \underline{\mathbf{E}} & \underline{\mathbf{b}} & \underline{\mathbf{b}} & \underline{\mathbf{b}} & \underline{\mathbf{b}} \\ \mathbf{E} & \underline{\mathbf{b}} & \underline{\mathbf{b}} & \underline{\mathbf{b}} & \underline{\mathbf{b}} \\ \mathbf{E} & \underline{\mathbf{b}} & \underline{\mathbf{b}} & \underline{\mathbf{b}} & \underline{\mathbf{b}} \\ \mathbf{E} & \underline{\mathbf{b}} & \underline{\mathbf{b}} & \underline{\mathbf{b}} & \underline{\mathbf{b}} \\ \mathbf{E} & \underline{\mathbf{b}} & \underline{\mathbf{b}} & \underline{\mathbf{b}} & \underline{\mathbf{b}} \\ \mathbf{E} & \underline{\mathbf{b}} & \underline{\mathbf{b}} & \underline{\mathbf{b}} & \underline{\mathbf{b}} \\ \mathbf{E} & \underline{\mathbf{b}} & \underline{\mathbf{b}} & \underline{\mathbf{b}} & \underline{\mathbf{b}} \\ \mathbf{E} & \underline{\mathbf{c}} & \underline{\mathbf{c}} & \underline{\mathbf{c}} & \underline{\mathbf{c}} & \underline{\mathbf{c}} \\ \mathbf{E} & \underline{\mathbf{c}} & \underline{\mathbf{c}} & \underline{\mathbf{c}} & \underline{\mathbf{c}} & \underline{\mathbf{c}} \\ \mathbf{E} & \underline{\mathbf{c}} & \underline{\mathbf{c}} & \underline{\mathbf{c}} & \underline{\mathbf{c}} & \underline{\mathbf{c}} & \underline{\mathbf{c}} \\ \mathbf{E} & \underline{\mathbf{c}} & \underline{\mathbf{c}} & \underline{\mathbf{c}} & \underline{\mathbf{c}} & \underline{\mathbf{c}} & \underline{\mathbf{c}} & \underline{\mathbf{c}} \\ \mathbf{E} & \underline{\mathbf{c}} & \underline{\mathbf{c}} & \underline{\mathbf{c}} & \underline{\mathbf{c}} & \underline{\mathbf{c}} & \underline{\mathbf{c}} & \underline{\mathbf{c}} \\ \mathbf{E} & \underline{\mathbf{c}} & \mathbf{$	383 393 438 338 527 486 346 346 346 427 427 337 337 427 823 888 888 456 466 662 888 888 456 418 424 602 563 418
$hTR\beta 1hPPAR\alphahPOR1hVDRdE75ArNGFI-BhRAR\gammahRXRarHNF-4dUSPhCOUP-TFIhERhPRhARhGRhMRhMRhMRhTR\beta 1hPPARahROR1hVDRdE75ArNOFI-BhRARYhRXRarHNF-4dUSP$	321 330 377 284 464 425 232 232 232 232 232 274 400 774 619 825 384 439 339 439 339 4487 339 528 487 386 296	$ \begin{array}{c} \underline{\mathbf{E}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} \\ \mathbf{E} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} \\ \mathbf{E} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} \\ \mathbf{E} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} \\ \mathbf{E} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} \\ \mathbf{E} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} \\ \mathbf{E} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} \\ \mathbf{E} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} \\ \mathbf{E} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} \\ \mathbf{E} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} \\ \mathbf{E} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} \\ \mathbf{E} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} \\ \mathbf{E} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} \\ \mathbf{E} & \underline{\mathbf{h}}$	383 393 393 385 527 427 337 427 337 427 337 427 823 682 888 837 823 682 563 456 511 424 602 563 418 457 367 367
$\label{eq:hardenergy} \begin{split} & hTR\beta 1 \\ hPPAR\alpha \\ hPPAR\alpha \\ hPPAR\alpha \\ hPPAR\alpha \\ rMOFI-B \\ hRAR\gamma \\ hRAR\gamma \\ hHRR\alpha \\ rHNF-4 \\ dUSP \\ hCOUP-TFI \\ hER \\ hPR \\ hAR \\ hGR \\ hMR \\ hMR \\ hIRR \\ hMR \\ hTR\beta 1 \\ hPPAR\alpha \\ hROR1 \\ hVDR \\ dE75A \\ rNGFI-B \\ hRAR\gamma \\ hRAR\gamma \\ rHNF-4 \\ dUSP \\ hCOUP \\ n=1 \end{split}$	321 330 377 284 464 425 222 284 222 284 222 274 400 774 60 619 825 384 394 439 339 528 487 347 386 426 296 426 296	$ \begin{array}{c} \underline{\mathbf{E}} & \underline{\mathbf{b}} & \underline{\mathbf{b}} & \underline{\mathbf{b}} & \underline{\mathbf{b}} \\ \\ \underline{\mathbf{E}} & \underline{\mathbf{b}} & \underline{\mathbf{b}} & \underline{\mathbf{b}} & \underline{\mathbf{b}} \\ \\ \underline{\mathbf{E}} & \underline{\mathbf{b}} & \underline{\mathbf{b}} & \underline{\mathbf{b}} & \underline{\mathbf{b}} \\ \\ \underline{\mathbf{E}} & \underline{\mathbf{b}} & \underline{\mathbf{b}} & \underline{\mathbf{b}} & \underline{\mathbf{b}} \\ \\ \underline{\mathbf{E}} & \underline{\mathbf{b}} & \underline{\mathbf{b}} & \underline{\mathbf{b}} & \underline{\mathbf{b}} \\ \\ \underline{\mathbf{E}} & \underline{\mathbf{c}} & \underline{\mathbf{c}} & \underline{\mathbf{c}} & \underline{\mathbf{c}} \\ \\ \underline{\mathbf{E}} & \underline{\mathbf{c}} & \underline{\mathbf{c}} & \underline{\mathbf{c}} & \underline{\mathbf{c}} \\ \\ \underline{\mathbf{E}} & \underline{\mathbf{c}} & \underline{\mathbf{c}} & \underline{\mathbf{c}} & \underline{\mathbf{c}} \\ \\ \underline{\mathbf{E}} & \underline{\mathbf{c}} & \underline{\mathbf{c}} & \underline{\mathbf{c}} & \underline{\mathbf{c}} \\ \\ \underline{\mathbf{E}} & \underline{\mathbf{c}} & \underline{\mathbf{c}} & \underline{\mathbf{c}} & \underline{\mathbf{c}} & \underline{\mathbf{c}} \\ \\ \underline{\mathbf{E}} & \underline{\mathbf{c}} & \underline{\mathbf{c}} & \underline{\mathbf{c}} & \underline{\mathbf{c}} & \underline{\mathbf{c}} \\ \\ \underline{\mathbf{E}} & \underline{\mathbf{c}} & \underline{\mathbf{c}} & \underline{\mathbf{c}} & \underline{\mathbf{c}} & \underline{\mathbf{c}} \\ \\ \underline{\mathbf{E}} & \underline{\mathbf{c}} & \underline{\mathbf{c}} & \underline{\mathbf{c}} & \underline{\mathbf{c}} & \underline{\mathbf{c}} & \underline{\mathbf{c}} \\ \\ \underline{\mathbf{E}} & \underline{\mathbf{c}} & \underline{\mathbf{c}} & \underline{\mathbf{c}} & \underline{\mathbf{c}} & \underline{\mathbf{c}} & \underline{\mathbf{c}} \\ \\ \underline{\mathbf{E}} & \underline{\mathbf{c}} \\ \\ \underline{\mathbf{E}} & \underline{\mathbf{c}} \\ \\ \underline{\mathbf{E}} & \underline{\mathbf{c}}	383 393 438 338 527 486 346 346 427 427 337 337 427 823 837 823 888 471 823 888 471 823 561 424 424 888 456 466 466 466 466 466 418 427 427 337 337 337 421 823 337 337 421 823 829 429 429 429 429 429 429 429 429 429 4
hTRβ1 hPPARα hPPARα hVDR dE75A rNGFI-B hRARγ hRXRα rNNF-4 dUSP hCOUP-TFI hER hPR hAR hGR hMR hMR hTRβ1 hPPARα hVDR dE75A rNGFI-B hRARγ hRXRα rNF-4 dUSP hRXRα rNF-4 dUSP hRXRα	321 330 377 284 464 425 232 232 232 232 274 400 774 400 619 825 384 439 528 487 339 528 487 3347 386 296 338	$ \begin{array}{c} \underline{\mathbf{E}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} \\ \\ \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} \\ \\ \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} \\ \\ \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} \\ \\ \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} \\ \\ \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} \\ \\ \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} \\ \\ \underline{\mathbf{h}} & \mathbf{$	383 393 438 338 527 486 527 427 427 427 427 427 427 427 427 427 4
$\label{eq:hardensity} \begin{split} & h TR\beta 1 \\ h PFPAR\alpha \\ h PFPAR\alpha \\ h CR1 \\ h VDR \\ dE75A \\ r NoF1-B \\ h RAR\gamma \\ h RAR\gamma \\ h MR \\ h COUP-TFI \\ h COUP-TFI \\ h COUP-TFI \\ h RAR \\ h GR \\ h M \\ h M \\ h $	321 330 377 284 464 425 362 232 232 232 232 232 232 232 232 232	$ \begin{array}{c} \underline{\mathbf{E}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} \\ \mathbf{h} \\ \mathbf{E} \\ \mathbf{L} \\ \mathbf{L} \\ \mathbf{L} \\ \mathbf{A} \\ \mathbf{X} \\ \mathbf{G} \\ \mathbf{K} \\ \mathbf{M} \\ M$	383 393 438 338 527 486 346 346 427 427 427 427 427 427 427 427 427 427
hTRβ1 hPPARα hPPARα hVDR dE75A rNGFI-B hRARγ hRXRα rNNF-4 dUSP hCOUP-TFI hER hPR hAR hGR hMR hTRβ1 hPPARα hVDR dE75A rNGFI-B hRARγ hRXRα tRVF-4 dUSP hCOUP-TFI hRXRα rNNF-4 dUSP hCOUP-TFI hRXRα	321 330 377 284 464 425 232 232 232 232 274 400 774 619 825 384 439 439 528 825 339 447 339 528 487 339 447 339 447 338 426 338 472 272	$ \begin{array}{c} \underline{\mathbf{E}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} \\ \mathbf{h} \\ h$	383 393 438 338 527 426 486 346 346 4297 427 337 427 337 427 337 427 337 427 427 427 427 337 427 427 427 427 429 429 429 429 429 429 429 429 429 429
$\label{eq:alpha} \begin{split} & hTR\beta 1 \\ hPPAR\alpha \\ hPPAR\alpha \\ hPPAR\alpha \\ hPPAR\alpha \\ hTRF -4 \\ dUSP \\ hCOUP-TFI \\ hER \\ hCOUP-TFI \\ hER \\ hRR \\ hGR \\ hMR \\ hRR \\$	321 330 377 284 464 425 362 232 232 232 232 232 232 232 232 232	$ \begin{array}{c} \underline{\mathbf{E}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} \\ \mathbf{h} \\ h$	383 393 438 338 527 486 346 346 4297 427 823 682 888 456 456 456 456 456 456 456 456 448 457 367 497 409 456 553 418
$\label{eq:hardenergy} \begin{split} & hTR\beta 1 \\ hPPAR\alpha \\ hPPAR\alpha \\ hPPAR\alpha \\ hPPAR\alpha \\ hTRS 1 \\ hRAR\gamma \\ hRAR\gamma \\ hRAR\gamma \\ hRAR\gamma \\ hRAR\gamma \\ hRC0UP-TFI \\ hER \\ hPR \\ hAR \\ hGR \\ hMR \\ hTR\beta 1 \\ hPPAR\alpha \\ hRC01 \\ hVDR \\ dZ75A \\ rMOFI-B \\ hRAR\gamma \\ hRXR\alpha \\ rHNF-4 \\ dUSP \\ hCOUP-TFI \\ hER \\ hPR \\ hAR \\ hAR \\ hRC01 \\ hRC01 \\ hRAR\gamma \\ hR$	321 330 377 284 425 284 425 232 232 232 232 274 400 774 619 825 384 394 439 439 439 439 439 439 439 439 43	$ \begin{array}{c} \underline{\mathbf{E}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} & \underline{\mathbf{h}} \\ \mathbf{h} \\ \mathbf{E} \\ \mathbf{L} \\ \mathbf{L} \\ \mathbf{L} \\ \mathbf{L} \\ \mathbf{A} \\ \mathbf{X} \\ \mathbf{C} \\ \mathbf{M} \\ M$	383 393 438 338 338 338 427 427 427 337 337 427 823 888 888 456 466 662 8888 456 418 457 409 409 546 915 901
$\label{eq:hardenergy} \begin{split} & hTR\beta 1 \\ hPPAR\alpha \\ hPPAR\alpha \\ hPPAR\alpha \\ hPPAR\alpha \\ hTRSR\alpha \\ rNGFI-B \\ hRAR\gamma \\ hRAR\gamma \\ hRAR\gamma \\ hCOUP-TFI \\ hER \\ hPR \\ hAR \\ hGR \\ hMR \\ hRCR1 \\ hPPAR\alpha \\ hROR1 \\ hPPAR\alpha \\ hROR1 \\ hPOR1 \\ hROR1 $	321 330 377 284 464 425 362 232 362 232 362 274 400 774 619 825 384 394 4339 339 439 339 487 347 347 347 347 386 228 426 426 426 338 447 425 426 426 426 426 426 426 426 426 426 426	$ \begin{array}{c} & \underline{\mathbf{b}} \ \underline{\mathbf{b}$	383 393 438 338 527 420 4297 422 337 427 427 427 427 427 427 427 427 427 42
$\label{eq:hardenergy} \begin{split} & hTR\beta 1 \\ hPPAR\alpha \\ hPPAR\alpha \\ hPPAR\alpha \\ hPPAR\alpha \\ rMOFI-B \\ hRAR\gamma \\ rHNF-4 \\ dUSP \\ hCOUP-TFI \\ hER \\ hPR \\ hAR \\ hGR \\ hMR \\ \end{split}$	321 330 377 284 464 425 222 284 522 284 400 774 60 619 825 825 84 394 439 339 528 487 347 347 386 426 426 338 472 888 824 463 889	E h h h E h h h h MLVAYGNOFTTREPLKSLRK-PCDIMEPKDPAMKFNA-LELDDDDISLFVALLCQDRPGLN MSWTCGNQDYKYKSL-GCBUFTSVFEFGKSLCG-MLITEDEIALFSAFULMSA	383 393 438 338 527 426 486 346 385 2297 427 337 337 423 888 888 456 466 662 8888 418 455 418 456 446 407 407 409 559 5915 906
hTRβ1 hPPARα hPPARα hPPARα hPPARα hPPARα hTPPA hRXRα rINF-4 dUSP hCOUP-TFI hER hPR hAR hGR hMR hMR hTRβ1 hPPARα hRCR1 hVDR dE75A rNGFI-B hRARγ hCOUP-TFI hRXRα rINF-4 dUSP hRARγ hCOUP-TFI hRXRα rINF-4 dUSP hRARγ hCOUP-TFI hER hPR hAR hCOUP-TFI hER hPR hAR hCOUP-TFI	321 330 377 284 464 425 272 272 274 400 774 400 774 619 825 384 394 439 339 528 487 386 296 438 487 386 296 426 338 447 388 483 889	$ \begin{array}{c} \underline{E} & \underline{h} & \underline$	383 393 438 3393 448 336 527 427 427 427 427 427 427 427 427 427 4

Similarités de séquences. Signification Biologique / Signification statistique

Sequences from single domain proteins with unrelated folds were aligned and sequence identity plotted as a function of the aligment lenght:



Comparaisons des structures 3D de protéines homologues Déformations autour d'un thème commun

Comparaison des structures 3D de protéines homologues. Relations entre la divergence en séquence et la divergence en structure 3D



Relation type exponentiel % identité / déformations
Modelling structures of protein complexes



Integrative computational modeling of protein interactions

João P. G. L. M. Rodrigues and Alexandre M. J. J. Bonvin

Computational Structural Biology Group, Bijvoet Center for Biomolecular Research, Utrecht University, the Netherlands

Table 1. List of top-performing docking approaches participating in CAPRI.			Table 1. (Continued).						
Name	Protocol	Strengths and weaknesses	Integration of data	Public Web Server	Name	Protocol	Strengths and weaknesses	Integration of data	Public Web Server
ATTRACT (39)	Energy minimization in translational and rotational degrees of freedom using a reduced protein model and normal-mode analysis to allow conformational chapters	+ Fast derivative-based search method. + Conformational changes upon binding (local and global motions)	Implements interface data by adding atom/residue-specific weights, which can be negative (repulsive). Also offers the option to dock using Cao EM desity more	None		correlations generates a very large number of putative models, which are then re-scored using a shape correlations or shape plus electrostatic correlations.	2 graphical processing units). – Cannot handle flexible complexes.	intermolecular axis defined by a pair of residues, one on each interacting monomer. The angular range of the search can also be defined.	
	upon binding.	 Hotons). Support for Cryo-EM density maps. Not available via a web server. 	using cryo-ewi density maps.		PatchDock [25]	The surface of the molecules is divided into patches (concave, convex and flat) and only those containing	 + Extremely efficient and fast protocol. + Integrated suite of docking tools [24]. 	Regions/residues in the binding partners can be introduced to bias the scoring of the models.	http://bioinfo3d.cs.tau.ac.il/ PatchDock/
ClusPro (61)	IlusPro [61] Rigid-body search via a FFT correlation approach (PIPER), followed by structural similarity (RMSD) based clustering to find the most popular interaction modes, and final refinement of	+ Best automated R + Best automated R CAPRI evaluation [12], + Fast and exhaustive protocol. + Several docking 'modes' depending on the biological function (antibody/antigen, multimer, others), – Carnot handle flexible complexes.	Regions/residues in the binding partners can be introduced to bias the scoring of the models. Noteworthy option of 'negative' (repulsion) contacts.	http://cluspro.bu.edu/ pyl		'hot-spot' residues are kept. The patches are then matched using geometric hashing and pose-clustering techniques and the candidate models are examined (to remove extreme clashes) and scored.	 Fragmentation of protocol in very specialized tools requires a priori knowledge of their limitations. 		
	selected structures using CHARMM.				pyDock [22]	Rigid-body search via a FFT search method (custom optimized FTDOCK) followed by scoring with a combined electrostatics and desolvation	 + Fast protocol. + Integration of different modules (pyDockSIPPER, pyDockRST, 	PyDockRST module scores models based on agreement with user-defined distances. Implements SAXS scoring by generating synthetic SAXS	http://life.bsc.es/serviet/ pydock/home/
GRAMM- X ^a [68]	GRAMM- Grid-based FFT rigid body docking approach using a softened Lennard-Jones potential function. The top predictions are minimized and re-corred using a soft Lennard-Jones potential, an evolutionary concervation term, a statistical residue- residue potential, and the volume of the local energy minima in the ard.	+ Fast. + Uses CONSURF to determine evolutionary conserved residues. - Cannot handle flexible complexes.	Regions/tesidues in the binding partners can be introduced to bias the scoring of the models.	http:// vakser.bioinformatics.ku.edu/ resources/gramm/grammx/		energy function.	pyDockSAXS) to improve predictions. – Cannot handle flexible complexes.	curves of the models to user-provided data.	
					RosettaDock [38]	Low-resolution, rigid-body, Monte Carlo search followed by simultaneous optimization of backbone displacement and side-chain conformations using Monte Carlo minimization.	Scoring based on the Rosetta energy function. Y Very powerful refinement protocol. Poor (direct) support of interface information. _ Computationality	Molecules can be positioned manually in space (e.g. using PyMOL). Interfaces with other Rosetta tools (RosettaInterface) to validate mutagenesis data.	http:// antbody.graylab.jhu.edu/ docking
HADDOCK [32]	Rigid-body energy minimization followed by semi-flexible (interface) refinement and final optimization in explicit solvent. Returns clusters of models ranked by HADDOCK score.	+ Best performing team in the latest CAPRI evaluation [12]. + Restraint-based integration of ambiguous experimental/ prediction data. + Explicit flexibility of the interface. + Powerful user- friendly web interface. - Sincer than EFT.	Several types of restraints allow integration of different sources of data: distances, orientations, radius of gration, symmetry type, etc. Directly integrates several NMR-derived data.	http://haddocking.org	SwarmDock	Local docking and particle	demanding and slow protocol.	Residues belonging to the	http://
					[33]	swarm optimization of partner position and orientation, using normal modes to model induced fit, and final energy minimization. Uses the DComplex sooring function but the final models are re- ranked with a centroid potential prior to clustering.	 Second method matching explicitly models global flexibility upon binding. Siow. Predictions can take days unless a local version and sufficient computational resources are available. 	his binding site can be selected to bias the starting positions of the binding partners.	bmm.cancerresearchuk.org/ SwarmDock
Hex [27]	Spherical polar Fourier approach using rotational	based methods. + Extremely fast approach (~ 15 s using	Can restrict the rotational search around an	http://hexserver.loria.fr/	ZDOCK [28]	FFT-based rigid-body search using a scoring function composed of desolvation energy, electrostatics, and	+ Fast protocol + Robust performance over several CAPRI rounds.	Allows user-defined selection of 'blocked' and 'binding' residues that influence the	http://zdock.umassmed.edu/

Comparative modeling of the Mtq2/Trm112 complex

zinc-binding (pink) and methyltransferase (orange) domains of RImA(I) (PDB 1P91).

Superimposition of of Mtq2 (red) and Trm112 (green) models on their respective structural homologous domains of RImA(I).



Distance restraints derived from Ca– Ca contacts between the domains of RImA(I) and mapped onto the Mtq2 and Trm112 models).

Superimposition of Mtq2 (red) /Trm112 (green) models on the experimental .structure (rmsd 3.3 Å

Rodrigues and Bonvin. 2013

Structural data used in integrative modeling

Atomic structures of parts of the system	X-ray and neutron crystallography, NMR, Cryo-EM/ET, Comparative modelling and molecular docking
Composition and components positions	Purification from source with gel analysis or MS, Electron microscopy and tomography, gold labelling, Super resolution microscopy, FRET imaging
Physical proximity	Co-purification of sub complexes, native MS, genetic methods, sequence convariance, Y2H, Chromosome conformation Capture and other data,
Size and shape	AUC, SAS, atomic force microscopy, ion mobility MS Fluorescence correlation spectroscopy or anisotropy
Atomic and protein distances	NMR, FRET, EPR, X-link/ w/o MS
Binding site mapping	NMR, FRET, H-D/MS, mutagenesis
Solvent accessiblity	Footprinting methods including H-DX/MS, NMR and chemical modifications

Observations of complexes in their native environment

Electron microscopy and tomography, Super resolution microscopy, FRET imaging

Localization of proteins by immuno-EM. Immuno-EM montages for Pom152–PrA nuclei and Ndc1–PrA nuclear envelopes with goldlabelled antibodies. (Alber, 2007)

Super resolution microscopy: PALM (PhotoActivated Localization Microscopy), GSDim (Ground State Depletion imaging followed by Individual Molecule return)

Observe individual proteins with a resolution down to 20 nm in intact cells, and second-order statistics to study the spatial interactions of the proteins.









Purification from cells and mapping

Information: subunit composition and stoickiometry

Method: purification (from endogenous source and analysis)



Quantitative densitometry

Genome editing (yeast, mammalian cells)



ATG Endogenous TAP-tagging with CRISPR/Cas9 VELAG 2v Stre or TALENS Expression from Homology-directed repair natural genomic IxFLAG 2xStrep context Affinity purification of native complexes Mass spectrometry Enzymatic Structural assays studies Ac-CoA CoA Post-translationa Interactions modifications

Double-strand DNA break

Analysis of the SEA complex relative stoichiometry by SYPRO Ruby staining (Algret et al. 2014)

Dalvail et al. 2015

Quantitation by MS using ion peak intensity



Measure peak intensities from Different samples

Relative and absolute quantification

Compare isotopically labelled samples $(^{14}N/^{15}N)$

Isobaric Tags for relative and Absolute Quantification (iTRAC)

Isotope-Coded Affinity Tags (iCAT)

Spiking sample with isotope-labeled reference peptides

Thermofisher web site

Label free quantitation by MS

Label-free quantification approaches aim to correlate the mass spectrometric signal of intact proteolytic peptides or the number of peptide sequencing events with the relative or absolute protein quantity directly.

Relative quantitation strategies compare the levels of individual peptides in a sample to those in an identical, but experimentally modified, sample.

Absolute quantification can be obtained *estimated* from analysis of several mass spectrometric signal (TOP3 where the intensity of the selected peaks is taken into account) or the number of peptide sequencing events (emPAI == exponentially modified Protein Abundance Index).



Physical proximity of components

√ In vitro

- Co-immunoprecipitation
- GFP, GST, His, Strep-pull down assays
- ChIPs Protein arrays
- TAP-MS

Bait – Prey model



does X bind

with a protein?

✓In vivo

- Yeast two-hybrid system
- Phage display

Physical interaction between protein binding domains

Pair wise analysis and purification of sub-complexes



Systematic dissection of protein-protein interactions from recombinant proteins co-expressed in insect cells



Analysis of the protein interaction network Identification of key regulatory interactions













RESEARCH ARTICLE

Architectures of multisubunit complexes revealed by a visible immunoprecipitation assay using fluorescent fusion proteins

Yohei Katoh*, Shohei Nozaki*, David Hartanto, Rie Miyano and Kazuhisa Nakayama[‡]









Yeast two-hybrid system

- Detecting protein-protein interactions in yeast
- Transcriptional regulator system
- "prey"-"bait" model :fusion proteins with a transcriptional activating domain (AD, prey), a DNAbinding domain (DBD, bait)
- Term "two-hybrid" derives from these two chimeric proteins.
- Most commonly used method for large scale, high-throughput identification of potential proteinprotein interactions



Native mass spectrometry of large entities

Subunit stoichiometry Dissociation pathways and identification of sub-complexes.



Radu et al., in prep

Topological models from MS-based hybrid analysis



Build an hypothetical contact map from experimental data Sampling/scoring (Monte Carlo search and scoring) Clustering (and minimization) to identify most the best-scoring models







Politis et al. 2015 (Robinson and Sali's labs)

The main limitation: sample preparation

buffer exchange or desalting procedure

Ultra centrifugation

- micro-concentrators :
- Microcon, Centricon, Amicon (Millipore)
- Vivaspin (Vivasciences)



- centrifuge
- 5 to 7 cycles at least
- takes time but very efficient procedure !

Size exclusion chromatography

gel filtration colums :NAP-10 et NAP-5 (GE Healthcare)



- Often 2 runs with a concentration step in between
- takes less time but dilutes the sample

Equilibrium dialysis

- dialysis or mini-dialysis units :
 - Slide-A-Lyzer minidialysis (Pierce)



- dilutes the samples
- very easy to perform (overnight)

Covariation of RNA sequences

- RNA typically produced as a single stranded molecule (unlike DNA)
- Strand folds upon itself to form base pairs & secondary structures
- Structure conservation is important
- RNA sequence analysis is different from DNA sequence



Variations in RNA sequence maintain base-pairing patterns for secondary structures (conserved patterns of base-pairing)

When a nucleotide in one base changes, the base it pairs to must also change to maintain the same structure

Such variation, referred to as *covariation* indicates base pairing (A:U, G-C or G-U (Wobbel)).

Structural data used in integrative modeling

Atomic structures of parts of the system X-ray and neutron crystallography, NMR, Cryo-EM/ET, Comparative modelling and molecular docking Purification from source with gel analysis or MS, Composition and components positions Electron microscopy and tomography, gold labelling, Super resolution microscopy, FRET imaging Physical proximity Co-purification of sub complexes, native MS, genetic methods, sequence convariance, Y2H, Chromosome conformation Capture and other data, Size and shape SAXS, SANS, AUC,, atomic force microscopy, ion mobility MS Fluorescence correlation spectroscopy or anisotropy Atomic and protein distances NMR, FRET, EPR, X-link/ w/o MS..... Binding site mapping NMR, FRET, H-D/MS, mutagenesis Solvent accessiblity Footprinting methods including H-DX/MS, NMR

and chemical modifications

Small Angle X-ray Scattering

With a crystaline sample, diffraction pattern forms from the constructive interference of light passing through a crystal

With a non crystaline sample, there is no prefferential orientation of the molecules and light is scattered in all directions resulting in averaged diffusion patterns

- q = 0: Determination of I(0)
- Small q Dimension of particules Guinier, Krazky
- Larger q Shape of particules (Ab-initio analysis, fitting)







$$I(q) = I(0) \exp\left(-\frac{R_g^2 q^2}{3}\right)$$
$$q \operatorname{Rg} \le 1.0$$



 $q = 4\pi sin\theta/\lambda$

Neutron scattering (SANS)

Photons in interact with the electronic shell Neutrons interact with the nucleus

The intensity of the interaction depends on the diffusion lenght b:

The diffusion lenght is fonction of:

the atomic number (Z) for X-ray diffusion the spin of the nucleus for neutron scattering

н	-0.374 . 10 ⁻¹² cm
D	0.667 . 10 ⁻¹² cm
с	0.665 . 10 ⁻¹² cm
N	0.940 . 10 ⁻¹² cm
0	0.580 . 10 ⁻¹² cm

Weak negative signal from H, positive for D

Varies as a function of the number of H/D atoms

No sample dammage

Contrast variation



$$I(\mathbf{q}) = \frac{1}{v} \left| \int_{v} (\rho(\vec{\mathbf{r}}) - \rho_{s}) e^{i\vec{\mathbf{q}}\cdot\vec{\mathbf{r}}} d^{3}\mathbf{r} \right|^{2}$$



	X-rays	Neutrons	UV/visible
λ	# 1.5 A	# 1-20 A	# 5000 A
Energy	# keV	# meV	# 10eV
Source	Synchrotron 10 ¹² to 10 ¹⁵ ph/cm ² /s	Reactor 10 ⁴ to 10 ⁸ n/cm²/s	Laser 10 ²² ph/cm²/s
	Structure	Structure Internal dynamics	(Structure) SLS Dynamics DLS

Dynamic Light scattering



The rate of intensity fluctuation is dependent upon the size of the particle/molecule



D = kT/f

f = 6pηr

Vs = (M/N).[ν2 + $ν1^{\circ}$. δ1]

Hydrodynamic radius

Static Light Scattering



The intensity is a function of the particule's molecular weight, its concentration, shape (form factor) and of the refractive index increment of the solution.

Form factor

In general, the scattered intensity varies with Θ , the angle between the incident beam and the detector.





$$I(\theta)_{\text{scattered}} \propto Mc \left(\frac{dn}{dc}\right)^2 P(\theta)$$

The form factor, which accounts for this dependence is also a function of the wavelenght and of the dimensions of the particule.

$$P(\theta) = 1 - \frac{16\pi^2 n_0^2}{3\lambda_0^2} \sin^2\left(\frac{\theta}{2}\right) \left\langle r_g^2 \right\rangle + \dots$$

Objects smaller than the wavelength of light

The form factor does not dependent on Θ if $r_g << \lambda$ and diffusion is isotropic.

$$P(\theta) = 1 - \frac{16\pi^2 n_0^2}{3\lambda_0^2} \sin^2\left(\frac{\theta}{2}\right) \langle r_g^2 \rangle + \dots \approx 1$$

$$\frac{KC}{\Delta R_{\theta}} = \left(\frac{1}{M} + 2A_2C\right)P_{\theta} = \left(\frac{1}{M} + 2A_2C\right)$$

When the size of objects is not negligeable

The scattered intensity not only depends on M, A2 but also varies as function of θ

$$I_{\text{scattered}}(\theta) \propto R(\theta) = K^* M c P(\theta) [1 - 2A_2 M c P(\theta)]$$



Fit of data to light scattering equation provides M, A₂ and Rg.



SEC-MALLS combines Light Scattering with fractionation



MALLS: Multi-Angle Laser Light scattering

SEC and SLS

Ovalbumin (expected MM) total mass in eluting peak	MM ± SD (5 analyses)	Precision SD (%)	Accuracy
Monomer (42.8 kDa) 178ug	43.0 ± 0.7	0.2%	0.4%
Dimer (85.6 kDa) 25ug	82.7 ± 0.4	0.5%	3%
Trimer (128.4 kDa) ? 5ug	114 ± 4	3%	11%



SEC, SLS and DLS



Analytical Ultracentrifugation

There are 3 forces acting on a sedimenting particle, buoyancy, viscous drag and centrifugal force. As soon as the rotor accelerates to a constant speed, the particle reaches terminal velocity and an equilibrium between these 3 forces is established. There are several experimental conditions and sample properties that influence the sedimentation behavior:

Experimental conditions:

- 1. rotor speed
- 2. distance from the rotor center
- 3. density of the solution
- 4. viscosity of the solution
- 5. temperature

Sample properties:

- 1. molecular weight
- 2. shape
- 3. partial specific volume



Analytical ultracentrifuge

Ultracentrifuge that posseses a detection system allowing the measure of the solute concentration as a function of the distance to the rotation axis (optical density, interferrométry).

Absorbance measurement: choice of the wavelenght:

NA and protein 260 et 280 nm

ligand eg 380

peptidic bond 220: low concentrations

Limitations: detector response: inf à 0.5 OD

Avantage: large choice in the experimental conditions: buffer and ionic strength



Sedimentation velocity

Principle: the protein migrates towards the bottom of the tube; The speed of the particule is measured

Determine sedimentation coefficient s (SVEDBERG) 1 S = 10^{-13} s

Speed of sedimentation v = dr/dt per acceleration unit

 $s = dr/dt . (1/w^2r)$

 $dr/dt = s \cdot w^2 r$

dr/dt = speed of the particuler = distance of the particule to the rotation axisw = angular speed of the rotor



Séparation en fonction de S

[MS⁻¹ S²M⁻¹] [S]



Sedimentation:

Forces at Equilibrium:

Fc - Fb - Fd = 0

Fb (buoyancy)	=	ω² r m _o
Fd (viscous drag)	=	fv
Fc (centrifugal force)	=	ω²rm

Explanation:

Fb is the buoyancy force - the force required to displace the buffer surrounding the solute, and m_o is the mass of the displaced solvent.
In practice, one can measure how the sedimentation boundary moves



Diffusion impacts on the shape of sedimentation boundaries They are recorded at regular, intervals



How does the midpoint of the boundary move?



Ln (r(t)/r(t_o)) = ω^2 s(t-t^o)

s = dr/dt . (1/ ω^2 r) dr/dt = s . ω^2 r $\frac{dr}{r} = \omega^2$ sdt





Volume (occupied by the macromolecule) = Mass * Specific Volume Mass of displaced solvant = Volume * Volumic Mass

Fb (buoyancy)= $\omega^2 r m_0$ Fd (viscous drag)= f vFc (centrifugal force)= $\omega^2 r m$

Substitute the mass of the solvent, m_{o} , with the mass of the solute, m

$$m_0 = m \bar{v} \rho, \ Fb = \omega^2 r m \bar{v} \rho$$

Rearrange the force equation: Fc - Fb - Fd = 0 and substitute

$$\omega^2 r m - \omega^2 r m \bar{v} \rho = f v$$

Place molecular parameter on one side and experimental parameters on the other $\frac{m\left(1-\bar{v}\rho\right)}{f} = \frac{v}{\omega^2 r}$ $\frac{M\left(1-\bar{v}\rho\right)}{Nf} = \frac{v}{\omega^2 r} = s$

Put into molar units by multiplying with Avogadro's number, N

Important: The sedimentation coefficient depends on both M and f

The sedimentation coefficient is directly proportional to the molecular weight of the solute, and inversely proportional to the frictional coefficient of the solute. A large molecular weight increases the sedimentation speed, while an irregular shape will slow it down.

A real case



Equilibrium Sedimentation

At low speed, diffusion is not negligeable

The system reaches an equilibrium: centrigulation force = diffusion force



$$\frac{\mathrm{d}\ln(C_r)}{\mathrm{d}r^2} = \frac{M(1-r\bar{\rho})\omega^2}{2\mathrm{RT}}$$

$$Ln(C_x/C_0)=M\omega^2(1-v\rho)(x^2-x_0^2)/2RT$$

Representation of Ln(Cx/Co) as a function of x^2-xo^2 yields a linear function

The slope is only a function of M, $\omega,$ v and ρ



<u>r</u> (cm³/g) volume specific partial (hydrated) of the macromolecule ρ volumic mass of the solvant (g/cm³)

cr macromolecule concentration

r distance to the axis

A real case



FIGURE 3

Sedimentation equilibrium data. Simulated data for a reversible monomer-dimer equilibrium: (—) total, (\cdots) monomer, (--) dimer. The concentration distribution of the dimer is steeper than that of the monomer, and the relative amounts of monomer and dimer at each radial point are determined by mass-action equilibrium.

Mixture of noninteracting solutes

$$a(r) = \sum_{n} c_{n,0} \varepsilon_n d \exp\left[\frac{M_n (1 - \overline{v}_n \rho)\omega^2}{2RT} (r^2 - r_0^2)\right] + \delta$$

Self-association

$$a(r) = \sum_{n} n\varepsilon_{1} dK_{n} (c_{1,0})^{n} \exp\left[\frac{nM_{1} (1 - \overline{\nu}\rho)\omega^{2}}{2RT} (r^{2} - r_{0}^{2})\right] + \delta \quad \text{with } K_{1} = 1$$
(10)

Hetero-association

$$a(r) = c_{A,o} \varepsilon_A d \exp\left[\frac{M_A^* \omega^2}{2RT} (r^2 - r_o^2)\right] + c_{B,o} \varepsilon_B d \exp\left[\frac{M_B^* \omega^2}{2RT} (r^2 - r_o^2)\right] + c_{A,o} c_{B,o} K_{AB} (\varepsilon_A + \varepsilon_B) d \exp\left[\frac{(M_A^* + M_B^*) \omega^2}{2RT} (r^2 - r_o^2)\right] + \delta$$

Structural data used in integrative modeling

Atomic structures of parts of the system	X-ray and neutron crystallography, NMR, Cryo-EM/ET, Comparative modelling and molecular docking
Composition and components positions	Purification from source with gel analysis or MS, Electron microscopy and tomography, gold labelling, Super resolution microscopy, FRET imaging
Physical proximity	Co-purification of sub complexes, native MS, genetic methods, sequence convariance, Y2H, Chromosome conformation Capture and other data,
Size and shape	SAXS, SANS, atomic force microscopy, ion mobility MS AUC, Fluorescence correlation spectroscopy or anisotropy
Atomic and protein distances	NMR, FRET, EPR, X-link/ w/o MS
Binding site mapping	NMR, FRET, H-D/MS, mutagenesis
Solvent accessiblity	Footprinting methods including H-DX/MS, NMR and chemical modifications

X-link/MS experiments



MW 572.43; Crosslink Mass 138.07 Spacer Arm 11.4 Å

"Na⁺

"Na

Na⁺0

Expansion of the genetic code and incorporation of non-natural amino acids

Incorporate photo-activable amino acids in proteins:





pTefb = cdk9/cyclinT/Hexim + 7SK RNA

Map hexim peptides that bind and inhibit cdk9

An evolutionary conserved Hexim1 peptide binds to the Cdk9 catalytic site to inhibit P-TEFb





Kobbi et al. 2016



H/D exchange MS experiments



Förster Resonance Energy Transfer (FRET)

The mechanism of FRET involves a donor fluorophore (D) in an excited electronic state, which may transfer its excitation energy to a nearby acceptor chromophore (A)

Non-radiative process through long-range dipole-dipole interactions that results in the emission of light by the acceptor

The absorption spectrum of the acceptor must overlap fluorescence emission spectrum of the donor



Wavelength

FRET strongly depends on:

The relative orientation of the transition moments of the Donor and the Acceptor
The distance between the fluorophores

Energy transfer studies give information about

- distance between groups
- orientation of two groups and
- the refractive index between two groups



$$\kappa = \vec{D} \cdot \vec{A} - 3\left(\vec{D} \cdot \vec{R}\right)\left(\vec{A} \cdot \vec{R}\right)$$

The efficiency of transfer varies with the inverse sixth power of the distance.



 R_0 in this example was set to 40 Å. When the *E* is 50%, $R=R_0$

Distances can generally be measured between $\sim 0.5 R_0$ and $\sim 1.5 R_0$



In vivo and in vitro FRET analysis

Screening for compounds that inhibit or modulate A/B interactions

Use or fluorescent proteins fused to the proteins of interest or of fluorescent probes that are chemically coupled to the donnor and to the acceptor molecules





Structural organization of the bacterial (Thermus aquaticus) RNA polymerasepromoter open complex obtained by FRET (Mekler et al., 2002) was subsequently validated by a crystal structure (Zhang et al., 2012).

Fluorescence properties that can be measured

- spectra (environmental effects)
- fluorescence life times
- polarization (orientation and dynamics)
- excitation transfer (distances -> dynamics)
- location of fluorescence











Structural data used in integrative modeling

Composition

Atomic structures of parts of the system

3D maps, 2D images, components positions

Atomic and protein distances

Binding site mapping

Size and shape

Physical proximity

Solvent accessiblity

Purification from source with gel analysis or MS

X-ray and neutron crystallography, NMR, Cryo-EM/ET, Comparative modelling and molecular docking

Electron microscopy and tomography, gold labelling, Super resolution microscopy, FRET imaging

NMR, FRET, EPR, X-link/ w/o MS.....

NMR, FRET, H-D/MS, mutagenesis

AUC, SAS, atomic force microscopy, ion mobility Fluorescence correlation spectroscopy or anisotropy

Co-purification, native MS, genetic methods, sequence convariance, Chromosome conformation Capture and other data, Y2H

Footprinting methods including H-DX/MS, NMR and chemical modifications



Perrakis et al., 2011

END